

# Microbial community composition controls carbon flux across litter types in early phase of litter decomposition

Marie E. Kroeger,<sup>1</sup> M. Rae DeVan,<sup>1</sup> Jaron Thompson,<sup>2</sup> Renee Johansen,<sup>1,3</sup> La Verne Gallegos-Graves,<sup>1</sup> Deanna Lopez,<sup>1</sup> Andreas Runde,<sup>1</sup> Thomas Yoshida,<sup>4</sup> Brian Munsky,<sup>2,5</sup> Sanna Sevanto,<sup>6</sup>

Michaeline B. N. Albright<sup>1</sup> and John Dunbar<sup>1\*</sup>

<sup>1</sup>Bioscience Division, Los Alamos National Laboratory, Mailstop M888, Los Alamos, NM, 87545.

<sup>2</sup>Department of Chemical and Biological Engineering, Colorado State University, Fort Collins, CO, 80523.

<sup>3</sup>Manaaki Whenua – Landcare Research, Private Bag 92170, Auckland Mail Centre, Auckland, New Zealand.

<sup>4</sup>Chemical Diagnostics and Engineering, Los Alamos National Laboratory, Mailstop K484, Los Alamos, NM, 87544.

<sup>5</sup>School of Biomedical Engineering, Colorado State University, Fort Collins, CO, 80523.

<sup>6</sup>Earth and Environmental Sciences Division, Los Alamos National Laboratory, Mailstop J495, Los Alamos, NM, 87545.

## Summary

**Leaf litter decomposition is a major carbon input to soil, making it a target for increasing soil carbon storage through microbiome engineering. We expand upon previous findings to show with multiple leaf litter types that microbial composition can drive variation in carbon flow from litter decomposition and specific microbial community features are associated with synonymous patterns of carbon flow among litter types. Although plant litter type selects for different decomposer communities, within a litter type, microbial composition drives variation in the quantity of dissolved organic carbon (DOC) measured at the end of the decomposition period. Bacterial richness was negatively correlated with DOC quantity, supporting our hypothesis that across multiple litter types there are common microbial traits linked to carbon flow patterns. Variation in DOC abundance (i.e. high versus low DOC) driven by microbial composition is tentatively due to differences in bacterial**

**metabolism of labile compounds, rather than catabolism of non-labile substrates such as lignin. The temporal asynchrony of metabolic processes across litter types may be a substantial impediment to discovering more microbial features common to synonymous patterns of carbon flow among litters. Overall, our findings support the concept that carbon flow may be programmed by manipulating microbial community composition.**

## Introduction

Soil has an enormous potential to store organic carbon (Schmidt *et al.*, 2011). Consequently, management strategies to increase soil carbon storage are of keen interest as a means to reduce atmospheric CO<sub>2</sub> pollution (Batjes, 1999; Prescott, 2010; Paustian *et al.*, 2019). Because plant litter decomposition releases large quantities of carbon to the atmosphere and soil (Cotrufo *et al.*, 2015), understanding the factors that control carbon flow from plant litter is a priority. Conventional factors affecting the rate and fate of carbon flow from litter decomposition include climate and litter chemistry (Aerts, 1997). More recently, microbial composition was identified as a key factor based on observed variation in litter decomposition rates even when conventional variables were constant (Strickland *et al.*, 2009; Cleveland *et al.*, 2014; Bradford *et al.*, 2016). The failure of abiotic factors to explain large variation (e.g. up to 70-fold) in litter decomposition in natural ecosystems (Bradford *et al.*, 2014; Bradford *et al.*, 2017) emphasizes the potential impact of microbial composition as a controlling factor. Since microbially derived products from plant litter decomposition are found to be a major component of persistent soil organic carbon (Grandy and Neff, 2008; Prescott, 2010), deciphering the microbial community features that can affect soil carbon should enable new management strategies. Manipulating microbial community features on litter at the soil surface in order to influence below-ground soil organic carbon accumulation is of particular interest because microbial communities in the litter layer are easily accessible for land management intervention.

Received 14 February, 2021; accepted 2 August, 2021. \*For correspondence. E-mail [dunbar@lanl.gov](mailto:dunbar@lanl.gov); Tel. 505 667-8806.

Several microbial features have been proposed to affect litter carbon flow and influence soil carbon storage. For example, the dominance of brown versus white-rot fungi is of interest owing to an observed correlation of these groups with soil carbon abundance in coniferous forests (Bai *et al.*, 2017), possibly linked to their differences in melanin production (Siletti *et al.*, 2017). Similarly, the ratio of fungi to bacteria has correlated with soil carbon abundance in some ecosystems (Waring *et al.*, 2013; Malik *et al.*, 2016). The ratio of oligotrophic to copiotrophic bacteria (Wieder *et al.*, 2015) or specialists to generalists (Lopez-Mondejar *et al.*, 2018) may affect the quantity of litter carbon retained in microbial biomass owing to differences among these guilds in carbon use efficiency, at least when carbon supplies are in excess (Saifuddin *et al.*, 2019). Microbial species richness is an emergent community property that has been examined as a factor (Nielsen *et al.*, 2011; Louis *et al.*, 2016) and is of particular interest because it is easily measured, well documented and amenable to manipulation.

Several recent microcosm studies have shown a strong link between microbial taxon richness and litter decomposition dynamics (CO<sub>2</sub> efflux and/or litter mass loss) (Juarez *et al.*, 2013; Maron *et al.*, 2018; Wagg *et al.*, 2019). These studies reduced richness up to five orders of magnitude by extreme dilution or by size fractionation of microbial communities prior to inoculation into microcosms (Juarez *et al.*, 2013, Maron *et al.*, 2018, Wagg *et al.*, 2019). Using an alternative approach that exploits naturally occurring variation in soil microbial community composition, we found a strong correlation between bacterial richness and the quantity and quality of dissolved organic carbon (DOC) from pine litter decomposition (Albright *et al.*, 2020a; Albright *et al.*, 2020b). The link between microbial richness and DOC abundance illuminates a means by which features of surface litter decomposer communities may affect organic matter abundance in the deep subsurface. DOC is a critical link between ephemeral carbon at the soil surface and persistence in the deep subsurface (e.g. 1 m belowground). Changes in the quantity or quality of DOC transported to the deep subsurface can alter the binding of organic carbon to mineral surfaces (Kaiser and Kalbitz, 2012; Newcomb *et al.*, 2017), which is key to carbon stabilization at the millennial-scale (Schöning and Kögel-Knabner, 2006; Rumpel and Kögel-Knabner, 2010).

To further explore the relationship between microbial richness and litter carbon flow mentioned above in Albright *et al.* (2020a) and Albright *et al.* (2020b), we tested the relationship with additional litter types inoculated with microbial communities from 100 different soils. Litter chemistry strongly affects both microbial community

and DOC composition (Cleveland *et al.*, 2004; Don and Kalbitz, 2005; Strickland *et al.*, 2009; Bray *et al.*, 2012; Wickings *et al.*, 2012). Consequently, the relationship between microbial richness and DOC abundance discovered in one litter type (e.g. pine) may not be evident in other litter types due to inherent differences in litter chemistry.

Using our prior pine litter decomposition study as a baseline (Albright *et al.*, 2020a; Albright *et al.*, 2020b), we applied the same approach with oak litter (*Quercus gambelii*) and a grass litter mix (one to one mix of *Hilaria jamesii* and *Stipa hymenoides*) that is common in the southwestern United States. A random subset of 100 soils out of the 206 soils from the pine study (Albright *et al.*, 2020b) were used to inoculate oak and grass litters. We measured carbon flow as cumulative CO<sub>2</sub> production throughout the 44-day experiment and DOC abundance at the end of the experiment. In addition to examining microbial richness using 16S and 28S rRNA amplicon sequencing, we asked if synonymous patterns of carbon flow (represented by the high versus low DOC abundance groups) among litter types shared a common metabolic signature. Metabolic signatures were represented as SEED. Subsystem functions documented in metatranscriptomes of six litter decomposer communities from each DOC group from the pine, oak and grass litter experiments. We hypothesized that (i) plant litter type selects for different decomposer communities, but within a litter type, microbial richness varies with DOC abundance, and (ii) synonymous carbon flow patterns among litter types share common metabolic features.

## Materials and methods

### *Initial soil collection for microbial inoculum*

Soil samples were collected from 208 locations throughout the southwestern United States between February and April 2015 as described previously (Albright *et al.*, 2020a) (Supplemental Table 1). Samples were typically collected at locations approximately 80 km apart, at least 15 m from roadways, from the top 3 cm of the soil surface after removal of surface litter. Samples were collected in sterile 50-ml screw-cap tubes and immediately stored on ice. The location of each sample was recorded by GPS and photographed to facilitate description of the major ecosystem types from which samples were obtained (Supplemental Table 1).

### *Microcosm construction and CO<sub>2</sub> sampling*

As described in Albright *et al.* (2020a), microcosms were constructed using 125 ml serum bottles, each containing approximately 5 g of sand and an initial dose of 0.02 g of

dried leaf litter that was milled in a Wiley Mill (Thomas Scientific, Swedesboro, NJ, USA). Litters used were either pine (*Pinus ponderosa*), oak (*Quercus gambelii*), or a grass mix (one to one mix of *Hilaria jamesii* and *Stipa hymenoides*). Oak and pine litters were collected in Los Alamos, NM and *Hilaria jamesii* and *Stipa hymenoides* litter were collected in Canyonlands, UT. The microcosms were sterilized by autoclaving three times for 1 h, with at least an 8-h resting interval between each autoclave cycle. Microbial community inoculum was extracted from each soil sample with the pine study using 206 soils, while the grass and oak study used a randomly selected subset of 100 soils each ( $n = 206$  for pine,  $n = 100$  for oak,  $n = 100$  for grass) (Supplemental Table 1). The microbial inoculum was extracted by suspending 1 g of soil in 9 ml of phosphate-buffered saline (PBS), then generating a 1000-fold dilution in PBS amended with  $\text{NH}_4\text{NO}_3$  at  $4.8 \text{ mg ml}^{-1}$  on the day of microcosm inoculation. The nitrogen addition was consistent with application rates used in field studies investigating the impacts of anthropogenic N deposition (Mueller *et al.*, 2015). The pine experiment used three microcosms replicates per soil sample ( $n = 618$  total) while the oak and grass experiments used two microcosms per soil sample ( $n = 200$  oak,  $n = 200$  grass). Each microcosm received 1.3 ml of inoculum, pipetted directly onto 0.02 g of litter. Negative control microcosms, used to confirm the efficacy of sterilization, received the same quantities of litter, PBS and  $\text{NH}_4\text{NO}_3$ , but no microbial communities. The microcosms were incubated in two phases. In the first phase, sealed microcosms were incubated at  $25^\circ\text{C}$  in the dark for 14 days to allow physiological equilibration of the diverse inocula on the leaf litter. To prevent oxygen depletion and excess  $\text{CO}_2$  accumulation, the headspace of each serum bottle was evacuated using a vacuum pump and replaced with sterile-filtered air on days 3 and 7. On day 14, the second phase was started by adding a further 0.1 g of litter sterilized by three rounds of autoclaving to each microcosm and sealing the microcosms with Teflon-lined crimp caps. The microcosms were incubated at  $25^\circ\text{C}$  in the dark for a further 30 days. During this time,  $\text{CO}_2$  was measured by gas chromatography using an Agilent Technologies 490 Micro GC (Santa Clara, CA, USA) on days 2, 5, 9, 16, 23 and 30. After each measurement, the headspace air was evacuated with a vacuum pump and replaced with sterile-filtered air. Cumulative  $\text{CO}_2$  was calculated by taking the sum of the  $\text{CO}_2$  quantities recorded at each measurement point.

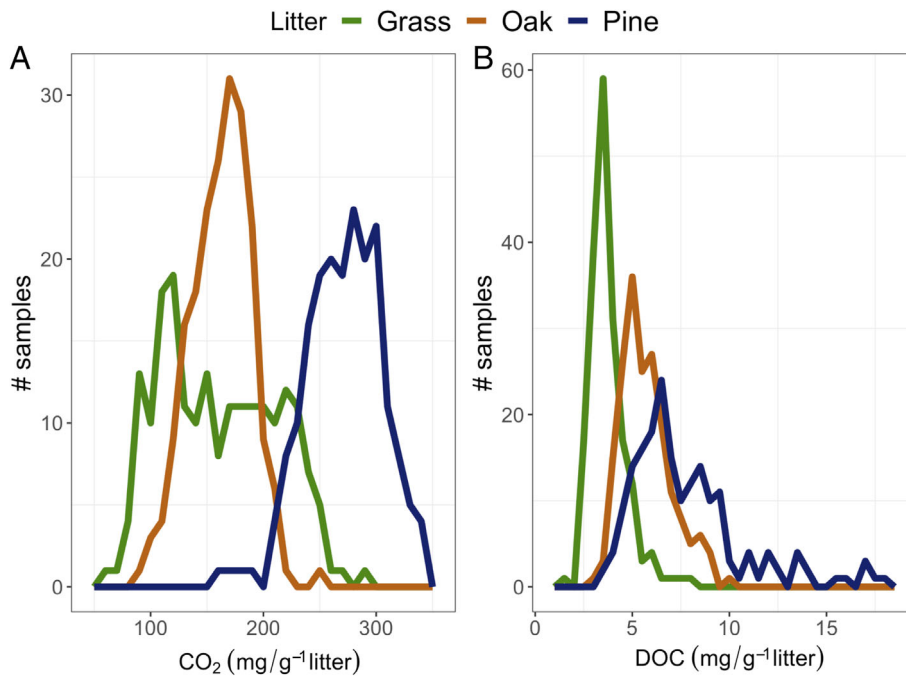
#### *Dissolved organic carbon and litter community sampling*

After the 44-day (total) incubation, microcosms were destructively sampled to measure DOC and community

composition as described in detail previously (Albright *et al.*, 2020a). The DOC extracts were analysed for organic carbon using a Shimadzu TOC-LCSH Carbon Analyser (Shimadzu Scientific Instruments, Columbia, MD, USA) employing a combustion catalytic oxidation method and non-dispersive infrared detector. Following DOC sampling, material (sand and litter) from each microcosm was frozen at  $-80^\circ\text{C}$  for DNA extraction.

#### *Bacterial and fungal community taxonomic profiling*

Since the goal of this study is to identify the direct impact of microbial community composition on carbon flow, we selected samples within each litter type from the extremes of the DOC spectrum. The mean DOC quantity for each set of microbial community replicates was calculated and the replicate microcosms with the highest and lowest mean DOC quantities, representing the two tails of the DOC distribution (Fig. 1B), were selected for DNA extraction and sequencing. For pine,  $n = 192$  from each tail (64 soils  $\times$  3 replicates), for grass and oak,  $n = 50$  from each tail (25 soils  $\times$  2 replicates). DNA extractions were performed using a DNeasy PowerSoil 96-well plate DNA extraction kit (Qiagen, Hilden, Germany). The standard protocol was used with the following two exceptions: (i) 0.3 g of material was used per extraction; (ii) bead beating was conducted using a SPEX Certiprep 2000 Geno/Grinder (SPEX SamplePrep, Metuchen, NJ, USA) for 3 min at 1900 strokes min. DNA samples were quantified with an Invitrogen Quant-iT™ ds DNA Assay Kit (Thermo Fisher Scientific, Eugene, OR, USA) on a BioTek Synergy HI Hybrid Reader (BioTek Instruments, Winooski, VT, USA). PCR templates were prepared by diluting an aliquot of each DNA stock in sterile water to  $1 \text{ ng } \mu\text{l}^{-1}$ . The bacterial (and archaeal) 16S rRNA gene (V3–V4 region) was amplified using primers 515f-R806 (Bates *et al.*, 2010). Hereafter, archaeal sequences were analysed with bacterial sequences. The fungal 28S rRNA gene (D2 hypervariable region) was amplified using the LR22R primer (Mueller *et al.*, 2016) and the reverse LR3 primer (Vilgalys and Hester, 1990). Preparation for Illumina high-throughput sequencing was undertaken using a two-step approach, similar to that performed by Mueller *et al.* (2015), with Phusion Hot Start II High Fidelity DNA polymerase (Thermo Fisher Scientific, Vilnius, Lithuania). In the first PCR, unique 6 bp barcodes were inserted into the forward and reverse primer in a combinatorial approach over 22 cycles with an annealing temperature of  $60^\circ\text{C}$  (Gloor *et al.*, 2010). The second PCR added Illumina-specific sequences over 10 cycles with an annealing temperature of  $65^\circ\text{C}$  (Illumina, San Diego, CA, USA). Amplicons were cleaned using a Mo Bio Ultra-Clean PCR clean-up kit (Carlsbad, CA, USA), quantified using the same procedure as for the extracted DNA, and



**Fig. 1.** Ecosystem functioning varied by litter type. Distribution of CO<sub>2</sub> (A) and DOC (B) across samples and litter types (grass = green, oak = orange, pine = blue).

then pooled at a concentration of 10 ng each. The pooled samples were further cleaned and concentrated using the Mo Bio UltraClean PCR clean-up kit. All clean ups were undertaken as per the manufacturer's instructions with the following modifications: binding buffer amount was reduced from 5× to 3× sample volume, and final elutions were performed in 50 μl Elution Buffer. A bio-analyzer was used to assess DNA quality, concentration was verified using qPCR, and paired-end 300 bp reads were obtained using an Illumina MiSeq sequencer at Los Alamos National Laboratory.

Bacterial and fungal sequences were processed following the UPARSE pipeline in usearch v11 (Edgar, 2013). Paired ends were merged with a 90% minimum similarity, truncating tails at the first base pair quality below 30, and a minimum merge length of 150 bp. Primers were stripped using fastx\_truncate. Sequences were quality filtered with a maximum expected error of 1.0 and globally trimming fungal reads at 250 base pairs and bacterial reads at 240 base pairs. After dereplication, singletons were removed. OTUs were picked using cluster\_otus for both fungal and bacterial sequences, which simultaneously identifies and removes chimeric sequences. OTU tables were generated using 97% similarity. Bacterial and fungal OTUs were classified using the Ribosomal Database Project classifier (Wang *et al.*, 2007). The OTUs that were not classified as bacteria or fungi with 100% confidence were removed from the dataset. Bacterial OTUs also had to have a phylum classification confidence level of at least 80% to remain in the dataset, as per Wang *et al.* (2007), which has become a convention

for the V3–V4 region. For pine, following quality control and classification, 9 579 215 sequences from 345 microcosms were obtained for bacteria and 12 986 765 sequences from 377 microcosms were obtained for fungi. These are sorted into 2913 OTUs for bacteria (an average of 281 per microcosm, SE = 8) and 829 OTUs for fungi (an average of 45 per microcosm, SE = 1). For oak, 3 208 923 sequences from 99 microcosms were obtained for bacteria and 5 508 105 sequences from 100 microcosms were obtained for fungi. These are sorted into 2491 OTUs for bacteria (an average of 359 per microcosm, SE = 12) and 489 OTUs for fungi (an average of 45 per microcosm, SE = 1.5). For grass, 1 047 697 sequences from 99 microcosms were obtained for bacteria and 1 954 949 sequences from 100 microcosms were obtained for fungi. These are sorted into 2214 OTUs for bacteria (an average of 230 per microcosm, SE = 11) and for 296 OTUs for fungi (an average of 18 per microcosm, SE = 0.75). Uneven sample sizes between bacteria and fungi for various litter types were due to quality control during processing. Sequences were only retained if they passed a filter with a maximum expected error of 1 and were at least 240 bp long for bacteria, and 250 bp for fungi, and for further statistical tests, samples had to contain at least 1000 reads, thus, in some cases a sample met this threshold for fungi but not bacteria and vice versa.

All analyses were completed in R v 4.0.2 (R Core Team, 2020). The 'phyloseq' package was used to rarefy samples with replacement to 1000 and determine

bacterial and fungal richness, as well as to complete NMDS analyses on relativized data using Bray–Curtis dissimilarity matrices and three axes (number of axes was determined using stress plots created in ‘vegan’ v 2.5-6 (Oksanen *et al.*, 2019). Beta dispersion of DOC groups and litter type were completed in ‘vegan’ v 2.5-6 (Oksanen *et al.*, 2019) using the group centroid and post hoc Tukey’s honest significant difference was performed in ‘stats’ (R Core Team, 2020). Variance in DOC abundance explained by microbial community composition was completed in vegan v2.5-6 using adonis. Richness correlation analyses with DOC abundance were done using ‘ggpubr’ v 0.4.0 (Kassambara, 2020) and the `stat_cor` command with method set to Pearson.

### Machine learning

The Random Forest Indicator species analysis Neural Network (RFINN) platform (Thompson *et al.*, 2019) was used to identify subsets of genera that were found to be important features for prediction of either CO<sub>2</sub> or DOC. The data were pre-processed with genera abundances computed as the average abundance of all OTUs within each genus. Additionally, only genera that were present among all three litter types were considered for feature selection because our goal was to identify a common microbial feature across the litter types. RFINN applies an ensemble of machine learning and standard statistical methods to identify a small consensus set of taxa whose relative abundances are robustly and significantly linked with a target variable. Feature selection using RFINN was applied separately to pine, oak and grass datasets to identify bacterial and fungal genera that were highly predictive of DOC or CO<sub>2</sub> in those data sets. For each feature selected by RFINN, the direction of the correlation with DOC or CO<sub>2</sub> was determined by indicator species analysis and the neural network. In all cases, the direction of the correlation determined by the NN and indicator species analysis agreed. For each litter type, the data set was partitioned into 15 folds of training and testing sets, which resulted in 15 iterations of feature selection results. The final set of selected features included all genera present in at least one of the 15 sets. This comprehensive analysis resulted in 12 sets of selected genera, one for each of the three litter types using either bacterial or fungal OTUs to predict either DOC or CO<sub>2</sub>.

A logistic regression model implemented using Scikit-Learn in Python (Pedregosa *et al.*, 2011) was used to predict DOC levels (high or low) using total biomass, bacterial richness and fungal richness as model features. To determine prediction performance on held-out data, k-fold cross-validation was used to determine out-of-fold predictions over 10-folds of the data set. The out-of-fold sets did not contain replicates of samples present in the

corresponding training set. The area under the receiver operating characteristic curve was used to evaluate relative prediction performance and a Z-test of two proportions was used to determine prediction significance by comparing the proportion of correct predictions to the most frequently occurring class. The Wald test was used to evaluate feature significance using a final logistic regression model fit to the entire data set using the `statsmodels` module in Python (Seabold and Perktold, 2010).

### Metatranscriptome sequencing and analysis

RNA was extracted from a subset of high ( $n = 6$ ) and low ( $n = 6$ ) microcosm samples from each plant litter ( $n = 3$ ) for a total of 36 samples representing the most extreme members of each DOC distribution. Extractions were performed as described previously (Albright *et al.*, 2020b). Briefly, RNA/DNA was coextracted from 1 g of sample (a mixture of sand and litter) as detailed in Hesse *et al.* (2015) followed by RNA clean-up with Ambion Turbo DNase kit (Ambion, Austin, TX, USA) and purification with the Qiagen RNeasy Mini kit. rRNA was removed using a combination of the Illumina RiboZero H/M/R and Bacteria kits. Libraries were prepared using an Illumina ScriptSeq v2 library preparation kit. Library validation was performed with a Qubit dsDNA HS assay, BioAnalyzer DNA high-sensitivity assay (Agilent Technologies) and library quantification kit (Roche, Basel, Switzerland). Libraries were run on a NextSeq 500 system (high-output v2 kit for 300-cycle sequencing).

Metatranscriptomes were processed and annotated on MG-RAST following the default pipeline (Meyer *et al.*, 2008; Keegan *et al.*, 2016). For the metatranscriptomes, an average of 1.45 million sequences per sample passed quality control. Of these sequences,  $20.5 \pm 12.6\%$  were known predicted proteins,  $27.8 \pm 10.0\%$  unknown predicted proteins and  $51.7 \pm 20.8\%$  ribosomal genes. From the MG-RAST server, RefSeq (O’Leary *et al.*, 2016) and Subsystems (Kanehisa *et al.*, 2016) data were downloaded for analysis of taxonomy and function assignments respectively (default parameters,  $E$ -value  $< 10^{-5}$ , identity  $> 60\%$ ). Only reads annotated as bacteria, virus, fungi, or archaea were analysed. Taxonomic and functional annotations for each sample were analysed in two different ways to determine differences between low and high DOC communities. For within plant litter types, samples were rarefied and log-transformed. Pine samples were rarefied to 98 500 and 18 800, oak samples were rarefied to 224 000 and 64 500 and grass samples were rarefied to 164 500 and 20 700 for taxonomic and functional annotations respectively. The non-parametric Kruskal–Wallis H test was then used in R v3.6.0 (R Core Team, 2019) to

determine significant differences between low and high DOC groups. Diversity metrics were determined using rarefied, log-transformed data with significant differences between low and high DOC groups determined by Pearson's product-moment correlation values in the 'stats' package (R Core Team, 2019). To determine high and low DOC traits regardless of litter type, we used the DESeq2 package in R with the design =  $\sim$ DOC + Litter (Love *et al.*, 2014). Bray-Curtis and Jaccard dissimilarity matrices were computed using DESeq2 normalized counts and a permutational multivariate analysis of variance was run to look for correspondence between taxonomic and functional gene expression and DOC group. To quantify the relative variability within each DOC group (i.e. high and low), we measured the average distance to the centroid within each DOC group using a test for homogeneity of dispersion [*vegan* v 2.4-3 package, R (Oksanen *et al.*, 2017)]. Plots were made in R using 'ggplot2' (v. 3.3.2) and 'ggpubr' (v. 0.4.0) packages (Wickham, 2016; R Core Team, 2019; Kassambara, 2020).

#### Litter chemistry

All plant litter chemical analyses were performed by the Colorado State University Soil, Water, and Plant Testing Laboratory (Fort Collins, CO; [www.soiltestinglab.colostate.edu](http://www.soiltestinglab.colostate.edu)). Milled plant litter (pine, oak and grass) was analysed for pH, EC, organic matter, NO<sub>3</sub>-N, P, K, Zn, Cu, Fe and Mn. The protocol for each laboratory analysis was as follows: Litter pH was measured with a glass electrode and EC was determined with an EC meter in a suspension of litter to water ratio of 1:1. Litter organic matter was determined using the Walkley-Black procedure by digesting soil with 1 N K<sub>2</sub>Cr<sub>2</sub>O<sub>7</sub> and H<sub>2</sub>SO<sub>4</sub>, adding distilled water, filtering through Whatman 1 paper into a Spect 20 tube, and reading on Spectronic 20 at 610 nm adjusted to 100% transmittance with the blank (Walkely and Black, 1934; Walkely, 1947). For the determination of NO<sub>3</sub>-N, P, K, Cu, Fe, Mn and Zn, litter was extracted with AB-DTPA solution and the extracted aliquot was stored in clean plastic bottles (Soltanpour and Workman, 1979). For estimation of NO<sub>3</sub>-N, absorbance of light in the extract was read at 540-nm wavelength on a spectrophotometer (Kamphake *et al.*, 1967). The concentration of P was measured at 880-nm wavelength using a spectrophotometer. The potassium in soil extract was determined directly with a flame photometer or by an atomic absorption using a potassium hollow cathode lamp. Micro-nutrients Zn, Cu, Fe and Mn were determined directly from AB-DTPA extract by atomic absorption. Total carbon and nitrogen were quantified on a LECO TruSpec CN analyser (LECO Corporation, St. Joseph, MI, USA).

#### Data availability

Amplicon sequence data have been deposited in the NCBI Sequence Read Archive (SRP151768). Metatranscriptomes are publicly available on MG-RAST under project IDs mgp90738, mgp90741 and mgp90765. All code and data files for logistic regression modelling and machine learning analyses are available at [https://github.com/MunskyGroup/Kroeger\\_et\\_al\\_2021](https://github.com/MunskyGroup/Kroeger_et_al_2021). All other data including OTU tables are available upon request.

## Results

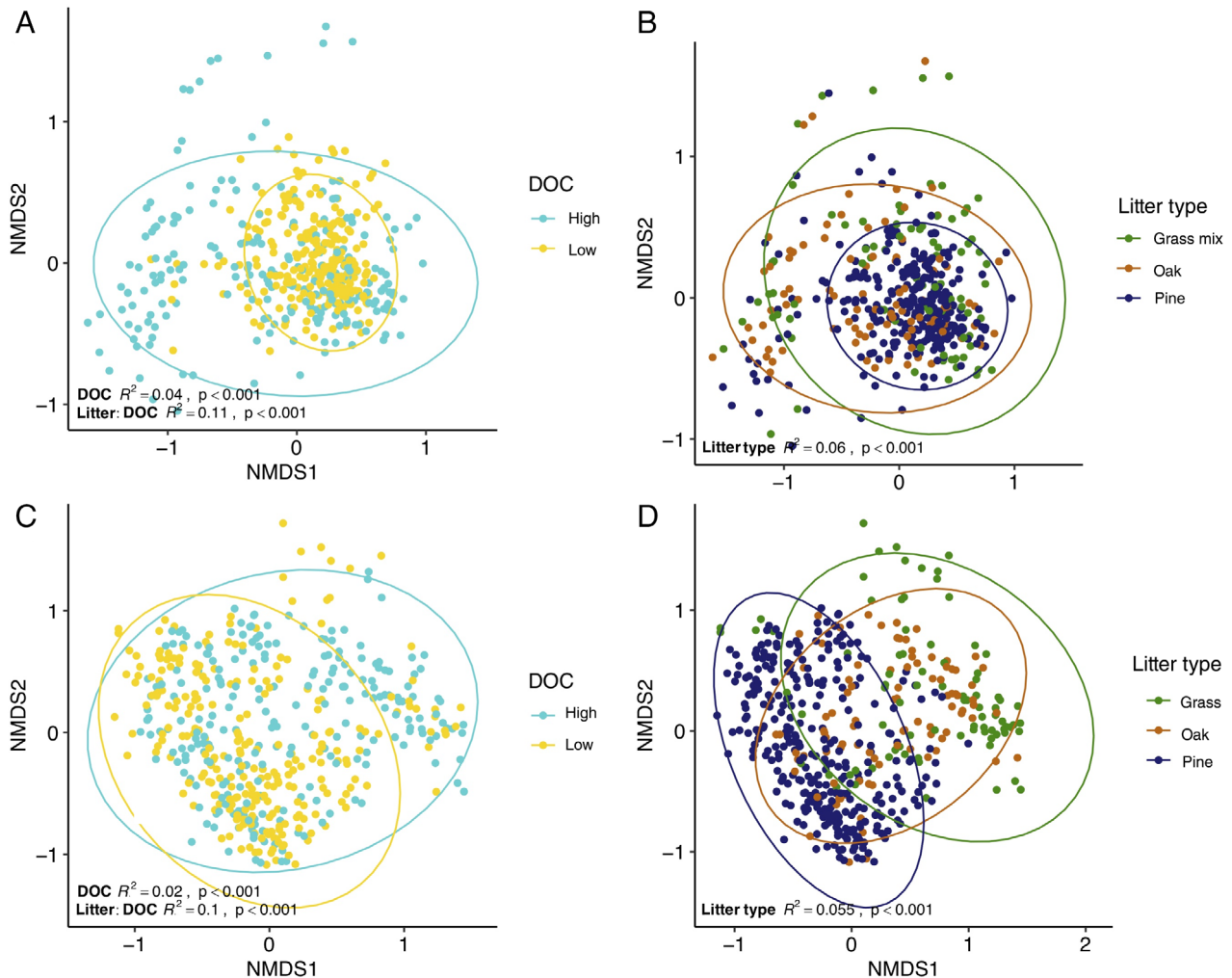
### CO<sub>2</sub> production and DOC abundance varied by litter type

Initial litter type affected CO<sub>2</sub> production and DOC abundance, as expected. At the end of the experiment, pine had a significantly higher cumulative CO<sub>2</sub> concentration compared to oak ( $p < 0.001$ ) and grass ( $p < 0.001$ ) litter types (Supplemental Fig. 1). All litters had significantly different DOC abundances ( $p < 0.001$ ) with pine having the highest mean (7.75 mg g<sup>-1</sup> litter) followed by oak (5.71) and grass (3.76) (Fig. 1, Supplemental Fig. 2). Each litter type exhibited a significant negative correlation between DOC abundance and CO<sub>2</sub> production (pine:  $R^2 = 0.16$ ,  $p < 0.001$ , oak:  $R^2 = 0.15$ ,  $p < 0.001$ , grass:  $R^2 = 0.024$ ,  $p = 0.028$ ) (Supplemental Fig. 3). The initial C:N ratio in the litter types was very similar for pine and oak (59.53 and 54.23 respectively) but much higher in the grass (92.06). Additionally, grass litter had a higher pH (5.9) than both pine (3.9) and oak (3.9). Oak and grass litter had similar levels of phosphorus and zinc (oak P = 243 ppm, grass P = 223 ppm, oak Zn = 11.6 ppm, grass Zn = 11.4 ppm). Grass had the highest abundance of iron (22.3 ppm), copper (78.2 ppm) and potassium (2855 ppm), while oak had the greatest amount of manganese (151 ppm), and pine had the greatest abundance of zinc (16.4 ppm) and phosphorus (380 ppm).

### Total microbial community composition explained significant variance in DOC groups

Microbial community composition determined by amplicon sequencing explained 4% and 2% (bacteria:  $p < 0.001$ , fungi:  $p < 0.001$ ) of the variance in DOC group (Fig. 2A and C). Across litter types, we found microbial composition explained 6% and 5.5% variance (bacteria:  $p < 0.001$ , fungi:  $p < 0.001$ ) for bacteria and fungi (Fig. 2B and D). When community composition was constrained by DOC and litter type, it explained 11% and 10% (bacteria:  $p < 0.001$ , fungi:  $p < 0.001$ ) of the variance for bacteria and fungi respectively. Overall, we observed significantly greater dispersion among high compared to low DOC communities ( $p < 0.001$ ) for both





**Fig. 2.** Variation in bacterial (A, B) and fungal (C, D) community composition based on DOC group (low, high) and litter type (grass, oak, pine). The plot shows the first two axes of a 3D scaling plot (stress: A and B = 0.198; C and D = 0.204) visualizing Bray–Curtis dissimilarities among communities. The  $p$ -values for adonis tests were calculated as implemented in vegan.

bacteria and fungi. Between litter communities, pine had significantly less bacterial dispersion than either oak (fungi:  $p = 0.051$ , bacteria:  $p < 0.001$ ) or grass (fungi:  $p = 0.059$ , bacteria:  $p < 0.001$ ).

#### Total microbial community features predict DOC abundance

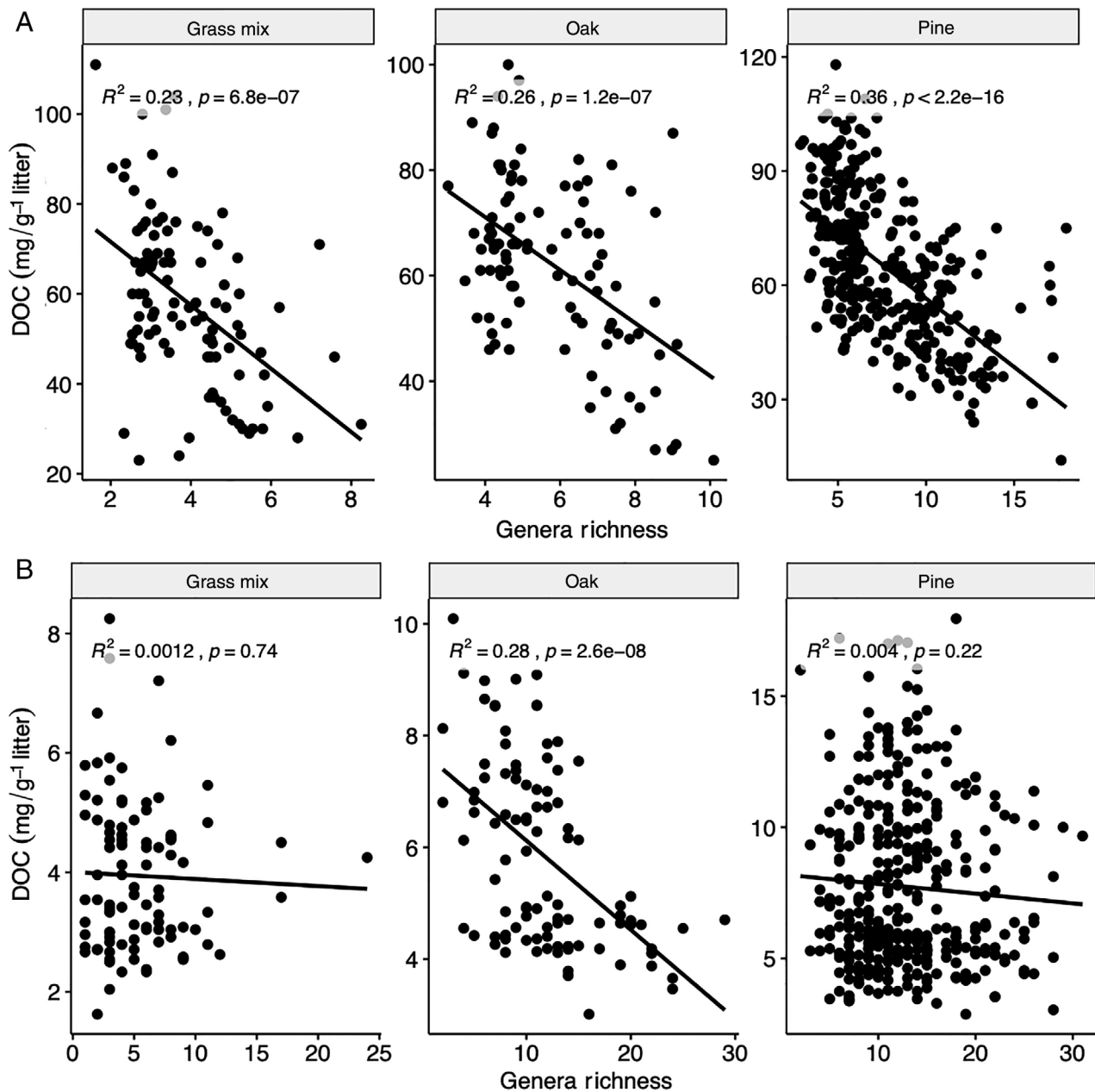
Bacterial richness was negatively correlated with DOC abundance across all litter types ( $p < 0.01$ ; grass:  $R^2 = 0.23$ , oak:  $R^2 = 0.26$ , pine:  $R^2 = 0.36$ ) (Fig. 3A). Fungal richness was also negatively correlated with DOC abundance for oak litter ( $R^2 = 0.28$ ,  $p = 2.6e-08$ ) but not grass ( $R^2 = 0.0012$ ,  $p = 0.74$ ) or pine litter ( $R^2 = 0.004$ ,  $p = 0.22$ ), (Fig. 3B). Additionally, logistic regression models predicted DOC abundance significantly better than chance for each litter type using total biomass, bacterial richness and fungal richness as community features

(grass:  $p = 2.67e-02$ , oak:  $p = 4.33e-06$ , pine:  $p = 1.65e-4$ ) (Supplemental Figs 4–6).

Out of over 500 bacterial and fungal amplicon profiles, RFINN identified only a single bacterial genus, *Microvirga*, and one fungal genus, *Plectosphaerella*, that was universal to all three litter types to predict DOC abundance. No fungal genera were common across all litter types to predict  $CO_2$  abundance; however, the bacteria genus *Rhizobium* was found to predict  $CO_2$  abundance in pine, oak and grass (Fig. 4; Supplemental Table 2).

#### Active microbial community explained significant variance across litter types and DOC groups

Litter type explained more of the variance in community composition ( $R^2 = 0.1249$ ,  $p = 0.001$ ) than DOC alone ( $R^2 = 0.0673$ ,  $p = 0.003$ ), but DOC constrained by litter type explained the most variance ( $R^2 = 0.3109$ ,  $p = 0.0001$ )



**Fig. 3.** Linear regression of bacterial (A) and fungal (B) genera richness with DOC abundance by litter type (grass, oak, pine). Pearson's correlation  $R^2$  and  $p$ -values are reported for each.

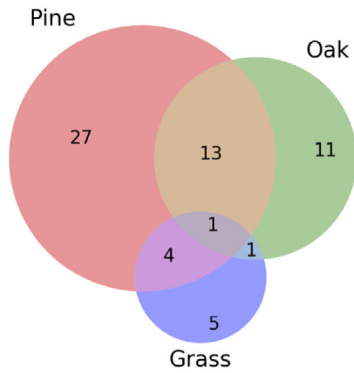
(Fig. 5A) like the amplicon data. However, we did not observe a significant difference in dispersion between litter types ( $p = 0.142$ ) or DOC group ( $p = 0.124$ ) like we did in the amplicon data. The functional composition also significantly differed between litter type ( $R^2 = 0.158$ ,  $p = 0.001$ ) and DOC group ( $R^2 = 0.0658$ ,  $p = 0.006$ ) with DOC constrained by litter type again explaining the most variance ( $R^2 = 0.2823$ ,  $p = 0.0001$ ) (Fig. 5B). Dispersion was significantly greater in high DOC ( $p = 0.003$ ) communities than low DOC for functional composition.

#### *Active community richness only decreased with DOC abundance in pine litter*

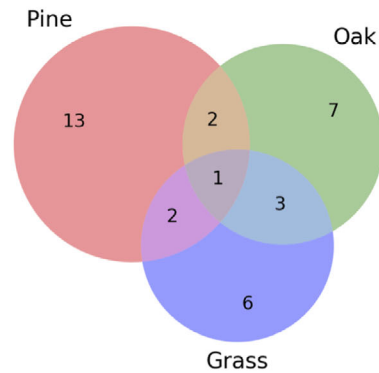
In contrast to amplicon sequence data of more than 500 samples, when metatranscriptome data from each litter type were analysed individually (12 samples each), the only significant correlation between taxonomic diversity and DOC abundance occurred with pine litter samples ( $R^2 = -0.781$ ,  $p = 0.003$ ; Supplemental Table 3). Similarly, the mean taxonomic richness was significantly different between DOC groups (high versus low) only for



Overlap in bacterial genera between litter types (DOC)



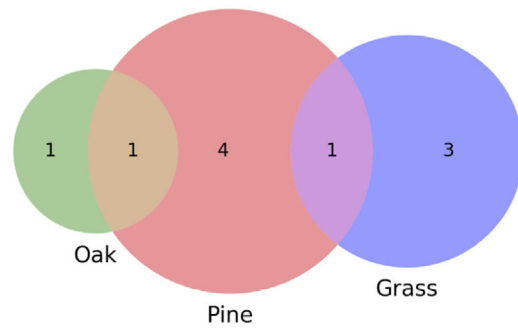
Overlap in bacterial genera between litter types (CO<sub>2</sub>)



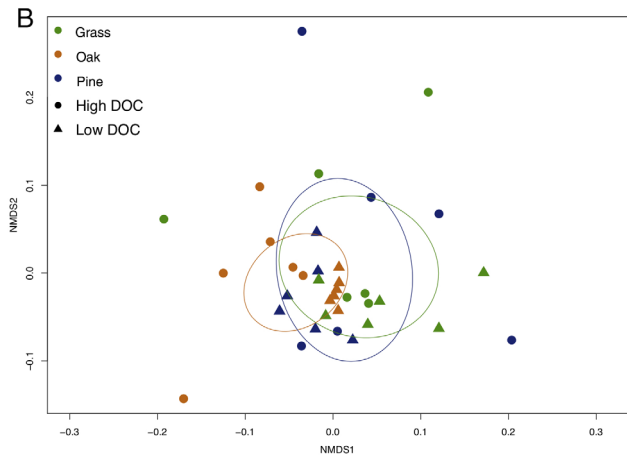
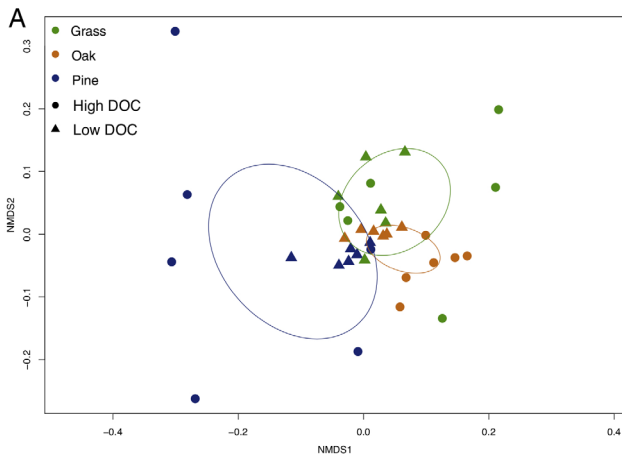
Overlap in fungal genera between litter types (DOC)



Overlap in fungal genera between litter types (CO<sub>2</sub>)



**Fig. 4.** Overlap across litter types (pine, oak, grass) of bacterial (A and B) and fungal (C and D) genera driving carbon flow (DOC or CO<sub>2</sub> abundance), down-selected by the RFINN machine learning software.

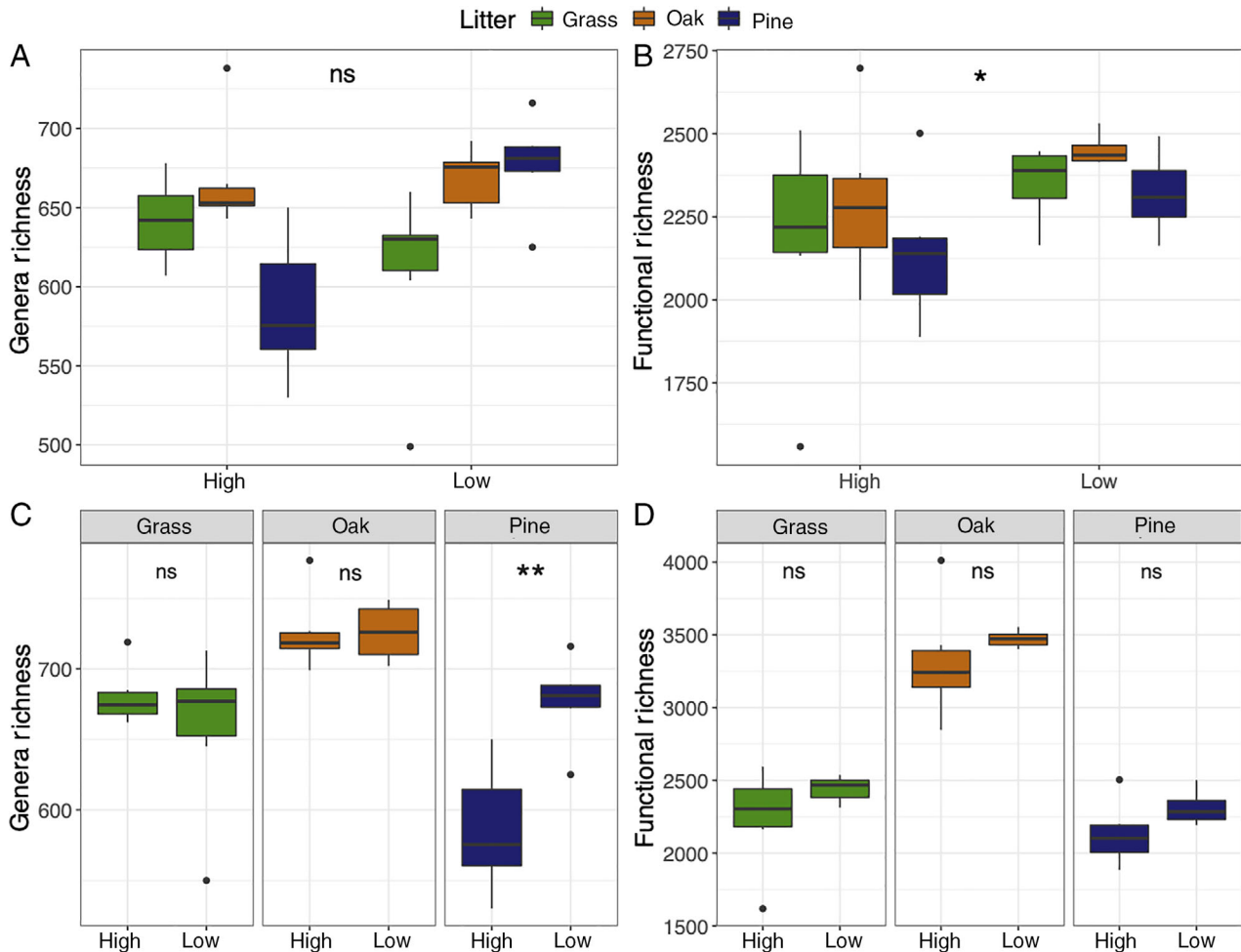


**Fig. 5.** Variation in microbial community composition at the genus level (A) and functional subsystem level 3 (B) using DESeq2 normalized counts. Bray–Curtis dissimilarities were visualized by non-metric dimensional scaling.

the 12 pine samples ( $p = 0.002$ ) (Fig. 6C), in contrast to amplicon sequence data. There was no significant difference in functional richness within any litter type (Fig. 6D).

However, when samples from all litter types (36 total) were analysed together, taxonomic and functional

richness inversely correlated with DOC abundance, which was driven exclusively by the pine samples, as seen in the amplicon diversity analyses of more than 500 samples (taxonomic:  $R^2 = -0.421$ ,  $p = 0.011$ ; functional:  $R^2 = -0.344$ ,  $p = 0.040$ ; Supplemental Table 3).



**Fig. 6.** Taxonomic richness at the genus level (A) and functional richness (B) when all samples are analysed together for low and high DOC groups. The taxonomic richness at the genus level (C) and functional richness (D) when each litter type is analysed separately for low and high DOC groups. When litter types were analysed together, samples were rarefied 18 800 annotations per sample for functional diversity and 98 500 annotations for taxonomic diversity. When litter types were analysed separately, pine litter was rarefied to 18 800 and 98 500, oak litter was rarefied to 64 600 and 224 000, and grass litter was rarefied to 20 700 and 164 000 for functional and taxonomic diversity respectively. Grass mix litter = green, Oak litter = orange, Pine litter = blue.

When DOC groups were compared (high versus low), there was no significant difference in taxonomic richness ( $p = 0.12$ ) between groups (Fig. 6A). Functional richness, on the other hand, was significantly different between DOC groups ( $p = 0.016$ ) (Fig. 6B).

#### Active microbial community composition changes across litter types and DOC groups

Proteobacteria and Actinobacteria were the dominant active bacterial phyla and Ascomycota was the dominant active fungi in all litter types and both DOC groups (Supplemental Fig. 7). When each plant litter was analysed separately, we found that no phyla significantly changed abundance in grass litter between high and low

DOC groups. In contrast, the abundance of 15 and 9 phyla changed in pine and oak litter respectively (Supplemental Table 4), and three of those phyla overlapped: Dictyoglomi (pine:  $p = 0.015$ , oak:  $p = 0.013$ ), Verrucomicrobia (pine:  $p = 0.025$ , oak:  $p = 0.010$ ) and Chytridiomycota (pine:  $p = 0.020$ , oak:  $p = 0.025$ ). For oak, the relative abundance of Verrucomicrobia and Chytridiomycota increased in low DOC while Dictyoglomi increased in high DOC. However, in pine all three phyla increased in the low DOC group. At the genus level, 154, 147 and 26 genera significantly changed abundance between high and low DOC for pine, oak and grass respectively. Of these genera, 19 significantly increased in the low DOC group for both oak and pine litter and five in the high DOC group for both oak and grass litter. Although six genera were shared between grass and

pine litter, none of these genera significantly increased abundance in the same DOC group between the two litter types (Supplemental Table 5).

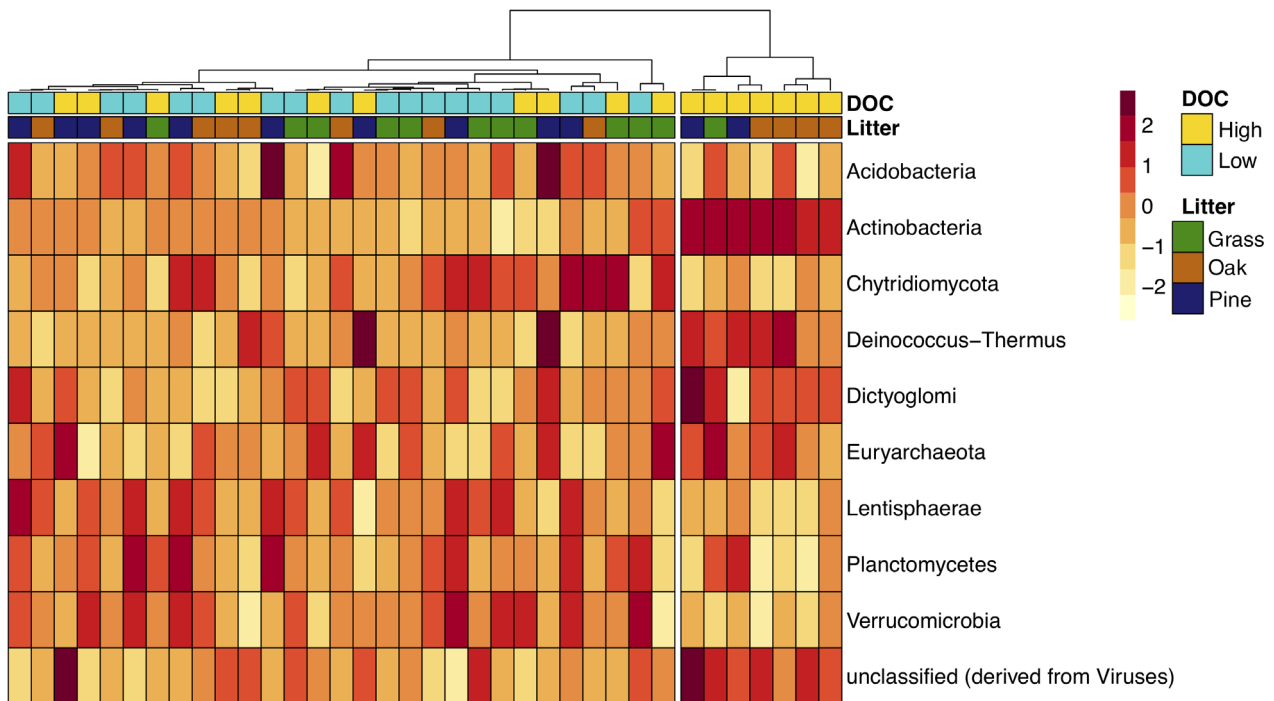
Next, to better understand how community composition and abundance of taxa changed between the low and high DOC groups regardless of litter type, we analysed all of the samples together using DESeq2. Thirty-nine phyla (bacteria, archaea, fungi and viruses) were analysed and out of those, 10 were found to be differentially expressed ( $p_{adj} < 0.05$ ) between low and high DOC (Supplemental Table 6; Fig. 7). At the genus level, 136 genera were differentially expressed with 39 associating with low DOC and 97 associating with high DOC (Supplemental Table 7). The majority of genera that were differentially expressed were bacteria: 82.1% and 82.5% of the differentially expressed genera in low or high DOC respectively. Some genera that had significant differential expression regardless of litter type were also found to be significant when litter types were analysed individually including *Opitutus*, *Lentisphaera*, *Methylacidiphilum*, *Asticcaulis* and *Methylotenera* in the low DOC group and *Collinsella*, *Synechococcus*, *Lactobacillus*, *Rothia* and *Deinococcus* in the high DOC group (Supplemental Table 7).

#### Minimal metabolic overlap occurred across litter types

No functional gene groups were found to overlap across all litter types in the same DOC group. At subsystem

level 2, we found that translation and biotin subsystems had increased gene expression in low DOC communities for both grass and pine. Two other functional annotations overlapped between grass and pine, coenzyme M biosynthesis and alanine, serine, and glycine subsystems, but they did not increase expression in the same DOC group (pine: low DOC, grass: high DOC). Oak and grass shared five subsystem level 2 categories that had higher expression within the same DOC group: putative isoquinoline-1-oxidoreductase with low DOC and catabolism of an unknown compound, isoprenoid cell wall biosynthesis, Gram-positive cell wall components and plant hormones with high DOC (Supplemental Table 8). Oak and pine shared seven subsystem categories, five of which increased expression within the same DOC group including phages and prophages that increased in high DOC samples while osmotic stress and fermentation increased expression in the low DOC communities (Supplemental Fig. 8).

More specific functional overlap was observed between oak and grass in subsystem level 3. For these two litters, a consistent response occurred with 12 subsystem categories with a  $p$ -value  $< 0.05$  and an additional 16 with a  $p$ -value  $< 0.1$  (Supplemental Table 9). Mannose metabolism, alginate biosynthesis, synthesis of osmoregulated periplasmic glucans, phospholipid and fatty acid biosynthesis, and terminal cytochrome O ubiquinol



**Fig. 7.** Heatmap of DESeq2 normalized abundances of phyla with differential activity between low and high DOC communities. The litter type and DOC group for each sample are coloured in the first two rows. DOC: high = gold, low = turquoise. Litter: grass = green, oak = orange, pine = dark blue. The dendrogram represents hierarchical clustering of samples with the default settings in the pheatmap package (v 1.0.12).

oxidase were some of the level 3 subsystem categories with higher expression in the low DOC group while sialic acid metabolism, NADPH quinone oxidoreductase, glucoside transport system and L-fucose utilization had higher expression in the high DOC group (Supplemental Table 9). For pine and grass, there was less functional overlap with only two categories with a  $p$ -value  $<0.05$  and 10 additional with a  $p$ -value  $<0.1$ . Protection from reactive oxygen species increased expression in the high DOC group for both pine ( $p = 0.01$ ) and grass ( $p = 0.02$ ). Additionally, both pine and grass had increased expression of sulfatases, zinc resistance, purine utilization and peptidoglycan biosynthesis genes (Supplemental Table 9). Oak and pine shared six functional categories with  $p < 0.05$  and an additional 19 with a  $p < 0.1$ . Of note were low DOC communities in both litter types having increased expression of genes involved in fatty acid biosynthesis, benzoate catabolism, biotin biosynthesis, acetyl-CoA fermentation to butyrate, and acetone, butanol, ethanol synthesis (Supplemental Table 9). Auxin biosynthesis was the only shared subsystem level 3 category that increased expression in the high DOC group for oak and pine.

Omitting litter type as a factor with DESeq2, 34 out of 1113 functions at subsystem level 3 were differentially expressed. In high DOC communities there was an increase in the expression of 22 functional categories including prophages ( $p_{\text{adj}} = 0.0010$ ), cytochrome biogenesis ( $p_{\text{adj}} = 0.0041$ ), protection from reactive oxygen species ( $p_{\text{adj}} = 0.0060$ ), mannitol utilization ( $p_{\text{adj}} = 0.0060$ ), DNA repair ( $p_{\text{adj}} = 0.0075$ ), L-fucose utilization ( $p_{\text{adj}} = 0.0236$ ), sialic acid metabolism ( $p_{\text{adj}} = 0.0236$ ), rRNA modification in bacteria ( $p_{\text{adj}} = 0.0236$ ), iron–sulfur cluster assembly ( $p_{\text{adj}} = 0.0269$ ), phosphate uptake ( $p_{\text{adj}} = 0.0278$ ), siderophore assembly ( $p_{\text{adj}} = 0.0336$ ) and the alpha-amylase locus ( $p_{\text{adj}} = 0.0361$ ) (Supplemental Table 10). Fewer functional categories increased expression in the low DOC group (12 total with  $p_{\text{adj}} < 0.05$ ). The few functions with significantly increased expression were related to the electron transport chain ( $p_{\text{adj}} = 0.0269$ ), benzoate catabolism ( $p_{\text{adj}} = 0.0278$ ), Vir-like Type 4 secretion system ( $p_{\text{adj}} = 0.0278$ ), transcription ( $p_{\text{adj}} = 0.0288$ ), oligosaccharide biosynthesis ( $p_{\text{adj}} = 0.0319$ ), taurine utilization ( $p_{\text{adj}} = 0.0367$ ), Ton and Tol transport systems ( $p_{\text{adj}} = 0.0387$ ), glucan synthesis ( $p_{\text{adj}} = 0.0387$ ), alginate biosynthesis and metabolism ( $p_{\text{adj}} = 0.0387$ ) and fatty acid biosynthesis ( $p_{\text{adj}} = 0.0387$ ) (Supplemental Table 9). At the gene level, additional functions were differentially expressed in high DOC communities including catalase ( $p_{\text{adj}} = 0.0012$ ), cysteine synthase ( $p_{\text{adj}} = 0.0353$ ), bacterial proteasome ( $p_{\text{adj}} = 0.0372$ ), nitrate/nitrite response regulator ( $p_{\text{adj}} = 0.03417$ ) and ATP-dependent Clp protease proteolytic subunit ( $p_{\text{adj}} = 0.03949$ ) (Supplemental Table 11).

## Discussion

In this study we showed that microbial community composition can drive large variation in carbon flow during short-term litter decomposition of different litter types, and we took a first step toward identifying common physiological markers. Consistent with a large body of prior research, litter type selected different decomposer communities in our study. But within a litter type, community composition still varied with DOC group (high or low DOC) (Figs 2 and 5). Both amplicon and metatranscriptome data supported this conclusion: more of the variance in community composition was explained by DOC constrained by litter type than by either variable alone. The only two prior studies that examined how microbial composition affects carbon flow (measured as  $\text{CO}_2$ ) across multiple plant litters also found that the greatest variance in community composition was explained when both carbon flow and litter type were taken into account (Strickland *et al.*, 2009; Cleveland *et al.*, 2014). We extended prior work by demonstrating that community composition not only alters respiration during litter decomposition but also DOC abundance. Compared to previous studies (Strickland *et al.*, 2009; Cleveland *et al.*, 2014), we observed over 10-fold more variation in respiration within each litter type that was driven by microbial inoculum. Additionally, there was greater than 300% variation in DOC abundance within each litter type driven by microbial inoculum. This microbial control of DOC abundance is important because the DOC pool can be transported into deeper mineral layers where stabilization over long timescales can occur (Kaiser and Kalbitz, 2012). Overall, these findings support the hypothesis that microbial communities control carbon flow across litter types. With further investigation into the mechanisms behind this microbially driven carbon flow during litter decomposition, soil and litter microbiomes can potentially be engineered to increase soil carbon sequestration to mitigate climate change.

To improve earth system models, recent research has sought to identify broad microbial community features that impact carbon cycling (Krause *et al.*, 2014; Graham *et al.*, 2016; Kallenbach *et al.*, 2019; Malik *et al.*, 2020). Similarly, we sought common features driving DOC abundance across pine, oak and grass litter types. A combination of biomass, bacterial richness and fungal richness strongly predicted DOC abundance from the pine litter (Albright *et al.*, 2020a). The combination of these features also predicted DOC abundance for oak and grass litters significantly better than chance (Supplemental Figs 1–3). With amplicon sequence data, we again found a negative correlation between bacterial richness and DOC abundance (Fig. 3A) as previously observed with pine litter

(Albright *et al.*, 2020a), but the active community (i.e. metatranscriptome sequencing data) did not show a consistent pattern (Fig. 6C). The small metatranscriptome sample size (12 per litter type) compared to amplicon data (~100 each) may explain the discrepancy. But a more likely factor is that metatranscriptome sequencing targets the active microbiome (Franzosa *et al.*, 2015), where richness patterns may be more obscure because they are more dynamic.

Digging deeper into the common microbial community features associating with high and low DOC groups across litter types, we found some overlap both in specific microbial taxa and functional gene categories across all litter substrates, but it was rather minimal for such a diverse system (Supplemental Tables 1, 4, 7, 8). This lack of a strong overlap across all litter substrates could be due to the initial differences in litter chemistry between grass, oak and pine such as C:N ratio that affects litter decomposition rates (Osono *et al.*, 2013). Alternatively, due to the single timepoint assessed, we are likely not seeing the complete physiological picture making it challenging to assess the overlap across litter substrates. Nonetheless, it is well established that substrate (i.e. plant litter type) influences decomposition rates and microbial community composition (Taylor *et al.*, 1991; Kunito and Nagaoka, 2009; Freschet *et al.*, 2012; Rahman *et al.*, 2013; Zhang and Wang, 2015).

The eco-physiological phenomenon that accounts for high versus low DOC abundance remains unclear. Based on the negative correlation between microbial richness and DOC abundance, we previously hypothesized that high DOC communities potentially lacked microbial taxa necessary for litter decomposition and thus were experiencing a lag in decomposition. However, many lines of evidence suggest that this is not the case. First, we observed an upregulation of Actinobacteria in high DOC microbial communities (Fig. 7; Supplemental Table 5), which are known to play a prominent role in litter decomposition and have been described as late-stage generalists in litter decomposition (Kirby, 2005; Snajdr *et al.*, 2011; Schneider *et al.*, 2012; Buresova *et al.*, 2019). Second, in the high DOC group, enzymes that create reactive oxygen species were upregulated and protection mechanisms against reactive oxygen species were upregulated (Supplemental Tables 8 and 9). Reactive oxygen species are created by microbial enzymes such as superoxide dismutase and catalase (Janusz *et al.*, 2017; Bissaro *et al.*, 2018), which had increased expression in the high DOC group (Supplemental Table 10). Higher activity of these enzymes creates more ROS in the environment, inducing greater expression of ROS protection mechanisms as we observed. These coupled observations may indicate greater lignin degradation in the high DOC group – a

phenomenon expected for later stages of decomposition, not the early phase. Third, the high DOC communities have active taxa that are known to degrade a complex suite of carbon compounds ranging from lignin to simple sugars, which suggests a food web not lacking functional abilities (Kageyama *et al.*, 1999; Liu *et al.*, 2013; Ramanan *et al.*, 2014; Asadu *et al.*, 2018). Lastly, both high and low DOC microbial communities appear to be capable of metabolizing less labile plant compounds such as cellulose since most of the enzymes responsible for these processes are not significantly different between groups and both communities have taxa known to utilize these substrates. For example, in the low DOC communities, *Optitutus* spp. increased abundance in the oak and pine and are known to anaerobically metabolize cellulose (Dai *et al.*, 2016; Wilhelm *et al.*, 2017; Lacerda-Junior *et al.*, 2019). In high DOC communities, *Rothia* spp. increased abundance in oak and grass and are known to degrade cellulose and lignin (Asadu *et al.*, 2018).

Based on the preceding evidence, we developed an alternative hypothesis that the microbial communities from high and low DOC groups differ in their carbon flow outcomes (DOC/CO<sub>2</sub> abundance) because of differences in their metabolism of labile carbon compounds driven primarily by bacteria. We observed that high and low DOC groups have distinct differences in labile carbon compound metabolism (Supplemental Tables 8 and 9). Support for this alternative hypothesis is evident from significant differences in the abundance of labile carbon metabolism pathways like sugar and aromatic compound metabolism genes. For example, mannose and benzoate metabolism genes increased abundance in the low DOC group while sialic acid, mannitol, L-fucose metabolism increased in the high DOC group (Supplemental Tables 8 and 9). A conceptual model was previously proposed that microbial decomposition products such as proteins, lipids, amino sugars and carbohydrates contribute more to persistent soil carbon than complex plant compounds like lignin (Grandy and Neff, 2008). Previous research by our group discovered that during short-term pine litter decomposition, proteins, lipids and amino sugars are associated with the high DOC group (unpublished data). Additionally, in the short-term pine litter decomposition experiment, we found that the DOC produced from the high DOC microbial communities had significantly higher mineral binding capacity compared to low DOC communities, which would support longer soil residence times (Albright *et al.*, 2020a). Our finding that labile carbon metabolism differentiates the low and high DOC groups during litter decomposition, the high DOC group in pine litter decomposition has a higher mineral binding capacity, along with the above conceptual model provides support that microbial community composition may be a key control point to increase soil carbon storage.

Overall, we identified common microbial community features across all litter types using amplicon sequencing data that predicted carbon flow outcomes including bacterial richness, *Microvirga*, *Plectosphaerella* and *Rhizobium* abundances. However, using metatranscriptome analysis of the active microbial community, we were unable to find universal microbial features across all litter types that explained disparate DOC groups. Finally, our new evidence suggests that decomposition of labile rather than non-labile compounds may be key in microbially driven differences in carbon flow. Further understanding the basis of this observation may be a starting point for improving soil carbon management. Future studies should expand on these results by investigating the temporal dynamics of microbial communities and carbon flow during litter decomposition. Additionally, in-depth characterization of how the DOC is produced by different microbial communities (high versus low DOC groups) should be conducted to understand the long-term soil carbon storage potential from manipulating microbial communities.

### Acknowledgements

This work was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Division, under award numbers F260LANL2015 and F255LANL2018.

### References

- Aerts, R. (1997) Climate, leaf litter chemistry, and leaf litter decomposition in terrestrial ecosystems: a triangular relationship. *Nordic Society Oikos* **79**: 439–449.
- Albright, M.B.N., Johansen, R., Thompson, J., Lopez, D., Gallegos-Graves, L.V., Kroeger, M.E., *et al.* (2020a) Soil bacterial and fungal richness forecast patterns of early pine litter decomposition. *Front Microbiol* **11**: 542220.
- Albright, M.B.N., Thompson, J., Kroeger, M.E., Johansen, R., Ulrich, D.E.M., Gallegos-Graves, V., *et al.* (2020b) Differences in substrate use linked to divergent carbon flow during litter decomposition. *FEMS Microbiol Ecol* **96**: fiae135.
- Asadu, C., Aneke, N., Agulanna, A., and Eze, C. (2018) Degradation of sawdust by thermo tolerant microorganisms for bio fertilizer synthesis. *Asian J Biotechnol Bioresour Technol* **2**: 1–7.
- Bai, Z., Ma, Q., Dai, Y., Yuan, H., Ye, J., and Yu, W. (2017) Spatial heterogeneity of SOM concentrations associated with white-rot versus Brown-rot Wood decay. *Sci Rep* **7**: 13758.
- Bates, S.T., Berg-Lyons, D., Caporaso, J.G., Walters, W.A., Knight, R., and Fierer, N. (2010) Examining the global distribution of dominant archaeal populations in soil. *ISME J* **5**: 908.
- Batjes, N. H. (1999) Management options for reducing CO<sub>2</sub>-concentrations in the atmosphere by increasing carbon sequestration in the soil. Report 410-200-031, International Soil Reference and Information Centre (ISRIC), Wageningen, Netherlands.
- Bissaro, B., Varnai, A., Rohr, A.K., and Eijssink, V.G.H. (2018) Oxidoreductases and reactive oxygen species in conversion of lignocellulosic biomass. *Microbiol Mol Biol Rev* **82**: e00029-18.
- Bradford, M.A., Berg, B., Maynard, D.S., Wieder, W.R., Wood, S.A., and Cornwell, W. (2016) Understanding the dominant controls on litter decomposition. *J Ecol* **104**: 229–238.
- Bradford, M.A., Veen, G.F.C., Bonis, A., Bradford, E.M., Classen, A.T., Cornelissen, J.H.C., *et al.* (2017) A test of the hierarchical model of litter decomposition. *Nat Ecol Evol* **1**: 1836–1845.
- Bradford, M.A., Warren li, R.J., Baldrian, P., Crowther, T.W., Maynard, D.S., Oldfield, E.E., *et al.* (2014) Climate fails to predict wood decomposition at regional scales. *Nat Clim Change* **4**: 625–630.
- Bray, S.R., Kitajima, K., and Mack, M.C. (2012) Temporal dynamics of microbial communities on decomposing leaf litter of 10 plant species in relation to decomposition rate. *Soil Biol Biochem* **49**: 30–37.
- Buresova, A., Kopecky, J., Hrdinkova, V., Kamenik, Z., Omelka, M., and Sagova-Mareckova, M. (2019) Succession of microbial decomposers is determined by type, but site conditions drive decomposition rates. *Appl Environ Microbiol* **85**: e01760-01719.
- Cleveland, C.C., Neff, J.C., Townsend, A.R., and Hood, E. (2004) Composition, dynamics, and fate of leached dissolved organic matter in terrestrial ecosystems: results from a decomposition experiment. *Ecosystems* **7**: 175–285.
- Cleveland, C.C., Reed, S.C., Keller, A.B., Nemergut, D.R., O'Neill, S.P., Ostertag, R., and Vitousek, P.M. (2014) Litter quality versus soil microbial community controls over decomposition: a quantitative analysis. *Oecologia* **174**: 283–294.
- Cotrufo, M.F., Soong, J.L., Horton, A.J., Campbell, E.E., Haddix, M.L., Wall, D.H., and Parton, W.J. (2015) Formation of soil organic matter via biochemical and physical pathways of litter mass loss. *Nat Geosci* **8**: 776–779.
- Dai, Y., Yan, Z., Jia, L., Zhang, S., Gao, L., Wei, X., *et al.* (2016) The composition, localization and function of low-temperature-adapted microbial communities involved in methanogenic degradations of cellulose and chitin from Qinghai-Tibetan plateau wetland soils. *J Appl Microbiol* **121**: 163–176.
- Don, A., and Kalbitz, K. (2005) Amounts and degradability of dissolved organic carbon from foliar litter at different decomposition stages. *Soil Biol Biochem* **37**: 2171–2179.
- Edgar, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* **10**: 996–998.
- Franzosa, E.A., Hsu, T., Sirota-Madi, A., Shafquat, A., Abu-Ali, G., Morgan, X.C., and Huttenhower, C. (2015) Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nat Rev Microbiol* **13**: 360–372.
- Freschet, G.T., Aerts, R., and Cornelissen, J.H.C. (2012) Multiple mechanisms for trait effects on litter

- decomposition: moving beyond home-field advantage with a new hypothesis. *J Ecol* **100**: 619–630.
- Gloor, G.B., Hummelen, R., Macklaim, J.M., Dickson, R.J., Fernandes, A.D., MacPhee, R., and Reid, G. (2010) Microbiome profiling by Illumina sequencing of combinatorial sequence-tagged PCR products. *PLoS One* **5**: e15406.
- Graham, E.B., Knelman, J.E., Schindlbacher, A., Siciliano, S., Breulmann, M., Yannarell, A., et al. (2016) Microbes as engines of ecosystem function: when does community structure enhance predictions of ecosystem processes? *Front Microbiol* **7**: 214.
- Grandy, A.S., and Neff, J.C. (2008) Molecular C dynamics downstream: the biochemical decomposition sequence and its impact on soil organic matter structure and function. *Sci Total Environ* **404**: 297–307.
- Hesse, C.N., Mueller, R.C., Vuyisich, M., Gallegos-Graves, L.V., Gleasner, C.D., Zak, D.R., and Kuske, C.R. (2015) Forest floor community metatranscriptomes identify fungal and bacterial responses to N deposition in two maple forests. *Front Microbiol* **6**: 337.
- Janusz, G., Pawlik, A., Sulej, J., Swiderska-Burek, U., Jarosz-Wilkolazka, A., and Paszczynski, A. (2017) Lignin degradation: microorganisms, enzymes involved, genomes analysis and evolution. *FEMS Microbiol Rev* **41**: 941–962.
- Juarez, S., Nunan, N., Duday, A.-C., Pouteau, V., and Chenu, C. (2013) Soil carbon mineralisation responses to alterations of microbial diversity and soil structure. *Biol Fertil Soils* **49**: 939–948.
- Kageyama, A., Benno, Y., and Nakase, T. (1999) Phylogenetic and phenotypic evidence for the transfer of *Eubacterium aerofaciens* to the genus *Collinsella* as *Collinsella aerofaciens* gen. nov., comb. nov. *Int J Syst Bacteriol* **49**: 557–565.
- Kaiser, K., and Kalbitz, K. (2012) Cycling downwards – dissolved organic matter in soils. *Soil Biol Biochem* **52**: 29–32.
- Kallenbach, C.M., Wallenstein, M.D., Schipanski, M.E., and Grandy, A.S. (2019) Managing agroecosystems for soil microbial carbon use efficiency: ecological unknowns, potential outcomes, and a path forward. *Front Microbiol* **10**: 1146.
- Kamphake, L.J., Hannah, S.A., and Cohen, J.M. (1967) Automated analysis for nitrate by hydrazine reduction. *Water Res* **1**: 205–216.
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**: D457–D462.
- Kassambara, A. (2020) Ggpubr: ‘ggplot2’ based publication ready plots.
- Keegan, K.P., Glass, E.M., and Meyer, F. (2016) MG-RAST, a metagenomics service for analysis of microbial community structure and function. In *Methods in Molecular Biology*, Martin, F., and Uroz, S. (eds). New York, NY: Humana Press.
- Kirby, R. (2005) Actinomycetes and lignin degradation. *Adv Appl Microbiol* **58**: 125–168.
- Krause, S., Le Roux, X., Niklaus, P.A., Van Bodegom, P.M., Lennon, J.T., Bertilsson, S., et al. (2014) Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. *Front Microbiol* **5**: 251.
- Kunito, T., and Nagaoka, K. (2009) Effects of plant litter type and additions of nitrogen and phosphorus on bacterial community-level physiological profiles in a brown forest soil. *Microbes Environ* **24**: 68–71.
- Lacerda-Junior, G.V., Noronha, M.F., Cabral, L., Delforno, T. P., de Sousa, S.T.P., Fernandes-Junior, P.I., et al. (2019) Land use and seasonal effects on the soil microbiome of a Brazilian dry Forest. *Front Microbiol* **10**: 648.
- Liu, Z.X., Yang, L.L., Huang, Y., Zhao, H., Liu, H., Tang, S. K., et al. (2013) *Rothia marina* sp. nov., isolated from an intertidal sediment of the South China Sea. *Antonie Van Leeuwenhoek* **104**: 331–337.
- Lopez-Mondejar, R., Brabcova, V., Stursova, M., Davidova, A., Jansa, J., Cajthaml, T., and Baldrian, P. (2018) Decomposer food web in a deciduous forest shows high share of generalist microorganisms and importance of microbial biomass recycling. *ISME J* **12**: 1768–1778.
- Louis, B.P., Maron, P.A., Viaud, V., Leterme, P., and Menasseri-Aubry, S. (2016) Soil C and N models that integrate microbial diversity. *Environ Chem Lett* **14**: 331–344.
- Love, M.I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.
- Malik, A.A., Chowdhury, S., Schlager, V., Oliver, A., Puissant, J., Vazquez, P.G., et al. (2016) Soil fungal:bacterial ratios are linked to altered carbon cycling. *Front Microbiol* **7**: 1247.
- Malik, A.A., Martiny, J.B.H., Brodie, E.L., Martiny, A.C., Treseder, K.K., and Allison, S.D. (2020) Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. *ISME J* **14**: 1–9.
- Maron, P.A., Sarr, A., Kaisermann, A., Leveque, J., Mathieu, O., Guigue, J., et al. (2018) High microbial diversity promotes soil ecosystem functioning. *Appl Environ Microbiol* **84**: e02738-17.
- Meyer, F., Paarmann, D., D’Souza, M., Olson, R., Glass, E. M., Kubal, M., et al. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.
- Mueller, R., Belnap, J., and Kuske, C. (2015) Soil bacterial and fungal community responses to nitrogen addition across soil depth and microhabitat in an arid shrubland. *Front Microbiol* **6**: 891.
- Mueller, R.C., Gallegos-Graves, L.V., and Kuske, C.R. (2016) A new fungal large subunit ribosomal RNA primer for high-throughput sequencing surveys. *FEMS Microbiology Ecology* **92**: fiv153.
- Newcomb, C.J., Qafoku, N.P., Grate, J.W., Bailey, V.L., and De Yoreo, J.J. (2017) Developing a molecular picture of soil organic matter-mineral interactions by quantifying organo-mineral binding. *Nat Commun* **8**: 396.
- Nielsen, U.N., Ayres, E., Wall, D.H., and Bardgett, R.D. (2011) Soil biodiversity and carbon cycling: a review and synthesis of studies examining diversity-function relationships. *Eur J Soil Sci* **62**: 105–116.
- O’Leary, N.A., Wright, M.W., Brister, J.R., Ciufu, S., Haddad, D., McVeigh, R., et al. (2016) Reference sequence (RefSeq) database at NCBI: current status,



- taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**: D733–D745.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., *et al.* (2019) ‘Vegan’: community ecology package.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., and McGlinn, D. (2017) vegan: Community Ecology Package. R package version 2.4-3.
- Osono, T., Azuma, J.-i., and Hirose, D. (2013) Plant species effect on the decomposition and chemical changes of leaf litter in grassland and pine and oak forest soils. *Plant Soil* **376**: 411–421.
- Paustian, K., Larson, E., Kent, J., Marx, E., and Swan, A. (2019) Soil C sequestration as a biological negative emission strategy. *Front Clim* **1**: 8.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., *et al.* (2011) Scikit-learn: machine learning in python. *J Mach Learn Res* **12**: 2825–2830.
- Prescott, C.E. (2010) Litter decomposition: what controls it and how can we alter it to sequester more carbon in forest soils? *Biogeochemistry* **101**: 133–149.
- R Core Team. (2019) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- R Core Team. (2020) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rahman, M.M., Tsukamoto, J., Rahman, M.M., Yoneyama, A., and Mostafa, K.M. (2013) Lignin and its effects on litter decomposition in forest ecosystems. *Chem Ecol* **29**: 540–553.
- Ramanan, P., Barreto, J.N., Osmon, D.R., and Tosh, P.K. (2014) *Rothia* bacteremia: a 10-year experience at Mayo Clinic, Rochester, Minnesota. *J Clin Microbiol* **52**: 3184–3189.
- Rumpel, C., and Kögel-Knabner, I. (2010) Deep soil organic matter – a key but poorly understood component of terrestrial C cycle. *Plant Soil* **338**: 143–158.
- Saifuddin, M., Bhatnagar, J.M., Segre, D., and Finzi, A.C. (2019) Microbial carbon use efficiency predicted from genome-scale metabolic models. *Nat Commun* **10**: 3568.
- Schmidt, M.W., Torn, M.S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I.A., *et al.* (2011) Persistence of soil organic matter as an ecosystem property. *Nature* **478**: 49–56.
- Schneider, T., Keiblinger, K.M., Schmid, E., Sterflinger-Gleixner, K., Ellersdorfer, G., Roschitzki, B., *et al.* (2012) Who is who in litter decomposition? Metaproteomics reveals major microbial players and their biogeochemical functions. *ISME J* **6**: 1749–1762.
- Schöning, I., and Kögel-Knabner, I. (2006) Chemical composition of young and old carbon pools throughout Cambisol and Luvisol profiles under forests. *Soil Biol Biochem* **38**: 2411–2424.
- Seabold, S., and Perktold, J. (2010) Statsmodels: econometric and statistical modeling with python. Proceedings of the 9th python in science conference.
- Siletti, C.E., Zeiner, C.A., and Bhatnagar, J.M. (2017) Distributions of fungal melanin across species and soils. *Soil Biol Biochem* **113**: 285–293.
- Snajdr, J., Cajthaml, T., Valaskova, V., Merhautova, V., Petrankova, M., Spetz, P., *et al.* (2011) Transformation of *Quercus petraea* litter: successive changes in litter chemistry are reflected in differential enzyme activity and changes in the microbial community composition. *FEMS Microbiol Ecol* **75**: 291–303.
- Soltanpour, P.N., and Workman, S. (1979) Modification of the NH<sub>4</sub> HCO<sub>3</sub>-DTPA soil test to omit carbon black. *Commun Soil Sci Plan Anal* **10**: 1411–1420.
- Strickland, M.S., Lauber, C., Fierer, N., and Bradford, M.A. (2009) Testing the functional significance of microbial community composition. *Ecology* **90**: 441–451.
- Taylor, B.R., Prescott, C.E., Parsons, W.J.F., and Parkinson, D. (1991) Substrate control of litter decomposition in four Rocky Mountain coniferous forests. *Canada J Bot* **69**: 2242–2250.
- Thompson, J., Johansen, R., Dunbar, J., and Munsky, B. (2019) Machine learning to predict microbial community functions: an analysis of dissolved organic carbon from litter decomposition. *PLoS One* **14**: e0215502.
- Vilgalys, R., and Hester, M. (1990) Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several *Cryptococcus* species. *J Bacteriol* **172**: 4238–4246.
- Wagg, C., Schlaeppi, K., Banerjee, S., Kuramae, E.E., and van der Heijden, M.G.A. (2019) Fungal-bacterial diversity and microbiome complexity predict ecosystem functioning. *Nat Commun* **10**: 4841.
- Walkely, A. (1947) A critical examination of a rapid method for determining organic carbon in soils: effect of variations in digestion conditions and of inorganic soil constituents. *Soil Sci* **63**: 251–263.
- Walkely, A., and Black, I.A. (1934) An examination of the Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. *Soil Sci* **37**: 29–38.
- Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.
- Waring, B.G., Averill, C., and Hawkes, C.V. (2013) Differences in fungal and bacterial physiology alter soil carbon and nitrogen cycling: insights from meta-analysis and theoretical models. *Ecol Lett* **16**: 887–894.
- Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
- Wickings, K., Grandy, A.S., Reed, S.C., and Cleveland, C.C. (2012) The origin of litter chemical complexity during decomposition. *Ecol Lett* **15**: 1180–1188.
- Wieder, W.R., Grandy, A.S., Kallenbach, C.M., Taylor, P.G., and Bonan, G.B. (2015) Representing life in the earth system with soil microbial functional traits in the MIMICS model. *Geosci Model Dev* **8**: 1789–1808.
- Wilhelm, R.C., Cardenas, E., Leung, H., Szeitz, A., Jensen, L.D., and Mohn, W.W. (2017) Long-term enrichment of stress-tolerant cellulolytic soil populations following timber harvesting evidenced by multi-omic stable isotope probing. *Front Microbiol* **8**: 537.
- Zhang, X., and Wang, W. (2015) Control of climate and litter quality on leaf litter decomposition in different climatic zones. *J Plant Res* **128**: 791–802.

## Supporting Information

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

**Supplemental Fig. 1.** Distribution of cumulative CO<sub>2</sub> (mg g<sup>-1</sup> litter) in each plant litter type. The global Kruskal-Wallis H test p-value comparing all litter types and each individual comparison are reported.

**Supplemental Fig. 2.** Distribution of cumulative DOC (mg g<sup>-1</sup> litter) in each plant litter type. The global Kruskal-Wallis H test p-value comparing all litter types and each individual comparison are reported.

**Supplemental Fig. 3.** Linear regressions of cumulative CO<sub>2</sub> and DOC (mg g<sup>-1</sup> litter) in (A) pine, (B) oak, and (C) grass mix. Pearson's correlation R, r<sup>2</sup> and p-values are reported for each.

**Supplemental Fig. 4.** Receiver operating characteristic (ROC) curve of out-of-fold predictions of pine litter DOC using a logistic regression model with total biomass, bacterial richness, and fungal richness as model features. The area under the ROC curve was 0.89, which reflects the model's ability to distinguish high DOC samples from low DOC samples. The proportion of correctly classified held-out samples was significantly greater than the proportion of the most frequently occurring class ( $p = 1.03 \times 10^{-10}$ , Z-test of two proportions).

**Supplemental Fig. 5.** Receiver operating characteristic (ROC) curve of out-of-fold predictions of oak litter DOC using a logistic regression model with total biomass, bacterial richness, and fungal richness as model features. The area under the ROC curve was 0.90, which reflects the model's ability to distinguish high DOC samples from low DOC samples. The proportion of correctly classified held-out samples was significantly greater than the proportion of the most frequently occurring class ( $p = 1.58 \times 10^{-5}$ , Z-test of two proportions).

**Supplemental Fig. 6.** Receiver operating characteristic (ROC) curve of out-of-fold predictions of grass litter DOC using a logistic regression model with total biomass, bacterial richness, and fungal richness as model features. The area under the ROC curve was 0.83, which reflects the model's ability to distinguish high DOC samples from low DOC samples. The proportion of correctly classified held-out samples was significantly greater than the proportion of the most frequently occurring class ( $p = 9.49 \times 10^{-6}$ , Z-test of two proportions).

**Supplemental Fig. 7.** Relative abundance of bacterial and fungal phyla in each litter type from the low DOC group (A) and high DOC group (B).

**Supplemental Fig. 8.** Heatmap of subsystem level 2 functional groups that significantly changed abundance between low and high DOC groups based on a Kruskal-Wallis H test using rarefied log-transformed abundances. The asterisk shows whether it was significantly higher in high or low DOC group.

**Supplemental Table 1.** Metadata on the soils used as inoculum in the microcosms for pine, oak, and grass mix. Study indicates which microcosm experiment it was used as an inoculum. Study date provides the time the soil was collected. Litter details what plant litter was used in the microcosms. Soil ID is the internal ID used for the soil collection.

Soil Ecosystem details the ecosystem from which the soil was collected. Lat and Long provide the GPS coordinates for the collection. Alt = the elevation of the location where the soil was collected.

**Supplemental Table 2.** RFINN output of the significant ( $p < 0.05$ ) bacterial and fungal genera predicting DOC group and CO<sub>2</sub> group from amplicon sequencing data within litter types and across multiple litters. The genera highlighted in yellow are found across all litter types with the same phenotype (high or low). Genera highlighted in green are found in 2 out of 3 litter types with the same phenotype (high or low). Genera highlighted in blue are found in 2 out of 3 litter types, but in different phenotypes (high or low).

**Supplemental Table 3.** Pearson's product-moment correlation values between diversity (functional and taxonomic) and DOC abundance for all samples. When litter types were analysed together, samples were rarefied 18,800 annotations per sample for functional diversity and 98,500 annotations for taxonomic diversity. When litter types were analysed separately, pine litter was rarefied to 18,800 and 98,500, oak litter was rarefied to 64,600 and 224,000, and grass litter was rarefied to 20,700 and 164,000 for functional and taxonomic diversity respectively. Diversity metrics used to analyse the samples include Shannon-Weaver index, Simpson index, inverse Simpson index, richness, Pielou's evenness, and the Chao1 richness estimator.

**Supplemental Table 4.** The abundance of phyla that significantly changed between low and high DOC communities within each litter type (pine, oak, and grass) based on a Kruskal-Wallis H test of rarefied log-transformed abundances. Only annotations classified in MG-RAST by RefSeq as Bacteria, Fungi, Viruses, or Archaea were included in these analyses. No phyla significantly ( $p$ -value  $< 0.05$ ) changed between low and high DOC for grass and therefore is not reported here.

**Supplemental Table 5.** Genera from oak, pine, and grass found to be significant in more than one litter type based on the Kruskal-Wallis H Test using rarefied log-transformed abundances. The statistic and p-value from the Kruskal-Wallis H Test are reported along with the average abundance in low and high DOC groups. The cells in green highlight the higher value for each genus within each litter.

**Supplemental Table 6.** Phyla with significant differential expression based on DESeq2 analysis between low and high DOC when all litter types (pine, oak, and grass) were analysed together. The baseMean is the normalized counts of all samples to account for sequencing depth. Log2FoldChange is the difference between low and high DOC normalized counts with  $> 0$  associating with low DOC and  $< 0$  associating with high DOC. lfcSE is the standard error for log2 fold change. Stat is the Wald statistic that divides the log2 fold change by its standard error which is used to calculate the p-value. Padj is the multi-test correct p-value using the Benjamin-Hochberg correction. DOC category reports whether the phylum associated with high or low DOC.

**Supplemental Table 7.** Genera determined to be differentially expressed between low and high DOC communities when all litter types (pine, oak, and grass) were analysed together using DESeq2. The baseMean is the normalized

counts of all samples to account for sequencing depth. Log2FoldChange is the difference between low and high DOC normalized counts with  $>0$  associating with low DOC and  $< 0$  associating with high DOC. lfcSE is the standard error for log2 fold change. Stat is the Wald statistic that divides the log2 fold change by its standard error which is used to calculate the p-value. Padj is the multi-test correct p-value using the Benjamin-Hochberg correction. DOC category reports whether the phylum associated with high or low DOC. The phylum, class, order, and family are reported for each genus under the appropriate header.

**Supplemental Table 8.** Subsystem level 2 functional genes from oak, pine, and grass found to be significant in more than one litter type based on the Kruskal-Wallis H Test using rarefied log-transformed abundances. The statistic and p-value from the Kruskal-Wallis H Test are reported along with the average abundance in low and high DOC groups. The cells in green highlight the higher value for each genus within each litter.

**Supplemental Table 9.** Subsystem level 3 from oak, pine, and grass found to be significant ( $p < 0.1$ ) in more than one litter type based on the Kruskal-Wallis H Test using rarefied log-transformed abundances. The statistic and p-value from the Kruskal-Wallis H Test are reported along with the average abundance in low and high DOC groups. The cells in green highlight the higher value for each genus within each litter.

**Supplemental Table 10.** Level 3 Subsystem functions determined to be differentially expressed between low and high DOC communities when all litter types (pine, oak, and grass) were analysed together using DESeq2. The baseMean is the normalized counts of all samples to account for sequencing depth. Log2FoldChange is the difference between low and high DOC normalized counts with  $>0$  associating with low DOC and  $< 0$  associating with high DOC. lfcSE is the standard error for log2 fold change. Stat is the Wald statistic that divides the log2 fold change by its standard error which is used to calculate the p-value. Padj is the multi-test correct p-value using the Benjamin-Hochberg correction. DOC category reports whether the functional group increased abundance in high or low DOC.

**Supplemental Table 11.** Subsystem functions determined to be differentially expressed between low and high DOC communities when all litter types (pine, oak, and grass) were analysed together using DESeq2. The baseMean is the normalized counts of all samples to account for sequencing depth. Log2FoldChange is the difference between low and high DOC normalized counts with  $>0$  associating with low DOC and  $< 0$  associating with high DOC. lfcSE is the standard error for log2 fold change. Stat is the Wald statistic that divides the log2 fold change by its standard error which is used to calculate the p-value. Padj is the multi-test correct p-value using the Benjamin-Hochberg correction. DOC category reports whether the functional group increased abundance in high or low DOC.