Supplemental Information

Contents

1 Supplemental Methods

1.1 Soil Collection and Preparation

We collected soils from an organic farm in Penn Yan, New York. Soils were Honoeye/Lima, a silty clay loam on calcareous bedrock. To get a field average, cores (5 cm diameter x 10 cm depth) were collected in duplicate from six different sampling locations around the field using a slide hammer bulk density sampler (coordinates: (1) N 42° 40.288 W 77° 02.438, (2) N 42 40.296 W 77° 02.438, (3) N 42° 40.309 W 77° 02.445, (4) N 42° 40.333 W 77° 02.425, (5) N 42° 40.340 W 77° 02.420, (6) N 42° 40.353 W 77° 02.417) on November 21, 2011. Soil cores were sieved (2mm), homogenized by mixing, and stored at 4C until preincubation (within 1-2 week of collection). Carbon and nitrogen content were previously measured for these soils [1]. Reported soil C values for the organic field were 12.15 (± s.d. 0.78) mg C g^{-1} dry soil and 1.16 (± s.d. 0.13) mg N g^{-1} dry soil.

1.2 Cellulose production

Bacterial cellulose was produced by *Gluconoacetobacter xylinus* grown in Heo and Son [2] minimal media (HS medium) made with 0.1% glucose and without inositol. For the production of 13C-cellulose, $^{13}C_6$-D-glucose, 99 atom % 13C (Isotec) was used. Cellulose was produced in 1L Erlenmeyer flasks containing 100 mL HS medium inoculated with three colonies of *Gluconoacetobacter xylinus* grown on HS agar plates. Flasks were incubated statically in the dark at 30°C for 2-3 weeks. Cellulose pellicules were decanted, rinsed with deionized water, suspended in two volumes of 1% alconox, and then autoclaved. Cellulose pellicules were purified by dialysis for 12 hr in 1 L deionized water and dialysis was repeated 10 times. Harvested pellicules were dried overnight (60°C), cut into pieces, and ground to 53 μm - 250 μm using 5100 Mixer/Mill (SPEX SamplePrep, Metuchen, NJ) and dry sieve. The particulate size range was selected to be representative of particulate organic matter in soils (3).

The purity of ground cellulose was checked by biological assay, Benedict's reducing sugars assay, Bradford assay, and isotopic analysis. *E.coli* is not able to use cellulose as a C source but is capable of growth on a variety of nutrients available in the Heo and Son medium. The biological assay consisted of *E. coli* inoculated into minimal M9 media which lacked a C source and was supplemented with either: (1) 0.01% glucose, (2) 2.5 mg purified, ground cellulose, (3) 25 mg purified, ground cellulose, (4) 25 mg purified, ground cellulose and 0.01% glucose. Growth in media was checked by spectrometer (OD_{450}). No measurable growth was observed with either 2 mg or 25 mg cellulose, indicating absence of contaminating nutrients that can support growth of *E. coli*. In addition, the presence of 25 mg cellulose did not inhibit the growth of *E.coli* cultures provided with glucose (relative to control), indicating the absence of compounds in the purified cellulose that could inhibit microbial growth.

Purified cellulose was also assayed for residual proteins and sugars using Bradford and Benedict's assays, respectively. Bradford assay was performed as in [3] with a standard curve ranging from 0 - 2000 μg ml $^{-1}$ BSA. Ground, purified cellulose contained 6.92 μg protein mg cellulose^{-1}(*i.e.* 99.31% purity). Reducing sugars were not detected in cellulose using Benedict's reducing sugar assay [4] tested at 10 mg cellulose ml^{-1}. Finally, ^{13}C-cellulose had an average 96% $\pm$ 5 (s.d.) degree of ^{13}C labeling as determined by isotopic analysis (UCDavis Stable Isotope Facility).

1.3 Soil microcosms

Microcosms were created by adding 10 g d.w. sieved soil to a 250 mL Erlenmeyer flask capped with a butyl rubber stopper. The headspace was flushed with air every 3 days which was sufficient to prevent anoxia (data not shown). Microcosms were pre-incubated at room temperature for 2 weeks until the soil respiration rate (determined by GCMS measurement of head space CO_2) had stabilized. Sieving causes a transient increase in soil respiration rate presumably due to the liberation of fresh labile soil organic matter [5]. Pre-incubation ensures that this labile organic matter is consumed and/or stabilized prior to the beginning of the experiment. Respiration rate (CO_2) stabilized after 10 days (data not shown).

Three parallel treatments were established. Each treatment received the same amendment, where the only difference was the isotopically labeled component in the amendment. Specifically, we made unlabeled control treatment and treatments that substituted either ^{13}C-cellulose (synthesized as described above) or $^{13}C_5$-D-xylose, 98 atom % ^{13}C (Isotec) for their unlabeled equivalents. The molecular composition of the amendment was designed to approximate switchgrass biomass with hemicellulose replaced by its constituent monomers [6, 7]. The amendment was added at 5.3

mg g^{-1} d.w. soil which is representative of natural concentrations in soil during early phases of decomposition [8]. The amendment contained by mass: 38% cellulose, 23% lignin, 20% xylose, 3% arabinose, 1% galactose, 1% glucose, and 0.5% mannose, with the remaining 13.5% composed of amino acids (Teknova C0705) and a basal salt mixture (Murashige and Skoog, Sigma Aldrich M5524). The amendment had a C:N ratio of 10. Cellulose (2 mg cellulose g^{-1} d.w. soil) and lignin (1.2 mg lignin g-1 d.w. soil) were uniformly distributed over the soil surface as a powder and the remaining constituents were added in solution in a volume of 0.12 ml g^{-1} d.w. soil. The volume of liquid was determined in relation to soil moisture to achieve 50% water holding capacity. Water holding capacity of 50% was chosen, in relation to the texture for this soil, to achieve approximately 70% water filled pore space, which is the optimal water content for respiration [9]. A total of 12 microcosms were established per treatment. Microcosms were sampled destructively on days 1, 3, 7, 14, and 30 and soils were frozen at -80 °C. The cellulose treatment was not sampled on day 1 because it was not expected that significant cellulose metabolism would have occurred within this time. The abbreviation 13CXPS refers to the 13C-xylose treatment (13C Xylose Plant Simulant), 13CCPS refers to the 13C-cellulose treatment and 12CCPS refers to the unlabeled control. A subset of soil from each sample was reserved for isotopic analysis at the Cornell University Stable Isotope Laboratory to determine the mass of ^{13}C remaining in soil.

1.4 Nucleic acid extraction

Nucleic acids were extracted from 0.25 g soil using a modified Griffiths protocol [10]. Cells were lysed by bead beating for 1 min at 5.5 ms^{-1} in 2mL lysis tubes containing 0.5 g of 0.1 mm diameter silica/zirconia beads (treated at 300C for 4 hours to remove RNases), 0.5 mL extraction buffer (240 mM Phosphate buffer 0.5% N-lauryl sarcosine), and 0.5 mL phenol-chloroform-isoamyl alcohol (25:24:1) for 1 min at 5.5 ms^{-1}. After lysis, 85 uL 5 M NaCl and 60 uL 10% hexadecyltriammonium bromide (CTAB)/0.7 M NaCl were added to lysis tube, vortexed, chilled for 1 min on ice, and centrifuged at 16,000 x g for 5 min at 4C. The aqueous layer was transferred to a new tube and reserved on ice. To increase DNA recovery, the pellet was back extracted with 85 uL 5 M NaCl and 0.5 mL extraction buffer. The aqueous extract was washed with 0.5 mL chloroform:isoamyl alcohol (24:1). Nucleic acids were precipitated by addition of 2 volumes polyethylene glycol solution (30% PEG 8000, 1.6 M NaCl) on ice for 2 hrs, followed by centrifugation at 16,000 x g, 4C for 30 min. The supernatant was discarded and pellets were washed with 1 mL ice cold 70% EtOH. Pellets were air dried, resuspended in 50 uL TE and stored at -20C. To prepare nucleic acid extracts for isopycnic centrifugation as previously described [11], DNA was size selected ($>$ 4kb) using 1% low melt agarose gel and β-agarase I enzyme extraction per manufacturers protocol (New England Biolab, M0392S). Final resuspension of DNA pellet was in 50 μL TE.

1.5 Isopycnic centrifugation and fractionation

We fractionated DNA on density gradients for ^{13}C-xylose treatments (days 1, 3, 7, 14, 30), ^{13}C-cellulose treatments (days 3, 7, 14, 30), and control treatments (days 1, 3 ,7, 14, 30). A total of 5 μg DNA was added to each 4.7 mL CsCl density gradient. Density gradient were composed of 1.69 g mL^{-1} CsCl ml^{-1} in gradient buffer solution (pH 8.0 15 mM Tris-HCl, 15 mM EDTA, 15 mM KCl). Centrifugation was performed at 55,000 rpm 20 °C for 66 hr using a TLA-110 rotor in a Bechman Coulter Optima MAX-E ultracentrifuge. Fractions of $\sim$100 μL were collected from below by displacing the DNA-CsCl-gradient buffer solution in the centrifugation tube with water

using a syringe pump at a flow rate of 3.3 μL s^{-1} [12] into Acroprep 96 filter plate (part no. 5035, Pall Life Sciences). The refractive index of each fraction was measured using a Reichart AR200 digital refractometer modified as previously described to measure a volume of 5 μL [11]. Buoyant density was calculated from the refractive index as previously described [11] using the equation ρ=$a\eta$-b, where ρ is the density of the CsCl (g ml^{-1}), η is the measured refractive index, and a and b are coefficient values of 10.9276 and 13.593, respectively, for CsCl at 20C [13] and correcting for non-CsCl salts in the gradient buffer. A total of 35 gractions were collected from each gradient and the average density between fractions was 0.0040 g mL^{-1}. The DNA was desalted by washing with TE (5X 200 μL) in the Acroprep filter wells. DNA was resuspeneded in 50 μL TE.

1.6 DNA Sequencing

1.6.1 PCR amplification of SSU rRNA genes

SSU rRNA genes were amplified from gradient fractions (n = 20 per gradient) and from non-fractionated DNA from soil. Barcoded primers consisted of: 454-specific adapter B, a 10 bp barcode (Reference 90), a 2 bp linker (5-CA-3), and 806R primer for reverse primer (BA806R); and 454-specific adapter A, a 2 bp linker (5-TC-3), and 515F primer for forward primer (BA515F). Each PCR contained 1.25 U l-1 AmpliTaq Gold (Life Technologies, Grand Island, NY; N8080243), 1X Buffer II (100 mM Tris-HCl, 500 mM KCl, pH 8.3), 2.5 mM MgCl2, 200 M of each dNTP, 0.5 mg ml-1 BSA, 0.2 M BA515F, 0.2 M BA806R, and 10 L of 1:30 DNA template in 25 l total volume). The PCR conditions were 95C for 5min followed by 22 cycles of 95C for 10s, 53C for 30s, and 72C for 30s, followed by a final elongation at 72C for 5 min. Amplification products were checked by 1% agarose gel. Reactions were performed in triplicate and pooled. Amplified DNA was gel purified (1% low melt agarose) using the Wizard SV gel and PCR clean-up system (Promega, Madison, WI; A9281) per manufacturers protocol. Samples were normalized by SequalPrepTM normalization plates (Invitrogen, Carlsbad, CA; A10510) and pooled in equimolar concentration. Amplicons were sequenced on Roche 454 FLX system using titanium chemistry at Selah Genomics (Columbia, SC).

1.6.2 DNA sequence quality control

SSU rRNA gene sequences were initially screened by maximum expected errors at a specific read length threshold [14]. Reads that had more than 0.5 expected errors at a length of 250 nt were discarded. The remaining reads were aligned to the Silva Reference Alignment as provided in the Mothur software package using the Mothur NAST aligner [15, 16]. Reads that did not align to the expected region of the SSU rRNA gene were discarded. After expected error and alignment based quality control. The remaining quality controlled reads were annotated using the UClust taxonomic annotation framework in [17, 18]. We used 97% cluster seeds from the Silva SSU rRNA database (release 111Ref) [19] as reference for taxonomic annotation (provided on the QIIME website) [19]. Quality control screening filtered out 344,472 or 1,720,480 raw sequencing reads. Reads annotated as "Chlorloplast", "Eukaryota", "Archaea", "Unassigned" or "mitochondria" were culled from the dataset.

1.6.3 OTU binning

Sequences were distributed into OTUs with a centroid based clustering algorithm (i.e. UPARSE [14]). The centroid selection also included robust chimera screening [14]. OTU centroids were

established at a threshold of 97% sequence identity and non-centroid sequences were mapped back to centroids. Reads that could not be mapped to an OTU centroid at greater than or equal to 97% sequence identity were discarded.

1.6.4 Phylogenetic reconstruction

We used SSU-Align [20, 21] to align SSU rRNA gene sequences. Columns in the alignment that were aligned with poor confidence ($< 95\%$ of characters had posterior probability $> 95\%$) were not considered when building the phylogenetic tree. Additionally, the alignment was trimmed to coordinates such that all sequences in the alignment began and ended at the same positions. FastTree [22] was used with default parameters to build the phylogeny. NMDS ordination was performed on weighted Unifrac [23] distances between samples. The Phyloseq [24] wrapper for Vegan [25] (both R packages) was used to compute sample values along NMDS axes. The 'adonis' function in Vegan was used to perform Adonis tests (default parameters) [26].

1.7 OTU characteristics

1.7.1 Identifying ^{13}C responders

Figures S11 and S12 demonstrate raw data for responder and non-responder OTUs, respectively. Responders increased in relative abundance in the heavy fractions due to ^{13}C-labeling of their DNA. As our data is compositional, often OTUs had consistent *relative* abundance across the density gradients. If OTU DNA is positioned in heavy or light fractions, however, due to G+C content and/or ^{13}C-labeling, it spikes in relative abundance near where it is centered. Thus, we identified responders by finding OTUs enriched in heavy fractions of ^{13}C treatment gradients relative to control. This technique accounts for the variation in OTU base abundance and the variation in OTU G+C content (and therefore natural buoyant density) because ^{13}C treatment abundances are always compared to appropriate control abundances.

1.7.2 Estimating rrn copy number

We estimated the *rrn* copy number for each OTU as described in **(author?)** [27] (i.e. we used the code and reference information provided in **(author?)** [27] directly). In brief, OTU centroid sequences were inserted into a reference SSU rRNA gene phylogeny [28] from organisms of known *rrn* copy number. The *rrn* copy number was then inferred from the phylogenetic placement in the reference phylogeny.

1.7.3 NRI, NTI, and consenTRAIT

NRI and NTI were calculated using the "picante" R package [29]. We used the "independentswap" null model for phylogenetic distribution. The consenTRAIT clade depth for xylose and cellulose responders was calculated using R code used to calculate the metric in **(author?)** [30] which employs the R "adephylo" package [31].

1.7.4 Buoyant density shift estimates

Upon labeling, DNA from an organism that incorporates exclusively ^{13}C will increase in BD more than DNA from an organism that does not exclusively utilize isotopically labeled C. Therefore, the

magnitude DNA $\Delta\hat{BD}$ indicates substrate specificity given our experimental design as only one substrate was labeled in each amendment (assuming all members of an OTU behave similarly with respect to ^{13}C incorporation). We measured $\Delta\hat{BD}$ as the change in an OTU's density profile center of mass between corresponding control and labeled gradients (Figure S11). Because all gradients did not span the same density range and gradient fractions cannot be taken at specific density positions, we limited our $\Delta\hat{BD}$ analysis to the density range for which fractions were taken for all gradients. Within this density range we linearly interpolated 20 evenly spaced relative abundance values. The center of mass for an OTU along the density gradient was then the density weighted average where weights were the linearly interpolated relative abundance values. $\Delta\hat{BD}$ should not be evaluated on an individual OTU basis as a small number of $\Delta\hat{BD}$ values are observed for each OTU and the variance of the density shift metric at the level of individual OTUs is unknown. It is therefore more informative to compare $\Delta\hat{BD}$ among substrate responder groups. Further, $\Delta\hat{BD}$ values are based on relative abundance profiles and would be distorted in comparison to $\Delta\hat{BD}$ based on absolute DNA concentration profiles and should be interpreted with this transformation in mind. It should also be noted that there was overlap in observed $\Delta\hat{BD}$ between ^{13}C-cellulose and ^{13}C-xylose responder groups.

1.7.5 Finding cultured relatives of OTUs

OTU centroids were compared (BLAST [32, 33]) to sequences in "The All-Species Living Tree" project (LTP). The LTP is a collection of SSU rRNA gene sequences for classified species of Archaea and Bacteria [34]. We used LTP version 115 for analyses in this paper.

1.7.6 OTU changes in relative abundance with time

We identified OTUs that changed in relative abundance over time using DESeq2 [35]. Specifically, we used day treated as an ordered factor as the regressor with LFC of the relative abundance in non-fractionated DNA as the outcome in the general linear model. We used the default DESeq2 base mean independent filtering and disabled the Cook's cutoff outlier detection. The null model was that abundance did not change with time and we assessed significance at a false discovery rate of 10%.

1.8 Sequencing and density fractionation statistics

Microcosm DNA was density fractionated on CsCl density gradients. We sequenced SSU rRNA gene amplicons from a total of 277 CsCl gradient fractions from 14 CsCl gradients and 12 bulk microcosm DNA samples. The SSU rRNA gene data set contained 1,102,685 total sequences. The average number of sequences per sample was 3,816 (sd 3,629) and 265 samples had over 1,000 sequences. We sequenced SSU rRNA gene amplicons from an average of 19.8 fractions per CsCl gradient (sd 0.57). The average density between fractions was 0.0040 g mL^{-1} The sequencing effort recovered a total of 5,940 OTUs. 2,943 of the total 5,940 OTUs were observed in bulk samples. We observed 33 unique phylum and 340 unique genus annotations.

References

[1] Berthrong S-T, Buckley D-H, Drinkwater L-E (2013) Agricultural Management and Labile Carbon Additions Affect Soil Microbial Community Structure and Interact with Carbon and Nitrogen Cycling. *Microb Ecol* 66(1): 158–170.

[2] Heo M-S, Son H-J (2002) Development of an optimized, simple chemically defined medium for bacterial cellulose production by Acetobacter sp. A9 in shaking cultures. *Biotechnol Appl Biochem* 36(1): 41.

[3] Bradford M-M (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 72(1-2): 248–254.

[4] Benedict S-R (1909) A reagent for the detection of reducing sugars. *Journal of Biological Chemistry* 5(5): 485–487.

[5] Datta R, Vranová V, Pavelka M, Rejšek K, Formánek P (2014) Effect of soil sieving on respiration induced by low-molecular-weight substrates. *International Agrophysics* 28(1): ???.

[6] Yan J, Hu Z, Pu Y, Brummer E-C, Ragauskas A-J (2010) Chemical compositions of four switchgrass populations. *Biomass and Bioenergy* 34(1): 48–53.

[7] David K, Ragauskas A-J (2010) Switchgrass as an energy crop for biofuel production: A review of its ligno-cellulosic chemical properties. *Energy Environ Sci* 3(9): 1182.

[8] Schneckenberger K, Demin D, Stahr K, Kuzyakov Y (2008) Microbial utilization and mineralization of ^{14}C glucose added in six orders of concentration to soil. *Soil Biol Biochem* 40(8): 1981–1988.

[9] Linn D-M, Doran J-W (1984) Aerobic and Anaerobic Microbial Populations in No-till and Plowed Soils1. *Soil Sci Soc Am J* 48(4): 794.

[10] Griffiths R-I, Whiteley A-S, O'Donnell A-G, Bailey M-J (2000) Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl Environ Microbiol* 66(12): 5488–5491.

[11] Buckley D-H, Huangyutitham V, Hsu S-F, Nelson T-A (2007) Stable isotope probing with ^{15}N achieved by disentangling the effects of genome G+C content and isotope enrichment on DNA density. *Appl Environ Microbiol* 73(10): 3189–3195.

[12] Manefield M, Whiteley A-S, Griffiths R-I, Bailey M-J (2002) RNA Stable isotope probing a novel means of linking microbial community function to phylogeny. *Appl Environ Microbiol* 68(11): 5367–5373.

[13] Birnie G-D (1978) Centrifugal separations in Molecular and cell biology. (Butterworth & Co Publishers Ltd, Boston)

[14] Edgar R-C (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10(10): 996–998.

[15] DeSantis T-Z, Hugenholtz P, Keller K, Brodie E-L, Larsen N, Piceno Y-M, *et al.* (2006) NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res* 34(suppl 2): W394–W399.

[16] Schloss P-D, Westcott S-L, Ryabin T, Hall J-R, Hartmann M, Hollister E-B, *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75(23): 7537–7541.

[17] Caporaso J-G, Kuczynski J, Stombaugh J, Bittinger K, Bushman F-D, Costello E-K, *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7(5): 335–336.

[18] Edgar R-C (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19): 2460–2461.

[19] Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, *et al.* (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41: D590–596.

[20] Nawrocki E-P, Kolbe D-L, Eddy S-R (2009) Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25(10): 1335–1337.

[21] Nawrocki E-P, Eddy S-R (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29(22): 2933–2935.

[22] Price M-N, Dehal P-S, Arkin A-P (2010) FastTree2 - approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5(3): e9490.

[23] Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71(12): 8228–8235.

[24] McMurdie P-J, Holmes S (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 8(4): e61217.

[25] Oksanen J, Blanchet F-G, Kindt R, Legendre P, Minchin P-R, O'Hara R-B, *et al.* (2015) vegan: Community Ecology Package.

[26] Anderson M-J (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 26(1): 32–46.

[27] Kembel S-W, Wu M, Eisen J-A, Green J-L (2012) Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Comput Biol* 8(10): e1002743.

[28] Matsen F-A, Kodner R-B, Armbrust E-V (2010) pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* 11: 538.

[29] Kembel S, Cowan P, Helmus M, Cornwell W, Morlon H, Ackerly D, *et al.* (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26: 1463–1464.

[30] Martiny A-C, Treseder K, Pusch G (2013) Phylogenetic conservatism of functional traits in microorganisms. *ISME J* 7(4): 830–838.

[31] Jombart T, Dray S (2010) adephylo: exploratory analyses for the phylogenetic comparative method.. *Bioinformatics* 26: 1907–1909.

[32] *Warning: citation key "altschul2990" is not in the bibliography database.*

[33] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421.

[34] Yarza P, Richter M, Peplies J, Euzeby J, Amann R, Schleifer K-H, *et al.* (2008) The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst Appl Microbiol* 31(4): 241–250.

[35] Love M-I, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12): 550.