

Video Article

Single-throughput Complementary High-resolution Analytical Techniques for Characterizing Complex Natural Organic Matter Mixtures

Malak M. Tfaily^{*1,2}, Rachel M. Wilson^{*3}, Heather M. Brewer¹, Rosalie K. Chu¹, Heino M. Heyman⁴, David W. Hoyt¹, Jennifer E. Kyle⁵, Samuel O. Purvine¹

¹Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory

²Department of Soil, Water and Environmental Science, University of Arizona

³Department of Earth Ocean and Atmospheric Sciences, Florida State University

⁴Bruker Daltonics Inc.

⁵Biological Sciences Division, Pacific Northwest National Laboratory

*These authors contributed equally

Correspondence to: Rachel M. Wilson at rmwilson@fsu.edu

URL: <https://www.jove.com/video/59035>

DOI: [doi:10.3791/59035](https://doi.org/10.3791/59035)

Keywords: Fourier transform ion cyclotron resonance mass spectrometry, high-resolution analytical techniques, peatlands, dissolved organic matter, microbial decomposition, environmental metabolomics, proteomics

Date Published: 12/14/2018

Citation: Tfaily, M.M., Wilson, R.M., Brewer, H.M., Chu, R.K., Heyman, H.M., Hoyt, D.W., Kyle, J.E., Purvine, S.O. Single-throughput Complementary High-resolution Analytical Techniques for Characterizing Complex Natural Organic Matter Mixtures. *J. Vis. Exp.* (), e59035, doi:10.3791/59035 (2018).

Abstract

Natural organic matter (NOM) is composed of a highly complex mixture of thousands of organic compounds which, historically, proved difficult to characterize. However, to understand the thermodynamic and kinetic controls on greenhouse gas (carbon dioxide [CO₂] and methane [CH₄]) production resulting from the decomposition of NOM, a molecular-level characterization coupled with microbial proteome analyses is necessary. Further, climate and environmental changes are expected to perturb natural ecosystems, potentially upsetting complex interactions that influence both the supply of organic matter substrates and the microorganisms performing the transformations. A detailed molecular characterization of the organic matter, microbial proteomics, and the pathways and transformations by which organic matter is decomposed will be necessary to predict the direction and magnitude of the effects of environmental changes. This article describes a methodological throughput for comprehensive metabolite characterization in a single sample by direct injection Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS), gas chromatography mass spectrometry (GC-MS), nuclear magnetic resonance (NMR) spectroscopy, liquid chromatography mass spectrometry (LC-MS), and proteomics analysis. This approach results in a fully-paired dataset which improves statistical confidence for inferring pathways of organic matter decomposition, the resulting CO₂ and CH₄ production rates, and their responses to environmental perturbation. Herein we present results of applying this method to NOM samples collected from peatlands; however, the protocol is applicable to any NOM sample (e.g., peat, forested soils, marine sediments, etc.).

Video Link

The video component of this article can be found at <https://www.jove.com/video/59035/>

Introduction

Globally, wetlands are estimated to contain 529 Pg of carbon (C), mostly as organic C buried in peat deposits¹. Currently, such peatlands act as a net C sink, sequestering 29 Tg C y⁻¹ in North America alone¹. However, environmental disturbance such as draining, fires, drought, and warmer temperatures can offset this C sink by increasing organic matter decomposition resulting in increased C losses via greenhouse gas (carbon dioxide [CO₂] and methane [CH₄]) production^{1,2}. Climate change may contribute to C loss if warmer temperatures or dryer conditions stimulate faster C decomposition by microorganisms. Alternatively, higher temperatures and air CO₂ concentrations may stimulate primary production to sequester more CO₂ as organic carbon (OC). To what extent and how fast that OC is then decomposed into CO₂ and CH₄ depends on the complex interactions between the electron donor substrates, the availability of electron acceptors, and the microorganisms that mediate the transformation. In many cases, the mechanisms are not well-characterized, thus their response to environmental perturbations is not well-constrained and it remains unclear what the net result of climate change will be on carbon balance in peatland ecosystems.

The complex nature of natural organic matter (NOM) has made even identifying the organic compounds present in the NOM mixtures historically difficult. Recent advances have greatly improved our ability to characterize compounds that traditionally and, to some extent continue to be, regarded as recalcitrant humic or fulvic compounds^{3,4,5}. We now understand that many of these compounds are actually microbially available and may be decomposed if a suitable terminal electron acceptor (TEA) is made available^{6,7}. Calculating the nominal oxidation state of the carbon (NOSC) for a compound provides a metric for predicting the potential for decomposition and the energy yield of the TEA required. However, it requires a molecular-level characterization of the organic matter⁷. NOSC is calculated from the molecular formula via the following

equation⁷: $NOSC = -((z + 4(\#C) + (\#H) - 3(\#N) - 2(\#O) + 5(\#P) - 2(\#S)) / (\#C)) + 4$, where z is the net charge. NOSC is correlated with the thermodynamic driving force⁸, wherein compounds with higher NOSC are easier to degrade, while compounds with lower NOSC require increasingly energetic TEAs in order to be reduced. Compounds with NOSC less than -2 require a high energy yielding TEAs such as O_2 , nitrate or Mn^{IV} , and cannot be degraded by commonly occurring lower energy yielding TEAs such as Fe^{III} or sulfate⁷. This is an important consideration in the waterlogged anoxic conditions found in wetlands where O_2 and other high energy yielding TEAs are scarce⁹ and therefore the degradation of lower NOSC compounds under these conditions are thermodynamically limited. Environmental perturbation can influence the thermodynamic state of the ecosystem through hydrologic changes that influence O_2 (the most energetic electron acceptor), changes in organic substrates and electron acceptors made available by primary production, and to a smaller extent by temperature. An important example of the temperature effects in wetland systems occurs with regard to the trade-off that occurs between homoacetogenesis (*i.e.*, acetate production from CO_2 and H_2) and hydrogenotrophic methanogenesis (*i.e.*, CH_4 production from CO_2 and H_2). At low temperatures it appears that homoacetogenesis is slightly favored, while warmer temperatures favor CH_4 production¹⁰. This temperature effect may have important implications for the response of ecosystems to changing climate, as CH_4 is a much stronger greenhouse gas than CO_2 ¹¹ and thus increasing production of CH_4 at the expense of CO_2 at warmer temperatures may contribute to a positive feedback with climate warming.

Peatlands produce globally significant quantities of CO_2 and CH_4 ⁶ via microbial respiration of naturally occurring organic matter. The NOSC of the organic carbon substrates determines the relative proportion of CO_2 : CH_4 produced which is a critical parameter because of the higher radiative forcing of CH_4 compared to CO_2 ¹¹, but also because modeling efforts have identified this ratio as a critical parameter for estimating C flux in peatlands¹². In the absence of terminal electron acceptors other than CO_2 , it can be shown by electron balance that organic C substrates with $NOSC > 0$ will produce CO_2 : $CH_4 > 1$, organic C with $NOSC = 0$ produces CO_2 and CH_4 in equimolar ratio, and organic C with $NOSC < 1$ will produce CO_2 : $CH_4 < 1$ ¹³. Decomposition of OC in natural ecosystems is mediated by microorganisms, so that even when degradation of a specific compound is thermodynamically feasible, it is kinetically limited by the activity of microbial enzymes and, under anoxic conditions, by the thermodynamic driving force (*i.e.*, NOSC)⁷. Until now it has been challenging to fully characterize the organic matter because the diversity of compounds present requires different complementary techniques for their characterization. Recent advances have closed the gap; using a suite of analytical techniques we can analyze a large range of organic compounds providing molecular-level characterization and, in some cases quantification, from small primary metabolites like glucose up to 800 Da poly-heterocycles. Previously such large complex molecules would have been characterized simply as lignin-like or tannin-like and assumed to have been recalcitrant. Molecular-level characterization, however, allows the calculation of NOSC for even these large complex molecules. These NOSC values are linearly correlated with the thermodynamic driving force allowing for an assessment of the quality of organic matter available for decomposition, which in many cases reveals that these complex molecules may actually be microbially degradable even under the anoxic conditions that prevail in wetlands.

Since introduction of O_2 allows organic matter of nearly all naturally observed NOSC values to be decomposed, herein we focus on changes in the organic matter and microbial proteomics which are likely to be the primary drivers in wetland (*i.e.*, limited O_2) systems. However, all the techniques that we will discuss can be applied to organic matter from any ecosystem. Commonly, bulk measurements based on optical and fluorescence analyses have been used to assess organic matter quality^{3,14}. When using bulk measurements such as these, however, fine details are lost as large numbers of molecules are categorized together under generic terms like humics or fulvics. The definitions of these categories are not well-constrained and, in fact, may vary from study to study making comparisons impossible. Further, bulk measurements do not provide the molecular detail necessary for calculating the thermodynamics governing the system and therefore fall short of truly assessing organic matter quality¹⁵.

Individual techniques such as Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS), nuclear magnetic resonance (NMR) spectroscopy, gas chromatography mass spectrometry (GC-MS), and liquid chromatography mass spectrometry (LC-MS) do provide such molecular-level detail. While each of these techniques presents its own limitations, they also bring their own strengths that can be leveraged in an integrated approach to achieve the fine molecular detail necessary for quantifying organic matter quality in a rigorous thermodynamic sense. GC-MS is useful for identifying critical small metabolites that are likely to have proximal influence on CO_2 and CH_4 production (*e.g.*, glucose, acetate, *etc.*); however, GC-MS requires verification against a standard and is therefore limited to already known compounds present in the database preventing identification of novel compounds. Furthermore, GC-MS is a semi-qualitative technique allowing inference about the changes in relative concentrations, but not providing the actual concentration information necessary for calculating Gibb's free energies for example. Finally, GC-MS requires derivatization of molecules prior to analysis which limits resolution to compounds smaller than ~ 400 Da and volatile alcohols are lost during the drying step.

One-dimensional (1D) 1H liquid state NMR allows highly quantitative characterization of small metabolites (including primary small molecular weight metabolites and volatiles like alcohols, acetate, acetone, formate, pyruvate, succinate, short-chained fatty acids, as well as a range of carbohydrates notoriously absent or compromised from MS-based methods) and their concentrations are particularly useful for calculating thermodynamic parameters. Yet, like GC-MS, 1D NMR of complex mixtures requires standardization relative to a database and therefore does not alone allow easy identification of novel compounds that are likely to be abundant in complex natural and changing ecosystems. Additionally, NMR is less sensitive than the MS-based techniques and therefore, quantitative metabolite profiling is achieved only above $1 \mu M$ using NMR systems equipped with helium-cooled cold-probes. Not widely appreciated, some NMR cold-probes are salt-tolerant and allow environmental mixture analysis in the presence of millimolar salt concentrations when used in smaller diameter (< 3 mm outer diameter) sample tubes¹⁶. However, a further complication of NMR is that high amounts of paramagnetic metals and minerals (*e.g.*, Fe and Mn above 1 - 3 wt%), which can be abundant in upland soils, can broaden spectral features and complicate interpretation of the NMR spectra. Using solid phase extraction (SPE) can aide in the interpretation of both NMR and MS-based metabolomics methods by reducing the mineral salts and increasing spectral quality.

FTICR-MS by direct injection is a highly sensitive technique capable of detecting tens of thousands of metabolites from a single sample, but it does not capture the critical small metabolites such as acetate, pyruvate, and succinate and is notoriously difficult to use for sugars and other carbohydrates¹⁷, nor does it provide quantitative information. However, unlike the other techniques, FTICR-MS excels at identifying and assigning molecular formula to novel compounds and therefore identifies the largest number of compounds providing more molecular information than any of the other described techniques. This is useful, because the molecular information provided by FTICR-MS (and other techniques) can be used to calculate NOSC which is related to the thermodynamic driving force governing the likelihood of certain reactions⁸ and their rates under certain conditions⁷. Furthermore, by coupling FTICR-MS with separation techniques, such as LC together with tandem MS, quantitative structural information can be attained, offsetting some of the disadvantages of this technique. LC-MS is useful for identifying lipid-like compounds

and other metabolites that are not well-characterized by any of the other methods. Coupling LC FTICR-MS or LC-MS with a fraction collector and collecting fractions of specific unknowns of interest for structural elucidation by two-dimensional (2D) liquid state NMR is the ideal situation for identifying and quantifying unknown compounds^{18,19}. However, this is a very time-consuming step that could be used if and when needed. Taken individually, each of these techniques provide a different snapshot of the organic matter, and by integrating them, we can achieve a more complete understanding than using any of the techniques in isolation.

While the thermodynamic considerations set the ultimate constraints on what transformations are possible in a system, organic matter decomposition is mediated by microorganisms whose enzyme activities control reaction rates. Thus, fully understanding the controls on organic matter decomposition and ultimately greenhouse gas (CO₂ and CH₄) production from wetlands requires an integrated omics approach to characterizing the microbial enzyme activities as well as the metabolites. In this article, we describe a method for achieving such a comprehensive analysis from a single sample using a sequential approach that results in a fully paired analysis. This approach expands on the metabolite, protein, and lipid extraction (MPLEX) protocol in which proteomics was coupled with GC-MS and LC-MS²⁰ to identify small metabolites, proteins, and lipids by incorporating quantitative metabolite information via NMR and identification of larger secondary metabolites via FTICR-MS. Slightly different to MPLEX, we begin the protocol with a water extraction and then use sequential extraction with increasingly non-polar solvents. All extractions are done on a single sample which conserves sample when volumes are limited or difficult to obtain and decreases experimental error introduced through variation among aliquots from heterogeneous sample matrices (e.g., soil and peat) or differences in storage conditions and duration.

Finally, by coupling the OM analyses with proteomics analyses of the microbial community, we can build metabolic networks that describe the pathways and transformation of organic matter decomposition. This allows us to test specific hypotheses about how perturbations to the system will influence ultimate CO₂ and CH₄ production through alteration to the available organic substrates, electron acceptors, and the microbial communities mediating the reactions via the activity of enzyme catalysts.

The overall goal of this method is to provide a single throughput protocol for analyzing metabolites, lipids, and microbial proteins from a single sample thereby creating a fully paired dataset for building metabolic networks while constraining analytical errors.

Protocol

1. Sequential Extraction of Organic Matter from Soil, Sediments, or Peat

1. Collect soil, sediments, or peat *via* coring and divide cores according to the hypothesis being tested (e.g., depth). Store samples in polytetrafluoroethylene coated containers and freeze at -80 °C for storage prior to analysis.
NOTE: Approximately 25 mg C is needed for this protocol. For peat (typically 45% C), 50 mg of dried peat is required. Larger amounts of sample may be needed for low organic samples like mineral or forested uplands soils depending on the C content (up to 5 g). Because extraction with organic solvents will pull any polyethylene glycol (PEG) into the extracts, which will negatively affect ionization during step 2.4, it is important to avoid allowing samples to contact plastic at any point during collection, storage or extraction.
2. When ready to analyze the samples, freeze dry to constant weight, then grind the samples in a high-speed ball mill using stainless steel grinding balls to homogenize and break up any aggregates.
NOTE: The protocol can be paused at this point and the material stored at -80 °C.
3. Using an ethanol-washed stainless-steel utensil, aliquot 50 mg of each of the dried samples into individual 2 mL glass vials. These samples will be sequentially extracted using a series of solvents to consecutively extract increasingly non-polar metabolites from each sample. Add 1 mL of distilled, degassed water (H₂O) to each sample, cap the vials and shake for 2 h on a shaker table.
4. After shaking allow the solutions to stand at room temperature (RT) for 20 min, then centrifuge at 15,000 x g for 30 min, decant and save the supernatant from each.
NOTE: These solutions will be injected into the FTICR-MS by direct injection.
5. Conduct a Folch extraction²¹ (also known as MPLEX²⁰) on the now water extracted residues by repeating steps 1.3 and 1.4 substituting 1 mL of a -20 °C 4:3 chloroform:methanol mixture for the water in step 1.3.
CAUTION: Chloroform and methanol are both highly flammable and toxic. Use appropriate personal protective equipment (PPE) to avoid skin contact and avoid open flame.
6. Carefully separate the two resulting solvent layers, which will be visually distinguishable, for separate analysis by FTICR-MS by using a separation funnel or simply remove the top layer by careful pipetting.
NOTE: The chloroform-containing fraction will be at the bottom while the methanol-containing fraction is less dense and will be on the top.
7. Dilute the chloroform extract (step 1.6) 1:1 in methanol and the water extract (step 1.4) 2:1 in methanol to improve the electrospray ionization (ESI) efficiency for FTICR-MS by direct injection.
NOTE: The methanol-containing fraction from step 1.7 does not need to be diluted further in methanol. The methanol layer will be run by direct injection on the FTICR-MS.

2. FTICR-MS Analysis

1. Calibrate the FTICR spectrometer by directly injecting 100 µL of a tuning solution (see **Table of Materials**) spanning a mass range of approximately 100 - 1300 Da into the FTICR-MS.
2. Prepare the Suwannee River Fulvic Acid standard (see **Table of Materials**) by making a 1 mg mL⁻¹ solution in ultrapure filtered water and then diluting the resulting solution to 20 µg mL⁻¹ in methanol.
3. Direct inject 23 µL of this final solution to the ESI source coupled to the FTICR spectrometer through a syringe pump set to a flow rate of 3.0 µL min⁻¹. Set needle voltage to +4.4 kV, Q1 to 150 m/z and glass capillary at 180 °C. Inspect resulting spectra using the analysis software (see **Table of Materials**) to confirm the quality of the data.

4. Use HPLC grade methanol before running samples and throughout the sampling to monitor carryover. Introduce 23 μL of each extract *via* direct injection to the ESI source coupled to the FTICR spectrometer through a syringe pump set to a flow rate of 3.0 $\mu\text{L min}^{-1}$. Set needle voltage to +4.4 kV, Q1 to 150 m/z and glass capillary at 180 $^{\circ}\text{C}$.
5. Adjust ion accumulation time (IAT) for each sample or group of samples to account for variation in C concentration. Typical values are between 0.1 to 0.3 s. Collect 144 scans for each sample, average the scans and then conduct an internal calibration using homologous CH_2 (i.e., 14 Da separation) series.
NOTE: After FTICR-MS analysis, the decision can be made to either combine the water and methanol extracted fractions in order to streamline the remaining steps or the fractions can be kept separate throughout subsequent steps. The advantages and disadvantages of each approach are described at length in the **Discussion**. If doing so, combine the water and methanol-extracted fractions.
6. Dry extracts using a concentrator and save the remainder of the extracts (~ 1 mL) for subsequent GC-MS (water, methanol or water + methanol), LC-MS (chloroform) and NMR (water + methanol) analysis.
NOTE: The protocol can be paused at this point and the material stored at -20 $^{\circ}\text{C}$.

3. FTICR-MS Data Processing

1. Co-align all sample peak lists for the entire dataset to reduce mass shifts and standardize peak assignments using Formularity software²² ahead of formula assignment.
2. Use the Formularity software²² to assign molecular formula using a signal to noise ratio greater than 7, mass measurement error < 1 ppm, and allowing C, H, O, N, S, and P while excluding all other elements.
3. If multiple candidate formulas are returned for a given mass (frequent above 500 Da) impose constraints consistent with the material being sampled. For example, in peat, typical constraints include: lowest mass error, lowest number of heteroatoms (N, S, P), and when present, P must be in oxidized form (i.e., there must be at least 4 O atoms for each P atom in the formula).

4. Chemical Derivatization for GC-MS

1. Prepare blank control samples of HPLC-grade hexane in GC-MS autosampler vials²³. Dissolve 100 mg fatty acid methyl esters (FAMES: C8 - C28) mixture retention time standard in 200 μL of hexane.
2. To protect carbonyl groups, add 20 μL of 30 mg mL^{-1} methoxyamine hydrochloride in pyridine to each of the methanol extracts and water extracts (or combined methanol/water extracts if using) from step 2.6, blanks and FAME calibration samples²³. Seal vials with caps.
CAUTION: Pyridine is both highly toxic and flammable. Wear appropriate PPE to prevent skin or eye contact and avoid open flame. In addition, pyridine volatilizes at RT causing harmful air contamination. Work only in well-ventilated spaces under a fume hood.
3. Vortex extracts for 20 s. Sonicate extracts for 60 s. Then, incubate extracts in a centrifuge at 37 $^{\circ}\text{C}$ for 90 min at 100 x g.
NOTE: An excessive amount of carbohydrates or salts can cause metabolites to crystallize after being dried down. Sonicating the samples will help to reconstitute the crystallized metabolites into the derivatizing reagent.
4. After incubation, add 80 μL of N-methyl-N-(trimethylsilyl)trifluoroacetamide with 1% trimethylchlorosilane to each sample, vortex extracts for 20 s. Sonicate extracts for 60 s and incubate extracts again in a centrifuge at 37 $^{\circ}\text{C}$ for 30 min at 100 x g.
5. Cool extracts to RT (20 - 24 $^{\circ}\text{C}$), then transfer into GC-MS autosampler vials.

5. GC-MS Analysis

1. Tune and calibrate MS according to vendor recommendations (see **Table of Materials**) before analysis to make sure the machine read MS data correctly. Check that helium gas pressure is within specified tolerance.
2. Separate polar metabolites on a GC column (30 m \times 0.25 mm \times 0.25 μm). Set the oven temperature protocol as follows: (1) initial temperature of 60 $^{\circ}\text{C}$ for 1 min, (2) ramp to 325 $^{\circ}\text{C}$ at a rate of 10 $^{\circ}\text{C min}^{-1}$, and (3) hold at 325 $^{\circ}\text{C}$ for 5 min.
3. Analyze extracts using a GC coupled to a single quadrupole MS. Set injection port temperature to a constant 250 $^{\circ}\text{C}$. Inject 1 μL of each derivatized extract in the splitless mode.

6. GC-MS Data Processing

1. Inspect all the data files to ensure that they were correctly captured. Pay attention to potential shifts with regards to internal standard retention times and intensities to confirm that the data was captured consistently throughout the analysis.
2. Convert the vendor specific MS data format to a general MS format if required. Process raw data files using MetaboliteDetector²⁴ calibrating retention indices based on the FAME internal standards. After aligning the retention times of all data files, continue with deconvolution and finally metabolite identification by matching retention indices and GC-MS spectra against the FiehnLib polar metabolite library²⁵.
3. Cross-check remaining unidentified metabolites against the NIST14 GC-MS library using spectral matching. Validate identifications individually to eliminate false identifications and reduce deconvolution errors.

7. Liquid State NMR Analysis

1. Dilute the remainder of the water extracts (~ 300 μL) by 10% (vol/vol) with a 5 mM 2,2-dimethyl-2-silapentane-5-sulfonate- d_6 internal standard. Alternatively, combine water and methanol extracts from step 1 then resubstitute in water. By doing so however, some of the volatile compounds may be lost during the freeze-drying step. Typical final sample volumes are in the range 180 - 300 μL .
2. Transfer the mixture into a high-quality 3 mm outer diameter (O.D.) borosilicate glass NMR tube.
3. Collect spectra using an NMR spectrometer (ideally at least 600 MHz) equipped with a 5 mm triple resonance salt-tolerant cold probe and a cold-carbon pre-amplifier.
4. Collect 1D ^1H spectra using a 1D nuclear Overhauser effect spectroscopy (NOESY) presaturation experiment at 298 K with a 4 s acquisition time, 1.5 s recycle delay, 65,536 complex points, and 512 scans.

5. Collect 2D ^1H - ^{13}C heteronuclear single-quantum correlation (HSQC) and ^1H - ^1H total correlation (TOCSY) spectra to help in assigning metabolites and validation.
6. Process, assign, and analyze all spectra using the NMR analysis software (see **Table of Materials**) to quantify intensities relative to the internal standard. Identify metabolites by matching chemical shift, J-coupling and intensity information in the samples against the library. NOTE: The library is further enhanced by custom targeted metabolite standards and metabolite additions made on a regular basis. The protocol can be paused at this point and the material stored at $-20\text{ }^{\circ}\text{C}$.

8. LC-MS Lipidomics Analysis

1. Rewet the dried chloroform extract generated during step 2.5 with 200 μL of methanol.
2. Inject 10 μL of each extract into an ultra-performance liquid chromatograph coupled to an Orbitrap mass spectrometer using a reversed phase charged surface hybrid column (3.0 mm \times 150 mm \times 1.7 μm particle size). Set a 34-min gradient (mobile phase A: acetonitrile/ H_2O (40:60) containing 10 mM ammonium acetate; mobile phase B: acetonitrile/isopropanol (10:90) containing 10 mM ammonium acetate) at a flow rate of 250 $\mu\text{L}/\text{min}$. Use both negative and positive ionization modes with higher-energy collision dissociation and collision induced dissociation.
CAUTION: Acetonitrile is an eye, skin, and respiratory irritant. Wear appropriate PPE. In addition, acetonitrile is combustible. Do not allow to contact oxidizing agents, toxic fumes may be produced during burning.
3. Upload the raw LC-MS/MS data files into LIQUID along with the target file (*i.e.*, list of > 25,000 lipid species) for the respective ionization mode (positive or negative). Process the raw file. Manually validate the resulting identifications by examining the MS/MS spectra for presence of diagnostic ions, if applicable, matching fragment ions (*e.g.*, fatty acyl chains), isotopic separation, mass ppm error of the precursors, and the retention time. Export resulting list of confidently identified lipids as a .tsv file.
NOTE: The protocol can be paused at this point and store the material at $-20\text{ }^{\circ}\text{C}$.

9. Proteomics Analysis

1. Extract proteins according to the MPLeX protocol²⁰ from the remainder of the methanol phase resulting from step 2.4, by washing the extract with 20 times the extract volume of additional cold ($-20\text{ }^{\circ}\text{C}$) methanol.
2. Use a 1 mL/50 mg silica-based sorbent (see **Table of Materials**) to condition C18 SPE columns with 3 mL methanol, 2 mL of 0.1% trifluoroacetic acid (TFA) in water, followed by the addition of the extract from step 9.1 at a rate no greater than 1 mL/min.
3. Following the addition of the sample, wash the column with 4 mL of 95:4.9:0.1 water:acetonitrile:TFA, then allow to dry. Place a 1.5 mL collection tube under the SPE column, and elute the sample with 1 mL of 80:19.9:0.1 methanol:water:TFA.
4. Concentrate the extracts to 100 μL under vacuum-assisted freeze dryer, then measure the protein concentration by bicinchoninic acid (BCA) colorimetric assay²⁶ at a wavelength of 562 nm.
5. Centrifuge extracts at 10,000 $\times g$ for 10 min at $4\text{ }^{\circ}\text{C}$. Discard the resulting supernatant and dry the remaining pellet under vacuum for 5 min. Resuspend the protein pellet in water to a final concentration of 0.1 μg peptide per μL .
6. Add dithiothreitol to a final concentration of 5 mM and incubate at $60\text{ }^{\circ}\text{C}$ for 30 min. Dilute 10-fold with 100 mM NH_4HCO_3 /8M urea solution and incubate at $37\text{ }^{\circ}\text{C}$ for 3 h in the presence of 1 mM CaCl_2 and porcine trypsin at a 1:50 enzyme to protein ratio.
NOTE: Protocol may be paused at this point and store material at $-20\text{ }^{\circ}\text{C}$.
7. Separate extracts via liquid chromatography using an exponential gradient of a 0.1% formic acid in water mobile phase (A) and a 0.1% formic acid in acetonitrile mobile phase (B) at 10 kpsi and 500 nL min^{-1} .
8. Introduce resulting eluent into an ESI-coupled mass spectrometer collecting spectra from 400-2000 m/z with 100,000 resolution at m/z 400 in a linear ion trap (LTQ) Orbitrap mass spectrometer.
9. Convert RAW spectra files to the mzML format using msConvert or ProteoWizard accepting all default parameters. Use the universal database search tool MSGFPlus to search resulting proteomes against a targeted protein database of protein sequences predicted from relevant metagenome assembled genomes.
10. Append common contaminants (*e.g.*, trypsin, keratin, albumin) and remove all exactly duplicated protein sequences to improve peptide-to-mass-spectrum match statistics. Evaluate the resulting MSGF spectral probability scores to determine which peptide-to-mass-spectrum match is best. Use the Q-Value from MSGFPlus to filter the entire data pile to 1% false discovery rate (FDR).

10. Metabolomics Analysis and Metabolic Network Building

1. Compile all metabolite molecular formula identified in steps 3, 6, 7, and 8 into a single database of metabolites present in the samples. Combine these metabolites with the Enzyme Commission (EC) or KEGG Orthology (KO) numbers of enzymes identified in step 9. Search this combined dataset against the KEGG²⁷ database, metabolic pathways section.
2. Manually evaluate results to identify most probable pathways and integrate into a metabolic model.

Representative Results

We performed the described complementary analysis protocol and compared peat with depth in the S1 bog in the Spruce and Peatlands Response Under Changing Environments (SPRUCE) site in Minnesota, USA. These results are compared to those from a permafrost bog and fen from northern Sweden to show how sites may vary in metabolite and enzyme activities. We identified 3,312 enzymes in the proteomics analysis. An analysis of the enzymes activities with depth reveals that the number of enzymes declines sharply between 15 cm and 45 cm in the SPRUCE bog (**Figure 1**).

While the proteomics results indicate which proteins are expressed, the metabolic data show which reactions are actually occurring. Overall, we identified 67,040 metabolites (including lipids) in all of the peat samples from the combination of FTICR-MS, NMR, GC-MS, and LC-MS analyses.

Of these, we were able to assign molecular formula to 15,385 compounds (**Figure 2**). The combined metabolic data spans a range of oxidation states, masses, and compound classes.

This is typically visualized through the use of a van Krevelen diagram in which the atomic H/C ratios of identified formula are plotted against their atomic O/C ratios²⁸. We have included an added dimension to the typical 2D format, depicting NOSC by means of color coding the symbols representing individual formula (**Figure 3** and **Figure 4**). With increasing depth in the SPRUCE bog, there is an increase in the total number of large secondary metabolites identified by FTICR-MS, specifically in highly condensed (lower left of the van Krevelen plot) and hydrogenated formulae (top of the plot) (**Figure 4**). Over the same depths there is a decrease in the number of small highly energetic compounds identified by GC-MS and in the lipids (**Figure 5**). This could suggest either that the lipids and small metabolites are consumed in the surface peat before reaching the deeper depths or that decomposition rates in the deep peat are faster than in the surface such that downward advecting compounds are rapidly consumed. Differentiating between these two competing hypotheses requires a process-level understanding of the C cycling at the site. Such a process-level understanding can only be gained by coupling the metabolomics and proteomics datasets. We accomplish this by cross-validating the combined FTICR-MS, GC-MS, LC-MS, and NMR identified metabolites against the KEGG database. In doing so, we find that the compounds we identified are involved in common metabolic pathways such as tricarboxylic acid (TCA) cycle, glycolysis, and sugar metabolism. Since individual metabolites can be involved in multiple pathways confirmation with the enzymes increases our confidence in assigning pathways (**Figure 6**).

Through this mapping we find evidence of sucrose and starch metabolism in the surface peat, while in the deeper depths pyruvate and other fermentation products build up (**Figure 6**). These results are consistent with the first hypothesis that sugars and other energetic metabolites are degraded in the surface peat and do not reach the deeper peat depths. As can be seen in **Figure 5**, sugars (NOSC = 0) and amino acids ($0 < \text{NOSC} < 1$) are consumed in the surface peat, while lipids ($-2 < \text{NOSC} < -1$) appear to accumulate with depth. This is consistent with expectations based on NOSC values that higher NOSC compounds are more readily degraded while lower NOSC compounds persist in the highly anaerobic (*i.e.*, TEA-limited) conditions of subsurface peat. This approach also successfully distinguishes between soil types from different environments. For example, soil organic matter from boreal peatland appears to be compositionally different than permafrost peatland, as well as within fen and bog within the permafrost region. These results are consistent with a previous study that showed differences in site geochemistry between these habitats²⁹, suggesting that site geochemistry have a big effect on microbial degradation of C below ground.

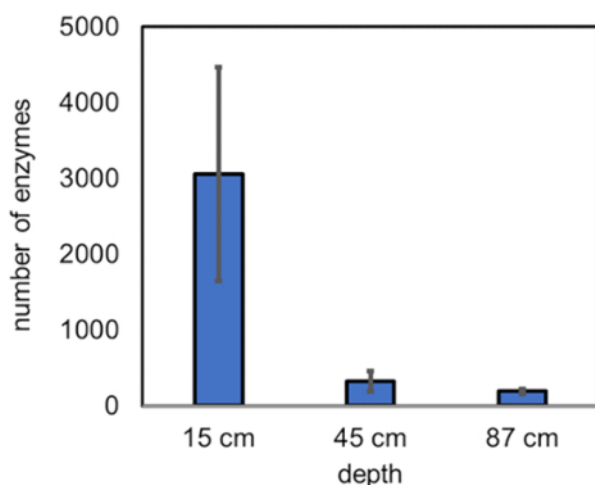


Figure 1: Proteomics analysis indicates strong depth stratification in the number of enzymes. Bars indicate averages and error bars indicate one standard deviation of abundances. This suggests that microorganisms are most active in the peat surface. [Please click here to view a larger version of this figure.](#)

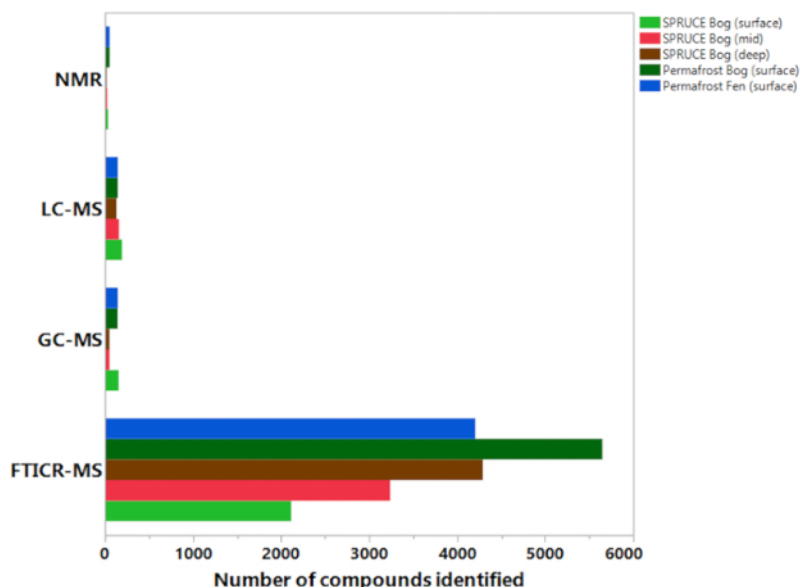


Figure 2: The number of compounds identified by each technique in the surface peat (< 30 cm) of each habitat as well as the mid (45 cm) and deep (87 cm) peat from the SPRUCE bog. [Please click here to view a larger version of this figure.](#)

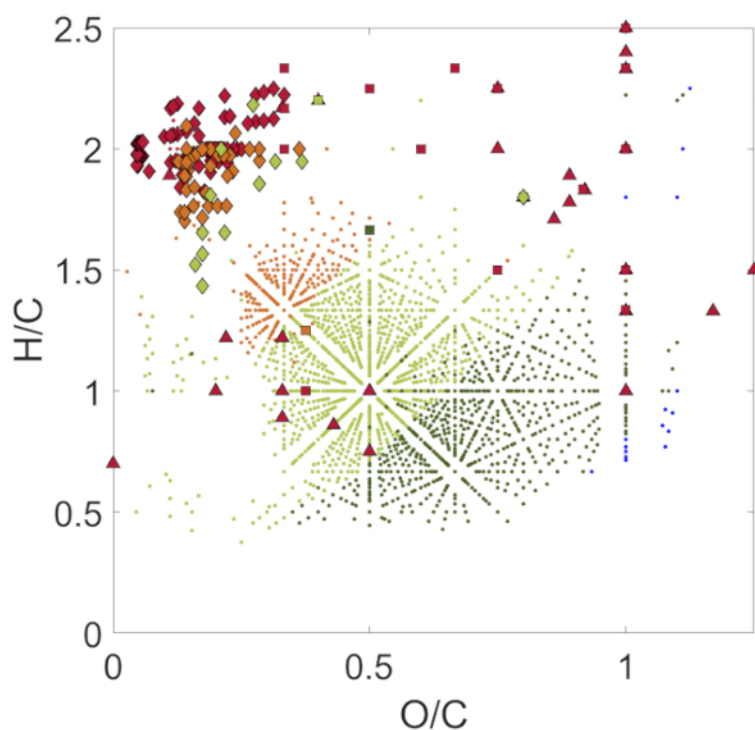


Figure 3: The van Krevelen diagram (atomic H/C vs. O/C of each identified molecular formula) for the surface (15 cm) SPRUCE bog depth to demonstrate the coverage of compounds characterized by each technique. FTICR-MS allows us to identify the largest number of compounds (small circles), while GC-MS (triangles) is good for differentiating and identifying sugars ($H/C = 2$ and $O/C = 1$). NMR (squares) spectroscopy provides quantitative information on energetically important compounds such as sugars, amino acids, pyruvate, etc. LC-MS (diamonds) provides information on lipids. [Please click here to view a larger version of this figure.](#)

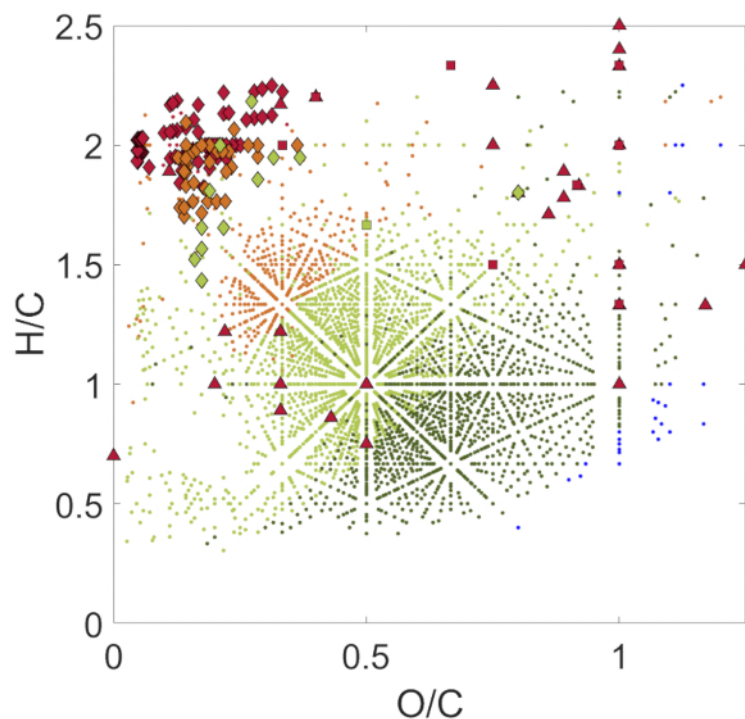


Figure 4: The van Krevelen diagram (atomic H/C vs. O/C of each identified molecular formula) for the deep (87 cm) SPRUCE bog depth to demonstrate the coverage of compounds characterized by each technique. [Please click here to view a larger version of this figure.](#)

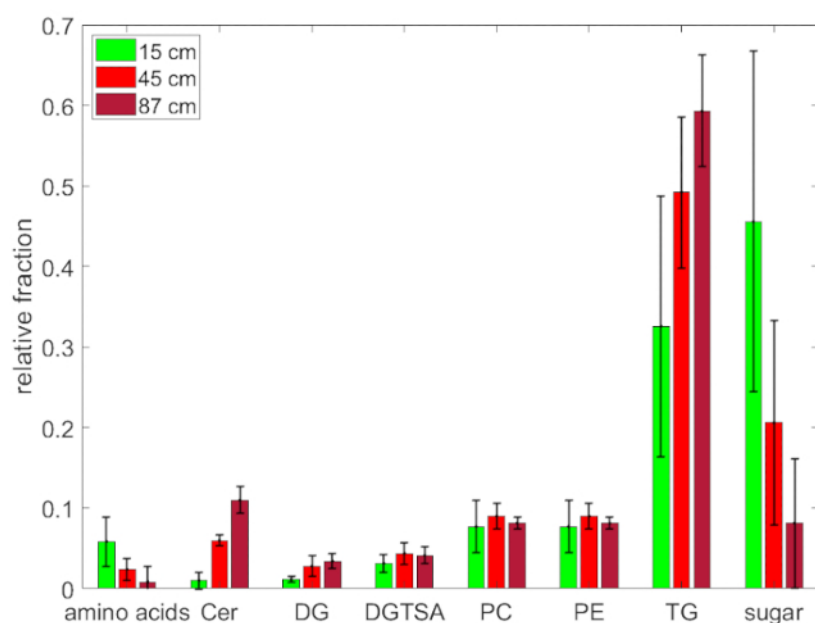


Figure 5: The relative fraction of different chemical classes identified via the various techniques in the different depths. Bars are plotted as averages for each depth \pm one standard deviation. Classes plotted include amino acids, sphingolipids (Cer), glycerophosphocholines (PC), phosphoethanolamines (PE), diacylglyceroltrihomoserine (DGTSA), diacylglycerol (DG), triacylglycerols (TG), and sugars. [Please click here to view a larger version of this figure.](#)

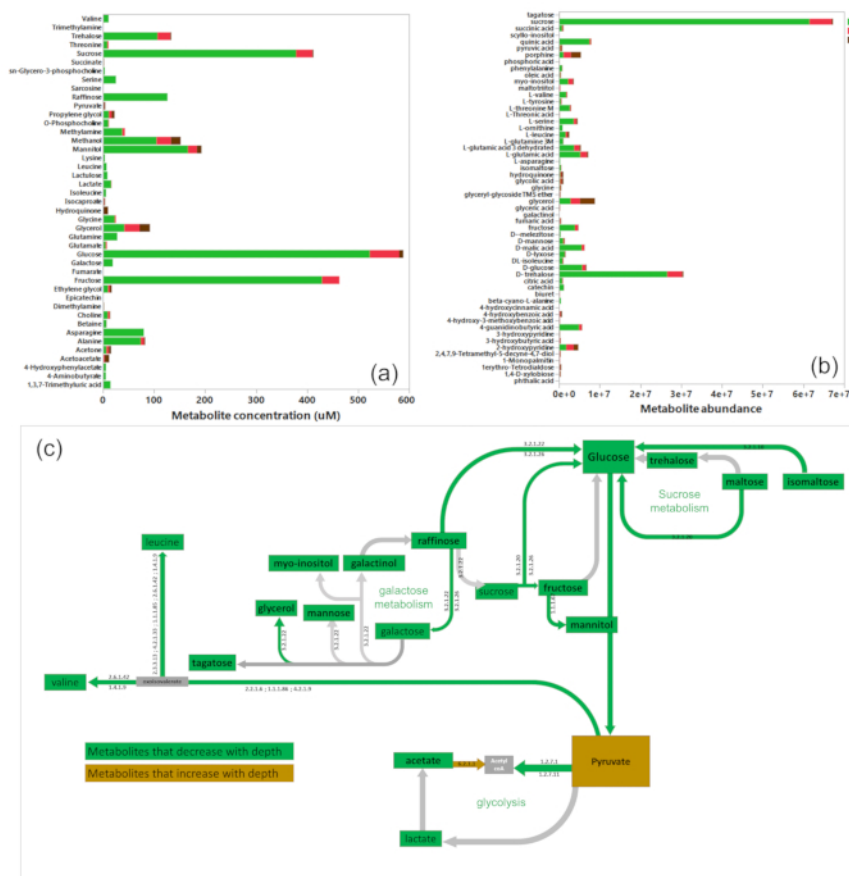


Figure 6: Combining results to create a metabolic network. Depth distributions of metabolites identified by NMR (a) and GC-MS (b), and a simplified metabolic map (c) showing a select number of transformations identified at SPRUCE bog. Green boxes indicate metabolites that decrease with depth, brown boxes indicate metabolites that increase with depth. Green connecting arrows indicate enzyme identified in our dataset that mediate the indicated transform (enzyme EC numbers indicated next to arrows). Gray connecting arrows indicate transforms for which enzymes were not identified in our dataset. [Please click here to view a larger version of this figure.](#)

Discussion

The single-throughput, fully-coupled analysis stream used to characterize metabolites and the proteome provides insights into the pathways by which C cycling is occurring in a complex ecosystem. Soil and peat are heterogeneous matrices, and therefore, one of the critical steps of this method occurs in the earliest steps in ensuring that the starting peat or soil material is highly homogenous. It is preferable to grind the sample well as aggregates can reduce extraction efficiency. This is a particular problem for aggregated soils and soils with low C and high mineral contents that may require the use of a stainless-steel ball grinder to adequately homogenize. Because spatial heterogeneity of soils and peat within an experimental site may be high, biological replicates are highly recommended.

This method utilizes three solvents: water, methanol, and chloroform, which bias the types of compounds it is possible to extract. In principle, this sequential extraction method should solubilize compounds covering a wide polarity range. However, these solvents are not optimized for extracting organo-mineral complexes or highly stabilized complexes. If such compounds are of interest, harsher solvents such as strong acids and bases are preferred, but harsher extraction processes could alter the chemistry of the samples. Incorporating such solvents at the end of the extraction sequence may minimize this effect. Additionally, chloroform will inactivate clinically important soil and peat bacterial and viral pathogens, and other pathogenic disease-causing biological agents by dissolving cell membrane lipids. Thus, incorporating chloroform in the extraction protocol will reduce the risks involved in biology studies of samples potentially infected by pathogens from different regions of the world. If there are concerns about dead microbial cells, debris, or particulates following extraction, filtering the extracts through a 0.2 µm glass fiber filter is recommended.

To streamline the process, the water and methanol extracts could be combined for GC-MS analysis during step 4.1. However, these extracts must remain separated for FTICR-MS analysis due to ionization efficiency issues with the ESI source. The advantage of running the combined extract on the GC-MS is that more metabolites (covering a larger range of polarity) will be identified. The disadvantage of combining the extracts at this point is that one of the major metabolites important in microbial processing of organic matter is methanol. Traces of methanol will always remain in the combined sample, even after drying, so if methanol is a metabolite of interest, the water extract must be run separately. Having good controls with no analytes will also help in identifying potential contamination in the samples.

In step 7.1 of preparing extracts for NMR analysis, as in the GC-MS analysis, either the water extract alone or the combined water and methanol extracts can be used. The disadvantages of doing this are similar to those enumerated in the preceding paragraph. The advantage of combining the methanol and water extracts is that some of the less polar metabolites that are solubilized in the methanol fraction will be identified. However, because of the freeze-drying step, many more volatile compounds will be lost. This is a particular disadvantage if quantifying volatile fatty acids is a critical component of the experiment.

In this procedure for identifying proteomics, only fully tryptic peptides are searched, thus endogenous peptidase activity and in-source fragmentations will be missed. On the other hand, oxidized methionine may be considered as a post-translational modification for the peptide candidates as this modification commonly occurs during sampling processing and handling. Quantification of enzymes using peptide elution areas can be done, but is outside the scope of this project.

Many recent advances in the analysis of NOM and microbial parameters are providing a wealth of techniques for understanding organic C cycling. By combining these techniques in a single streamlined protocol, we gain a novel view of the processes that are taking place. Coupling the proteomics analysis with the metabolic analyses provides compelling corroborative evidence that a reaction is actually occurring. Metabolic analysis alone is limited in that it informs us which metabolites are higher or lower in concentration in comparisons among sites or after a treatment effect, but can't discern the causes for those changes. For example, we identified declining sugar concentrations with depth in the bog, but without further information it is unclear whether the decline with depth is due to slow inputs or fast degradation in the deep peat. Analysis of the proteome and the declining enzyme expression with depth allows us to reject the second hypothesis. Rather our results are consistent with fast sugar degradation in the surface peat limiting inputs to the deeper depths. These particular insights are only possible when all techniques are considered in tandem as each one provides a unique, but critical piece of the C cycling puzzle.

Disclosures

The authors have nothing to disclose.

Acknowledgements

We would like to thank J.P. Chanton, J.E. Kostka, and M.M. Kolton for assistance with collecting peat samples. Portions of this work were conducted at the Environmental Molecular Sciences Laboratory, a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research. PNNL is operated by Battelle for the DOE under Contract DE-AC05-76RL01830. This work was supported by the U.S. Department of Energy, Office of Science, and Office of Biological and Environmental research (grants: DE-AC05-00OR22725, DE-SC0004632, DESC0010580, DE-SC0012088, and DE-SC0014416).

References

1. Bridgman, S. D., Megonigal, P. J., Keller, J. K., Bliss, N. B., Trettin, C. The carbon balance of North American wetlands. *Wetlands*. **26** (4), 889-916 (2006).
2. Wilson, R. M., et al. Greenhouse gas balance over thaw-freeze cycles in discontinuous zone permafrost. *Journal of Geophysical Research: Biogeosciences*. **122** (2), 387-404 (2017).
3. Broder, T., Knorr, K. H., Biester, H. Changes in dissolved organic matter quality in a peatland and forest headwater stream as a function of seasonality and hydrologic conditions. *Hydrology and Earth System Sciences*. **21** (4), 2035-2051 (2017).
4. Ejarque, E., et al. Quality and reactivity of dissolved organic matter in a Mediterranean river across hydrological and spatial gradients. *Science of The Total Environment*. **599**, 1802-1812 (2017).
5. Valenzuela, E. I., et al. Anaerobic methane oxidation driven by microbial reduction of natural organic matter in a tropical wetland. *Applied and Environmental Microbiology*. **83** (11) e00645-e00617 (2017).
6. Lehmann, J., Kleber, M. The contentious nature of soil organic matter. *Nature*. **528** (7580), 60-68 (2015).
7. Keiluweit, M., Nico, P. S., Kleber, M., Fendorf, S. Are oxygen limitations under recognized regulators of organic carbon turnover in upland soils? *Biogeochemistry*. **127** (2-3), 157-171 (2016).
8. LaRowe, D. E., Van Cappellen, P. Degradation of natural organic matter: A thermodynamic analysis. *Geochimica et Cosmochimica Acta*. **75** (8), 2030-2042 (2011).
9. Wilson, R. M., et al. Hydrogenation of organic matter as a terminal electron sink sustains high CO₂: CH₄ production ratios during anaerobic decomposition. *Organic Geochemistry*. **112**, 22-32 (2017).
10. Ye, R., Jin, Q., Bohannan, B., Keller, J. K., Bridgman, S. D. Homoacetogenesis: A potentially underappreciated carbon pathway in peatlands. *Soil Biology and Biochemistry*. **68**, 385-391 (2014).
11. Neubauer, S. C., Megonigal, J. P. Moving beyond global warming potentials to quantify the climatic role of ecosystems. *Ecosystems*. **18** (6), 1000-1013. (2015).
12. Ma, S., et al. Data-Constrained Projections of Methane Fluxes in a Northern Minnesota Peatland in Response to Elevated CO₂ and Warming. *Journal of Geophysical Research: Biogeosciences*. **122** (11), 2841-2861 (2017).
13. Conrad, R. Contribution of hydrogen to methane production and control of hydrogen concentrations in methanogenic soils and sediments. *Federation of European Microbiological Societies Microbiology Ecology*. **28** (3), 193-202 (1999).
14. Cunada, C. L., Lesack, L. F. W., Tank, S. E. Seasonal dynamics of dissolved methane in lakes of the Mackenzie Delta and the role of carbon substrate quality. *Journal of Geophysical Research: Biogeosciences*. **123** (2), 591-609 (2018).
15. Wilson, R. M., Tfaily, M. M. Advanced molecular techniques provide new rigorous tools for characterizing organic matter quality in complex systems. *Journal of Geophysical Research: Biogeosciences*. **123** (6), 1790-1795 (2018).
16. Borton, M. A., et al. Coupled laboratory and field investigations resolve microbial interactions that underpin persistence in hydraulically fractured shales. *Proceedings of the National Academy of Sciences*. **115** (28), E6585-E659 (2018).

17. Tang, K., Page, J. S., Smith, R. D. Charge competition and the linear dynamic range of detection in electrospray ionization mass spectrometry. *Journal of the American Society for Mass Spectrometry*. **15** (10), 1416-1423 (2004).
18. Boiteau, R. M., *et al.* Structure Elucidation of Unknown Metabolites in Metabolomics by Combined NMR and MS/MS Prediction. *Metabolites*. **8** (1), 8 (2018).
19. Walker, L. R., *et al.* Unambiguous Metabolite Identification in High-throughput Metabolomics by Hybrid 1D NMR/ESI MS Approach. *Magnetic Resonance in Chemistry*. **54** (12), 998-1003 (2016).
20. Nicora, C. D., Burnum-Johnson, K. E., Nakayasu, E. S., Casey, C. P., White III, R. A., Roy Chowdhury, T., Kyle, J. E., Kim, Y. M., Smith, R. D., Metz, T. O., Jansson, J. K., Baker, E. S. The MPLEx Protocol for Multi-omic Analyses of Soil Samples. *J. Vis. Exp.* (135), e57343 (2018).
21. Folch, J., Lees, M., Sloane-Stanley, G. H. Extraction of fatty acid. *Journal of Biological Chemistry*. **226**, 497-509 (1957).
22. Tolic, N., *et al.* Formularity: software for automated formula assignment of natural and other organic matter from ultrahigh-resolution mass spectra. *Analytical Chemistry*. **89** (23), 12659-12665 (2017).
23. Kim, Y. M., *et al.* Diel metabolomics analysis of a hot spring chlorophototrophic microbial mat leads to new hypotheses of community member metabolisms. *Frontiers in microbiology*. **6**, 209 (2015).
24. Hiller, K., *et al.* MetaboliteDetector: comprehensive analysis tool for targeted and nontargeted GC/MS based metabolome analysis. *Analytical Chemistry*. **81** (9), 3429-3439 (2009).
25. Kind, T., *et al.* FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Analytical Chemistry*. **81** (24), 10038-10048 (2009).
26. Kyle, J. E., *et al.* LIQUID: an-open source software for identifying lipids in LC-MS/MS-based lipidomics data. *Bioinformatics*. **33** (11), 1744-1746 (2017).
27. Kanehisa, M. Enzyme annotation and metabolic reconstruction using KEGG. *Protein Function Prediction: Methods and Protocols*. **1611**, 135-145 (2017).
28. Van Krevelen, D. W. Graphical-statistical method for the study of structure and reaction processes of coal. *Fuel*. **29**, 269-284 (1950).
29. Hodgkins, S. B., *et al.* Changes in peat chemistry associated with permafrost thaw increase greenhouse gas production. *Proceedings of the National Academy of Sciences*. **111** (16), 5819-5824 (2014).