1
2
3
4
5
6 **Untargeted soil metabolomics methods for analysis of extractable organic matter**
7
8

9 Tami L. Swenson, Stefan Jenkins, Benjamin P. Bowen, Trent R. Northen*
10
11
12
13
14
15
16

17 *Life Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley,*
18 *CA 94720, USA*

19 **Corresponding author: E-mail: TRNorthen@lbl.gov; Fax: +1 (510) 486-4545; Tel: +1 (510)*
20 *486-5240*

ABSTRACT

The cycling of soil organic matter (SOM) by microorganisms is a critical component of the global carbon cycle but remains poorly understood. There is an emerging view that much of SOM, and especially the dissolved fraction (DOM), is composed of small molecules of plant and microbial origin resulting from lysed cells and released metabolites. Unfortunately, little is known about the small molecule composition of soils and how these molecules are cycled (by microbes or plants or by adsorption to mineral surfaces). The water-extractable organic matter (WEOM) fraction is of particular interest given that this is presumably the most biologically-accessible component of SOM. Here we describe the development of a simple soil metabolomics workflow and a novel spike recovery approach using ^{13}C bacterial lysates to assess the types of metabolites remaining in the WEOM fraction. Soil samples were extracted with multiple mass spectrometry-compatible extraction buffers (water, 10 mM K_2SO_4 or NH_4HCO_3 , 10-100% methanol or isopropanol/methanol/water [3:3:2 v/v/v]) with and without prior chloroform vapor fumigation. Profiling of derivatized extracts was performed using gas chromatography/ mass spectrometry (GC/MS) with 55 metabolites identified by comparing fragmentation patterns and retention times with authentic standards. As expected, fumigation, which is thought to lyse microbial cells, significantly increased the range and abundance of metabolites relative to unfumigated samples. To assess the types of microbial metabolites from lysed bacterial cells that remain in the WEOM fraction, an extract was prepared from the soil bacterium *Pseudomonas stutzerii* RCH2 grown on ^{13}C acetate. This approach produced highly labeled metabolites that were easily discriminated from the endogenous soil metabolites. Comparing the composition of the fresh bacterial extract with what was recovered following a 15 min incubation with soil revealed that only 27% of the metabolites showed >50% recovery in the WEOM. Many,

especially cations (polyamines) and anions showed <10% recovery. These represent metabolites that may be inaccessible to microbes in this environment and would be most likely to accumulate as SOM presumably due to binding with minerals and negatively-charged clay particles. This study presents a simple untargeted metabolomics workflow for extractable organic matter and an approach to estimate microbial metabolite availability in soils. These methods can be used to further our understanding of SOM and DOM composition and examine the link between metabolic pathways and microbial communities to terrestrial carbon cycling.

Keywords: soil organic matter, metabolomics, gas chromatography/ mass spectrometry, fumigation

1. INTRODUCTION

Microbial products are an important source of soil organic matter

Over two-thirds of carbon in the terrestrial biosphere is stored as soil organic matter (SOM) originating from plant, animal and microbial sources (Johnston et al., 2004). Microorganisms are a critical component in soil as they are actively involved in the cycling of SOM (*i.e.* production of microbial products and metabolic decomposition) (Schmidt et al., 2012). The component of SOM that is accessible to microbial processing is the soluble fraction (Gregorich et al., 2000; Blagadotsky et al., 2010), referred to here as dissolved organic matter (DOM) and is physically defined as what can pass through a 0.45 μm filter. As with SOM, a significant portion of this complex pool of metabolites is thought to be derived from soil microbes with the flow of this carbon fraction mediating microbial biomass turnover (Baldock and Nelson, 2000; Gregorich et al., 2000; Kalbitz et al., 2000).

DOM is in a constant state of flux by soil microbes (Schimel and Schaeffer, 2012; Xu et al., 2014, Schmidt et al., 2012). However, microbial decomposition, and therefore the resultant composition of DOM, depends on the availability of small molecule substrates *in situ*. A number of biotic and abiotic factors affect microbial access to these molecules (such as leaching or diffusion), but one of the most critical factors is adsorption to mineral surfaces (Kalbitz et al., 2000). The types of biologically accessible substrates likely depend on the specific interactions between DOM components and mineral surfaces. In turn, the composition of the available substrate pool will be a major determinant of microbial community structure and metabolic activities (Judd et al., 2006). Hence, elucidation of the chemical composition of microbe-accessible substrates is a critical step toward understanding the complex dynamics of soil nutrients and microbial communities.

Traditional soil extraction methods and DOM analyses

To analyze DOM, a standard method is to obtain the water-extractable organic matter (WEOM) fraction by using aqueous extractants, while microbial biomass is measured by fumigating soil with chloroform vapors to release intracellular metabolites prior to extraction (Brookes, et al., 1985; Vance et al., 1987). Comparisons are then made to an extraction of unfumigated soil in order to estimate the labile or microbe-accessible fraction. Given technical limitations on measuring its molecular composition, DOM has typically been quantified by oxidizing or combusting the soil sample and analyzing its elemental composition to determine dissolved organic (or inorganic) carbon or total dissolved nitrogen (Jones and Willett, 2006). In some cases, DOM and SOM are resolved into more specific classes of metabolites such as neutral sugars, amino sugars, amino acids, fatty acids and other biomolecules using colorimetric methods or compound-specific derivatization followed by gas chromatography/ mass spectrometry (GC/MS) (Amelung et al., 2008; Kakumanu et al., 2013). Such groupings into broad classes of biomolecules essentially preclude linkage with microbial genomics since biochemical specificity requires knowledge of molecular composition.

The use of untargeted metabolomics to understand DOM composition and availability

Untargeted metabolomics is a rapidly-growing and robust method that has become an important approach in biomedical science by providing comprehensive data-driven metabolism analyses of complex extracts (Fiehn, 2002; Garcia et al., 2008; Baran et al., 2009; 2013). As mass spectrometry (MS) instruments such as Fourier Transform Ion Cyclotron Resonance (FTICR/MS) (Hirai et al., 2004), capillary electrophoresis (CE/MS) (Soga et al., 2003; Edwards

et al., 2006), and liquid chromatography (LC/MS) (Baran et al., 2011; 2013) are commonly used, they are not widely available. In contrast, GC/MS is a widely-used and generally more-accessible instrument to microbiologists and soil scientists due its low operational cost, availability of curated metabolite spectral databases, broad analytical scope with good coverage of metabolite classes (carbohydrates, alcohols, sterols, amino acids, fatty acids, *etc*), as well as widespread application to phospholipid fatty acid analysis (Buyer and Sasser, 2012). It would therefore be desirable to develop workflows for GC/MS-based soil metabolomics for examination of extractable organic matter composition (Fiehn et al., 2000; Roessner et al., 2000; Koek et al., 2006; Lee et al., 2012).

Metabolomics approaches are beginning to have a major impact on improving our understanding of marine and freshwater communities and the significant role DOM plays in these environments. While the majority of marine metabolomics studies have involved controlled cultures, metabolite profiling has been successful with marine bacteria, microalgae, macroalgae and animals (Mopper et al., 2007; Minor et al., 2014). Metabolomics analyses of spent media from these systems have shed light on unique chemical defense mechanisms and production of secondary metabolites (Goulitquer et al., 2012). Marine metabolomics has also contributed to our understanding of DOM lability and recalcitrance as a function of microbial carbon pumps in aquatic systems (Jiao et al., 2010). Studies such as these are allowing us to grasp the value of DOM in shaping microbial communities and to begin to extrapolate its role in global climate change. Unfortunately, the field of soil metabolomics (and our understanding of SOM) lags behind, but is emerging as an equally important area of study.

Soils present many challenges to untargeted metabolomics methods

123 Recently a few papers have emerged relating to metabolomics of soils, many pertaining
124 to the production of osmolytes during drying and re-wetting conditions (Kakumanu et al., 2013;
125 Warren, 2013; 2014; Jones et al., 2014). FTICR/MS is one of the most established methods for
126 DOM analysis and routinely resolves thousands of unique chemical formulas. However, these
127 methods are typically used to discriminate soils and soil treatments based on changes in
128 abundance and composition of these chemical formulas (Hockaday et al., 2006; Ohno et al.,
129 2014). Understanding the soil biochemistry requires identifying the soil metabolites such that
130 they can be linked to enzymes and microorganisms. Towards this goal, a recent pioneering study
131 by Warren describes important methods for identifying and quantifying soil metabolites using
132 CE/MS and GC/MS (Warren, 2014). Findings from this study provide key insights into the
133 drying response of soil bacteria in terms of osmolyte production and accumulation.

134 The comparatively sparse literature on metabolomics of soils versus marine systems is
135 likely attributable to the unique challenges soils present for untargeted metabolomics. To
136 evaluate the chemical composition of SOM or DOM, methods typically require the development
137 of analytical approaches similar to the classical soil extraction methods involving chloroform
138 fumigation followed by extraction with buffers containing high concentrations of salts (*e.g.* 500
139 mM K₂SO₄) (Brookes et al., 1985; Vance et al., 1987; Tate et al., 1988; Murage and Voroney,
140 2007; Makarov et al., 2013). Unfortunately the salt content of the resultant samples complicate
141 metabolite analysis because of salt crystal formation during dry-down and ion suppression during
142 electrospray ionization mass spectrometry analysis (Annesley, 2003). To overcome these
143 challenges and setbacks, optimizing detection of a large range of soil metabolites by GC/MS
144 requires careful consideration of numerous variables during method development including
145 extractant selection, extraction time, sample concentration and derivatization methods.

Untargeted global metabolite profiling with GC/MS

As is evident, classical soil extraction methods involve long extraction times followed by compound-specific analyses. However, as efforts aimed at understanding the effects of climate change on soil will inevitably increase, there is a critical need to elucidate the small molecule composition of soils, their dynamics and microbial substrate availability. Our initial goal here is to establish a rapid and simple workflow to measure the small molecule composition of WEOM, the most representative fraction of labile SOM (Ohno et al., 2014), using GC/MS. This untargeted global soil profiling is a means to obtain unbiased investigations of soil metabolites in a qualitative rather than quantitative manner. To develop a standardized method for extractable organic matter, we used conventional soil extraction approaches and compared metabolite pools that are extracted from unfumigated and chloroform-fumigated soils using water with and without a series of additives: K_2SO_4 , NH_4HCO_3 , methanol or isopropanol. Another major focus is to evaluate compound-specific soil adsorption (and therefore substrate accessibility) of microbial metabolites. In attempt to extrapolate to *in situ* conditions, ^{13}C -labeled lysates from a soil bacterium were spiked into soil then analyzed for the components that remained in the WEOM fraction.

Overall, this work lays a technical foundation for soil metabolomics including detailed examination of the challenges that must be carefully considered when performing untargeted soil metabolite profiling. Given the wide-spread availability of GC/MS instrumentation, this workflow including stable isotope approaches has the potential to enable a wide range of researchers to characterize metabolites in soils ultimately improving our understanding of metabolic processes within soils.

2. METHODS

2.1. Chemicals

Ethanol-free chloroform (CAS 67-66-3), LC/MS Chromasolv isopropanol (CAS 67-63-0), 2-amino-3-bromo-5-methylbenzoic acid (ABMBA; CAS 13091-43-5), pyridine (CAS 110-86-1), 98% methoxyamine hydrochloride (CAS 593-56-6), sodium acetate (CAS 127-09-3) and ¹³C₂-sodium acetate (99 atom %, CAS 56374-56-2) were from Sigma (St. Louis, MO). LC/MS-grade water and LC/MS-grade methanol (CAS 67-56-1) were from Honeywell Burdick & Jackson (Morristown, NJ) and N-methyl-N-trimethylsilyltrifluoroacetamide (MSTFA) containing 1% trimethylchlorosilane (TMCS) was from Restek (Bellafonte, PA).

2.2. Soil collection, characterization and fumigation

Soil was collected from the Angelo Coast Range Reserve in Mendocino County, CA in September 2013, prior to the first autumn rainfall. Dominant soils from this region are alfisols, inceptisols, and ultisols with mixed mineralogy and a sandy loam texture as described earlier (Cruz-Martínez et al., 2012). After removal of the residual litter layer, soil was sampled to 40 cm depth and stored at -80°C in our laboratory until analysis. To determine the water content, a subset of soil was oven-dried at 105°C until a constant weight. Soil pH (using a saturated paste) and particle size (Sheldrick and Wang, 1993) were measured by the University of California Davis Analytical Laboratory. Total dissolved organic carbon (TOC) was measured on a filtered 1:4 soil:water extract of unfumigated and fumigated soil (following the method of section 2.3) (Shimadzu TOC-V CSH analyzer, Kyoto, Japan). Immediately prior to fumigation and

extractions, soil was sieved <2-mm using an ethanol-sterilized sieve (AS 200 sieve shaker, Retsch, Haan, Germany) at 4°C. To measure total metabolites (defined here as microbial and extracellular), soil was fumigated with ethanol-free chloroform for 24 h by the method of Vance et al. (Vance et al., 1987).

2.3. Soil extraction conditions

The following soil extraction method was based on a preliminary investigation of optimizing experimental variables for soil metabolite extraction (see supplemental information). Extractions were performed using either LC/MS-grade water (pH 7.4), 10 mM K₂SO₄ (pH 6.3), 10 mM NH₄HCO₃ (pH 8.5), isopropanol/methanol/water (3:3:2 v/v/v) or 10-100% methanol in water. All extraction solutions were sterilized by filtration through 0.2 µm cellulose acetate filters (Nalgene Rapid-Flow SFCA, Thermo Scientific, Waltham, MA). Field-moist soil (2 g, unfumigated or fumigated) was added to 50 mL polypropylene Falcon tubes and kept on ice. Ice-cold extraction solution (8 mL) was added followed by a spike with the internal standard, ABMBA (5 µg). Samples were shaken on an orbital shaker (Orbital-Genie, Scientific Industries, Bohemia, NY) at 200 rpm for 1 h at 4°C then centrifuged at 3220 x g for 15 min. The supernatant was filtered through 0.45 µm syringe filters (Pall Acrodisc Supor membrane) into 15 mL tubes and dried in a Savant SpeedVac SPD111V (Thermo Scientific, Waltham, MA) for 6-8 h. Dried residues were resuspended in 200 µl methanol, vortexed and sonicated for 5 min followed by a final filtration through 0.22 µm centrifugal membranes (Ultrafree-MC-GV PVDF, Millipore, Billerica, MA) and an aliquot (100 µl) dried for GC/MS derivatization. Soil extractions were performed in quadruplicate.

2.4. Aqueous microbial metabolite spike-recovery

For the spike-recovery experiment, soil was sterilized in an oven at 105°C for 6 h and checked for sterility by plating 100 µl aliquots (in triplicate) of centrifuged water extracts (2 g soil + 8 mL water) on LB agar plates where no growth was observed. The soil isolate, *Pseudomonas stutzeri* RCH2 was obtained from Romy Chakraborty, LBNL and grown at 37°C in M9 minimal media (per liter: 6 g Na₂HPO₄, 3 g KH₂PO₄, 0.5 g NaCl, 1 g NH₄Cl, 2 mM MgSO₄, 100 µM CaCl₂). For unlabeled (¹²C) cultures, the minimal media was supplemented with 0.2% sodium acetate as the carbon source, while the labeled (¹³C) cultures contained ¹³C₂-sodium acetate. Cultures were grown until an optical density of 0.5 (at 600 nm) was reached. Cells were distributed into twelve 50 mL Falcon tubes and pelleted (by centrifuging at 3220 x g for 10 min), washed with cold phosphate-buffered saline (pH 7.4) and re-pelleted. The supernatant was discarded and pellets resuspended in 1 mL cold methanol. Cells were lysed by sonicating for 2 x 20 s using a Q125 QSonica sonicator followed by 5 min in a sonicating water bath (VWR symphony), centrifuged at 2348 x g for 5 min and the supernatant dried. Unlabeled lysates were resuspended in 200 µl methanol and 100 µl were re-dried for GC/MS derivatization. The ¹³C-labeled lysates were resuspended in 1 mL sterile LC/MS-grade water and pooled together resulting in a total organic carbon content of 383 mg/L (Shimadzu TOC-V CSH analyzer, Kyoto, Japan). Aliquots (1 mL) were incubated with sterile soil (2 g dry weight) in 50 mL Falcon tubes (or added to tubes not containing soil as a negative control) and shaken at 200 rpm at 4°C for 15 min. Extraction solution (8 mL, either LC/MS-grade H₂O or 10 mM K₂SO₄) was added to samples and shaken for an additional 1 h and extracted as in section 2.3. Three biological replicates were performed for each condition. Dried extracts were resuspended in 200 µl methanol and 100 µl were re-dried and derivatized for GC/MS.

2.5. Sample Derivatization for GC/MS

Metabolite analysis was performed following methods described by Kind et al (2009). Briefly, 10 µl of a solution containing 40 mg/mL of 98% methoxyamine hydrochloride in pyridine was added and shaken for 90 min at 60°C. A mixture of internal retention time markers, used to monitor retention time stability across runs, was prepared using fatty acid methyl esters (FAMES) of C8, C9, C10, C12, C14, C16, C18, C20, C22, C24, C26, C28 and C30 linear chain length, dissolved in chloroform at a concentration of 0.8 mg/mL (C8-C16) and 0.4 mg/mL (C18-C30). The FAME mixture (20 µl) was added to 1 mL of MSTFA (containing 1% TMCS). To each extract was then added 92 µl of MSTFA spiked with the FAME mixture followed by mixing for 30 min at 37° C.

2.6. GC/MS Data Acquisition

Derivatized samples were manipulated using a Gerstel automatic liner exchange MPS system (Gerstel, Muehlheim, Germany) controlled by Maestro software to inject 0.5 µl of sample into a Gerstel Cooled Injection System (CIS4). The injector was operated in splitless mode, and the split vent was opened after 60 s. Samples were injected into the 50°C injector port which is ramped to 270°C in a 12°C/s thermal gradient and held for 3 min. Volatilized metabolites were separated using an Agilent 7890 gas chromatograph (Agilent Technologies, Santa Clara, CA), controlled by Agilent GC/MS MassHunter Acquisition software. The gas chromatograph was equipped with a 30 m long, 0.25 mm ID Rtx5Sil-MS column (Restek, Bellefonte, PA), 0.25 mm 5% diphenyl film with a 10 m integrated guard column. Initial oven temperature was 50°C with the following gradient applied: ramp at 5°C min⁻¹ to 65°C, held for 0.2 min; ramp at 15°C min⁻¹

to 80°C, held for 0.2 min; ramp at 15°C min⁻¹ to 310°C, held for 12 min. Mass spectrometry was performed by an Agilent 5977 single quadrupole mass spectrometer (MS) with 250°C transfer line temperature, electron ionization at 70 eV and an ion source temperature of 230°C. Mass spectra were acquired from 50–700 *m/z* at 8 spectra s⁻¹. Blanks (prepared as in section 2.5, but starting with an empty Eppendorf tube) were run between every 6 samples and a quality control mix was run as five replicates before and after the sample set.

2.7. GC/MS data analysis

Initial data analysis was done with Agilent MassHunter Qualitative Analysis software (Agilent Technologies, Santa Clara, CA). Total ion chromatograms were integrated, with each integrated peak was associated a cleaned spectra post background subtraction. Retention index markers (FAMES) were monitored across runs to ensure consistent retention time alignment and instrument sensitivity. Compound identification was based on an orthogonal approach in accordance with the Metabolomics Standards Initiative (Sumner et al., 2007). First, peak lists were queried against the Agilent Fiehn Metabolomics mass spectral library of over 1,000 curated metabolite spectra with a required minimum match score of 70%. Secondly, the identification of these matches was confirmed by comparing retention times and spectra with authentic standards. In the cases where matches could not be confirmed due to retention time discrepancies (± 10 sec or more) with an authentic standard or unavailability of an authentic standard, the compound identities are considered putatively characterized (based solely on spectral similarity with the Fiehn library) as indicated in the figures and tables. Resultant compound lists were aligned by retention times using custom software developed in-house using the Matlab (Mathworks, Natick, MA) programming language followed by manual curation for verifying proper data alignment

and compound annotation. For retention times with more than one compound hit, the final compound identification was determined based on manual investigation of the data (presence in the largest number of samples and highest abundance). Significant differences between unfumigated and fumigated samples were analyzed by Student's *t*-test with a *p* value < 0.05 considered statistically significant.

For the spike-recovery experiment, spectra from unlabeled RCH2 (¹²C) were compared to labeled RCH2 (¹³C) for the identification of *m/z* shifts corresponding to ¹³C labeling of fragments. These ¹³C-*m/z* (target ions) originating from the labeled RCH2 were quantified in all samples within the corresponding retention time window by subtracting the ¹³C-*m/z* contribution from soil:

$$[^{13}\text{C-}m/z \text{ peak area}_{\text{soil+lysates}}] - [^{13}\text{C-}m/z \text{ peak area}_{\text{soil}}] = \text{recovered } ^{13}\text{C-}m/z_{\text{lysates}}$$

Percent recovery of labeled metabolites were calculated by:

$$[\text{recovered } ^{13}\text{C-}m/z_{\text{lysates}} / \text{added } ^{13}\text{C-}m/z_{\text{lysates, no soil}}] \times 100 = \% \text{ recovered}$$

Compounds for the spike-recovery experiment were identified based on multiple approaches: comparing spectra from the unlabeled RCH2 with the Fiehn and National Institute of Standards and Technology mass spectral libraries, confirming retention times and spectra with authentic standards, and based on ¹³C patterns in spectra (and comparing to ¹²C spectra) to determine the number of carbons in the compound.

3. RESULTS

3.1. Soil properties

TOC of unfumigated Angelo soil extract (with water) was 10.10 ± 1.88 mg/L and fumigated was 21.10 ± 0.58 mg/L. The percentage water weight was approximately 10%, soil pH (saturated paste) was 5.4 and particle size (% sand/ silt/ clay) was 51/ 28/ 21.

3.2. Detection of a wide range of soil metabolites using GC/MS

Our untargeted profiling soil extraction method (Fig. 1) and GC/MS analyses revealed a wide range of metabolites (Fig. 2, Supp. Table 1) in the Angelo sandy loam soil. Initial analyses led to the detection of approximately 300 molecular features at unique retention times. However, many of these features were determined to be contaminants, derivatization artifacts and/or unidentified metabolites. After further investigation, the list was narrowed down to 55 features that were annotated to the Fiehn spectral metabolite database (Kind et al., 2009). Most metabolites (80%) were verified with authentic standards in accordance with the guidelines of the Metabolomics Standards Initiative (Sumner et al., 2007) and those that could not be confirmed are considered putative and indicated by italics (Fig. 2). Metabolites that were detected in all extracts and were among the most intense peaks include sugars (sucrose, glucose, trehalose, fructose) and the sugar alcohol, mannitol. Nine amino acids and many amino acid metabolites were also consistently detected. Other metabolites, more dependent on extraction conditions, include other sugars and sugar alcohols, fatty acids, dicarboxylic acids, sterols and nucleobases (Supp. Table 1). As is common in GC/MS, multiple peaks were detected for some metabolites, which is especially problematic for sugars due to the many isomers and silylation positions (Blau and Halket, 1993; Bertozzi and Rabuka, 2009; Ruiz-Matute et al., 2011). In these cases, we report the most intense peak and in the few situations where we could not discriminate between isomers they are presented as two identities (*e.g.* lyxose/ xylose).

3.3. Effects of extractant solutions on soil metabolite extraction

Extractant comparison included a small set of common additives to determine what, if any, additional types of soil metabolites can be detected with our extraction and GC/MS workflow. Additions included a classical soil extractant salt (K_2SO_4), a volatile salt (NH_4HCO_3), and standard metabolomics organic solvents (methanol and isopropanol) commonly included to induce cell lysis and/or to improve recovery of non-polar metabolites. Although 500 mM K_2SO_4 is commonly used, we used 10 mM since higher concentrations interfered with GC/MS derivatization in our studies and low-salt concentrations have been reported as effective in earlier studies (Gaillard et al., 1999, Haney et al., 2001, Zsolnay, 2003). Hierarchical clustering was used to compare soil metabolite peak areas (normalized to the most intense peak for each metabolite) across all 16 extraction conditions (eight extractants of unfumigated or fumigated soil). This approach grouped compounds together with similar extraction patterns as shown in Figure 2. Some metabolite peaks were more intense in the aqueous extractions (*e.g.* serine, threonine, glycine, isoleucine, alanine, valine, 3-hydroxybutyric acid, malonic acid and many sugars and sugar alcohols; lower half of Fig. 2) while others were only observed in soil extracted with high organic solvent content (ergosterol, fucose, stigmasterol, oleic acid, linoleic acid, palmitoleic acid; top of Fig. 2). The total number of compounds detected in each extraction condition was surprisingly similar (Supp. Fig. 1). Notably, the addition of 10 mM K_2SO_4 or 10 mM NH_4HCO_3 did not increase the number of compounds compared to water alone.

3.4. Effects of soil fumigation on metabolite extraction

As indicated by the largest cluster in Figure 2 (mid-lower half), a subset of metabolites were significantly more abundant in fumigated samples than unfumigated (Supp. Table 2) with many not even detected unless fumigated prior to extraction. Valine, glycerol, lyxose/ xylose, ribonic acid-gamma-lactone, fructose, glucose, methyl-beta-galactopyranoside, N-acetyl-mannosamine, allo-inositol, sophorose and maltose/ isomaltose peak areas were all significantly higher after fumigation regardless of extractant (Supp. Table 2). Some compounds only detected in fumigated soil include 2'-deoxycytidine, proline, thymine, threose, ribose, galactose, gluconic acid and allose. A few metabolites were significantly higher in unfumigated soil than fumigated (trehalose, ergosterol and 5-hydroxy-tryptophan) possibly due to modification during the fumigation process.

3.5. Microbial metabolite adsorption and recovery

To explore the types of microbial metabolites that adsorb to soil and therefore may represent those that are unavailable to microbes *in situ*, ¹³C-labeled RCH2 bacterial lysates were incubated with soil and extracted with water (Fig. 3). K₂SO₄ (10 mM) was compared to water to determine if this classical extractant was more efficient at metabolite recovery. The most intense peaks from the unlabeled RCH2 lysates (grown with ¹²C-acetate) were selected for further analysis in labeled samples (compound identification, ¹³C incorporation and recovery calculations). Spectra from unlabeled RCH2 were compared to labeled cultures, which resulted in the identification of 45 fully-labeled metabolites (Fig. 4, Supp. Table 3). The ¹³C target ion for each metabolite was quantified for all samples and recovery was calculated. Only 27% of the quantified metabolites showed >50% recovery in the WEOM, with the majority of those being neutral metabolites. All of the polyamine cations identified in this experiment (putrescine,

cadaverine, lysine, spermine and spermidine) and most of the anionic metabolites were among those that adsorbed rapidly to soil (Fig. 4). The inclusion of 10 mM K₂SO₄ as a recovery extractant did not drastically improve or reduce recovery compared to water alone.

4. DISCUSSION

GC/MS is a simple and suitable approach for untargeted soil metabolomics

GC/MS is a popular analytical technique for analyses of metabolites in a variety of biological systems. In this study we aimed to develop a simple and robust GC/MS-based metabolomics workflow to characterize the chemical composition of the extractable fraction of soil, which consists of important microbial products, and to assess microbial metabolite recovery from soil. These foundational studies present the ability to perform untargeted global profiling of WEOM in soils and to detect and track microbial metabolites as a function of their soil-adsorptive properties (which can be expanded to other soil types and minerals).

Detected metabolites play important roles in soil microbial communities

Numerous studies have shown that SOM composition has a major impact on the activity, abundance and composition of microbial biomass in soils (Xu et al., 2014). Specifically, in addition to microbial products, plant exudates (consisting mostly of sugars, amino acids and organic acids) stimulate microbial activity in soils (Jones et al., 2004). Overall, our soil extraction and metabolomics workflow detected a wide range of small molecules that have critical roles in the biological activity of soils. Studies have shown that when glucose is administered to soil it is rapidly assimilated into microbial biomass (Sugai and Schimel, 1993; Kuzyakov and Jones, 2006) an indication that microbial communities will respond quickly to

changes in SOM composition resulting from fluctuating inputs from aboveground (*e.g.* plant diversity and productivity) and belowground (*e.g.* root exudates, microbial population). Production of small molecule osmolytes (*e.g.* proline, trehalose, fructose, sucrose) allow soil microbes to respond to altered water conditions (Warren, 2014) and monitoring these small molecules in DOM may provide insight into the physiological status of soil systems. Microbial growth has also been shown to be mediated more by nitrogen rather than soil carbon content (Eisenhauer et al., 2010) indicating that amino acids play a key role in regulating these microbial communities.

Addition of dilute salts has a weak effect on metabolite coverage

In order to demonstrate the suitability of an untargeted metabolite profiling approach for detection of a range of compounds, we developed a rapid workflow to evaluate WEOM from soil and compared a limited set of conditions. Water is an advantageous extractant in that it is simple and may be more reflective of actual microbe-accessible DOM in soils (Blagodatsky et al., 2010). However, the addition of salts (typically K_2SO_4) is thought to be required to release metabolites bound to mineral surfaces or clay particles somewhat analogous to ion exchange chromatography. Our preliminary studies of more concentrated salt solutions showed that concentrations greater than 10 mM were incompatible with GC/MS by forming excessive salt crystals that likely trap metabolites and also retain water causing interference with derivatization. The use of less than the standard salt concentration (500 mM) was still in agreement with earlier soil extraction methods (Gregorich et al., 2000; Haney et al., 2001). However, even with 10 mM K_2SO_4 , salt crystals formed during dry-down which may have contributed to a slightly weaker metabolite profile compared to water. Although we did not expand our experiments to include

419 buffers with a range of ionic strengths, it is likely that the efficiency of salt-based extractants
420 compared with water is dependent on the ionic strength of the solution (Gregorich et al., 2000;
421 Makarov et al., 2013) and on soil properties such as acidity and carbon/ nitrogen content (Haney
422 et al., 2001; Makarov et al., 2013). Specifically, in Haney et al (2001) and confirmed here, the
423 extractable yield is higher using water compared to K_2SO_4 for acidic soils such as the one
424 examined in this study. Overall, these studies show that 10 mM K_2SO_4 and 10 mM NH_4HCO_3
425 are compatible with mass spectrometry instrumentation, but are not more efficient at extracting
426 metabolites than water alone.

428 ***Water is a suitable extractant for exometabolomics***

429 Traditionally, another reason for including salts in extractants is to reduce osmotic shock
430 and prevent cell lysis or metabolite leakage during extraction (Gregorich et al., 2000) and this
431 may be another explanation for the richer metabolite profile in the water extracts. However, the
432 significant differences observed between the fumigated versus unfumigated treatments in the
433 water-extracted group suggest that lysis in the unfumigated samples is far from complete. This
434 result is consistent with the common interpretation of water-extractable metabolites as
435 extracellular metabolites (Gregorich et al., 2000), *i.e.* exometabolites, and it may be that soil
436 itself is sufficient for balancing salt concentrations. In situations where it is necessary to include
437 a buffering reagent, NH_4HCO_3 may be the salt of choice when combining soil extractions with
438 mass spectrometry-based metabolomics since it can be removed by drying. In some cases, salt
439 concentrations can be decreased by using low molecular weight dialysis (Murage and Voroney,
440 2007), but was not used in our studies due to concerns about metabolite loss.

As expected, we found that aqueous solutions, due to their polar nature, were better at extracting small, polar metabolites while organic solvents-based solutions extracted fatty acids and sterols. Our results for 3:3:2 extractions support earlier studies in demonstrating a broad coverage of metabolites (Lee et al., 2012), but extracted a similar number of compounds as water and the other extractants (Supp. Fig. 1). While water appears to be an excellent soil extractant, if the goal is to analyze fatty acids or sterols, higher concentrations of organic solvent are required.

Fumigation prior to extraction allows detection of significantly more microbial metabolites

The original chloroform vapor fumigation method by Vance et al. (1987) was used rather than direct extraction with chloroform since it would have interfered with comparing the extraction efficiency of methanol and isopropanol due to miscibility. We found that fumigation significantly increased the number and peak intensity of metabolites detected from soils. In fact, hierarchical clustering based on metabolite patterns clustered fumigated samples with the 100% methanol and 3:3:2 extracts of unfumigated soil (Supp. Fig. 2) supporting the notion that chloroform vapor lyses microbial cells (just as organic solvent does when used as an extractant). Many of the metabolites that increased after fumigation were also among the most recoverable RCH2 bacterial metabolites detected in the spike-recovery experiments (*e.g.* alanine, valine, isoleucine, glycine, threonine, serine) providing further evidence that these are microbial metabolites. Compounds observed at higher abundances in unfumigated soils compared to fumigated (trehalose, ergosterol and 5-hydroxy-tryptophan) could be the result of continued metabolic activity and/ or increased enzymatic degradation of those metabolites during 24 h of fumigation. The use of water as an extractant of fumigated soil is a suitable approach to examine both (intracellular) microbial metabolites and exometabolites (Supp. Fig. 3), but some drawbacks

of fumigation need to be carefully considered (difficulty in fumigating dry soil, recolonization of microbial communities, altered microbial enzymatic activity, chloroform exposure time differences) (Brookes et al., 1985; Tate et al., 1988; Haubensak et al., 2002; Renella et al., 2002).

Some microbial metabolites rapidly adsorb to soil

Soil physical properties such as mineral composition, surface area, shape and porosity present a series of unique challenges to soil extraction and metabolomics approaches (Skopp, 2000). To gain insights into these confounding variables and to establish a method to estimate what types of metabolites may be available to microbes *in situ*, we added bacterial extracts to soil samples and determined the fractional recovery with water. Our results indicate that even after a short incubation, potentially mimicking immediately after the onset of rainfall, many metabolites are adsorbed to soil. Cationic and anionic metabolites were most affected, while the majority of neutral metabolites were recovered. An important consideration is that the metabolites not recovered represent molecules that are often overlooked in aqueous soil extraction studies aimed at examining total SOM. Interestingly, the polyamines that were detected in the bacterial lysates were among the least-recoverable from soil. The loss of these compounds is unlikely due to microbial metabolism since the soil was sterilized by oven-heating (105 °C). These findings are consistent with numerous reports of amines adsorbing to clay material in soil (Zhang et al., 1993; Lee et al., 1997) as well as other metabolites, especially carboxylic acids, potentially forming covalent bonds with organic matter or being chelated by metals or other ions (originating from soil or from the extractant) (Jones et al., 2003; Fischer et al., 2010). Therefore, cationic compounds such as polyamines may represent an inaccessible biological class of substrates that are not present in DOM *in situ*. While we focused on RCH₂

metabolite adsorption to a sandy loam soil, the system described could be further parameterized: microbial metabolites (from other species), soil types, metabolite-soil contact time, incubation temperature and recovery extractants (with ionic strength or pH similar to rainwater). This spike-recovery method holds great promise to further our understanding of how specific metabolites will behave in response to future climatic changes (*e.g.* precipitation) thereby affecting substrate availability and the structure of soil microbial communities.

CONCLUSION

Here we report a GC/MS soil metabolomics workflow that was used to profile a diverse set of metabolites in a sandy loam soil using a variety of extractants and to examine microbial metabolite adsorption. We found that water is an excellent starting point to examine both microbe-accessible extracellular soil nutrients (using unfumigated soil) and total extractable organic matter (using fumigated soil) while the addition of organic solvent can expand lipid coverage. Our spike-recovery studies revealed a set of metabolites (mostly cationic polyamines and anions) that are not recovered using aqueous extractants and represent an important set of biological compounds that may be inaccessible to microbes in the environment. Together this work provides a rapid and simple foundation for GC/MS soil metabolite profiling methods that are readily accessible to microbiologists and soil scientists as efforts continue in exploring the chemical composition and dynamics of complex soil systems.

ACKNOWLEDGEMENTS

We thank Jill Banfield (University of California, Berkeley) and her lab (especially Susan Spaulding and David Burstein) for assisting in soil collection, Zachary Aanderud (Brigham

Young University) for providing essential guidance on soil extraction methodology, Peter Nico (LBNL) for providing critical feedback on this manuscript and Alexandra Walling for preparing and analyzing the authentic standards. This work was funded by the US Department of Energy grants DE-SC0010566 and DE-AC02-05CH11231.

REFERENCES

- Amelung, W., Brodowski, S., Sandhage-Hofmann, A., Bol, R., 2008. Chapter 6. Combining Biomarker with Stable Isotope Analyses for Assessing the Transformation and Turnover of Soil Organic Matter. In: Sparks, D.L. (Ed.), *Advances in Agronomy*, Vol 100. Academic Press, Waltham, MA, pp. 155-250
- Annesley, T.M., 2003. Ion Suppression in Mass Spectrometry. *Clinical Chemistry* 49, 1041–1044.
- Baldock, J.A., Nelson, P.N., 2000. Soil Organic Matter. In: Sumner, M.E. (Ed.), *Handbook of Soil Science*. CRC Press, LLC, New York, pp. B25-B84.
- Baran, R., Reindl, W., Northen, T.R., 2009. Mass spectrometry based metabolomics and enzymatic assays for functional genomics. *Current Opinion in Microbiology* 12, 547–552.
- Baran, R., Bowen, B.P., Northen, T.R., 2011. Untargeted metabolic footprinting reveals a surprising breadth of metabolite uptake and release by *Synechococcus sp. PCC 7002*. *Molecular BioSystems* 7, 3200-3206.
- Baran, R., Ivanova, N., Jose, N., Garcia-Pichel, F., Kyrpides, N., Gugger, M., Northen, T., 2013. Functional genomics of novel secondary metabolites from diverse cyanobacteria using untargeted metabolomics. *Marine Drugs* 11, 3617–3631.

533 Bertozzi, C.R., Rabuka, D., 2009. Chapter 2. Structural Basis of Glycan Diversity. In: Varki A.,
 534 Cummings R.D., Esko J.D., Freeze H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., Etzler
 535 M.E. (Eds.), *Essentials of Glycobiology*, Second Edition. Cold Spring Harbor Laboratory
 536 Press, Cold Spring Harbor Laboratory Press, New York. pp. 23-36.

537 Blagodatsky, S., Blagodatskaya, E., Yuyukina, T., Kuzyakov, Y., 2010. Model of apparent and
 538 real priming effects: Linking microbial activity with soil organic matter decomposition. *Soil
 539 Biology and Biochemistry* 42, 1275–1283.

540 Blau, K., Halket, J., 1993. *Handbook of Derivatives for Chromatography*, Second Edition. John
 541 Wiley & Sons Ltd., Chichester, U.K., pp. 70.

542 Brookes, P.C., Landman, A., Pruden, G., Jenkinson, D.S., 1985. Chloroform fumigation and the
 543 release of soil nitrogen: a rapid direct extraction method to measure microbial biomass
 544 nitrogen in soil. *Soil Biology and Biochemistry* 17, 837–842.

545 Buyer, J.S., Sasser, M., 2012. High throughput phospholipid fatty acid analysis of soils. *Applied
 546 Soil Ecology* 61, 127–130.

547 Cruz-Martínez, K., Rosling, A., Zhang, Y., Song, M., Andersen, G.L., Banfield, J.F., 2012.
 548 Effect of rainfall-induced soil geochemistry dynamics on grassland soil microbial
 549 communities. *Applied and Environmental Microbiology* 78, 7587–7595.

550 Edwards, J.L., Chisolm, C.N., Shackman, J.G., Kennedy, R.T., 2006. Negative mode sheathless
 551 capillary electrophoresis electrospray ionization-mass spectrometry for metabolite analysis
 552 of prokaryotes. *Journal of Chromatography A* 1106, 80–88.

553 Eisenhauer, N., Beßler, H., Engels, C., Gleixner, G., Habekost, M., Milcu, A., Partsch, S.,
 554 Sabais, A.C.S., Scherber, C., Steinbeiss, S., Weigelt, A., Weisser, W.W., Scheu, S. 2010.
 555 Plant diversity effects on soil microorganisms support the singular hypothesis. *Ecology* 91,

556 485–496.

557 Fiehn, O., Kopka, J., Trethewey, R.N., Willmitzer, L., 2000. Identification of uncommon plant
 558 metabolites based on calculation of elemental compositions using gas chromatography and
 559 quadrupole mass spectrometry. *Analytical Chemistry* 72, 3573–3580.

560 Fiehn, O., 2002. Metabolomics- the link between genotypes and phenotypes. *Plant Molecular*
 561 *Biology* 48, 155–171.

562 Fischer, H., Ingwersen, J., Kuzyakov, Y., 2010. Microbial uptake of low-molecular-weight
 563 organic substances out-competes sorption in soil. *European Journal of Soil Science* 61, 504–
 564 513.

565 Gaillard, V., Chenu, C., Recous, S., Richard, G., 1999. Carbon, nitrogen and microbial gradients
 566 induced by plant residues decomposing in soil. *Eur J Soil Sci* 50, 567-578.

567 Garcia, D.E., Baidoo, E.E., Benke, P.I., Pingitore, F., Tang, Y.J., Villa, S., Keasling, J.D., 2008.
 568 Separation and mass spectrometry in microbial metabolomics. *Current Opinion in*
 569 *Microbiology* 11, 233–239.

570 Goulitquer, S., Potin, P., Tonon, T., 2012. Mass spectrometry-based metabolomics to elucidate
 571 functions in marine organisms and ecosystems. *Marine Drugs* 10, 849–880.

572 Gregorich, E.G., Liang, B.C., Drury, C.F., Mackenzie, A.F., McGill, W.B., 2000. Elucidation of
 573 the source and turnover of water soluble and microbial biomass carbon in agricultural soils.
 574 *Soil Biology and Biochemistry* 32, 581–587.

575 Haney, R.L., Franzluebbers, A.J., Hons, F.M., Hossner, L.R., Zuberer, D.A., 2001. Molar
 576 concentration of K₂SO₄ and soil pH affect estimation of extractable C with chloroform
 577 fumigation–extraction. *Soil Biology and Biochemistry* 33, 1501–1507.

578 Haubensak, K.A., Hart, S.C., Stark, J.M., 2002. Influences of chloroform exposure time and soil

579 water content on C and N release in forest soils. *Soil Biology and Biochemistry* 34, 1549-
580 1562.

581 Hirai, M.Y., Yano, M., Goodenowe, D.B., Kanaya, S., Kimura, T., Awazuhara, M., Arita, M.,
582 Fujiwara, T., Saito, K., 2004. Integration of transcriptomics and metabolomics for
583 understanding of global responses to nutritional stresses in *Arabidopsis thaliana*.
584 *Proceedings of the National Academy of Sciences USA* 101, 10205–10210.

585 Hockaday, W.C., Grannas, A.M., Kim, S., Hatcher, P.G., 2006. Direct molecular evidence for
586 the degradation and mobility of black carbon in soils from ultrahigh-resolution mass spectral
587 analysis of dissolved organic matter from a fire-impacted forest soil. *Organic Geochemistry*
588 37, 501–510.

589 Jiao, N., Herndl, G.J., Hansell, D.A., Benner, R., Kattner, G., Wilhelm, S.W., Kirchman, D.L.,
590 Weinbauer, M.G., Luo, T., Chen, F., Azam, F., 2010. Microbial production of recalcitrant
591 dissolved organic matter: long-term carbon storage in the global ocean. *Nature Reviews*
592 *Microbiology* 8, 593–599.

593 Johnston, C.A., Groffman, P., Breshears, D.D., Cardon, Z.G., Currie, W., Emanuel, W.,
594 Gaudinski, J., Jackson, R.B., Lajtha, K., Nadelhoffer, K., Nelson, D., Post, W.M., Retallack,
595 G., Wielopolski, L., 2004. Carbon cycling in soil. *Frontiers in Ecology and the Environment*
596 2, 522–528.

597 Jones, D.L., Dennis, P.G., Owen, A.G., van Hees, P.A.W., 2003. Organic acid behavior in soils -
598 misconceptions and knowledge gaps. *Plant and Soil* 248, 31-41.

599 Jones, D.L., Hodge, A., Kuzyakov, Y., 2004. Plant and mycorrhizal regulation of
600 rhizodeposition. *New Phytologist* 163, 459–480.

601 Jones, D.L., Willett, V.B., 2006. Experimental evaluation of methods to quantify dissolved

602 organic nitrogen (DON) and dissolved organic carbon (DOC) in soil. *Soil Biology and*
 603 *Biochemistry* 38, 991–999.

604 Jones, O.A.H., Sdepanian, S., Lofts, S., Svendsen, C., Spurgeon, D.J., Maguire, M.L., Griffin,
 605 J.L., 2014. Metabolomic analysis of soil communities can be used for pollution assessment.
 606 *Environmental Toxicology and Chemistry* 33, 61–64.

607 Judd, K.E., Crump, B.C., Kling, G.W., 2006. Variation in dissolved organic matter controls
 608 bacterial production and community composition. *Ecology* 87, 2068–2079.

609 Kakumanu, M.L., Cantrell, C.L., Williams, M.A., 2013. Microbial community response to
 610 varying magnitudes of dessication in soil: a test of the osmolyte accumulation hypothesis.
 611 *Soil Biology and Biochemistry* 57, 644–653.

612 Kalbitz, K., Solinger, S., Park, J.-H., Michalzik, B., Matzner, E., 2000. Controls on the dynamics
 613 of dissolved organic matter in soils: a review. *Soil Science* 165, 277–304.

614 Kind, T., Wohlgemuth, G., Lee, D.Y., Lu, Y., Palazoglu, M., Shahbaz, S., Fiehn, O., 2009.
 615 FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole
 616 and time-of-flight gas chromatography/mass spectrometry. *Analytical Chemistry* 81, 10038–
 617 10048.

618 Koek, M.M., Muilwijk, B., van der Werf, M.J., Hankemeier, T., 2006. Microbial metabolomics
 619 with gas chromatography/mass spectrometry. *Analytical Chemistry* 78, 1272–1281.

620 Kuzyakov, Y., Jones, D.L., 2006. Glucose uptake by maize roots and its transformation in the
 621 rhizosphere. *Soil Biology Biochemistry* 38, 851–860.

622 Lee, L.S., Nyman, A.K., Li, H., Nyman, M.C., Jafvert, C., 1997. Initial sorption of aromatic
 623 amines to surface soils. *Environmental Toxicology and Chemistry* 16, 1575–1582.

624 Lee, D.Y., Bowen, B.P., Nguyen, D.H., Parsa, S., Huang, Y., Mao, J.-H., Northen, T.R., 2012.

625 Low-dose ionizing radiation-induced blood plasma metabolic response in a diverse genetic
626 mouse population. *Radiation Research* 178, 551–555.

627 Makarov, M.I., Shuleva, M.S., Malysheva, T.I., Menyailo, O.V., 2013. Solubility of the labile
628 forms of soil carbon and nitrogen in K₂SO₄ of different concentrations. *Eurasian Soil*
629 *Science* 46, 369–374.

630 Minor, E.C., Swenson, M.M., Mattson, B.M., Oyler, A.R., 2014. Structural characterization of
631 dissolved organic matter: a review of current techniques for isolation and analysis.
632 *Environmental Science Processes & Impacts*, advance online publication. doi:
633 10.1039/C4EM00062E

634 Mopper, K., Stubbins, A., Ritchie, J.D., Bialk, H.M., Hatcher, P.G., 2007. Advanced
635 instrumental approaches for characterization of marine dissolved organic matter: extraction
636 techniques, mass spectrometry, and nuclear magnetic resonance spectroscopy. *Chemical*
637 *Reviews* 107, 419–442.

638 Murage, E.W., Voroney, P.R., 2007. Modification of the original chloroform fumigation
639 extraction technique to allow measurement of $\delta^{13}\text{C}$ of soil microbial biomass carbon. *Soil*
640 *Biology and Biochemistry* 39, 1724–1729.

641 Ohno, T., Parr, T.B., Gruselle, M.-C. I., Fernandez, I.J., Sleighter, R.L., Hatcher, P.G., 2014.
642 Molecular composition and biodegradability of soil organic matter: a case study comparing
643 two new england forest types. *Environ. Sci. Technol.* 48, 7229-7236.

644 Renella, G., Landi, L., Nannipieri, P., 2002. Hydrolase activities during and after the chloroform
645 fumigation of soil as affected by protease activity. *Soil Biology and Biochemistry* 34, 51–60.

646 Roessner, U., Wagner, C., Kopka, J., Trethewey, R.N., Willmitzer, L., 2000. Simultaneous
647 analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *The Plant*

Journal 23, 131–142.

Ruiz-Matute, A.I., Hernández-Hernández, O., Rodríguez-Sánchez, S., Sanz, M.L., Martínez-Castro, I., 2011. Derivatization of carbohydrates for GC and GC-MS analyses. *J. Chromatogr. B* 879, 1226–1240.

Schimel, J.P., Schaeffer, S.M., 2012. Microbial control over carbon cycling in soil. *Frontiers in Microbiology* 3, 348.

Schmidt, M.W.I., Torn, M.S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I.A., Kleber, M., Kögel-Knabner, I., Lehmann, J., Manning, D.A.C., Nannipieri, P., Rasse, D.P., Weiner, S., Trumbore, S.E., 2012. Persistence of soil organic matter as an ecosystem property. *Nature* 478, 49–56.

Sheldrick, B.H., Wang, C., 1993. Particle-size Distribution. In: Carter, M.R. (Ed.), *Soil Sampling and Methods of Analysis*, Canadian Society of Soil Science. Lewis Publishers, Ann Arbor, MI, pp. 499-511.

Skopp, 2000. Physical Properties of Primary Particles. In: Sumner, M.E. (Ed.), *Handbook of Soil Science*. CRC Press, LLC, New York, pp. A3-A17.

Soga, T., Ohashi, Y., Ueno, Y., Naraoka, H., Tomita, M., Nishioka, T., 2003. Quantitative Metabolome Analysis Using Capillary Electrophoresis Mass Spectrometry. *Journal of Proteome Research* 2, 488–494.

Sugai, S.F., Schimel, J.P., 1993. Decomposition and biomass incorporation of ¹⁴C-labeled glucose and phenolics in taiga forest floor: effect of substrate quality, successional state, and season. *Soil Biology Biochemistry* 25, 1379–1389.

Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., Fan, T.W.M., Fiehn, O., Goodacre, R., Griffin, J.L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R.,

- Kopka, J., Lane, A.N., Lindon, J.C., Marriott, P., Nicholls, A.W., Reily, M.D., Thaden, J.J., Viant, M.R., 2007. Proposed minimum reporting standards for chemical analysis. *Metabolomics* 3, 211–221.
- Tate, K.R., Ross, D.J., Feltham, C.W., 1988. A direct extraction method to estimate soil microbial C- effects of experimental variables and some different calibration procedures. *Soil Biology and Biochemistry* 20, 329–335.
- Vance, E.D., Brookes, P.C., Jenkinson, D.S., 1987. An extraction method for measuring soil microbial biomass C. *Soil Biology and Biochemistry* 19, 703-707.
- Warren, C.R., 2013. High diversity of small organic N observed in soil water. *Soil Biology and Biochemistry* 57, 444-450.
- Warren, C.R., 2014. Response of osmolytes in soil to drying and rewetting. *Soil Biology and Biochemistry* 70, 22–32.
- Xu, X., Schimel, J.P., Thornton, P.E., Song, X., Yuan, F., Goswami, S., 2014. Substrate and environmental controls on microbial assimilation of soil organic carbon: a framework for Earth system models. *Ecology Letters* 17, 547-555.
- Zhang, Z.Z., Sparks, D.L., Scrivner, N.C., 1993. Sorption and desorption of quaternary amine cations on clays. *Environmental Science and Technology* 27, 1625–1631.
- Zsolnay, Á., 2003. Dissolved organic matter: artefacts, definitions, and functions. *Geoderma* 113, 187–209.

FIGURES

Figure 1. Workflow for soil WEOM analysis with GC/MS. **A.** Soil is sieved through 2 mm and fumigated with chloroform for 24 h or left unfumigated, **B.** extracted with the appropriate solution in a 1:4 ratio (2 g soil: 8 mL extractant) on an orbital shaker for 1 h at 4°C, **C.** centrifuged to pellet soil and the supernatant filtered through 0.45 µm filter discs, **D.** dried down and derivatized for GC/MS. **E.** Data are analyzed for metabolite identification.

Figure 2. Comparison of extraction conditions and relative intensity of detected soil metabolites. The dendrogram on the left clusters similarly extracted metabolites based on hierarchical clustering and the heat map displays the intensity of metabolites normalized to the most intense peak within each row (metabolite). Differences in metabolite peak intensities across extraction conditions are especially apparent in the mid-lower portion (unfumigated vs. fumigated soil) and the top portion (polar vs. nonpolar extractions). Putative metabolite annotations are shown by italics (retention time and/or spectra could not be confirmed with an authentic standard). n=4 for each extraction condition.

Figure 3. Experimental workflow for microbial metabolite spike-recovery experiment. Soil (brown flask), ¹³C-labeled RCH2 bacterial lysates (green flask) or soil spiked with lysates (blue flask) were incubated for 15 min and then extracted for 1 h with either water or 10 mM K₂SO₄, centrifuged and filtered. Dried extracts were derivatized and analyzed by GC/MS to determine the aqueous recovery of ¹³C metabolites.

Figure 4. Aqueous recovery or adsorption of microbial metabolites. ¹³C-labeled metabolites (from RCH2 bacterial lysates) were added to soil, incubated for 15 min, and extracted with water (or 10 mM K₂SO₄, not shown here) for 1 h. There were no obvious differences between recovery with water versus 10 mM K₂SO₄. Data shown are normalized to the largest percent recovery (serine). Putative metabolite annotations are shown by italics (retention time and/or spectra could not be confirmed with an authentic standard). Most notably, cations, consisting of biologically-important polyamines, and most anions were not recovered from soil compared to neutral compounds. The large error bars are likely a result of the inherent complexity and heterogeneous nature of soil. n=3.

Figure 1

**A. soil preparation-
sieving and fumigation**

**B. metabolite
extraction**

**C. centrifugation
and filtration**

**D. sample dry-down
and derivatization**

**E. GC/MS data acquisition
and analysis**

Figure 2

Figure 3

Figure 4

Supplementary Figure 1

[Click here to download Supplementary Material for online publication only: Supp Figure 1.pdf](#)

Supplementary Figure 2

[Click here to download Supplementary Material for online publication only: Supp Figure 2 v2.pdf](#)

Supplementary Figure 3

[Click here to download Supplementary Material for online publication only: Supp Figure 3 v2.pdf](#)

Supplementary Figure 4

[Click here to download Supplementary Material for online publication only: Supp Figure 4.pdf](#)

Supplementary Table 1

[Click here to download Supplementary Material for online publication only: Supp Table 1 v2 final.xlsx](#)