Construction of Viable Soil Defined Media Using Quantitative Metabolomics Analysis of Soil Metabolites

Stefan Jenkins [1†], Tami L. Swenson [1], Rebecca Lau [1], Andrea M. Rocha [2,3], Alex Aaring [4], Terry C. Hazen [2,5,6,7], Romy Chakraborty [4] and Trent R. Northen [1,8*]

[1] Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA, United States, [2] Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, United States, [3] Geosyntec Consultants, Knoxville, TN, United States, [4] Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA, United States, [5] Department of Civil and Environmental Engineering, University of Tennessee-Knoxville, Knoxville, TN, United States, [6] Department of Earth and Planetary Sciences, University of Tennessee-Knoxville, Knoxville, TN, United States, [7] Department of Microbiology, University of Tennessee-Knoxville, Knoxville, TN, United States, [8] Joint Genome Institute, Walnut Creek, CA, United States

Exometabolomics enables analysis of metabolite utilization of low molecular weight organic substances by soil bacteria. Environmentally-based defined media are needed to examine ecologically relevant patterns of substrate utilization. Here, we describe an approach for the construction of defined media using untargeted characterization of water soluble soil microbial metabolites from a saprolite soil collected from the Oak Ridge Field Research Center (ORFRC). To broadly characterize metabolites, both liquid chromatography mass spectrometry (LC/MS) and gas chromatography mass spectrometry (GC/MS) were used. With this approach, 96 metabolites were identified, including amino acids, amino acid derivatives, sugars, sugar alcohols, mono- and di-carboxylic acids, nucleobases, and nucleosides. From this pool of metabolites, 25 were quantified. Molecular weight cut-off filtration determined the fraction of carbon accounted for by the quantified metabolites and revealed that these soil metabolites have an uneven quantitative distribution (e.g., trehalose accounted for 9.9% of the <1 kDa fraction). This quantitative information was used to formulate two soil defined media (SDM), one containing 23 metabolites (SDM1) and one containing 46 (SDM2). To evaluate the viability of the SDM, we examined the growth of 30 phylogenetically diverse soil bacterial isolates from the ORFRC field site. The simpler SDM1 supported the growth of 13 isolates while the more complex SDM2 supported 15 isolates. To investigate SDM1 substrate preferences, one isolate, *Pseudomonas corrugata* strain FW300-N2E2 was selected for a time-series exometabolomics analysis. Interestingly, it was found that this organism preferred lower-abundance substrates such as guanine, glycine, proline and arginine and glucose and did not utilize the more abundant substrates maltose, mannitol,

trehalose and uridine. These results demonstrate the viability and utility of using exometabolomics to construct a tractable environmentally relevant media. We anticipate that this approach can be expanded to other environments to enhance isolation and characterization of diverse microbial communities.

INTRODUCTION

Soil organic matter, historically considered to be composed of large polymeric humic substances, is now thought to largely consist of microbial products (Schmidt et al., 2011). Traditionally, the water soluble fraction of soil carbon, known as dissolved organic carbon or matter, is defined as the fraction that passes through a 0.45 μm filter (Ohno et al., 2014). Many microbes process macromolecular substrates such as plant biomass extracellularly and uptake the resulting low molecular weight organic substances (LMWOS). Thus, it is not surprising that the water extractable organic carbon (WEOC) (Boyer and Groffman, 1996; Guigue et al., 2014) is associated with high microbial activity and soil respiration (Haney et al., 2012). This reinforces long-standing views that it is desirable for culture media to approximate the conditions, especially in terms of the composition and quantity of metabolites.

Most of our understanding of the biology of soil microbes is based on the small fraction of microbes that have been successfully cultivated in isolation. Reasoner's 2A agar medium (R2A) is one of the most widely used for isolations and was developed for the cultivation of bacteria found in potable water (Reasoner and Geldreich, 1985). Its effectiveness may be attributable to its rich nutrients (peptone, casamino acids and yeast extract). However, R2A was not designed to be ecologically relevant for soil environments and thus is not intended to assess soil bacterial processes. Mass spectrometry-based metabolomics methods now enable the examination of LMWOS composition, which has revealed a diversity of small molecule metabolites in soils (Warren, 2014; Swenson et al., 2015b). Soil metabolomics methods such as these may have utility in informing the development of relevant culture media to support laboratory studies of soil microorganisms.

Another metabolomics approach that has proven useful for obtaining bacterial phenotype information is exometabolomics. This method enables the study of the transformation of the small molecule environment by bacteria and other microorganisms by comparing spent media to non-inoculated controls. This direct assessment of phenotype is capable of resolving metabolite depletion against production and is emerging as a powerful complement to existing broader techniques such as bulk soil respiration and carbon transformation measurements (Butler et al., 2003; Miller et al., 2005; Schimel and Mikan, 2005; Fernandes et al., 2013; Tucker et al., 2014). Previous studies have demonstrated the capability of exometabolomics in providing a functional complement to genomic and transcriptomic data (Baran et al., 2013; Liebeke and Lalk, 2014). Recently, exometabolomics was used to discover exometabolite niche partitioning for seven biological soil crust isolates using microbial extract media derived from either six heterotrophs or a single cyanobacterium (Baran et al., 2015). However, a limitation of this approach is that many of the metabolites in these environmentally relevant rich media could not be identified and presumably many others were not detected. For many exometabolite profiling experiments it would be desirable to have an environmentally relevant defined medium such that all of the metabolites are accounted for and the observed patterns of metabolite utilization are relevant to the field site of interest. Linking together microbes with function would provide critical insights into the specific role and impact of organisms within an environment.
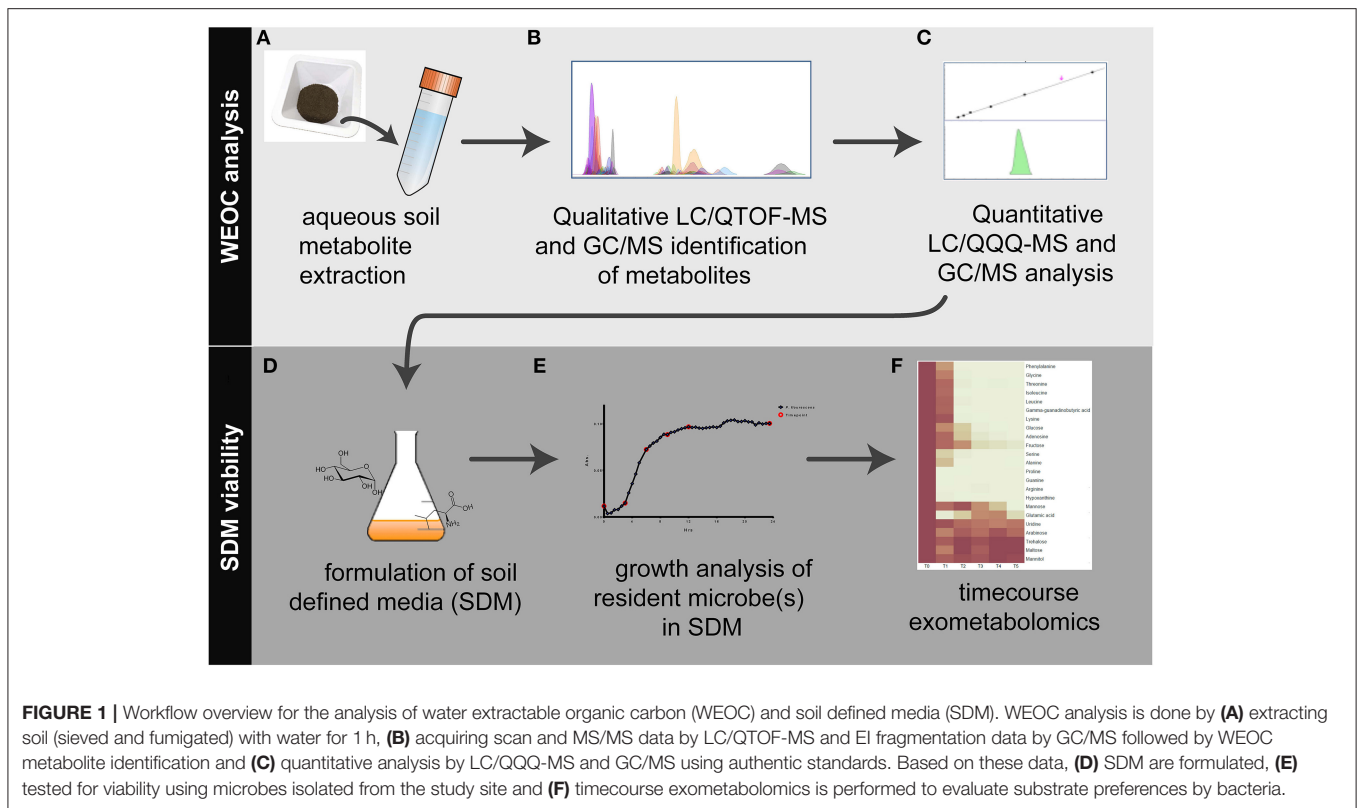
Here we describe the construction of defined media for exometabolomics experiments based on untargeted analysis of metabolites from a soil of interest (**Figure 1**). The extraction procedure focused specifically on metabolites present in a "whole soil" extract (i.e., intracellular and extracellular) to reflect potentially accessible microbial metabolites that may be present in rich soils. These water extractable metabolites, referred to here as WEOC, were first qualitatively characterized using both liquid chromatography mass spectrometry (LC/MS) and gas chromatography mass spectrometry (GC/MS). Application of these complementary technologies provided orthogonal confirmation of metabolite identification while also expanding the analytical scope of the study. From these data, a subset of metabolites was selected for absolute quantification to assist in formulating defined media that approximates their composition and quantity of metabolites within the soil. We then evaluated the ability of the resulting media to support the growth of 30 phylogenetically diverse isolates from the Oak Ridge Field Research Center (ORFRC) and performed time series characterization of the substrate preferences of a single isolate.

MATERIALS AND METHODS

Chemicals

Glucose (CAS 50-99-7) was from Amresco (Solon, OH). LC/MS-grade methanol (CAS 67-56-1) and water were from J.T. Baker (Avantor Performance Materials, Center Valley, PA). N-methyl-N-trimethylsilyltrifluoroacetamide (MSTFA) containing 1% trimethylchlorosilane (TMCS) was from Restek (Bellafonte,

---

Abbreviations: WEOC, water extractable organic carbon; LMWOS, low molecular weight organic substances; SDM, soil defined media; SDM(1/2), soil defined medium (1 and 2); R2A, Reasoner's 2A agar medium; ORFRC, Oak Ridge Field Research Center; LC/MS, liquid chromatography mass spectrometry; GC/MS, gas chromatography/mass spectrometry; TOC, total organic carbon; HILIC, hydrophilic interaction liquid chromatography.

FIGURE 1 | Workflow overview for the analysis of water extractable organic carbon (WEOC) and soil defined media (SDM). WEOC analysis is done by **(A)** extracting soil (sieved and fumigated) with water for 1 h, **(B)** acquiring scan and MS/MS data by LC/QTOF-MS and EI fragmentation data by GC/MS followed by WEOC metabolite identification and **(C)** quantitative analysis by LC/QQQ-MS and GC/MS using authentic standards. Based on these data, **(D)** SDM are formulated, **(E)** tested for viability using microbes isolated from the study site and **(F)** timecourse exometabolomics is performed to evaluate substrate preferences by bacteria.

PA). Acetonitrile (CAS 75-05-8), adenosine (CAS 58-61-7), arabinose (CAS 28697-53-2), biotin (CAS 58-85-5), calcium chloride (CAS 10043-52-4), calcium pantothenate (CAS 137-08-6), cobalt nitrate (CAS 10026-22-9), d$_{27}$-myristic acid (CAS 60658-41-5), ethylenediaminetetraacetic acid (EDTA; CAS 6381-92-6), folic acid (CAS 59-30-3), fructose (CAS 57-48-7), galactose (CAS 59-23-4), gamma-guanidinobutyric acid (CAS 463-00-3), guanine (CAS 73-40-5), hypoxanthine (CAS 68-94-0), iron sulfate (CAS 7782-63-0), magnesium sulfate (CAS 10034-99-8), maltitol (CAS 585-88-6), maltose (CAS 6363-53-7), manganese sulfate (CAS 10034-96-5), mannose (CAS 3458-28-4), mannitol (69-65-8), methoxyamine (CAS 593-56-6), nicotinic acid (CAS 59-67-6), p-aminobenzoic acid (CAS 150-13-0), monobasic potassium phosphate (CAS 7778-77-0), pyridine (CAS 110-86-1), pyridoxine hydrochloride (CAS 58-56-0), raffinose (CAS 17629-30-0), riboflavin (CAS 83-88-5), rutinose (CAS 90-74-4), sodium chloride (CAS 7647-14-5), sucrose (CAS 57-50-1), thiamine (CAS 67-03-8), thioctic acid (CAS 1077-28-7), trehalose dehydrate (CAS 6138-23-4), uridine (CAS 58-96-8), vitamin B12 (CAS 68-19-9), xylitol (CAS 87-99-0), xylobiose (CAS 6860-47-5), ^{13}C-^{15}N-L-phenylalanine (CAS 878339-23-2) and amino acid kits #A6407 and A6282 were from Sigma-Aldrich (St. Louis, MO)

Aqueous Soil Extraction and Total Organic Carbon Analysis

Upper B-horizon soil was collected on September 24, 2013 from the background area of the ORFRC in Oak Ridge, Tennessee. Specifically, the collection site was located approximately 3.5 m

from groundwater well FW-300, in an area where grass, surface roots, and surface rocks were not abundantly present. To reduce the presence of debris and non-soil constituents from the soil sample, rocks and surface roots present on the surface were physically removed prior to sampling. Surface soil was collected by shoveling soil to a depth of 30 cm into sterile whirl pack bags, which were immediately stored on ice. The soil at this location is an unconsolidated saprolite with organic rich clay soil that ranges from 0.5 to 3 m approximating the depth of the root zone. Other soil properties including total nitrogen, oxidizable organic matter, cation exchange capacity, pH, and particle size were determined by the UC Davis Analytical Laboratory (Davis, CA).

Aqueous extraction was performed as previously described (Swenson et al., 2015b). Briefly, soil was lyophilized to dryness and sieved to 2 mm prior to 24 h chloroform fumigation (Vance et al., 1987). Fumigation was used to lyse microbial cells to identify metabolites present in this fraction and to limit microbial processing of metabolites during subsequent steps. Six grams of soil were extracted in 24 mL of water at 4°C for 1 h while shaking on an orbital shaker (Orbital-Genie, Scientific Industries, Bohemia, NY). Tubes were then centrifuged at 4°C for 5 min at 3,200 × g. Supernatants were collected in fresh 50 mL conical tubes and centrifuged again after which supernatants were filtered through 0.45 μm syringe filters (Pall Acrodisc Supor membrane). These final WEOC samples were lyophilized to dryness in a Labconco Freezone 2.5 (Labconco, Kansas City, MO) and resuspended in water at a final equivalent concentration of 2 g soil/mL. WEOC with and without further 1 kDa molecular weight cutoff (MWCO) filtration was analyzed for total organic

carbon (TOC) by acidification and measurement of the non-purgeable organic carbon. WEOC samples were prepared as above, but at a final equivalent concentration of 1 g/mL (w/v of soil to water) and injected (50 μL) and run via an NPOC method using a Shimadzu (Kyoto, Japan) TOC-L series CSH/H-type TIC/TOC analyzer. Samples were individually mixed with a 1.5% v/v of 1 M HCl for the conversion of inorganic carbon to CO_2 followed by a 2.5 min purging time and combustion of the remaining organic carbon at an oven temperature of 680°C. The carrier gas used was synthetic air at a flow rate of 80 mL/min. Values reported for TOC are in ppm (μg C per mL extract).

LC/Q-TOF Scan MS and Tandem MS WEOC Profiling

Initial profiling of WEOC extracts was done by scan MS with a subset of metabolite identifications enhanced by acquiring tandem MS data and matching fragmentation spectra to spectral libraries. One hundred microliters of the WEOC extract were lyophilized in triplicate to dryness and resuspended in 100 μL of methanol containing internal standards 3,6-dihydroxy-4-methylpyridazine, 4-(3,3-dimethyl-ureido)benzoic acid and 9-anthracene carboxylic acid at 25 μM. Samples were separated by hydrophilic interaction liquid chromatography (HILIC) on an Agilent 1290 UHPLC (Agilent Technologies, Santa Clara CA) using a Merck Sequant ZIC-pHILIC column (Merck KGaA, Darmstadt, Germany) of dimensions 150 x 2.1 mm and 4.6 μm particle size. Two microliters of sample were injected with the following mobile phase: A- 5 mM ammonium acetate; B- 9:1 acetonitrile: 50 mM ammonium acetate. The following gradient was used with a flow rate of 0.4 mL/min: 0.0–1.5 min, 100% B; 25.0 min, 50% B; 26.0–32.0 min, 35% B; 33.0–40.0 min, 100% B. The column temperature was set to 40°C and the autosampler was maintained at 4°C. Flow was directed to the electrospray ionization source of an Agilent 6550 quadrupole time-of-flight (Q-TOF) mass spectrometer with the following settings for both positive and negative polarity ionization modes; drying gas temperature and flow 275°C and 14 L/min; sheath gas temperature and flow were 275°C and 9 L/min, respectively; nebulizer pressure was 30 psi; capillary voltage was 3,500 V and nozzle voltage 1,000 V. Mass spectra were acquired from 50 to 1,700 m/z at 4 spectra s^{-1}. Procedural blanks were used to generate precursor exclusion lists of compounds and added to a data-dependent auto MS/MS method with a spectral scan range of 30–1,200 amu and speed of 2 spectra s^{-1}. The quadrupole was set to a narrow isolation width of 1.3 amu and the collision cell fixed to 10, 20, and 40 eV. The auto MS/MS algorithm was restricted to a max of 4 precursor ions per MS/MS scan with a required precursor ion count threshold of 7,500 and a target of 50,000. Initial MS/MS spectra were then added within a retention time window of 0.25 min to the precursor exclusion list ensuring they were not selected for MS/MS in a subsequent run. This process was iterated in triplicate to ensure that MS/MS spectra of low intensity precursor ions were captured.

LC/QTOF-MS Data Analysis

MassHunter Qualitative Analysis v.B.07.00 SP1 (Agilent Technologies, Santa Clara CA) was used to query scan data against a custom LC accurate mass and retention time metabolite database based on previously acquired authentic standard data. Initial metabolite identifications based on scan MS data were augmented by querying spectral libraries such as METLIN (Smith et al., 2005), MassBank (Horai et al., 2010), and HMDB (Wishart et al., 2013) with subsequently acquired tandem MS spectra. For metabolites only identified by LC/MS, the highest confidence identifications (ranked as "1" in Supplementary Table 1) required the unknown to be within 5 ppm m/z, 0.5 min retention time and to share dominant fragment ions with the standard. Lower confidence identifications matched standards in accurate mass and retention time but not fragmentation (or MS/MS was not obtained) and are ranked as a level "2" in Supplementary Table 1. For some metabolites, authentic standards were not available, but were identified based on matching fragmentation patterns with spectral libraries and are indicated by a level "3."

GC/MS WEOC Metabolite Profiling and Sugar Quantitation

GC/MS was applied in addition to LC/MS to expand the analytical scope of the study by utilizing retention time indexed electron ionization (EI) spectral libraries for metabolite identification. Sample extracts were dried and derivatized by oximation-silylation and data acquired by GC/MS as previously described (Swenson et al., 2015b). A C8-C30 fatty acid methyl ester ladder was added to each sample for the establishment of retention indices (RI). The EI source of an Agilent 5977 GC/MS was maintained at 70 eV for spectral library matching. Metabolite profiling data were deconvoluted using Unknowns Analysis v.B.07.00 from Agilent Technologies (Santa Clara, CA) followed by matching spectral fragmentation by forward matching quality score of 70 to the Agilent Fiehn GC/MS Metabolomics RTL Library (Kind et al., 2009) within ±20 RI units. The level of chromatographic resolution imparted by gas phase chromatography was determined capable of separation and improved identification of the WEOC sugar pool. A set of sugar standards were initially used for qualitative profiling and included pentose, hexose, di-hexose and pentose alcohols (cellobiose, D-arabinose, D-cellobiose, D-galactose, D-mannose, D-xylose, fructose, glucose, L-arabinose, maltitol, maltose, mannitol, mannose, raffinose, rutinose, sucrose, trehalose, xylitol and xylobiose). Quantitation of the WEOC sugar pool was then done by preparing standards by single replicate at concentrations of 1, 5, 10, 25, and 50 μM with d$_{27}$-myristic acid used as an internal standard at a concentration of 25 μM (Supplementary Figure 1A). Quantitative data analysis was done using Agilent MassHunter Quantitative Analysis v.B.07.00. GC/MS identifications are provided in Supplementary Table 1.

LC/QQQ-MS for Metabolite Quantitation

For quantitation of metabolites initially detected in scan Q-TOF MS data, an Agilent 6460 triple quadrupole (QQQ) MS (Agilent Technologies, Santa Clara CA) was used with the same LC, column and ESI source conditions. Authentic standards were prepared for a subset of metabolites identified in the WEOC (i.e., amino acids, nucleobases, and nucleosides)

to generate compound specific transitions. Multiple reaction monitoring (MRM) transitions were scheduled by retention time segment as established using Agilent Optimizer v.B.07.00 and are provided in Supplementary Table 1. For the amino acid assay, a standard mix (Sigma-Aldrich amino acid standards kit, product#A6407, A6282) was diluted to concentrations of 0.1, 0.5, 1, and 10 μg/mL and run as single replicates (Supplementary Figure 1B). For the nucleotide/nucleobase assay, authentic standards were prepared at 1, 5, 10, and 25 μM concentrations in triplicate with ^{13}C-phenylalanine used as an internal standard at 25 μM (Supplementary Figure 1C). Soil extracts were then run at a concentration of 2 g/ml (soil/water) for determining concentrations of these metabolites. Quantitative data analysis was performed using Agilent MassHunter Quantitative Analysis version B.07.00. For each metabolite, quantitative data (mg/mL for each metabolite) are converted to carbon ppm based on the percent weight carbon within each metabolite and are reported as mg kg^{-1} extracted soil.

Formulation of Soil Defined Media and Comparison of Growth of 30 Phylogenetically Diverse ORFRC Isolates

Soil defined medium 1 (SDM1) was prepared by adding 23 metabolites quantified from WEOC (Section LC/Q-TOF Scan MS and Tandem MS WEOC Profiling) at the observed absolute concentrations, to a base medium composed of 1x Wolfe's mineral and 1x Wolfe's vitamin solutions, potassium phosphate and ammonium chloride (**Table 1**). For soil defined medium 2 (SDM2), the number of metabolites was expanded to 46 to include additional observed, but not quantified WEOC compounds. These additional metabolites (including many amino acids, nucleobases, and nucleosides) were added to the medium at the lowest concentration of a quantified metabolite within the same class (Supplementary Table 1). The soil defined media (SDM) at 10x concentrations were compared to R2A medium at 1x concentration (Tecknova, Hollister CA) in their ability to support the growth of a broad range of ORFRC isolates, each in triplicate. An existing collection of 30 phylogenetically diverse isolates from the ORFRC site were revived in liquid R2A from frozen glycerol stocks. DNA was obtained from overnight cultures to verify 16S rDNA based identifications. Aliquots were washed (3 times) with the test medium prior to inoculation into 96-well microtitre plates. For each fresh medium, 20 uL of starter culture was added to 180 uL and plates were incubated under aerobic conditions for 46 h at 28°C, and shaken at 365 rpm. Growth data were collected by measuring OD$_{600}$ on a BioTek Eon Microplate Spectrophotometer (BioTek, Winooski, VT) at 15 min intervals. Growth was defined as an increase of 0.05 or greater from the first time point (max OD- initial OD) after subtraction of the media control.

Time Series Exometabolomics Analysis with *Pseudomonas* sp. FW300-N2E2

Using SDM1-10x liquid medium, triplicate 2 ml cultures (including medium and water blank controls) of *Pseudomonas* sp. FW300-N2E2, a bacterial isolate from the ORFRC (16S

rDNA sequencing closest to *Pseudomonas corrugata* strain P94-EF153018.1; Thorgersen et al. (2015) were prepared in 24-well spectrophotometer plates using the inoculation technique described in 2.7. Time points were sampled from separate wells at early/mid log phase (3, 6 h) early/late stationary phase (9, 12 h) and finally at 24 h (Supplementary Figure 2). Culture fractions were removed for LC/MS and GC/MS analysis at 0.5 and 0.2 mL volumes, respectively, and centrifuged in a Savant centrifuge (Thermo Scientific, San Jose CA) for 5 min at 2,823 × g and the resultant supernatant filtered through 0.2 μm centrifugal filters (Pall Corporation, Port Washington NY) for 5 min at 11292 x g. Filtrates were then frozen at −80°C, lyophilized to dryness with LC/MS samples resuspended in 100 μL methanol containing 25 μM of the internal standards and analyzed as described in 2.3. GC/MS samples were derivatized and also analyzed as described in 2.6.

RESULTS

Soil Preparation and Properties

Organic rich saprolite soil was collected and immediately frozen to limit microbial activity. Prior to extraction for metabolites, soil samples were fumigated with chloroform in order to create a "whole soil extract" containing metabolites that microbes may have access to in this environment (i.e., both extracellular and intracellular metabolites). Furthermore, fumigation was used to limit metabolite processing by microbes that may occur during the extraction process. The TOC of the WEOC was 129.1 ppm (or 129.1 mg C kg^{-1} soil). Other physical properties of the ORFRC soil were: total N of 0.08%, organic matter (oxidizable) content of 1.5%, cation exchange capacity of 14.5 meq/100 g, soil pH of 7.7 and particle size (% sand/silt/clay) of 29/54/17.

Integrating GC/MS with LC/MS to Analyze Soil WEOC Composition and Concentrations

In order to expand on GC/MS WEOC annotations and composition previously reported from fumigated soil (Swenson et al., 2015b), this approach was complemented with HILIC LC/MS (Swenson et al., 2015a). Using both approaches here, a total of 96 water extractable soil metabolites were identified representing multiple classes of compounds such as amino acids, amino acid derivatives, mono- and di-carboxylic acids, nucleobases, nucleosides, osmolytes, sugars, sugar acids, and sugar alcohols. Using LC/MS, 85 metabolites were identified (Supplementary Table 1) with 66 of these confirmed by analytical standards. Eighteen of the metabolites identified using LC/MS were also identified by GC/MS. In addition, there were 11 metabolites identified using GC/MS that were not detected by LC/MS. Sugars were best chromatographically resolved by GC/MS versus LC/MS. Using authentic standards, this sugar pool was found to be comprised of hexoses (glucose, fructose, arabinose, mannose), dihexoses (trehalose, maltose) and the sugar alcohol mannitol.

Quantitative methods were developed for 36 compounds representing the biochemical classes with the most abundant MS

TABLE 1 | Metabolite formulation for soil defined medium 1 (SDM1).

Metabolite amendment	Formula	Molecular weight (g/mol)	Metabolite Class	WEOC mg/L	WEOC C ppm	SDM1 mg/L	SDM1 C ppm
Trehalose	$C_{12}H_{22}O_{11}$	342.3	dihexose	12.53	5.28	20	8.4
Fructose	$C_6H_{12}O_6$	180.2	hexose	6.70	2.68	15	5.81
Glucose	$C_6H_{12}O_6$	182.2	hexose	12.55	4.96	15	5.93
Mannitol	$C_6H_{14}O_6$	182.2	sugar alcohol	8.17	3.23	15	5.93
Maltose	$C_{12}H_{22}O_{11}$	342.3	dihexose	2.95	1.24	5	2.11
Alanine	$C_3H_7NO_2$	89.1	amino acid	0.93	0.37	2	0.81
Arabinose	$C_5H_{10}O_5$	150.1	pentose	0.52	0.21	2	0.8
Leucine	$C_6H_{13}NO_2$	131.2	amino acid	0.7	0.38	1	0.55
Mannose	$C_6H_{12}O_6$	180.2	hexose	0.54	0.21	1	0.4
Isoleucine	$C_6H_{13}NO_2$	131.2	amino acid	0.46	0.25	1	0.55
Arginine	$C_6H_{14}N_4O_2$	174.2	amino acid	0.22	0.09	0.5	0.21
Proline	$C_5H_9NO_2$	115.1	amino acid	0.2	0.1	0.5	0.26
Threonine	$C_4H_9NO_3$	119.1	amino acid	0.2	0.08	0.5	0.2
Lysine	$C_6H_{14}N_2O_2$	146.2	amino acid	0.2	0.1	0.5	0.25
Phenylalanine	$C_9H_{11}NO_2$	165.2	amino acid	0.19	0.13	0.5	0.33
Glycine	$C_2H_5NO_2$	75.1	amino acid	0.17	0.05	0.5	0.16
Uridine	$C_9H_{12}N_2O_6$	244.2	nucleoside	0.16	0.07	0.5	0.22
Glutamate	$C_5H_9NO_4$	147.1	amino acid	0.11	0.04	0.1	0.04
Serine	$C_3H_7NO_3$	105.1	amino acid	0.09	0.03	0.1	0.03
Adenosine	$C_{10}H_{13}N_5O_4$	267.2	nucleoside	0.08	0.04	0.1	0.04
Gamma-guanidinobutyric acid	$C_5H_{11}N_3O_2$	145.2	amino acid derivative	0.05	0.02	0.1	0.04
Hypoxanthine	$C_5H_4N_4O$	136.1	purine	0.04	0.02	0.1	0.04
Guanine	$C_5H_5N_5O$	151.1	nucleobase	0.004	0.002	0.1	0.04
TOTAL				55.59	19.58	81.1	33.15

For each metabolite, concentrations are shown for soil WEOC (in mg/L and C ppm), the amount added to SDM1 (mg/L) and the equivalent SDM1 carbon ppm.

ion signal (amino acids, sugars, nucleobases, and nucleosides). LC triple quadrupole MS was used for quantification of amino acids, nucleobases and nucleosides while sugars were quantified by GC/MS. Of these 36 quantified metabolites, 25 were above the limit of quantitation (Supplementary Table 1). Converting the metabolite concentrations into carbon concentrations, these 25 metabolites together accounted for a total of 20.0 mg C kg^{-1} soil with sugars accounting for 89.2% of the quantified metabolites. Trehalose, glucose, mannitol, fructose and maltose each accounted for 5.3, 5.0, 3.2, 2.7, and 1.2 mg C kg^{-1} soil respectively (Supplementary Table 1). Of the remainder, 9.8% were amino acids and only 0.8% were nucleobases and nucleosides.

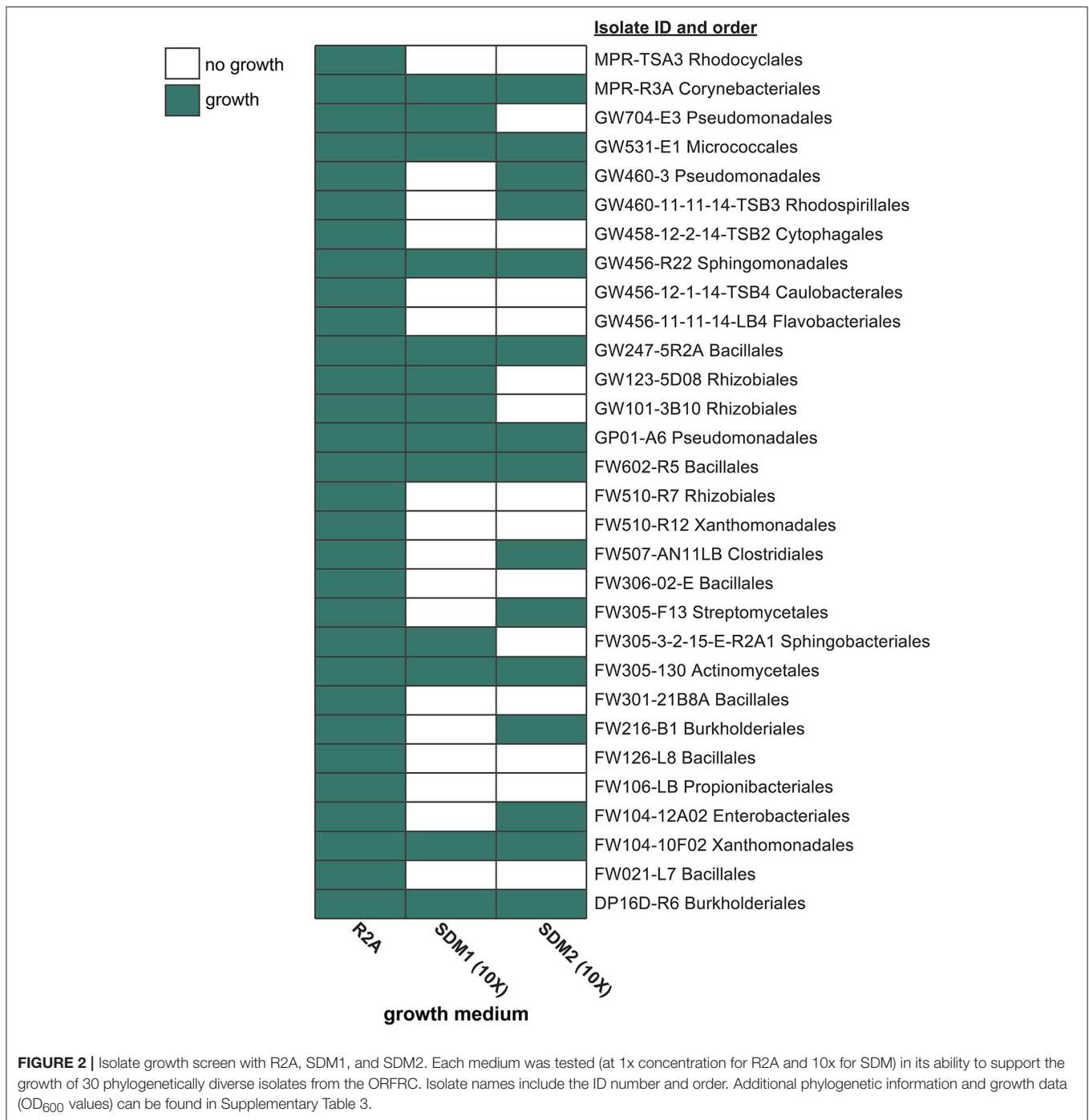## Molecular Weight Cutoff Filtration and TOC Quantitation of WEOC

Since the metabolomic methods used here would not be expected to detect all WEOC components, molecular weight cutoff filtration was used to help estimate the fraction of the WEOC pool quantified by MS. The TOC of the initial soil WEOC was 129.1 mg C kg^{-1} soil, which, after filtration through 1 kDa filters, was reduced to 53.3 mg C kg^{-1} soil. This reduction in TOC is consistent with the removal of biopolymers, colloids, trace lignin and cellulosic debris. Based on these TOC data, the 25

metabolites that were quantified from the WEOC, accounting for 20.0 mg C kg^{-1} soil, represented approximately 15.5% of the initial WEOC fraction and 37.5% of the <1 kDa filtered fraction.

Comparison of Soil Defined Media and Growth of Bacterial Isolates

Based on soil WEOC data, two SDM were formulated. The first (SDM1) was based on 23 abundant (quantified) WEOC metabolites (**Table 1**). The second (SDM2) contained twice as many compounds (Supplementary Table 2) to determine if expanding the number of metabolites present in SDM supported additional growth of bacterial isolates. These additional metabolites (including many amino acids, nucleobases and nucleosides) were detected in the soil WEOC, but were not quantified. Because of this, they were added at the lowest concentration of a quantified metabolite within the same class. For both SDM1 and SDM2, added metabolites served as microbial carbon sources and were supplemented with 1x Wolfe's vitamins and 1x Wolfe's mineral solutions (**Table 1**).

A broad assessment was then performed to compare the viability of these two media with 30 native ORFRC bacterial isolates (**Figure 2**, Supplementary Table 3A). In order to benchmark the viability of these media, isolate growth was also compared to the isolation medium, R2A, which supported

FIGURE 2 | Isolate growth screen with R2A, SDM1, and SDM2. Each medium was tested (at 1x concentration for R2A and 10x for SDM) in its ability to support the growth of 30 phylogenetically diverse isolates from the ORFRC. Isolate names include the ID number and order. Additional phylogenetic information and growth data (OD$_{600}$ values) can be found in Supplementary Table 3.

the growth of all 30 isolates (Supplementary Table 3B). At 10x concentrations, SDM1 supported the growth of 13 isolates (Supplementary Table 3C) while 15 isolates grew on SDM2 (Supplementary Table 3D) (**Figure 2**).

Isolate Exometabolomics Analysis Using SDM1

Time-series exometabolomics studies can help delineate substrate preference of isolates as an additional measure of resource partitioning. Since SDM1 appeared to be a

simple, yet viable medium for many isolates, this medium was selected to test its suitability for exometabolomics with *Pseudomonas* sp. FW300-N2E2. Distinct patterns of substrate depletion were observed as shown in **Figure 3**. Compounds such as arginine, proline, glutamic acid, guanine and hypoxathine were already depleted by the first time point of 3 h. During later exponential growth phase (6 and 9 h), a second pool of compounds were depleted including alanine, isoleucine, gamma-guanadinobutyric acid, glycine, leucine, lysine, phenyalanine, serine, and threonine. Interestingly, some

typical bacterial nutrients were not depleted until stationary phase with glucose detected up to 12 h and mannose and fructose up to the final 24 h sampling. Several metabolites persisted at near initial concentrations for the duration of the experiment including the abundant dihexoses trehalose and maltose, the hexose arabinose, the sugar alcohol mannitol and the nucleoside uridine.

DISCUSSION

Integrating Qualitative GC/MS with LC/QTOF-MS to Analyze WEOC Composition

The objective of this study was to develop viable and relevant defined media based on metabolomics analyses of an environment of interest to enable exometabolomic characterization of bacterial isolate resource partitioning. Here we used a "whole soil" approach to examine all of the water soluble metabolites that could be extracted from fumigated soil. Fumigation, used to prevent microbial activity during the extraction process also lyses cells. Hence, this extract represents the total water extractable soil metabolite profile of potentially accessible metabolites. This "whole soil" analysis is also supported by the finding that much of soil organic matter consists of microbial products and lysed cells (Kogel-Knabner, 2002; Schmidt et al., 2011). Here we combined two metabolomics approaches, GC/MS (Swenson et al., 2015b) and HILIC LC/MS (Swenson et al., 2015a) for qualitative and quantitative analysis of soil. In total, 96 metabolites were identified using a combination of authentic standards, MS/MS and spectral library matching (Supplementary Table 1).

While this work focused on a single soil for the development of a metabolomics workflow for defined media preparation, it is informative to see how our metabolites detected lie within the context of other soil metabolomics studies. Overall, the range of metabolite classes detected, including amino acids, amino acid derivatives, mono- and di-carboxylic acids, nucleobases, nucleosides, osmolytes, sugars, sugar acids and sugar alcohols are consistent across many soil studies. The detection of trehalose and other compatible solutes (betaine and proline betaine) as well as additional quaternary ammonium compounds (acetylcarnitine, carnitine, and choline) have been reported for other soils (Baran et al., 2013; Warren, 2014; Bouskill et al., 2016). Metabolites such as acetylcarnitine, citrulline, cytosine, gamma-aminobutyric acid, and nicotinic acid are consistent with a key finding from Warren (2013) who showed that a diverse pool of nonpeptide organic N exists in soils.

The pool of organic acids in our WEOC samples was found to be diverse, composed of aliphatic di-carboxylic acids (malic, maleic, succinic) and aromatic carboxylic acids (benzoic, salicylic and shikimic). This is consistent with the observation of these low molecular weight carboxylic acids in many soils as reviewed by Strobel (2001) and the ability of aqueous shaking extraction to desorb higher concentrations of these metabolites than solution displacement techniques such as lysimeter collection or soil centrifugation. Our samples, which originated from an upper B horizon subsoil, may be associated with limited amounts of aliphatic mono-carboxylic acids (not detected in our WEOC extracts) potentially due to rapid microbial turnover of these compounds. Furthermore, the di-carboxylic and aromatic carboxylic acids that were detected here may be associated with a limited nutrient pool *in situ* that became desorbed during the aqueous shaking extraction (Strobel, 2001).
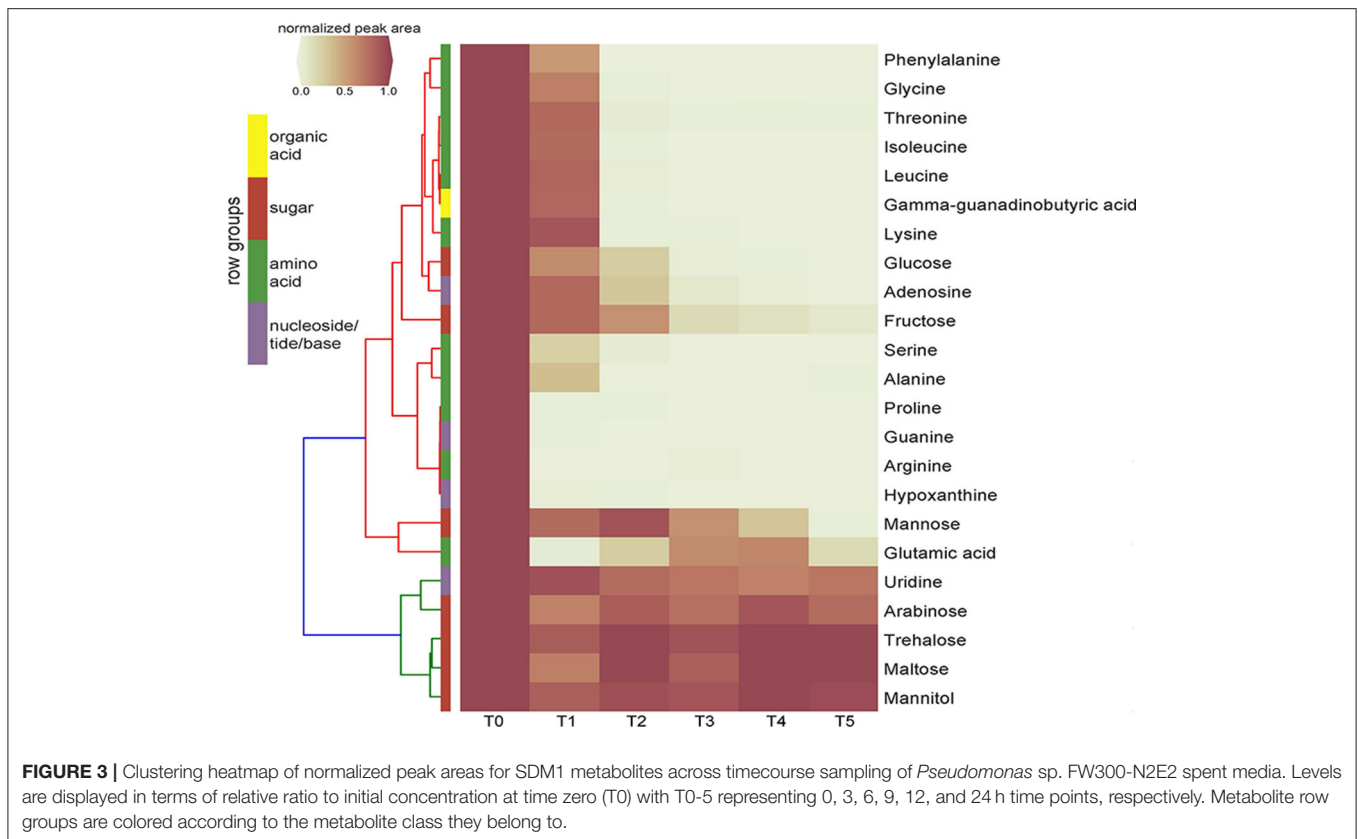
Analyzing the TOC Content of Soil WEOC

The TOC level observed in our soil WEOC (129.1 mg C kg^{-1} soil) was found to be consistent with many previous reports on soil extracts, noting that differences in soil types and extraction methods (including fumigation) may dramatically affect these values. With this caveat, our TOC level is surprisingly close to the 146 ppm reported for a soil (leachate) solution from a eutric cambisol grassland (Jones and Willett, 2006). However, this study displays the effect of extraction techniques on these values by reporting lower TOC values (55–70 mg kg^{-1} soil) for the same soil following aqueous shaking (rather than leachate collection). Another study that is consistent with our TOC results focused on evaluating dissolved organic matter dynamics in Greek vineyard soils (Christou et al., 2005). They report drastic seasonal variation of dissolved organic carbon levels in soil solutions ranging from approximately 100–400 ppm over a 12 month period (Christou et al., 2005) and displayed substantial differences between topsoil and subsoil with WEOC levels decreasing from 89 to 58 mg C kg^{-1} soil (Christou et al., 2006).

Since soluble polymers and small particles presumably account for a large fraction of the WEOC, we wanted to determine the fraction of this total carbon pool accounted for by the 25 quantified metabolites. The TOC level of the soil WEOC was found to be reduced by more than half (129.1–53.3 ppm) following 1 kDa filtration, representing the most relevant fraction to our metabolomics methods given that that one operational definition of metabolites is molecules less than <1 kDa (Holmes et al., 2008). Since microbes are limited in their ability to directly uptake macromolecules and are dependent on extracellular deconstruction followed by transporting the resulting metabolites, these molecules smaller than 1 kDa likely represent the most directly accessible fraction of the WEOC for microbes.

Characterizing the LMWOS Fraction of WEOC by Quantitative LC/QQQ-MS and GC/MS Analysis

Based on the molar concentrations of the 25 quantified metabolites in the WEOC (consisting mostly of carbohydrates and amino acids) and the number of carbons in each metabolite, we determined that these quantified metabolites accounted for 20.0 mg C kg^{-1} in the fumigated soil sample. This represents 15.5% of the WEOC and 37.5% of the <1 kDa metabolite pool, indicating that even using two highly-sensitive analytical approaches, the majority of soil metabolites remain rare, unidentified or undetected. This pool of metabolites could represent ones for which analytical standards are not

FIGURE 3 | Clustering heatmap of normalized peak areas for SDM1 metabolites across timecourse sampling of *Pseudomonas* sp. FW300-N2E2 spent media. Levels are displayed in terms of relative ratio to initial concentration at time zero (T0) with T0-5 representing 0, 3, 6, 9, 12, and 24 h time points, respectively. Metabolite row groups are colored according to the metabolite class they belong to.

readily available (thus making identification difficult) including secondary metabolites, antibiotics, peptides, or simply those undetectable by LC/MS or GC/MS. Numerous other studies support this assertion of the many unannotated chemical features present in soil LMWOS (Ohno et al., 2010; Warren, 2013; Baran et al., 2015) highlighting the value of using defined metabolite mixtures for exometabolomic characterization of microbial resource use.

The soil metabolites that we detected have a very uneven abundance distribution. Specifically, of the quantified metabolites, sugars represented approximately 89.2% of the quantified metabolite TOC pool, with a single metabolite, trehalose accounting for 5.3 mg C kg^{-1} soil (29.7% of the carbon within the quantified sugar pool and 9.9% of the TOC of the <1 kDa WEOC fraction). The abundance of trehalose is not too surprising given its role as an important bacterial carbon source and osmoprotectant (Argüelles, 2000). Of the amino acids, only alanine, valine, leucine and isoleucine were at levels above 200 μg C kg^{-1} soil while all nucleosides and nucleobases except for uridine were at trace levels at or below 50 μg C kg^{-1} soil. While this type of metabolite distribution is specific to the ORFRC soil, we can compare these values with previous reports. Where we found total quantified amino acids to be 4.08 mg kg^{-1} soil, Fischer et al. (2007) examined the composition of Haplic Luvisol soil leachate and reports an amino acid content of 281.1 μg kg^{-1} soil, considerably lower than ours likely because fumigation in our study released many of these intracellular metabolites. However, they observed a

similar ranked composition (e.g., alanine, leucine and isoleucine ranked among the highest; Fischer et al., 2007). We found the ratio of carbohydrates to amino acids to be 10.8 by weight, which is consistent with Hertenberger et al. (2002) who reports ratios ranging from 6.4 to 17.4, but does not align with more amino acids as reported by Fischer et al. (2007) with a ratio of 0.4. While these differences may simply be due to actual differences between soils, bias introduced during soil preparation and extraction is likely another important factor given that extraction techniques ranged from mild soil leaching used by Fischer et al. (2007) to the fumigation-aqueous extraction (this study) to a more aggressive acetone-water extraction used by Hertenberger et al. (2002).

Preparation and Evaluation of Defined Media Based on Soil WEOC Composition

Both SDM were found to support the rapid growth of a range of taxa (**Figure 2**). SDM1 (10x), which contained 23 abundant WEOC metabolites, supported the growth of 13 out of the 30 isolates tested. Although doubling the number of metabolites in the medium (SDM2) slightly increased the number of isolates that grew (15 out of 30), it is important to note that the additional metabolites added to SDM2 were at relatively low concentrations (Supplementary Table 2). The finding that both SDM were found to be less viable than R2A is not surprising given that the bacterial strains analyzed here were isolated using R2A. Furthermore, R2A is rich in amino acids and peptides while SDM1 and SDM2 have a high sugar content, indicating potential unique and developed substrate preferences of these bacteria. Regardless, the

isolate screen demonstrates the viability of soil-based synthetic media that can likely be enhanced further by addition of certain metabolites or by alternate formulations that increase the relative concentrations of specific classes of compounds.

The application of SDM1 to investigate the substrate preferences of *Pseudomonas* sp. FW300-N2E2 revealed an interesting pattern of substrate utilization. We observed rapid depletion of arginine, glutamate, proline, guanine, and hypoxanthine by the 3 h time point associated with the onset of logarithmic growth followed by consumption of alanine, isoleucine, gamma-guanadinobutyric acid, glycine, leucine, lysine, phenyalanine, serine, and threonine (depleted by 6 h) then finally depletion of sugars including glucose. Several sugars such as arabinose, maltose, mannitol, and trehalose were not depleted at all suggesting that either uptake is equivalent to efflux or that this bacterium lacks a sufficient set of transporters to use these abundant resources, supporting the view that it is important to perform exometabolomics experiments on media relevant, whenever possible, to their native environment. The utilization profile of this *Pseudomonas* sp. is consistent with a previous report for *Pseudomonas aeruginosa* PAO1 in which growth on a complex tryptone medium resolved an initial preference for select amino acids including leucine, proline, serine, and threonine while the carbohydrate utilization was very similar to our results, namely, while glucose was metabolized, maltose, mannitol, mannose, and trehalose were not (Frimmersdorf et al., 2010). Since *Pseudomonas* sp. FW300-N2E2 was isolated using R2A, which is rich in many of the preferentially consumed compounds (such as amino acids) perhaps it is not surprising that it preferentially uses resources that, while relatively rare in the soil environment, are abundant in the isolation medium. We anticipate that media prepared based on soil metabolite analyses may have additional utility for isolating organisms that utilize the major measurable carbon sources within those environments.

The success of the SDM demonstrates the ability to use soil metabolomics to develop defined media based on metabolites known to be abundant in soils (including both intracellular and extracellular) and at concentrations relevant to the native environment. This also indicates the potential, in agreement with previous approaches, of using single or complex amendment of minimal media to isolate previously unculturable soil bacteria (Sait et al., 2002; Joseph et al., 2003). While R2A was suitable for the isolation of the 30 tested bacterial isolates, this medium may lack certain nutrients or appropriate (lower) concentrations of metabolites that could enable isolation of unique microbes. Specifically, microbial transport and regulatory systems, particularly catabolite repression, are responsive to the absolute concentrations of metabolites in the environment. Thus, evaluating substrate utilization in defined media relevant to their habitat can greatly improve exometabolomic studies.

There are a number of costs or challenges associated with creating media that reflect microbial habitats for laboratory experimentation that must be considered. First, we estimate that the cost of SDM is higher than R2A given that it is a defined media based on pure compounds. However, for certain experiments, such as exometabolomic analysis, the advantages outweigh the increased costs where using defined media allows complete control and tracking of metabolites of interest. In other cases it may be desirable to simply add components of SDM to R2A. Secondly, soil fumigation undoubtedly over-represents the accessible metabolite pool and this type of bulk analysis creates averaged metabolite compositions unlike any particular niche within the soil. In addition, as we have shown here, many metabolites cannot be identified or go undetected and with the tools used here, we did not analyze inorganic components, which may be essential for microbial growth. Complementing mass spectrometry with other types of spectroscopy, ICP-MS, may help address this. Finally, due to the extreme heterogeneity between and within soil ecosystems, our particular findings on soil metabolite abundances should not be generalized except for the dozen or so metabolites that we have found to be consistent with other reports.

Implications for Other Environments

These methods, using exometabolomics analysis of environmental samples to prepare defined media for the study of organisms from that environment, are likely applicable to a diversity of environments. Extension of this approach to diverse soil types offers both the exciting possibility of helping connect soil metabolite composition to soil microbial community composition and development of generalizable defined soil media. While these defined media were focused on carbon, an important extension would be to also account for other critical elements especially organic nitrogen and phosphorus. The great advantage of these defined media is that they enable quantitative analysis of resource utilization by soil microorganisms.

One exciting implication of this analysis is that rare metabolites that were not quantified, together may account for a significant portion of the carbon (WEOC) in this sample, similar to how rare microbes collectively account for a large portion of microbiomes. Synthesis of this view with previous results showing that bacteria from this soil (and biological soil crusts) utilize largely non-overlapping metabolites suggests there is coupling between microbial diversity and soil metabolite diversity. These results showing the unevenness of water soluble metabolites may further support the traditional view of copiotrophic and oligotrophic organisms. Specifically, there are a low-diversity of organisms competing for the abundant resources (copiotrophs) and a high diversity of rare organisms using low-abundance substrates (oligotrophs) (Upton and Nedwell, 1989; Konopka et al., 1998). It should be emphasized that this is highly speculative and studies of multiple soil samples and multiple sites would be required to support generalization to both this particular site and to soils in general.

CONCLUSION

This study used soil metabolomic analyses to characterize the low molecular weight organic matter composition to formulate defined media intended to approximate the qualitative and quantitative composition of microbe bioavailable carbon from a specific study site. Composition and carbon concentrations

were found to align well with related studies with these soil metabolites having a very uneven quantitative distribution (e.g., trehalose accounting for 9.9% of the <1 kDa WEOC fraction), analogous to the uneven soil microbial community structure. The defined media that were synthesized to reflect the soil WEOC composition were found to support the growth of up to 15 out of 30 phylogenetically diverse isolates. A detailed study of a single isolate, *Pseudomonas* sp. FW300-N2E2 showed that this isolate rapidly depleted guanine, serine, leucine and hypoxanthine while several metabolites including the most abundant disaccharides were not utilized from SDM1. We anticipate that this approach of preparing environmentally relevant defined media will be applicable to diverse environments to enable more ecologically relevant isolation and examination of microbial substrate utilization.

AUTHOR CONTRIBUTIONS

SJ and TN designed the study and the experiments. AR and TH performed soil sampling. SJ and RL performed soil extractions, NPOC analysis and isolate cultivation for longitudinal exometabolomics. AA and RC performed isolate growth screen on defined media supplied by RL. SJ acquired and analyzed all mass spectrometry data. SJ, RL, and TS interpreted results. SJ, TS, and TN wrote the manuscript, with contributions from all co-authors.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02618/full#supplementary-material

REFERENCES

Argüelles, J. C. (2000). Physiological roles of trehalose in bacteria and yeasts: a comparative analysis. *Arch. Microbiol.* 174, 217–224. doi: 10.1007/s002030000192

Baran, R., Bowen, B. P., Price, M. N., Arkin, A. P., Deutschbauer, A. M., and Northen, T. R. (2013). Metabolic footprinting of mutant libraries to map metabolite utilization to genotype. *ACS Chem. Biol.* 8, 189–199. doi: 10.1021/cb300477w

Baran, R., Brodie, E. L., Mayberry-Lewis, J., Hummel, E., Da Rocha, U. N., Chakraborty, R., et al. (2015). Exometabolite niche partitioning among sympatric soil bacteria. *Nat. Commun.* 6:8289. doi: 10.1038/ncomms9289

Bouskill, N. J., Wood, T. E., Baran, R., Ye, Z., Bowen, B. P., Lim, H., et al. (2016). Belowground response to drought in a tropical forest soil. I. changes in microbial functional potential and metabolism. *Front. Microbiol.* 7:525. doi: 10.3389/fmicb.2016.00525

Boyer, J. N., and Groffman, P. M. (1996). Bioavailability of water extractable organic carbon fractions in forest and agricultural soil profiles. *Soil Biol. Biochem.* 28, 783–790. doi: 10.1016/0038-0717(96)00015-6

Butler, J. L., Williams, M. A., Bottomley, P. J., and Myrold, D. D. (2003). Microbial community dynamics associated with rhizosphere carbon flow. *Appl. Environ. Microbiol.* 69, 6793–6800. doi: 10.1128/AEM.69.11.6793-6800.2003

Christou, M., Avramides, E. J., and Jones, D. L. (2006). Dissolved organic nitrogen dynamics in a Mediterranean vineyard soil. *Soil Biol. Biochem.* 38, 2265–2277. doi: 10.1016/j.soilbio.2006.01.025

Christou, M., Avramides, E. J., Roberts, J. P., and Jones, D. L. (2005). Dissolved organic nitrogen in contrasting agricultural ecosystems. *Soil Biol. Biochem.* 37, 1560–1563. doi: 10.1016/j.soilbio.2005.01.025

Fernandes, M. F., Saxena, J., and Dick, R. P. (2013). Comparison of whole-cell fatty acid (MIDI) or phospholipid fatty acid (PLFA) extractants as biomarkers to profile soil microbial communities. *Microb. Ecol.* 66, 145–157. doi: 10.1007/s00248-013-0195-2

Fischer, H., Meyer, A., Fischer, K., and Kuzyakov, Y. (2007). Carbohydrate and amino acid composition of dissolved organic matter leached from soil. *Soil Biol. Biochem.* 39, 2926–2935. doi: 10.1016/j.soilbio.2007.06.014

Frimmersdorf, E., Horatzek, S., Pelnikevich, A., Wiehlmann, L., and Schomburg, D. (2010). How *Pseudomonas aeruginosa* adapts to various environments: a metabolomic approach. *Environ. Microbiol.* 12, 1734–1747. doi: 10.1111/j.1462-2920.2010.02253.x

Guigue, J., Mathieu, O., Lévêque, J., Mounier, S., Laffont, R., Maron, P. A., et al. (2014). A comparison of extraction procedures for water-extractable organic matter in soils. *Eur. J. Soil Sci.* 65, 520–530. doi: 10.1111/ejss.12156

Haney, R. L., Franzluebbers, A. J., Jin, V. L., Johnson, M.-V., Haney, E. B., White, M. J., et al. (2012). Soil organic C:N vs. water-extractable organic C:N. *Open J. Soil Sci.* 2, 269–274. doi: 10.4236/ojss.2012.23032

Hertenberger, G., Zampach, P., and Bachmann, G. (2002). Plant species affect the concentration of free sugars and free amino acids in different types of soil. *J. Plant Nutr. Soil Sci.* 165, 557–565. doi: 10.1002/1522-2624(200210)165:5<557::AID-JPLN1111557>3.0.CO;2-G

Holmes, E., Wilson, I. D., and Nicholson, J. K. (2008). Metabolic phenotyping in health and disease. *Cell* 134, 714–717. doi: 10.1016/j.cell.2008.08.026

Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., et al. (2010). MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* 45, 703–714. doi: 10.1002/jms.1777

Jones, D., and Willett, V. (2006). Experimental evaluation of methods to quantify dissolved organic nitrogen (DON) and dissolved organic carbon (DOC) in soil. *Soil Biol. Biochem.* 38, 991–999. doi: 10.1016/j.soilbio.2005.08.012

Joseph, S. J., Hugenholtz, P., Sangwan, P., Osborne, C. A., and Janssen, P. H. (2003). Laboratory cultivation of widespread and previously uncultured soil bacteria. *Appl. Environ. Microbiol.* 69, 7210–7215. doi: 10.1128/AEM.69.12.7210-7215.2003

Kind, T., Wohlgemuth, G., Lee do, Y., Lu, Y., Palazoglu, M., Shahbaz, S., et al. (2009). FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal. Chem.* 81, 10038–10048. doi: 10.1021/ac9019522

Kogel-Knabner, I. (2002). The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter. *Soil Biol. Biochem.* 34, 139–162. doi: 10.1016/S0038-0717(01)00158-4

Konopka, A., Oliver, L., and Turco, R. F. Jr. (1998). The use of carbon substrate utilization patterns in environmental and ecological microbiology. *Microb. Ecol.* 35, 103–115. doi: 10.1007/s002489900065

Liebeke, M., and Lalk, M. (2014). *Staphylococcus aureus* metabolic response to changing environmental conditions - a metabolomics perspective. *Int. J. Med. Microbiol.* 304, 222–229. doi: 10.1016/j.ijmm.2013.11.017

Miller, A., Schimel, J., Meixner, T., Sickman, J., and Melack, J. (2005). Episodic rewetting enhances carbon and nitrogen release from chaparral soils. *Soil Biol. Biochem.* 37, 2195–2204. doi: 10.1016/j.soilbio.2005.03.021

Ohno, T., He, Z., Sleighter, R. L., Honeycutt, C. W., and Hatcher, P. G. (2010). Ultrahigh resolution mass spectrometry and indicator species analysis to identify marker components of soil- and plant biomass- derived organic matter fractions. *Environ. Sci. Technol.* 44, 8594–8600. doi: 10.1021/es101089t

Ohno, T., Parr, T. B., Gruselle, M. C., Fernandez, I. J., Sleighter, R. L., and Hatcher, P. G. (2014). Molecular composition and biodegradability of soil organic matter: a case study comparing two new England forest types. *Environ. Sci. Technol.* 48, 7229–7236. doi: 10.1021/es405570c

Reasoner, D. J., and Geldreich, E. E. (1985). A new medium for the enumeration and subculture of bacteria from potable water. *Appl. Environ. Microbiol.* 49, 1–7.

Sait, M., Hugenholtz, P., and Janssen, P. H. (2002). Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. *Environ. Microbiol.* 4, 654–666. doi: 10.1046/j.1462-2920.2002.00352.x

Schimel, J. P., and Mikan, C. (2005). Changing microbial substrate use in Arctic tundra soils through a freeze-thaw cycle. *Soil Biol. Biochem.* 37, 1411–1418. doi: 10.1016/j.soilbio.2004.12.011

Schmidt, M. W., Torn, M. S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I. A., et al. (2011). Persistence of soil organic matter as an ecosystem property. *Nature* 478, 49–56. doi: 10.1038/nature10386

Smith, C. A., O'Maille, G., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., et al. (2005). METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* 27, 747–751. doi: 10.1097/01.ftd.0000179845.53213.39

Strobel, B. W. (2001). Influence of vegetation on low-molecular-weight carboxylic acids in soil solution—a review. *Geoderma* 99, 169–198. doi: 10.1016/S0016-7061(00)00102-6

Swenson, T. L., Bowen, B. P., Nico, P. S., and Northen, T. R. (2015a). Competitive sorption of microbial metabolites on an iron oxide mineral. *Soil Biol. Biochem.* 90, 34–41. doi: 10.1016/j.soilbio.2015.07.022

Swenson, T. L., Jenkins, S., Bowen, B. P., and Northen, T. R. (2015b). Untargeted soil metabolomics methods for analysis of extractable organic matter. *Soil Biol. Biochem.* 80, 189–198. doi: 10.1016/j.soilbio.2014.10.007

Thorgersen, M. P., Lancaster, W. A., Vaccaro, B. J., Poole, F. L., Rocha, A. M., Mehlhorn, T., et al. (2015). Molybdenum availability is key to nitrate removal in contaminated soil groundwater environments. *Appl. Environ. Microbiol.* 81, 4976–4983. doi: 10.1128/AEM.00917-15

Tucker, C. L., Young, J. M., Williams, D. G., and Ogle, K. (2014). Process-based isotope partitioning of winter soil respiration in a subalpine ecosystem reveals importance of rhizospheric respiration. *Biogeochemistry* 121, 389–408. doi: 10.1007/s10533-014-0008-9

Upton, A. C., and Nedwell, D. B. (1989). Nutritional flexibility of oligotrophic and copiotrophic Antarctic bacteria with respect to organic substrates. *FEMS Microbiol. Lett.* 62, 1–6. doi: 10.1111/j.1574-6968.1989.tb03651.x

Vance, E. D., Brookes, P. C., and Jenkinson, D. S. (1987). An extraction method for measuring soil microbial biomass C. *Soil Biol. Biochem.* 19, 703–707. doi: 10.1016/0038-0717(87)90052-6

Warren, C. R. (2013). High diversity of small organic N observed in soil water. *Soil Biol. Biochem.* 57, 444–450. doi: 10.1016/j.soilbio.2012.09.025

Warren, C. R. (2014). Response of osmolytes in soil to drying and rewetting. *Soil Biol. Biochem.* 70, 22–32. doi: 10.1016/j.soilbio.2013.12.008

Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y., et al. (2013). HMDB 3.0–the human metabolome database in 2013. *Nucleic Acids Res.* 41, D801–D807. doi: 10.1093/nar/gks1065