bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 # Construction of soil defined media using quantitative exometabolomic
2 # analysis of soil metabolites

3

4 Stefan Jenkins[1#], Tami L. Swenson[1], Rebecca Lau[1], Andrea Rocha[2,3], Alex Aaring[1], Terry C. Hazen[2,4],
5 Romy Chakraborty[1], Trent Northen[1,5*]

6

7

8

9 **Affiliations**

10 1. Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley CA 94720
11 2. Oak Ridge National Laboratory, 1 Bethel Valley Rd, Oak Ridge TN 37831
12 3. Geosyntec Consultants, 180A Market Place Blvd., Knoxville, TN 37922
13 4. University of Tennessee, 676 Dabney Hall, Knoxville TN 37996
14 5. Joint Genome Institute, Department of Energy, 2800 Mitchell Dr, Walnut Creek CA 94598

15

16 *Corresponding author: trnorthen@lbl.gov
17 #Current address: Intrexon Corporation, 1750 Kraft Dr, Blacksburg VA 24060

18

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

19 **Abstract**

20 Exometabolomics enables analysis of metabolite utilization of low molecular weight organic
21 substances by soil isolates. Environmentally-based defined media are needed to examine ecologically
22 relevant patterns of substrate utilization. Here, we describe an approach for the construction of defined
23 media using untargeted characterization of water soluble soil metabolites. To broadly characterize soil
24 metabolites, both liquid chromatography mass spectrometry (LC/MS) and gas chromatography mass
25 spectrometry (GC/MS) were used. With this approach, 96 metabolites were identified, including amino
26 acids, amino acid derivatives, sugars, sugar alcohols, mono- and di-carboxylic acids, osmolytes,
27 nucleobases, and nucleosides. From this pool of metabolites, 25 were quantified. Water soluble organic
28 carbon was fractionated by molecular weight and measured to determine the fraction of carbon accounted
29 for by the quantified metabolites. This revealed that, much like soil microbial community structures, these
30 soil metabolites have an uneven quantitative distribution, with a single metabolite, trehalose accounting
31 for 9.9 percent of the (< 1 kDa) water extractable organic carbon. This quantitative information was used
32 to formulate two soil defined media (SDM), one containing 23 metabolites (SDM1) and one containing
33 46 (SDM2). To evaluate SDM for supporting the growth of bacteria found at this field site, we examined
34 the growth of 30 phylogenetically diverse soil isolates obtained using standard R2A medium. The simpler
35 SDM1 supported the growth of up to 13 isolates while the more complex SDM2 supported up to 25
36 isolates. One isolate, *Pseudomonas corrugata* strain FW300-N2E2 was selected for a time-series
37 exometabolomics analysis to investigate SDM1 substrate preferences. Interestingly, it was found that this
38 organism preferred lower-abundance substrates such as guanine, glycine, proline and arginine and glucose
39 and did not utilize the more abundant substrates maltose, mannitol, trehalose and uridine. These results
40 demonstrate the viability and utility of using exometabolomics to construct a tractable environmentally
41 relevant media. We anticipate that this approach can be expanded to other environments to enhance
42 isolation and characterization of diverse microbial communities.

43

44

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

45 **Highlights**

46 • LC/MS and GC/MS analyses of soil extracts revealed a diversity of 96 metabolites.
47 • Soil defined media were constructed based on water extractable soil metabolomics data.
48 • The defined media supported the growth of 25 out of 30 bacterial isolates.
49 • Exometabolomics demonstrated preferential consumption of amino acids for one isolate.
50 • These media can be used to understand environmentally relevant microbial substrate preferences.

51

52 **Keywords**

53 soil defined media, water extractable organic matter, exometabolomics, liquid chromatography mass
54 spectrometry, gas chromatography mass spectrometry

55

56 **Abbreviations**

57 DOM dissolved organic matter; SOM soil organic matter; WEOC water extractable organic carbon;
58 LMWOS low molecular weight organic substances; SDM(1/2) soil defined media (1 and 2); R2A
59 Reasoner's 2A agar medium; ORFRC Oak Ridge Field Research Center; LC/MS liquid chromatography
60 mass spectrometry; GC/MS gas chromatography/ mass spectrometry; TOC total organic carbon; HILIC
61 hydrophilic interaction liquid chromatography

62

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

63 **1. Introduction**

64 Soil organic matter, historically considered to be composed of large polymeric humic substances,
65 is now thought to largely consist of microbial products (Schmidt et al., 2011). Traditionally, the water
66 soluble fraction of soil carbon, known as dissolved organic carbon or matter (DOC/DOM), is defined as
67 the fraction that passes through a 0.45 μm filter (Ohno et al., 2014). Many microbes process
68 macromolecular substrates such as plant biomass extracellularly and uptake the resulting low molecular
69 weight organic substances (LMWOS). Thus, it is not surprising that the water extractable organic carbon
70 (WEOC)(Boyer and Groffman, 1996; Guigue et al., 2014) is associated with high microbial activity and
71 soil respiration (Haney et al., 2012). Examination of the composition of LMWOS using metabolomic
72 methods has revealed a diversity of small molecule metabolites (Warren, 2014; Swenson et al., 2015b).
73 Recently, we observed exometabolite niche partitioning in sympatric soil microbes indicating that there
74 can be a strong linkage between the composition of soil exometabolites and microbial community
75 structure (Baran et al., 2015). This reinforces long-standing views that it is desirable for culture media to
76 approximate the conditions, especially the quality and abundance of metabolites found within their
77 habitat. This raises the exciting possibility that soil metabolomics methods can help inform the
78 development of relevant culture media to enable laboratory studies investigating microbial resource
79 partitioning.

80 Most of our understanding of the biology of soil microbes is based on the small fraction of
81 microbes that have been successfully cultivated in isolation. Reasoner's 2A agar medium (R2A) is one of
82 the most widely used for isolations and was developed for the cultivation of bacteria found in potable
83 water (Reasoner and Geldreich, 1985). Its effectiveness may be attributable to its rich nutrients (peptone,
84 casamino acids and yeast extract). However, the composition of R2A is dramatically different from the
85 composition of soil LMWOS, and therefore has limited ecological relevance. There are many promising
86 new technologies for isolating soil microbes with direct connection to the native small molecule
87 environment (Pham and Kim, 2012) such as the use of transwell plates, soil substrate membranes, and

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

88 recently the iCHIP that enables diffusion of native DOM into the plates for isolations (Svenning et al.,
89 2003; Ferrari et al., 2008; Ling et al., 2015).
90 Exometabolomics enables the study of the transformation of the small molecule environment by
91 bacteria and other microorganisms by comparing spent media to non-inoculated controls, typically using
92 mass spectrometry. This direct assessment of phenotype provides high-resolution of metabolite depletion
93 which is emerging as a powerful complement to existing broader techniques such as bulk soil respiration
94 and carbon transformation measurements (Butler et al., 2003; Miller et al., 2005; Schimel and Mikan,
95 2005; Fernandes et al., 2013; Tucker et al., 2014). Previous studies have demonstrated the capability of
96 exometabolomics in providing a functional complement to genomic and transcriptomic data (Baran et al.,
97 2013; Liebeke and Lalk, 2014). Recently, exometabolomics was used with soil microbes and a microbial
98 extract media to discover exometabolite niche partitioning for the 13 isolates and 2 environments
99 examined (Baran et al., 2015). However, a limitation of this approach is that many of the metabolites in
100 these environmentally relevant rich media could not be identified and presumably many others were not
101 detected. For many exometabolite profiling experiments it would be desirable to have an environmentally
102 relevant defined medium such that all of the metabolites are accounted for and the observed patterns of
103 metabolite utilization are relevant to the field site.
104 Here we describe the construction of defined media for exometabolomics experiments based on
105 untargeted analysis of metabolites from a specific soil sample of interest (Figure 1). Water extractable
106 metabolites, referred to here as WEOC, were first qualitatively characterized using both liquid
107 chromatography mass spectrometry (LC/MS) and gas chromatography mass spectrometry (GC/MS).
108 From these data, a subset of metabolites was selected for absolute quantification to assist in formulating
109 defined media that approximates the quality and quantities of metabolites within the soil. We evaluated
110 the ability of the resulting media to support the growth of 30 phylogenetically diverse isolates from the
111 Oak Ridge Field Research Center (ORFRC) and performed time series characterization of the substrate
112 preferences of a single isolate.
113

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

114 **2 Methods and Materials**

115 **2.1 Chemicals**

116 Glucose (CAS 50-99-7) was from Amresco (Solon, OH). LC/MS-grade methanol (CAS 67-56-1)
117 and water were from J.T. Baker (Avantor Performance Materials, Center Valley, PA). N-methyl-N-
118 trimethylsilytrifluoroacetamide (MSTFA) containing 1% trimethylchlorosilane (TMCS) was from Restek
119 (Bellafonte, PA). Acetonitrile (CAS 75-05-8), adenosine (CAS 58-61-7), arabinose (CAS 28697-53-2),
120 biotin (CAS 58-85-5), calcium chloride (CAS 10043-52-4), calcium pantothenate (CAS 137-08-6), cobalt
121 nitrate (CAS 10026-22-9), d_{27}-myristic acid (CAS 60658-41-5), ethylenediaminetetraacetic acid (EDTA;
122 CAS 6381-92-6), folic acid (CAS 59-30-3), fructose (CAS 57-48-7), galactose (CAS 59-23-4), gamma-
123 guanidinobutyric acid (CAS 463-00-3), guanine (CAS 73-40-5), hypoxanthine (CAS 68-94-0), iron
124 sulfate (CAS 7782-63-0), magnesium sulfate (CAS 10034-99-8), maltitol (CAS 585-88-6), maltose (CAS
125 6363-53-7), manganese sulfate (CAS 10034-96-5), mannose (CAS 3458-28-4), mannitol (69-65-8),
126 methoxyamine (CAS 593-56-6), nicotinic acid (CAS 59-67-6), p-aminobenzoic acid (CAS 150-13-0),
127 monobasic potassium phosphate (CAS 7778-77-0), pyridine (CAS 110-86-1), pyridoxine hydrochloride
128 (CAS 58-56-0), raffinose (CAS 17629-30-0), riboflavin (CAS 83-88-5), rutinose (CAS 90-74-4), sodium
129 chloride (CAS 7647-14-5), sucrose (CAS 57-50-1), thiamine (CAS 67-03-8), thioctic acid (CAS 1077-28-
130 7), trehalose dehydrate (CAS 6138-23-4), uridine (CAS 58-96-8), vitamin B12 (CAS 68-19-9), xylitol
131 (CAS 87-99-0), xylobiose (CAS 6860-47-5), ^{13}C-^{15}N-L-phenylalanine (CAS 878339-23-2) and amino
132 acid kits #A6407 and A6282 were from Sigma-Aldrich (St. Louis, MO)

133

134 **2.2 Aqueous soil extraction and total organic carbon analysis**

135 Upper B-horizon soil was collected on September 24, 2013 from the background area of the
136 ORFRC in Oak Ridge, Tennessee. Specifically, the collection site was located approximately 12 feet
137 from groundwater well FW-300, in an area where grass, surface roots, and surface rocks were not
138 abundantly present. To reduce the presence of debris and non-soil constituents from the soil sample, rocks
139 and surface roots present on the surface, were physically removed prior to sampling. Surface soil was

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

140 collected by shoveling soil to a depth of 12-inches into sterile whirl pack bags, which were immediately
141 stored on blue ice. The soil at this location is an unconsolidated saprolite with organic rich clay soil that
142 ranges from 0.5-3 m approximating the depth of the root zone. Other soil properties including total
143 nitrogen, oxidizable organic matter, cation exchange capacity, pH, and particle size were determined by
144 the UC Davis Analytical Laboratory (Davis, CA).

145 Aqueous extraction was performed as previously described (Swenson et al., 2015b). Briefly, soil
146 was lyophilized to dryness and sieved to 2 mm prior to 24 h chloroform fumigation (Vance, 1987). Six
147 grams of soil were extracted in 24 mL of water at 4°C for 1 h while shaking on an orbital shaker (Orbital-
148 Genie, Scientific Industries, Bohemia, NY). Tubes were then centrifuged at 4°C for 5 min at 4000 rpm.
149 Supernatants were collected in fresh 50 mL conical tubes and centrifuged again after which supernatants
150 were filtered through 0.45 µm syringe filters (Pall Acrodisc Supor membrane). These final WEOC
151 samples were lyophilized to dryness in a Labconco Freezone 2.5 (Labconco, Kansas City, MO) and
152 resuspended in water at a final equivalent concentration of 2 g soil/mL. WEOC with and without further
153 1 kDa molecular weight cutoff (MWCO) filtration was analyzed for total organic carbon (TOC) by
154 acidification and measurement of the non-purgeable organic carbon. WEOC samples were prepared as
155 above, but at a final equivalent concentration of 1 g/mL (w/v of soil to water) and injected (50 µL) and
156 run via an NPOC method using a Shimadzu (Kyoto, Japan) TOC-L series CSH/H-type TIC/TOC
157 analyzer. Samples were individually mixed with a 1.5 % v/v of 1 M HCl for the conversion of inorganic
158 carbon to CO_2 followed by a 2.5 minute purging time and combustion of the remaining organic carbon at
159 an oven temperature of 680°C. The carrier gas used was synthetic air at a flow rate of 80 mL/min. Values
160 reported for TOC are in ppm (µg C per mL extract).

161

162 **2.3 LC/Q-TOF Scan MS WEOC Profiling**

163 One hundred microliters of the WEOC extract was lyophilized in triplicate to dryness and
164 resuspended in 100 µL of methanol containing internal standards 3,6-dihydroxy-4-methylpyridazine, 4-
165 (3,3-dimethyl-ureido)benzoic acid and 9-anthracene carboxylic acid at 25 µM . Samples were separated

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

166 by hydrophilic interaction liquid chromatography (HILIC) on an Agilent 1290 UHPLC (Agilent
167 Technologies, Santa Clara CA) using a Merck Sequant ZIC-pHILIC column (Merck KGaA, Darmstadt,
168 Germany) of dimensions 150 x 2.1 mm and 4.6 µm particle size. Two microliters of sample were injected
169 with the following mobile phase: A- 5mM ammonium acetate; B- 9:1 acetonitrile: 50 mM ammonium
170 acetate. The following gradient was used with a flow rate of 0.4 mL/min: 0.0-1.5min, 100% B; 25.0 min,
171 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
172 autosampler was maintained at 4°C. Flow was directed to the electrospray ionization source of an Agilent
173 6550 quadrupole time-of-flight (Q-TOF) mass spectrometer with the following settings for both positive
174 and negative polarity ionization modes; drying gas temperature and flow 275°C and 14 L/min; sheath gas
175 temperature and flow were 275°C and 9 L/min, respectively; nebulizer pressure was 30 psi; capillary
176 voltage was 3500 V and nozzle voltage 1000 V. Mass spectra were acquired from 50 to 1700 m/z at 4
177 spectra s^{-1}.

178

179 **2.4 LC/Q-TOF Tandem MS Analysis**

180 Procedural blanks were used to generate exclusion lists of compounds and added to a data-
181 dependent auto MS/MS method with a spectral scan range of 30-1200 amu and speed of 2 spectra sec^{-1}.
182 The quadrupole was set to a narrow isolation width of 1.3 amu and the collision cell fixed to 10, 20 and
183 40 eV. The auto MS/MS algorithm was restricted to a max of 4 precursor ions per MS/MS scan with a
184 required precursor ion count threshold of 7,500 and a target of 50,000. Initial MS/MS spectra were then
185 added within a retention time window of 0.25 min to the precursor exclusion list ensuring they were not
186 selected for MS/MS in a subsequent run, a process that captured MS/MS spectra of low intensity
187 precursor ions.

188

189 **2.5 LC/QTOF-MS Data Analysis**

190 MassHunter Qualitative Analysis v.B.07.00 SP1 was used to query acquired scan data against a
191 custom LC method specific accurate mass and retention time metabolite database based on acquired

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

192 authentic standard data. These initial metabolite identifications from scan MS data were augmented with
193 tandem MS by searching at specific collision energies against spectral libraries such as METLIN (Smith
194 et al., 2005), MassBank (Horai et al., 2010) and HMDB (Wishart et al., 2013). For metabolites only
195 identified by LC/MS, the highest confidence identifications (ranked as "1" in Supplementary Table 1)
196 required the unknown to be within 5 ppm m/z, 0.5 min retention time and to share dominant fragment
197 ions with the standard. Lower confidence identifications matched standards in accurate mass and retention
198 time but not fragmentation (or MS/MS was not obtained) and are ranked as a level "2" in Supplementary
199 Table 1. For some metabolites, authentic standards were not available, but were identified based on
200 matching fragmentation patterns with spectral libraries and are indicated by a level "3".
201

202 **2.6 GC/MS WEOC metabolite profiling**

203 Sample extracts were dried and derivatized by oximation-silylation and data acquired by GC/MS
204 as previously described (Swenson et al., 2015b). A C8-C30 fatty acid methyl ester ladder was added to
205 each sample to enable orthogonal compound identification within a retention index window of +/- 20 with
206 the electron ionization (EI) source of an Agilent 5977 GC/MS maintained at 70 eV for spectral library
207 matching. Metabolite profiling data were deconvoluted using Unknowns Analysis v.B.07.00 from Agilent
208 Technologies (Santa Clara, CA) followed by matching spectral fragmentation by quality score of 70 to the
209 Agilent Fiehn GC/MS Metabolomics RTL Library (Kind et al., 2009). A set of sugar standards comprised
210 of pentose, hexose, di-hexose and pentose alcohols (cellobiose, D-arabinose, D-arabinose, D-cellobiose,
211 D-galactose, D-mannose, D-xylose, fructose, glucose, L-arabinose, maltitol, maltose, mannitol, mannose,
212 raffinose, rutinose, sucrose, trehalose, xylitol and xylobiose) were used to resolve sugar pool composition
213 for initial qualitative metabolite profiling and then prepared by single replicate at 1, 5, 10, 25 and 50 µM
214 for determining sugar concentrations detected in the WEOC with d_{27}-myristic acid used as an internal
215 standard at a concentration of 25 µM (Supplementary Figure 1A). Quantitative data analysis was done
216 using Agilent MassHunter Quantitative Analysis v.B.07.00. GC/MS identifications are provided in
217 Supplementary Table 1.

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

218

219 **2.7 LC/QQQ-MS for metabolite quantitation**

220 For LC/MS-analyzed metabolites, quantitative assays were performed on an Agilent 6460 triple
221 quadrupole (QQQ) MS (Agilent Technologies, Santa Clara CA) using the same LC, column and ESI
222 source conditions as described for Q-TOF MS analysis. Authentic standards were prepared for amino
223 acids, nucleobases and nucleosides for a subset of metabolites identified in the WEOC to generate
224 compound specific transitions. Multiple reaction monitoring (MRM) transitions were scheduled by
225 retention time segment as established using Agilent Optimizer v.B.07.00 and are provided in
226 Supplementary Table 1. For the amino acid assay, a standard mix (Sigma-Aldrich amino acid standards
227 kit, product#A6407, A6282) was diluted to concentrations of 0.1, 0.5, 1 and 10 µg/mL and run as single
228 replicates (Supplementary Figure 1B). For the nucleotide/nucleobase assay, authentic standards were
229 prepared at 1, 5, 10 and 25 µM concentrations in triplicate with ^{13}C-phenylalanine used as an internal
230 standard at 25 µM (Supplementary Figure 1C). Soil extracts were then run at a concentration of 2 g/ml
231 (soil/water) for determining concentrations of these metabolites. Quantitative data analysis was performed
232 using Agilent MassHunter Quantitative Analysis version B.07.00. For each metabolite, quantitative data
233 (mg/mL for each metabolite) are converted to carbon ppm based on the percent weight carbon within each
234 metabolite and are reported as mg kg^{-1} extracted soil.

235

236 **2.8 Formulation of soil defined media and comparison of growth of 30 phylogenetically diverse**
237 **ORFRC isolates**

238 Soil defined medium 1 (SDM1) was prepared by adding 23 metabolites quantified from WEOC
239 (Section 2.3) at the observed absolute concentrations, to a base media composed of 1x Wolfe's mineral
240 and 1x Wolfe's vitamin solutions, potassium phosphate and ammonium chloride (Table 1). For soil
241 defined medium 2 (SDM2), the number of metabolites was expanded to 46 to include additional
242 observed, but not quantified WEOC compounds. These additional metabolites were added to the medium
243 at the lowest concentration of a quantified metabolite within the same class (Supplemental Table 1). SDM

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

244 (both at 1x and 10x concentrations) were compared to R2A medium at 1x concentration (Tecknova,
245 Hollister CA) in their ability to support the growth of a broad range of ORFRC isolates. An existing
246 collection of 30 phylogenetically diverse isolates from the ORFRC site were revived in liquid R2A from
247 frozen glycerol stocks. DNA was obtained from overnight cultures to verify 16S rDNA based
248 identifications. Aliquots were washed (3 times) prior to inoculation into 96-well microtitre plates. For
249 each fresh medium, 20 uL of starter culture was added to 180 uL and plates were incubated under aerobic
250 conditions for 110 h at 28°C, and shaken at 365 rpm. To facilitate growth monitoring, starting volumes of
251 each isolate were normalized (200 uL total). Growth data were collected by measuring OD_{600} on a BioTek
252 Eon Microplate Spectrophotometer at 15 min intervals.
253

254 **2.9 Time series exometabolomics analysis with *Pseudomonas sp.*** FW300-N2E2

255 Using SDM1-10x liquid medium, triplicate 2 ml cultures (including media and water blank
256 controls) of *Pseudomonas sp.* FW300-N2E2, a bacterial isolate from the ORFRC (16S rDNA sequencing
257 closest to *Pseudomonas corrugata* strain P94- EF153018.1; Thorgersen et al. (2015)) were prepared in
258 24-well spectrophotometer plates using the inoculation technique described in 2.8. Time points were
259 sampled early/mid log phase (3,6 hours) early/late stationary phase (9,12 hours) and finally at 24 h
260 (Supplementary Figure 2). Culture fractions were removed for LC/MS and GC/MS analysis at 0.5 and 0.2
261 mL volumes, respectively, and centrifuged in a Savant centrifuge (Thermo Scientific, San Jose CA) for 5
262 min at 5000 rpm and the resultant supernatant filtered through 0.2 µm centrifugal filters (Pall
263 Corporation, Port Washington NY) for 5 min at 10000 rpm. Filtrates were then frozen at -80 °C,
264 lyophilized to dryness with LC/MS samples resuspended in 100 µL methanol containing 25 µM of the
265 internal standards and analyzed as described in 2.3. GC/MS samples were derivatized and also analyzed
266 as described in 2.6.
267

268 **3 Results**

269 **3.1 Soil properties**

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

270 The TOC of the WEOC was 129.1 ppm (or 129.1 mg C kg^{-1} soil). Other physical properties of the
271 ORFRC soil were: total N of 0.08%, organic matter (oxidizable) content of 1.5%, cation exchange
272 capacity of 14.5 meq/100 g, soil pH of 7.7 and particle size (% sand/silt/clay) of 29/54/17.
273

274 **3.2 Integrating GC/MS with LC/MS to analyze soil WEOC composition and concentrations**

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

286 Quantitative methods were developed for 36 compounds representing the biochemical classes
287 with the most abundant MS ion signal (amino acids, sugars, nucleobases and nucleosides). LC triple
288 quadrupole MS was used for quantification of amino acids, nucleobases and nucleosides while sugars
289 were quantified by GC/MS. Of these 36 quantified metabolites, 25 were within the quantifiable (linear)
290 range (Supplementary Table 1). Converting the metabolite concentrations into carbon concentrations,
291 these 25 metabolites together accounted for a total of 20.0 mg C kg^{-1} soil with sugars accounting for
292 89.2% of the quantified metabolites. Trehalose, glucose, mannitol, fructose and maltose each accounted
293 for 5.3, 5.0, 3.2, 2.7,1.2 mg C kg^{-1} soil respectively (Supplementary Table 1). Of the remainder, 9.8%
294 were amino acids and only 0.8% were nucleobases and nucleosides.
295

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

296 **3.3 Molecular weight cutoff filtration and TOC quantitation of WEOC**

297 Since the metabolomic methods used here would not be expected to detect all WEOC
298 components, molecular weight cutoff filtration was used to help estimate the fraction of the WEOC pool
299 quantified by MS. The TOC of the initial soil WEOC was 129.1 mg C kg^{-1} soil, which, after filtration
300 through 1 kDa filters, was reduced to 53.3 mg C kg^{-1} soil. This reduction in TOC is consistent with the
301 removal of biopolymers, colloids, trace lignin and cellulosic debris. Based on these TOC data, the 25
302 metabolites that were quantified from the WEOC, accounting for 20.0 mg C kg^{-1} soil, represented
303 approximately 15.5% of the initial WEOC fraction and 37.5% of the <1 kDa filtered fraction.

304

305 **3.4 Comparison of soil defined media and growth of bacterial isolates**

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

314 A broad assessment was then performed to compare the viability of these two media, at two
315 different concentrations, with 30 native ORFRC bacterial isolates (Figure 2). In order to benchmark the
316 viability of these media, isolate growth was also compared to the isolation medium, R2A, which
317 supported the growth of all 30 isolates. At 1x concentrations, SDM1 was able to support the growth of 11
318 isolates and at 10x SDM1 supported the growth of 13 isolates. While only five isolates grew on 1x
319 SDM2, excitingly, 25 isolates grew on this medium at 10x concentration (Figure 2 and Supplementary
320 Table 3).

321

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

322 **3.5 Isolate exometabolomics analysis using SDM1**

323 Time-series exometabolomics studies can help delineate substrate preference of isolates as an
324 additional measure of resource partitioning. Since SDM1 appeared to be a simple, yet viable medium for
325 many isolates, this medium was selected to test its suitability for exometabolomics. *Pseudomonas sp.*
326 FW300-N2E2 was grown in SDM1-5x and samples were collected at 3, 6, 9 and 12 h for targeted
327 exometabolomics analysis by LC/QQQ-MS and GC/MS as described. Distinct patterns of substrate
328 depletion were observed as shown in Figure 3. Compounds such as arginine, proline, glutamic acid,
329 guanine and hypoxathine were already depleted by the first time point of 3 h. During later exponential
330 growth phase (6 and 9 h), a second pool of compounds were depleted including alanine, isoleucine,
331 gamma-guanadinobutyric acid, glycine, leucine, lysine, phenyalanine, serine and threonine. Interestingly,
332 presumably high value substrates were not depleted until stationary phase with glucose detected up to 12
333 h and mannose and fructose up to the final 24 h sampling. Several metabolites persisted at near initial
334 concentrations for the duration of the experiment including the abundant dihexoses trehalose and maltose,
335 the hexose arabinose, the sugar alcohol mannitol and the nucleoside uridine.

336

337 **4 Discussion**

338 **4.1 Integrating qualitative GC/MS with LC/QTOF-MS to analyze WEOC composition**

339 The objective of this study was to develop a viable and relevant defined media based on
340 metabolomics analyses of an environment of interest to enable exometabolomic characterization of isolate
341 resource partitioning. Here we combined two metabolomics approaches, GC/MS (Swenson et al., 2015b)
342 and HILIC LC/MS (Swenson et al., 2015a) for qualitative and quantitative analysis. In total, 96
343 metabolites were identified using a combination of authentic standards, MS/MS and spectral library
344 matching (Supplementary Table 1).

345 While this work focused on a single soil for the development of a metabolomics workflow for
346 defined media preparation, it is informative to see how our metabolites detected lie within the context of
347 other soil metabolomics studies. Overall, the range of metabolite classes detected, including amino acids,

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

348 amino acid derivatives, mono- and di-carboxylic acids, nucleobases, nucleosides, osmolytes, sugars, sugar
349 acids and sugar alcohols are consistent across many soil studies. The detection of trehalose and other
350 compatible solutes (betaine and proline betaine) as well as additional quaternary ammonium compounds
351 (acetylcarnitine, carnitine and choline) have been reported for other soils (Baran et al., 2013; Warren,
352 2014; Bouskill et al., 2016). Metabolites such as acetylcarnitine, citrulline, cytosine, gamma-aminobutyric
353 acid and nicotinic acid are consistent with a key finding from Warren (2013) who showed that a diverse
354 pool of nonpeptide organic N exists in soils.

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

365

366 **4.2. Analyzing the TOC content of soil WEOC**

367 The TOC level observed in our soil WEOC (129.1 mg C kg^{-1} soil) was found to be consistent
368 with many previous reports on soil DOC, noting that differences in soil types and extraction methods may
369 dramatically affect these values. With this caveat, our TOC level is close to the 146 ppm reported for a
370 soil (leachate) solution from a eutric cambisol grassland (Jones and Willett (2006). However, this study
371 displays the affect of extraction techniques on these values by reporting lower TOC values (55-70 mg kg^{-1}
372 soil) for the same soil following aqueous shaking (rather than leachate collection). Another study that is
373 consistent with our TOC results focused on evaluating dissolved organic matter dynamics in Greek

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

374 vineyard soils (Christou et al., 2005). They report drastic seasonal variation of DOC levels in soil
375 solutions ranging from approximately 100-400 ppm over a 12 month period (Christou et al., 2005) and
376 displayed substantial differences between topsoil and subsoil with WEOC levels decreasing from 89 to 58
377 mg C kg^{-1} soil (Christou et al., 2006).

378 Since soluble polymers and small particles presumably account for a large fraction of the WEOC,
379 we wanted to determine the fraction of this total carbon pool accounted for by the 25 quantified
380 metabolites. The TOC level of the soil WEOC was found to be reduced by more than half (129.1 ppm to
381 53.3 ppm) following 1 kDa filtration. Since microbes are limited in their ability to directly uptake
382 macromolecules and are dependent on extracellular deconstruction followed by transporting the resulting
383 metabolites, these molecules smaller than 1 kDa likely represent the most directly accessible fraction of
384 the WEOC for microbes. This fraction is most relevant to our metabolomics methods given that that one
385 operational definition of metabolites is molecules less than <1 kDa (Holmes et al., 2008).

386

387 **4.3. Characterizing the LMWOS fraction of WEOC by quantitative LC/QQQ-MS and GC/MS**
388 **analysis**

389 Based on the molar concentrations of the 25 quantified metabolites in the WEOC (consisting
390 mostly of carbohydrates and amino acids) and the number of carbons in each metabolite, we determined
391 that these quantified metabolites accounted for 20.0 mg C kg^{-1} in the soil sample. This represents 15.5%
392 of the WEOC and 37.5% of the < 1 kDa metabolite pool, indicating that even using two highly-sensitive
393 analytical approaches, the majority of soil metabolites remain rare, unidentified or undetected. Numerous
394 other studies support this assertion of the many unannotated chemical features present in soil LMWOS
395 (Ohno, et al., 2010; Warren 2013; Baran et al., 2015) highlighting the value of using defined metabolite
396 mixtures for exometabolomic characterization of microbial resource use.

397 The soil metabolites that we detected have a very uneven abundance distribution. Specifically, of
398 the quantified metabolites, sugars represented approximately 89.2% of the quantified metabolite TOC
399 pool, with a single metabolite, trehalose accounting for 5.3 mg C kg^{-1} soil (29.7% of the quantified sugar

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

400 TOC pool and 9.9% of the < 1 kDa WEOC). Of the amino acids, only alanine, valine, leucine and
401 isoleucine were at levels above 200 μg C kg^{-1} soil while all nucleosides and nucleobases except for
402 uridine were at trace levels at or below 50 μg C kg^{-1} soil. While this type of metabolite distribution is
403 specific to the ORFRC soil, we can compare these values with previous reports. Where we found total
404 quantified amino acids to be 4.08 mg kg^{-1} soil, Fischer et al. (2007) examined the composition of Haplic
405 Luvisol soil leachate and reports an amino acid content of 281.1 μg kg^{-1} soil, considerably lower than
406 ours, though with a similar ranked composition (*e.g.* alanine, leucine and isoleucine ranked among the
407 highest)(Fischer et al., 2007). We found the ratio of carbohydrates to amino acids to be 10.8 (g/g; more
408 carbohydrates), which is consistent with Hertenberger et al (2002) who reports ratios ranging from 6.4-
409 17.4, but Fisher et al. (2007) reports 0.4 (more amino acids). While these differences may simply be due
410 to actual differences between soils, bias introduced during extraction is likely another important factor.
411 For example extraction techniques range from mild soil leaching used by Fisher et al (Fischer et al.,
412 2007) to the fumigation-aqueous extraction (this study) to a more aggressive acetone-water extraction
413 used by Hertenberger et al. (2002).
414

415 **4.4. Preparation and evaluation of defined media based on soil WEOC composition**

416 SDM1, which contained 23 abundant WEOC metabolites, was found to support the growth of 13
417 out of the 30 isolates tested. Doubling the number of metabolites in the media (SDM2) dramatically
418 increased the number of isolates that grew (25 out of 30) demonstrating similar viability to the widely
419 used rich R2A media. Interestingly however, at reduced concentrations (1x), SDM1 supported growth of
420 more isolates compared to SDM2. It could be that at 1x concentrations, SDM1 has a more optimal carbon
421 concentration or C:N than SDM2, but at 10x concentrations, metabolite-specific transporters become
422 active, promoting the survival in the more rich SDM2 (Button, 1993). While, the medium R2A was still
423 more viable than the both SDM, this is not surprising given that the bacterial isolates analyzed here were

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

424 isolated using R2A. Furthermore, R2A is rich in amino acids and peptides while SDM1 and SDM2 have a
425 high sugar content, indicating potential unique and developed substrate preferences of these bacteria.

426 The application of SDM1 to investigate the substrate preferences of *Pseudomonas sp.* FW300-
427 N2E2 revealed an interesting pattern of substrate utilization. We observed rapid depletion of arginine,
428 glutamate, proline and guanine and hypoxanthine by the 3 h timepoint associated with the onset of
429 logarithmic growth followed by consumption of alanine, isoleucine, gamma-guanadinobutyric acid,
430 glycine, leucine, lysine, phenyalanine, serine and threonine (depleted by 6 h) then finally depletion of
431 sugars including glucose. Several sugars such as arabinose, maltose, mannitol and trehalose were not used
432 at all suggesting that this microbe may lack a sufficient set of transporters to use these abundant
433 resources, supporting the view that it is important to perform exometabolomics experiments on media
434 relevant, whenever possible, to their native environment. The utilization profile of this *Pseudomonas sp.*
435 is consistent with a previous report for *Pseudomonas aeruginosa* PAO1 in which growth on a complex
436 tryptone medium resolved an initial preference for select amino acids including leucine, proline, serine
437 and threonine while the carbohydrate utilization was very similar to our results, namely, while glucose
438 was metabolized, maltose, mannitol, mannose and trehalose were not (Frimmersdorf, Horatzek et al.
439 2010). Since, *Pseudomonas sp.* FW300-N2E2 was isolated using R2A, which is rich in many of the
440 preferentially consumed compounds (such as amino acids) perhaps it is not surprising that it preferentially
441 uses resources that, while relatively rare in the soil environment, are abundant in the isolation medium.
442 We anticipate that media prepared based on soil metabolite analyses may have additional utility for
443 isolating organisms that utilize the major measurable carbon sources within those environments.

444 The success of SDM demonstrates the ability to use soil metabolomics to develop defined media
445 based on metabolites known to be abundant in soils and at concentrations relevant to the native
446 environment. This also indicates the potential, in agreement with previous approaches, of using single or
447 complex amendment of minimal media to isolate previously unculturable soil bacteria (Sait et al., 2002;
448 Joseph et al., 2003). The approach used here is especially important for exometabolite profiling studies.
449 Specifically, microbial transport and regulatory systems, particularly catabolite repression, are responsive

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

450 to the absolute concentrations of metabolites in the environment. Thus, evaluating substrate utilization in
451 defined media relevant to their habitat can greatly improve exometabolomic studies. There are a number
452 of challenges associated with creating media that reflect microbial habitats for laboratory experimentation
453 must be considered. Specifically, bulk analysis of soil no doubt creates averaged metabolite compositions
454 unlike any particular niche within the soil. In addition, as we have shown here, many metabolites cannot
455 be identified or go undetected. Complementing mass spectrometry with other types of spectroscopy, such
456 as NMR, may help address this. Finally, due to the extreme heterogeneity between and within soil
457 ecosystems, our particular findings on soil metabolite abundances should not be generalized except for
458 the dozen or so metabolites that we have found to be consistent with other reports.
459
460 **4.5. Implications for other environments**
461 These methods, using exometabolomics analysis of environmental samples to prepare defined
462 media for the study of organisms from that environment, are likely applicable to a diversity of
463 environments. Extension of this approach to diverse soil types offers both the exciting possibility of
464 helping connect soil metabolite composition to soil microbial community composition and development
465 of generalizable defined soil media. While these defined media were focused on carbon, an important
466 extension would be to also account for other critical elements especially organic nitrogen and phosphorus.
467 The great advantage of these defined media is that they enable quantitative analysis of resource utilization
468 by soil microorganisms.
469 One exciting implication of this analysis is that rare metabolites that were not analyzed may
470 together account for a significant portion of the carbon (WEOC) in this sample, similar to how rare
471 microbes collectively account for a large portion of microbiomes. Synthesis of this view with previous
472 results showing that microbes from this soil (and biological soil crusts) utilize largely non-overlapping
473 metabolites suggests there is coupling between microbial diversity and soil metabolite diversity. These
474 results showing the unevenness of water soluble metabolites may further support the traditional view of
475 copiotrophic and oligotrophic organisms. Specifically, there are a low-diversity of organisms competing

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

476 for the abundant resources (copiotrophs) and a high diversity of rare organisms using low-abundance
477 substrates (Upton and Nedwell, 1989; Konopka et al., 1998). It should be emphasized that this is highly
478 speculative and studies of multiple soil samples and multiple sites would be required to support
479 generalization to both this particular site and to soils in general.
480

481 5. Conclusion

482 This study used soil metabolomic analyses to characterize the low molecular weight organic
483 matter composition to formulate defined media intended to approximate the qualitative and quantitative
484 composition of microbe bioavailable carbon from a specific study site. Composition and carbon
485 concentrations were found to align well with related studies with these soil metabolites having a very
486 uneven quantitative distribution (e.g. trehalose accounting for 9.9% of the <1 kDa WEOC fraction),
487 analogous to the uneven soil microbial community structure. The defined media that were synthesized to
488 reflect the soil WEOC composition were found to support the growth of 13 or 25 (for SDM1 and SDM2,
489 respectively) out of 30 phylogenetically diverse isolates. A detailed study of a single isolate,
490 *Pseudomonas sp.* FW300-N2E2 showed that this isolate rapidly depleted guanine, serine, leucine and
491 hypoxanthine while several metabolites including the most abundant disaccharides were not utilized from
492 SDM1. We anticipate that this approach of preparing environmentally relevant defined media will be
493 applicable to diverse environments to enable more ecologically relevant isolation and examination of
494 microbial substrate utilization.
495

496 Acknowledgements

497 The authors gratefully acknowledge funding from Ecosystems and Networks Integrated with
498 Genes and Molecular Assemblies (http://enigma.lbl.gov), a Scientific Focus Area Program at Lawrence
499 Berkeley National Laboratory for the U.S. Department of Energy, Office of Science, Office of Biological
500 and Environmental Research under contract number DE-AC02-05CH11231.

501 REFERENCES

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

502 Baran, R., Bowen, B.P., Price, M.N., Arkin, A.P., Deutschbauer, A.M., Northen, T.R., 2013. Metabolic
503 footprinting of mutant libraries to map metabolite utilization to genotype. ACS Chem Biol 8, 189-199.
504
505 Baran, R., Brodie, E.L., Mayberry-Lewis, J., Hummel, E., Da Rocha, U.N., Chakraborty, R., Bowen,
506 B.P., Karaoz, U., Cadillo-Quiroz, H., Garcia-Pichel, F., Northen, T.R., 2015. Exometabolite niche
507 partitioning among sympatric soil bacteria. Nat Commun 6, 8289.
508
509 Bouskill, N.J., Wood, T.E., Baran, R., Ye, Z., Bowen, B.P., Lim, H., Zhou, J., Nostrand, J.D., Nico, P.,
510 Northen, T.R., Silver, W.L., Brodie, E.L., 2016. Belowground Response to Drought in a Tropical Forest
511 Soil. I. Changes in Microbial Functional Potential and Metabolism. Front Microbiol 7, 525.
512
513 Boyer, J.N., Groffman, P.M., 1996. Bioavailability of water extractable organic carbon fractions in forest
514 and agricultural soil profiles. Soil Biology and Biochemistry 28, 783-790.
515
516 Butler, J.L., Williams, M.A., Bottomley, P.J., Myrold, D.D., 2003. Microbial Community Dynamics
517 Associated with Rhizosphere Carbon Flow. Appl Environ Microbiol 69, 6793-6800.
518
519 Button, D.K., 1993. Nutrient-limited microbial growth kinetics: overview and recent advances. Antonie
520 Van Leeuwenhoek 63, 225-235.
521
522 Christou, M., Avramides, E.J., Jones, D.L., 2006. Dissolved organic nitrogen dynamics in a
523 Mediterranean vineyard soil. Soil Biology and Biochemistry 38, 2265-2277.
524
525 Christou, M., Avramides, E.J., Roberts, J.P., Jones, D.L., 2005. Dissolved organic nitrogen in contrasting
526 agricultural ecosystems. Soil Biology and Biochemistry 37, 1560-1563.
527

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

528 Fernandes, M.F., Saxena, J., Dick, R.P., 2013. Comparison of whole-cell fatty acid (MIDI) or
529 phospholipid fatty acid (PLFA) extractants as biomarkers to profile soil microbial communities. Microb
530 Ecol 66, 145-157.
531
532 Ferrari, B.C., Winsley, T., Gillings, M., Binnerup, S., 2008. Cultivating previously uncultured soil
533 bacteria using a soil substrate membrane system. Nat Protoc 3, 1261-1269.
534
535 Fischer, H., Meyer, A., Fischer, K., Kuzyakov, Y., 2007. Carbohydrate and amino acid composition of
536 dissolved organic matter leached from soil. Soil Biology and Biochemistry 39, 2926-2935.
537
538 Guigue, J., Mathieu, O., Lévêque, J., Mounier, S., Laffont, R., Maron, P.A., Navarro, N., Chateau, C.,
539 Amiotte-Suchet, P., Lucas, Y., 2014. A comparison of extraction procedures for water-extractable organic
540 matter in soils. European Journal of Soil Science 65, 520-530.
541
542 Haney, R.L., Franzluebbers, A.J., Jin, V.L., Johnson, M.-V., Haney, E.B., White, M.J., Harmel, R.D.,
543 2012. Soil Organic C:N vs. Water-Extractable Organic C:N. Open Journal of Soil Science 02, 269-274.
544
545 Hertenberger, G., Zampach, P., Bachmann, G., 2002. Plant species affect the concentration of free sugars
546 and free amino acids in different types of soil. Journal of Plant Nutrition and Soil Science 165, 557-565.
547
548 Holmes, E., Wilson, I.D., Nicholson, J.K., 2008. Metabolic phenotyping in health and disease. Cell 134,
549 714-717.
550
551 Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K., Tanaka, S.,
552 Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda, F., Sawada, Y., Hirai, M.Y.,
553 Nakanishi, H., Ikeda, K., Akimoto, N., Maoka, T., Takahashi, H., Ara, T., Sakurai, N., Suzuki, H.,

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

554 Shibata, D., Neumann, S., Iida, T., Tanaka, K., Funatsu, K., Matsuura, F., Soga, T., Taguchi, R., Saito,
555 K., Nishioka, T., 2010. MassBank: a public repository for sharing mass spectral data for life sciences. J
556 Mass Spectrom 45, 703-714.
557
558 Jones, D., Willett, V., 2006. Experimental evaluation of methods to quantify dissolved organic nitrogen
559 (DON) and dissolved organic carbon (DOC) in soil. Soil Biology and Biochemistry 38, 991-999.
560
561 Joseph, S.J., Hugenholtz, P., Sangwan, P., Osborne, C.A., Janssen, P.H., 2003. Laboratory Cultivation of
562 Widespread and Previously Uncultured Soil Bacteria. Appl Environ Microbiol 69, 7210-7215.
563
564 Kind, T., Wohlgemuth, G., Lee do, Y., Lu, Y., Palazoglu, M., Shahbaz, S., Fiehn, O., 2009. FiehnLib:
565 mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas
566 chromatography/mass spectrometry. Anal Chem 81, 10038-10048.
567
568 Konopka, A., Oliver, L., Jr, R.F.T., 1998. The Use of Carbon Substrate Utilization Patterns in
569 Environmental and Ecological Microbiology. Microb Ecol 35, 103-115.
570
571 Liebeke, M., Lalk, M., 2014. Staphylococcus aureus metabolic response to changing environmental
572 conditions - a metabolomics perspective. Int J Med Microbiol 304, 222-229.
573
574 Ling, L.L., Schneider, T., Peoples, A.J., Spoering, A.L., Engels, I., Conlon, B.P., Mueller, A., Schaberle,
575 T.F., Hughes, D.E., Epstein, S., Jones, M., Lazarides, L., Steadman, V.A., Cohen, D.R., Felix, C.R.,
576 Fetterman, K.A., Millett, W.P., Nitti, A.G., Zullo, A.M., Chen, C., Lewis, K., 2015. A new antibiotic kills
577 pathogens without detectable resistance. Nature 517, 455-459.
578

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

579 Miller, A., Schimel, J., Meixner, T., Sickman, J., Melack, J., 2005. Episodic rewetting enhances carbon
580 and nitrogen release from chaparral soils. Soil Biology and Biochemistry 37, 2195-2204.
581
582 Ohno, T., He, Z., Sleighter, R. L., Honeycutt, C. W., Hatcher, P. G., 2010. Ultrahigh resolution mass
583 spectrometry and indicator species analysis to identify marker components of soil- and plant biomass-
584 derived organic matter fractions. Environ Sci Technol 44, 8594-8600.
585
586 Ohno, T., Parr, T.B., Gruselle, M.C., Fernandez, I.J., Sleighter, R.L., Hatcher, P.G., 2014. Molecular
587 composition and biodegradability of soil organic matter: a case study comparing two new England forest
588 types. Environ Sci Technol 48, 7229-7236.
589
590 Pham, V.H., Kim, J., 2012. Cultivation of unculturable soil bacteria. Trends Biotechnol 30, 475-484.
591
592 Reasoner, D.J., Geldreich, E.E., 1985. A new medium for the enumeration and subculture of bacteria
593 from potable water. Appl Environ Microbiol 49, 1-7.
594
595 Sait, M., Hugenholtz, P., Janssen, P.H., 2002. Cultivation of globally distributed soil bacteria from
596 phylogenetic lineages previously only detected in cultivation-independent surveys. Environ Microbiol 4,
597 654-666.
598
599 Schimel, J.P., Mikan, C., 2005. Changing microbial substrate use in Arctic tundra soils through a freeze-
600 thaw cycle. Soil Biology and Biochemistry 37, 1411-1418.
601
602 Schmidt, M.W., Torn, M.S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I.A., Kleber, M.,
603 Kogel-Knabner, I., Lehmann, J., Manning, D.A., Nannipieri, P., Rasse, D.P., Weiner, S., Trumbore, S.E.,
604 2011. Persistence of soil organic matter as an ecosystem property. Nature 478, 49-56.

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

605

606 Smith, C.A., O'Maille, G., Want, E.J., Qin, C., Trauger, S.A., Brandon, T.R., Custodio, D.E., Abagyan,
607 R., Siuzdak, G., 2005. METLIN: a metabolite mass spectral database. Ther Drug Monit 27, 747-751.
608

609 Strobel, B.W., 2001. Influence of vegetation on low-molecular-weight carboxylic acids in soil solution—
610 a review. Geoderma 99, 169-198.
611

612 Svenning, M.M., Wartiainen, I., Hestnes, A.G., Binnerup, S.J., 2003. Isolation of methane oxidising
613 bacteria from soil by use of a soil substrate membrane system. FEMS Microbiology Ecology 44, 347-354.
614

615 Swenson, T.L., Bowen, B.P., Nico, P.S., Northen, T.R., 2015a. Competitive sorption of microbial
616 metabolites on an iron oxide mineral. Soil Biology and Biochemistry 90, 34-41.
617

618 Swenson, T.L., Jenkins, S., Bowen, B.P., Northen, T.R., 2015b. Untargeted soil metabolomics methods
619 for analysis of extractable organic matter. Soil Biology and Biochemistry 80, 189-198.
620

621 Thorgersen, M.P., Lancaster, W.A., Vaccaro, B.J., Poole, F.L., Rocha, A.M., Mehlhorn, T., Pettenato, A.,
622 Ray, J., Waters, R.J., Melnyk, R.A., Chakraborty, R., Hazen, T.C., Deutschbauer, A.M., Arkin, A.P.,
623 Adams, M.W., 2015. Molybdenum Availability Is Key to Nitrate Removal in Contaminated Groundwater
624 Environments. Appl Environ Microbiol 81, 4976-4983.
625

626 Tucker, C.L., Young, J.M., Williams, D.G., Ogle, K., 2014. Process-based isotope partitioning of winter
627 soil respiration in a subalpine ecosystem reveals importance of rhizospheric respiration. Biogeochemistry
628 121, 389-408.
629

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

630 Upton, A.C., Nedwell, D.B., 1989. Nutritional flexibility of oligotrophic and copiotrophic Antarctic
631 bacteria with respect to organic substrates. FEMS Microbiol Lett 62, 1-6.
632
633 Vance, E.D., Brookes, P.C., Jenkinson, D.S., 1987. An extraction method for measuring soil microbial
634 biomass C. Soil Biology and Biochemistry 19, 703-707.
635
636 Warren, C.R., 2013. High diversity of small organic N observed in soil water. Soil Biology and
637 Biochemistry 57, 444-450.
638
639 Warren, C.R., 2014. Response of osmolytes in soil to drying and rewetting. Soil Biology and
640 Biochemistry 70, 22-32.
641
642 Wishart, D.S., Jewison, T., Guo, A.C., Wilson, M., Knox, C., Liu, Y., Djoumbou, Y., Mandal, R., Aziat,
643 F., Dong, E., Bouatra, S., Sinelnikov, I., Arndt, D., Xia, J., Liu, P., Yallou, F., Bjorndahl, T., Perez-
644 Pineiro, R., Eisner, R., Allen, F., Neveu, V., Greiner, R., Scalbert, A., 2013. HMDB 3.0--The Human
645 Metabolome Database in 2013. Nucleic Acids Res 41, D801-807.

646

647

648

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

649 **TABLES.**

Metabolite amendment	Formula	Molecular weight (g/mol)	Metabolite Class	WEOC mg/L	SDM1 mg/L	SDM1 C ppm
Trehalose	$C_{12}H_{22}O_{11}$	342.3	dihexose	12.53	20	8.4
Fructose	$C_6H_{12}O_6$	180.2	hexose	14.53	15	5.81
Glucose	$C_6H_{12}O_6$	182.2	hexose	12.55	15	5.93
Mannitol	$C_6H_{14}O_6$	182.2	sugar alcohol	8.17	15	5.93
Maltose	$C_{12}H_{22}O_{11}$	342.3	dihexose	2.95	5	2.11
Alanine	$C_3H_7NO_2$	89.1	amino acid	0.93	2	0.81
Arabinose	$C_5H_{10}O_5$	150.1	pentose	0.52	2	0.8
Leucine	$C_6H_{13}NO_2$	131.2	amino acid	0.7	1	0.55
Mannose	$C_6H_{12}O_6$	180.2	hexose	0.54	1	0.4
Isoleucine	$C_6H_{13}NO_2$	131.2	amino acid	0.46	1	0.55
Arginine	$C_6H_{14}N_4O_2$	174.2	amino acid	0.22	0.5	0.21
Proline	$C_5H_9NO_2$	115.1	amino acid	0.2	0.5	0.26
Threonine	$C_4H_9NO_3$	119.1	amino acid	0.2	0.5	0.2
Lysine	$C_6H_{14}N_2O_2$	146.2	amino acid	0.2	0.5	0.25
Phenylalanine	$C_9H_{11}NO_2$	165.2	amino acid	0.19	0.5	0.33
Glycine	$C_2H_5NO_2$	75.1	amino acid	0.17	0.5	0.16
Uridine	$C_9H_{12}N_2O_6$	244.2	nucleoside	0.16	0.5	0.22
Glutamate	$C_5H_9NO_4$	147.1	amino acid	0.11	0.1	0.04
Serine	$C_3H_7NO_3$	105.1	amino acid	0.09	0.1	0.03
Adenosine	$C_{10}H_{13}N_5O_4$	267.2	nucleoside	0.08	0.1	0.04
Gamma-guanidino-butyric acid	$C_5H_{11}N_3O_2$	145.2	amino acid derivative	0.05	0.1	0.04
Hypoxanthine	$C_5H_4N_4O$	136.1	purine	0.04	0.1	0.04
Guanine	$C_5H_5N_5O$	151.1	nucleobase	0.004	0.1	0.04
TOTAL				**55.59**	**81.1**	**33.15**

650
651

652 **Table 1. Metabolite formulation for soil defined medium 1 (SDM1).** For each metabolite,
653 concentrations are shown for soil WEOC (in mg/L), the amount added to SDM1 (mg/L) and the
654 equivalent carbon ppm.

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURES

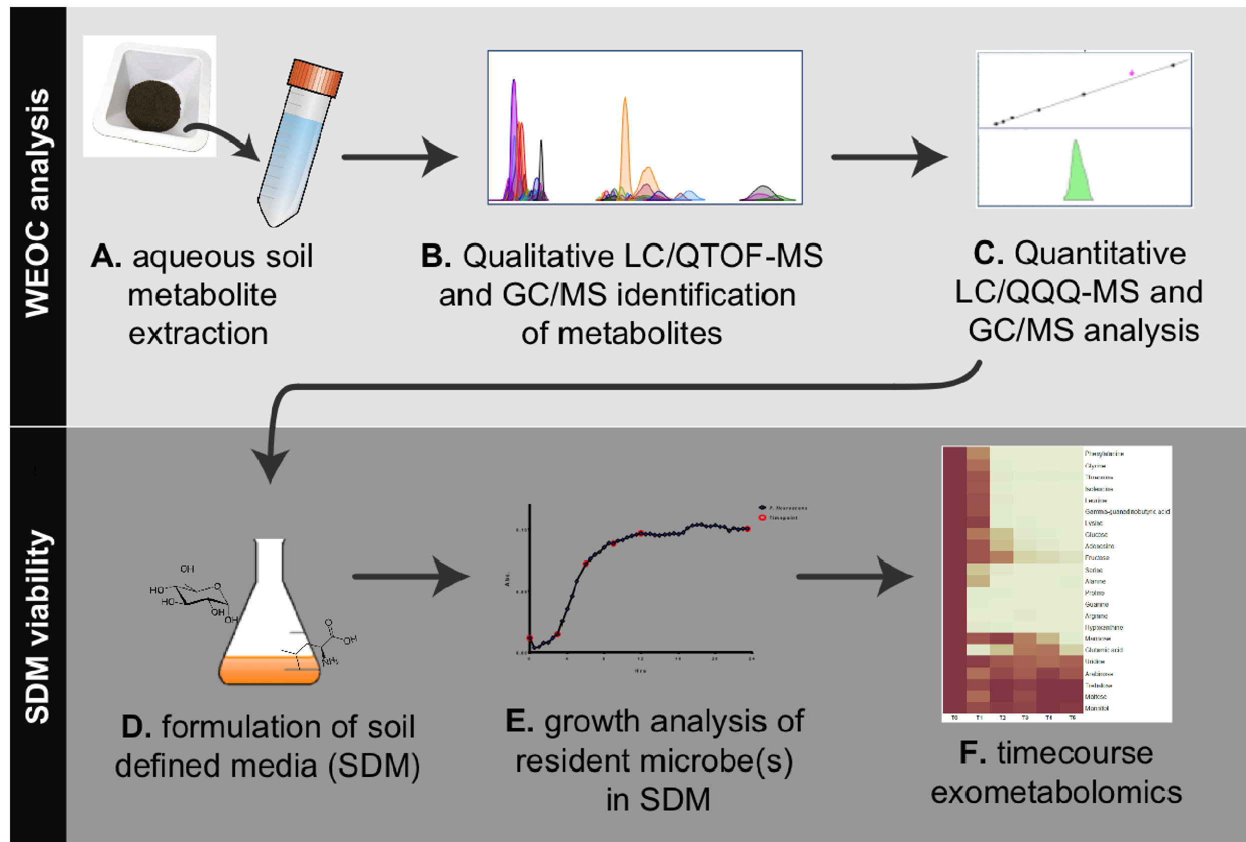

Figure 1. Workflow overview for the analysis of water extractable organic carbon (WEOC) and soil defined media (SDM). WEOC analysis is done by **(A)** extracting soil (sieved and fumigated) with water for 1 h, **(B)** acquiring scan and MS/MS data by LC/QTOF-MS and EI fragmentation data by GC/MS followed 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 monitor substrate utilization by bacteria.

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

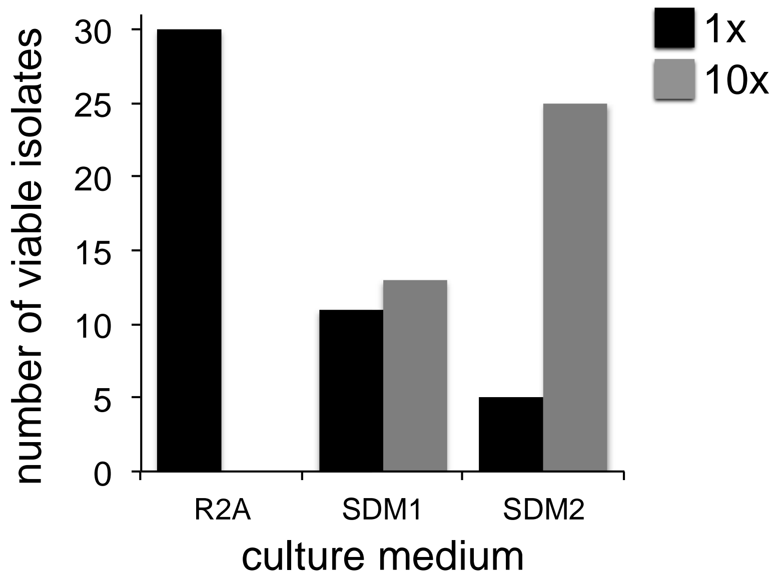

Figure 2. Isolate growth screen with R2A, SDM1 and SDM2. Each medium was tested (at 1x concentration for R2A and SDM and 10x for SDM) in its ability to support the growth of 30 phylogenetically diverse isolates from the ORFRC.

bioRxiv preprint doi: https://doi.org/10.1101/151282; this version posted June 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

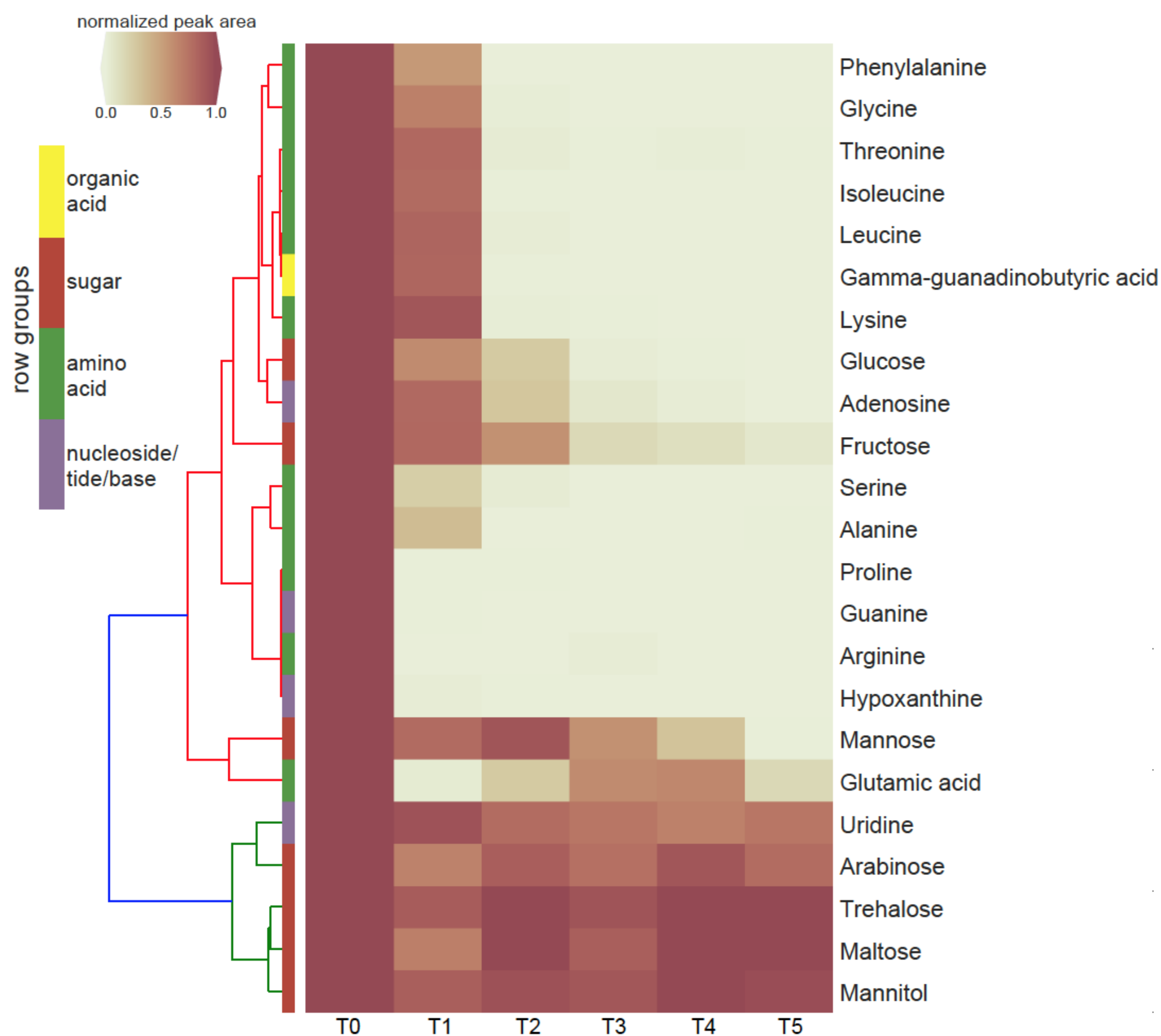

Figure 3. Clustering heatmap of normalized peak areas for SDM1 metabolites across timecourse sampling of *Pseudomonas sp.* FW300-N2E2 spent media. Levels are displayed in terms of relative ratio to initial concentration at time zero (T0) with T0-5 representing 0, 3, 6, 9, 12 and 24 h time points, respectively.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