Marine metabolomics: a method for the non-targeted measurement of metabolites in seawater by gas-chromatography mass spectrometry.

Emilia M Sogin[1*], Erik Puskas[1], Nicole Dubilier[1], Manuel Liebeke[1*]

1-Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, Bremen, Germany

[*]Corresponding authors

Maggie Sogin, email: esogin@mpi-bremen.de, tel: +49 421 2028-823

Manuel Liebeke email: mliebeke@mpi-bremen.de, tel: +49 421 2028-822

Abstract

Microbial communities exchange molecules with their environment that play a major role in global biogeochemical cycles and climate. While extracellular metabolites are commonly measured in terrestrial and limnic ecosystems, the presence of salt in marine habitats has hampered non-targeted analyses of the ocean exo-metabolome. To overcome these limitations, we developed SeaMet, a gas chromatography-mass spectrometry (GC-MS) method that detects at minimum 107 metabolites down to nano-molar concentrations in less than one milliliter of seawater, and improves signal detection by 324 fold compared to standard methods for marine samples. To showcase the strengths of SeaMet, we used it to explore marine metabolomes *in vitro* and *in vivo*. For the former, we measured the production and consumption of metabolites during culture of a heterotrophic bacterium that is widespread in the North Sea. Our approach revealed successional uptake of amino acids, while sugars were not consumed, and highlight the power of exocellular metabolomics in providing insights into nutrient uptake and energy

conservation in marine microorganisms. For *in vivo* analyses, we applied SeaMet to explore the *in situ* metabolome of coral reef and mangrove sediment porewaters. Despite the fact that these ecosystems occur in nutrient-poor waters, we uncovered high concentrations of many different sugars and fatty acids, compounds predicted to play a key role for the abundant and diverse microbial communities in coral-reef and mangrove sediments. Our data demonstrate that SeaMet advances marine metabolomics by enabling a non-targeted and quantitative analysis of marine metabolites, thus providing new insights into nutrient cycles in the oceans.

Importance

The non-targeted, hypothesis-free approach using metabolomics to analyzing metabolites that occur in the oceans is less developed than for terrestrial and limnic ecosystems. The central challenge in marine metabolomics is that salt prevents the comprehensive analysis of metabolites in seawater. Building on previous sample preparation methods for metabolomics, we developed SeaMet, which overcomes the limitations of salt on metabolite detection. Considering the oceans contain the largest organic carbon pool on Earth, describing the marine metabolome using non-targeted approaches is critical for understanding the drivers behind element cycles, biotic interactions, ecosystem function, and atmospheric CO_2 storage. Our method complements both targeted marine metabolomic investigations as well as other 'omics' (e.g., genomics, transcriptomics and proteomics) level approaches by providing an avenue for studying the chemical interaction between marine microbes and their habitats.

42 ## Introduction

43 Marine microorganisms produce and stabilize the largest pool of organic carbon on Earth
44 by exchanging molecules with their environment (1, 2). Marine microbes are also the basis for
45 maintaining the long term storage of carbon dioxide (CO_2) in the oceans, which plays a complex
46 role in biogeochemical cycles with uncertain implications for global climate (3). While
47 metagenomic and metatranscriptomic studies of the ocean, driven by low sequencing costs and
48 projects like Tara Oceans (4), have deepened our knowledge of the identity and activity of
49 marine microbes, these studies are limited in their ability to determine the molecules that
50 contribute to the chemical complexity of marine habitats. New approaches are needed to permit
51 equivalent surveys of the extracellular metabolome, or exometabolome of the ocean.

52 Exometabolomics provides an opportunity to directly characterize the molecular interaction
53 between microbes and their environment by profiling the types of molecules cellular organisms
54 secrete (5). In terrestrial and limnic systems, these studies have advanced our understanding of
55 microbial communities in soil organic matter cycling (6, 7), overflow metabolism of cultivable
56 microorganisms (8, 9) and chemical ecology of the environment (10, 11). While intracellular
57 metabolomic analyses of tissues from marine microbial cells to invertebrates is becoming
58 increasingly more common (12-14), the defining characteristic of marine habitats - high salt
59 concentration - limits exometabolomic analyses of the oceans to studies that require salt removal
60 prior to metabolite extraction (10, 15, 16).

61 Our knowledge of the metabolite composition of ocean habitats is restricted to methods that
62 require sample preparation techniques that alter their molecular composition, or targeted
63 approaches that measure a defined group of metabolites (17-19). The most common
64 environmental profiling strategies in marine ecosystems rely on solid phase extraction (SPE)

65 techniques to remove salt prior to mass-spectrometry (MS) analyses (20, 21). These studies have
66 demonstrated the role of microbial communites in producing recalcitrant dissolved organic
67 matter (DOM) and provided insights into their role in long term carbon storage (22). However,
68 the removal of salt from marine samples using SPE is accompanied by the co-removal of small
69 polar compounds, which are the primary components of the liable organic matter pool (17).
70 Consequently, SPE-based studies can only detect about 50% of the compounds that make up the
71 DOM pool from the ocean, and fail to detect the majority of compounds involved in the central
72 metabolism of cells. Furthermore, current DOM analytical approaches remain largely
73 inaccessible for the majority of research institutions and projects. This is largely due to high
74 instrumentation costs for high-resolution MS (coupled to liquid-chromatography or with direct-
75 infusion), large sample volume requirements, and the relatively low-throughput in data
76 acquisition.

77 Gas chromatography (GC)- MS analysis, on the other hand, is a high-throughput and
78 widely available analytical method that allows for the detection of primary metabolites, small
79 molecules that occur in central metabolic pathways across biological systems (23, 24). High
80 reproducibility coupled to the widespread availability of annotation resources make GC-MS the
81 "workhorse" of analytical chemistry facilities. GC-MS has allowed the identification of
82 metabolites associated with human disease (25), detection of compounds that serve as
83 environmental cues in foraging (26), description of metabolic fluxes within and between cells
84 (27), and is used for environmental profiling of soils and microbial activity on land (6, 28).
85 Despite its the power of detecting metabolites involved in central metabolism, exometabolomic
86 studies using GC-MS from marine habitats are absent due to the inhibitory effects of salt on
87 sample analysis.

88 The ocean metabolome remains largely undefined, despite a growing field of research
89 exploring the molecular composition of DOM (1, 2, 20, 21). To more efficiently decipher ocean
90 metabolism, cost-effective, high-throughput, and untargeted workflows that can readily identify
91 and quantify molecules from high salinity environments are critical. Here, we present SeaMet, a
92 marine metabolomics method that builds on previous GC-MS sample derivatization methods to
93 enable metabolite detection in seawater. Using SeaMet, we demonstrate how our method can
94 enhance our understanding of microbial metabolism in culture experiments and profiling of
95 marine habitats.
96

97 ## Results and Discussion

98 SeaMet modifies the well-established two-step derivatization procedure, which permits
99 the detection of non-volatile primary metabolites using GC-MS, and involves methoximation
100 followed by trimethylsilylation (29). Like other GC-MS sample preparation techniques (30, 31),
101 SeaMet removes liquid through vacuum drying prior to derivatization - a process that results in a
102 salt pellet when working with marine samples, which restricts MS analysis. Our preliminary tests
103 suggested that water locked within the dried salt crystals hindered the chemical reactions needed
104 for GC-MS (**Fig. S1**). Our method overcomes this limitation by first eliminating residual water
105 within the salt crystals and then extracting metabolites into the derivatization reagents (**Fig. 1A**).

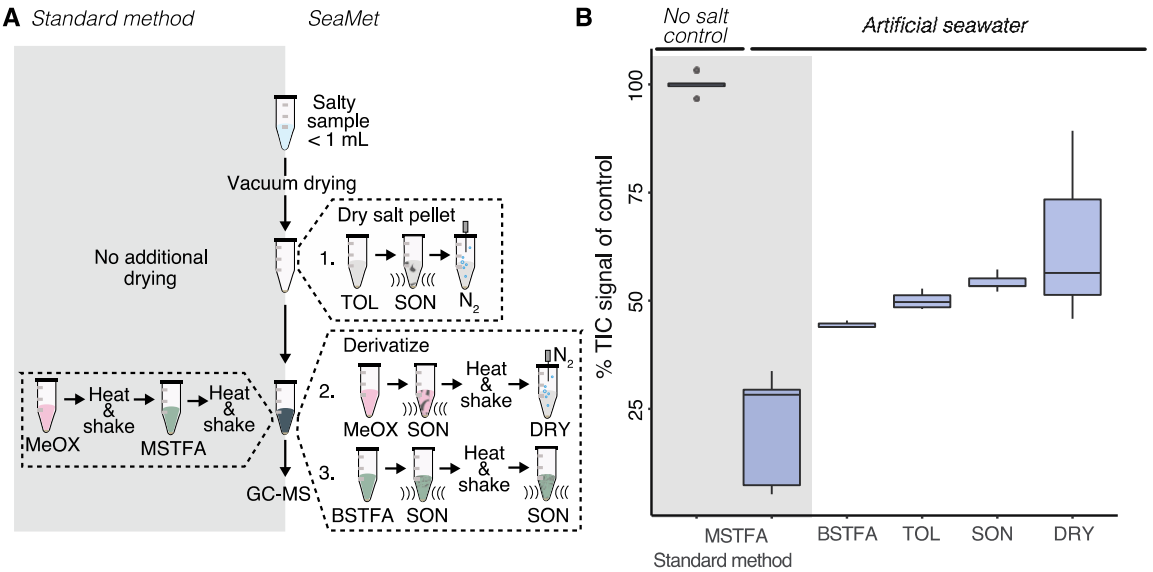

Figure 1. How SeaMet works. A, Modifications to the standard two-step methoximation (MeOX)-trimethylsilylation (TMS) derivatization protocol include key steps that enhance metabolite signal detection in seawater as shown in **B.** Steps modified from the standard method include a switch in derivatization reagents from MSTFA to BSTFA, further drying of the salt pellet using toluene (TOL) to remove water azeotropically, ultrasonication (SON) after the addition of TOL, MeOX, BSTFA, and after BSTFA derivatization, and drying (DRY) of the pyridine after the MeOX derivatization prior to BSTFA addition. **B,** Box plots showing changes in total ion chromatogram (TIC) signals after GC-MS data acquisition. Results are from a synthetic mixture of 45 metabolites representing a broad scope of metabolite classes (**Table S1**) dissolved in 0.5 mL of seawater (n = 5) relative to average of the no salt control.

We used a mixture of 45 different metabolites (**Table S1**) dissolved in artificial seawater to 0.4 mM to document the performance in metabolite detection of our method. Overall, SeaMet increased total signal intensity on average by 42% and up to 89% for high salinity samples in comparison to the standard GC-MS sample preparation (**Fig. 1B**). We first replaced the most commonly used trimethylsilylation reagent, N-methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA)(31), with one that is less susceptible to inhibition by water, N, O-Bistrifluoroacetamide (BSTFA), which resulted in higher metabolite signals (**Fig. S1B**). To

125 eliminate water from the samples, we increased the speed-vacuum drying time from four to eight
126 hours, and integrated a toluene drying step that is used in urine-based metabolomic analyses (30).
127 We further enhanced metabolite signals by treating the salt pellet to a combination of
128 ultrasonication and vortexing after the addition of toluene and both derivatization reagents, and
129 following completion of the trimethylsilylation reaction. These steps break apart the salt crystals
130 and release water into the toluene to enhance salt drying and metabolite extraction. Finally,
131 following a recently described method for improving GC-MS metabolite detection regardless of
132 sample type (32), we included an additional step between the methoximation and
133 trimethylsilylation derivatization reactions and evaporated the first derivatization reagent under
134 N_2 gas (see **Fig. 1B** for total signal improvements of each step).

135 Overall, SeaMet allowed us to detect significant increases in metabolite abundances
136 across molecular classes when compared to the standard method (adjusted p-value < 0.05; mean
137 fold change across all ions = 323; **Fig. 2A; Fig. S2**). This included measurement of organic
138 acids, amino acids, and fatty acids, as well as sugars (and their stereoisomers), sugar alcohols,
139 and sterols (**Table S1**).

140 To determine the quantitative capabilities of SeaMet, we used a metabolite mixture (45
141 metabolites spanning 9 compound classes) and added different concentrations (from 0.0039 mM-
142 to 0.4 mM) to seawater. Our detection limits were in the nano-molar range and comparable to
143 those of targeted techniques for marine ecosystems that were developed to quantify single
144 compounds from specific molecular classes (**Tables S2; Table S3**). In contrast to previously
145 published techniques, which require at least an order of magnitude higher sample volumes,
146 SeaMet only requires 0.5 mL to 1 mL of seawater for metabolite detection (17, 33). Using
147 SeaMet, we measured 107 metabolite standards in seawater, representing major metabolite

148 groups involved in primary metabolic pathways (**Table S4**). Our method provides reproducible
149 quantification across metabolite classes ($r^2 > 0.7$), and gives similar linearity and dynamic range
150 in seawater samples compared to salt-free samples prepared with the standard GC-MS
151 derivatization method (**Fig. 2A-H; Fig. S3)**. Moreover, we demonstrate that SeaMet reduces
152 variation in ion detection for individual metabolites (Welches t-test *p*-value < 0.01 across all ions
153 at 4 nmols; average % $CV_{salt} = 20.2 \pm 0.78$, average % $CV_{salt\text{-}free} = 23.5 \pm 0.72$) compared to
154 salt-free samples prepared with the standard GC-MS derivatization procedure (**Table S2**). The
155 analytical characteristics of the 107 metabolites (**Table S4**) can be used for more sensitive,
156 targeted GC-MS analyses or help in identifying metabolites in untargeted applications.
157

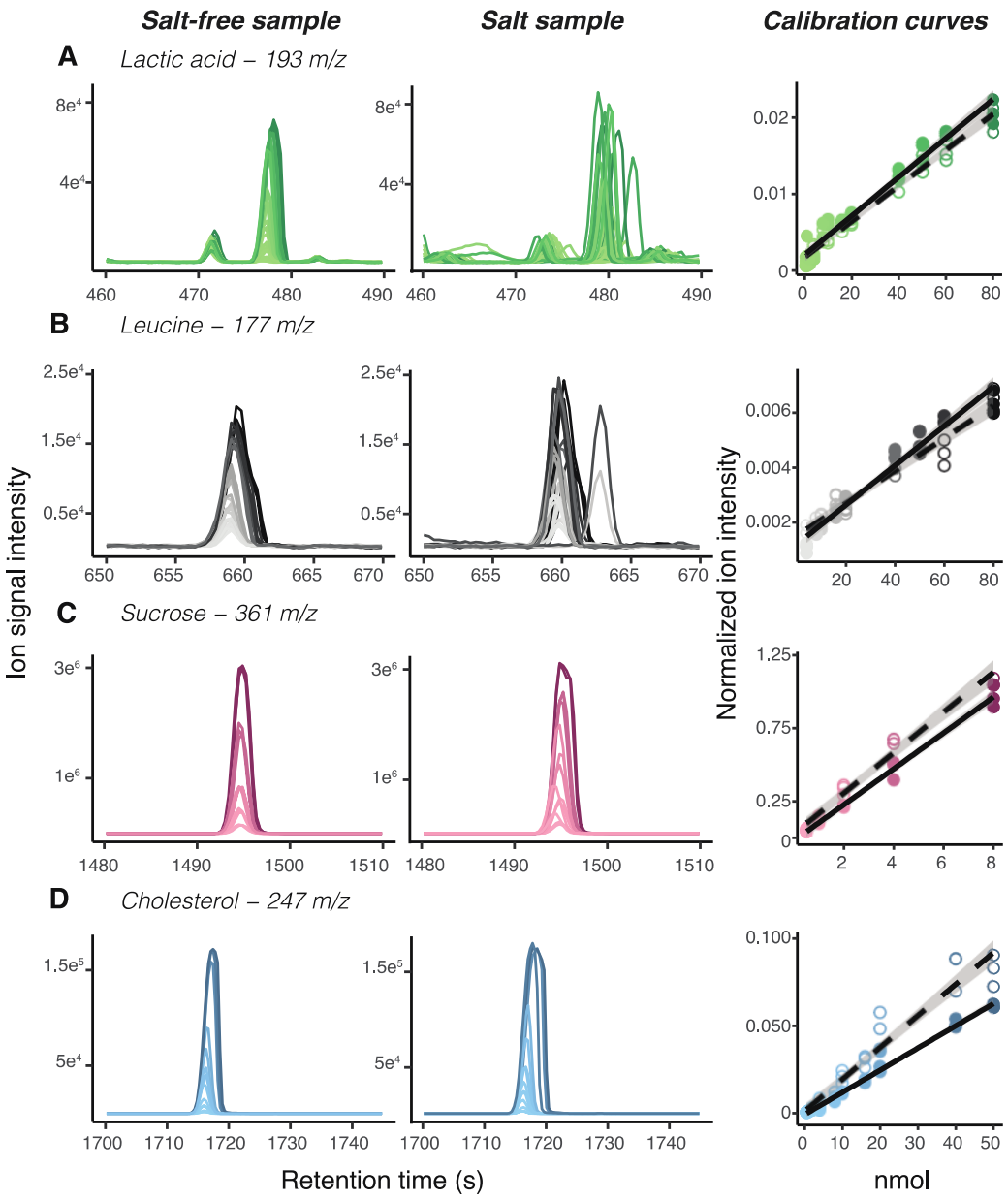

Figure 2. Metabolite detection and quantification in seawater using SeaMet in comparison to salt-free water. A-D, Extracted ion chromatograms for select metabolites in salt-free and artificial seawater demonstrate reproducible metabolite detection across concentration gradients, as shown in associated calibration curves on the right. Spectra and points are shaded in scale with sample concentrations, where more concentrated samples are represented by darker colors. Open circles = salt samples; filled circles = salt-free samples.

167 Given that SeaMet avoids SPE, we assessed how SPE sample treatment affects the ability
168 to detect compounds in marine samples. We compared GC-MS profiles measured with SeaMet
169 before and after salt removal using the most commonly used Bond Elut styrene-divinylbenzene
170 (PPL) SPE-columns. Our analyses revealed that small polar compounds, such as sugars, sugar
171 alcohols, amino acids, and organic acids, were co-removed with salt during SPE sample
172 preparation (**Fig. 3B**). These results provide evidence that SeaMet captures compounds
173 commonly missed by SPE-based exometabolomic approaches for marine samples. SeaMet thus
174 expands the range of metabolites that can be measured by untargeted approaches beyond those
175 currently used to characterize marine DOM, and contributes to advancing marine metabolomics.

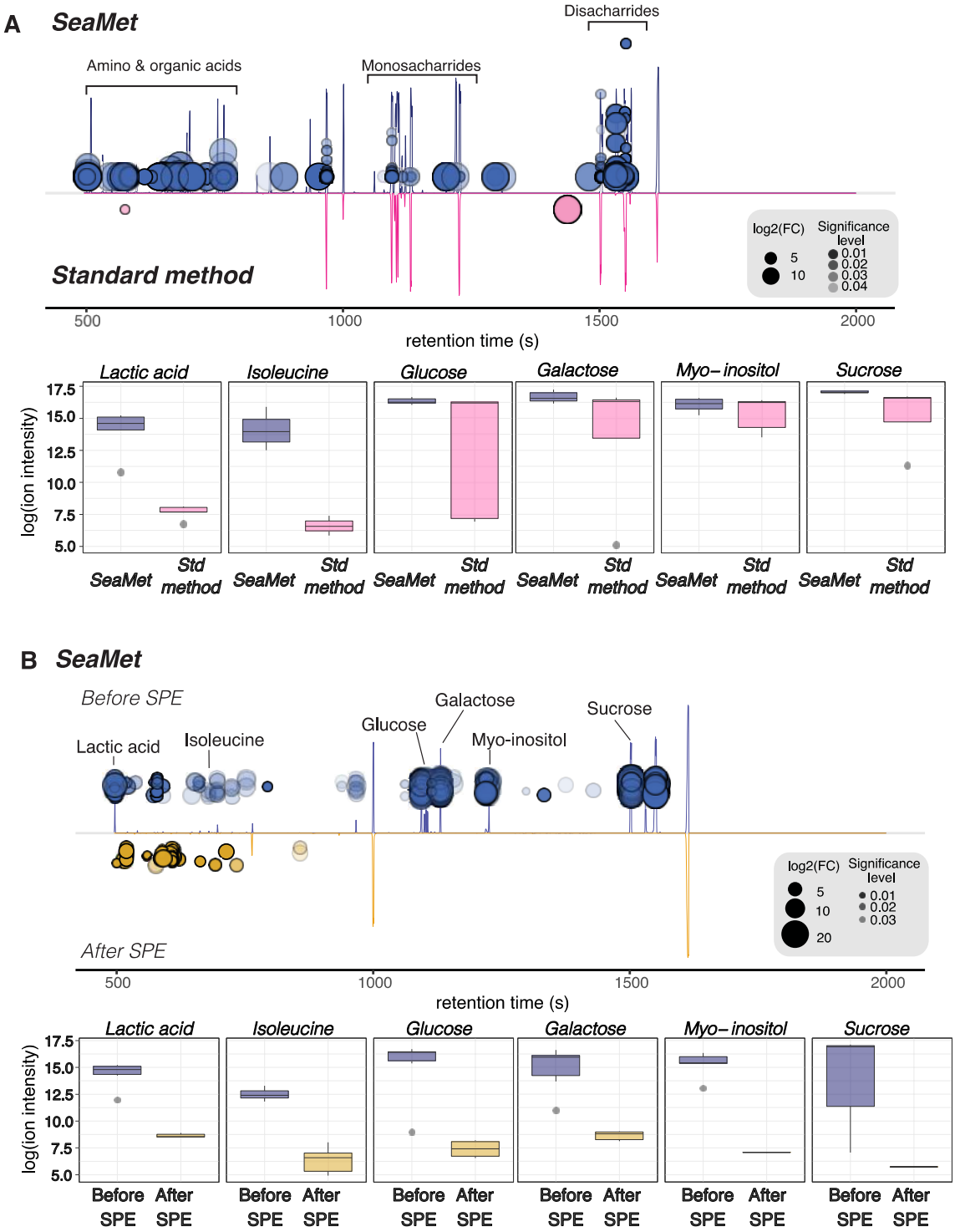

Figure 3. SeaMet enhances the detection of metabolites in marine samples. A, B Total ion chromatogram cloud plots from GC-MS profiles of metabolite mixtures indicate significant differences (Benjamini-Hochberg adjusted *p-value* < 0.05) between ion abundances when comparing **A,** SeaMet

180 (blue; top) to the standard metabolite derivatization (pink; bottom) protocol for GC-MS samples and **B,**
181 chromatograms using SeaMet on marine samples before (blue; top) and after (yellow; bottom) solid phase
182 extraction (SPE). Individual compound box plots are also shown in **A** and **B** to highlight improvements in
183 metabolite detection using SeaMet. For the cloud plots, larger bubbles indicate higher log2(fold changes)
184 between groups and more intense colors represent lower t-test p-values when comparing individual
185 feature (*m*/*z* ions) intensities. Samples prepared with SeaMet had high abundances of organic acids (lactic
186 acid, succinic acid, and fumarate), amino acids (isoleucine, leucine, threonine and valine), sugar alcohols
187 (myo-inositol and mannitol), and sugars (fructose, glucose, cellobiose, maltose, ribose, galactose, and
188 sucrose) in comparison to SPE-based sample preparation. Representatives of each class are indicated in **B**.
189 To show signal improvement using SeaMet, samples for both comparisons included authentic metabolite
190 standards representing multiple chemical classes.
191

192 To demonstrate the power of SeaMet in characterizing the metabolism of marine bacteria,
193 we monitored changes in the extracellular metabolome during growth of a heterotrophic
194 Gammaproteobacterium, *Marinobacter adhaerens* that occurs in aggregation with diatoms
195 throughout the North Sea. Using SeaMet, we simultaneously observed hundreds of metabolites
196 and detected significant changes in the metabolite composition of marine culture medium during
197 the bacteria's initial growth phase (adjusted p-value < 0.05; **Fig. 4; Fig. S3).** The bacteria took
198 up different carbon and nitrogen resources in a cascade-like fashion**,** and later in growth, began
199 excretion of an undescribed compound (**Fig. 4C, D; Fig. S3**). By measuring multiple metabolite
200 classes in a single analytical run, our results revealed that *M. adhaerens* preferentially took up
201 amino acids over readily available sugar compounds (e.g., trehalose, **Fig. S3**). Previous
202 proteomic results indicated that *M. adhaerens* had a high number of expressed amino acid uptake
203 transporters (34). Our results expand on these findings by i) highlighting which amino acids the
204 *M. adhaerens* prefers, ii) providing experimental evidence that this heterotroph does not take up
205 sugars, despite the genomic ability to use them in their metabolism (35) , and iii), showcasing
206 that *M. adhaerens* participates in the successional uptake of resources. Successional dynamics in

207 substrate use is a common energy conservation mechanism in bacteria (36) and affects central
208 carbon and nitrogen dynamics during growth. *M. adhaerens,* like many other bacteria,
209 participates in the release of organic carbon, which can be metabolized by other microorganisms
210 or will contribute to the complexity of refractory DOM.

211 Given that other exometabolomic methods for marine samples either miss major
212 compound classes due to sample pre-treatment (e.g., SPE based sample preparation) or are
213 targeted approaches that can only measure a few metabolite groups in a given run, it is likely
214 these observations in *M. adhaerens* physiology would have been obscured. Give the ease in
215 applying our method to culture studies, it is possible to integrate SeaMet with other "omics"
216 approaches to help illuminate microbial physiology in the marine environment. By identifying
217 and quantifying metabolites that are consumed and excreted in cultivable marine bacteria, our
218 method expands our understanding of key primary compounds involved in the transformation of
219 organic matter in the ocean.
220

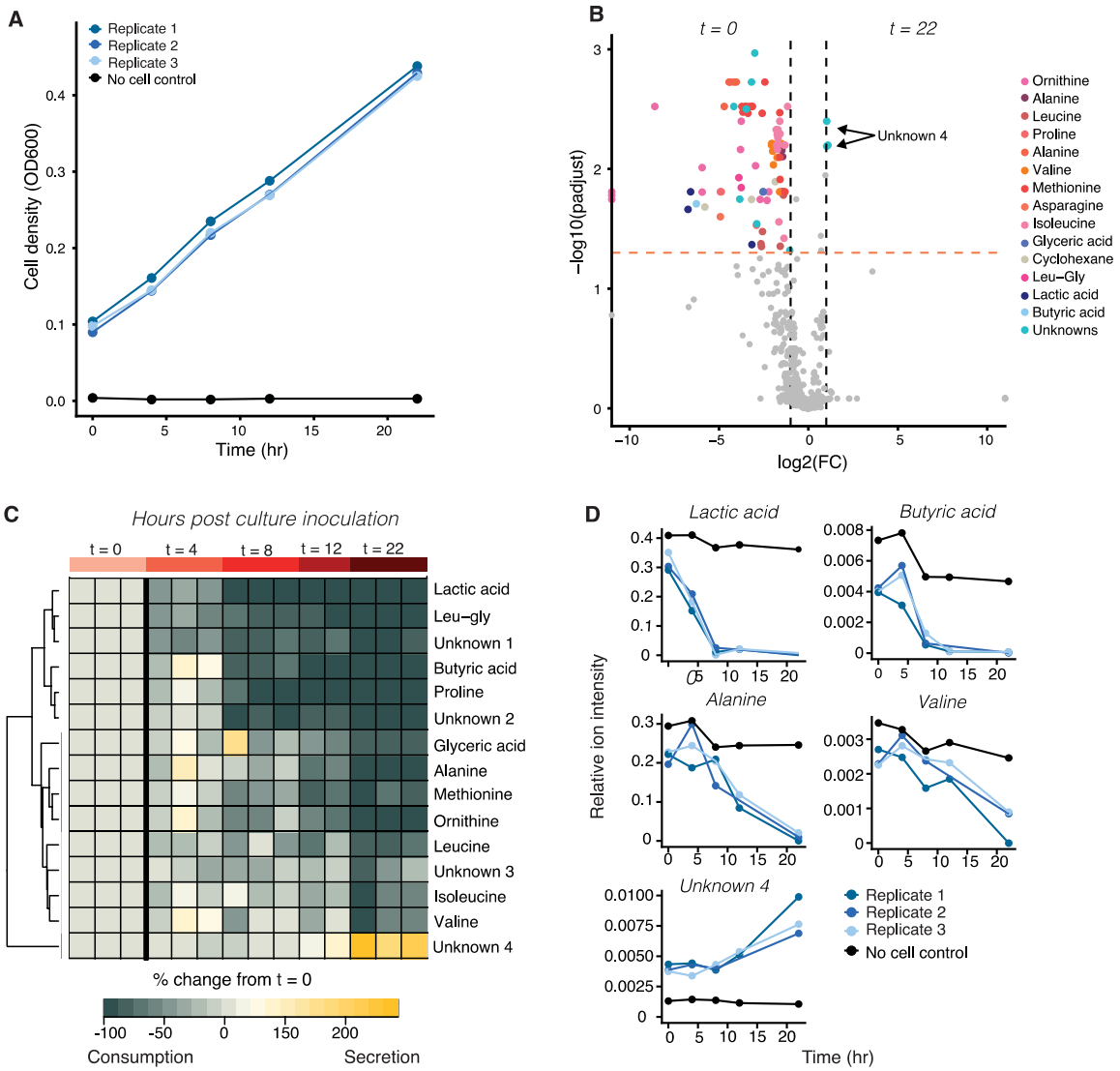

Figure 4. Metabolite consumption and excretion during culture of the marine heterotroph *Marinobacter adhaerens*. **A,** Cell densities increased during the first 22 hours of culture growth in Marine Broth. **B,** Volcano plot showing differences in ion abundances in cell growth media between the initial and final (22 hour) sampling time points. Variables exhibiting high fold change values (log2(fold change) > 2) and significant differences (p-adjusted < 0.05) between the two sampling time points are colored according to their metabolite database (NIST) annotation. **C,** A heatmap of metabolite abundances after 22 h relative to starting conditions indicates some compounds, like the dipeptide leucine-glycine (leu-gly), and lactic acid were taken up before others, such as branch chain amino acids. After 12–22 hours of growth, the bacteria excreted an unknown

231 compound (Unknown 4). Hierarchical clustering shows groups of metabolites that changed significantly
232 during growth (left hand colored bars, B.H. adjusted p-value < 0.05; fold change > 2). These metabolite
233 groups represent successive stages in *M. adhaerens* consumption and production of marine broth
234 components. **D,** Relative ion abundances over time for select metabolites from each cluster group shown
235 in **C.** The blue lines represent biological replicate cultures while the black line shows results from a
236 control sample with no cell addition. Low variation among biological replicates highlights the
237 reproducibility of SeaMet.
238

239 To test the ability of our workflow to assess complex environmental metabolomes, we
240 applied SeaMet to porewater samples from coralline and mangrove sediments. Coral reefs and
241 mangroves, two globally important coastal ecosystems, contain many biological compounds that
242 remain undescribed. It is essential to characterize the metabolome of these habitats to understand
243 the role of these ecosystems in biogeochemical cycling.

244 Our approach detected 295 and 428 metabolite peaks from coralline and mangrove
245 porewater profiles (**Fig. 5**), including sugars, amino acids, organic acids, fatty acids, and
246 background signals. Diverse and abundant sugars from sediment porewaters adjacent to corals,
247 as well as fatty acids from porewaters next to mangroves drove the observed significant
248 differences between habitats (ADONIS p-value < 0.001, $R^2 = 0.514$; **Fig. 5 and Table S5**).

249 Given that corals and mangroves thrive in oligotrophic waters and their associated
250 sediments harbor diverse, abundant and metabolically active microorganisms (37, 38), we were
251 surprised to measure high levels of metabolites that are typically consumed in primary
252 metabolism. Metabolomic analyses of marine sediments (in bulk) have also detected high
253 abundances of primary metabolites (39, 40), suggesting sediment habitats – which are globally
254 home to an estimated 2.9×10^{29} microbial cells (41) – contain many different types of
255 metabolites that drive microbial community metabolism. These data call for a reexamination of

carbon sequestration in coastal sediments using techniques that can identify and quantify the accumulation of liable metabolites.

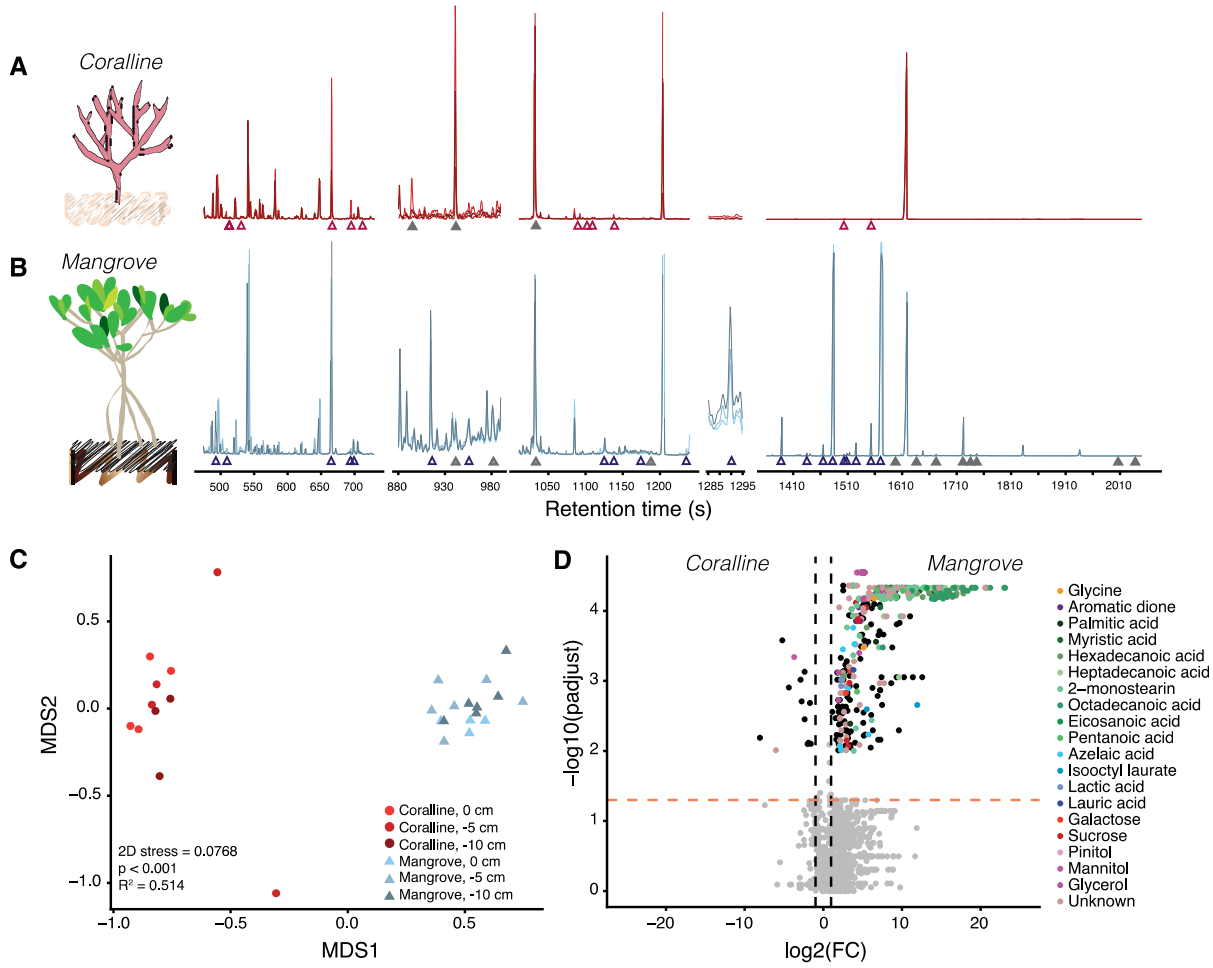

Figure 5. Metabolite profiles from marine habitats acquired with SeaMet. GC-MS metabolomic profiles from **A,** coralline and **B,** mangrove sediment porewaters showed high concentrations of identified metabolites (open triangles), e.g. fatty acids and sugars that explain multivariate differences in composition in **C** . Profiles also revealed unknown peaks (filled triangles) for which no matches were found in public databases (**Table S5**). **C,** Bray-Curtis informed non-metric multidimensional scaling analysis of sediment porewater metabolomic profiles from coralline (red) and mangrove (blue) habitats across sediment depths. ADONIS p-value and R^2 showed a significant correlation between sampling location and metabolite composition. **D,** Volcano plot showing differences in ion abundances between habitats. Significant ions (p-adjusted < 0.05) with log2-fold change > 2 are shaded according to their metabolite database (NIST) annotation.

Due to the technical difficulties of detecting metabolites in seawater, a large portion of ocean chemistry remains unannotated, reflecting one of the central challenges in metabolomics research (42). By providing a new method to measure a broad scope of the marine metabolome, we offer an avenue to identify molecules from marine environments and expand existing mass spectrometry databases that aim to characterize chemical space across ecosystems. As an example, our samples from sediment porewaters of mangroves and coral reefs revealed 11 metabolites driving variation between habitats that did not match public database entries (Fig. S6 and **Table S5**) (43, 44).

Conclusions

SeaMet is a marine metabolomics workflow that enables the analysis of primary metabolism in the oceans. It is time efficient, allows the detection of diverse metabolite classes in a single run, and expands the analytical window for molecules that can be detected within marine samples. This advance enables untargeted metabolomics for marine ecosystems using a low-cost, easy to use GC-MS platform. Moreover, SeaMet is independent of MS instrumentation, allowing it to be combined with time-of-flight or Orbitrap MS detectors to provide faster analysis time and higher mass resolving power to improve metabolite identification. We expect our marine metabolomics workflow will enable the exploratory analysis of metabolites occurring in seawater and thereby advance our understanding of the ocean's vast and largely unexplored metabolome.

293 ## Materials and Methods

294 **Data availability.** All metabolite profile data will be made publicly available at Metabolights
295 (https://www.ebi.ac.uk/metabolights/) under identification numbers MTBLS826, MTBLS839,
296 MTBLS843, MTBLS844, MTBLS848, and MTBLS849 (currently IN REVIEW) or by contact
297 with the authors. Reviewer links:
298 https://www.ebi.ac.uk/metabolights/reviewer5eb6b480436b019d9f1351a828ee7c3d
299 https://www.ebi.ac.uk/metabolights/reviewerd923ea1c3a53d000b97ccf383991032d
300 https://www.ebi.ac.uk/metabolights/reviewera9be9cf4-9a7d-4fff-98d5-c3d574c3b7f5
301 https://www.ebi.ac.uk/metabolights/reviewer08ce2d89-3945-45be-8a9b-4ea872fc86bf
302 https://www.ebi.ac.uk/metabolights/reviewer07a4ce73-1e8e-46aa-80c8-0c2f26411174
303 https://www.ebi.ac.uk/metabolights/reviewer2878413e6f8a6a883b27bdee8c1bbba6
304

305 **Reagents and experimental sample preparation.** The derivatization chemicals, trimethylsilyl-
306 N-methyl trifluoroacetamide (MSTFA) and N,O-Bis(trimethylsilyl)trifluoroacetamide (BSTFA)
307 were obtained from CS-Chromatographie Service and pyridine from Sigma-Aldrich at >99.98%
308 purity. Methoxyamine hydrochloride (MeOX; Sigma-Aldrich) aliquots were further dried at
309 60 °C in a drying oven for 1 h to remove residual moisture. Artificial seawater (ASW) was
310 prepared within the range of natural salinity (36‰) by dissolving (per L of water) 26.37 g
311 sodium chloride, 6.8 g magnesium sulfate heptahydrate, 5.67 g magnesium chloride hexahydrate,
312 1.47 g calcium chloride, 0.6 g potassium chloride, and 0.09 g potassium bromide. Following
313 autoclave sterilization, pH was adjusted to 7.7 using sodium hydroxide. 1 mL of the following
314 supplements and solutions were added: 150 mM monopotassium phosphate, 500 mM ammonium
315 chloride pH 7.5, trace element solution, selenite-tungstate solution, vitamin solution, thiamine

316 solution, B12 solution and 0.21 g sodium bicarbonate (45). Ultra-pure water (MQ) was prepared
317 by purifying deionized water with an Astacus membraPure system (Astacus membraPure, 18.3
318 mΩ × cm 25 °C).

319 Metabolite standards were obtained from commercial sources (**Table S4**) and combined
320 into mixtures in which each compound had a final concentration of 0.4 mM. Metabolite mixtures
321 were prepared to (a) test the effect of salt and water on metabolite detection, (b) develop SeaMet,
322 our marine metabolomics workflow, (c) compare metabolite detection before and after solid
323 phase extraction (SPE) based sample preparation, and (d) to quantify the detection limits of
324 specific compound classes (**Table S6**). Finally, multiple mixtures were prepared to document the
325 retention times of 107 standards dissolved in ASW using SeaMet (**Table S4**). Sample aliquots
326 for the above mentioned experiments were prepared by drying down 200 µL of the mixture in a
327 speed vacuum concentrator (Eppendorf Concentrator Plus$^{(R)}$, 2.5 h, 45°C, V-AQ) for all
328 experiments except SPE comparison and quantification of detection limits. For the SPE
329 comparison experiment, 400 µL of the mix were dried down. For the quantification of metabolite
330 classes, a serial dilution of the mix was prepared to obtain concentrations between 0.5 nmol and
331 80 nmol of each compound. All dried mixture samples were stored at 4 °C.

332

333 **SeaMet metabolite derivatization.** To prepare marine samples for gas chromatography-mass
334 spectrometry (GC-MS) analysis, 0.5 to 1 mL of a saltwater sample or experimental mixture
335 dissolved in ASW was dried in a speed vacuum concentrator for 8 hours (Eppendorf
336 Concentrator Plus$^{(R)}$, 45°C, V-AQ). To further remove residual water locked within the salt
337 pellet, 250 µL of toluene (99.8%, < 0.2 % water) was added to each sample and the mixture was
338 ultrasonicated for 10 min at maximum intensity. The toluene was subsequently removed under a

339 gentle flow of N_2 gas. Metabolite derivatization was performed by adding 80 µL of MeOX
340 dissolved in pyridine (20 mg × mL^{-1}) to the dried pellet. The mixture was ultrasonicated (EMag
341 Emmi-12HC®) for 10 min at maximum intensity, briefly vortexed to dissolve the pellet into
342 solution, and subsequently incubated for 90 min at 37 °C using a thermal rotating incubator
343 under constant rotation at 1350 rpm. The pyridine was removed from the sample at room
344 temperature under a gentle flow of N_2 gas (approximately 1 hour). Following the addition of 100
345 µL of BSTFA, the mixture was ultrasonicated for 10 min at maximum intensity, vortexed, and
346 incubated for 30 min at 37 °C using a thermal rotating incubator under constant rotation at 1350
347 rpm. The derivatized mixture was ultrasonicated for 10 min at maximum intensity. Remaining
348 salt in each sample was pelleted through centrifugation at 21.1 *g* for 2 min at 4 °C. 100 µL was
349 transferred to a GC-MS vial for analysis. The full proposed method is publicly available at
350 dx.doi.org/10.17504/protocols.io.nyxdfxn.
351
352 **GC-MS data acquisition.** All derivatized samples were analyzed on an Agilent 7890B GC
353 coupled to an Agilent 5977A single quadrupole mass selective detector. Using an Agilent 7693
354 autosampler, 1 µL was injected in splitless mode through a GC inlet liner (ultra inert, splitless,
355 single taper, glass wool, Agilent) onto a DB-5MS column (30 m × 0.25 mm, film thickness 0.25
356 µm; including 10 m DuraGuard column, Agilent). The inlet liner was changed every 50 samples
357 to avoid damage to the GC column and associated shifts in retention times. The injector
358 temperature was set at 290 °C. Chromatography was achieved with an initial column oven
359 temperature set at 60 °C followed by a ramp of 20 °C min^{-1} until 325 °C, then held for 2 mins.
360 Helium carrier gas was used at a constant flow rate of 1 mL min^{-1}. Mass spectra were acquired in
361 electron ionization mode at 70 eV across the mass range of 50–600 m/z and a scan rate of 2

362 scans s^{-1}. The retention time for the method locked using standard mixture of fatty acid methyl
363 esters (Sigma Aldrich).
364
365 **Data processing and analysis.** Raw Agilent data files were converted to mzXML files using
366 Msconvert (46) and imported into XCMS (v. 2.99.6)(47) within the R software environment (v.
367 3.4.2) for data processing and analysis. Total ion chromatograms (TIC) were obtained using the
368 xcmsRaw function. TICs comparing sample preparation steps were expressed as a percentage of
369 the MQ control. For environmental and cell culture GC-MS profiles, peaks were picked using the
370 matchedFilter algorithm in XCMS with a full width at half maximum set to 8.4, signal to noise
371 threshold at 1, *m*/*z* width of 0.25 (step parameter), and *m*/*z* difference between overlapping peaks
372 at 1 (**Supplemental Text 1**). Resulting peaks were grouped, retention times corrected and
373 regrouped using the density (bandwidth parameter set to 2) and obiwarp methods. Following
374 peak filling, the CAMERA (v.1.32.0)(48) package was used to place *m*/*z* peaks into pseudo-
375 spectra by grouping similar peaks with the groupFWHM function. Masses below 150 *m*/*z* were
376 removed from the resulting peak table and all profiles were normalized to the ribitol internal
377 standard. Peaks occurring in run blanks and those with higher relative standard deviation scores
378 (% RSD > 25) in quality control samples (cell culture experiment only) were removed from the
379 dataset. To determine differences in metabolite abundances between sediment habitats,
380 metabolite peak data were analyzed using a Bray-Curtis informed non-metric multidimensional
381 scaling analysis followed by an analysis of variance using distance matrices (ADONIS) to test if
382 there are significant differences in metabolite composition between sites. To identify individual
383 peaks that differed significantly between sediment habitats and between cell culture sampling
384 time points, resulting peaks tables were also log transformed and compared using a one-way

385 analysis of variance. All p-values were adjusted using the Benjamini-Hochberg (B.H.) method to
386 control for false positives (49). Significant variables exhibiting large fold-change differences
387 between starting and ending conditions were further investigated. CAMERA grouped peaks from
388 the environmental survey, and those important to shifts in the cell culture experiment were
389 identified using AMDIS (50). Peaks with NIST hits below 800 were compared to the online data
390 repositories, BinVestigate (44) and Golm (43) using the calcualted Kovats retention indices (51)
391 based on a reference n-alkane standard (C7-C40 Saturated Alkanes Standards, Sigme-Aldrich). If
392 no hit was provided, these were considered unknowns.
393
394 **The effect of salt and water on metabolite detection.** To test the effect of salt on metabolite
395 derivatization, metabolite mix aliquots were resuspended in 1 mL of ASW ranging in salinity
396 from 0 to 34‰ and dried as described above. Methoxamine-trimethylsilylation (TMS) two step
397 derivatization was performed by resuspending each sample in 80 µL of MeOX in pyridine (20
398 mg mL^{-1}) and incubating for 90 min at 37 °C using a thermal rotating incubator under constant
399 rotation at 1350 rpm. MSTFA was subsequently added to the mixture, and the mixture incubated
400 under the same conditions for 90 min (29). Derivatized samples were centrifuged to pellet salt
401 and the supernatant was transferred to a GC-MS vial for analysis. To test the independent effect
402 of water on metabolite derivatization reactions, MQ was added to dried mixture aliquots in steps
403 of 1 µL from 0 to 10 µL. Replicate water gradient samples were subsequently derivatized as
404 before using MeOX and MSTFA or by replacing the MSTFA reagent with BSTFA.
405
406 **Marine metabolomics method development.** To show how each method development step
407 increased signal intensity and reduced variation in metabolite detection, replicate mixture

408 aliquots (n = 5) were resuspended in 0.5 mL of ASW. Mixture aliquots (n = 5) were also
409 resuspended in MQ as a no-salt control to highlight the effects of saltwater on metabolite
410 derivatization. 40 µL ribitol (0.2 mM) and 100 µL cholestane (1 mM) were added to each aliquot
411 as internal standards. MQ and ASW samples were first derivatized following the (i) two-step
412 methoxamine-trimethylsilylation (TMS) previously described. Successive steps in the proposed
413 protocol were then applied to ASW samples to demonstrate the combined effects on metabolite
414 detection: (ii) exchange of MSTFA for BSTFA, (iii) removal of residual water from the salt
415 pellet by increasing the speed vacuum drying time and by introducing a toluene drying step to
416 help extract water from the salt pellet, (iv) ultrasonication of the samples after the steps involving
417 addition of toluene, MeOX, BSTFA and following the last derivatization step, and (v) drying the
418 MeOX in pyridine reagent between derivatization reactions. Resulting GC-MS profiles were
419 used to show increases in total signals detected with successive changes in the proposed
420 protocol. Additionally, a cloud plot (using processed peak integration data) was generated to
421 compare compounds dissolved in seawater and to show which metabolite ions exhibited
422 significant (B.H. adjusted $p < 0.05$) and large fold changes (log2(FC) > 2) between the standard
423 and the SeaMet method.
424
425 **Solid phase extraction.** Replicate metabolite mix aliquots (n = 6) were resuspended in 2 mL of
426 artificial seawater. 0.5 mL was reserved from each sample to compare GC-MS profiles before
427 and after SPE sample concentration. Inorganic salts were eluted and metabolites extracted from
428 the remaining 1.5 mL mixture following a SPE based technique using Bond Elut styrene-
429 divinylbenzene (PPL) columns (17). The internal standards ribitol and cholestane were added to
430 both, the reserved sample (before SPE) and the resulting SPE-concentrated sample (after SPE).

431 All samples were prepared for GC-MS analysis following the proposed marine metabolomics
432 method. Resulting profiles were compared using a cloud plot to show which metabolite ions
433 exhibited significant (B.H. adjusted $p < 0.05$) and large fold changes (log2(FC) > 2) between the
434 pre- and post- SPE treatments.
435
436 **Environmental sampling.** Replicate porewater profiles were collected from coralline ($n = 4$)
437 and mangrove ($n = 6$) sediments from Carrie Bow Cay (N 16° 04' 59", W 88° 04' 55") and Twin
438 Cayes, Belize (N 16° 50' 3", W 88° 6' 23") using a 1 m steel lance with a 2 µm inner diameter
439 covered by 0.063 mm steel mesh. Samples (2 mL water) were collected every 5 cm from the
440 sediment surface to 15 cm depth. Samples were immediately frozen at -20 °C until further
441 analysis. Directly before preparation for GC-MS, the internal standards ribitol and cholestane
442 were added to 0.5 mL of each environmental sample. The mixture was subsequently prepared for
443 GC-MS analysis using the SeaMet method described above.
444
445 **Cell culture sampling.** Replicate cultures ($n = 3$) *of Marinobacter adhaerens* HP15 DsRed were
446 cultivated in Marine Broth media at 18 °C and 240 rpm as previously described (34). Media
447 samples from the cell cultures and a no-bacteria control media were collected at 0, 4, 8, 12, and
448 22 h post culture inoculation. Cell counts were monitored at each time point by measuring the
449 optical density at 600 nm (OD_{600}). Sampling was carried out by collecting 2 mL of each culture
450 and pelleting the cells through centrifugation for 10 min, at 21.1 *g* and 4 °C. The supernatant was
451 immediately stored at -20 °C until preparation for GC-MS analysis. Prior to sample
452 derivatization using SeaMet, ribitol (0.2 mM; 40 µL) and cholestane (100 mM; 100 µL) were
453 added to 0.5 mL of each experimental sample and subsequently dried down in a speed vacuum

concentrator (8 hr, 45 °C, VA-Q). To control for technical variation, quality control (QC) samples (n = 3) were prepared by combining 0.25 µL of each culture supernatant and an extraction blank generated by drying down 0.5 mL of MQ.

Supporting information

Supporting information includes supporting tables, figures, references and XCMS peak picking script.

Acknowledgements

We thank T. Gulstad, K. Caspersen, F. Fojt, and M. Meyer (MPI-Bremen) for support with data acquisition and sample preparation, N. Böttcher (Jacobs University Bremen) for providing culture samples, D. Michellod for sediment pore water sample collection, and B. Geier and J. Beckmann (MPI-Bremen) for valuable discussions. We acknowledge the Max-Planck Society and the Gordon and Betty Moore Foundation (Marine Microbial Initiative Investigator Award to ND, Grant #GBMF3811) for financial support. This work is contribution *XXX* from the Carrie Bow Cay Laboratory, Caribbean Coral Reef Ecosystem Program, National Museum of Natural History, Washington DC.

References

1. Moran MA, *et al.* (2016) Deciphering ocean carbon in a changing world. *Proc Natl Acad Sci U S A* 113(12):3143-3151.

2. Lechtenfeld OJ, Hertkorn N, Shen Y, Witt M, & Benner R (2015) Marine sequestration of carbon in bacteria metabolites. *Nat Comm* 6:6711.

3. Sabine CL, *et al.* (2004) The oceanic sink for anthropogenic CO_2. *Science* 305:367-371.

4. Sunagawa S, *et al.* (2015) Structure and function of the global ocean microbiome. *Science* 384(6237).

5. Kell DB, *et al.* (2005) Metabolic footprinting and systems biology: the medium is the message. *Nat Rev Microbiol* 3(7):557-565.

6. Swenson TL, Jenkins S, Bowen BP, & Northen TR (2015) Untargeted soil metabolomics methods for analysis of extractable organic matter. *Soil Biol and Biochem* 80:189-198.

7. Swenson TL, Karaoz U, Swenson JM, Bowen BP, & Northen TR (2018) Linking soil biology and chemistry in biological soil crust using isolate exometabolomics. *Nat Commun* 9(1):19.

8. Allen J, *et al.* (2003) High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat Biotech* 21(6):692-696.

9. Paczia N, *et al.* (2012) Extensive exometabolome analysis reveals extended overflow metabolism in various microorganisms. *Microbe Cell Fact* 11(22).

10. Ternon E, *et al.* (2016) Spherulization as a process for the exudation of chemical cues by the encrusting sponge *C. crambe*. *Sci Rep* 6:29474.

11. Kuhlisch C & Pohnert G (2015) Metabolomics in chemical ecology. *Nat Prod Rep* 32(7):937-955.

12. May MA, Bishop KD, & Rawson PD (2017) NMR profiling of metabolites in larval and juvenile blue mussels (*Mytilus edulis*) under ambient and low salinity conditions. *Metabolites* 7(3).

13. Sogin EM, Putnam HM, Nelson CE, Anderson P, & Gates RD (2017) Correspondence of coral holobiont metabolome with symbiotic bacteria, archaea and Symbiodinium communities. *Environ Microbiol Rep* 9(3):310-315.

14. Goulitquer S, Potin P, & Tonon T (2012) Mass spectrometry-based metabolomics to elucidate functions in marine organisms and ecosystems. *Mar Drugs* 10(4):849-880.

15. Selander E, *et al.* (2016) Solid phase extraction and metabolic profiling of exudates from living copepods. *PeerJ* 4.

16. Wienhausen G, Noriega-Ortega BE, Niggemann J, Dittmar T, & Simon M (2017) The exometabolome of two model strains of the roseobacter group: a marketplace of microbial metabolites. *Front Microbiol* 8:1985.

17. Dittmar T, Koch B, Hertkorn N, & Kattner G (2008) A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol Oceanogr* 6:230-235.

18. Amin SA, *et al.* (2015) Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* 522(7554):98-101.

19. Sanudo-Wilhelmy SA, *et al.* (2012) Multiple B-vitamin depletion in large areas of the coastal ocean. *Proc Natl Acad Sci U S A* 109(35):14041-14045.

20. Koch BP, Ludwichowski K-U, Kattner G, Dittmar T, & Witt M (2008) Advanced characterization of marine dissolved organic matter by combining reversed-phase liquid chromatography and FT-ICR-MS. *Mar Chem* 111(3-4):233-241.

21. Petras D*, et al.* (2017) High-Resolution Liquid Chromatography Tandem Mass Spectrometry enables large scale molecular characterization of dissolved organic matter. *Front Mar Sci* 4. DOI:10.3389/fmars.2017.00405

22. Jiao N*, et al.* (2010) Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nat Rev Microbiol* 8(8):593-599.

23. A J*, et al.* (2005) Extraction and GC/MS analysis of the human blood plasma metabolome. *Anal Chem* 77(24):8086-8094.

24. Fiehn O*, et al.* (2000) Metabolite profiling for plant functional genomics. *Nat Biotechnol* 18:1157-1161.

25. Theriot CM*, et al.* (2014) Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. *Nat Commun* 5:3114.

26. Bruce TJ*, et al.* (2008) cis-Jasmone induces *Arabidopsis* genes that affect the chemical ecology of multitrophic interactions with aphids and their parasitoids. *Proc Natl Acad Sci U S A* 105(12):4553-4558.

27. Fischer E & Sauer U (2003) Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS. *Euro J Biochem* 270(5):880-891.

28. Villas-Boas SG, Noel S, Lane GA, Attwood G, & Cookson A (2006) Extracellular metabolomics: a metabolic footprinting approach to assess fiber degradation in complex media. *Anal Biochem* 349(2):297-305.

29. Roessner U, Wagner C, Kopka J, Trethewey RN, & Willmitzer L (2000) Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *Plant J* 23(1):131-142.

30. Chan EC, Pasikanti KK, & Nicholson JK (2011) Global urinary metabolic profiling procedures using gas chromatography-mass spectrometry. *Nat Protoc* 6(10):1483-1499.

31. Kind T*, 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.

32. Liebeke M & Puskas E (2019) Drying enhances signal intensity for global GC-MS metabolomics. *Metabolites* 9(4).

33. Johnson WM, Kido Soule MC, & Kujawinski EB (2017) Extraction efficiency and quantification of dissolved metabolites in targeted marine metabolomics. *Limnol Oceanogr* 15(4):417-428.

34. Stahl A & Ullrich MS (2016) Proteomics analysis of the response of the marine bacterium *Marinobacter adhaerens* HP15 to the diatom *Thalassiosira weissflogii*. *Aquat Microb Ecol* 78(2):65-79.

583 35. Gardes A, *et al.* (2010) Complete genome sequence of *Marinobacter adhaerens* type strain
584 (HP15), a diatom-interacting marine microorganism. *Stand Genomic Sci* 3(2):97-107.
585
586 36. Behrends V, Ebbels TM, Williams HD, & Bundy JG (2009) Time-resolved metabolic
587 footprinting for nonlinear modeling of bacterial substrate utilization. *Appl Environ Microbiol*
588 75(8):2453-2463.
589
590 37. Ian H & Jed AF (2006) Spatial and vertical biogeography of coral reef sediment bacterial and
591 diazotroph communities. *MEPS* 306:79-86.
592
593 38. Andreote FD, *et al.* (2012) The microbiome of Brazilian mangrove sediments as revealed by
594 metagenomics. *PLoS One* 7(6):e38600.
595
596 39. Beale DJ, *et al.* (2018) Seasonal metabolic analysis of marine sediments collected from Moreton
597 Bay in South East Queensland, Australia, using a multi-omics-based approach. *Sci Total Environ*
598 631-632:1328-1341.
599
600 40. Kimes NE, *et al.* (2013) Metagenomic analysis and metabolite profiling of deep-sea sediments
601 from the Gulf of Mexico following the Deepwater Horizon oil spill. *Front Microbiol* 4:50.
602
603 41. Kallmeyer J, Pockalny R, Adhikari RR, Smith DC, & D'Hondt S (2012) Global distribution of
604 microbial abundance and biomass in subseafloor sediment. *Proc Natl Acad Sci U S A*
605 109(40):16213-16216.
606
607 42. Luzzatto-Knaan T, *et al.* (2017) Digitizing mass spectrometry data to explore the chemical
608 diversity and distribution of marine cyanobacteria and algae. *Elife* 6. DOI:10.7554/eLife.24214
609
610 43. Kopka J, *et al.* (2005) GMD@CSB.DB: the Golm metabolome database. *Bioinformatics*
611 21(8):1635-1638.
612
613 44. Lai Z, *et al.* (2018) identifying metabolites by integrating metabolome databases with mass
614 spectrometry cheminformatics. *Nat Methods* 15(1):53-56.
615
616 45. Widdel F & Bak F (1992) Gram-Negative Mesophilic Sulfate-Reducing Bacteria. *The*
617 *Prokaryotes: A Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Identification,*
618 *Applications*, eds Balows A, Trüper HG, Dworkin M, Harder W, & Schleifer KH (Springer, New
619 York, NY).
620
621 46. Holman JD, Tabb DL, & Mallick P (2014) Employing ProteoWizard to Convert Raw Mass
622 Spectrometry Data. *Curr Protoc Bioinformatics* 46:13 24 11-19.
623
624 47. Smith CA, Want EJ, O'Maille G, Abagyan R, & Siuzdak G (2006) XCMS: Processing mass
625 spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and
626 identification. *Anal Chem* 78 (3): 779-787.
627
628 48. Kuhl C, Tautenhahn R, Bottcher C, Larson TR, & Neumann S (2012) CAMERA: an integrated
629 strategy for compound spectra extraction and annotation of liquid chromatography/mass
630 spectrometry data sets. *Anal Chem* 84(1):283-289.
631
632 49. Benjamini Y & Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful
633 approach to multiple testing. *J R Stat Soc Series B* 57(1):289-300.

50. Stein SE (1999) An integrated method for spectrum extraction and cCompound identification from gas chromatography/mass spectrometry data. *JASMS* 10:770-781.

51. Kovats E (1958) Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone. *Helv. Chim. Acta.* 41(7):1915-1932.

Supplementary Information for

Marine metabolomics: a method for the non-targeted measurement of metabolites in seawater by gas-chromatography mass spectrometry.

E. Maggie Sogin[1*], Erik Puskas[1], Nicole Dubilier[1], Manuel Liebeke[1*]

[*]Corresponding authors

Maggie Sogin, email: esogin@mpi-bremen.de, tel: +49 421 2028-823

Manuel Liebeke email: mliebeke@mpi-bremen.de, tel: +49 421 2028-822

This PDF file includes

Figures S1-S3

Tables S1-S6

Supplementary Text 1

Supplementary references

Supporting Figures and Tables

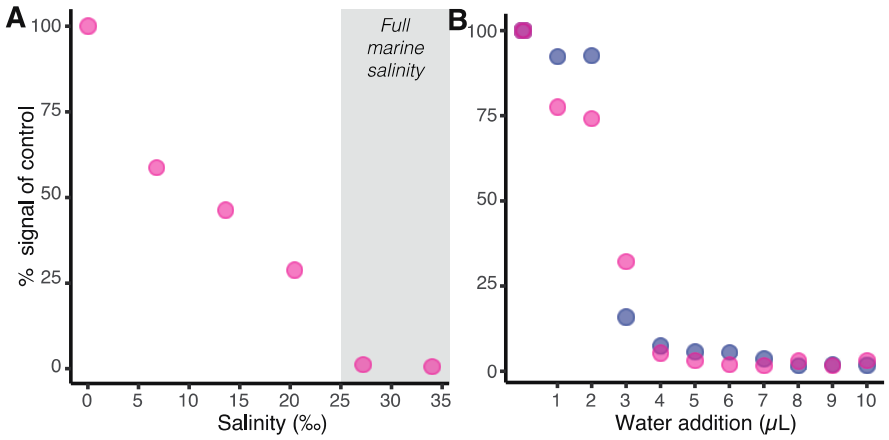

Figure S1. Salt and water inhibit metabolite derivatization reactions. Total detected ion signals are negatively related to increasing concentrations of **A,** salt and **B,** water for both MSTFA (red circles) and BSTFA (blue circles) derivatization reagents. Signals intensities from the metabolite mixture (**Table S6**) are relative to control samples (**A,** no salt; **B,** no water). To avoid damage to the GC-MS instrumentation, only 1 replicate / condition was analyzed.

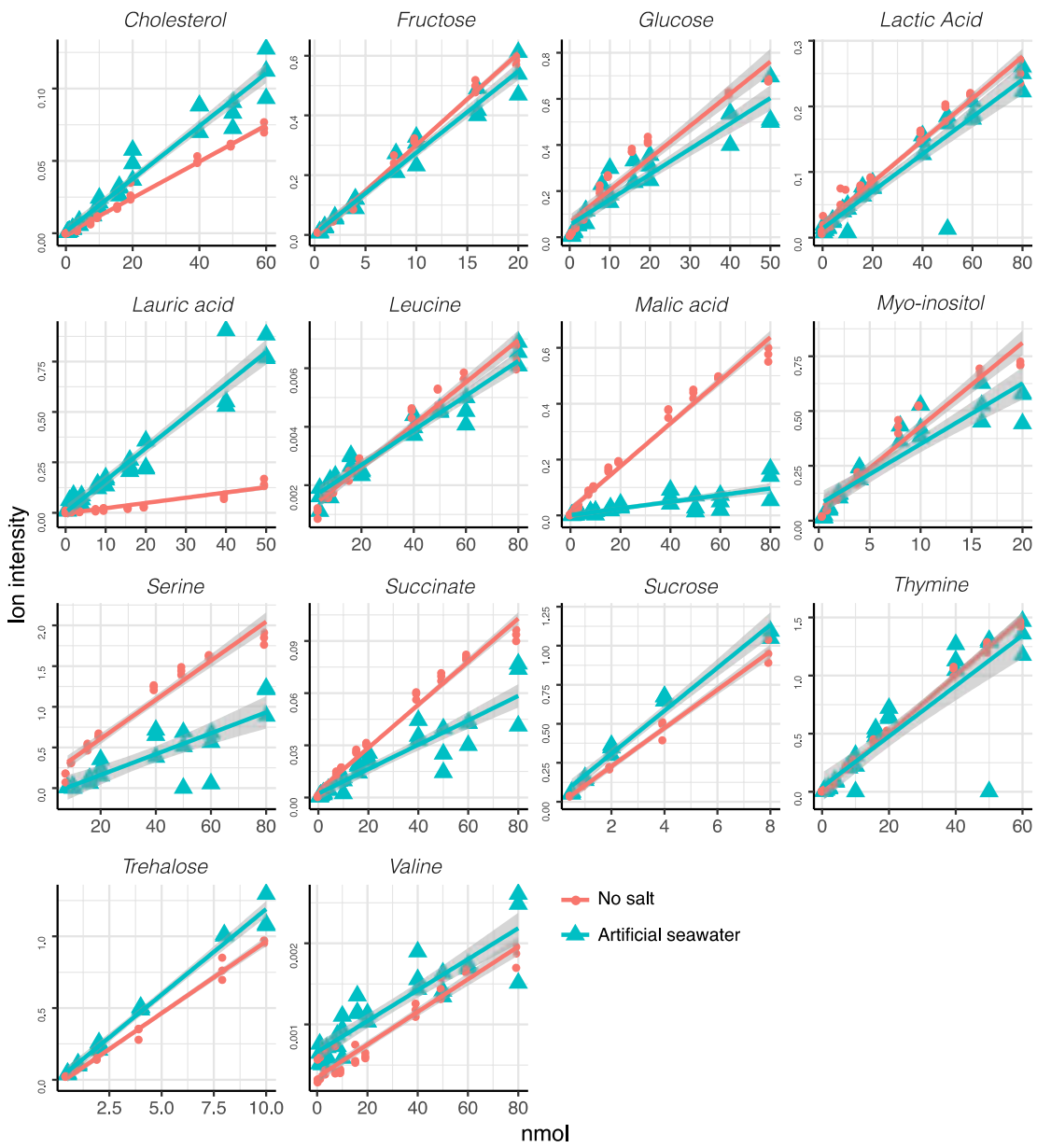

Figure S2. Calibration curves for individual metabolites in salt free and artificial seawater (ASW). Calculated calibration curves were compared for compounds that were detected in both salt-free and ASW conditions (n=3 for each concentration). Gray shading represents 90% confidence intervals and points are fitted using a linear regression. Model results are reported in Table S2.

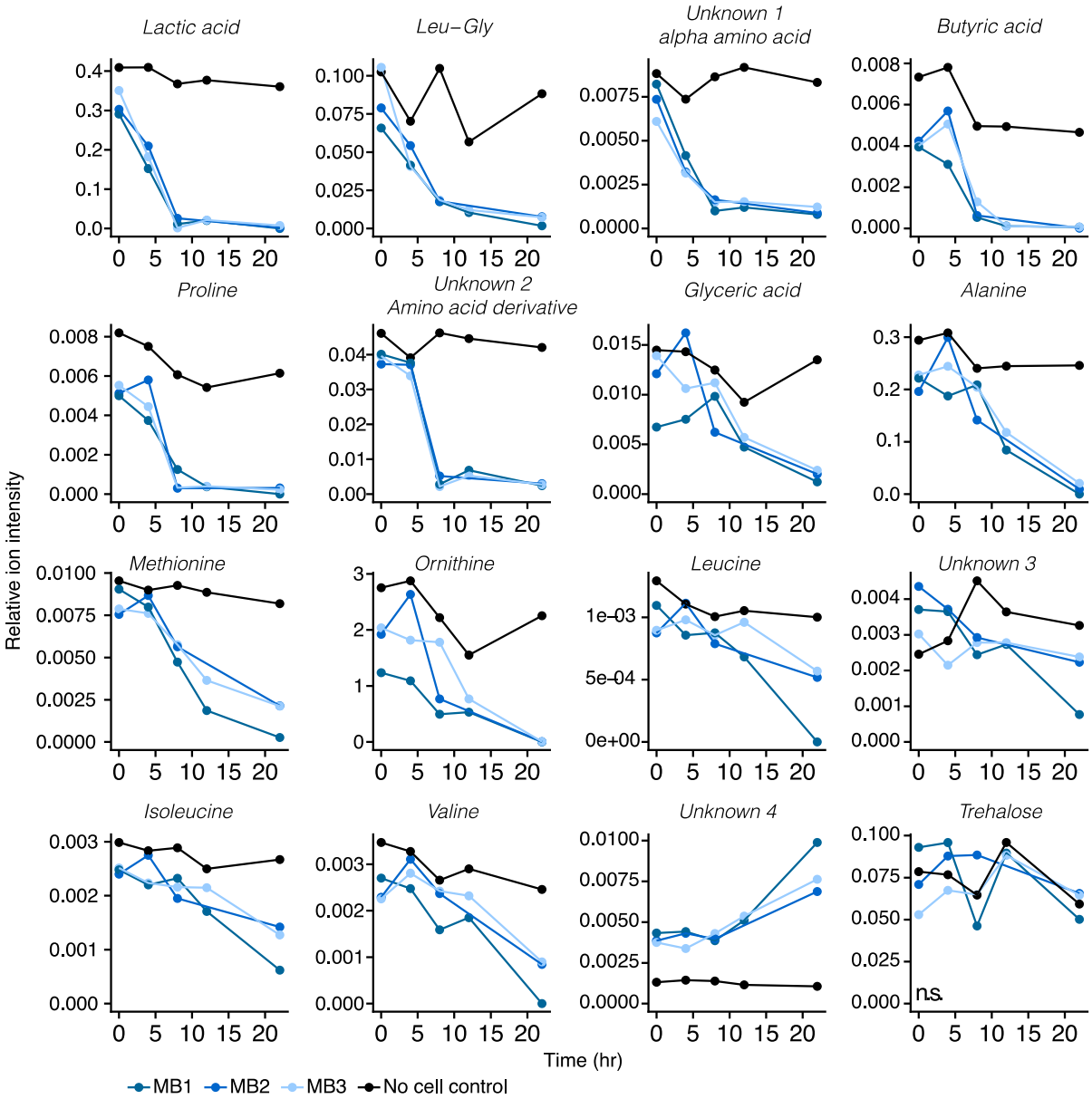

Figure S3. Extracellular metabolite levels shift with cell culture density. Metabolite relative abundances for each cell culture and no-cell control are plotted through time. Only metabolites that significantly (adjusted p-value < 0.05) varied with time in replicate culture experiments are plotted for clarity.

Table S1. Mixture of 45 metabolites used for the development and testing of the sample preparation method, with their retention times as detected in samples dissolved in artificial seawater using GC-MS. Metabolites with multiple retention times represent different TMS derivatives.

Compound	Retention time (s)	Compound class
Alanine	883	Amino acid
Arginine	Not detected	Amino acid
Aspartic acid	788	Amino acid
Carnitine	Not detected	Amino acid
Cellobiose	1536	Disaccharides
Citric acid	Not detected	Organic acid
Cysteine	Not detected	Amino acid
Dimethylsulfoniopropionate	Not detected	Organic compound
Fructose	1099/1106	Monosaccharides
Fumerate	732	Organic acid
Galactose	1109	Aldoses
Glucose	1112	Monosaccharides
Glutamate	Not detected	Amino acid
Glutamic acid	931	Amino acid
Glutamine	824	Amino acid
Glycerophosphate	Not detected	Organic compound
Glycine	700	Amino acid
Histidine	Not detected	Amino acid
Isoleucine	687	Amino acid
Lactic acid	512	Organic acid
Lauric acid	956	Fatty acid
Leucine	582/669	Amino acid
Lysine	1130	Amino acid
Maleic acid	695	Dicarboxylic acid
Malic acid	835	Dicarboxylic acid
Maltose	1557/1565	Disaccharides
Mannitol	1138	Sugar alcohol
Mannose	1118	Monosaccharides
Methionine	781/860	Amino acid
Myo-inositol	1231	Sugar alcohol
NAG	1223	Sugar
Ornithine	931/1064	Amino acid
Oxalic Acid	558	Dicarboxylic acid
Phenylalanine	938	Amino acid
Proline	692	Amino acid
Pyruvate	Not detected	Organic acid
Ribose	971	Aldoses
Serine	659/738	Amino acid
Succinic acid	704	Organic acid
Sucrose	1508	Disaccharides
Threonine	785	Amino acid
Thymine	Not detected	Amino acid
Trehalose	1557	Disaccharides
Tryptophan	1289	Amino acid
Urea	648	Organic compound
Valine	534	Amino acid

Table S2. Quantification ions, calibration coefficients, and retention of metabolites from each compound class in no salt and salt conditions. The minimum and maximum calibration points of select compounds from representatives of each metabolite class in artificial seawater are reported. The lowest concentration at which signals were observed is reported for select compounds.

			No salt			Salt					
Compound	**Category**	**quantification ion (m/z)**	**Retention time (s)**	**r^2**	**%CV 4 nmol**	**Retention time (s)**	**r^2**	**% CV 4 nmol**	**Min calibration point (nmol)***	**Max calibration point (nmol)***	**Lowest conc.****
Glycine 3TMS	Amino acid	204.1	689	0.955	17.12	689	**< 0.7**	4.68	0.5	80	
Histidine 2TMS	Amino acid	218.1	1126	0.973	20.73	*Not detected*					
Leucine 2 TMS	Amino acid	177.06	659	0.961	17.24	660	0.939	25.93	4	80	
Lysine 3TMS	Amino acid	156.1	989	0.973	31.31	*Not detected*					
Lysine 4TMS	Amino acid	174.1	*Not detected*			1127	**< 0.7**	15.28	0.5	80	
Ornithine 3TMS	Amino acid	174.1	928	0.992	18.19	*Not detected*					
Ornithine 4TMS	Amino acid	174.1	*Not detected*			1061	**< 0.7**	17.38	0.5	80	
Serine 3TMS	Amino acid	204.16	732	0.957	NA	733	**< 0.7**		8	80	
Thymine 2TMS	Amino acid	255.1	765	0.996	22.65	766	0.775	7.11	0.5	60	0.25
Valine 2TMS	Amino acid	188.09	612	0.94	32.15	614	0.828	10.94	1	80	0.25
Citrate/citric acid 3TMS	Organic acid	244.14	1061	0.954	16.68	*Not detected*					
Glutamic acid 3TMS	Organic acid	339.13	933	**< 0.7**	46.64	*Not detected*					
Glutamic acid 3TMS	Organic acid	246.1	*Not detected*			929	**< 0.7**	18.1	1	80	
Glutamic acid dervative - *Hydroxproline*	Organic acid	216.09	860	0.995	18.33	*Not detected*					
Glutamic acid dervative - *Hydroxproline*	Organic acid	221.09	860	**< 0.7**	14.13	859	**< 0.7**	4.68	0.5	80	
Lactic Acid 2TMS	Organic acid	191.1	478	0.971	20.21	480	0.877	7.28	0.5	80	0.0312
Lauric acid TMS	Organic acid	257.2	955	0.903	34.18	955	0.938	37.16	0.5	50	0.0312
Malic acid 3TMS	Organic acid	233.1	832	0.98	14.04	832	**< 0.7**	11.45	0.5	80	
Succinate 2TMS	Organic acid	172.1	696	0.982	15.94	696	0.835	5.71	0.5	80	0.25
Cholesterol TMS	Sterol	247.25	1716	0.99	34.42	1717	0.954	26.18	0.5	60	0.5
Fructose 5TMS	Sugar	307.2	1093	0.994	12.6	1093	0.971	16.89	0.5	20	0.0312
Glucose 5TMS	Sugar	205.1	1121	0.982	24.61	1120	0.804	15.18	0.5	80	0.125
Sucrose 8TMS	Sugar	361.27	1495	0.989	13.5	1495	0.975	2.44	0.5	8	0.0312

Trehalose 8TMS	Sugar	361.27	1543	0.991	12.91	1543	0.985	2.07	0.5	10	0.0312
Myo-inositol 6TMS	Sugar alcohol	305	1223	0.947	14.19	1224	0.853	14.34	0.5	20	0.0312

CV = % coefficient of variation

*nmol in 0.5 mL of artificial seawater

** For selected compounds in artificial seawater

Table S3. Targeted techniques for the quantification of specific metabolite classes in seawater

Compound class	Sample pre-treatment	Detection method	Limit of detection	Amount of sample required	Ref
Amino acids	o-phthaldialdehyde/2-mercaptoethanol derivatization	High performance liquid chromatography	fM	25 µL	**1**
Sugars: Monosaccharides and disaccharides	Acid hydrolysis, desalting	High performance anion exchange chromatography-pulsed amperometric detector	nM	~12 mL	**2**
Sterols	Derivatization Solid phase extraction	GC-MS LC/ESI-MS/MS	nM pM	1 L 5 L	**3** **4**
Fatty acids	Solvent extraction	GC-MS	µM	90 L	**5, 6**
Organic acids	Derivatization with 2-nitrophenyl hydrazine and carbodiimine hydrocholoride	High pressure liquid chromatography with ion pairing	nM	2 mL	**7**

1 **Table S4.** Retention times, quantification and qualification ions of the 107 metabolite standards
2 detected in artificial seawater using SeaMet. Kyoto Encyclopedia of Genes and Genomes
3 (KEGG ID) ids and associated BRITE functional hierarchy or metabolite class of each
4 compound are indicated.

Metabolite standard	Manufacture company	KEGG ID	BRITE hierarchy/ General category	Retention time (s)	Quantification ion	Qualifying ions
β-D-allose	Biolog	NA	Aldohexose sugar	1111	205	189; 147
3-methyl glucose	Biolog	NA	Sugar	1086	204	217; 147
D-galactonic acid-γ-lactone	Biolog	NA	Sugar	1066	261	217; 160
Lactulose	Biolog	C07064	Sugar	1553	307	361; 217
N-acetyl-β-D-gannosamine	Biolog	NA	Sugar	1221	205	319; 274
α-methyl-D-galactoside	Biolog	NA	Sugar	1095	307	217; 103
β-methyl-D-glucoside	Biolog	NA	Sugar	1113	331	215; 129
β-methyl-D-glucuronic acid	Biolog	NA	Sugar	1173	292	305; 319
β-methyl-D-xyloside	Biolog	NA	Sugar	997	205	319; 147
D-arabinose	Biolog	C00216	Carbohydrates; Monosaccharides; Aldoses	962	307	217; 189
D-mannose	Biolog	C00159	Carbohydrates; Monosaccharides; Aldoses	1126	203	217; 301
D-ribose	Biolog	C00121	Carbohydrates; Monosaccharides; Aldoses	973	204	217; 147
D-xylose	Biolog	C00181	Carbohydrates; Monosaccharides; Aldoses	959	217	307; 103
Galactose	Sigma-Aldrich	C00124	Carbohydrates; Monosaccharides; Aldoses	1124	319	147; 205; 217
Glucose	AppliChem	C00031	Carbohydrates; Monosaccharides; Aldoses	1120	321	147; 160; 205
L-arabinose	Biolog	C00259	Carbohydrates; Monosaccharides; Aldoses	969	217	307; 277
L-lyxose	Biolog	C01508	Carbohydrates; Monosaccharides; Aldoses	956	217	307; 189
D-glucosamine	Biolog	C00329	Carbohydrates; Monosaccharides; Amino sugars	1141	217	307; 319
N-acetyl-D-galactosamine	Biolog	C01132	Carbohydrates; Monosaccharides; Amino sugars	1505	230	245; 217
N-acetyl-D-glucosamine	Biolog	C00140	Carbohydrates; Monosaccharides; Amino sugars	1234	319	205; 333
N-acetyl-D-glucosaminitol	Biolog	C00140	Carbohydrates; Monosaccharides; Amino sugars	1220	319	205; 202

Metabolite standard	Manufacture company	KEGG ID	BRITE hierarchy/ General category	Retention time (s)	Quantification ion	Qualifying ions
2-deoxy-D-ribose	Biolog	C01801	Carbohydrates; Monosaccharides; Deoxy sugars	898	174	250; 100
D-mannosamine	Biolog	C03570	Carbohydrates; Monosaccharides; Deoxy sugars	999	217	319; 205
L-fucose	Biolog	C01019	Carbohydrates; Monosaccharides; Deoxy sugars	1015	174	214; 200
L-rhamnose	Biolog	C00507	Carbohydrates; Monosaccharides; Deoxy sugars	1032	189	217; 147
D-psicose	Biolog	C06568	Carbohydrates; Monosaccharides; Ketoses	1097	307	217; 277
D-tagatose	Biolog	C00795	Carbohydrates; Monosaccharides; Ketoses	1110	319	217; 205
Fructose	Sigma-Aldrich	C00095	Carbohydrates; Monosaccharides; Ketoses	1098	307	147; 217; 277
L-sorbose	Biolog	C00247	Carbohydrates; Monosaccharides; Ketoses	1113	204	217; 147
Dulcitol	Biolog	C01697	Sugar alcohol	1145	217	252; 233
I-Erythritol	Biolog	C00503	Sugar alcohol	869	304	174; 147
D-arabitol	Biolog	C01904	Sugar alcohol	1001	205	307; 319
L-arabitol	Biolog	C00532	Sugar alcohol	1002	319	307; 205
Lactitol	Biolog	NA	Sugar alochol	1584	361	204. 217
Maltitol	Biolog	NA	Sugar alochol	1596	361	217; 204
Mannitol	Fluka	C00392	Carbohydrates; Monosaccharides; Sugar alcohols	1135	319	147; 205; 217
Myo-inositol	Fluka	C00137	Carbohydrates; Monosaccharides; Sugar alcohols	1224	305	147; 217; 305
Sorbitol	Merck	C00794	Carbohydrates; Monosaccharides; Sugar alcohols	1134	319	205; 217; 307
D-melibiose	Biolog	C05402	Carbohydrates; Oligosaccharides; Disaccharides	1764	363	469; 334
Gentiobiose	Biolog	C08240	Carbohydrates; Oligosaccharides; Disaccharides	1593	361	345; 217
Maltose	Sigma-Aldrich	C00208	Carbohydrates; Oligosaccharides; Disaccharides	1550	361	204; 217; 319
Palatinose	Biolog	C01742	Carbohydrates; Oligosaccharides; Disaccharides	1603	361	204; 217
Sucrose	Fluka	C00089	Carbohydrates; Oligosaccharides; Disaccharides	1495	361	73; 147; 217

Metabolite standard	Manufacture company	KEGG ID	BRITE hierarchy/ General category	Retention time (s)	Quantification ion	Qualifying ions
Trehalose	Fluka	C01083	Carbohydrates; Oligosaccharides; Disaccharides	1543	377	147; 191; 217
Turanose	Biolog	C19636	Carbohydrates; Oligosaccharides; Disaccharides	1561	361	345; 319
α-D-Lactose	Biolog	C00243	Carbohydrates; Oligosaccharides; Disaccharides	1530	361	217; 204
Lauric acid	Fluka	C002679	Lipids; Fatty acids; Saturated fatty acids	955	215	117; 201; 257
Capric acid	Biolog	C01571	Lipids; Fatty acids; Saturated fatty acids	823	247	359; 147
Sebacic acid	Biolog	C08277	Lipids; Fatty acyls; Fatty acids and conjugates; Dicarboxylic acids	1116	319	205; 217
Beta-sitosterol	Biolog	C01753	Lipids; Sterol lipids; Sterols	1176	333	292; 319
Cholesterol	Sigma-Aldrich	C00187	Lipids; Sterol lipids; Sterols	1717	247	129; 329; 443
Ergosterol	Biolog	C01694	Lipids; Sterol lipids; Sterols	1784	485	394; 255
Stigmasterol	Biolog	C05442	Lipids; Sterol lipids; Sterols	1814	357	396; 487
Thymine	Fluka	C00178	Nucleic acids; Bases; Pyrimidine	766	255	113; 147; 270
Uracil	Biolog	C00106	Nucleic acids; Bases; Pyrimidine	724	259	215; 147
Adenosine	Biolog	C00212	Nucleic acids; Nucleosides; Ribonucleosides	1527	361	204; 217
2-hydroxybenzoic acid	Biolog	C00805	Organic acid	737	146	103; 73
4-hydroxybenzoic acid	Biolog	C00156	Organic acid	941	292	219; 189
Citraconic acid	Biolog	C02226	Organic acid	733	195	177; 120
Citramalic acid	Biolog	C00851	Organic acid	827	158	260; 68
D-tartaric acid	Biolog	C02107	Organic acid	948	252	281; 296
D,L-α-amino-N-butyric acid	Biolog	NA	Organic acid	516	177	205; 161
Glycolic acid	Biolog	C00160	Organic acid	528	174	202; 116
m-hydroxyphenylacetic acid	Biolog	NA	Organic acid	939	174	318; 200
Mucic acid	Biolog	C00879	Organic acid	1217	378	319; 246
p-hydroxyphenylacetic Acid	Biolog	NA	Organic acid	949	217	147; 307
Sorbic acid	Biolog	NA	Organic acid	665	174	147; 100
α-hydroxybutyric acid	Biolog	NA	Organic acid	578	191	233; 147
α-hydroxyglutaric acid(-γ-lactone)	Biolog	NA	Organic acid	927	296	281; 164
δ-amino-N-valeric acid	Biolog	C00803	Organic acid; Carboxylic acid; Monocarboxylic acids/ ALT: Lipids; Fatty acyls; Fatty acid sand	941	267	223; 193

Metabolite standard	Manufacture company	KEGG ID	BRITE hierarchy/ General category	Retention time (s)	Quantification ion	Qualifying ions
			conjugates; Straight chain fatty acids			
Acetic acid		C00033	Organic acids; Carboxylic acids; Monocarboxylic acids	509	147	73; 137
Propionic acid		C00163	Organic acids; Carboxylic acids; Monocarboxylic acids	497	147	117; 191
Fumarate	Fluka	C00122	Dicarboxylic acid	727	245	143; 147; 217
Itaconic acid	Biolog	C00490	Dicarboxylic acid	729	147	259; 73
Maleic acid	Fluka	C01384	Dicarboxylic acid	692	245	73; 147; 215
Succinate	Merck	C00042	Organic acids; Carboxylic acids; Dicarboxylic acids	696	262	129; 147; 247
Lactic acid	Fluka	C00186	Organic acids; Carboxylic acids; Hydroxycarboxylic acids	480	193	117; 147; 191
Malic acid	Sigma-Aldrich	C00149	Organic acids; Carboxylic acids; Hydroxycarboxylic acids	832	233	147; 245; 307
β-hydroxybutyric acid	Biolog	C01089	Organic acids; Carboxylic acids; Hydroxycarboxylic acids	580	86	188; 75
N-butylamine	Biolog	C18706	Organic compound	555	205	190; 233
Urea	Fluka	C00086	Organic compound	644	189	73; 147; 171
D,L-octopamine	Biolog	C04227	Organic compound	1200	333	292; 305
Ethanolamine	Biolog	C00189	Peptides; Amines; Biogenic amines	690	219	130; 117
Phenylethyl-amine	Biolog	C05332	Peptides; Amines; Biogenic amines	899	247	203; 147
Putrescine	Biolog	C00134	Peptides; Amines; Biogenic amines	1017	117	133; 160
Alanine	Fluka	C00041	Peptides; Amino acids; Common amino acids	531	218	116; 233; 258
Aspartic acid	Fluka	C00049	Peptides; Amino acids; Common amino acids	858	232	147; 218; 292
Cysteine	Fluka	C00097	Peptides; Amino acids; Common amino acids	885	220	147; 204; 246
Glutamic acid	Merck	C00025	Peptides; Amino acids; Common amino acids	929	177	73; 147; 246
Glycine	Fluka	C00037	Peptides; Amino acids; Common amino acids	689	204	147; 174; 248
Isoleucine	Fluka	C00407	Peptides; Amino acids; Common amino acids	684	232	147; 158; 218
L-phenylalanine	Biolog	C00079	Peptides; Amino acids; Common amino acids	897	221	321; 147
Leucine	Biolog	C00123	Peptides; Amino acids; Common amino acids	598 660	169 177	125; 95 86; 146; 188
Lysine	Fluka	C00047	Peptides; Amino acids; Common amino acids	1127	174	218; 230; 317
Methionine	Sigma-Aldrich	C00073	Peptides; Amino acids; Common amino acids	779 860	221 176	104; 178; 206 128; 73

Metabolite standard	Manufacture company	KEGG ID	BRITE hierarchy/ General category	Retention time (s)	Quantification ion	Qualifying ions
Proline		C00148	Peptides; Amino acids; Common amino acids	689	216	142; 189; 244
Serine	Fluka	C00065	Peptides; Amino acids; Common amino acids	733	204	73; 116; 132
Threonine	Biolog	C00188	Peptides; Amino acids; Common amino acids	720 756	255 291	147; 219; 320
Valine	Sigma-Aldrich	C00183	Peptides; Amino acids; Common amino acids	614	188	130; 146; 174
Hydroxy-L-proline	Biolog	C01157	Peptides; Amino acids; Other amino acids	849	307	205; 217
L-homoserine	Biolog	C00263	Peptides; Amino acids; Other amino acids	816	229	117; 129
Ornithine	Merck	C00077	Peptides; Amino acids; Other amino acids	1061	162	73; 142; 174
γ-aminobutyric acid	Biolog	C00334	Peptides; Amino acids; Other amino acids	897	120	146; 91
Gly-Glu	Biolog	NA	Dipeptide	1193	267	174; 426

Table S5. Major metabolite peaks found in marine sediment porewaters and their retention times. Compounds that did not match NIST database enteries are labeled as "unknown" followed by their retention time. Mass spectra from these compounds were also compared to both the Golm and BinBase databases using the Kovats retention time index adjustment. Golm predicted functional groups and BinBase splash id's are reported when unknowns matched database hits.

Habitat	Annotation	Retention time (s)	Class	Unknown Compounds		
				Kovats retention index	Golm predicted functional groups	BinBase related splash ID
Coralline	Lactic acid	499	Organic acid			
	Acetate	512	Organic acid			
	Alanine	533	Amino acid			
	Glycerol	666	Sugar alcohol			
	Glycine	695	Amino acid			
	Succinate	699	Organic acid			
	Propanoic acid	713	Organic acid			
	Unknown 893	895		1563	No hits	No hits
	Unknown 1031	1031		1759	Primary alcohol Secondary alcohol Alcohol 1,2, diol Hydroxy	splash10-0fvj-0920000000-2b50a374508a5be92557
	Azelaic acid	1048	Fatty Acid			
	Unknown compound 1073	1073		1823	Hydroxy Alcohol Carboxylic acid	No hits
	Pinitol	1089	Sugar alcohol			
	Fructose	1101	Sugar			
	Galactose	1108	Sugar			
	Galactose	1112	Sugar			
	Mannitol	1138	Sugar alcohol			
	Sucrose	1503	Sugar			
	Trehalose	1553	Sugar			
Mangrove	Lactic acid	499	Organic acid			
	Acetate	512	Organic acid			
	Glycerol	666	Sugar alcohol			
	Succinate	699	Organic acid			
	Pentanoic acid	914	Organic acid			
	Lauric acid	937	Straight chain fatty acid			
	Lauric acid	956	Straight chain fatty acid			
	Unknown 1031	1031		1759	Primary alcohol Secondary alcohol Alcohol 1,2, diol Hydroxy	splash10-0fvj-0920000000-2b50a374508a5be92557
	Azelaic acid	1048	Fatty Acid			
	Aromatic dione	1126				
	Mannitol	1138	Sugar alcohol			
	Palmitic acid	1177	Straight chain fatty acid			
	Isooctyl laurate	1235	Straight chain fatty acid			
	Lauric acid	1289	Straight chain fatty acid			
	Myristic acid	1371	Straight chain fatty acid			
	Myristic acid	1388	Straight chain fatty acid			

Habitat	Annotation	Retention time (s)	Class	Unknown Compounds		
				Kovats retention index	Golm predicted functional groups	BinBase related splash ID
	Pentadecanoic acid	1434	Straight chain fatty acid			
	Hexadecanoic acid	1464	Straight chain fatty acid			
	Hexadecanoic acid	1481	Straight chain fatty acid			
	Sucrose	1503	Sugar			
	Heptadecanoic acid	1509	Straight chain fatty acid			
	Heptadecanoic acid	1524	Straight chain fatty acid			
	2-Monostearin	1552	Straight chain fatty acid			
	Octadecanoic acid	1569	Straight chain fatty acid			
	Unknown peak 1595	1595		2832	Alcohol	No hits
	Nonadecanoic acid	1607	Straight chain fatty acid			
	Unknown 1634	1635		2930	Alcohol Hydroxy Primary alcohol	No hits
	Eicosanoic acid	1647	Straight chain fatty acid			
	Unkonwn 1669	1672		3023	Alcohol Hydroxy Primary alcohol	No hits
	Unknown Compound 1720	1721		3151	Hydroxy Alcohol Carboxylic acid Primary alcohol	splash10-05o0-1910000000-49e59c211689b393e2a6
	Unknown compound 1732	1734		3185	Alcohol Hydroxy Primary alcohol	No hits
	Unknown compound 1746	1746		3215	Alcohol Hydroxy Primary alcohol	No hits
	Unknown 2005	2005		3691	Alcohol Hydroxy Alkene	No hits
	Unknown 2030	2035		NA	Hydroxy	No hits

1 **Table S6.** Metabolite mixtures used across experiments. All mixtures contain a diverse range of
2 compounds representing multiple metabolite classes. A reduced set of compounds were
3 combined to both show the effects of salt and water on metabolite detection and create
4 calibration curves for specific compounds.

Compound	Class	Experiment Mixture
Pyruvate	Alpha keto-acid	Method development and SPE
Alanine	Amino acid	Method development and SPE
Arginine	Amino acid	Method development and SPE
Aspartic acid	Amino acid	Method development and SPE
Carnitine	Amino acid	Method development and SPE
Cysteine	Amino acid	Method development and SPE
Glutamate	Amino acid	Method development and SPE
Glutamic acid	Amino acid	Method development and SPE
Glutamine	Amino acid	Method development and SPE
Glycine	Amino acid	Method development and SPE
Histidine	Amino acid	Method development and SPE
Isoleucine	Amino acid	Method development and SPE
Lactic acid	Amino acid	Method development and SPE
Leucine	Amino acid	Method development and SPE
Lysine	Amino acid	Method development and SPE
Methionine	Amino acid	Method development and SPE
Ornithine	Amino acid	Method development and SPE
Phenylalanine	Amino acid	Method development and SPE
Proline	Amino acid	Method development and SPE
Serine	Amino acid	Method development and SPE
Threonine	Amino acid	Method development and SPE
Thymine	Amino acid	Method development and SPE
Tryptophan	Amino acid	Method development and SPE
Valine	Amino acid	Method development and SPE
Citric acid	Organic acid	Method development and SPE
Fumerate	Organic acid	Method development and SPE
Lauric acid	Organic acid	Method development and SPE
Maleic acid	Organic acid	Method development and SPE
Malic acid	Organic acid	Method development and SPE
Succinic acid	Organic acid	Method development and SPE
DMSP	Organic compound	Method development and SPE
Oxalic Acid	Organic compound	Method development and SPE
Urea	Organic compound	Method development and SPE
Cellobiose	Sugar	Method development and SPE
Fructose	Sugar	Method development and SPE
Galactose	Sugar	Method development and SPE
Glucose	Sugar	Method development and SPE
Maltose	Sugar	Method development and SPE
Mannose	Sugar	Method development and SPE
NAG	Sugar	Method development and SPE
Ribose	Sugar	Method development and SPE

Sucrose	Sugar	Method development and SPE
Trehalose	Sugar	Method development and SPE
Mannitol	Sugar alcohol	Method development and SPE
Myo-inositol	Sugar alcohol	Method development and SPE
Glycerophosphate	Sugar phosphate	Method development and SPE
Alanine	Amino acid	Quantifying salt-water effect
Glutamine	Amino acid	Quantifying salt-water effect
Glycine	Amino acid	Quantifying salt-water effect
Lactate	Amino acid	Quantifying salt-water effect
Leucine	Amino acid	Quantifying salt-water effect
Lysine	Amino acid	Quantifying salt-water effect
Ornithine	Amino acid	Quantifying salt-water effect
Serine	Amino acid	Quantifying salt-water effect
Succinate	Amino acid	Quantifying salt-water effect
Thymine	Amino acid	Quantifying salt-water effect
Valine	Amino acid	Quantifying salt-water effect
Citric acid	Organic acid	Quantifying salt-water effect
Glutamic acid	Organic acid	Quantifying salt-water effect
Malic acid	Organic acid	Quantifying salt-water effect
Oxalic acid	Organic acid	Quantifying salt-water effect
Fructose	Sugar	Quantifying salt-water effect
Glucose	Sugar	Quantifying salt-water effect
NAG	Sugar	Quantifying salt-water effect
Sucrose	Sugar	Quantifying salt-water effect
Trehalose	Sugar	Quantifying salt-water effect
Myo-inositol	Sugar alcohol	Quantifying salt-water effect
Glycine	Amino acid	Quantify detection limits
Histidine	Amino acid	Quantify detection limits
Leucine	Amino acid	Quantify detection limits
Lysine	Amino acid	Quantify detection limits
Ornithine	Amino acid	Quantify detection limits
Serine	Amino acid	Quantify detection limits
Thymine	Amino acid	Quantify detection limits
Valine	Amino acid	Quantify detection limits
Citate/citric acid	Organic acid	Quantify detection limits
Glutamic acid	Organic acid	Quantify detection limits
Lactic Acid	Organic acid	Quantify detection limits
Lauric acid	Organic acid	Quantify detection limits
Malic acid	Organic acid	Quantify detection limits
Succinate	Amino acid	Quantify detection limits
Cholesterol	Sterol	Quantify detection limits
Fructose	Sugar	Quantify detection limits
Glucose	Sugar	Quantify detection limits
Sucrose	Sugar	Quantify detection limits
Trehalose	Sugar	Quantify detection limits
Myo-inositol	Sugar alcohol	Quantify detection limits

Supplementary Text 1. R script for peak picking for GC-MS data

```
# Peak Picking.R
# EM Sogin
# Description: R script to pick peaks from GC-MS data

library(xcms)
library(CAMERA)

## PEAK PICKING, RETENTION TIME GROUPING & CORRECTION WITH XCMS
setwd('home/path/to/files')
files<-list.files(pattern='.mzXML', recursive = T, full.names=T)
xs <- xcmsSet(files, method = "matchedFilter", fwhm = 8.4, snthresh = 1,step= 0.25, steps= 2,sigma =
3.56718192627824, max= 500, mzdiff= 1,index= FALSE)
xset1 <- group(xs,method  = "density", bw=2, mzwid= 1, minfrac = 0.3,  minsamp = 1,max = 500) ##
Initial peak grouping
xset2 <- retcor(xset1)
xset2 <- group(xset2,method  = "density", bw=2, mzwid= 1, minfrac = 0.3,  minsamp = 1,max = 500)
xset<-fillPeaks(xset2)

## Group peaks in to pseudo-spectra using CAMERA
an<-xsAnnotate(xset)
xsF<-groupFWHM(an, perfwhm=3)

peaks<-getPeaklist(xsF)
peaks[is.na(peaks)]<-0

save.image('Peak_Picking_Results.RData')
# End
```

Supplementary references

1. Lindroth P & Mopper K (1979) High performance liquid chromatographic determination of subpicomole amounts of amino acids by precolumn fluorescence derivatization with o-Phthaldialdehyde. *Anal Chem* 51(11):1667-1674.
2. Engel A & Händel N (2011) A novel protocol for determining the concentration and composition of sugars in particulate and in high molecular weight dissolved organic matter (HMW-DOM) in seawater. *Mar Chem*127(1-4):180-191.
3. Rocha MJ, Cruzeiro C, & Rocha E (2013) Development and validation of a GC-MS method for the evaluation of 17 endocrine disruptor compounds, including phytoestrogens and sitosterol, in coastal waters - their spatial and seasonal levels in Porto costal region (Portugal). *J Water Health* 11(2):281-296.
4. Ronan JM & McHugh B (2013) A sensitive liquid chromatography/tandem mass spectrometry method for the determination of natural and synthetic steroid estrogens in seawater and marine biota, with a focus on proposed Water Framework Directive Environmental Quality Standards. *Rapid Commun Mass Spectrom* 27(7):738-746.
5. Kennicutt MC & Jeffrey LM (1981) Chemical and GC-MS characterization of marine dissolved lipids. *Mar Chem* 10:367-387.
6. Mannino A & Harvey HR (1999) Lipid composition in particulate and dissolved organic matter in the Delaware Estuary: Sources and diagenetic patterns. *Geochimica et Cosmochimica Acta* 63(15):2219-2235.
7. Albert DB & Martens CS (1997) Determination of low-molecular-weight organic acid concentrations in seawater and pore-water samples via HPLC *Mar Chem* 56:27-37.