Revealing the ocean metabolome with mass spectrometry

Classification: BIOLOGICAL SCIENCES – environmental sciences

E. Maggie 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

Significance

Metabolites are the chemical currency of cellular metabolism across all domains of life. However, describing metabolites that occur in the oceans is lagging behind similar studies conducted on land. The central challenge in marine metabolomics is that salt prevents the comprehensive analysis of metabolites in seawater. We developed a method, SeaMet, that overcomes the limitations of salt on metabolite detection. SeaMet provides a time and cost efficient method, using gas chromatography-mass spectrometry, for the reproducible identification and quantification of a broad range of marine compounds. Considering the oceans contain the largest organic carbon pool on Earth, describing the marine metabolome is critical for understanding the drivers behind element cycles, biotic interactions, ecosystem function, and atmospheric CO_2 storage.

Abstract

All life exchanges molecules with its environment. While these metabolites are commonly measured in terrestrial and limnic ecosystems, the presence of salt in marine habitats has hampered quantitative analyses of the ocean metabolome. To overcome these limitations, we developed SeaMet, a gas chromatography-mass spectrometry (GC-MS) method that detects hundreds of metabolites down to nano-molar concentrations in less than one milliliter of seawater. Using a set of metabolites dissolved in artificial seawater to benchmark our method, we show metabolite signal detection increased on average across ions by 324 fold in comparison to standard GC-MS methods. Our observed signal improvement occurred across tested metabolite classes and provides reproducible and quantifiable results. To showcase the capabilities of our method, we used SeaMet to explore the production and consumption of metabolites during culture of a heterotrophic bacteria 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 metabolomics in providing insights into nutrient uptake and energy conservation in marine microorganisms. We also 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 a remarkable diversity and abundance of sugars and fatty acids, compounds predicted to be rapidly consumed by marine microorganisms. Our method advances marine metabolomics by enabling the unbiased, and quantitative analysis of marine metabolites, and will help provide new insights into carbon cycle dynamics and ocean biogeochemistry.

Keywords: Marine; Metabolomics; Microbiology; GC-MS

Gas chromatography-mass spectrometry (GC-MS) analysis enables the detection of primary metabolites, small molecules that occur in central metabolic pathways across biological systems[1, 2]. High reproducibility coupled to the widespread availability of annotation resources make GC-MS the "workhorse" of analytical chemistry facilities. GC-MS has allowed the identification of metabolites associated with human disease[3], detection of compounds that serve as environmental cues in foraging[4], and description of metabolic fluxes within and between cells[5]. GC-MS is also widely used for environmental profiling of soils and microbial activity on land[6, 7].

In contrast to terrestrial or limnic systems, the defining characteristic of marine habitats - high salt concentration - inhibits MS-based metabolomic analysis. Consequently, our knowledge of metabolite composition of the oceans is severely limited. This, in turn, restricts our understanding of global biogeochemical cycles, as the oceans contain the largest organic carbon pool on Earth[8]. While metagenomic and metatranscriptomic studies of the ocean, driven by low sequencing costs and projects like Tara Oceans[9], have deepened our knowledge of the identity and activity of marine microbes, current metabolomic techniques do not permit equivalent surveys of the ocean metabolome.

The ocean metabolome remains largely undefined, despite a growing field of research exploring the molecular composition of dissolved organic matter [8, 10, 11, 12]. Current efforts rely on solid phase extraction (SPE) methods to remove salt prior to MS analyses[10, 11]. SPE successful removal of salt from marine samples is accopanied by co-removal of small polar compounds, which are the primary components of the liable organic matter pool (**Fig. 1A**)[13]. Consequently, while SPE-based studies can measure up to 50% of the organic carbon from the ocean, they fail to detect the majority of compounds involved in the central metabolism of cells. Furthermore, current DOM analytical approaches remain largely inaccessible for the majority of research institutions and projects. This is largely due to high instrumentation costs, large sample volume requirements, and the long amount of time needed for metabolome analysis (**Fig. 1B)**. To more efficiently decipher ocean metabolism, it is necessary to develop cost-effective, high-throughput,

and untargeted workflows that can readily identify and quantify molecules from high salinity environments. Here, we present SeaMet, a marine metabolomics method that uses sample derivatization to enable metabolite detection in seawater using GC-MS.

Results and Discussion

SeaMet modifies the well-established two-step metabolite derivatization procedure, which permits the detection of non-volatile compounds using GC-MS, and involves methoximation followed by trimethylsilylation[14]. Like other GC-MS sample preparation techniques[15, 16], SeaMet removes liquid through vacuum drying prior to derivatization - a process that results in a salt pellet when working with marine samples, which restricts MS analysis. Our initial results suggested that water locked within the dried salt crystals hindered the chemical reactions needed for GC-MS (***SI Appendix* Fig. S1A, B**). Our method overcomes this limitation by first eliminating residual water within the salt crystals and then extracting small compounds into the derivatization reagents (**Fig. 1A**).

A promising GC-MS method to advance marine metabolomics research

We used a synthetic mixture of 45 metabolites (**Table S1**) dissolved in artificial seawater 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. 2B**). We first replaced the most commonly used trimethylsilylation reagent, N-methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA)[16], with one that is less susceptible to inhibition by water, N, O-Bistrifluoroacetamide (BSTFA), which resulted in higher metabolite signals (***SI Appendix* Fig. S1B**). To eliminate water from the samples, we increased the speed-vacuum drying time from four to eight hours, and integrated a toluene drying step that is used in urine-based metabolomic analyses[15]. We further enhanced metabolite signals by treating the salt pellet to a combination of ultrasonication and vortexing after the additions of toluene and derivatization reagents, and following completion of the

trimethylsilylation reaction. These steps break apart the salt crystals and release water into the toluene to enhance salt drying and metabolite extraction. Finally, following a recently described method for improving metabolite detection regardless of sample type[17], we dried the samples between the methoximation and trimethylsilylation derivatization reactions under N_2 gas (see **Fig. 2B** for signal improvements of each step).

Overall, SeaMet allowed us to detect significant increases in metabolite abundances across molecular classes when compared to the standard method (adjusted *p*-value < 0.05; mean fold change across all ions = 323; **Fig. 3A**). This included measurement of organic acids, amino acids, and fatty acids, as well as sugars (and their stereoisomers), sugar alcohols, and sterols (**Table S1**). Our detection limits in the nano-molar range are comparable to those of targeted techniques for marine ecosystems that were developed to quantify single compounds from specific molecular classes (**Tables S2 and S3**). Moreover, our workflow only requires 0.5 mL to 1 mL of seawater for metabolite detection, at least an order of magnitude less sample volumes than previously published techniques[13, 18]. To date, we have successfully measured 107 metabolite standards dissolved in artificial seawater, representing major metabolite groups involved in primary metabolic pathways (**Table S4**), highlighting that SeaMet can be used for both targeted and untargeted applications. Our method provides reproducible quantification and gives similar linearity, dynamic range, and coefficient of variation values as salt-free samples prepared with the standard GC-MS derivatization procedure (**Table S2**; ***SI Appendix* Fig. S2A-H**).

To demonstrate that our method captures compounds commonly missed by SPE-based metabolomic approaches, we compared GC-MS profiles using SeaMet before and after SPE sample treatment. Our method detected small polar compounds that are co-removed with salt during SPE preparation, such as sugars, sugar alcohols, amino acids, and organic acids (**Fig. 3B**). Given our results, SeaMet is complementary to current marine DOM characterization approaches that require SPE preparation and high-resolution MS for metabolite detection. Thus, SeaMet expands the possibilities for metabolite

analysis in the ocean and advances marine metabolomics by recovering compounds that previously could not be analyzed by current platforms in a single analysis.

SeaMet can monitor cell culture physiology

To demonstrate the power of our approach in characterizing metabolite profiles of marine bacteria, we monitored changes in the exo-cellular metabolome during growth of a heterotrophic Proteobacterium, *Marinobacter adhaerens*. Using SeaMet, we observed significant changes in the metabolite composition of marine culture medium during the bacteria's initial growth phase (adjusted p-value < 0.05; **Fig. 4A; *SI Appendix* Fig. S3A, B).** The bacteria took up different carbon and nitrogen resources in a cascade-like fashion**,** and later in growth, began excretion of an undescribed compound (**Fig. 4A, B; *SI Appendix* Fig. S4**). Successional dynamics in substrate use is a common energy conservation mechanism in bacteria[19] and affects central carbon and nitrogen dynamics during growth. With SeaMet, we show that like many bacteria, *M. adhaerens* participates in the release of organic carbon, which can be metabolized by other microorganisms or will contribute to the complexity of refractory DOM. By identifying and quantifying metabolites that are consumed and excreted in cultivable marine bacteria, our method expands our understanding of key primary compounds involved in the transformation of organic matter in the ocean.

SeaMet is a powerful tool for exploring ocean chemistry

To test the ability of our workflow to assess complex environmental metabolomes, we applied SeaMet to samples from coralline and mangrove sediment porewaters. Coral reefs and mangroves, two globally important coastal ecosystems, contain many biological compounds that remain undescribed. It is essential to characterize the metabolome of these habitats to understand the role of these ecosystems in biogeochemical cycling. Our approach detected 295 and 428 metabolite peaks from coralline and mangrove sediment profiles (**Fig. 5 A, B**), including sugars and amino acids, organic acids, fatty acids, and background signals. Diverse and abundant sugars from coralline sediments as well as fatty acids from mangroves drove the observed significant differences between habitats (ADONIS p-value < 0.001, R^2 =

0.514; ***SI Appendix* Fig. S5 A, B and Table S5**). Considering corals and mangroves thrive in oligotrophic waters and their associated sediments harbor diverse and abundant microorganisms, we were surprised to measure high abundances of energy-rich metabolites that are typically consumed in primary 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.

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[20]. 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 (**Table S5)**[21, 22].

Conclusions

We have presented SeaMet, 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.

Materials and methods

Data availability. All metabolite profile data will be made publicly available at Metabolights (https://www.ebi.ac.uk/metabolights/) under identification numbers MTBLS826, MTBLS839, MTBLS843, MTBLS844, MTBLS848, and MTBLS849 (currently IN REVIEW) or by contact with the authors. Reviewer links:

https://www.ebi.ac.uk/metabolights/reviewer5eb6b480436b019d9f1351a828ee7c3d
https://www.ebi.ac.uk/metabolights/reviewerd923ea1c3a53d000b97ccf383991032d
https://www.ebi.ac.uk/metabolights/reviewera9be9cf4-9a7d-4fff-98d5-c3d574c3b7f5
https://www.ebi.ac.uk/metabolights/reviewer08ce2d89-3945-45be-8a9b-4ea872fc86bf
https://www.ebi.ac.uk/metabolights/reviewer07a4ce73-1e8e-46aa-80c8-0c2f26411174
https://www.ebi.ac.uk/metabolights/reviewer2878413e6f8a6a883b27bdee8c1bbba6

Reagents and experimental sample preparation. The derivatization chemicals, trimethylsilyl-N-methyl trifluoroacetamide (MSTFA) and N,O-Bis(trimethylsilyl)trifluoroacetamide (BSTFA) were obtained from CS-Chromatographie Service and pyridine from Sigma-Aldrich at >99.98% purity. Methoxyamine hydrochloride (MeOX; Sigma-Aldrich) aliquots were further dried at 60 °C in a drying oven for 1 h to remove residual moisture. Artificial seawater (ASW) was prepared within the range of natural salinity (36‰) by dissolving (per L of water) 26.37 g sodium chloride, 6.8 g magnesium sulfate heptahydrate, 5.67 g magnesium chloride hexahydrate, 1.47 g calcium chloride, 0.6 g potassium chloride, and 0.09 g potassium bromide. Following autoclave sterilization, pH was adjusted to 7.7 using sodium hydroxide. 1 mL of the following supplements and solutions were added: 150 mM monopotassium phosphate, 500 mM ammonium chloride pH 7.5, trace element solution, selenite-tungstate solution, vitamin solution, thiamine solution, B12 solution and 0.21 g sodium bicarbonate[23]. Ultra-pure water (MQ) was prepared by purifying deionized water with an Astacus membraPure system (Astacus membraPure, 18.3 mΩ × cm 25 °C).

Metabolite standards were obtained from commercial sources (**Table S4**) and combined into mixtures in which each compound had a final concentration of 0.4 mM. Metabolite mixtures were prepared to (a) test the effect of salt and water on metabolite detection, (b) develop SeaMet, our marine metabolomics workflow, (c) compare metabolite detection before and after solid phase extraction (SPE) based sample preparation, and (d) to quantify the detection limits of specific compound classes (**Table**

S6). Finally, multiple mixtures were prepared to document the retention times of 107 standards dissolved in ASW using SeaMet (**Table S4**). Sample aliquots for the above mentioned experiments were prepared by drying down 200 µL of the mixture in a speed vacuum concentrator (Eppendorf Concentrator Plus$^{(R)}$, 2.5 h, 45°C, V-AQ) for all experiments except SPE comparison and quantification of detection limits. For the SPE comparison experiment, 400 µL of the mix were dried down. For the quantification of metabolite classes, a serial dilution of the mix was prepared to obtain concentrations between 0.5 nmol and 80 nmol of each compound. All dried mixture samples were stored at 4 °C.

SeaMet metabolite derivatization. To prepare marine samples for gas chromatography-mass spectrometry (GC-MS) analysis, 0.5 to 1 mL of a saltwater sample or experimental mixture dissolved in ASW was dried in a speed vacuum concentrator for 8 hours (Eppendorf Concentrator Plus$^{(R)}$, 45°C, V-AQ). To further remove residual water locked within the salt pellet, 250 µL of toluene (99.8%, < 0.2 % water) was added to each sample and the mixture was ultrasonicated for 10 min at maximum intensity. The toluene was subsequently removed under a gentle flow of N_2 gas. Metabolite derivatization was performed by adding 80 µL of MeOX dissolved in pyridine (20 mg × mL^{-1}) to the dried pellet. The mixture was ultrasonicated (EMag Emmi-12HC®) for 10 min at maximum intensity, briefly vortexed to dissolve the pellet into solution, and subsequently incubated for 90 min at 37 °C using a thermal rotating incubator under constant rotation at 1350 rpm. The pyridine was removed from the sample at room temperature under a gentle flow of N_2 gas (approximately 1 hour). Following the addition of 100 µL of BSTFA, the mixture was ultrasonicated for 10 min at maximum intensity, vortexed, and incubated for 30 min at 37 °C using a thermal rotating incubator under constant rotation at 1350 rpm. The derivatized mixture was ultrasonicated for 10 min at maximum intensity. Remaining salt in each sample was pelleted through centrifugation at 21.1 *g* for 2 min at 4 °C. 100 µL was transferred to a GC-MS vial for analysis. The full proposed method is publicly available at dx.doi.org/10.17504/protocols.io.nyxdfxn.

GC-MS data acquisition. All derivatized samples were analyzed on an Agilent 7890B GC coupled to an Agilent 5977A single quadrupole mass selective detector. Using an Agilent 7693 autosampler, 1 µL was injected in splitless mode through a GC inlet liner (ultra inert, splitless, single taper, glass wool, Agilent) onto a DB-5MS column (30 m × 0.25 mm, film thickness 0.25 µm; including 10 m DuraGuard column, Agilent). The inlet liner was changed every 50 samples to avoid damage to the GC column and associated shifts in retention times. The injector temperature was set at 290 °C. Chromatography was achieved with an initial column oven temperature set at 60 °C followed by a ramp of 20 °C min^{-1} until 325 °C, then held for 2 mins. Helium carrier gas was used at a constant flow rate of 1 mL min^{-1}. Mass spectra were acquired in electron ionization mode at 70 eV across the mass range of 50–600 m/z and a scan rate of 2 scans s^{-1}. The retention time for the method locked using standard mixture of fatty acid methyl esters (Sigma Aldrich).

Data processing and analysis. Raw Agilent data files were converted to mzXML files using Msconvert[24] and imported into XCMS (v. 2.99.6)[25] within the R software environment (v. 3.4.2) for data processing and analysis. Total ion chromatograms (TIC) were obtained using the xcmsRaw function. TICs comparing sample preparation steps were expressed as a percentage of the MQ control. For environmental and cell culture GC-MS profiles, peaks were picked using the matchedFilter algorithm in XCMS with a full width at half maximum set to 8.4, signal to noise threshold at 1, *m*/*z* width of 0.25 (step parameter), and *m*/*z* difference between overlapping peaks at 1 (***SI Appendix*** **Text 1**). Resulting peaks were grouped, retention times corrected and regrouped using the density (bandwidth parameter set to 2) and obiwarp methods. Following peak filling, the CAMERA (v.1.32.0)[26] package was used to place *m*/*z* peaks into pseudo-spectra by grouping similar peaks with the groupFWHM function. Masses below 150 *m*/*z* were removed from the resulting peak table and all profiles were normalized to the ribitol internal standard. Peaks occurring in run blanks and those with higher relative standard deviation scores (%RSD > 25) in quality control samples (cell culture experiment only) were removed from the dataset. To determine differences in metabolite abundances between sediment habitats, metabolite peak data were

analyzed using a Bray-Curtis informed non-metric multidimensional scaling analysis followed by an analysis of variance using distance matrices (ADONIS) to test if there are significant differences in metabolite composition between sites. To identify individual peaks that differed significantly between sediment habitats and between cell culture sampling time points, resulting peaks tables were also log transformed and compared using a one-way analysis of variance. All p-values were adjusted using the Benjamini-Hochberg method to control for false positives[27]. Significant variables exhibiting large fold-change differences between starting and ending conditions were further investigated. CAMERA grouped peaks from the environmental survey, and those important to shifts in the cell culture experiment were identified using AMDIS[28]. Peaks with NIST hits below 800 were compared to online BinVestigate[29] and Golm[21] data repositories. If no hit was provided, these were considered unknowns.

The effect of salt and water on metabolite detection. To test the effect of salt on metabolite derivatization, metabolite mix aliquots were resuspended in 1 mL of ASW ranging in salinity from 0 to 34‰ and dried as described above. Methoxamine-trimethylsilylation (TMS) two step derivatization was performed by resuspending each sample in 80 µL of MeOX in pyridine (20 mg mL^{-1}) and incubating for 90 min at 37 °C using a thermal rotating incubator under constant rotation at 1350 rpm. MSTFA was subsequently added to the mixture, and the mixture incubated under the same conditions for 90 min[14]. Derivatized samples were centrifuged to pellet salt and the supernatant was transferred to a GC-MS vial for analysis. To test the independent effect of water on metabolite derivatization reactions, MQ was added to dried mixture aliquots in steps of 1 µL from 0 to 10 µL. Replicate water gradient samples were subsequently derivatized as before using MeOX and MSTFA or by replacing the MSTFA reagent with BSTFA.

Marine metabolomics method development. To show how each method development step increased signal intensity and reduced variation in metabolite detection, replicate mixture aliquots (n = 5) were resuspended in 0.5 mL of ASW. Mixture aliquots (n = 5) were also resuspended in MQ as a no-salt

control to highlight the effects of saltwater on metabolite derivatization. 40 µL ribitol (0.2 mM) and 100 µL cholestane (1 mM) were added to each aliquot as internal standards. MQ and ASW samples were first derivatized following the (i) two-step methoxamine-trimethylsilylation (TMS) previously described. Successive steps in the proposed protocol were then applied to ASW samples to demonstrate the combined effects on metabolite detection: (ii) exchange of MSTFA for BSTFA, (iii) removal of residual water from the salt pellet by increasing the speed vacuum drying time and by introducing a toluene drying step to help extract water from the salt pellet, (iv) ultrasonication of the samples after the steps involving addition of toluene, MeOX, BSTFA and following the last derivatization step, and (v) drying the MeOX in pyridine reagent between derivatization reactions. Resulting GC-MS profiles were used to show increases in total signals detected with successive changes in the proposed protocol. Additionally, a cloud plot (using processed peak integration data) was generated to compare compounds dissolved in seawater and to show which metabolite ions exhibited significant (B.H. adjusted $p < 0.05$) and large fold changes (log2(FC) > 2) between the standard and the SeaMet method.

Solid phase extraction. Replicate metabolite mix aliquots ($n = 6$) were resuspended in 2 mL of artificial seawater. 0.5 mL was reserved from each sample to compare GC-MS profiles before and after SPE sample concentration. Inorganic salts were eluted and metabolites extracted from the remaining 1.5 mL mixture following a SPE based technique using Bond Elut styrene-divinylbenzene (PPL) columns[13]. The internal standards ribitol and cholestane were added to both, the reserved sample (before SPE) and the resulting SPE-concentrated sample (after SPE). All samples were prepared for GC-MS analysis following the proposed marine metabolomics method. Resulting profiles were compared using a cloud plot to show which metabolite ions exhibited significant (B.H. adjusted $p < 0.05$) and large fold changes (log2(FC) > 2) between the pre- and post- SPE treatments.

Environmental sampling. Replicate porewater profiles were collected from coralline ($n = 4$) and mangrove ($n = 6$) sediments from Carrie Bow Cay (N 16° 04’ 59”, W 88° 04’ 55”) and Twin Cayes,

Belize (N 16° 50’ 3”, W 88° 6’ 23”) using a 1 m steel lance with a 2 µm inner diameter covered by 0.063 mm steel mesh. Samples (2 mL water) were collected every 5 cm from the sediment surface to 15 cm depth. Samples were immediately frozen at -20 °C until further analysis. Directly before preparation for GC-MS, the internal standards ribitol and cholestane were added to 0.5 mL of each environmental sample. The mixture was subsequently prepared for GC-MS analysis using the SeaMet method described above.

Cell culture sampling. Replicate cultures (*n* = 3) *of Marinobacter adhaerens* HP15 DsRed were cultivated in Marine Broth media at 18 °C and 240 rpm as previously described[30]. Media samples from the cell cultures and a no-bacteria control media were collected at 0, 4, 8, 12, and 22 h post culture inoculation. Cell counts were monitored at each time point by measuring the optical density at 600 nm (OD_{600}). Sampling was carried out by collecting 2 mL of each culture and pelleting the cells through centrifugation for 10 min, at 21.1 *g* and 4 °C. The supernatant was immediately stored at -20 °C until preparation for GC-MS analysis. Prior to sample derivatization using SeaMet, ribitol (0.2 mM; 40 µL) and cholestane (100 mM; 100 µL) were 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.

Acknowledgements

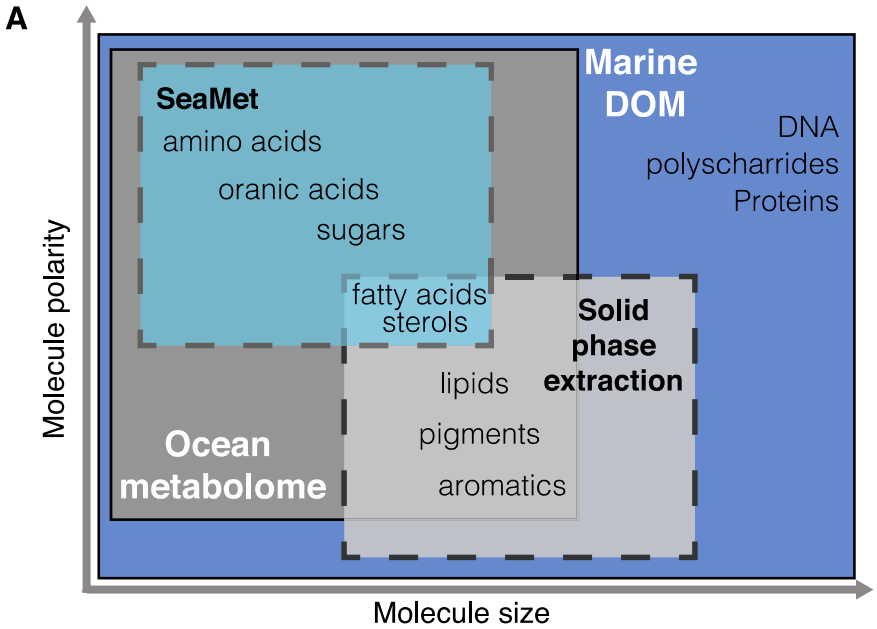
We thank T. Gulstad, K. Caspersen, F. Fojt, and M. Meyer (MPI-Bremen) for support with data acquisition and sample preparation, N. Böttcher (Jacobs University Bremen) for providing culture samples, D. Michellod for sediment pore water sample collection, and B. Geier and J. Beckmann (MPI-Bremen) for valuable 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. A J, *et al.* (2005) Extraction and GC/MS analysis of the human blood plasma metabolome. *Anal Chem* 77(24):8086-8094.
2. Fiehn O, *et al.* (2000) Metabolite profiling for plant functional genomics. *Nat Biotechnol* 18:1157-1161.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. Moran MA, *et al.* (2016) Deciphering ocean carbon in a changing world. *Proc Natl Acad Sci U S A* 113(12):3143-3151.
9. Sunagawa S, *et al.* (2015) Structure and function of the global ocean microbiome. *Science* 384(6237).
10. 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.
11. 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
12. Lechtenfeld OJ, Hertkorn N, Shen Y, Witt M, & Benner R (2015) Marine sequestration of carbon in bacteria metabolites. *Nat Comm* 6(6711).
13. 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.
14. 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.
15. Chan EC, Pasikanti KK, & Nicholson JK (2011) Global urinary metabolic profiling procedures using gas chromatography-mass spectrometry. *Nat Protoc* 6(10):1483-1499.
16. 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.
17. Liebeke M & Ppuskas E (2019) Drying enhances signal intensity for global GC-MS metabolomics. BioaRxiv:doi.org/10.1101/544916. Preprint, posted February 8, 2019.
18. 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.

19. Behrends V, Ebbels TM, Williams HD, & Bundy JG (2009) Time-resolved metabolic footprinting for nonlinear modeling of bacterial substrate utilization. *Appl Environ Microbiol* 75(8):2453-2463.
20. Luzzatto-Knaan T, *et al.* (2017) Digitizing mass spectrometry data to explore the chemical diversity and distribution of marine cyanobacteria and algae. *Elife* 6. DOI:10.7554/eLife.24214
21. Kopka J, *et al.* (2005) GMD@CSB.DB: the Golm metabolome database. *Bioinformatics* 21(8):1635-1638.
22. Lai Z, *et al.* (2018) identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nat Methods* 15(1):53-56.
23. Widdel F & Bak F (1992) Gram-negative mesophilic sulfate-reducing bacteria. *The Prokaryotes: A Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Identification, Applications*, eds Balows A, Trüper HG, Dworkin M, Harder W, & Schleifer KH (Springer, New York, NY).
24. Holman JD, Tabb DL, & Mallick P (2014) Employing proteoWizard to convert raw mass spectrometry data. *Curr Protoc Bioinformatics* 46:13 24 11-19.
25. Smith CA, Want EJ, O'Maille G, Abagyan R, & Siuzdak G (2006) XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 78(3):779-787.
26. Kuhl C, Tautenhahn R, Bottcher C, Larson TR, & Neumann S (2012) CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal Chem* 84(1):283-289.
27. Benjamini Y & Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J Royal Stat Soc* 57:289-300.
28. Stein SE (1999) An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J Am Soc Mass Spectrom* 10:770-781.
29. Lai Z, *et al.* (2018) identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nat Methods* 15(1):53-56.
30. Stahl A & Ullrich MS (2016) Proteomics analysis of the response of the marine bacterium Marinobacter adhaerens HP15 to the diatom Thalassiosira weissflogii. *Aquatic Microbial Ecology* 78(2):65-79.



B

	GC-MS	LC-MS	FT-ICR-MS
Sample preparation	*SeaMet*	*SPE*	*SPE*
% of peaks identified	+++++	+++	+
Quantification	+++++	+++	+
Cost	$	$$$	$$$$$
Time invested			
Sample volume			

Figure 1. SeaMet fills a gap in ocean metabolomics. A. SeaMet is a sample preparation method complementary to solid phase extraction (SPE) and can characterize the liable fraction of the ocean metabolome. The ocean metabolome is captured within marine dissolved organic matter (DOM) and is represented by compounds with varying molecule size and polarity. **B.** Gas chromatography-mass spectrometry is more accessible then other techniques used to investigate the ocean metabolome, such as liquid chromatography(LC)-MS and Fourier-transform ion cyclotron resonance (FT-ICR) MS. GC-MS can identify a higher percentage of peaks within marine samples, can more efficiently quantify molecule abundance, is cheaper in relation to both cost and time invested for metabolite analysis, and requires less sample volume (µL) in comparison to other techniques (mL to L).

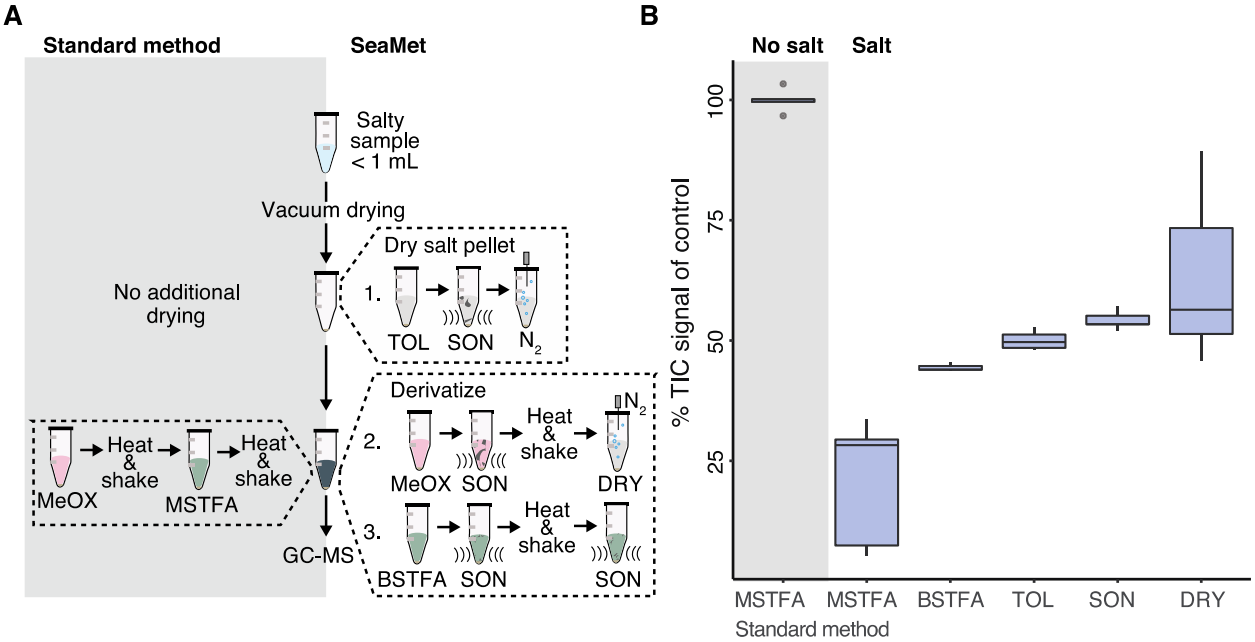


Figure 2. How SeaMet works. A, Modifications to the standard two-step methoximation (MeOX)-trimethylsilylation (TMS) derivatization protocol include key steps that enhance metabolite detection in seawater as shown in **b.** Steps modified from the standard method include a switch in derivatization reagents 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 the BSTFA reaction, and drying (DRY) of the pyridine after the MeOX reaction prior to BSTFA addition. **B,** Boxplots showing the total detected signals from each chromatogram after GC-MS from a synthetic mixture of 45 metabolites (**Table S1**) dissolved in 0.5 mL of seawater (n = 5) relative to the no salt control average.

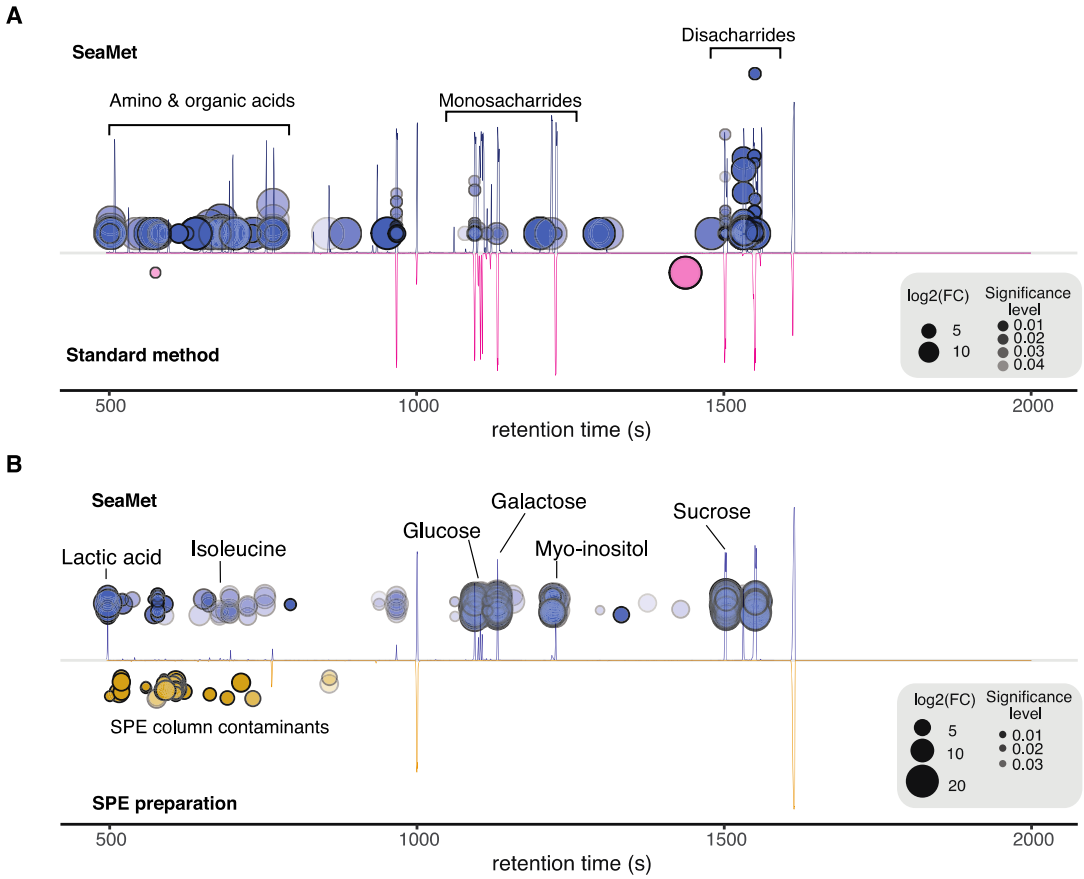


Figure 3. SeaMet enhances the detection of metabolites in marine samples. A, B Total ion chromatogram cloud plots from metabolite mixture samples indicate significant differences (B.H. $p < 0.05$) between ion abundances when comparing **A,** SeaMet (top) to the standard (bottom) protocol and **B,** chromatograms using SeaMet before (top) and after (bottom) solid phase extraction (SPE). Larger bubbles indicate higher log2(fold changes) between groups and more intense colors represent lower t-test p-values when comparing individual feature (*m/z* ions) intensities. Samples prepared with SeaMet had high abundances of organic acids (lactic acid, succinic acid, and fumarate), amino acids (isoleucine, leucine, threonine and valine), sugar alcohols (myo-inositol and mannitol), and sugars (fructose, glucose, cellobiose, maltose, ribose, galactose, and sucrose) in comparison to SPE-based sample preparation. Representatives of each class are indicated in **B**.

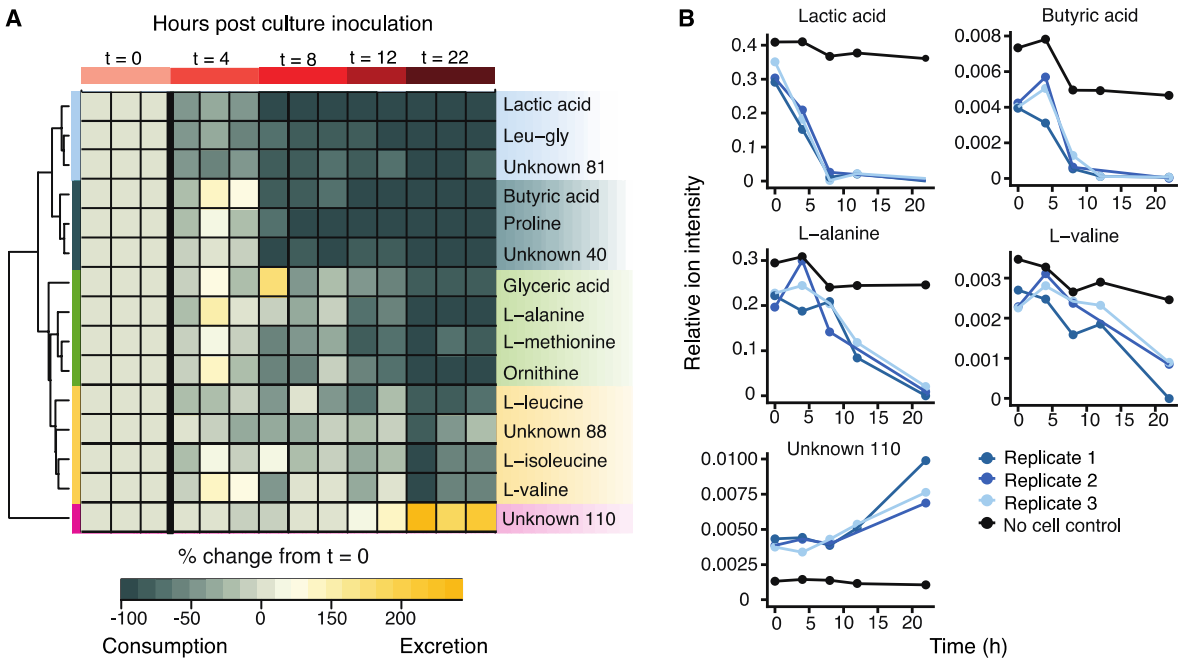


Figure 4. Metabolite consumption and excretion during culture of a marine bacterium. A, A heatmap of metabolite abundances relative to starting conditions during the culture of the heterotroph *Marinobacter adhaerens* in marine broth indicates some compounds, like the dipeptide leucine-glycine (leu-gly), and lactic acid are taken up before others, such as branch chain amino acids. After 12–22 hours of growth, the bacterial culture excretes an unknown compound (Unknown 110). Hierarchical clustering shows groups of metabolites that changed significantly during the experiment (left hand colored bars, B.H. adjusted p-value < 0.05; fold change > 2). These metabolite groups represent successive stages in *M. adhaerens* consumption and production of marine broth components. **B,** Relative ion abundances over time for select metabolites from each cluster group depicted in **A.** The blue lines represent biological replicate cultures while the black line shows results from a control sample with no cell addition. Low variation among biological replicates highlights the reproducibility of our marine metabolomics method.

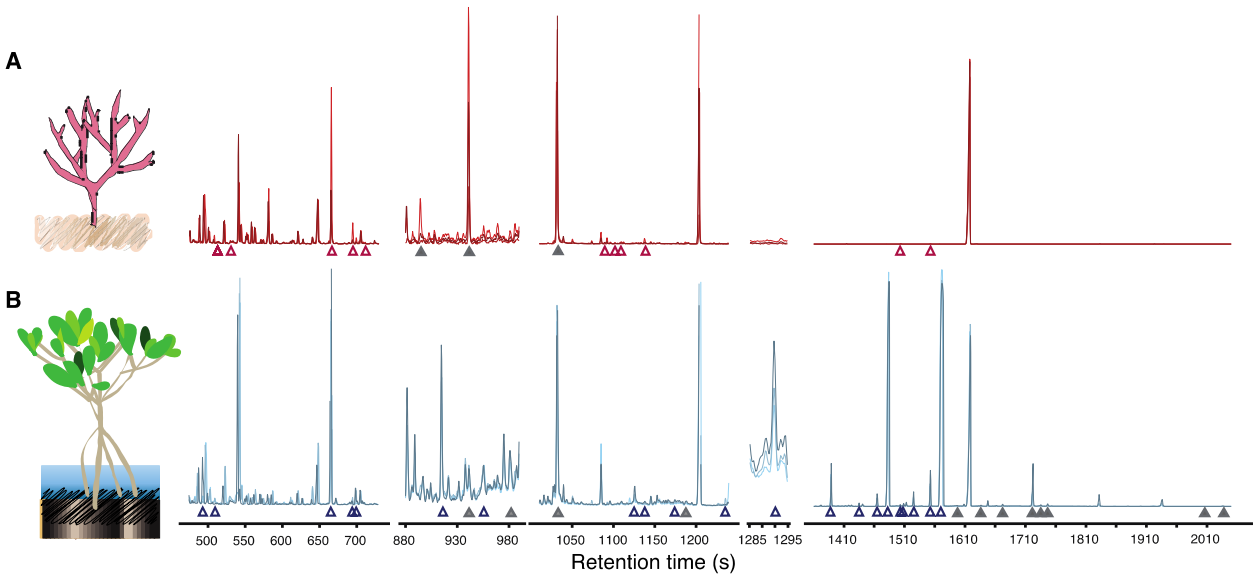


Figure 5. Metabolite profiles from marine habitats. SeaMet GC-MS metabolomic profiles from **A,** coralline and **B,** mangrove sediment porewaters show high concentrations of identified (open triangles) fatty acids and sugars that are driving multivariate differences in composition (***SI Appendix* Fig. S5A**). Profiles also reveal unknown peaks (filled triangles) for which no matches were found in public databases (**Table S5**).

Supplementary Information for

Revealing the ocean metabolome with 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

Figs. S1-S5

Tables S1-S6

Supplementary Text 1

Supplementary refereces

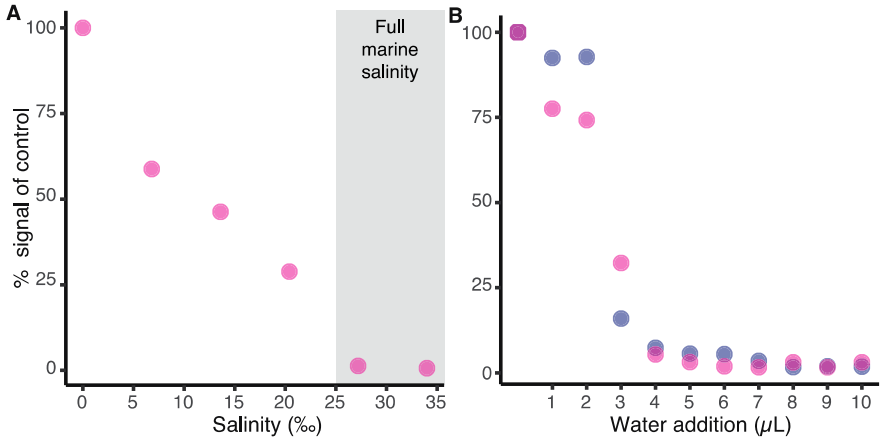


Figure S1. Salt and water inhibit metabolite derivatization reactions. Total detected ion signals are negatively related to increasing concentrations of **A,** salt and **B,** water for both MSTFA (red circles) and BSTFA (blue circles) derivatization reagents. Signals intensities from the metabolite mixture (**Table S6**) are relative to control samples (**A,** no salt; **B,** no water).

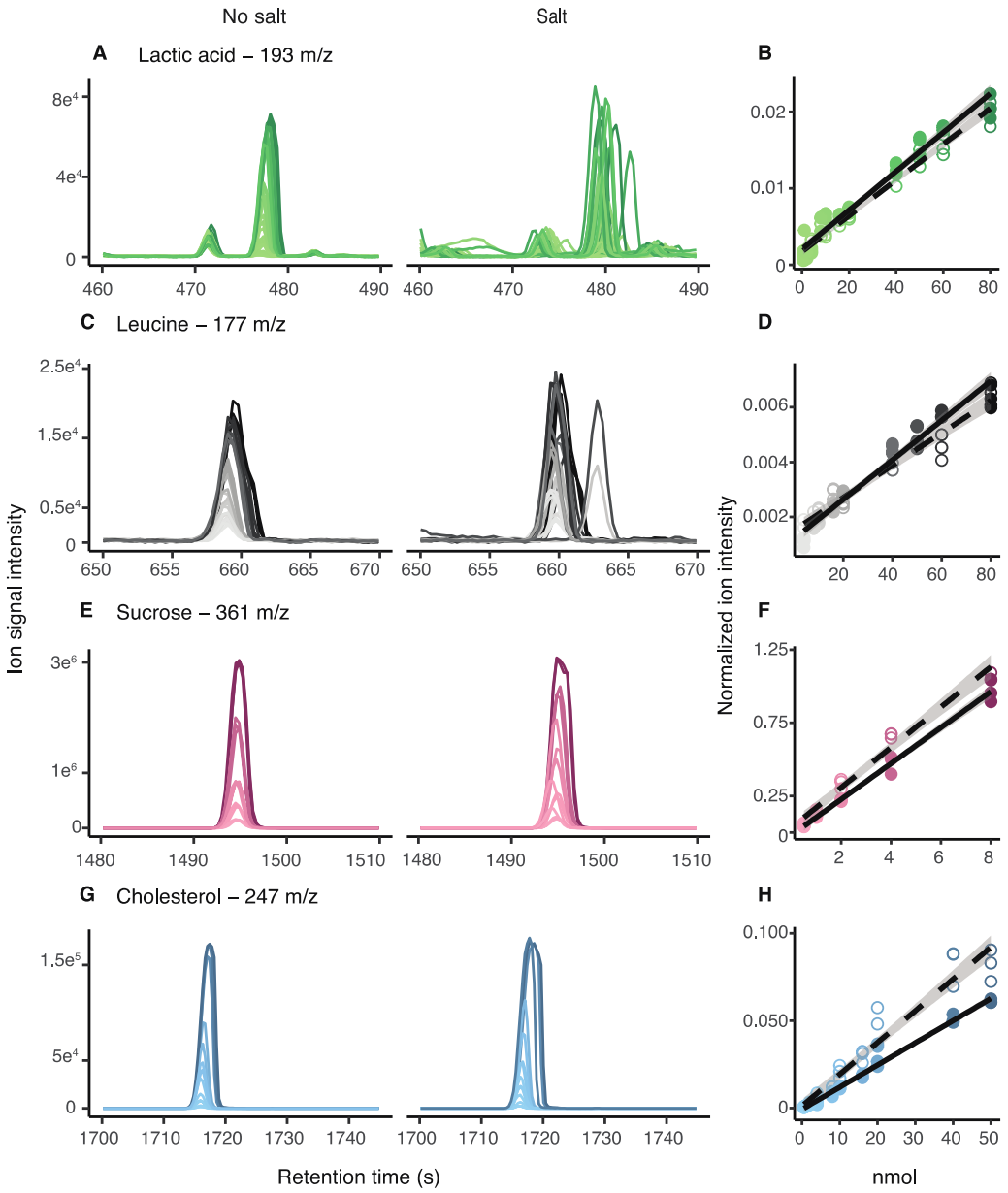


Figure S2. Metabolite detection and quantification in seawater using SeaMet is comparable to no-salt conditions. A, C, E, G, Extracted ion chromatograms for select metabolites in no salt and artificial seawater demonstrate reproducible metabolite detection across concentration gradients, as depicted in associated calibration curves in **B, D, F, H**. 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 = no salt samples.

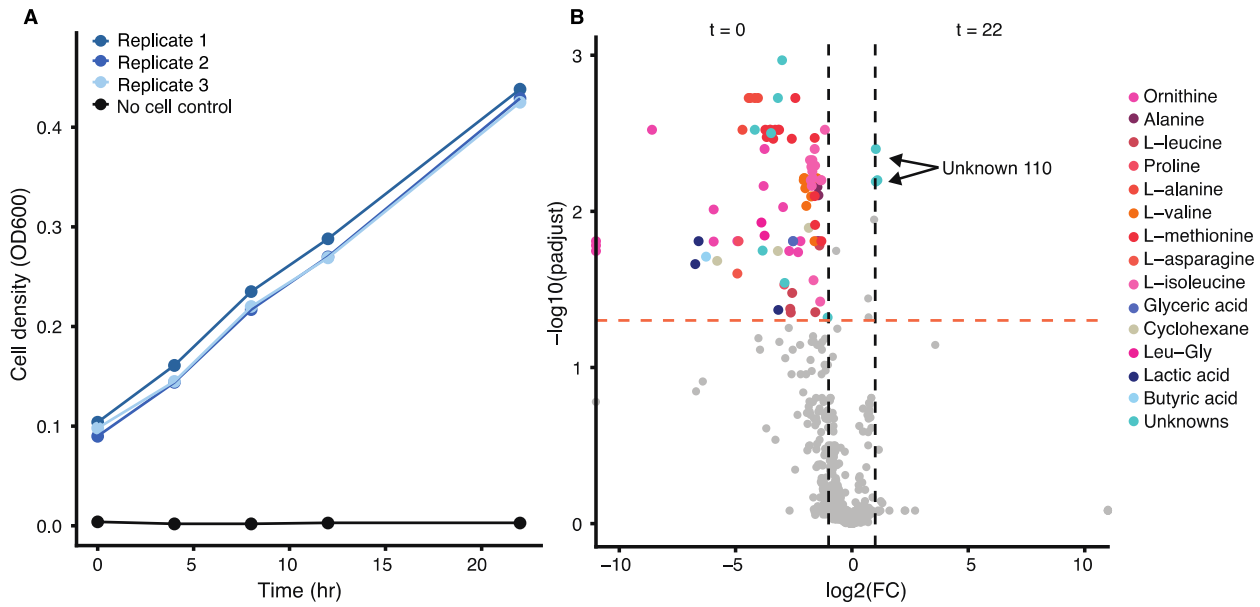


Figure S3. Marine broth metabolite composition changes with cellular growth of *Marinobacter adhaerens*. A, Cell densities increase during the first 22 hours of culture growth in Marine Broth. **B,** Volcano plot showing differences in ion abundances in cell growth media between the initial and final (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.

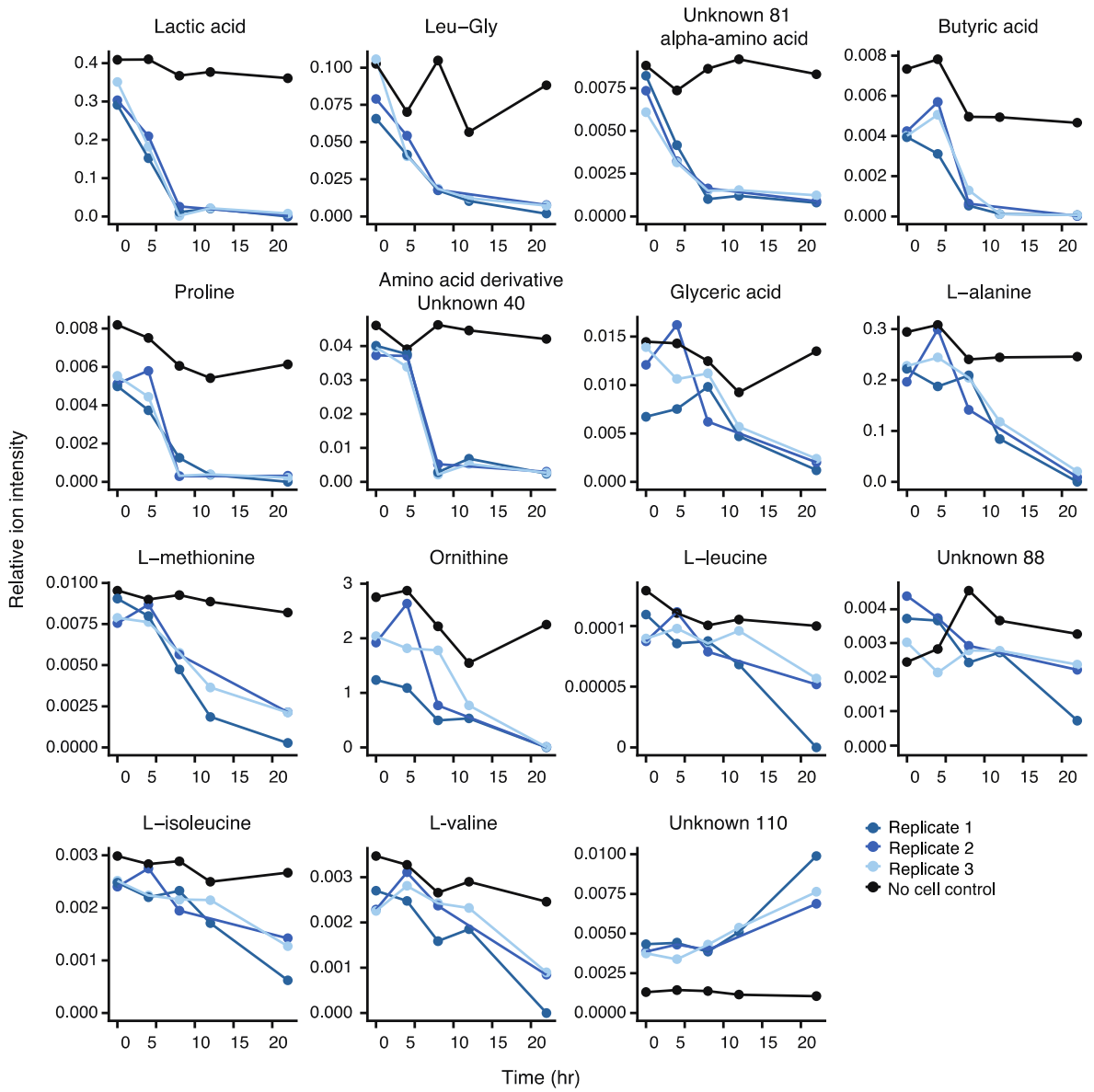


Figure S4. Extracellular metabolite levels shift with cell culture density. Metabolite relative abundances for each cell culture and no-cell control are plotted through time. Only metabolites that significantly (adjusted *p*-value < 0.05) varied with time in replicate culture experiments are plotted for clarity.

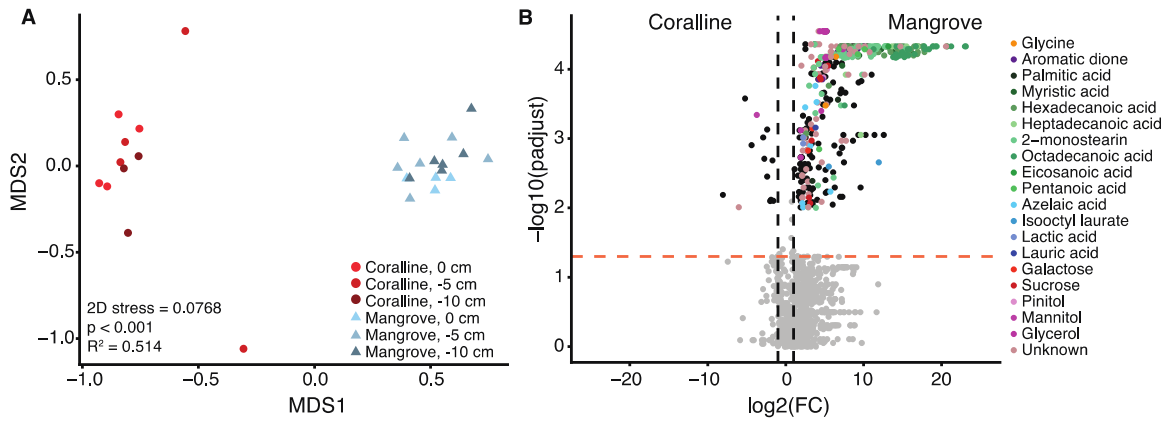


Figure S5. Coralline and mangrove sediment porewaters are chemically distinct. A, Bray-Curtis informed non-metric multidimensional scaling analysis of sediment porewater metabolomic profiles from coralline (red) and mangrove (blue) habitats across sediment depths. ADONIS p-value and R^2 show there is a significant correlation between sampling location and metabolite composition. **B,** 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.

Table S1. Mixture of 45 metabolites used for the development and testing of the preparation method, with their retention times as detected in samples from artificial seawater using GC-MS. Metabolites with multiple retention times represent different 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 detection limits of select compounds in artificial seawater are presented as well as the minimum and maximum calibration points.

			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	GC-MS	nM	1 L	3
	Solid phase extraction	LC/ESI-MS/MS	pM	5 L	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

Table S4. Retention times and quantification ions of the 107 metabolite standards detected in artificial seawater using SeaMet.

Metabolite standard	Manufacture company	KEGG ID	BRITE hierarchy/ General category	Retention time (s)	Quantification ion	Qualifiers
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
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
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 conjugates; Straight chain fatty acids	941	267	223; 193
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
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

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			
	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 122	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

Table S6. Metabolite mixtures used across experiments to develop SeaMet.

Metabolite	Experimental Mixture
Alanine	Quantifying salt-water effect
Citric acid	
Fructose	
Glucose	
Glutamic acid	
Glutamine	
Glycine	
Lactate	
Leucine	
Lysine	
Malic acid	
Myo-Inositol	
NAG	
Ornithine	
Oxalic acid	
Serine	
Succinate	
Sucrose	
Thymine	
Trehalose	
Valine	
Lactic acid	Method development and SPE
Valine	
Oxalic Acid	
Leucine	
Urea	
Serine	
Isoleucine	
Proline	
Maleic acid	
Glycine	
Succinic acid	
Fumerate	
Methionine	
Threonine	
Aspartic acid	
Glutamine	
Malic acid	
Alanine	
Glutamic acid	
Ornithine	
Phenylalanine	
Lauric acid	
Ribose	
Fructose	
Galactose	
Glucose	
Mannose	
Lysine	
Mannitol	
NAG	
Myo-inositol	
Tryptophan	
Sucrose	
Cellobiose	
Trehalose	
Maltose	

Arginine Carnitine Citric acid Cysteine DMSP Glutamate Glycerophosphate Histidine Pyruvate Thymine	
Glycine Histidine Leucine Lysine Ornithine Serine Thymine Valine Citate/citric acid Glutamic acid Lactic Acid Lauric acid Malic acid Succinate Cholesterol Fructose Glucose Sucrose Trehalose Myo-inositol	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.