# UC San Diego

## UC San Diego Previously Published Works

**Title**

Non-targeted tandem mass spectrometry enables the visualization of organic matter chemotype shifts in coastal seawater

**Permalink**

https://escholarship.org/uc/item/984440pd

**Authors**

Petras, Daniel
Minich, Jeremiah J
Cancelada, Lucia B
et al.

**Publication Date**

2021-05-01

**DOI**

10.1016/j.chemosphere.2020.129450

Peer reviewed

# Non-targeted tandem mass spectrometry enables the visualization of organic matter chemotype shifts in coastal seawater

**Daniel Petras**[a,b,*], **Jeremiah J. Minich**[b], **Lucia B. Cancelada**[c], **Ralph R. Torres**[b], **Emily Kunselman**[b], **Mingxun Wang**[a], **Margot E. White**[b], **Eric E. Allen**[b,d], **Kimberly A. Prather**[b,c], **Lihini I. Aluwihare**[b], **Pieter C. Dorrestein**[a,c]

[a]University of California San Diego, Collaborative Mass Spectrometry Innovation Center, 9500 Gilman Drive, La Jolla, USA

[b]University of California San Diego, Scripps Institution of Oceanography, 9500 Gilman Drive, La Jolla, USA

[c]University of California San Diego, Department of Chemistry, 9500 Gilman Drive, La Jolla, USA

[d]University of California San Diego, Center for Microbiome Innovation, 9500 Gilman Drive, La Jolla, USA

## Abstract

Urbanization along coastlines alters marine ecosystems including contributing molecules of anthropogenic origin to the coastal dissolved organic matter (DOM) pool. A broad assessment of the nature and extent of anthropogenic impacts on coastal ecosystems is urgently needed to inform regulatory guidelines and ecosystem management. Recently, non-targeted tandem mass spectrometry approaches are gaining momentum for the analysis of global organic matter composition (chemotypes) including a wide array of natural and anthropogenic compounds. In line with these efforts, we developed a non-targeted liquid chromatography tandem mass spectrometry (LC-MS/MS) workflow that utilizes advanced data analysis approaches such as feature-based molecular networking and repository-scale spectrum searches. This workflow allows the scalable comparison and mapping of seawater chemotypes from large-scale spatial surveys as well as molecular family level annotation of unknown compounds. As a case study, we visualized organic matter chemotype shifts in coastal environments in northern San Diego, USA, after notable rain

---

[*]Correspondence should be addressed to: dpetras@ucsd.edu; Postal Address: University of California, 9500 Gilman Drive #0751, La Jolla CA 92093, USA; Phone: +1 858 337 4899.

Conflict of Interest Statement

Pieter C. Dorrestein is a scientific advisor for Sirenas LLC. Mingxun Wang is the founder of Ometa labs LLC.

fall in winter 2017/2018 and highlight potential anthropogenic impacts. The observed seawater chemotype, consisting of 4,384 LC-MS/MS features, shifted significantly after a major rain event. Molecular drivers of this shift could be attributed to multiple anthropogenic compounds, including pesticides (Imazapyr and Isoxaben), cleaning products (Benzyl-tetradecyl-dimethylammonium) and chemical additives (Hexa(methoxymethyl)melamine) and potential degradation products. By expanding the search of identified xenobiotics to other public tandem mass spectrometry datasets, we further contextualized their possible origin and show their importance in other ecosystems. The mass spectrometry and data analysis pipelines applied here offer a scalable framework for future molecular mapping and monitoring of marine ecosystems, which will contribute to a deliberate assessment of how chemical pollution impacts our oceans.

## Graphical Abstract
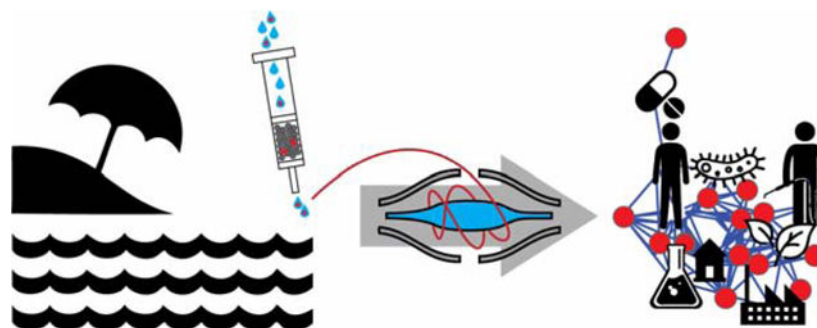


## Keywords

Dissolved Organic Matter; Tandem Mass Spectrometry; Non-targeted Screening; Molecular Networking; Coastal Environments; Anthropogenic Pollution

## 1.  Introduction

Population growth and urbanization place increasing stress on coastal marine ecosystems, resulting in numerous negative impacts. In a meta-analysis of 216 studies, chemical contaminants were associated with a 40% reduction in biodiversity richness in marine communities(Johnston and Roberts, 2009). Chemical pollutants come in numerous organic and inorganic forms, and new compounds are constantly detected in the environment. Currently, there are over 157 million unique organic and inorganic chemical substances registered in the Chemical Abstract Service (CAS), with 30,000 to 70,000 estimated to be in daily use through human activities (Schwarzenbach et al., 2006). This incredible chemical diversity, including a myriad of potential degradation products, makes it challenging for eco-toxicologists to monitor and evaluate the impacts of all possible pollutants. In spite of this chemical diversity, targeted mass spectrometry (MS) methods, that target a limited number of a priori known compounds, are typically applied for determining the presence and abundance of chemical pollutants in routine monitoring (Magi and Di Carro, 2018; Petrie et al., 2016). For example, the Clean Water Act of the Environmental Protection Agency (EPA) in the United States recommends monitoring 126 compounds(US EPA, 2013). However, as mentioned above, these compounds occupy only a small percentage of the chemical space of

potential pollutants and their degradation products (Kolpin et al., 2002; Schwarzenbach et al., 2006).

In order to account for a broader range of possible chemical compounds, the use of high resolution mass spectrometry (HR-MS), including bioinformatic infrastructures, has gained momentum in the field of non-targeted analysis (Albergamo et al., 2019; Alygizakis et al., 2019; Gago-Ferrero et al., 2015; Hernández et al., 2019; Hollender et al., 2017; Huntscha et al., 2014; Schlüsener et al., 2015; Schymanski et al., 2014b; Ulrich et al., 2019; Verkh et al., 2018). HR-MS, such as direct infusion Electro Spray Ionization (ESI) Fourier-transform ion cyclotron resonance (FT-ICR) MS and orbital ion traps have also been widely applied to analyze organic compounds in various biogeochemical contexts (Dittmar and Paeng, 2009; Hawkes et al., 2016; Osterholz et al., 2016). Traditional biogeochemical approaches typically utilize direct infusion (DI) HR-MS and aim for a comprehensive analysis, including refractory dissolved organic matter (DOM) components that may show high isomeric complexity and are typically low in abundance and challenging to resolve by liquid chromatography. Recent advances in non-targeted liquid chromatography high resolution tandem mass spectrometry (LC-HR-MS/MS) for the analysis of complex organic mixtures could also be applied for detecting organic pollutants in marine systems that are accessible by solid phase extraction and ESI (Hawkes et al., 2018; Longnecker and Kujawinski, 2017; Lu et al., 2018a; Petras et al., 2017). MS/MS based approaches target primarily high abundant and chromatographically well resolved features (e.g., metabolites and xenobiotics) and are limited in their coverage by chromatographic and MS/MS precursor isolation as well as duty cycle time and total number of MS/MS scans. While DI-HR-MS offers comprehensive information on elemental composition which is central to the study of elemental fluxes, structural annotations remain elusive. Whereas, LC-MS/MS based annotation allows level 1 and 2 structural annotations that are essential for addressing chemical ecology and environmental toxicology questions.

However, MS/MS data analysis and especially the annotation of compounds that are not covered by spectral libraries remain as major challenges.

Here, new developments in MS software tools, including molecular networking, comprehensive spectral and structure databases, statistical significance estimation, as well *in silico* annotation and community based annotation criteria have significantly improved the annotation of MS/MS spectra (Aron et al., 2019; Dührkop et al., 2019; Horai et al., 2010; Nothias et al., 2020; Ruttkies et al., 2016; Scheubert et al., 2017; Silva et al., 2018; Steen et al., 2020; Wang et al., 2016; Watrous et al., 2012). These technological improvements enable the annotation of a multitude of known and unknown chemical compounds which is central for a mechanistic understanding and prediction of how human actions influence ecosystems. This case study serves as a first large scale application of Feature-based Molecular Networking for the analysis of marine organic matter composition. In this study, we assessed the chemical impacts of a major rain event in northern San Diego, California (USA) in Winter 2017/2018. While Southern California normally experiences a dry climate, rainfall is typically observed in Winter months and can cause runoff into waterways around San Diego. Conversely, for the dry Summer month, we expect fewer terrestrial molecular drivers to be detected due to less run-off. However, from previous studies (Stephens et al., 2020), we

know the performance of our SPE and LC-MS/MS workflow to be of similar quality during the summer month, with typically higher water temperature. Our data provide insights into global chemical impacts on coastal environments in Northern San Diego after notable rain in the winter season and offer complementary insights to microbiology (He and He, 2008; Steele et al., 2018) and chemical analysis of anthropogenic pollutants in the area (Patterson et al., 1976; Tran et al., 1997). Our results show a clear shift in the observed organic matter composition after the rain event that could be attributed in part to the increased presence of multiple anthropogenic pollutants, some from identifiable point sources. The results serve thereby as strong case for the use of non-targeted LC-MS/MS and advanced data-analysis methods such as Feature-based Molecular Networking, and the sharing and reuse of MS/MS datasets (Jarmusch et al., 2020; Nothias et al., 2020; Wang et al., 2020).

## 2. Experimental Section

A detailed experimental procedure, including mass spectrometry setting can be found in the Supporting Information (SI). Between 10:00 and 13:00 (PST) on December 8th, 2017 and January 12th, 2018, we collected surface seawater samples (−10 cm) at 30 sites spaced approximately 300 meters apart and 50–100 m offshore along the San Diego coastline from Torrey Pines State Beach to Mission Bay. The water depth at sampling sites was between ~2 m (Mission Bay) and ~ 5 m (Torrey Pines, La Jolla Shores, La Jolla Reefs, Pacific and Mission Beach). Geographical locations of sampling stations are shown in Figure 1. Aliquots of seawater were taken for nutrient (50 mL), total organic carbon (TOC) analysis (40 ml) and non-targeted LC-MS/MS ($2 \times 1$L). For concentration and desalting of organic compounds, we performed solid phase extraction (200 mg PPL cartridges) of seawater adjusted to pH ~2 (1.2 mL 37% HCl) and eluted in methanol (Dittmar et al., 2008). Eluted samples were concentrated *in vacuo* and re-dissolved in 100 μL MeOH/$H_2$O/Formic acid (80/19/1) and 10 μL were injected for reverse phase ultra-high-performance liquid chromatography high-resolution tandem mass-spectrometry (RP UHPLC-MS/MS) analysis using positive mode electrospray ionization (ESI) quadrupole-Orbitrap mass spectrometer as described before (Petras et al., 2017). While DI-HR-MS analysis of DOM are typically performed in negative ESI mode, we chose to run the samples in positive mode as the larger part of MS/MS libraries are acquired in positive ESI and nitrogen containing xenobiotics and metabolites have typically higher ionization efficiencies in positive mode. Retention time and m/z drifts were controlled with a quality control mix of 6 standards (Sulfamethazine, Sulfamethizole, Sulfachloropyridazine, Sulfadimethoxine, Amitryptilin, Coumarin-314) between runs and within samples with commonly observed contaminants and typical components of DOM (Dibutyl phthalate, pheophorbide A and tryptophan) which was below 0.1 min and 5 ppm (Figure S1). For level 1 annotations, authentic standards (Irgarol, Isoxaben, Imazapyr, Hexa(methoxymethyl)melamine) were dissolved in methanol/water/ formic acid (80/19/1) with final concentrations of 0.01, 0.10, 1.00 and 10.00 μg/mL. To estimate extraction efficiencies during SPE, we spiked a serial dilution (0, 1, 10, 100 ng) of authentic standards that were detected in this study (Irgarol, Isoxaben, Imazapyr, Hexa(methoxymethyl)melamine) as well as other commonly observed xenobiotics in coastal ecosystems (Carbamazepine, Cocaine)(Pereira et al., 2016) to 1L seawater samples collected at Scripps Institution of Oceanography (SIO) Pier, California, USA (Feb 15 2020).

For LC-MS/MS data analysis, ion features were generated from extracted ion chromatograms (XIC) using MZmine2 and linked to their tandem mass spectra, which provides a combination of relative abundance (XIC) and qualitative (MS/MS) information (Nothias et al., 2019). Detailed settings are provided in the supplemental information. Subsequently, mass spectra were analyzed by molecular networking to establish molecular (MS/MS similarity) relationships between all detected compounds and annotated against the *GNPS* library, which currently contains 74,044 MS/MS spectra (Dec 18, 2019, including Mass Bank (Horai et al., 2010), ReSpect (Sawada et al., 2012) and HMDB (Wishart et al., 2007)), as well as against the commercial NIST17 MS/MS spectral library. The False Discovery Rate (FDR), estimated through a target-decoy approach using the tool Passatuto (Scheubert et al., 2017), gave rise to approximately 1% false positive matches at the settings used. Besides level 2 MS/MS matches, we confirmed the annotations discussed in the main text with authentic standards as level 1 annotation. In addition to spectrum library matching, molecular formulas were calculated based on exact mass, isotope pattern, MS/MS fragmentation trees and network topology using *ZODIAC* within *SIRIUS4* (Dührkop et al., 2019; Ludwig et al., 2020) which was run as a GNPS workflow. Features from potential contaminants that were observed in the PPL process blanks with a relative peak area > 30% in comparison to the sample average were filtered out. The resulting molecular feature table was then compared to different metadata categories by means of multivariate statistical analysis and spatial visualization in a geographic heat map. A flow chart of the sample processing and LC-MS/MS analysis is shown in Figure 1. The feature table, including all levels of MS/MS annotations are provided a .csv file in the supplemental information.

Besides total organic matter composition, we analyzed macronutrient concentrations including $NH_4^+$, $NO_3^-$, $NO_2^-$, $PO_4^{3-}$, and $SiO_4^{4-}$ along with TOC (Figure S3), correlated their values against organic matter compositional shifts (Bray-Curtis Dissimilarity, ADONIS), and compared nutrient concentrations across samples taken during the two time points (Mann-Whitney) using *Qiime* (Bolyen et al., 2019; Caporaso et al., 2010). Selected MS/MS spectra from features that were annotated as potential pollutants and that showed distinct spatial patterns were then searched against all public *GNPS* datasets in the *MassIVE* spectral repository (massive.ucsd.edu) using the *Mass Spectrometry Search Tool* (*MASST*) (Wang et al., 2019). The first pass search matched the query MS/MS against the clustered MS/MS across all public GNPS datasets and then de-clustered the match results to reveal the exact underlying mass spectrometry files the matches were found in. To aid in sensitivity, we additionally queried the MS/MS spectra against the non-clustered data. The total number of MS/MS spectra considered was over 400 million.

## 3. Results and Discussion

### 3.1. Field Work and Auxiliary Measurements.

In December 2017, no precipitation was reported in San Diego, CA, USA, whereas three days before our sampling in January 2018, a major rainfall event was measured (January 9th 2018, 39.7 mm of precipitation in San Diego). At the Scripps Pier station, salinity was measured to be 33.24 PSU on December 9th 2017 and 31.9 PSU on January 9th 2018, which was lower than any measurements made in the previous year (2017) at this site. This

decrease in salinity is consistent with modest freshwater inputs following rainfall (Nezlin et al., 2008; Reifel et al., 2009). Macronutrient concentrations including $NH_4^+$ (P<0.0001), $NO_3^-$ (P<0.0001, $NO_2^-$ (P<0.0001), $PO_4^{3-}$ (P<0.0001), and $SiO_4^{4-}$ (P<0.01) along with Total Organic Carbon (P<0.01) were significantly higher following the rain event (Mann-Whitney, Figure S3) which we attributed to input from river water and/or runoff. At both time points, samples taken from the southern end of our study area (Mission Bay) were noticeably higher in all measured nutrient concentrations than other samples (Figure S3), which could be explained by increased influence of runoff at Mission Bay through the San Diego River. Looking at geographical differences from the samples taken before the rain, we noted that $NO_3^-$ and $NH_4^+$ concentrations were generally higher in Mission Bay as compared to other sites. In the samples taken after the rain, $NH_4^+$, $NO_3^-$, $NO_2^-$, $PO_4^{3-}$ as well as TOC were also higher in Mission Bay and the Los Penasquitos Lagoon outlet as compared to most other sites. As inorganic salts will be efficiently removed during solid phase extraction as part of the sample preparation for LC-MS/MS analysis, we did not expect differences in matrix effects based on nutrient concentrations.

### 3.2. Non-Targeted MS/MS Analysis.

Following solid phase extraction of seawater with hydrophobic PPL resin, which typically captured between 35–40% of dissolved organic carbon (Petras et al., 2017), we performed non-targeted LC-MS/MS analysis of the extracts. Compound specific extraction efficiencies were estimated with filtered seawater spiked with authentic standards for six compounds and ranged from 0.3%–76%. After feature extraction from LC-MS/MS data, we obtained 5,521 MS1 ion features (defined as single chromatographic peak at a given 5 ppm m/z window, of which 90% had a peak width < 15 sec and an asymmetry factor between 0.5 and 5, Figure S2) with assigned MS/MS spectra, which decreased to 4,384 MS1 ion features after PPL process blank subtraction (the feature matrix is provided in the supporting information). From these MS/MS spectra, we created a molecular network using the Feature-based Molecular Networking workflow in *GNPS* (Nothias et al., 2020; Wang et al., 2016). All MS/MS spectra were thereby searched against the *GNPS, Massbank* and *NIST17* spectral reference libraries, which resulted in 92 annotations after blank subtraction (142 in total). The observed feature annotation rate of 2.1% after blank subtraction is in a similar range to other non-targeted environmental metabolomics studies (Floros et al., 2017; Petras et al., 2017, 2016), leaving the vast majority of ions detected as not yet assignable. Low annotation rates are still driven by the sparse availability of reference MS/MS spectra in the public domain and highlights the tremendous knowledge gap in chemical space and the need to expand non-targeted approaches. In addition to limited spectral library coverage, the high isomeric complexity of environmental samples can result in chimeric MS/MS spectra (i.e., spectra containing fragments from more than one compound). The 'contamination' of MS/MS spectra by co-eluting compounds limits spectral matching and poses a major challenge for an in depth chemical analysis of DOM and other complex samples such a tryptic protein digests in proteomics studies (Hawkes et al., 2018; Houel et al., 2010; Petras et al., 2017). Therefore, high chromatographic resolution through the use of UHPLC and ideally the development and application of multi-dimensional approaches such as 2D-UHPLC and Ion Mobility will improve the resolution of ultra-complex samples such as DOM for in depth MS/MS analysis (Lu et al., 2018b; Spranger et al., 2019). However, in

large scale studies with contemporary reversed phase UHPLC methods, the MS/MS spectra of medium and high abundant features are typically of sufficient quality to enable annotation by MS/MS matching which can be propagated between samples at MS1 feature level. (The extracted ion chromatograms of the main features discussed below are shows for a set of representative samples in the supplemental information in Figure S5.) These library matches are considered level two annotations according to the 2007 Metabolomics Standards Initiative (Sumner et al., 2007) and the environmental science community (Schymanski et al., 2014a) and should be considered as putative identifications. The library matching score cut-off used in this study give rise to an 1 % false discovery rate for spectral matching (Scheubert et al., 2017; Wang et al., 2016). MS/MS mirror plots of the sample spectrum and the reference spectrum for either library or analog matches discussed here are shown in Figure S4. Precursor mass deviation and MS/MS matching score (cosine score, (Watrous et al., 2012)) of all annotations are displayed in the Feature Table .csv file in the supplemental information and can be found online under https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=5ea78acca01b42efaa8464398c631cf6. Expanding the library search to putative analogs, we were able to annotate 1404 spectra (32%) to molecular families. The analog search algorithm matches similar MS/MS spectra with different precursor masses to library spectra, where the maximum precursor and fragment mass difference between putative analogs and library compounds is restricted by the user. Through the extended search space, analog annotations are more prone to false positive matches, but matches with high MS/MS matching scores (cosine score, (Watrous et al., 2012)) and delta masses that correspond to known modifications such as methylation, hydroxylation or acetylation (Hartmann et al., 2017) can provide insight into the putative molecular family of a compound. Both spectrum library matching with strict and expanded precursor mass tolerance (analog search) that are discussed here were manually inspected for precursor mass deviation as well as number of matching fragments ions. For analog matches (spectrum library matching with expanded precursor and corresponding fragment mass tolerance(Cooper et al., 2019; Wang et al., 2016)) exact mass offsets of precursor and fragment ions where only considered if they could be explained by common chemical modifications (e.g. methylation, acetylation oxidation etc.)(Hartmann et al., 2017).

### 3.3. Global Comparison of Seawater Chemotypes.

After feature detection and annotation, we initially displayed sample-to-sample similarity by multivariate statistical analysis of their LC-MS/MS-based chemical composition with the Bray-Curtis dissimilarity metric. Sampling date (December vs. January) significantly influenced the seawater organic matter composition (Adonis, P<0.001, Figure 2C). Besides the clear time dependency, the separation of different seawater chemotypes could also be attributed to the benthic substrate (sandy bottom vs. rocky reef bottom) of particular samples sites. Together with macronutrient concentrations, including $NH_4^+$, $NO_3^-$, $NO_2^-$, and $SiO_4^{4-}$, sampling period and location explain the differentiation of the chemical seawater composition (p<0.001, Figure 2C). In a principle coordinate analysis (PCoA) shown in Figure 2A, the samples (represented as dots) are separated based on their compositional similarity with the Bray-Curtis dissimilarity metric. This separation could be attributed to sampling date (blue vs. orange), with January samples being the furthest outlier. Before the rain (December), Mission Bay samples clustered closer to samples from other locations

(Figure 2 B, red vs. orange and blue) when compared to the January samples. A similar trend was also observed for the four samples taken after the rain at the northern end of Torrey Pines State Beach (Figure 2 A, orange, Sand; Station 1 and 2, proximal to the Los Penasquitos Lagoon outlet). The compositional sample-to-sample distance in PCoA between those two sites and the other sites with sandy substrate were more dissimilar to each other after the rain than before the rain. Increased allochthonous inputs to particular stations (1, 2, 26–30) at each end of the sampling transect are consistent with their proximity to freshwater input through the Los Penasquitos Lagoon (north) or San Diego River and Rose Creek (south).

### 3.4.    Identification of Chemical Drivers.

In order to identify the main molecular drivers responsible for the differences in seawater chemotypes, we performed a Random Forest classification, a machine learning approach for compositional sample classification and regression. Two classifications between the sample groups were performed based on collection date (January vs. December) and subsequently on distance to freshwater outflows in the January samples only (close vs. far, e.g., sites 1, 2, 26–30 vs. 3–25).

Taking all features into account, the top two molecular drivers for temporal differences (December vs. January, Figure S6) did not result in any spectral library annotations. The two top drivers were a feature with $m/z$ 353.2316, RT 9.0min, $C_{20}H_{32}O_5$ ( $m/z = 1.8$ ppm), followed by a feature with $m/z$ 416.2473, RT 4.3 min $C_{18}H_{33}N_5O_6$ ( $m/z = 8.6$ ppm).

The top two drivers from the Random Forest classification that separated different sample sites in January (Figure S6) were two unknown compounds with $m/z$ 1129.3145, RT 11.8 and 12.6 min (no confident molecular formula assignment), which were organized in a molecular network (based on MS/MS similarity) with other features were also only found in January. The number of above discussed top-driving features is arbitrary. We chose to discuss only the two first drivers as we had no confident structural annotation in the top-ranked features.

Besides unknowns and putative analogs, we observed a feature within the top drivers with matching MS/MS spectra to benzyl-tetradecyl-dimethylammonium, a quaternary ammonium salt, commonly used as a disinfection agent and sanitizer (all MS/MS mirror plots of spectrum library matches discussed are shown in Figure S4). In order to prioritize the differentially abundant features with confident annotations, we filtered the Random Forest classification to features which resulted in library matches with precursor mass differences < 10 ppm. The top feature with a library match (cosine = 0.94) was Hexa(methoxymethyl)melamine (GNPS Library ID: CCMSLIB00000841617) which was confirmed by MS/MS and retention time match to an authentic standard (level1). Hexa(methoxymethyl)melamine appeared to drive the compositional distance between samples taken close to freshwater outflows. Hexa(methoxymethyl)melamine is a chemical component in resins used for coatings and plastics and has been described as a pollutant in rivers with acute fish and daphnia toxicity (Dsikowitzky and Schwarzbauer, 2015; Labunska et al., 2012). Based on an external calibration curve (Figure S8) and a standard addition experiment (Figure S9), we conclude that the extraction efficiency is only around 0.3% for

Hexa(methoxymethyl)melamine. Considering the relatively poor extraction efficiency, we assume that the low recovery can be attributed to the high content of amino groups in Hexa(methoxymethyl)melamine, that most likely undergo protonation at low pH and hence show lower affinity to the hydrophobic PPL resin.

However, at pH 8 the extraction efficiency was also low, ~3% (Figure S9) and considerably lower than for the other compounds tested, which is most likely due to the low hydrophobicity (logP = 1, Computed by XLogP3 3.0 from PubChem release 2019.06.18). Considering the low extraction efficiency, we estimate the highest concentrations observed in our study (Peak Area = 2.1E8) to be between 0.3 and 3 μg/L. Besides Hexa(methoxymethyl)melamine ($m/z$ = 391.2308), we detected several other compounds with analog matches to Hexa(methoxymethyl)melamine ($m/z$ = 377.2128; 359.2052; 333.1842; 301.1630; 303.1730), all of which were connected in a molecular network (Figure 3 B and Figure S10).

Investigating the retention times of the putative analogs, we assume that the features with m/z 359.2052, 301.1630, as well as two of the features with m/z 377.2128, are most likely in source fragments. However, derivatives with $m/z$ = 377.2128; 333.1842 and 303.1730 show independent retention times (7.0, 5.8, 4.8 and 4.4 min respectively). Due to the similar spatial patterns (Figure 4 A and B and Figure S11), we assume that those compounds are most likely degradation products that underwent subsequent demethylation, dehydration and demethoxylation ( m/z = 14.0162, $CH_2$; 18.0089, $H_2O$; 32.0251, $COH_4$; 44.0299, $C_2OH_4$. Interestingly, the presence of Hexa(methoxymethyl)melamine and some of the here described derivatives (Hexamethylolmelamine pentamethyl ether, m/z 377.2128 and Tetra(methoxymethyl)melamine, m/z 303.1730) have recently also been described as biotransformation products in municipal wastewater (Alhelou et al., 2019) and storm water run-off (Peter et al., 2018) and their MS/MS spectra show high similarity to the MS/MS spectra recorded in this study. It would be interesting to further decipher if these derivatives are formed due to biological degradation, formed as side products during polymerization, or whether they are a result of ageing of Hexa(methoxymethyl)melamine-containing polymers.

After the ranking in the Random Forest classification, we focused next on compounds that showed relevant patterns in our spatial survey (higher abundance at the northern and southern extreme of the study area) and that are known as potential pollutants. For example, Imazapyr (cosine = 0.93, GNPS Library ID: CCMSLIB00003723206, level 1 annotation), a water-soluble total herbicide used for the control of a broad range of weeds, was mainly found after the rain. The spatial distribution showed several orders of magnitude higher intensities in samples taken close to the outflows of Los Penasquitos Lagoon (north) and Mission Bay (south). Based on an external calibration curve and extraction efficiency of around 65%, the highest concentration of Imazapyr in our study (Peak Area = 2.6E8) was estimated to be between 1.5 and 15 ng/L.

Besides Imazapyr, we detected several other pesticides, such as Isoxaben (cosine = 0.98, CCMSLIB00003562269, level 1) with an estimated concentration range between 0.3 and 3 ng/L (extraction efficiency ~ 35%, Figure S8 and S9), Dimethenamide-P (cosine = 0.97, CCMSLIB00003638780) and Metalaxyl (cosine = 0.95, GNPS Library ID:

CCMSLIB00003135403) with similar spatial-temporal distributions (Figure S12), indicating that the Los Penasquitos Lagoon (north end) and Mission Bay, including Rose Creek and the San Diego River outlet (south end), are potential point sources. Agricultural activity in the San Diego metropolitan area mainly consists of residential gardening and parks and farming activity in San Diego County, e.g. Ramona and Oceanside(Scurlock, 2019; Sokolow et al., 2010). Imazapyr, Isoxaben, Dimethenamide-P are active ingredients in commercial herbicides used for weed control and could have been used in this context. Metalaxyl on the other hand is a fungicide that is typically used to treat root rot in vegetable crops and could originate from farmland located in north eastern San Diego (e.g., through the San Diego River).

Other xenobiotics, such as the herbicide and anti-biofouling agent Irgarol (cosine = 0.99, GNPS Library ID: CCMSLIB00000208263, level 1 annotation), were detected both in December and January (Figure 3 B and 4 F), indicating that the rain event did not have a major impact on the presence of these particular compounds. The spatial distribution further indicates that the marina in Mission Bay may also be a potential point source, which is consistent with the usage of Irgarol as an antifouling agent in boat paint. Interestingly, the relative abundance of Irgarol was higher in December before the rain (max Peak Area = 1.4E8) but only in the sample sites directly in the marina. Based on an external calibration curve (Figure S8 and S9, extraction efficiency ~40%), we estimate the highest concentrations to be in the range between 0.3 and 3 ng/L. After the rain, Irgarol was detected with higher abundance at the sample sites outside of the marina and the Mission Bay channel. The wider distribution could be explained by more freshwater influx into Mission Bay and hence higher dilution and efflux into the ocean.

Another compound in our dataset ($m/z$ = 214.1097) was annotated as desisopropyl-Irgarol (Figure 3 B). Desisopropyl-Irgarol has been described as a microbial and photo degradation product from Irgarol(Liu et al., 1997; Zhang et al., 2008), which is in line with the compounds' identical spatio-temporal patterns (Figure 4 G) in our survey.

In addition to the above described xenobiotics, we detected and annotated multiple natural products in this dataset. The fungal non-ribosomal cyclo-depsipeptides sporidesmolide I and II (Bertaud et al., 1965), for example, were detected with almost identical spatial patterns (Figure 4 G and H). Highest relative abundances were observed in the January samples, which were collected close to outlets of the Los Penasquitos Lagoon and Mission Bay. The two derivatives differ by one amino acid (isoleucine vs. valine) both of which are produced by the same biosynthetic machinery and organisms (Süssmuth et al., 2011; Wang et al., 2018).

### 3.5. Repository-Scale Meta-Analysis of identified pesticides.

In order to further investigate potential origins of the pesticides discussed above, we searched their MS/MS spectra against the entire *GNPS* spectral database, containing 1,239 public available MS/MS dataset with more than 400 million MS/MS spectra and 30 terabytes of data (1st August 2019, gnps.ucsd.edu) using the *Mass Spectrometry Search Tool* (*MASST*) (Wang et al., 2019). The matching datasets are listed in Table S1.

The MS/MS spectra from Irgarol for example was exclusively found in datasets from different coastal environments from California and Hawaii. All samples, in which Irgarol was present, were collected near marinas, ports or anchor sides. The spectra from the degradation product Desisopropyl-Irgarol was only present in datasets from San Diego (this study) and Hawaii. An interesting question resulting from this observation is whether Desisopropyl-Irgarol was not detected in other studies due to an overall lower concentration of Irgarol and hence a lower abundance of its breakdown product or whether differences in microbial and environmental conditions lead to less breakdown in other studies. The mass spectrum of Imazapyr was only found in the MassIVE dataset from this study. Dimethenamide-P was additionally found in samples from Imperial Beach (San Diego/Tijuana) close to the Tijuana River estuary, an area known to suffer from extreme anthropogenic pollution(Johnsen, 2018). The mass spectrum from Isoxabene was found in the dataset from this study, in data collected from office and human skin swab samples, and from *Euphorbiaceae* plants from New Caledonia. The mass spectra of the fungicide Metalaxyl on the other hand was found in more than 10 datasets including samples from citrus trees, corn, food as well as metabolomics studies with mice and human subjects. These results indicate that the pesticides found in our study could originate from direct agricultural deployment but also that these compounds are transferred through fruits and other plant-based foods to humans and animals. Nevertheless, it should be noted that MASST searches are inherently biased to environments and samples types that are present in datasets in the public domain and relies on the scientific community to share their MS/MS data in publicly accessible databases.

## 4.  Conclusion

We explored the spatio-temporal shift of the organic seawater chemotype along the northern San Diego coastline before and after a major rain event in winter 2017/2018 by means of non-targeted LC-MS/MS. We detected several thousand distinct ion features contributing to the total organic matter composition. Compositional changes in the seawater chemotype could be attributed to seasonal and spatial differences. Within this diverse and complex mixture of compounds, we could annotate several xenobiotics and potential derivatives and degradation products, and pinpoint the Mission Bay Marina, freshwater influx from Rose Creek and San Diego River, and the Los Penasquitos Lagoon as potential point sources. The repository scale meta-analysis of selected MS/MS spectra from our study could further contextualize the presence and potential origin of certain compounds and highlight other potential anthropogenic stress on environments. These results thus offer an effective example of how our data analysis pipeline and repository scale meta-analysis can assist in prioritizing, contextualizing, and tracking potential sources of anthropogenic compounds and complement other community efforts such as NORMAN (Alygizakis et al., 2019). The deposition of mass spectrometry data in public databases not only facilitates the accessibility of data and reproducibility of data analysis by other researchers, but also allows for constant retro-analysis and annotation. Data from this study for example, were first analyzed via *GNPS* in January 2018, resulting in a total of 142 library IDs (2.5% annotation rate). In August 2019, the same dataset resulted in 190 library IDs (4.6 % annotation rate, Figure S13).

We anticipate that the use of non-targeted LC-MS/MS workflows to measure and monitor anthropogenic pollution in marine environments will further increase in the future. We anticipate that the analytical and data analysis approaches presented here will contribute to obtaining new insights into the cycling of xenobiotics and their interactions in complex environmental systems.

## 5. Data Sharing

All MS/MS data can be found on the *Mass spectrometry Interactive Virtual Environment* (*MassIVE*) at https://massive.ucsd.edu/ with the identifier MSV000082312. Molecular Networking and Spectrum Library Matching results can be found online at GNPS under the following links: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=5ea78acca01b42efaa8464398c631cf6 https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=f9ae547fe4a141a58d3214462e97f9e0 https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=a17b86fa6240493fbd4b906cb00ff1fc

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgement

## 8. References

Albergamo V, Schollée JE, Schymanski EL, Helmus R, Timmer H, Hollender J, de Voogt P, 2019. Nontarget Screening Reveals Time Trends of Polar Micropollutants in a Riverbank Filtration System. Environ. Sci. Technol 53, 7584–7594. 10.1021/acs.est.9b01750 [PubMed: 31244084]

Alhelou R, Seiwert B, Reemtsma T, 2019. Hexamethoxymethylmelamine – A precursor of persistent and mobile contaminants in municipal wastewater and the water cycle. Water Res. 165, 114973. 10.1016/j.watres.2019.114973 [PubMed: 31430651]

Alygizakis NA, Oswald P, Thomaidis NS, Schymanski EL, Aalizadeh R, Schulze T, Oswaldova M, Slobodnik J, 2019. NORMAN digital sample freezing platform: A European virtual platform to exchange liquid chromatography high resolution-mass spectrometry data and screen suspects in "digitally frozen" environmental samples. TrAC Trends Anal. Chem 115, 129–137. 10.1016/j.trac.2019.04.008

Aron AT, Gentry E, McPhail KL, Nothias LF, Nothias-Esposito M, Bouslimani A, Petras D, Gauglitz JM, Sikora N, Vargas F, 2019. Reproducible Molecular Networking Of Untargeted Mass Spectrometry Data Using GNPS.

Bertaud WS, Probine MC, Shannon JS, Taylor A, 1965. Isolation of a new depsipeptide from pithomyces chartarum. Tetrahedron 21, 677–680. 10.1016/S0040-4020(01)82239-0

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol 37, 852–857. [PubMed: 31341288]

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, 2010. QIIME allows analysis of high- throughput community sequencing data. Nat. Methods 7, 335. [PubMed: 20383131]

Cooper BT, Yan X, Simón-Manso Y, Tchekhovskoi DV, Mirokhin YA, Stein SE, 2019. Hybrid search: a method for identifying metabolites absent from tandem mass spectrometry libraries. Anal. Chem 91, 13924–13932. [PubMed: 31600070]

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. Methods 6, 230–235.

Dittmar T, Paeng J, 2009. A heat-induced molecular signature in marine dissolved organic matter. Nat. Geosci 2, 175–179. 10.1038/ngeo440

Dsikowitzky L, Schwarzbauer J, 2015. Hexa(methoxymethyl)melamine: An Emerging Contaminant in German Rivers. Water Environ. Res 87, 461–469. 10.2175/106143014X14060523640919 [PubMed: 26460466]

Dührkop K, Fleischauer M, Ludwig M, Aksenov AA, Melnik AV, Meusel M, Dorrestein PC, Rousu J, Böcker S, 2019. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. Nat. Methods 299–302. [PubMed: 30886413]

Floros DJ, Petras D, Kapono CA, Melnik AV, Ling T-J, Knight R, Dorrestein PC, 2017. Mass Spectrometry Based Molecular 3D-Cartography of Plant Metabolites. Front. Plant Sci 8, 429. 10.3389/fpls.2017.00429 [PubMed: 28405197]

Gago-Ferrero P, Schymanski EL, Bletsou AA, Aalizadeh R, Hollender J, Thomaidis NS, 2015. Extended Suspect and Non-Target Strategies to Characterize Emerging Polar Organic Contaminants in Raw Wastewater with LC-HRMS/MS. Environ. Sci. Technol 49, 12333–12341. 10.1021/acs.est.5b03454 [PubMed: 26418421]

Hartmann AC, Petras D, Quinn RA, Protsyuk I, Archer FI, Ransome E, Williams GJ, Bailey BA, Vermeij MJ, Alexandrov T, 2017. Meta-mass shift chemical profiling of metabolomes from coral reefs. Proc. Natl. Acad. Sci 114, 11685–11690. [PubMed: 29078340]

Hawkes JA, Dittmar T, Patriarca C, Tranvik L, Bergquist J, 2016. Evaluation of the Orbitrap Mass Spectrometer for the Molecular Fingerprinting Analysis of Natural Dissolved Organic Matter. Anal. Chem 88, 7698–7704. 10.1021/acs.analchem.6b01624 [PubMed: 27400998]

Hawkes JA, Patriarca C, Sjöberg PJR, Tranvik LJ, Bergquist J, 2018. Extreme isomeric complexity of dissolved organic matter found across aquatic environments. Limnol. Oceanogr. Lett 3, 21–30. 10.1002/lol2.10064

He L-ML, He Z-L, 2008. Water quality prediction of marine recreational beaches receiving watershed baseflow and stormwater runoff in southern California, USA. Water Res. 42, 2563–2573. [PubMed: 18242661]

Hernández F, Bakker J, Bijlsma L, Boer J. de, Botero-Coy AM, Bruin Y.B. de, Fischer S, Hollender J, Kasprzyk-Hordern B, Lamoree M, López FJ, Laak T.L. ter, Leerdam J.A. van, Sancho JV, Schymanski EL, Voogt P. de, Hogendoorn EA, 2019. The role of analytical chemistry in exposure science: Focus on the aquatic environment. Chemosphere 222, 564–583. 10.1016/j.chemosphere.2019.01.118 [PubMed: 30726704]

Hollender J, Schymanski EL, Singer HP, Ferguson PL, 2017. Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? Environ. Sci. Technol 51, 11505–11512. 10.1021/acs.est.7b02184 [PubMed: 28877430]

Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, Ojima Y, Tanaka K, Tanaka S, Aoshima K, 2010. MassBank: a public repository for sharing mass spectral data for life sciences. J. Mass Spectrom 45, 703–714. [PubMed: 20623627]

Houel S, Abernathy R, Renganathan K, Meyer-Arendt K, Ahn NG, Old WM, 2010. Quantifying the Impact of Chimera MS/MS Spectra on Peptide Identification in Large- Scale Proteomics Studies. J. Proteome Res 9, 4152–4160. 10.1021/pr1003856 [PubMed: 20578722]

Huntscha S, Hofstetter TB, Schymanski EL, Spahr S, Hollender J, 2014. Biotransformation of benzotriazoles: insights from transformation product identification and compound-specific isotope analysis. Environ. Sci. Technol 48, 4435–4443. [PubMed: 24621328]

Jarmusch AK, Wang M, Aceves CM, Advani RS, Aguirre S, Aksenov AA, Aleti G, Aron AT, Bauermeister A, Bolleddu S, Bouslimani A, Caraballo Rodriguez AM, Chaar R, Coras R, Elijah EO, Ernst M, Gauglitz JM, Gentry EC, Husband M, Jarmusch SA, Jones KL, Kamenik Z, Le Gouellec A, Lu A, McCall L-I, McPhail KL, Meehan MJ, Melnik AV, Menezes RC, Montoya Giraldo YA, Nguyen NH, Nothias LF, Nothias-Esposito M, Panitchpakdi M, Petras D, Quinn RA, Sikora N, van der Hooft JJJ, Vargas F, Vrbanac A, Weldon KC, Knight R, Bandeira N, Dorrestein PC, 2020. ReDU: a framework to find and reanalyze public mass spectrometry data. Nat. Methods 17, 901–904. 10.1038/s41592-020-0916-7 [PubMed: 32807955]

Johnsen C, 2018. Countries Divided, Shared Pollution: Whose Mess is it? A Case Study on the issue of Transnational Border Sewage in the Tijuana River Valley Watershed.

Johnston EL, Roberts DA, 2009. Contaminants reduce the richness and evenness of marine communities: A review and meta-analysis. Environ. Pollut 157, 1745–1752. 10.1016/j.envpol.2009.02.017 [PubMed: 19286291]

Kolpin DW, Furlong ET, Meyer MT, Thurman EM, Zaugg SD, Barber LB, Buxton HT, 2002. Pharmaceuticals, hormones, and other organic wastewater contaminants in U.S. streams, 1999–2000: a national reconnaissance. Environ. Sci. Technol 36, 1202–1211. [PubMed: 11944670]

Labunska I, Brigden K, Santillo D, Johnston P, 2012. Organic chemical and heavy metal contaminants in wastewater discharged from three textile manufacturing and washing facilities in Mexico. Technical Report 08/2012, Greenpeace Research Laboratories, Exeter, UK.

Liu D, Maguire R, Lau Y, Pacepavicius G, Okamura H, Aoyama I, 1997. Transformation of the new antifouling compound Irgarol 1051 by Phanerochaete chrysosporium. Water Res. 31, 2363–2369.

Longnecker K, Kujawinski EB, 2017. Mining mass spectrometry data: Using new computational tools to find novel organic compounds in complex environmental mixtures. Org. Geochem 110, 92–99.

Lu K, Gardner WS, Liu Z, 2018a. Molecular Structure Characterization of Riverine and Coastal Dissolved Organic Matter with Ion Mobility Quadrupole Time-of-Flight LCMS (IM Q-TOF LCMS). Environ. Sci. Technol 52, 7182–7191. 10.1021/acs.est.8b00999 [PubMed: 29870664]

Lu K, Gardner WS, Liu Z, 2018b. Molecular Structure Characterization of Riverine and Coastal Dissolved Organic Matter with Ion Mobility Quadrupole Time-of-Flight LCMS (IM Q-TOF LCMS). Environ. Sci. Technol 52, 7182–7191. 10.1021/acs.est.8b00999 [PubMed: 29870664]

Ludwig M, Nothias L-F, Dührkop K, Koester I, Fleischauer M, Hoffmann MA, Petras D, Vargas F, Morsy M, Aluwihare L, 2020. Database-independent molecular formula annotation using Gibbs sampling through ZODIAC. Nat. Mach. Intell 2, 629–641.

Magi E, Di Carro M, 2018. Marine environment pollution: The contribution of mass spectrometry to the study of seawater. Mass Spectrom. Rev 37, 492–512. 10.1002/mas.21521 [PubMed: 27611504]

Nezlin NP, DiGiacomo PM, Diehl DW, Jones BH, Johnson SC, Mengel MJ, Reifel KM, Warrick JA, Wang M, 2008. Stormwater plume detection by MODIS imagery in the southern California coastal ocean. Estuar. Coast. Shelf Sci 80, 141–152.

Nothias LF, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A, Protsyuk I, Ernst M, Tsugawa H, Fleischauer M, Aicheler F, Aksenov A, Alka O, Allard P-M, Barsch A, Cachet X, Caraballo M, Da Silva RR, Dang T, Garg N, Gauglitz JM, Gurevich A, Isaac G, Jarmusch AK, Kameník Z, Kang KB, Kessler N, Koester I, Korf A, Gouellec AL, Ludwig M, Christian MH, McCall L-I, McSayles J, Meyer SW, Mohimani H, Morsy M, Moyne O, Neumann S, Neuweger H, Nguyen NH, Nothias-Esposito M, Paolini J, Phelan VV, Pluskal T, Quinn RA, Rogers S, Shrestha B, Tripathi A, van der Hooft JJJ, Vargas F, Weldon KC, Witting M, Yang H, Zhang Z, Zubeil F, Kohlbacher O, Böcker S, Alexandrov T, Bandeira N, Wang M, Dorrestein PC, 2019. Feature-based Molecular Networking in the GNPS Analysis Environment. bioRxiv 812404. 10.1101/812404

Nothias L-F, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A, Protsyuk I, Ernst M, Tsugawa H, Fleischauer M, Aicheler F, Aksenov AA, Alka O, Allard P-M, Barsch A, Cachet X, Caraballo-Rodriguez AM, Da Silva RR, Dang T, Garg N, Gauglitz JM, Gurevich A, Isaac G, Jarmusch AK, Kameník Z, Kang KB, Kessler N, Koester I, Korf A, Le Gouellec A, Ludwig M, Martin H,C, McCall L-I, McSayles J, Meyer SW, Mohimani H, Morsy M, Moyne O, Neumann S, Neuweger H, Nguyen NH, Nothias-Esposito M, Paolini J, Phelan VV, Pluskal T, Quinn RA, Rogers S, Shrestha B, Tripathi A, van der Hooft JJJ, Vargas F, Weldon KC, Witting M, Yang H, Zhang Z, Zubeil F, Kohlbacher O, Böcker S, Alexandrov T, Bandeira N, Wang M, Dorrestein PC, 2020. Feature-based

molecular networking in the GNPS analysis environment. Nat. Methods 17, 905–908. 10.1038/s41592-020-0933-6 [PubMed: 32839597]

Osterholz H, Singer G, Wemheuer B, Daniel R, Simon M, Niggemann J, Dittmar T, 2016. Deciphering associations between dissolved organic molecules and bacterial communities in a pelagic marine system. ISME J. 10, 1717–1730. 10.1038/ismej.2015.231 [PubMed: 26800236]

Patterson C, Settle D, Glover B, 1976. Analysis of lead in polluted coastal seawater. Mar. Chem 4, 305–319.

Pereira CDS, Maranho LA, Cortez FS, Pusceddu FH, Santos AR, Ribeiro DA, Cesar A, Guimarães LL, 2016. Occurrence of pharmaceuticals and cocaine in a Brazilian coastal zone. Sci. Total Environ 548, 148–154. [PubMed: 26802343]

Peter KT, Tian Z, Wu C, Lin P, White S, Du B, McIntyre JK, Scholz NL, Kolodziej EP, 2018. Using High-Resolution Mass Spectrometry to Identify Organic Contaminants Linked to Urban Stormwater Mortality Syndrome in Coho Salmon. Environ. Sci. Technol 52, 10317–10327. 10.1021/acs.est.8b03287 [PubMed: 30192129]

Petras D, Koester I, Da Silva R, Stephens BM, Haas AF, Nelson CE, Kelly LW, Aluwihare LI, Dorrestein PC, 2017. High-Resolution Liquid Chromatography Tandem Mass Spectrometry Enables Large Scale Molecular Characterization of Dissolved Organic Matter. Front. Mar. Sci 4. 10.3389/fmars.2017.00405

Petras D, Nothias L-F, Quinn RA, Alexandrov T, Bandeira N, Bouslimani A, Castro-Falcón G, Chen L, Dang T, Floros DJ, Hook V, Garg N, Hoffner N, Jiang Y, Kapono CA, Koester I, Knight R, Leber CA, Ling T-J, Luzzatto-Knaan T, McCall L-I, McGrath AP, Meehan MJ, Merritt JK, Mills RH, Morton J, Podvin S, Protsyuk I, Purdy T, Satterfield K, Searles S, Shah S, Shires S, Steffen D, White M, Todoric J, Tuttle R, Wojnicz A, Sapp V, Vargas F, Yang J, Zhang C, Dorrestein PC, 2016. Mass Spectrometry-Based Visualization of Molecules Associated with Human Habitats. Anal. Chem 88, 10775–10784. 10.1021/acs.analchem.6b03456 [PubMed: 27732780]

Petrie B, Youdan J, Barden R, Kasprzyk-Hordern B, 2016. Multi-residue analysis of 90 emerging contaminants in liquid and solid environmental matrices by ultra-high-performance liquid chromatography tandem mass spectrometry. J. Chromatogr. A 1431, 64–78. 10.1016/j.chroma.2015.12.036 [PubMed: 26792447]

Reifel KM, Johnson SC, DiGiacomo PM, Mengel MJ, Nezlin NP, Warrick JA, Jones BH, 2009. Impacts of stormwater runoff in the Southern California Bight: Relationships among plume constituents. Cont. Shelf Res 29, 1821–1835.

Ruttkies C, Schymanski EL, Wolf S, Hollender J, Neumann S, 2016. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. J. Cheminformatics 8, 3.

Sawada Y, Nakabayashi R, Yamada Y, Suzuki M, Sato M, Sakata A, Akiyama K, Sakurai T, Matsuda F, Aoki T, 2012. RIKEN tandem mass spectral database (ReSpect) for phytochemicals: a plant-specific MS/MS-based data resource and database. Phytochemistry 82, 38–45. [PubMed: 22867903]

Scheubert K, Hufsky F, Petras D, Wang M, Nothias L-F, Dührkop K, Bandeira N, Dorrestein PC, Böcker S, 2017. Significance estimation for large scale metabolomics annotations by spectral matching. Nat. Commun 8, 1494. 10.1038/s41467-017-01318-5 [PubMed: 29133785]

Schlüsener MP, Kunkel U, Ternes TA, 2015. Quaternary Triphenylphosphonium Compounds: A New Class of Environmental Pollutants. Environ. Sci. Technol 49, 14282–14291. 10.1021/acs.est.5b03926 [PubMed: 26555603]

Schwarzenbach RP, Escher BI, Fenner K, Hofstetter TB, Johnson CA, von Gunten U, Wehrli B, 2006. The challenge of micropollutants in aquatic systems. Science 313, 1072–1077. 10.1126/science.1127291 [PubMed: 16931750]

Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, Hollender J, 2014a. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. Environ. Sci. Technol 48, 2097–2098. 10.1021/es5002105 [PubMed: 24476540]

Schymanski EL, Singer HP, Longrée P, Loos M, Ruff M, Stravs MA, Ripollés Vidal C, Hollender J, 2014b. Strategies to characterize polar organic contamination in wastewater: exploring the capability of high resolution mass spectrometry. Environ. Sci. Technol 48, 1811–1818. 10.1021/es4044374 [PubMed: 24417318]

Scurlock AN, 2019. Urban Agriculture in San Diego: Garden Characteristics, Practices, and Environmental Outcomes.

Silva R.R. da, Wang M, Nothias L-F, Hooft J.J.J. van der, Caraballo-Rodríguez AM, Fox E, Balunas MJ, Klassen JL, Lopes NP, Dorrestein PC, 2018. Propagating annotations of molecular networks using in silico fragmentation. PLOS Comput. Biol 14, e1006089. 10.1371/journal.pcbi.1006089 [PubMed: 29668671]

Sokolow A, Hammond S, Norton M, Schmidt E, 2010. California communities deal with conflict and adjustment at the urban-agricultural edge. Calif. Agric 64, 121–128.

Spranger T, Pinxteren D. van, Reemtsma T, Lechtenfeld OJ, Herrmann H, 2019. 2D Liquid Chromatographic Fractionation with Ultra-high Resolution MS Analysis Resolves a Vast Molecular Diversity of Tropospheric Particle Organics. Environ. Sci. Technol 53, 11353–11363. [PubMed: 31478645]

Steele JA, Blackwood AD, Griffith JF, Noble RT, Schiff KC, 2018. Quantification of pathogens and markers of fecal contamination during storm events along popular surfing beaches in San Diego, California. Water Res. 136, 137–149. 10.1016/j.watres.2018.01.056 [PubMed: 29501758]

Steen AD, Kusch S, Abdulla HA, Caki N, Coffinet S, Dittmar T, Fulton JM, Galy V, Hinrichs K-U, Ingalls AE, 2020. Analytical and Computational Advances, Opportunities, and Challenges in Marine Organic Biogeochemistry in an Era of "Omics." Front. Mar. Sci 7.

Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW-M, Fiehn O, Goodacre R, Griffin JL, 2007. Proposed minimum reporting standards for chemical analysis. Metabolomics 3, 211–221. [PubMed: 24039616]

Süssmuth R, Müller J, Von Döhren H, Molnár I, 2011. Fungal cyclooligomer depsipeptides: from classical biochemistry to combinatorial biosynthesis. Nat. Prod. Rep 28, 99–124. [PubMed: 20959929]

Tran K, Yu CC, Zeng EY, 1997. Organic pollutants in the coastal environment off San Diego, California. 2. Petrogenic and biogenic sources of aliphatic hydrocarbons. Environ. Toxicol. Chem. Int. J 16, 189–195.

Ulrich EM, Sobus JR, Grulke CM, Richard AM, Newton SR, Strynar MJ, Mansouri K, Williams AJ, 2019. EPA's non-targeted analysis collaborative trial (ENTACT): genesis, design, and initial findings. Anal. Bioanal. Chem 411, 853–866. 10.1007/s00216-018-1435-6 [PubMed: 30519961]

US EPA, O., 2013. Summary of the Clean Water Act [WWW Document]. US EPA. URL https://www.epa.gov/laws-regulations/summary-clean-water-act (accessed 10.18.18).

Verkh Y, Rozman M, Petrovic M, 2018. A non-targeted high-resolution mass spectrometry data analysis of dissolved organic matter in wastewater treatment. Chemosphere 200, 397–404. 10.1016/j.chemosphere.2018.02.095. [PubMed: 29499520]

Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu W-T, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu C-C, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw C-C, Yang Y-L, Humpf H-U, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, P CAB, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque, Pociute E, O'Neill EC, Briand, Helfrich EJN, Granatosky EA, Glukhov, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard P-M, Phapale P, Nothias L-F, Alexandrov T, Litaudon M, Wolfender J-L, Kyle JE, Metz TO, Peryea T, Nguyen D-T, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BØ, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N, 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat. Biotechnol 34, 828–837. 10.1038/nbt.3597 [PubMed: 27504778]

Wang M, Jarmusch AK, Vargas F, Aksenov AA, Gauglitz JM, Weldon K, Petras D, da Silva R, Quinn R, Melnik AV, van der Hooft JJJ, Caraballo-Rodríguez AM, Nothias LF, Aceves CM, Panitchpakdi M, Brown E, Di Ottavio F, Sikora N, Elijah EO, Labarta-Bajo L, Gentry EC, Shalapour S, Kyle KE, Puckett SP, Watrous JD, Carpenter CS, Bouslimani A, Ernst M, Swafford AD, Zúñiga EI, Balunas MJ, Klassen JL, Loomba R, Knight R, Bandeira N, Dorrestein PC, 2020. Mass spectrometry searches using MASST. Nat. Biotechnol 38, 23–26. 10.1038/s41587-019-0375-9 [PubMed: 31894142]

Wang M, Jarmusch AK, Vargas F, Aksenov AA, Gauglitz JM, Weldon K, Petras D, Silva R da, Quinn R, Melnik AV, van der Hooft JJJ, Caraballo Rodríguez AM, Nothias LF, Aceves CM, Panitchpakdi M, Brown E, Di Ottavio F, Sikora N, Elijah EO, Labarta-Bajo L, Gentry EC, Shalapour S, Kyle KE, Puckett SP, Watrous JD, Carpenter CS, Bouslimani A, Ernst M, Swafford AD, Zúñiga EI, Balunas MJ, Klassen JL, Loomba R, Knight R, Bandeira N, Dorrestein PC, 2019. MASST: A Web-based Basic Mass Spectrometry Search Tool for Molecules to Search Public Data. bioRxiv 591016. 10.1101/591016

Wang X, Gong X, Li P, Lai D, Zhou L, 2018. Structural diversity and biological activities of cyclic depsipeptides from fungi. Molecules 23, 169.

Watrous J, Roach P, Alexandrov T, Heath BS, Yang JY, Kersten RD, Voort M. van der, Pogliano K, Gross H, Raaijmakers JM, Moore BS, Laskin J, Bandeira N, Dorrestein PC, 2012. Mass spectral molecular networking of living microbial colonies. Proc. Natl. Acad. Sci 109, E1743–E1752. 10.1073/pnas.1203689109 [PubMed: 22586093]

Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, 2007. HMDB: the human metabolome database. Nucleic Acids Res. 35, D521–D526. [PubMed: 17202168]

Zhang AQ, Leung KMY, Kwok KWH, Bao VWW, Lam MHW, 2008. Toxicities of antifouling biocide Irgarol 1051 and its major degraded product to marine primary producers. Mar. Pollut. Bull 57, 575–586. 10.1016/j.marpolbul.2008.01.021 [PubMed: 18314144]

## HIGHLIGHTS

- Feature-based Molecular Networking enables large-scale analysis of marine DOM

- Organic matter chemotype in coastal San Diego shifted significantly after rain

- Molecular drivers could be attributed to multiple anthropogenic compounds

- Spatial mapping highlighted different point sources as potential origin

- Repository-scale meta-analysis can further contextualize origin and importance
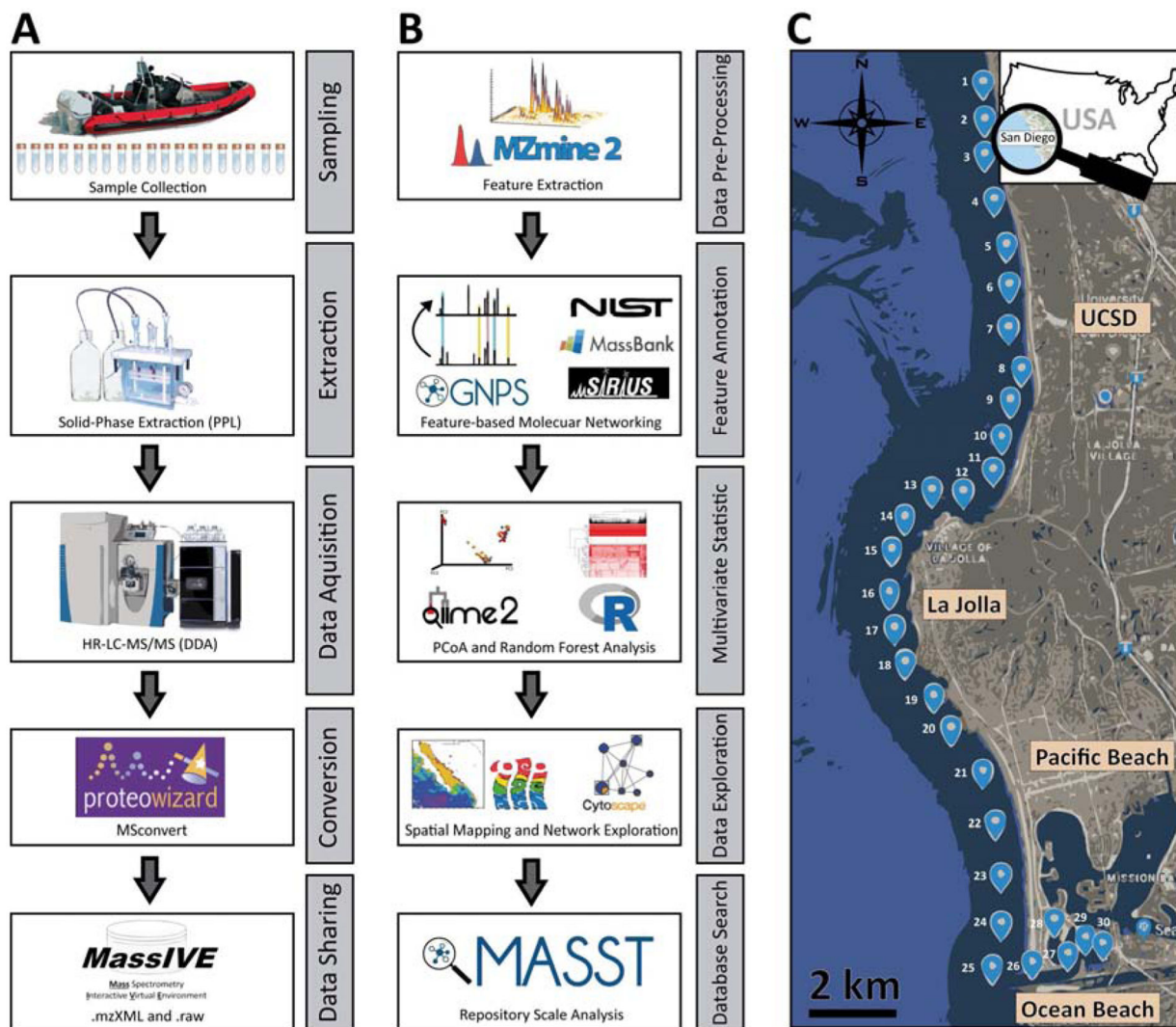
**Figure 1. Experimental design and study area.**
Panel A shows the LC-MS/MS data acquisition and data sharing steps. Panel B shows the MS/MS data analysis workflow. Panel C indicates the geographic positions of the sample sites along the San Diego coastline.
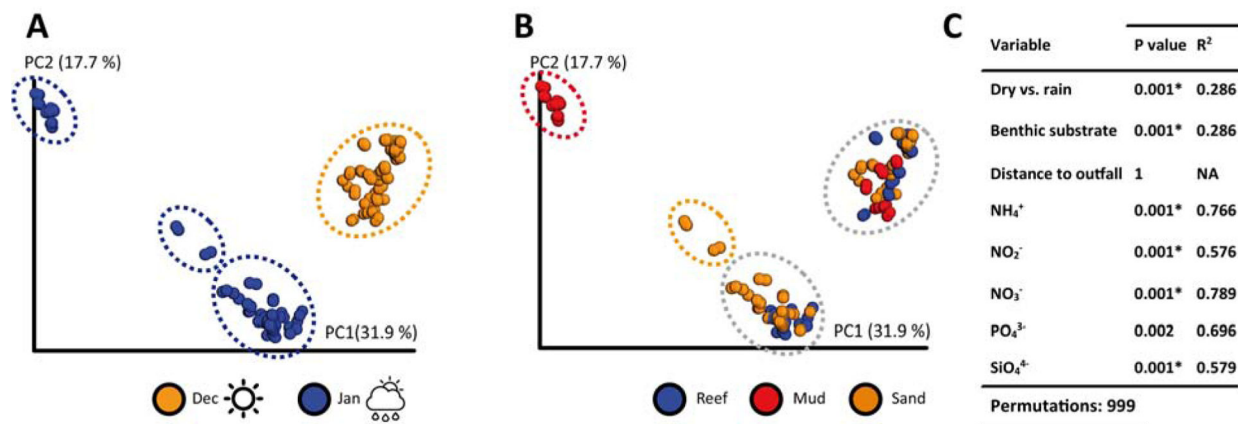
**Figure 2. Principal Coordinate Analysis (PCoA, Bray-Curtis) and Statistical Analysis (ADONIS).**

PCoA shows compositional sample to sample distance. The color coding in A (orange: before, blue: after rain) indicates that the rain event drives chemical composition of nearshore coastal seawater. In B samples are color coded based on benthic substrate (correlating to geographical differences, red: Mud/Sludge, Orange: Sand, Blue: Rock/Reef). The dashed circles indicate clustering of similar chemotypes. C shows the results of the multivariate statistical analysis of the effects of environmental variables on chemical composition in seawater. The ADONIS permutation-based analysis indicates the significance of explanation of multivariate distance (Bray-Curtis) by different meta-data categories.
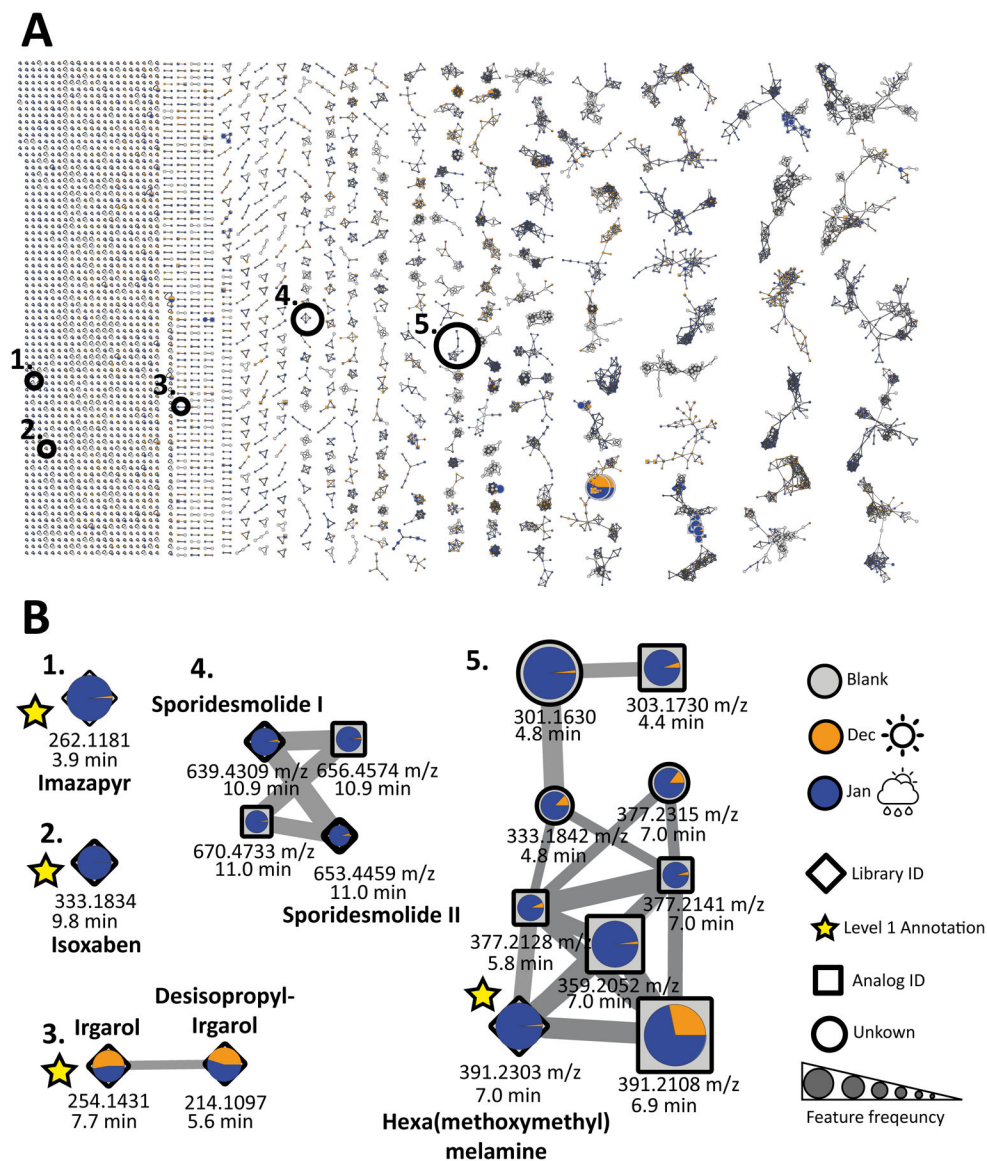
**Figure 3. Molecular Network of Seawater Chemotypes.**
Nodes represent MS/MS spectra which are connected based on their spectral similarity (cosine > 0.7). Node size indicates the relative average abundance of a feature in the data set and the pie chart indicates the relative distribution between samples taken before and after the rain event (based on MS1 XIC). Feature shape indicates the annotation level (circle: unknown, square: analog match, diamond: level 2 library match) The star symbol indicates level 1 annotations (confirmation of retention time and MS/MS spectrum with authentic standards).
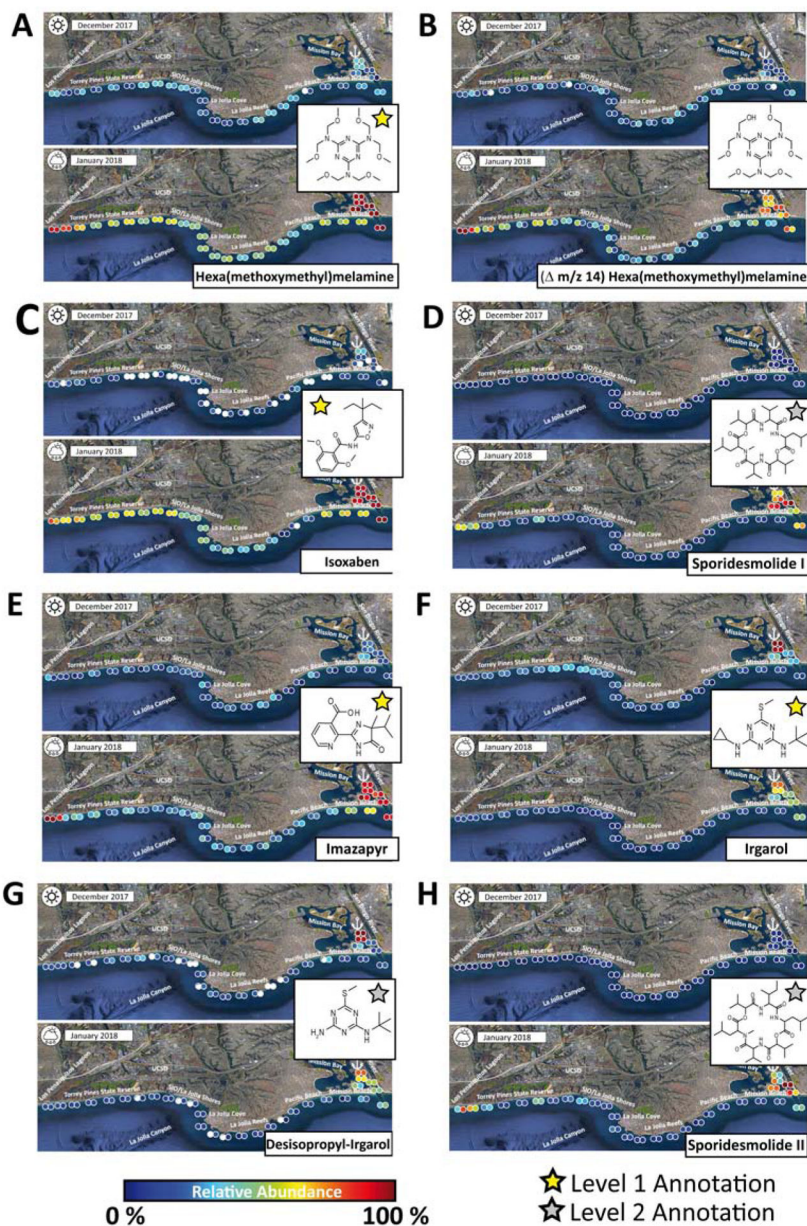
**Figure 4. Extracted Ion Maps (EIM) of relative xenobiotic and natural product concentrations in seawater samples.**

The spatial maps indicate the relative abundance of 8 selected molecular features (Xenobiotics: Hexa(methoxymethyl)-Melamine (CCMSLIB00000841617), Desmethyl-Hexa(methoxymethyl)-Melamine, Isoxaben (CCMSLIB00003562269), Imazapyr (CCMSLIB00003723206), Irgarol (CCMSLIB00000208263), Desisopropyl-Irgarol and Sporidesmolide I and II (CCMSLIB00000507910 and CCMSLIB00000577642) from two replicate samples. Absolute peak areas can be found in the feature table in the supplemental information.