

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

Digitizing mass spectrometry data to explore the chemical diversity and distribution of marine cyanobacteria and algae

Authors: Tal Luzzatto-Knaan^{1*}, Neha Garg^{1*}, Mingxun Wang², Evgenia Glukhov³, Yao Peng⁴, Gail Ackermann⁵, Amnon Amir⁵, Brendan M. Duggan¹, Sergey Ryazanov⁶, Lena Gerwick³, Rob Knight⁵, Theodore Alexandrov^{1,6}, Nuno Bandeira^{1,2}, William H. Gerwick^{1,3}, and Pieter C. Dorrestein^{1,2,3}

Affiliation:

- ¹Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, CA 92093, USA
- ²Center for Computational Mass Spectrometry and Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA 92093, USA
- ³Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California, San Diego, CA 92037, USA
- ⁴Department of Chemistry and Biochemistry, University of California San Diego, CA 92093, USA
- ⁵Departments of Pediatrics and Computer Science & Engineering, University of California San Diego, La Jolla, CA 92023, USA
- ⁶European Molecular Biology Laboratory, 69117 Heidelberg, Germany

*These authors contributed equally to this work.

Correspondence: For the cyanobacterial collections and structure elucidation contact wgerwick@ucsd.edu; for the study itself, the collection of the mass spectrometry data, molecular networking, the analysis of the mass spectrometry data, and the interpretation of the mass spectrometry data, contact pdorrestein@ucsd.edu.

Abstract

Natural product screening programs have uncovered molecules from diverse natural sources with various biological activities and unique structures. However, much is yet underexplored and additional information is hidden in these exceptional collections. We applied untargeted mass spectrometry approaches to capture the chemical space and dispersal patterns of metabolites from an in-house library of marine cyanobacterial and algal collections. Remarkably, 86% of the metabolomics signals detected were not found in other available datasets of similar nature, supporting the hypothesis that marine cyanobacteria and algae possess distinctive metabolomes. The data were plotted onto a world map representing 8 major sampling sites, and revealed potential geographic locations with high chemical diversity. We demonstrate the use of these inventories as a tool to explore the diversity and distribution of natural products. Finally, we utilized this tool to guide the isolation of a new cyclic lipopeptide, yuvalamide A, from a marine cyanobacterium.

.

Introduction

The marine environment is extraordinarily rich in both biological and chemical diversity. It is estimated that nearly 90% of all organisms living on earth inhabit marine and coastal environments (Blunt, et al., 2012; Kiuru, et al., 2014). Moreover, marine organisms are recognized as an extremely rich source of novel chemical entities that have potential utility in medicine, personal health care, cosmetics, biotechnology and agriculture (Gerwick and Moore, 2012; Kiuru, et al., 2014; La Barre, 2014; Rastogi and Sinha, 2009). Marine cyanobacteria ("blue-green algae") are considered the most ancient group of oxygenic photosynthetic organisms, having appeared on earth some three billion years ago. It is thought that their extensive evolutionary history explains, at least in part, the extensive nature of their structurally-unique and biologically-active natural products (Burja, et al., 2001; Chlipala, et al., 2011; Schopf, 2012; Whitton and Potts, 2012). Indeed, since the 1970s, several hundred unique chemical entities have been discovered from marine cyanobacteria, and these have been found to possess a remarkable spectrum of biological activities including anticancer, anti-inflammatory, antibacterial, anti-parasitic, neuromodulatory, and antiviral, among others (Blunt, et al., 2015; Nunnery, et al., 2010; Tan, 2007; Villa and Gerwick, 2010). Examples of marine cyanobacterial natural products that are under investigation for anticancer applications include the cyclodepsipeptides apratoxin A and F, the nitrogen-containing lipid curacin A and the lipopeptides dolastatin 10 and carmaphycin B (Bai, et al., 1990; Blokhin, et al., 1995; Luesch, et al., 2001; Pereira, et al., 2012; Tidgewell, et al., 2010). These natural product templates have inspired the synthesis of lead compounds using medicinal chemistry approaches, and have resulted in one clinically approved drug (Brentuximab vedotin) for the treatment of cancer; several others are in various stages of clinical and preclinical evaluation (Gerwick and Moore, 2012; Luesch, et al., 2001; Luesch, et al., 2001; Tan, et al., 2010; Tidgewell, et al., 2010; Uzair,

et al., 2012). Unfortunately, such fruitful natural product materials, rich with so much unexplored potential, oftentimes disappear upon completion of academic careers and closing of laboratories.

In this study, we used a mass spectrometry based approach to digitize (convert to data format that can be stored, shared and analyzed by computational tools) the chemical inventory of an established marine cyanobacteria and algae collection in order to better evaluate its diversity and probe for novel natural products. Additionally, while a number of biologically potent natural products have been isolated from marine cyanobacteria and algae, the overall molecular profile of these organisms has not been compared to other classes of bacteria or terrestrial/freshwater cyanobacteria (Chlipala, et al., 2011). To address this, we analyzed the crude extracts of a relatively large number of algae and cyanobacteria, as well as their derived chromatographic fractions, by high resolution liquid chromatography tandem mass spectrometry (HR-LC-MS/MS). We then compared this metabolomics dataset to four existing public datasets [Mass spectrometry Interactive Virtual Environment (MassIVE) repository] within the Global Natural Product Social (GNPS) platform (Wang, et al., 2016).

The marine samples used in this study were derived from over 300 field collections from locations in the Indian, Indo-Pacific, Central and South Pacific, and Western Atlantic (Caribbean) oceans with initial field taxonomic identifications as predominantly cyanobacteria and macro-algae. It should be noted that these are environmental samples, and therefore are inherently communities of organisms and are not single axenic cultures. Additionally, three samples were from non-axenic laboratory cultures of the cyanobacteria *Phormidium* (No.1646) and *Lyngbya* (No. 1933, 1963); the latter two were recently classified as *Moorea*

(Supplementary file 1) (Engene, et al., 2012). From both a chemical and biosynthetic perspective, these three cultures represent the most intensively studied organisms in this dataset (Engene, et al., 2012; Kleigrew, et al., 2015; Mevers, et al., 2014; Pereira, et al., 2010; Williamson, et al., 2002). While these and other of our cultured filamentous cyanobacteria cannot yet be grown axenically, their MS-derived molecular data represent marine cyanobacterial metabolomics markers that aid in data analysis. Furthermore, in exploring the chemical diversity of marine cyanobacterial and algal assemblages, we established a cartographic platform that combines LC-MS data and geographic locations that facilitates the discovery of new chemical scaffolds as well as identifies geographical areas of high chemical diversity (Boeuf and Kornprobst, 2009). The discovery of such “hotspots” may reveal new patterns of phylogenetic diversity as well as identify geographical areas with enhanced bioprospecting opportunities.

Assessment of the chemical diversity of these marine cyanobacterial and algal collections involved four major steps: 1) **Collection** - including permits, field collection, transport, extraction, fractionation and metadata recording; 2) **Data acquisition and digitization in public repository** - by LC-HRMS/MS to generate molecular fingerprints; 3) **Data analysis and visualization** - clustering similarly structured compounds as molecular families within the GNPS platform, identification of known molecules and assessing the richness and diversity between as well as within samples; 4) **Discovery** - identification of geographical patterns of distribution, differentiating common from regiospecific natural products, dereplication of new derivatives and the discovery of previously uncharacterized natural products.

Results

Over the past 30 years, a significant number of cyanobacterial and algal collections were obtained under the appropriate governmental permits using scuba diving and snorkeling in the Caribbean (Puerto Rico, Grenada, Panama), Central and South Pacific (Hawaii, Palmyra Atoll, French Polynesia, Fiji), Indo-Pacific (Indonesia, Papua New Guinea), and Indian Oceans (Madagascar, South Africa). These collections represent natural assemblages of benthic filamentous marine cyanobacteria and various classes of macro-algae (Rhodophyta, Chlorophyta and Phaeophyceae). Each collection was extracted ($\text{CH}_2\text{Cl}_2/\text{MeOH}$, 2:1) and fractionated using a standardized vacuum liquid chromatography protocol (Supplementary file 1). Approximately 2,600 fractions originating from 317 marine collections, including the unfractionated crude samples, were analyzed by reversed phase ultra-performance liquid chromatography (RP-UPLC) coupled with high resolution quadrupole time-of-flight mass spectrometry (HR-qTOF-MS) to obtain retention times and MS and tandem MS/MS fragmentation spectra (LC-MS/MS). Nearly 6,000 spectra were collected for each LC-MS/MS run, generating in excess of 15.6 million spectra for the samples in this study.

These data were analyzed using the GNPS platform enabling the extensive organization of the LC-MS/MS data (Wang, et al., 2016). GNPS detected features (i.e. molecules) based on MS/MS spectra and MS intensities. Further, some of these MS/MS spectra were identified by matching to spectral libraries available on GNPS. Even in the absence of MS/MS spectral matching to known reference MS/MS spectra, GNPS molecular networking can associate structurally related molecules that exhibit similar MS/MS fragmentation patterns into molecular families (Watrous, et al., 2012; Yang, et al., 2013).

Comparative metabolomics and chemical diversity of large scale datasets

Few tools allow assessment of the chemical diversity within large and diverse MS datasets such as those in the current study (Bouslimani, et al., 2014; Charlop-Powers, et al., 2014; Luzzatto-Knaan, et al., 2015; Purves, et al., 2016). To determine whether this collection of marine cyanobacterial and algal communities possessed unique chemical diversity, the LC-MS/MS data were analyzed with publicly available data sets accessed via the GNPS-MassIVE database (For datasets see methods) (Wang, et al., 2016). These datasets include LC-MS/MS data from terrestrial and marine actinobacteria (1000 samples), lichens (132 samples), marine sponges and corals (260 samples) and freshwater cyanobacteria (535 samples). Collection of all datasets was acquired using tandem mass spectrometry to obtain unique fragmentation fingerprints for each detected molecule.

To evaluate the chemical diversity of these source materials, we employed diverse MS-based informatic approaches, including multivariate analyses and molecular networking. Multivariate approaches, such as principal component analysis (PCA) and partial least squares-discriminant analysis, (PLS-DA), have been widely employed to assess the chemical diversity of metabolomics data (Worley and Powers, 2013). Here, for comprehensive metabolomics analysis, we applied principal coordinate analysis (PCoA), a statistical method that provides an overview of the dissimilarities between samples based on the metabolome of an individual sample relative to all other samples in the analysis (He, et al., 2015) The relatedness between samples is calculated by the dissimilarity distance matrix and displayed in a three-dimensional plot where each sphere represents a sample with specific MS/MS features. We utilized the Bray-Curtis distance matrix applied to all MS/MS features, as exported from GNPS, using the

QIIME platform (Caporaso, et al., 2010). The separation of the samples in PCoA space highlights the differences between the various origins of the samples (Figure 1A). The lichen samples separate from the marine associated samples, although they overlap the spatial distributions of some cyanobacteria populations, perhaps because most lichens contain cyanobacteria, bacteria and fungi (Mushegian, et al., 2011). Interestingly, among the five environmental samples, the marine cyanobacteria/algae and freshwater cyanobacteria are more similar in their metabolomics profiles than any other of the source materials, even though the two aquatic environments are quite distinct (Figure 1A). Nevertheless, while presenting the highest similarity compared to the other datasets, the marine cyanobacteria/algae and freshwater cyanobacteria samples show a clear differentiation in their chemical inventories (Figure 1B). To quantify the common and differential chemical features driving these distances, we generated a Venn diagram using the MS/MS features. (Figure 1C,). This analysis revealed that only 13.7% of the marine cyanobacteria/algae molecular features overlapped with other datasets, highlighting that 86.3% were features unique to these collections. Among the molecular commonalities, a number of the mass spectrometry signals were identified as matches to reference spectra in GNPS. These included a mixture of primary metabolites such as amino acids, fatty acids, structural lipids, pheophytin and pheophorbide A (chlorophyll derived products), as well as common mass spectrometry background signals such as formate clusters, plasticizers and polymers. Remarkably, despite being most similar by PCoA analysis, no reference MS/MS spectra of the known freshwater cyanobacterial natural products matched with our marine cyanobacteria and algal collection dataset.

In exploring the chemical diversity of the marine cyanobacteria and algae communities, we hypothesized that either the organism collected (cyanobacterial or algal) or the geographical

location was responsible for sample-to-sample metabolomics differences (Figure 2). Our PCoA analysis did not indicate differentiation based on field annotations as algal versus cyanobacterial (Figure 2A), and the non-significant clustering based on geographical origin, highlights the vast chemical diversity of samples within the same region (Figure 2B). PCA, the most common dimensionality reduction and visualization method used for metabolomics analysis, displayed a parallel outcome with minor trends being observed (Figure 1-figure supplement 1, Figure 2-figure supplement 1).

Construction of fraction libraries has become a common strategy to make more efficient the discovery of bioactive molecules in natural products studies so as to reduce sample complexity for biological screening as well as increase the titer of minor constituents (Bugni, et al., 2008; Harvey, et al., 2015). In the current analysis, the fraction library yielded over five times the number of molecular features compared to the crude extracts, thereby enabling a more thorough assessment of chemical richness and diversity (Figure 3A). The fraction library resulted in 1094 molecular families compared to 132 in the crude extract analysis. This tenfold difference highlights the advantage of using fraction libraries for chemical and biological assay screens. By comparing the number of features as well as the number of underlying spectra for a given molecular family (Nguyen, et al., 2013), we could identify particular chemical scaffolds that are more diverse in their structural variations and are more abundant in our collections. For example, the barbamide molecular family (Figure 3B) is comprised of 40 molecular features represented in 985 individual spectra and found in 75 collections. By contrast, the palmyrolide A molecular family has only 2 associated features and consists of 32 spectra that are found in 5 collections (Figure 3B). Using a rarefaction analysis, both the crude extracts and the fraction

library were found to level off under the current experimental conditions of analysis, indicating that the sample collections are saturating the chemical diversity of these materials (Figure 3A).

Mapping the distribution of marine natural products

To identify locations of high chemical diversity among sampling sites represented in this study (locations with more unique natural products compared to other regions), we compared the MS/MS features found only in a single location to the overall diversity of the entire dataset (Figure 3C, 3D). To do this, we normalized the contribution of features from a single location to a percentage contribution as some locations were sampled more frequently than others based on metadata information collected across 3 decades (Figure 3C). For example, approximately 4% of all MS/MS features contributing to the overall diversity are from Panama-Bocas del Toro (PAB) whereas about 7-8% are each contributed from Panama-Portobelo (PAP) and Palmyra Atoll (PAL), and 20% are from Papua New Guinea (PNG) (Figure 3D).

To explore the chemical nature of differentially produced metabolites in the marine cyanobacterial and algal sample collections, GNPS based molecular networking was used and the network was visualized in Cytoscape (For molecular network and Cytoscape file see methods, , Figure 3-figure supplement 1) (Purves, et al., 2016; Shannon, et al., 2003; Watrous, et al., 2012). As MS/MS spectral alignment detects molecular features, a set of structurally related features connected by edges is termed a 'molecular family'. Experimental MS/MS spectra were dereplicated (also described as 'finding known unknowns' by the metabolomics community) against the community contributed mass spectral 'GNPS library' and third party spectral libraries including MassBank, HMDB, ReSpect and NIST 2014 housed within GNPS analysis infrastructure (Horai, et al., 2010; Johnson and Lange, 2015; Sawada, et al., 2012;

Wang, et al., 2016; Wishart, et al., 2013). This is currently the largest molecular network generated and only possible through the GNPS interface (Wang, et al., 2016). From the more than 15.6 million spectra collected, 6,249 MS/MS spectra were dereplicated to existing publicly accessible reference data, with 91 of the matches to previously characterized cyanobacterial and algal derived natural products from 30 molecular families (Nguyen, et al., 2013). The spectral library matches from the marine cyanobacteria/algae collection included: 74 matches to the GNPS library, 7 matches to the NIH pharmacologically active compounds library, 3 matches to the NIH natural product library, 1 match to the NIH clinical collection, 1 match to the FDA natural products library and 5 matches to the Faulkner legacy library. This is a match rate of about 0.04% which is much lower than database matches of the average untargeted metabolomics experiments (1.8%, (Wang, et al., 2016)) and metabolomics analyses of model samples such as *E. coli*, human cell lines, mice or humans (da Silva, et al., 2015). The low match rate to existing natural product libraries within GNPS highlights that these marine cyanobacterial/algal communities are significantly underexplored from a chemical perspective.

The hundreds of unannotated molecular families from these samples suggest that they contain a significant number of natural products that remain uncharacterized. Molecular networking allows putative structural propagation through a molecular family because most molecules of similar structure fragment similarly. By investigating selected molecular families, we identified previously characterized metabolites and previously undescribed derivatives (Supplementary file 1 and Figure 3-figure supplement 2/3). For example, we identified the spectral cluster representing the apratoxin molecular family by a spectral match to the apratoxin B library standard. Further dereplication of this molecular family using literature information and manual analysis of MS/MS spectra revealed several known apratoxin derivatives (A, D, F, G, H)

(Figure 3-figure supplement 2) (Luesch, et al., 2001; Tidgewell, et al., 2010; Watrous, et al., 2012; Yang, et al., 2013). Additional apratoxin derivatives are suggested on the basis of mass differences of common functional groups (Figure 3-figure supplement 2). Further, GNPS identified a molecular family comprised of the barbamides (Orjala and Gerwick, 1996). Within this molecular family, we were able to deduce several candidate derivatives based on fragmentation patterns and presence of chlorine isotopes (Figure 3-figure supplement 1 and Figure 3-figure supplement 3 (1)) (Orjala and Gerwick, 1996; Yang, et al., 2013). For example, a 14 Da decrease from the y ion of barbamide most likely results from loss of the amide methyl group, leading to *N*-desmethyl barbamide (Figure 3-figure supplement 3 (2)). *O*-desmethyl barbamide was previously published and is identifiable by the loss of 14 Da from the characteristic MS/MS fragment containing the three chlorine atoms (m/z 242.98) (Kim, et al., 2012). The 34 Da reduction from the y ion appears to be associated with the substitution of the Phe group by Leu, a conservative hydrophobic replacement for A-domains that accept Phe in the NRPS-based biosynthesis of barbamide. Such a conservative substitution has previously been observed in other NRPS biosynthetic systems (Figure 3-figure supplement 3 (3)) (Flatt, et al., 2006). This exchange of amino acids gives rise to a new molecular entity; however, it is reminiscent of related compounds, such as dysidin, which was isolated from the cyanobacteria-harboring sponge *Dysidea herbacea* (Ilardi and Zakarian, 2011).

Because a majority (55%, Figure 3D) of molecules detected in this dataset were found in more than one sampling location, we explored the spatial distribution and abundance of specific natural products using binning networks. For this visualization, the abundance/presence of a metabolite was correlated to the 8 collection sites. In the network, each node represents a distribution pattern of molecules, where the node size and color correspond to the relative

number of features that share the same distribution (Figure 4). Of the 256 conceivable regiospecific patterns (based on the 8 collection sites) some molecules share similar distribution patterns, while some distribution patterns are absent or very minor in their abundance (light colored nodes). To visualize the distribution of select molecules identified via molecular networking, in our 8 major sampling sites, MS/MS feature intensity values were projected onto a 2D geographical map based on GPS coordinates using the open-source tool 'ili (an open-source tool for 2D and 3D data visualization created by our laboratories, <https://github.com/ili-toolbox/ili>, available as a Google Chrome app) (Figure 4-figure supplement 1). For example, out of the 33,481 detected features, dolastatin 10 along with 72 other chemical features share the same widespread distribution pattern (Curaçao (NAC), PAB, Panama-Coiba (PAC), PAL, PAP, PNG), while palmyramide A and 3033 other chemical features are endemic to PAL, at least within the scope of this data set, and are not found in any of the other sampling locations. Additionally, while barbamide is widely distributed, putative analogs are encountered relatively infrequently (Figure 4B, Figure 4-figure supplement 1 and Figure 3-figure supplement 3).

Discovery of new natural products

To emphasize the application of MS-based geographic network analyses for the discovery of uncharacterized natural products, we examined several chemical features from locations of high chemical diversity, namely the PAP, PAL and PNG sites. One sample was chosen from PAP and targeted for the isolation of an m/z 536.333 molecule from a cluster with no described family members (For molecular network see methods, molecular family #483). This specific molecular feature was observed in several cyanobacterial collections of *Lyngbya* from PAP obtained in different years from 2007-2013 (collections in Dec 2007, Sept 2010 and

Jan 2013). The MS/MS data indicated that this compound was peptidic and contained the amino acids glycine, valine (or *N*-methyl glycine), leucine or isoleucine, and 2-hydroxyvaleric acid (Hiv). The presence of this latter residue suggested it was the product of a non-ribosomal peptide synthetase (NRPS) (Kehr, et al., 2011). An additional residue was observed corresponding to the fragment mass m/z 149.094, compatible with the formula $C_{10}H_{13}O^+$. Further analysis of the sequential fragment losses suggested a putative structure of [Gly-Ile/Leu-Hiv-Val- $C_{10}H_{13}O$] $^+$. A search for a putative peptide with a molecular mass m/z 536.333 in various chemical databases (Blunt, et al., 2015; Blunt, et al., 2012; Luzzatto-Knaan, et al., 2015; MarinLit) was unsuccessful, suggesting that this natural product was uncharacterized.

In order to confirm the mass spectrometry-based partial structural assignment, we isolated and structurally characterized this natural product of mass m/z 536.333 (Figure 5, Figure 5-figure supplement 9, Supplementary file 2 (**4**)). Isolation of compound **4**, given the common name of 'yuvalamide A', was guided by LC-MS analyses. NMR analysis (DMSO- d_6) of the pure compound confirmed the presence of Gly, Val, Ile, and hydroxyisovaleric acid, and revealed the presence of 2,2-dimethyl-3-hydroxy-7-octynoic (Dhoya) as the $C_{10}H_{13}O^+$ fragment (Figure 5, Figure 5-figure supplement 1/2, Supplementary file 3). These NMR-based determinations correlate well with the structural predictions obtained by mass spectrometry.

The other chemical entities in the yuvalamide A molecular family were present at very low abundance, preventing their isolation and structural analysis by NMR. However, based on the structure of yuvalamide A and its fragmentation pattern by LC-ESI-MS/MS, we deduced putative structures of these co-metabolites. An analog of yuvalamide A (Figure 5-figure supplement 3 compound **5**, yuvalamide B, $[M+H]^+$ m/z 538.3487, $[M+Na]^+$ m/z 560.3320) was

detected by HRMS. Its fragmentation was nearly identical to compound **4**, differing only in the fragment belonging to Dhoya, suggesting this to be the alkene analog 2,2-dimethyl-3-hydroxy-7-octenoic acid (Dhoea). Co-occurring analogs possessing the Dhoya or Dhoea unit have been observed previously in marine cyanobacterial natural extracts. The Dhoya subunit is usually the more abundant of the two, as observed in this case for yuvalamide A (Boudreau, et al., 2012; Han, et al., 2005; Han, et al., 2011; Luesch, et al., 2001; Sitachitta, et al., 2000; Wan and Erickson, 2001). Moreover, the geographical distribution of yuvalamide B is the same as yuvalamide A (Figure 5), supporting the proposal that these share a common biosynthetic origin. Two additional yuvalamide analogs were observed and tentative structures were assigned based on MS (Figure 5-figure supplement 3).

Discussion

Over the last 40 years, a number of academic natural product discovery programs have generated unique and significant collections from a variety of sources. However, the samples as well as the associated collection and characterization information is often buried in individual laboratory materials and documentation, and frequently this disappears with academic turnovers. To explore a new paradigm for characterizing and preserving such information over a longer time period for the community at large, we harnessed recent advances and metabolomics approaches to deeply annotate the metabolomes of these collections. We then evaluated this rich dataset along with collection information for its diversity, distribution and discovery of new natural products.

To date, biodiversity studies have mostly relied on available genomic information (Bowen, et al., 2013; Passarini, et al., 2014). While organisms might be phylogenetically related, their metabolomic repertoire may vary greatly. The detection of biosynthetic gene clusters indicates natural product potential, and has been successfully employed to predict new molecules, especially of the core structure (Medema and Fischbach, 2015; Medema, et al., 2015; Owen, et al., 2015; Walsh, 2015). However, post assembly modifications (e.g. halogenations, oxidations, methylations) are often times difficult to predict without the structures in hand, and therefore genomic analyses do not provide a complete assessment of the secondary metabolomes. Chemical diversity represents the expression of that genetic potential and offers a new strategy to explore differences and to evaluate the diversity encrypted in complex samples and various environments. Herein, we broadly inventoried the molecules that are present in a large library of field collected marine algae and cyanobacteria using LC-MS/MS. Using the mass spectrometry detectable natural products from each sample, the chemodiversity and the regional patterns of these samples were characterized.

It is reported in the literature that marine cyanobacteria are an exceptional source for diverse new natural products (Blunt, et al., 2015; Burja, et al., 2001; Engene, et al., 2011; Kiuru, et al., 2014; Tan, 2007). However, this perception is difficult to assess quantitatively in comparison to other marine or terrestrial life forms. To gain insights into how unique are the secondary metabolites of marine cyanobacteria and algae, we compared these MS data with other metabolomic data sets that are available from the public domain.

The mass spectrometry data from marine cyanobacteria/algae collections obtained over the past 30 years indicate that, at a global level, this group is rich in metabolite features that do

not overlap with other available metabolomic data sets (Figure 1C). Common multivariate analyses used for metabolomics were successful in distinguishing the different datasets based on their unique metabolomic inventories.

Key to a chemical pattern analysis is the ability to organize molecules on the basis of structural features along with their associated metadata attributes. GNPS molecular networking generates molecular features that are based upon fragmentation similarity, thereby providing resolution at the individual molecular level. This approach also enables the rapid matching of spectra to known molecules by comparison with various natural product MS/MS libraries (Nguyen, et al., 2016; Wang, et al., 2016; Yang, et al., 2013). In this regard, the MS/MS spectrum of a molecule is a unique molecular signature or barcode. When structurally similar molecules are fragmented, the resulting MS/MS spectra are very often quite similar because they contain similarly positioned labile bonds (Wang, et al., 2016). Using molecular networking, a visual representation of these “barcodes” is produced by grouping similar patterns. Molecular networking enables the evaluation of the MS/MS based detectable chemical space, and groups molecules that have similar biosynthetic origins and resultant structures. At the organism level (e.g. invertebrates, plants, corals), marine environments such as the Papua New Guinea archipelago (PNG) are considered to be biodiversity “hotspots” (Bowen, et al., 2013). However, not only is PNG a biodiversity hotspot, molecular networking quantitatively revealed that this region is rich in the presence of unique natural products compared to other sampled locations in this study (Figure 3) (Bowen, et al., 2013; La Barre, 2014). Although this is demonstrated here for these marine algal and cyanobacterial collections, it would be interesting to evaluate if high biodiversity always correlates with high chemical diversity, and as a corollary, how chemical diversity changes upon a shift in biodiversity. The geographical patterns of molecule distribution

could be further investigated via the tools presented here, given the appropriate sampling strategy. While molecular networking revealed the locations with lowest and highest chemical diversity, visualizing these data on a world map enabled assessment of the distribution patterns of specific molecules within the eight sampled regions. This analysis also enabled targeting samples with higher abundance of molecules of interest, and may therefore assist in characterizing exceptionally high-producers of natural products. We have demonstrated that some chemical entities were found in common to multiple locations (e.g. barbamide, dolastatin 10) whereas others appeared to be unique and endemic to specific sampling sites (e.g. hoiamide B, palmyramide) (Figure 4). This observation is supported by site resampling over several years, an insight gained as a distinctive attribute of this unique collection of marine samples.

Additional tracing of the metadata associated with a given molecular family can assist in identification of the producing organism, and facilitate future biosynthetic and ecological investigations (Coates, et al., 2014; Harvey, et al., 2015; Kleigrew, et al., 2015; Leal, et al., 2016; Owen, et al., 2013). Finally, we demonstrated the utility of this overall approach through the discovery and subsequent structural elucidation and geographical distribution of yuvalamide A (**4**), reported here for the first time.

Conclusion

The integration of molecular networking with geographical mapping revealed that marine algae and cyanobacteria have widely varying metabolite distributions. As we can see in this dataset most molecules are broadly distributed in disparate physical locations, whereas others are highly specific to particular locations. Whether this specificity to a given location is due to

unique species with distinct metabolic capabilities, or to specific environmental factors regulating expression of these compounds, is unknown at present. Moreover, field collections are a complex mixture of microorganisms that hold a unique repertoire of natural products that may not be discoverable under laboratory conditions, even with sequenced strains. Therefore, our approach is uniquely capable of organizing the vast chemical diversity present in these complex environmental samples. Further, this approach could be used to identify locations with a higher level of chemical diversity as well as distribution patterns of natural products, potentially giving insights into their ecological roles. In turn, these perceptions may facilitate the search for novel natural products of utility to biomedicine and biotechnology.

Materials and Methods

Collection and culturing: A total of 317 diverse marine cyanobacterial and benthic algae collections were hand-collected in Panama, Papua New Guinea, Hawaii, Madagascar, Palmyra Atoll, Curaçao, and a few other tropical marine sites at depths from 0.3-20 m with the aid of snorkel or scuba gear (Supplementary file 1). GPS location, preliminary field taxonomic identification of the specimens, and depth of collection were noted for most collections. Samples for subsequent chemical extraction were strained through a mesh bag to remove excess seawater, preserved in equal volumes of seawater and ethanol, transported to the laboratory, and stored at -4 °C or -20 °C until workup. Live samples of marine cyanobacteria were brought back to the laboratory in vented tissue culture flasks with 0.2 µm filtered native seawater and subsequently isolated to mono-cultures. These were cultured in SWBG-11 media with 35 g/L

Instant Ocean (Aquarium Systems Inc.). Mono-cultures were grown at 28 °C in a 16 h light/8 h dark cycle with a light intensity of ~7 $\mu\text{mol photon/s/m}^2$ provided by 40 W cool white fluorescent lights. Specific collection information is described in the Supporting Information.

Extraction Library: Chemistry samples were extracted repetitively with CH_2Cl_2 :MeOH 2:1, dried *in vacuo*, and in most cases, fractionated into nine fractions (A-I) by silica gel vacuum liquid chromatography (VLC) using a stepwise gradient of hexanes/EtOAc and EtOAc/MeOH. Each fraction and crude extract was re-suspended at 5 mg/mL in pure DMSO and stored in 96-well plates at -20 °C until analysis.

UHPLC-MS/MS analysis: The extracted metabolites were analyzed with an UltiMate 3000 UHPLC system (Thermo Scientific) using a KinetexTM 1.7 μm C18 reversed phase UHPLC column (50 X 2.1 mm) and Maxis Impact Q-TOF mass spectrometer (Bruker Daltonics) equipped with an ESI source. The data were acquired in positive ionization mode for parent mass (MS^1) and tandem MS for molecular fragmentation (MS^2). Amount of material injected from each sample was 3.3 μg . The data were acquired in positive ionization mode for parent mass (MS^1) and tandem MS for molecular fragmentation (MS^2). The gradient employed for chromatographic separation was 5% solvent B (ACN/ H_2O /formic acid 98%/2%/0.1%) with solvent A (H_2O /ACN/formic acid 98%/2%/0.1%) for 1.5 min, a step gradient of 5% to 50% B in 0.5 min, held at 50% B for 2 min, a second gradient of 50%-100% B in 6 min, held at 100% B for 0.5 min, 100%-5% B in 0.5 min and kept at 5% B for 0.5 min at a flow rate of 0.5 mL/min throughout the run. The MS analysis was performed on a Maxis QTOF mass spectrometer (Bruker Daltonics), controlled by the Otof Control and Hystar software packages (Bruker

Daltonics), and equipped with ESI source. MS spectra were acquired as previously described (Garg, et al., 2014).

Molecular networking: A molecular network was created using the online workflow at GNPS (Wang and al., 2016). To de-noise, the data was filtered by removing all MS/MS peaks within +/- 17 Da of the precursor m/z . MS/MS spectra were filtered by choosing only the top 6 peaks in the +/- 50 Da window throughout the spectrum. The data were clustered with MS-Cluster with a parent mass tolerance of 1.0 Da and a MS/MS fragment ion tolerance of 0.5 Da to create consensus spectra. Consensus spectra that contained less than 3 spectra were discarded. A network was then created where edges were filtered to have a cosine score above 0.6 and more than 4 matched peaks. Further edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. The spectra in the network were then searched against GNPS spectral libraries. The library spectra were filtered in the same manner as the input data. The networking output was visualized using Cytoscape 2.8.3 free software (Shannon, et al., 2003) and organized using the FM3 force directed layout (Hachul and Junger, 2004). Follow the link to the GNPS molecular network of multiple datasets: <http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=e69c25e3fbcf4c09b6e0c75633565f95>. Links to the cyanobacteria/algae datasets: <ftp://massive.ucsd.edu/MSV000078568>, <ftp://massive.ucsd.edu/MSV000078892> and the corresponding network on GNPS: <http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=0837888b3dab43efa5bf9a50254c7c8f>. For the Cytoscape file <ftp://massive.ucsd.edu/MSV000078568/other/>.

461

462

463

464 **Statistics and diversity analysis:** The MS/MS data tables generated by GNPS for all analyzed
465 extracts and fractions was converted to a BIOM table (<http://biom-format.org/>) (McDonald, et al.,
466 2012). Each of the tables was then used for calculation the sample-sample distance metrics
467 using binary metric (Binary-Jaccard) using QIIME (Caporaso, et al., 2010) version 1.9. Principal
468 coordinates were calculated and visualized based on various metadata categories (region,
469 organism etc.) using EMPeror (Vazquez-Baeza, et al., 2013). Rarefaction curves of 20 random
470 iterations were generated by GNPS.

471 **Spatial mapping of mass spectrometry data:** Spatial molecular maps were created with
472 background map from a Google Map screenshot with several regions of interest enlarged and
473 plotted as insets to the overview map. By visual examination, we assigned to each sampling
474 spot the (x,y) pixel coordinates on the background map based on GPS coordinates taken at the
475 time of collection. For multiple samples from the same location, spots were placed in adjacent
476 locations. For each node of the molecular network generated in GNPS, we considered the area
477 under the curve of all mapped spots, scaled them from 0 to 100% and assigned colors to all
478 intensities by using the “jet” color map, color-coding low intensities with blue, high intensities
479 red, and other intensities with a colored gradient between blue and red. The rendering of the
480 maps was performed by the ‘ili tool for 2D and 3D spatial mapping ([https://github.com/ili-](https://github.com/ili-toolbox/ili)
481 [toolbox/ili](https://github.com/ili-toolbox/ili)). The color was made gradually disappearing with highest intensity at the center of the
482 sampling location and almost no intensity at the boundary of the spot. For this, a logarithmic

combination of the default color and assigned color was used, with the coefficient exponentially depending on the distance from the sampling spot center.

Structure elucidation of yuvalamide A (4): A preliminary structure was deduced based on a HRMS ($C_{28}H_{45}N_3O_7$, 8° unsaturation) and HRMS/MS (Bruker Daltonics Maxis Impact qTOF) fragment analysis. Exact masses observed were $[M+H]^+$ m/z 536.3335 and $[M+Na]^+$ m/z 558.3155 corresponding to a molecular formula of $C_{28}H_{45}N_3O_7$ (<0.9 ppm). Neutral losses of glycine, leucine/isoleucine and valine were observed. An additional loss matching the formula of 2-hydroxyisovaleric acid (Hiv) was observed suggesting yuvalamide was a non-ribosomal peptide. This MS-based fragment analysis suggested a Gly-Ile/Leu-Hiv-Val sequence with an additional $C_{10}H_{13}O$ fragment. The peptide appeared to be cyclic based on degrees of unsaturation, and by database query, was a new compound.

Analysis of 2D NMR data (gCOSY, TOCSY, HSQC, HMBC) confirmed three of the five substructures (Val, Hiv, Gly) and clarified the structures of the remaining two (Figure 5-figure supplement 9, Supplementary file 2). The Ile/Leu substructure was identified as Ile from COSY correlations between the beta protons at 1.96 ppm and a methyl group at δ 0.86, the beta protons at δ 1.96 and a pair of germinal protons at δ 1.49 and 1.13, and between the geminal protons to a second methyl group at δ 0.81. Finally, a 2,2-dimethyl-3-hydroxyoctynoic acid (Dhoya) residue was similarly identified. The *gem*-dimethyl arrangement was revealed by key HMBC correlations from CH_3 -9/10_{Dhoya} to the carbonyl carbon C-1_{Dhoya}, quaternary carbon C-2_{Dhoya}, and oxygenated methine CH-3_{Dhoya}. This latter resonance was connected by COSY to a linear sequence of three methylene groups, the terminal methylene of which was long range coupled to alkynyl carbons (δ_C 71.7 and 84.5), one of which (δ_C 84.5) showed a HMBC

correlation to a terminal alkynyl proton (δ_{H} 2.78). The assembly of these substructures into an overall cyclic depsipeptide structure was enabled by a key HMBC correlation between Dhoya H3 (δ_{H} 4.71) and the Gly carbonyl (δ_{C} 168.6) (Figure 5-figure supplement 9). Additional HMBC correlations linking all the substructures verified the MS/MS-deduced sequence described above (Figure 5-figure supplement 9).

Acknowledgements

The authors would like to thank Dr. Gabriel Navarro for his help with the NMR structure elucidation and Dr. Vanessa V. Phelan for her valuable insights on the manuscript. TLK was supported, in part, by BARD, the United States – Israel Binational Agricultural Research and Development Fund, Vaadia-BARD Postdoctoral Fellowship Award no. FI-494-13. This work was supported by NIH grant GM107550 to PCD, LG and WHG and European Union FP7 and H2020 programmes grants No. 305259 and 634402. The authors declare no conflict of interest.

References

- Bai, R., Pettit, G.R., and Hamel, E. (1990). Dolastatin 10, a powerful cytostatic peptide derived from a marine animal. Inhibition of tubulin polymerization mediated through the vinca alkaloid binding domain. *Biochem Pharmacol* 39, 1941-1949.
- Blokhin, A.V., Yoo, H.D., Gerald, R.S., Nagle, D.G., Gerwick, W.H., and Hamel, E. (1995). Characterization of the interaction of the marine cyanobacterial natural product curacin A with the colchicine site of tubulin and initial structure-activity studies with analogues. *Mol Pharmacol* 48, 523-531.

527 Blunt, J.W., Copp, B.R., Keyzers, R.A., Munro, M.H., and Prinsep, M.R. (2015). Marine natural
 528 products. *Nat Prod Rep* 32, 116-211.

529 Blunt, J.W., Copp, B.R., Keyzers, R.A., Munro, M.H.G., and Prinsep, M.R. (2012). Marine
 530 natural products. *Nat Prod Rep* 29, 144-222.

531 Boeuf, G., and Kornprobst, J.M. (2009). Biodiversity and marine chemodiversity. *Biofutur*, 28-32.

532 Boudreau, P.D., Byrum, T., Liu, W.T., Dorrestein, P.C., and Gerwick, W.H. (2012). Viequeamide
 533 A, a cytotoxic member of the kulolide superfamily of cyclic depsipeptides from a marine
 534 button cyanobacterium. *J Nat Prod* 75, 1560-1570.

535 Bouslimani, A., Sanchez, L.M., Garg, N., and Dorrestein, P.C. (2014). Mass spectrometry of
 536 natural products: current, emerging and future technologies. *Nat Prod Rep*.

537 Bowen, B.W., Rocha, L.A., Toonen, R.J., Karl, S.A., and Lab, T. (2013). The origins of tropical
 538 marine biodiversity. *Trends Ecol Evol* 28, 359-366.

539 Bugni, T.S., Richards, B., Bhoite, L., Cimbora, D., Harper, M.K., and Ireland, C.M. (2008).
 540 Marine natural product libraries for high-throughput screening and rapid drug discovery. *J*
 541 *Nat Prod* 71, 1095-1098.

542 Burja, A.M., Banaigs, B., Abou-Mansour, E., Burgess, J.G., and Wright, P.C. (2001). Marine
 543 cyanobacteria - a prolific source of natural products. *Tetrahedron* 57, 9347-9377.

544 Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K.,
 545 Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of
 546 high-throughput community sequencing data. *Nat Methods* 7, 335-336.

547 Cardellina, J.H., 2nd, Marner, F.J., and Moore, R.E. (1979). Seaweed dermatitis: structure of
 548 lyngbyatoxin A. *Science* 204, 193-195.

549 Charlop-Powers, Z., Owen, J.G., Reddy, B.V., Ternei, M.A., Guimaraes, D.O., de Frias, U.A.,
 550 Pupo, M.T., Seepe, P., Feng, Z., and Brady, S.F. (2014). Global biogeographic sampling of
 551 bacterial secondary metabolism. *Elife* 4, e05048.

552 Chlipala, G.E., Mo, S., and Orjala, J. (2011). Chemodiversity in freshwater and terrestrial
 553 cyanobacteria - a source for drug discovery. *Curr Drug Targets* 12, 1654-1673.

554 Choi, H., Pereira, A.R., Cao, Z., Shuman, C.F., Engene, N., Byrum, T., Matainaho, T., Murray,
 555 T.F., Mangoni, A., and Gerwick, W.H. (2010). The hoiamides, structurally intriguing
 556 neurotoxic lipopeptides from Papua New Guinea marine cyanobacteria. *J Nat Prod* 73,
 557 1411-1421.

558 Coates, R.C., Podell, S., Korobeynikov, A., Lapidus, A., Pevzner, P., Sherman, D.H., Allen,
 559 E.E., Gerwick, L., and Gerwick, W.H. (2014). Characterization of Cyanobacterial
 560 Hydrocarbon Composition and Distribution of Biosynthetic Pathways. *PLoS One* 9, e85140.

561 Crawford, J.M., Portmann, C., Kontnik, R., Walsh, C.T., and Clardy, J. (2011). NRPS substrate
 562 promiscuity diversifies the xenematides. *Org Lett* 13, 5144-5147.

563 da Silva, R.R., Dorrestein, P.C., and Quinn, R.A. (2015). Illuminating the dark matter in
 564 metabolomics. *Proc Natl Acad Sci U S A* 112, 12549-12550.

565 Engene, N., Choi, H., Esquenazi, E., Rottacker, E.C., Ellisman, M.H., Dorrestein, P.C., and
 566 Gerwick, W.H. (2011). Underestimated biodiversity as a major explanation for the perceived
 567 rich secondary metabolite capacity of the cyanobacterial genus *Lyngbya*. *Environ Microbiol*
 568 13, 1601-1610.

Engene, N., Rottacker, E.C., Kastovsky, J., Byrum, T., Choi, H., Ellisman, M.H., Komarek, J.,
and Gerwick, W.H. (2012). *Moorea producens* gen. nov., sp. nov. and *Moorea bouillonii*
comb. nov., tropical marine cyanobacteria rich in bioactive secondary metabolites. *Int J*
Syst Evol Microbiol 62, 1171-1178.

Esquenazi, E., Coates, C., Simmons, L., Gonzalez, D., Gerwick, W.H., and Dorrestein, P.C.
(2008). Visualizing the spatial distribution of secondary metabolites produced by marine
cyanobacteria and sponges via MALDI-TOF imaging. *Mol Biosyst* 4, 562-570.

Flatt, P.M., O'Connell, S.J., McPhail, K.L., Zeller, G., Willis, C.L., Sherman, D.H., and Gerwick,
W.H. (2006). Characterization of the initial enzymatic steps of barbamide biosynthesis. *J*
Nat Prod 69, 938-944.

Garcia-Pichel, F., Sherry, N.D., and Castenholz, R.W. (1992). Evidence for an ultraviolet
sunscreen role of the extracellular pigment scytonemin in the terrestrial cyanobacterium
Chlorogloeopsis sp. *Photochem Photobiol* 56, 17-23.

Garg, N., Kapono, C.A., Lim, Y.W., Koyama, N., Vermeij, M.J.A., Conrad, D., Rohwer, F., and
Dorrestein, P.C. (2014). Mass spectral similarity for untargeted metabolomics data analysis
of complex mixtures. *International Journal of Mass Spectrometry*.

Gerwick, W.H., and Moore, B.S. (2012). Lessons from the past and charting the future of marine
natural products drug discovery and chemical biology. *Chem Biol* 19, 85-98.

Hachul, S., and Junger, M. (2004). Drawing large graphs with a potential-field-based multilevel
algorithm. *Proceedings of the 12th international conference on Graph Drawing*, 285-295.

589 Han, B., Goeger, D., Maier, C.S., and Gerwick, W.H. (2005). The wewakpeptins, cyclic
 590 depsipeptides from a Papua New Guinea collection of the marine cyanobacterium *Lyngbya*
 591 *sempilena*. *J Org Chem* 70, 3133-3139.

592 Han, B., Gross, H., McPhail, K.L., Goeger, D., Maier, C.S., and Gerwick, W.H. (2011).
 593 Wewakamide A and guineamide G, cyclic depsipeptides from the marine cyanobacteria
 594 *Lyngbya semiplena* and *Lyngbya majuscula*. *J Microbiol Biotechnol* 21, 930-936.

595 Harvey, A.L., Edrada-Ebel, R., and Quinn, R.J. (2015). The re-emergence of natural products
 596 for drug discovery in the genomics era. *Nat Rev Drug Discov* 14, 111-129.

597 Harrigan, G.G., Luesch, H., Yoshida, W.Y., Moore, R.E., Nagle, D.G., Biggs, J., Park, P.U., and
 598 Paul, V.J. (1999). Tumonoic acids, novel metabolites from a cyanobacterial assemblage of
 599 *Lyngbya majuscula* and *Schizothrix calcicola*. *J Nat Prod* 62, 464-467.

600 Harrigan, G.G., Yoshida, W.Y., Moore, R.E., Nagle, D.G., Park, P.U., Biggs, J., Paul, V.J.,
 601 Mooberry, S.L., Corbett, T.H., and Valeriote, F.A. (1998). Isolation, structure determination,
 602 and biological activity of dolastatin 12 and lyngbyastatin 1 from *Lyngbya*
 603 *majuscula*/*Schizothrix calcicola* cyanobacterial assemblages. *J Nat Prod* 61, 1221-1225.

604 He, Y., Caporaso, J.G., Jiang, X.-T., Sheng, H.-F., Huse, S.M., Rideout, J.R., Edgar, R.C.,
 605 Kopylova, E., Walters, W.A., Knight, R., et al. (2015). Stability of operational taxonomic
 606 units: an important but neglected property for analyzing microbial diversity. *Microbiome* 3,
 607 20.

608 Hooper, G.J., Orjala, J., Schatzman, R.C., and Gerwick, W.H. (1998). Carmabins A and B, new
 609 lipopeptides from the Caribbean cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 61, 529-
 610 533.

611 Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K., Tanaka,
612 S., Aoshima, K., et al. (2010). MassBank: a public repository for sharing mass spectral data
613 for life sciences. *Journal of Mass Spectrometry* 45, 703-714.

614 Ilardi, E.A., and Zakarian, A. (2011). Efficient total synthesis of dysidenin, dysidin, and
615 barbamide. *Chem Asian J* 6, 2260-2263.

616 Johnson, S.R., and Lange, B.M. (2015). Open-access metabolomics databases for natural
617 product research: present capabilities and future potential. *Front Bioeng Biotechnol* 3, 22.

618 Kehr, J.C., Gatte Picchi, D., and Dittmann, E. (2011). Natural product biosyntheses in
619 cyanobacteria: A treasure trove of unique enzymes. *Beilstein J Org Chem* 7, 1622-1635.

620 Kim, E.J., Lee, J.H., Choi, H., Pereira, A.R., Ban, Y.H., Yoo, Y.J., Kim, E., Park, J.W., Sherman,
621 D.H., Gerwick, W.H., et al. (2012). Heterologous production of 4-O-demethylbarbamide, a
622 marine cyanobacterial natural product. *Org Lett* 14, 5824-5827.

623 Kiuru, P., D'Auria, M.V., Muller, C.D., Tammela, P., Vuorela, H., and Yli-Kauhaluoma, J. (2014).
624 Exploring marine resources for bioactive compounds. *Planta Med* 80, 1234-1246.

625 Kleigrew, K., Almaliti, J., Tian, I.Y., Kinnel, R.B., Korobeynikov, A., Monroe, E.A., Duggan,
626 B.M., Di Marzo, V., Sherman, D.H., Dorrestein, P.C., et al. (2015). Combining Mass
627 Spectrometric Metabolic Profiling with Genomic Analysis: A Powerful Approach for
628 Discovering Natural Products from Cyanobacteria. *J Nat Prod* 78, 1671-1682.

629 La Barre, S. (2014). Marine Biodiversity and Chemodiversity — The Treasure Troves of the
630 Future. In Biodiversity - The Dynamic Balance of the Planet, O. Grillo, ed. (InTech), pp. 1-
631 26.

632 Leal, M.C., Hilario, A., Munro, M.H., Blunt, J.W., and Calado, R. (2016). Natural products
633 discovery needs improved taxonomic and geographic information. Nat Prod Rep.

634 Luesch, H., Moore, R.E., Paul, V.J., Mooberry, S.L., and Corbett, T.H. (2001). Isolation of
635 dolastatin 10 from the marine cyanobacterium *Symploca* species VP642 and total
636 stereochemistry and biological evaluation of its analogue symplostatin 1. J Nat Prod 64,
637 907-910.

638 Luesch, H., Pangilinan, R., Yoshida, W.Y., Moore, R.E., and Paul, V.J. (2001). Pitipeptolides A
639 and B, new cyclodepsipeptides from the marine cyanobacterium *Lyngbya majuscula*. J Nat
640 Prod 64, 304-307.

641 Luesch, H., Yoshida, W.Y., Moore, R.E., Paul, V.J., and Corbett, T.H. (2001). Total structure
642 determination of apratoxin A, a potent novel cytotoxin from the marine cyanobacterium
643 *Lyngbya majuscula*. J Am Chem Soc 123, 5418-5423.

644 Luzzatto-Knaan, T., Melnik, A.V., and Dorrestein, P.C. (2015). Mass spectrometry tools and
645 workflows for revealing microbial chemistry. Analyst 140, 4949-4966.

646 MarinLit. Database. (<http://www.chem.canterbury.ac.nz/marinlit/marinlit.shtml>, Department of
647 Chemistry, University of Canterbury).

648 Marquez, B.L., Watts, K.S., Yokochi, A., Roberts, M.A., Verdier-Pinard, P., Jimenez, J.I., Hamel,
 649 E., Scheuer, P.J., and Gerwick, W.H. (2002). Structure and absolute stereochemistry of
 650 hectochlorin, a potent stimulator of actin assembly. *J Nat Prod* 65, 866-871.
 651

652 McDonald, D., Clemente, J.C., Kuczynski, J., Rideout, J.R., Stombaugh, J., Wendel, D., Wilke,
 653 A., Huse, S., Hufnagle, J., Meyer, F., et al. (2012). The Biological Observation Matrix
 654 (BIOM) format or: how I learned to stop worrying and love the ome-ome. *Gigascience* 1, 7.

655 Medema, M.H., and Fischbach, M.A. (2015). Computational approaches to natural product
 656 discovery. *Nat Chem Biol* 11, 639-648.

657 Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K., de Bruijn, I.,
 658 Chooi, Y.H., Claesen, J., Coates, R.C., et al. (2015). Minimum Information about a
 659 Biosynthetic Gene cluster. *Nat Chem Biol* 11, 625-631.

660 Mevers, E., Matainaho, T., Allara, M., Di Marzo, V., and Gerwick, W.H. (2014). Mooreamide A: a
 661 cannabinomimetic lipid from the marine cyanobacterium *Moorea bouillonii*. *Lipids* 49, 1127-
 662 1132.

663 Mushegian, A.A., Peterson, C.N., Baker, C.C., and Pringle, A. (2011). Bacterial diversity across
 664 individual lichens. *Appl Environ Microbiol* 77, 4249-4252.

665 Nguyen, D.D., Melnik, A.V., Koyama, N., Lu, X., Schorn, M., Fang, J., Aguinaldo, K., Lincecum,
 666 T.L., Jr., Ghequire, M.G., Carrion, V.J., et al. (2016). Indexing the *Pseudomonas*
 667 specialized metabolome enabled the discovery of poaeamide B and the bananamides. *Nat*
 668 *Microbiol* 2, 16197.

669 Nguyen, D.D., Wu, C.H., Moree, W.J., Lamsa, A., Medema, M.H., Zhao, X., Gavilan, R.G.,
 670 Aparicio, M., Atencio, L., Jackson, C., et al. (2013). MS/MS networking guided analysis of
 671 molecule and gene cluster families. *Proc Natl Acad Sci U S A* 110, E2611-2620.

672 Nunnery, J.K., Engene, N., Byrum, T., Cao, Z., Jabba, S.V., Pereira, A.R., Matainaho, T.,
 673 Murray, T.F., and Gerwick, W.H. (2012). Biosynthetically intriguing chlorinated lipophilic
 674 metabolites from geographically distant tropical marine cyanobacteria. *The Journal of*
 675 *Organic Chemistry* 77, 4198-4208.

676 Nunnery, J.K., Mevers, E., and Gerwick, W.H. (2010). Biologically active secondary metabolites
 677 from marine cyanobacteria. *Curr Opin Biotechnol* 21, 787-793.

678 Orjala, J., and Gerwick, W.H. (1996). Barbamide, a chlorinated metabolite with molluscicidal
 679 activity from the Caribbean cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 59, 427-430.

680 Orjala, J., Nagle, D., and Gerwick, W.H. (1995). Malyngamide H, an ichthyotoxic amide
 681 possessing a new carbon skeleton from the Caribbean cyanobacterium *Lyngbya majuscula*.
 682 *J Nat Prod* 58, 764-768.

683 Owen, J.G., Charlop-Powers, Z., Smith, A.G., Ternei, M.A., Calle, P.Y., Reddy, B.V., Montiel,
 684 D., and Brady, S.F. (2015). Multiplexed metagenome mining using short DNA sequence
 685 tags facilitates targeted discovery of epoxyketone proteasome inhibitors. *Proc Natl Acad Sci*
 686 *U S A* 112, 4221-4226.

687 Owen, J.G., Reddy, B.V., Ternei, M.A., Charlop-Powers, Z., Calle, P.Y., Kim, J.H., and Brady,
 688 S.F. (2013). Mapping gene clusters within arrayed metagenomic libraries to expand the
 689 structural diversity of biomedically relevant natural products. *Proc Natl Acad Sci U S A* 110,
 690 11797-11802.

691 Passarini, M.R.Z., Miqueletto, P.B., de Oliveira, V.M., and Sette, L.D. (2014). Molecular diversity
 692 of fungal and bacterial communities in the marine sponge *Dragmacidon reticulatum*. *Journal*
 693 *of Basic Microbiology*, n/a-n/a.

694 Pereira, A., Cao, Z., Murray, T.F., and Gerwick, W.H. (2009). Hoiamide a, a sodium channel
 695 activator of unusual architecture from a consortium of two papua new Guinea
 696 cyanobacteria. *Chem Biol* 16, 893-906.

697

698 Pereira, A.R., Cao, Z., Engene, N., Soria-Mercado, I.E., Murray, T.F., and Gerwick, W.H.
 699 (2010). Palmyrolide A, an unusually stabilized neuroactive macrolide from Palmyra Atoll
 700 cyanobacteria. *Org Lett* 12, 4490-4493.

701 Pereira, A.R., Kale, A.J., Fenley, A.T., Byrum, T., Deboni, H.M., Gilson, M.K., Valeriote, F.A.,
 702 Moore, B.S., and Gerwick, W.H. (2012). The carmaphyccins: new proteasome inhibitors
 703 exhibiting an alpha,beta-epoxyketone warhead from a marine cyanobacterium.
 704 *Chembiochem* 13, 810-817.

705 Pereira, A.R., McCue, C.F., and Gerwick, W.H. (2010). Cyanolide A, a glycosidic macrolide with
 706 potent Molluscicidal activity from the Papua New Guinea cyanobacterium *Lyngbya*
 707 *bouillonii*. *J Nat Prod* 73, 217-220.

708 Purves, K., Macintyre, L., Brennan, D., Hreggviethsson, G.O., Kuttner, E., Asgeirsdottir, M.E.,
 709 Young, L.C., Green, D.H., Edrada-Ebel, R., and Duncan, K.R. (2016). Using Molecular
 710 Networking for Microbial Secondary Metabolite Bioprospecting. *Metabolites* 6.

711 Rastogi, R.P., and Sinha, R.P. (2009). Biotechnological and industrial significance of
 712 cyanobacterial secondary metabolites. *Biotechnol Adv* 27, 521-539.

713 Sawada, Y., Nakabayashi, R., Yamada, Y., Suzuki, M., Sato, M., Sakata, A., Akiyama, K.,
 714 Sakurai, T., Matsuda, F., Aoki, T., et al. (2012). RIKEN tandem mass spectral database
 715 (ReSpect) for phytochemicals: A plant-specific MS/MS-based data resource and database.
 716 *Phytochemistry* 82, 38-45.

717 Schopf, J.W. (2012). The Fossil Record of Cyanobacteria. In *Ecology of cyanobacteria II: their*
 718 *diversity in space and time* (B.A. Whitton, ed.). (Springer: London, UK). pp. 15-38.

719 Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N.,
 720 Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated
 721 models of biomolecular interaction networks. *Genome Res* 13, 2498-2504.

722 Sitachitta, N., Williamson, R.T., and Gerwick, W.H. (2000). Yanucamides A and B, two new
 723 depsipeptides from an assemblage of the marine cyanobacteria *Lyngbya majuscula* and
 724 *Schizothrix* species. *J Nat Prod* 63, 197-200.

725 Tan, L.T. (2007). Bioactive natural products from marine cyanobacteria for drug discovery.
 726 *Phytochemistry* 68, 954-979.

727 Tan, L.T., Goh, B.P., Tripathi, A., Lim, M.G., Dickinson, G.H., Lee, S.S., and Teo, S.L. (2010).
 728 Natural antifoulants from the marine cyanobacterium *Lyngbya majuscula*. *Biofouling* 26,
 729 685-695.

730 Tan, L.T., Okino, T., and Gerwick, W.H. (2013). Bouillonamide: a mixed polyketide-peptide
 731 cytotoxin from the marine cyanobacterium *Moorea bouillonii*. *Mar Drugs* 11, 3015-3024.

732 Taniguchi, M., Nunnery, J.K., Engene, N., Esquenazi, E., Byrum, T., Dorrestein, P.C., and
 733 Gerwick, W.H. (2010). Palmyramide A, a cyclic depsipeptide from a Palmyra Atoll collection
 734 of the marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 73, 393-398.
 735 Tidgewell, K., Engene, N., Byrum, T., Media, J., Doi, T., Valeriote, F.A., and Gerwick, W.H.
 736 (2010). Evolved diversification of a modular natural product pathway: apratoxins F and G,
 737 two cytotoxic cyclic depsipeptides from a Palmyra collection of *Lyngbya bouillonii*.
 738 *Chembiochem* 11, 1458-1466.
 739 Uzair, B., Tabassum, S., Rasheed, M., and Rehman, S.F. (2012). Exploring marine
 740 cyanobacteria for lead compounds of pharmaceutical importance. *ScientificWorldJournal*
 741 2012, 179782.
 742 Vazquez-Baeza, Y., Pirrung, M., Gonzalez, A., and Knight, R. (2013). EMPeror: a tool for
 743 visualizing high-throughput microbial community data. *Gigascience* 2, 16.
 744 Villa, F.A., and Gerwick, L. (2010). Marine natural product drug discovery: Leads for treatment
 745 of inflammation, cancer, infections, and neurological disorders. *Immunopharmacol*
 746 *Immunotoxicol* 32, 228-237.
 747 Walsh, C.T. (2015). A chemocentric view of the natural product inventory. *Nat Chem Biol* 11,
 748 620-624.
 749 Wan, F., and Erickson, K.L. (2001). Georgamide, a new cyclic depsipeptide with an alkynoic
 750 acid residue from an Australian cyanobacterium. *J Nat Prod* 64, 143-146.
 751 Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D.,
 752 Watrous, J., Kapon, C.A., Luzzatto-Knaan, T., et al. (2016). Sharing and community

curation of mass spectrometry data with Global Natural Products Social Molecular
Networking. *Nat Biotech* 34, 828-837.

Watrous, J., Roach, P., Alexandrov, T., Heath, B.S., Yang, J.Y., Kersten, R.D., van der Voort,
M., Pogliano, K., Gross, H., Raaijmakers, J.M., et al. (2012). Mass spectral molecular
networking of living microbial colonies. *Proc Natl Acad Sci U S A* 109, E1743-1752.

Whitton, B.A., and Potts, M. (2012). Introduction to the Cyanobacteria. In *Ecology of
cyanobacteria II: their diversity in space and time* (B.A. Whitton, ed.). (Springer: London,
UK). pp. 1-13.

Williamson, R.T., Boulanger, A., Vulpanovici, A., Roberts, M.A., and Gerwick, W.H. (2002).
Structure and absolute stereochemistry of phormidolide, a new toxic metabolite from the
marine cyanobacterium *Phormidium* sp. *J Org Chem* 67, 7927-7936.

Wishart, D.S., Jewison, T., Guo, A.C., Wilson, M., Knox, C., Liu, Y.F., Djombou, Y., Mandal,
R., Aziat, F., Dong, E., et al. (2013). HMDB 3.0-The Human Metabolome Database in 2013.
Nucleic Acids Res 41, D801-D807.

Worley, B., and Powers, R. (2013). Multivariate Analysis in Metabolomics. *Curr Metabolomics* 1,
92-107.

Yang, J.Y., Sanchez, L.M., Rath, C.M., Liu, X., Boudreau, P.D., Bruns, N., Glukhov, E., Wodtke,
A., de Felicio, R., Fenner, A., et al. (2013). Molecular networking as a dereplication
strategy. *J Nat Prod* 76, 1686-1699.

Figures

Figure 1. MS/MS features as generated in GNPS are shown for marine cyanobacteria and algae (green), marine and terrestrial actinobacteria (blue), lichens (red), freshwater cyanobacteria (yellow) and corals (pink). (A) PCoA plots of 300 samples randomly selected from each dataset display the distance between samples based on molecular features using Bray-Curtis dissimilarity matrix. (B) Marine cyanobacteria and algae and freshwater cyanobacteria. Laboratory cultures of *Phormidium* 1646, *Lyngbya* 1933 are highlighted in red circles. Each sphere represents the full sample metabolome (C) Venn diagram display of overlapping MS/MS features. Percentage of overlapping features with the marine cyanobacteria and algae dataset are given in parenthesis in their respective colors (Figure 1-figure supplement 4).

Supplementary file 1: Samples list and metadata – Sample list and metadata are freely available upon registration at <http://qiita.microbio.me> Study ID:10125. Also refers to figures 2,3,4 and 5.

Figure 1-figure supplement 1: PCoA Plots- Please extract Emperor file from the link and open in Google Chrome – Also refers to figure 2.

<ftp://massive.ucsd.edu/MSV000078568/other/>

S4_6way PCoA_emperor-bc.n.txt

S4_5way PCoA_emperor-bc.300.n.txt

S4_2way_PCoA_emperor-cyano-bc.txt

Figure 2. MS/MS feature diversity between and within sampling sites. PCoA plots of crude extract molecular features using Bray-Curtis, dissimilarity matrix. Each point represents a single sample and points are colored by metadata. (A) PCoA plot shows the distance between samples based on field identified classification. (B) PCoAs color coded by geographical origin shows the distance of all locations together (B: upper panel) and the individual PCoA plots of the four collection sites with most samples (B: lower panel). Laboratory cultures of *Phormidium* 1646, *Lyngbya* 1933 and *Lyngbya* 1963 are highlighted in orange circles.

Figure 2- figure supplement 1: PCA plot of crude extracts of cyanobacteria and algae collections. PCA was performed using Bruker Profile Analysis software version 2.1. Samples are displayed by collection location as described in the legend. NAC-Curacao, PAB-Panama Bocas del Toro, PAC-Panama Coiba, PAG-Panama Gulf of Chiriqui, PAP-Panama Portobelo, PAL-Palmyra Atoll, PNG-Papua New Guinea.

804

805 **Figure 3.** Chemodiversity and richness of molecular features based on MS/MS data. (A) Rarefaction curve of
806 cyanobacteria collection library showing the chemical richness of crude extracts vs. the fraction library. (B)
807 Abundance of molecular families: Bars represent the number of features and spectra comprising each molecular
808 family clustered by GNPS molecular networking. (C) Bar chart depicting both the total number of extracts from a
809 given location and the percent contribution of unique features to the entire dataset. (D) Pie chart representing the
810 percentage of unique molecular features attributed to origin of the sample. HI-Hawaii, NAC-Curacao, PAB-Panama
811 Bocas del Toro, PAC-Panama Coiba, PAG-Panama Gulf of Chiriqui, PAP-Panama Portobelo, PAL-Palmyra Atoll,
812 PNG-Papua New Guinea.

813 **Figure 3- figure supplement 1:** Molecular network of marine cyanobacterial natural products with annotated
814 molecular families. (A) Molecular networking generated using MSV000078568/MSV000078892 datasets by
815 <http://gnps.ucsd.edu> and visualized in Cytoscape. A-malyngamides [M+H]⁺, B-barbamides [M+H]⁺, C-curacins
816 [M+H]⁺, D-hectochlorins [M+H]⁺, E-viequeamides [M+Na]⁺, F-carmabins [M+H]⁺, G-palmyramides [M+Na]⁺, H-
817 palmyramides [M+H]⁺, I-dolastatins [M+2H]⁺, J- cyanolides [M+H]⁺, K-aprattoxins [M+H]⁺, L-hoiamides [M+H]⁺, M-
818 hoiamides [M+Na]⁺, N-majusculamides [M+Na]⁺, O-hectochlorins [M+Na]⁺, P-palmyrolides [M+H]⁺. Spectral data
819 dereplicated by fragmentation patterns present in spectral libraries available on GNPS. Nodes are color-coded by
820 the geographical origin of the sample and labeled with parent mass (*m/z*). Edge thickness represents the cosine
821 similarity score. Insets on right hand side depict the **(B)** barbamide and **(C)** cyanolide molecular families. Shown for
822 each are the standard library spectrum (upper panel) and the matching sample spectrum from within the same
823 node (lower panel).

824 **Figure 3-figure supplement 2:** Dereplication of the apratoxin molecular family. Identification of apratoxin B based
825 on library standards enabled dereplication of 5 known aprattoxins (A, D, F, G, and H). Additional apratoxin
826 derivatives were characterized based on mass differences of functional groups.

827 **Figure 3-figure supplement 3.** Dereplication and spatial distribution for new derivatives of the barbamide
828 molecular family. Based on barbamide **(1)** dereplication, peaks assigned to fragments enable the characterization
829 of N-demethylbarbamide **(2)** and a substitution of the Phe and Leu groups as a new barbamide derivative **(3)**.

830 **Supplementary file 2 - Supplemental table 1: GNPS molecular networking MS/MS based identification of**
831 **molecules and molecular families – Refers to Figure 3 and 4.**

832

833 **Figure 4.** Spatial maps showing the geographic distribution of selected cyanobacterial natural products. (A) Features
834 were binned by their distribution patterns across the 8 main geographical locations, each bin represented by a node
835 with edges linking bins differing by one location. Number of features in each bin is presented according to size and
836 color scale from white (0) to red (10,000) as indicated by the scale bar on the top. Spatial patterns are represented for

selected natural products within these bins. Inserts in each map display zoomed-in sections of Panama, Curaçao and Papua New Guinea. Each sample is designated to a specific coordinate based on GPS coordinates of the collection site (multiple samples are represented by spots placed around the collection site). See Supplemental data 8. (B) Spatial maps display chemogeographical distribution and abundance of barbamide (**1**) and two barbamide analogs (**2**, **3**). Relative abundance is presented by Jet color scale from low (blue) to high (red). HI-Hawaii, NAC-Curaçao, PAB-Panama Bocas del Toro, PAC-Panama Coiba, PAG-Panama Gulf of Chiriqui, PAP-Panama Portobelo, PAL-Palmyra Atoll, PNG-Papua New Guinea.

Figure 4- figure supplement 1: Natural products global mapping. (Please open the link listed below in Google Chrome, wait till the data is loaded and visualized, then click at the Mapping submenu, and change the scale to Logarithmic, and Color map to Jet. For flipping through maps corresponding to different molecules, please click at the name of the molecule shown above the colorbar and select another molecule either with a click or with up-down arrow keys. In case you experience any problems, please contact theodore.alexandrov@embl.de.) <http://ili-toolbox.github.io/?cyano/bg.png;cyano/intensities.csv>

Figure 5. The molecular family and structure elucidation of a novel natural product yuvalamide A (**4**), isolated from a Panama-Portobelo (PAP) cyanobacterial collection. (A) Molecular families and the spatial distribution of yuvalamide A $[M+H]^+$ and $[M+Na]^+$ ions are highlighted within yellow circles. MS/MS spectra display the linear structure for the non-ribosomal peptide (NRP) fragments Gly-Ile-Hiv-Val. (B) for the $[M+H]^+$ and the (C) $[M+Na]^+$ ions. (D) The full elucidated structure of yuvalamide A (**4**) as confirmed by NMR analysis (Supplemental data 9).

Supplementary file 3. 2D NMR spectroscopic data for amino acids residues of yuvalamide A. ^{13}C and 1H chemical shifts were determined by HSQC and HMBC spectra. Hiv = 2-hydroxyisovaleric acid (600 MHz for 1H , 125 for ^{13}C), the solvent (DMSO- d_6) and the temperature (298K).

Figure 5-figure supplement 1: Structure identification of yuvalamide A. (A) 1H NMR spectrum of yuvalamide A with COSY and key HMBC correlations (600 Hz, DMSO- d_6). (B) HMBC correlation of Dhoya-H3 - Gly-CO. (C) HMBC correlation of Dhoya-H6 - C7 and H6-C8.

Figure 5-figure supplement 2. MS/MS fragmentation of yuvalamide A. The theoretical masses of predicted MS/MS fragments of yuvalamide A are shown in the top panel. The observed masses of these MS/MS fragments are labelled in the acquired spectrum.

Figure 5-figure supplement 3: Dereplication of yuvalamide molecular family. Additional putative yuvalamide analogs observed in the molecular family. One analog with an $[M+H]^+$ m/z 522.3156 (14 Da smaller than compound **4**) displayed a similar, but not identical fragmentation pattern to yuvalamide A (Figure 5-Figure supplement 3). The linear order we propose is Gly-[C10H17NO3]-Val-Dhoya, as the -

14 Da change is located in the Ile - Hiv residue region when compared to yuvalamide A. This is likely due to the promiscuous behavior of A domains in the NRPS biosynthetic assembly. Such promiscuous behavior where one amino acid is substituted by another, is quite common in non-ribosomal peptides, and results from similar amino acids being recognized by the cognate adenylation domain (Crawford, et al., 2011). Another yuvalamide analog with a 572.3303 m/z (Figure 5-Figure supplement 3) that clusters with the sodiated yuvalamide molecular family possess a similar fragmentation pattern to the yuvalamide A but with an increase in 14 Da. This shift is likely due to presence of additional methyl group and the fragmentation suggests this methyl is near the Hiv residue.

Figure 5-figure supplement 3: Dereplication of yuvalamide molecular family. Molecular families of [M+H]⁺ and [M+Na]⁺ ions highlighting yuvalamide A (4) (Gly-Ile-Hiv-Val-Dhoya) in yellow and the putative yuvalamide B (5) characterized as Dhoea analogue (Gly-Ile-Hiv-Val-Dhoea) in green circles. Two additional putative analogs characterized 522.3 m/z [M+H]⁺ (Gly-C10H17NO3-Val-Dhoya) and 572.3 m/z [M+Na]⁺ (Gly-Ile-MeHiv-Val-Dhoya).

A 3D scatter plot showing the distribution of data points across three principal components: PC1 (9.19%), PC2 (7.15%), and PC3 (5.74%). The plot displays several distinct clusters of points, each represented by a different color (red, blue, green, yellow, magenta, and cyan). The clusters are well-separated, indicating significant differences between the groups along the principal components.

actinobacteria (2.2%)

lichen (1.02%)

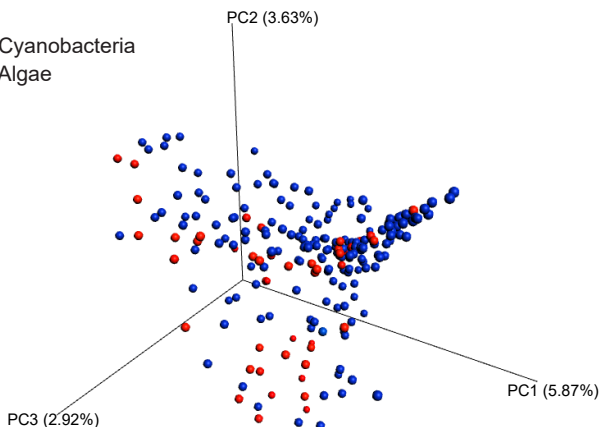
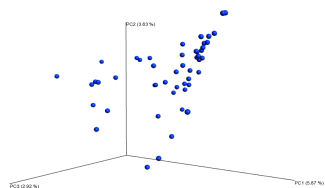
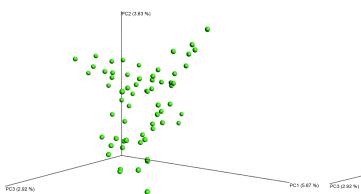
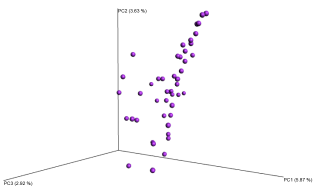
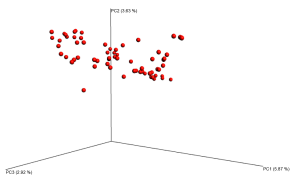
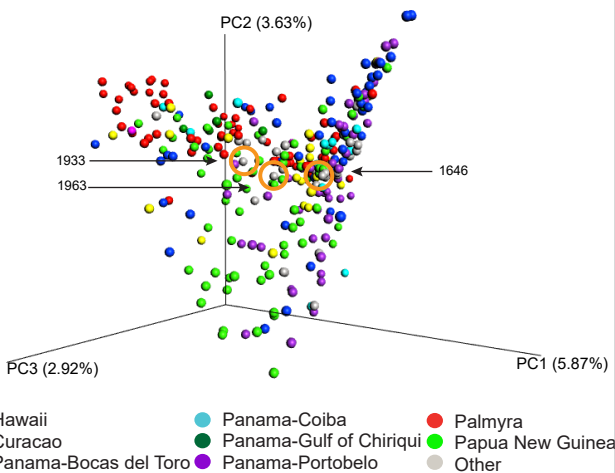
freshwater cyanobacteria (8.4%)

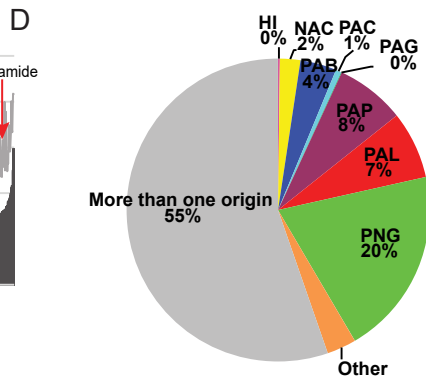
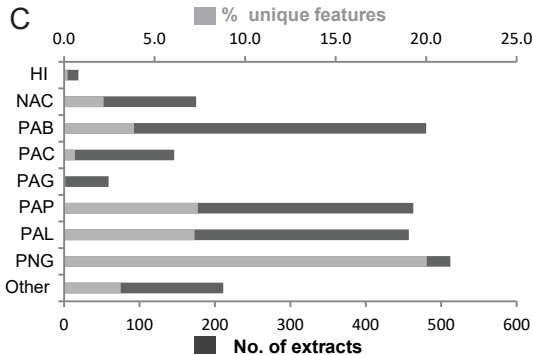
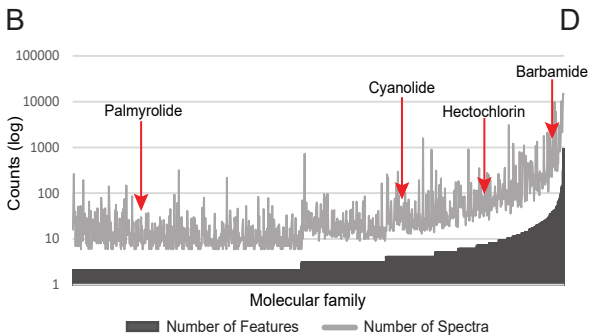
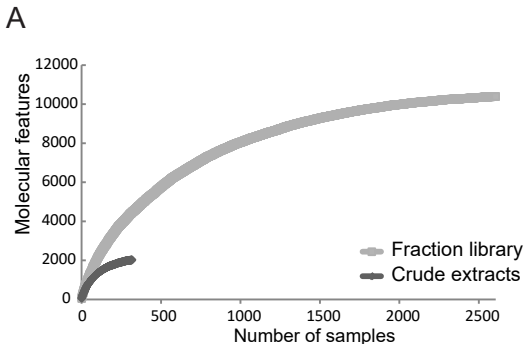
corals (3.7%)

marine cyanobacteria/algae

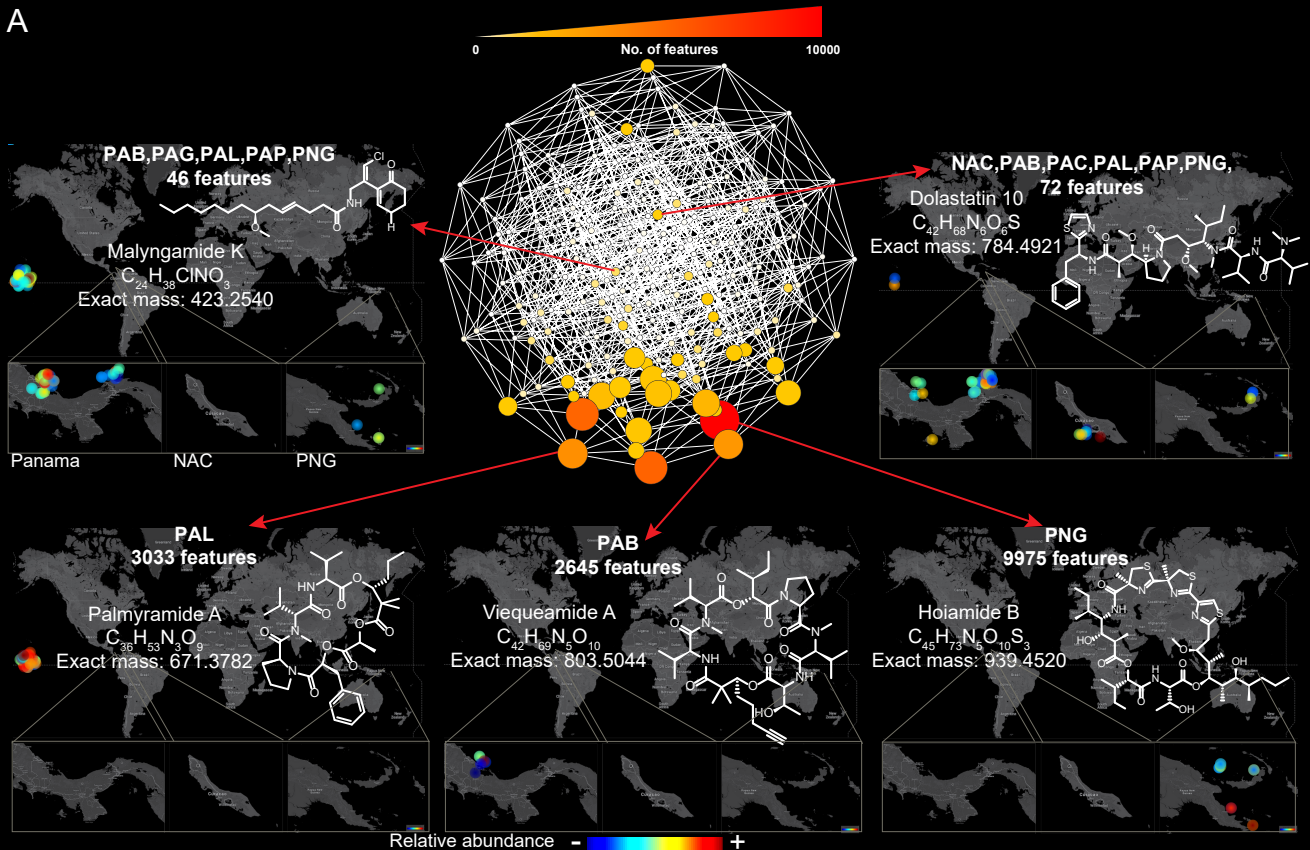
A

● Cyanobacteria
● Algae

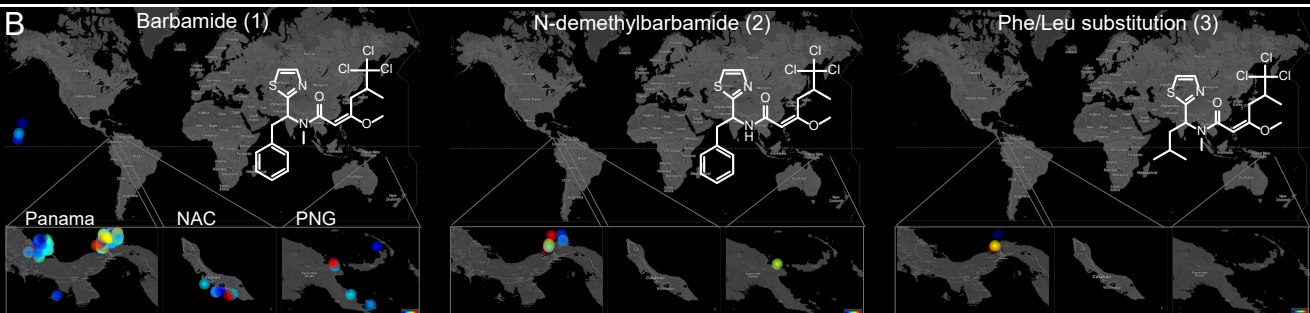
**B**

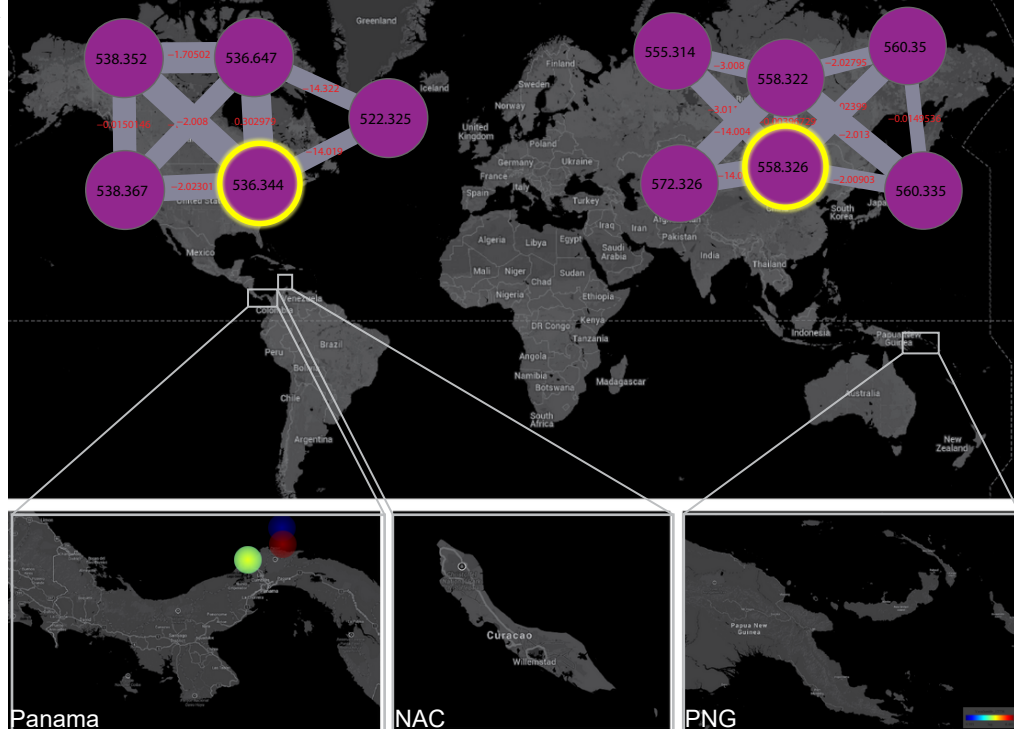


A

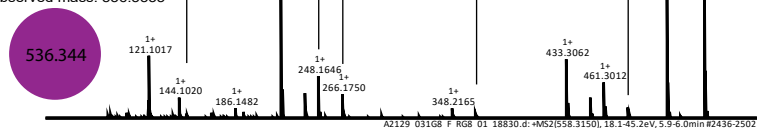


B

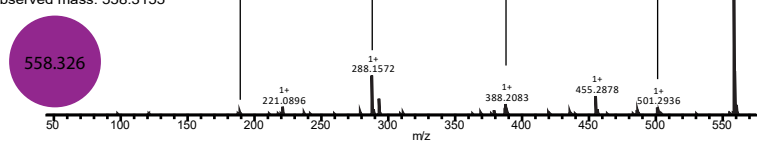
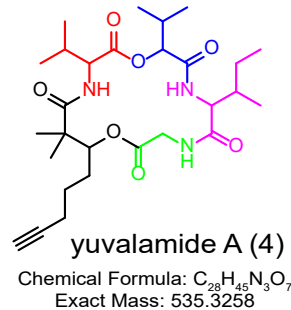


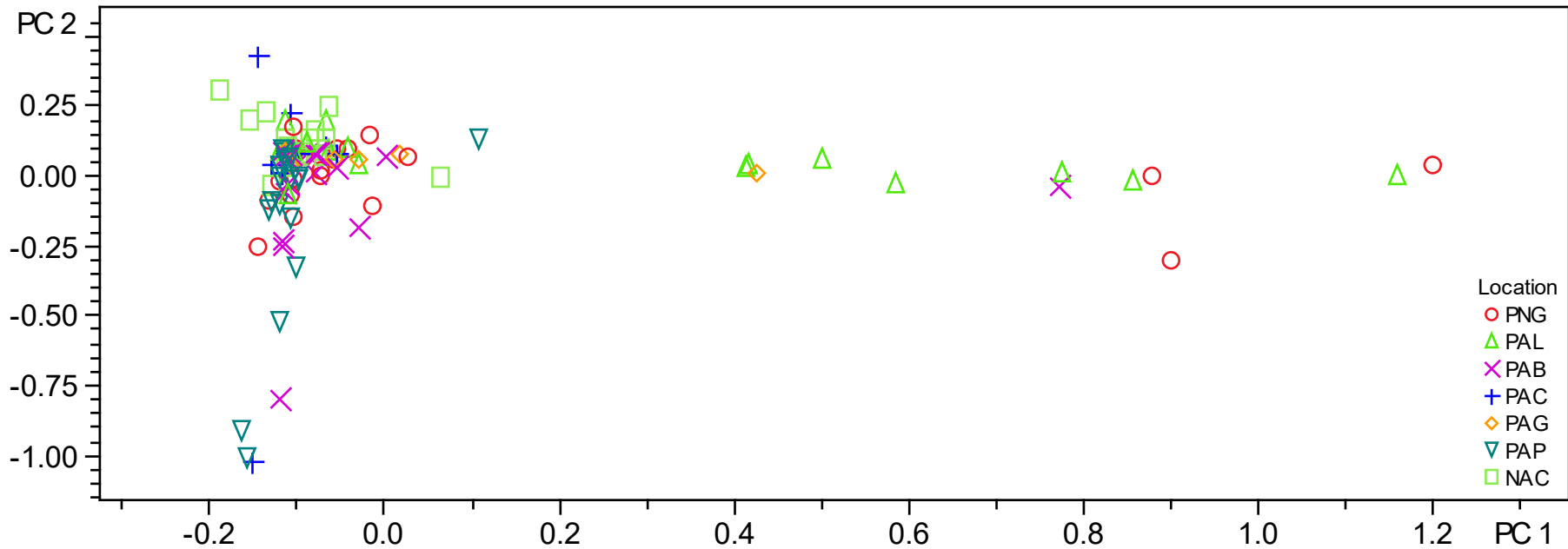
A**B****[M+H]⁺**

Theoretical Mass: 536.3330
Observed mass: 536.3335

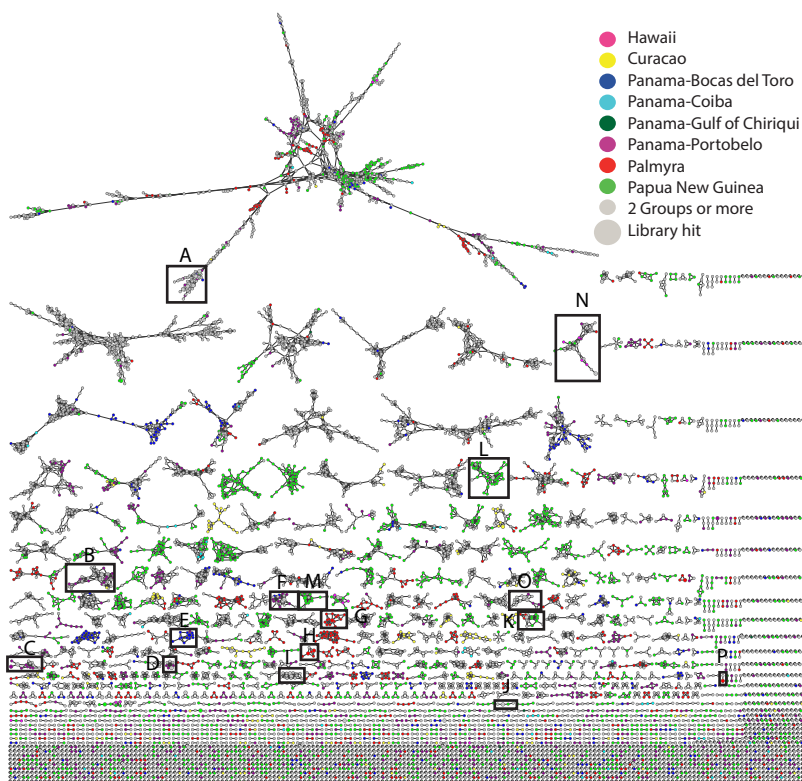
**C****[M+Na]⁺**

Theoretical Mass: 558.3150
Observed mass: 558.3155

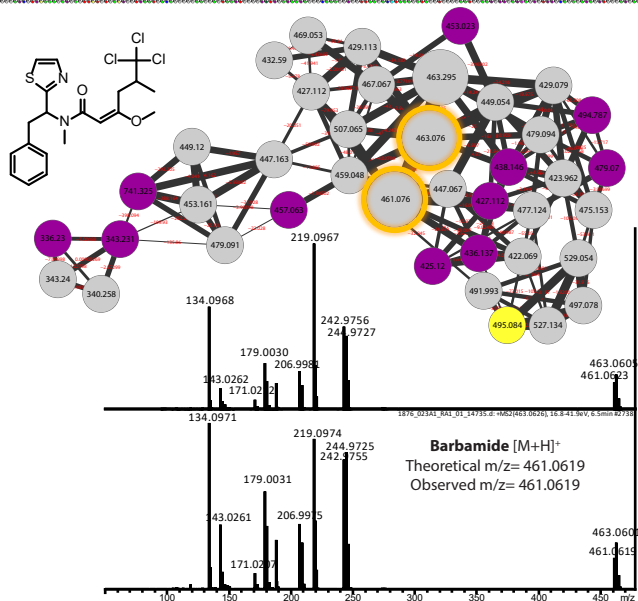
**D**



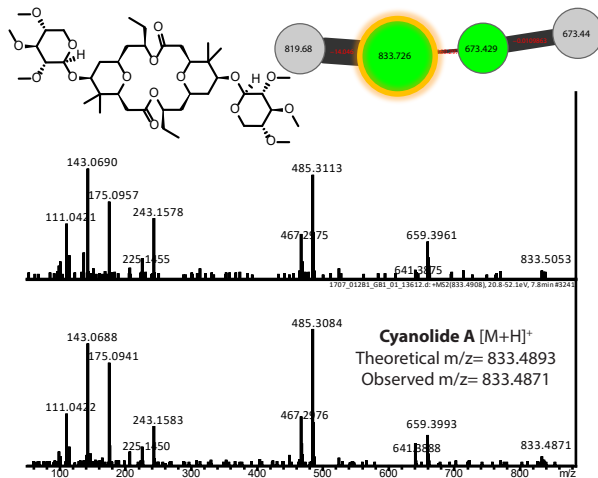
A



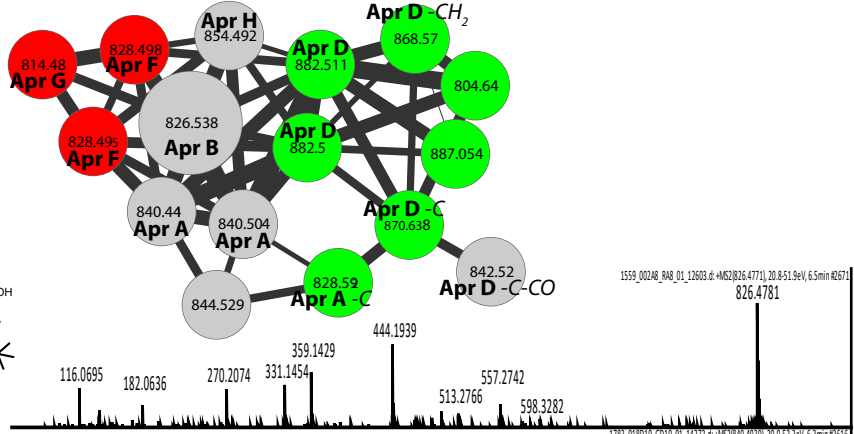
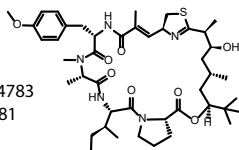
B



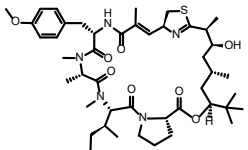
C



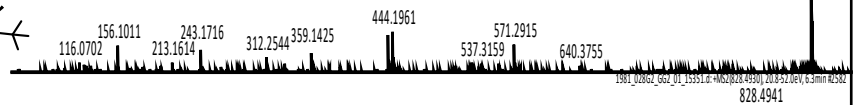
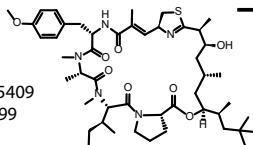
Apratoxin B [M+H]
Theoretical m/z = 826.4783
Observed m/z = 826.4781



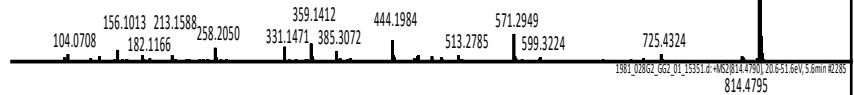
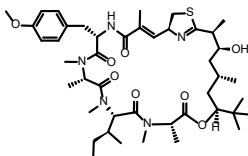
Apratoxin A [M+H]
Theoretical m/z = 840.4940
Observed m/z = 840.4936



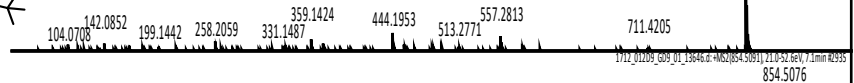
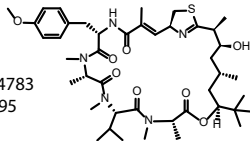
Apratoxin D [M+H]
Theoretical m/z = 882.5409
Observed m/z = 882.5399



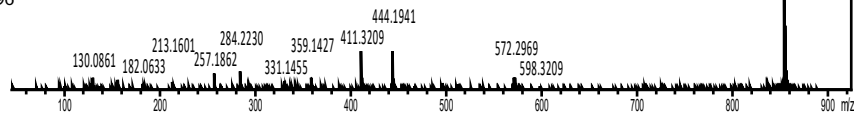
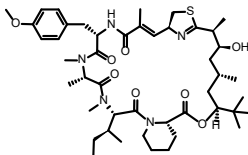
Apratoxin F [M+H]
Theoretical m/z = 828.4940
Observed m/z = 828.4941

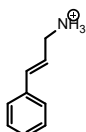
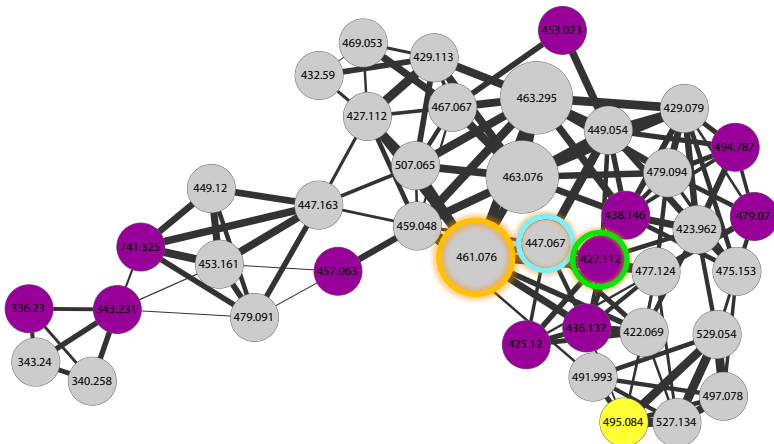


Apratoxin G [M+H]
Theoretical m/z = 814.4783
Observed m/z = 814.4795

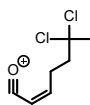


Apratoxin H [M+H]
Theoretical m/z = 854.5096
Observed m/z = 854.5076

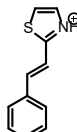




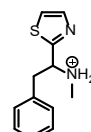
Exact Mass: 134.0964



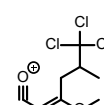
Exact Mass: 179.0025



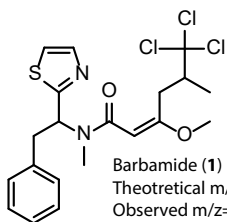
Exact Mass: 188.0528



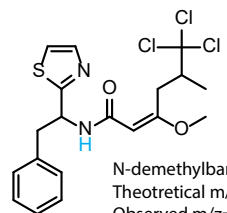
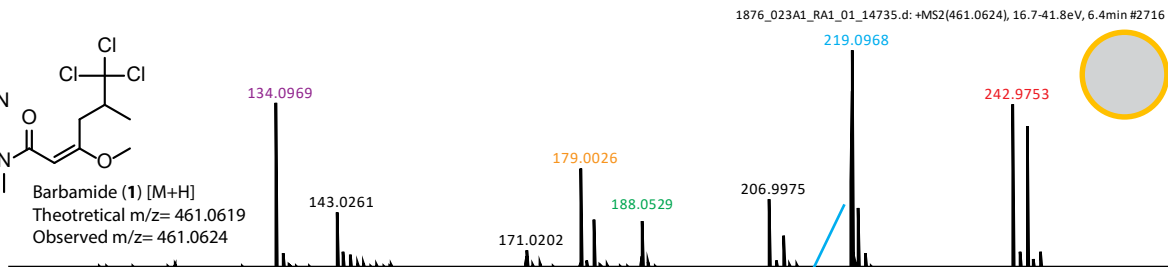
Exact Mass: 219.0950



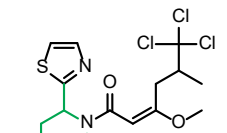
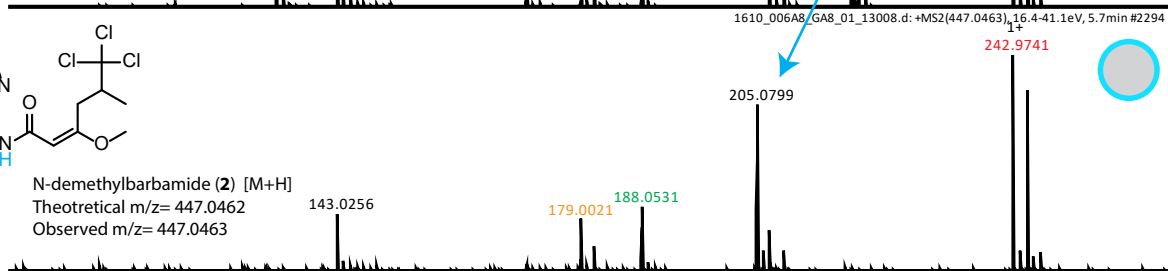
Exact Mass: 242.9741



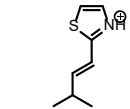
Barbamide (1) [M+H]
Theoretical m/z= 461.0619
Observed m/z= 461.0624



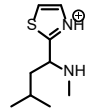
N-demethylbarbamide (2) [M+H]
Theoretical m/z= 447.0462
Observed m/z= 447.0463



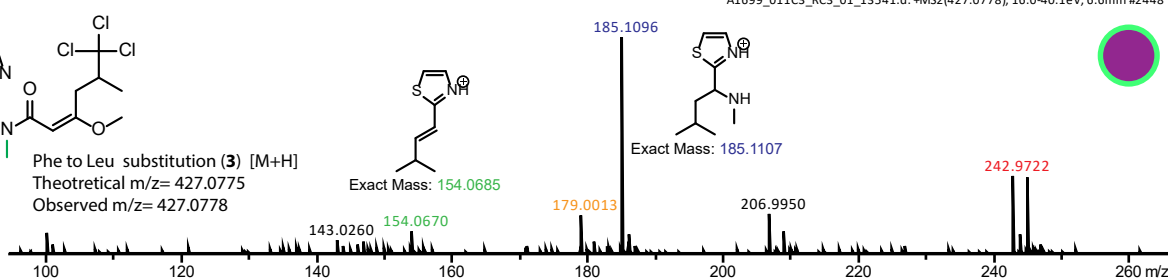
Phe to Leu substitution (3) [M+H]
Theoretical m/z= 427.0775
Observed m/z= 427.0778



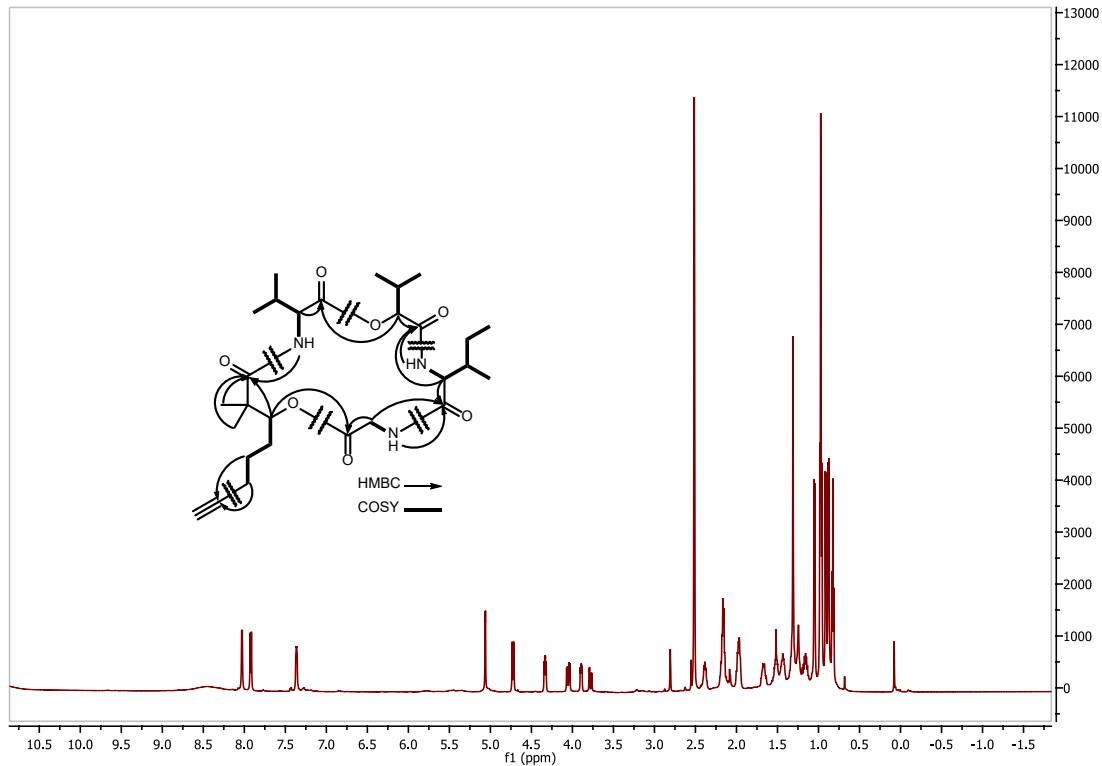
Exact Mass: 154.0685



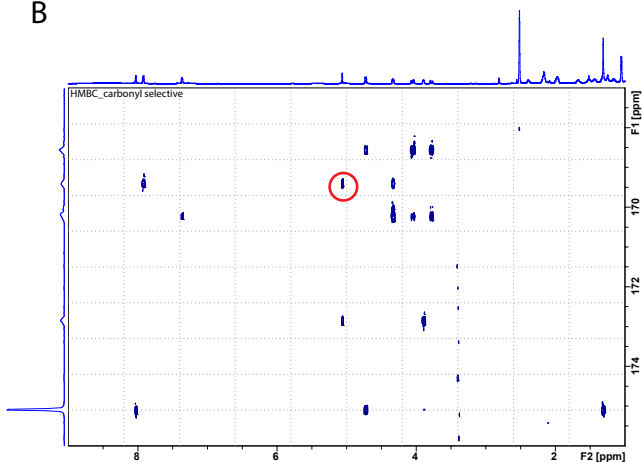
Exact Mass: 185.1107



A



B



C

