Illuminating the dark metabolome of *Pseudo-nitzschia*-microbiome associations

Irina Koester[1*], Zachary A. Quinlan[1], Louis-Félix Nothias[2,3], Margot E. White[1], Ariel Rabines[1,4], Daniel Petras[1,2,3], John K. Brunson[1,4], Kai Dührkop[5], Marcus Ludwig[5], Sebastian Böcker[5], Farooq Azam[1], Andrew E. Allen[1, 4], Pieter C. Dorrestein[2,3] and Lihini I. Aluwihare[1]

[1]Scripps Institution of Oceanography, University of California San Diego, La Jolla, USA
[2]Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, USA
[3]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, USA
[4]Microbial and Environmental Genomics Group, J. Craig Venter Institute, La Jolla, CA 92037, USA
[5]Chair for Bioinformatics, Friedrich Schiller University, Jena, Germany

*Corresponding author:
Irina Koester
University of California San Diego
Scripps Institution of Oceanography
8622 Kennel Way, La Jolla, CA, USA
ikoester@ucsd.edu

Running title: Illuminating Pseudo-nitzschia-culture metabolomes

Originality-Significance Statement

The exchange of organic molecules between microorganisms in the surface ocean is imperative for facilitating bloom maintenance, grazing deterrence, microbiome specificity, and in the case of harmful algae, toxin production. Previous work has linked a small number of specific compounds to cell-cell communication or nutrient exchange, but there are thousands of unidentified primary and secondary metabolites produced by an algal host and its associated microbiome. In this study, we applied state-of-the-art cheminformatic tools to mass spectrometry data in order to assign structural information to 56% of the chemical features produced by five different species of *Pseudo-nitzschia*. Some *Pseudo-nitzschia* are notorious for producing the neurotoxin domoic acid, but toxin production was not the only driver of the observed differences. The most prominent finding was the central role of organic nitrogen compounds in distinguishing the chemical milieu of each species. Furthermore, the microbiomes were species-specific and our analyses grouped specific bacteria and metabolites by *Pseudo-nitzschia* species indicating that our analysis has the potential to uncover novel metabolites that are involved in algal-bacterial interactions. Our study represents the most comprehensive characterization of the metabolite profile of a marine algae and demonstrates that this approach can differentiate the metabolomes of closely-related species (same genus). Ultimately, these results have the potential to transform our understanding of niche differentiation between these *Pseudo-nitzschia* species.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1462-2920.16242

Summary

The exchange of metabolites mediates algal and bacterial interactions that maintain ecosystem function. Yet, while 1000s of metabolites are produced, only a few molecules have been identified in these associations. Using the ubiquitous microalgae *Pseudo-nitzschia* sp., as a model, we employed an untargeted metabolomics strategy to assign structural characteristics to the metabolites that distinguished specific diatom-microbiome associations. We cultured five species of *Pseudo-nitzschia,* including two species that produced the toxin domoic acid, and examined their microbiomes and metabolomes. A total of 4826 molecular features were detected by tandem mass spectrometry. Only 229 of these could be annotated using available mass spectral libraries, but by applying new *in-silico* annotation tools, characterization was expanded to 2710 features. The metabolomes of the *Pseudo-nitzschia*-microbiome associations were distinct and distinguished by structurally diverse nitrogen compounds, ranging from simple amines and amides to cyclic compounds such as imidazoles, pyrrolidines, and lactams. By illuminating the dark metabolomes, this study expands our capacity to discover new chemical targets that facilitate microbial partnerships and uncovers the chemical diversity that underpins algae-bacteria interactions.

1. Introduction

Numerous organic molecules have been shown to facilitate ecologically important partnerships, including signaling molecules used for cell-cell communication and the exchange of organic nutrients as sources of energy, carbon and essential metabolites (Armbrust *et al.*, 2004; Amin *et al.*, 2012, 2015). Diatoms have received particular attention in this context as several studies have identified important microbial partners and some of the metabolites that maintain these

partnerships (Amin *et al.*, 2015; Shibl *et al.*, 2020). However, microalgae cells, their phycosphere (the surrounding exudate layer) and the external metabolite pool (dissolved metabolites in seawater) consist of thousands of diverse compounds, many of which are inaccessible to current analytical approaches. This has hampered our ability to assign structural characteristics to this "chemical dark matter" (da Silva *et al.*, 2015), keeping their ecological role invisible. The goal of this study was to design an approach to describe the chemical diversity of complex metabolite pools more comprehensively than is afforded by common analytical approaches in an effort to catalyze the discovery of new chemical tokens that mediate algae-bacteria interactions.

This work was done with the ecologically significant diatom genus *Pseudo-nitzschia*; some members of this genus produce the potent neurotoxin domoic acid (DA), which bioaccumulates in the aquatic food web endangering the health of apex marine predators and humans (Lelong *et al.*, 2012; Trainer *et al.*, 2012). Recent studies show that interactions between *Pseudo-nitzschia* and its associated microbiome, which varies by *Pseudo-nitzschia* species, can promote *Pseudo-nitzschia* growth and DA production (Bates *et al.*, 1995; Guannel *et al.*, 2011; Sison-Mangus *et al.*, 2014; Amin *et al.*, 2015). Amin *et al.* (2015) showed that these microbial interactions are highly complex and can be very exclusive. In particular, marine bacteria of the genus *Sulfitobacter* (strain SA11) were shown to enhance the growth of one *Pseudo-nitzschia* strain, while having no impact on other strains. This mutualistic partnership appeared to consist of the exchange of nutrients, such as ammonia and organosulfur compounds, and cell-cell communication via the signaling molecules tryptophan and indole-3-acetic acid.

In general, metabolomes are dictated by genetics and environmental growth conditions. Metabolites found in the dissolved phase could derive from excess production by the algae, cell lysis or are intended for intra- and interspecies interactions. On the other hand, distinct microbiomes may produce different metabolites themselves or they influence the physiology and therefore, the intra- and exometabolites produced by algae. The bacteria might also act as "gate

keepers" for the release of metabolites into the surrounding media and as "transformers" of metabolites. Since previous studies have suggested that metabolites secreted by *Pseudo-nitzschia* regulate relationships between its species-specific bacterial assemblages (Sison-Mangus *et al.*, 2014), we hypothesized that metabolomes of individual *Pseudo-nitzschia* cultures must be species-specific as well. Five different species of *Pseudo-nitzschia* were cultured and their associated microbiomes were identified using 16S ribosomal RNA (rRNA) sequencing. We postulated that both dissolved and intracellular metabolites could be relevant for maintaining and differentiating microbial interactions and partners, which led to the analysis of the dissolved metabolome and whole metabolome (intracellular plus dissolved).

Our approach employed an untargeted metabolomics strategy, coupling liquid chromatography (LC) with high resolution tandem mass spectrometry (MS/MS), to characterize the chemodiversity within *Pseudo-nitzschia* cultures. In order to compensate for the limited applicability of mass spectral libraries to the marine environment, we advanced the use of a number of recently developed *in-silico* tools to better characterize unknown molecules. This approach enabled the most comprehensive cataloging to date of the broad chemical classes present in a laboratory algal culture and in doing so, illuminated the chemical characteristics of metabolites that differentiate metabolomes of *Pseudo-nitzschia* cultures. Our findings provide a foundation for a more complete understanding of microbe-microbe interactions, including insights into how these interactions may contribute to maintaining and sustaining harmful algal blooms.

2. Experimental Procedures

Detailed methods are described in the Supplementary Information; workflows demonstrating the experimental setup and computational tools used to analyze and annotate MS/MS data are visualized in Fig. 1.

2.1 Experimental setup and data acquisition

Five unialgal *Pseudo-nitzschia* cultures isolated from Californian coastal waters were grown in duplicate in natural seawater media until stationary phase (Fig. S1). *Pseudo-nitzschia* cells and attached and free-living bacteria were also enumerated during growth by epifluorescence microscopy following 4′,6-diamidino-2-phenylindole (DAPI) staining (Fig. S2). To determine the active microbial community, we extracted RNA from cell pellets (centrifuged 90 mL of culture samples) and sequenced the v4-v5 region of 16S rRNA gene. To determine how much carbon was accumulating as dissolved metabolites in each culture, dissolved organic carbon (DOC) concentration was determined for the natural seawater growth media blank and each culture at the time of harvesting. *Pseudo-nitzschia*-microbiome associated metabolomes were examined in both dissolved and whole metabolome samples: dissolved metabolites were isolated in duplicate 50 mL culture samples after filtration through GF/F filters, and whole metabolites were isolated in duplicates of 50 mL unfiltered culture samples that were ultrasonicated for 30 minutes. Duplicates of media blanks for each treatment, dissolved and whole metabolome, were taken as well. Samples were acidified (pH2, hydrochloric acid), solid phase extracted (Bond Elut PPL resin) and eluted in methanol (Dittmar *et al.*, 2008; Petras *et al.*, 2017). The extracts were dried down, re-dissolved in 100 µL methanol with 1% formic acid of which 10 µL were subjected to ultra-high performance liquid chromatography (UPLC; C18 core-shell column) electrospray ionization (ESI+) mass spectrometry (MS). A Q-Exactive orbitrap mass spectrometer (Thermo Fisher

Scientific) was used in data dependent acquisition MS/MS mode (range 150-1500 mass-to-charge ratio (m/z)) (Petras *et al.*, 2017).

2.2 Analyses and annotation of MS/MS data

We use “feature” to refer to an ion signal (identified by its m/z) detected from a molecule eluting off the LC column at a specific time (retention time), for which an MS/MS spectrum is also available. The software MZmine 2 (Pluskal *et al.*, 2010) was used for feature extraction and alignment across all samples. Alignment was based on unique features in multiple samples being linked to a specific LC retention time and m/z value. In total 9904 features were detected and assigned an identification number. Feature intensities in each sample were calculated based on peak areas of extracted-ion chromatograms (XICs). Procedure and media blanks of this study were used to identify and remove “background features” from the data, resulting in 4826 features. These features are considered as newly produced in the cultures and were used for subsequent statistical analysis. Note that this blank removal procedure would remove features that were not produced in the culture but present in the seawater that was used to prepare the media.

To build subnetworks of structurally related features, we used feature-based molecular networking via GNPS (Wang *et al.*, 2016; Nothias *et al.*, 2020), which calculates cosine similarity scores between all MS/MS spectra of features in a dataset visualized using Cytoscape software (Shannon *et al.*, 2003). GNPS also searches public and commercial standard libraries for spectral matches - referred to as either library IDs or analogs depending on the specified parameters. While library IDs are exact matches, analogs are designated as partial matches to a library spectrum, which can successfully characterize features on the chemical class level. ClassyFire was used to assign chemical classifications to all structures annotated by spectral matches (Djoumbou Feunang *et al.*, 2016) (Fig. S3 A-C).

To assign molecular formulas (MFs) to each feature we used the SIRIUS workflow, which works independently from the spectral matches. This workflow uses MS/MS spectra and MS1 (parent ion) isotope patterns (Böcker *et al.*, 2009; Dührkop *et al.*, 2019). Subsequently, ZODIAC improved the correct MF annotation by reranking possible MF based on the MFs determined for our entire dataset (Ludwig *et al.*, 2020). Based on the reranked ZODIAC MFs and their associated MS/MS fragmentation trees, CSI:FingerID searches molecular structure databases to predict the structure of compounds (Dührkop *et al.*, 2015). Finally, the recently developed machine learning-based tool CANOPUS was used to predict the probability (0-1) that each feature (known or unknown) belonged to one or more of the 2497 CANOPUS compound classes (Fig. S3D) (Dührkop *et al.*, 2020). While these *in-silico* annotation tools created outputs for 91% of all features in this dataset, we only considered MFs and compound class predictions with a ZODIAC score of >0.98, which resulted in outputs of chemical information for 2710 of the 4826 features (56%). Furthermore, rare compound classes with consistently low CANOPUS probabilities (< 0.5) in the dataset were discarded, leaving 532 compound classes for further analysis.

2.3 Deposition of code and data

Statistical analysis of metabolomics and 16s rRNA data was performed using an in-house developed R script, which is available publicly and deposited here: https://github.com/Zquinlan/pseudonitzschia_Koester. The mass spectrometry data are deposited on the Mass Spectrometry Interactive Virtual Environment (MassIVE) repository (massive.ucsd.edu) with accession number MSV000081731 and are publicly available. The GNPS job is public and can be found on the GNPS site at the following link: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=96722ff3fec843969ca8b673b5e7fdaf and for analog search

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=ee40a3b66854448ca2b7f117f6b1c22a. The 16S rRNA data is deposited at the National Center for Biotechnology Information (NCBI) with accession number PRJNA776274.

3. Results

3.1. The distinctive microbiomes and metabolomes of *Pseudo-nitzschia* cultures

Generally, *Pseudo-nitzschia* are classified into two groups based on their valve width or transapical axis: delicatissima-size class (< 0.3 µm) and the seriata-size (> 0.3 µm) (Tomas and Hasle, 1997). Here, the cultures of the delicatissima-size class - *P. hasleana, P. galaxiae* and *P. delicatissima* - grew to higher densities than the seriata-size class species - *P. multiseries* and *P. subpacifica* (Fig S1 and S2A; classification according to Lelong *et al.* (2012)). The cultures of the delicatissima-size class also had higher bacterial counts (attached and free-living) at the time of harvesting (Fig. S2C and S2E). The total dissolved metabolite concentrations (measured as DOC, Fig. S2F) were only elevated in the samples of *P. hasleana, P. galaxiae*, which is also reflected by the TIC (total ion chromatogram = sum of all XICs in a sample) of the LC-MS/MS analysis of their dissolved metabolomes (Fig. S2H). In all cases, the TIC of the whole metabolome samples is higher than the TIC of dissolved metabolomes (Fig. S2G). This is consistent with the whole metabolome including cellular material and dissolved compounds. Whole metabolome TICs are highest in the densely growing delicatissima-size class cultures.

Both the microbiomes and the metabolomes of each of the five *Pseudo-nitzschia* species are distinct (PERMANOVA analyses, $p < 0.001$), despite their morphological size class classifications

and the fact that some species were isolated from seawater on the same day and at the same location (see Supplementary Information 1.1). The 16S rRNA sequences clearly distinguished the microbiomes of each *Pseudo-nitzschia* culture (Fig. 2A). The microbiomes were species-specific even when including data from an experiment in 2016, where the same cultures of *P. multiseries*, *P. subpacifica* and *P. delicatissima* as well as two additional strains of *P. subpacifica* and *P. delicatissima* were cultured (Fig. S4A; PERMANOVA analysis, $p < 0.001$). The interannual comparison means that we were unable to observe any significant temporal shift in microbiomes and that microbiomes were stable for time periods of up to one year. The metabolomes of the experiment in 2016 were also species-specific (Fig. S4B; PERMANOVA analysis, $p < 0.001$). Further, microbiomes and metabolomes of the faster-growing delicatissima-sized cultures are distinct from the slower-growing seriata-size cultures (PERMANOVA analysis of microbiome, $p < 0.05$; PERMANOVA analysis of metabolome, $p < 0.001$). The difference in the metabolomes is equally prominent within the whole metabolome and the dissolved metabolome samples (Fig. 2B).

Principle Coordinate Analysis (PCoA) of all metabolome samples separated the *Pseudo-nitzschia* species on Axis 1 (54%), whereas the dissolved exudates were differentiated from whole metabolome samples along Axis 2 (21%; Fig. 2B). The distinction between dissolved vs. whole metabolome (PERMANOVA analysis, $p < 0.001$) was analyzed the same way as the differences between the cultures described in the following and the results are presented and discussed in detail in the Supplementary Information (Fig. S5B, Fig. S6, Fig. S7, Fig. S8, Fig. S9D and E and Table S1). At a very simple level, these metabolite data can consider differences in presence and abundance of features. To account for differences in biomass between the cultures and therefore in the relative amount of sample injected into and detected by the mass spectrometer, we relativized the metabolite data to TIC and chlorophyll *a*. Thus, we do not attribute the observed differences along Axis 1 to cell densities and associated increases in the absolute metabolite

concentration but rather to culture-specific metabolomes. The separation between whole and dissolved metabolomes on Axis 2 can be explained by metabolites produced by the diatom-bacteria association with differing intra- and extracellular abundances.

3.2. The microbiome of *Pseudo-nitzschia* cultures

In this experiment, 47 Amplicon sequence variants (ASVs) were identified in the *Pseudo-nitzschia* cultures. A list of all ASVs and their percent abundance in the experiment can be found in Table S2. The table additionally identifies if bacteria observed in this study were found to be associated with *Pseudo-nitzschia* previously in the literature. Out of the 47 ASVs, 40 were previously found in studies describing the *Pseudo-nitzschia* microbiome on the genus or family level (Guannel *et al.*, 2011; Sison-Mangus *et al.*, 2014, 2016; Amin *et al.*, 2015; Needham and Fuhrman, 2016). Based on the mean percent abundance across all culture samples, Gammaproteobacteria was the most common class in the *Pseudo-nitzschia* microbiome of this study (58 ± 21%), within which *Alteromonadaceae* (40 ± 30%) was the most abundant family. The next most abundant class was Alphaproteobacteria (26 ± 17%), where *Rhodobacteraceae* (26 ± 17%) were the dominant family. Bacteroidia (11 ± 13%) was the third most common class and consisted exclusively of *Flavobacteriaceae*. Betaproteobacteria were only found in low abundances in the cultures (4.5 ± 6.2%).

Comparing the differences in microbial communities between *Pseudo-nitzschia* cultures (Fig. 2C), we observed that Gammaproteobacteria dominated the microbiome of *P. multiseries* and *P. subpacifica* (68% and 90% respectively for the mean of duplicates) but made up only 32% - 57% of the community in *P. delicatissima, P. galaxiae and P. hasleana.* The family *Alteromonadaceae* was most abundant in the *P. subpacifica* cultures (85%), and also abundant in *P. galaxiae* (52%)*, P. delicatissima* (33%), *and P. hasleana* (30%), but was completely absent in *P. multiseries*.

Alphaproteobacteria were omnipresent, but they were relatively more abundant in the cultures of the delicatissima-size class (35% - 44%), than in the seriata-size class (14% and 2.6%). In these cultures, the Alphaproteobacteria class was almost exclusively made up of bacteria of the family *Rhodobacteraceae*, but *Hyphomonadaceae* also appeared in low abundance in the cultures of *P. delicatissima* (0.9%), *P. multiseries* (0.4%), and *P. subpacifica* (0.2%). *Flavobacteriaceae* were highly abundant in the cultures of *P. delicatissima* (21%) and *P. hasleana* (31%), while rare in *P. multiseries* (4.3%) and *P. subpacifica* (0.4%), and completely absent in *P. galaxiae*.

In this experiment 29 of the 47 ASVs (62%) were significantly different between the microbiomes of the five *Pseudo-nitzschia* species (ANOVA, false discovery rate (FDR) adjusted p-value < 0.05; Fig. 4). Random forest analysis revealed the importance of the individual ASVs responsible for driving the separation between the *Pseudo-nitzschia* microbiomes. We present FDR adjusted p-values of ANOVA (Table S2) and random forest variable importance known as mean decrease accuracies (MDAs in Table S2; see MDA ranking in Fig. S9A).

3.3 Chemical features that distinguish *Pseudo-nitzschia*-culture metabolomes

To identify features that were either common to or distinguishing between the *Pseudo-nitzschia*-culture metabolomes, whole and dissolved metabolome samples were examined and treated as replicates. Out of the 4826 features produced in the experiment, there were 843 features not significantly different between the cultures (ANOVA, FDR adjusted p-value > 0.05) meaning they occurred in similar relative abundance in all cultures and are therefore referred to as common features. Of these, 40 had spectral matches (Table S3; see mirror plots of following features in Fig. S10) including numerous lipid-like compounds such as fatty acyls, glycerolipids, glycerophospholipids, and glycerophosphocholines, which are typically associated with cell

membranes and expected to be present in high abundance in all cultures. β-nicotinamide adenine dinucleotide [M+H] and [M+2H] participates in multiple cellular redox processes and is also in high abundance in every culture. Further, dipeptides (Glu-Leu [M+H], Lys-Phe [M+H], Glu-Tyr [M+H]) and amino acids (tryptophan [2M+H], phenylalanine [M+H], and its derivative homophenylalanine [M+H]) are common features. Tryptophan can act as a signaling molecule between *Pseudo-nitzschia* and its microbiome (Amin *et al.*, 2015). Other common metabolites were glutathione [M+H], an antioxidant found in algae (Dupont *et al.*, 2004; Boysen *et al.*, 2021), protoporphyrin [M+H], a precursor for chlorophyll *a* and heme, and biliverdin [M+H], a green bile pigment.

In contrast, 3983 features (83% of all features produced in the experiment) were significantly different between the species-specific metabolomes (ANOVA, FDR adjusted p-value < 0.05) and were further considered. A random forest model identified 347 features as primarily driving the distinction between the *Pseudo-nitzschia*-culture metabolomes. These features are referred to as distinguishing features (MDAs and cut-off shown in Fig. S9B) and their distribution across all samples is shown in Fig. 3. For ease of discussion, all other features produced in the experiment are referred to as non-distinguishing features. The separation between cultures was unambiguous (random forest prediction error known as out-of-bag error = 0%) for all species except *P. hasleana* and *P. galaxiae* which could not be separated as firmly (out-of-bag error = 20%; also see Fig. 3).

Of the 347 distinguishing features, only 12 features were annotated as library ID spectral matches (Fig. 3; putative names and Classyfire compound classifications labeled in bold). Three library IDs are related to DA, which is the toxin produced by *Pseudo-nitzschia* sp., and in our cultures, it was exclusively present in *P. multiseries* and *P. subpacifica*. DA was detected as the protonated compound [M+H], the [M+H-CH_2O_2] adduct, and methyl-DA [M+H] (Fig. 5A). All DA features were identified as distinguishing features between *Pseudo-nitzschia* cultures and were more abundant in the whole metabolome samples compared to the dissolved metabolomes, which is expected

since the toxin is produced by the algal cells. In the toxic cultures we also detected 7'-carboxy-*N*-geranyl-L-gluatamic acid [M+H], which is a precursor in the biosynthesis of DA (Brunson *et al.*, 2018). Pantothenic acid [M+H-H_2O], also known as vitamin B5, was found primarily in *P. delicatissima.* This vitamin is thought to be ubiquitous (Williams *et al.*, 1933) and was not added to the growth media, unlike B1 (thiamine), B7 (biotin) and B12 (cobalamin), which can be limiting nutrients. Pantothenic acid is known to be an essential vitamin required for the synthesis of coenzyme A and can be produced by eukaryotes as well as prokaryotes (Morimura, 1959; Shah and Vaidya, 1977; Fabregas and Herrero, 1990; Marek-Kozaczuk and Skorupska, 2001). In this study pantothenic acid and related compounds (Fig. 5C) were relatively more abundant in the dissolved metabolome compared to the whole metabolome samples (Fig. 3), which could indicate bacterial origin. A more detailed table of all library IDs as well as mirror plots of the query spectra compared against the spectra of library standards can be found in Table S3 and Fig. S11.

An additional 31 features had spectral matches as analogs (Fig. 3; Classyfire compound classifications labeled). Compounds classified as lysophosphatidylcholines, amino acids and derivatives, 2-benzopyrans and dialkyl ethers were enriched in the cultures of *P. multiseries* and *P. subpacifica*, while peptides and prostaglandins were found in higher abundance in *P. hasleana* and *P. galaxiae*.

Based on the assigned MFs we can investigate the elemental composition (number and ratios of elements) and other chemical properties such as the m/z, retention time (i.e., polarity), double bond equivalent (DBE) and aromaticity index (AI). A one-tailed t-test revealed that the distinguishing features are significantly enriched in nitrogen atoms and have a higher nitrogen-to-carbon ratio than non-distinguishing features (FDR adjusted p-value < 0.001). Although the positive ionization method is expected to be selective for some nitrogen-containing compounds, the nitrogen enrichment in distinguishing features as revealed by the t-test is independent of any

ionization bias. While the t-test identifies distinguishing features as enriched in nitrogen, it cannot reveal the distribution of these features among the *Pseudo-nitzschia* cultures. In order to determine the evenness of nitrogen-containing compounds across the *Pseudo-nitzschia* cultures, we isolated the distinguishing features and chose the most abundant features in each culture. Features were considered abundant if their abundance exceeded the mean abundance of all distinguishing features in a sample. Of the non-distinguishing features, 61% of the MFs contained one or more nitrogen atoms, while the percent of abundant distinguishing features that contained nitrogen in each *Pseudo-nitzschia* culture ranged from 65-83% (Fig. S12).

3.4 Compound classes that distinguish *Pseudo-nitzschia* culture metabolomes

The MFs provide important information about the elemental composition of the metabolome, but to obtain structural information about the detected features, compound class probabilities predicted by *in-silico* annotations issued through CANOPUS were used. The latter revealed chemical-class differences between the algae cultures.

Out of the 532 CANOPUS compound classes, 61 classes were found to have significantly higher probabilities for the distinguishing features compared to the non-distinguishing features (one-tailed t-test FDR adjusted p-value < 0.001) while their probability exceeds 0.5 in at least one distinguishing feature. It stands out that 50 of the 61 compound classes (82%) are nitrogen-containing by definition (Table S4), whereas only 44% of the total 532 are organonitrogen compound classes (Table S5). Next, we determined whether the chemical features separating species-specific metabolomes contained unique chemical characteristics. Therefore, we calculated the mean compound class probability for the abundant distinguishing features in each sample for the 61 CANOPUS compound classes. The PCoA of the resulting data table (Fig. S5A)

conserved the trends of culture separation observed when examining all features produced in the culture (Fig. 2B). A PERMANOVA confirmed that this compound class-based metabolome was also species-specific (p-value < 0.001), and in fact, all 61 compound classes were significantly different between the cultures (ANOVA, FDR adjusted p-value < 0.05). Based on a subsequent random forest analysis most of the top compound classes (22 of 25) responsible for driving differences between the *Pseudo-nitzschia* species are nitrogen-containing by definition (see Fig. 4; MDAs and cut-off shown in Fig. S9C). While there are distinctive organonitrogen compound classes across all *Pseudo-nitzschia* species, nitrogen heterocyclic compound classes have their highest probabilities in the delicatissima-size class cultures (Fig. 4).

4. Discussion

The *Pseudo-nitzschia* cultures grown in this experiment showed different characteristics in growth, associated microbiomes, and whole and dissolved metabolomes. The chemical diversity of the metabolome could thus be driven by *Pseudo-nitzschia* size-class similarities, ecosystem role and microbiome composition. Common features of all five cultures were metabolites involved in basic cell function, such as lipids, cofactors, nucleotides, dipeptides and amino acids, antioxidants, and chlorophyll catabolites. Other common compounds are known to influence diatom-bacteria interactions such as the signaling molecule tryptophan (Amin *et al.*, 2015). Since metabolomes and microbiomes are species-specific, exploring the distinguishing features might reveal new compounds involved in diatom-bacteria interaction. We examined the chemodiversity of the distinct metabolomes using an untargeted approach. As a result, only 4.7% of features could be reliably identified based on matches to existing compound libraries. Thus, we applied an *in-silico* approach that illuminated the compound class affiliations of 56% of detected features. Following this analysis, we found that particular compound classes were differentially distributed among the *Pseudo-nitzschia* cultures and that their composition was independent of cell density.

Furthermore, nitrogen-containing compounds were overrepresented within these distinguishing compounds. Finally, a hierarchical cluster analysis enabled the visualization of compound classes and microbiome members that shared a similar abundance pattern across the cultures, providing a catalog of compound classes that can be targeted in future experiments to be further explored as potential mediators in microbial interactions of this ecologically relevant diatom.

4.1 Different lifestyles of *Pseudo-nitzschia* size classes

Even though culture conditions were the same for all cultures, the delicatissima-size class (*P. delicatissima*, *P. hasleana* and *P. galaxiae*) and seriata-size class (*P. multiseries* and *P. subpacifica*) were distinct in terms of growth rate and density, the microbiome, number of attached and free-living bacteria cells, toxicity, and their metabolomes. The delicatissima-size species grew faster and more densely than the seriata-class species, as reflected in chlorophyll *a* concentration and *Pseudo-nitzschia* cell counts (Fig. S1 and S2A). Of the five *Pseudo-nitzschia* species cultured here, only *P. subpacifica* and *P. multiseries* produced the neurotoxin DA, its derivatives and precursors. Both have been previously identified as toxin-producers (Lelong *et al.*, 2012; Fernandes *et al.*, 2014; Bates *et al.*, 2018), but the ability to produce DA is not generally dependent on size class (Lelong *et al.*, 2012) or cell density. For example, in our experiment conducted in 2016 *P. subpacifica* and *P. multiseries* attained similar densities and had well-matched growth curves (Fig. S1C), but only *P. multiseries* produced DA. In contrast, these cultures diverged in their density and growth rates during the experiment in 2017 (Fig. S1B), but both produced DA. It is possible that *P. subpacifica* started producing DA in response to stress associated with aging, because about one month after the experiment was completed the *P. subpacifica* stock culture died, likely as a result of normal cell cycle processes. DA production has been described to be triggered by numerous stressors, such as nutrient availability, temperature, salinity, pH, and irradiance (Lelong *et al.*, 2012), but age and cell cycle processes

also influence DA production (Sauvey *et al.*, 2019). The ability of *P. subpacifica* to switch on the biosynthesis of DA is interesting and may be widespread. DA production is not dictated by phylogenetic relationships either, as *P. subpacifica* is more closely related to *P. hasleana* than to *P. multiseries* (Quijano-Scheggia *et al.*, 2020). While the role of DA for *Pseudo-nitzschia* or its microbiome is still not well understood, DA and derivatives are important compositional drivers differentiating species-specific metabolomes in this study, in particular by separating toxic from non-toxic *Pseudo-nitzschia* species of this genus.

Overall, the metabolomes of the *Pseudo-nitzschia* cultures are distinct between the toxin vs. non-toxin producers. Since the toxin producers tested in this study included only seriata-size species, we cannot determine whether size class or toxin production drives the observed similarity of these metabolomes. The delicatissima-size cultures also hosted a higher abundance of both, attached and free-living bacteria, than the seriata-size classes (Fig. S2C and E) that could be traced to a greater flux of dissolved organic matter to support heterotrophic bacteria. However, the ratio of attached bacteria to *Pseudo-nitzschia* cells appeared to be independent of measured DOC concentrations since it was similar across cultures (Fig. S2B and F), but the trends during the algal growth cycle were variable. The ratio of attached to free-living bacteria shifted to more attached bacteria than free-living bacteria when DOC was enriched in the cultures (Fig. S2D). It is possible that bacteria in the *Pseudo-nitzschia* cultures may cycle through attached and free-living life stages as suggested by Guannel *et al*. (2011), who showed that the community composition of the attached and free-living bacteria was similar. This may be consistent with the presence of *Flavobacteriaceae* for example, in *P. hasleana*, which have the potential to be motile (e.g., Johansen *et al.*, 2002).

4.2 Microbiomes and metabolomes are species-specific

Our isolation and growth conditions resulted in *Pseudo-nitzschia* associated microbiomes that were host-specific, independent of when or where the initial diatom cells were isolated, as demonstrated by previous work (Guannel *et al.*, 2011; Sison-Mangus *et al.*, 2014). Further, Guannel *et al.* (2011) showed that *Pseudo-nitzschia* microbiomes are stable during *Pseudo-nitzschia* growth cycles and over a long time period (up to a year). Here, we confirm that the microbiomes of cultures isolated in 2016 were also stable over a year. The vast majority of identified bacteria associated with our *Pseudo-nitzschia* cultures has been described previously in microbiomes of *Pseudo-nitzschia* cultures or natural blooms, at least on the genus or family level (Table S2). Similar microbial communities have also been reported more generally in diatom microbiomes (e.g., Fu *et al.*, 2020; Shibl *et al.*, 2020). Overall, the most common families were *Alteromonadaceae, Rhodobacteraceae,* and *Flavobacteriaceae*, which are generally considered to be copiotrophic organisms previously identified as important partners of diatoms (Grossart *et al.*, 2005; Kaczmarska *et al.*, 2005; Sapp, Schwaderer, *et al.*, 2007; Sapp, Wichels, *et al.*, 2007; Sison-Mangus *et al.*, 2016). Copiotrophic bacteria are anticipated to thrive under high organic matter regimes, such as bloom conditions or the phycosphere of diatoms. For example, *Rhodobacteraceae* are well known to be closely associated with diatom phycospheres, and recent work has shown *Alteromonadaceae* can also become enriched in diatom cultures over longer periods of time (Shibl *et al.*, 2020). *Flavobacteriaceae* are commonly found during and following algal blooms, where they play important roles in the recycling of nutrients and degradation of organic matter including complex carbohydrates (Teeling *et al.*, 2012; Buchan *et al.*, 2014). We propose that bacteria growing in our cultures were attached to the phycosphere when *Pseudo-nitzschia* cells were isolated or were free-living and were isolated in the micropipette used to isolate individual diatom cells. In either case, these communities must have been well adapted to thriving under our specific culturing conditions.

Our results for *Pseudo-nitzschia* cultures showed that metabolomes of the same genus, cultured under the exact same conditions, displayed distinct chemotypes. It may be anticipated that the majority of compounds produced by *Pseudo-nitzschia* are common across species, but this study reveals that 83% of features were significantly different in abundance across the five *Pseudo-nitzschia* cultures. Here we show both, whole and dissolved metabolomes, are distinct implying that extra- and intracellular metabolites are species-specific. We also observed that metabolites released into the environment do not simply reflect intracellular composition (also see Johnson *et al.*, 2021). While bacterial partners may produce compounds themselves, they also "filter" the metabolites that exits the algal cell. Furthermore, hydrophilicity will also determine which compounds accumulate in the media. However, the distinct chemotypes could also be the result of the unique ecological niches that these different *Pseudo-nitzschia* species occupy in the environment and could drive the distinct microbiomes.

4.3 Structurally diverse nitrogen metabolites distinguish species-specific metabolomes

Following the comprehensive *in-silico* assignment of chemical information, we were able to determine that metabolites largely responsible for the distinction of species-specific metabolomes of *Pseudo-nitzschia* were enriched in both the number of nitrogen atoms and their nitrogen-to-carbon ratio. In fact, all spectral library matches and the majority of analog matches are identified as nitrogenous and are classified as amino acids and derivatives, peptides, lysophosphatidylcholines, vitamin related compounds or other amines and amides (Fig. 3). We would expect freshly produced organic matter to be enriched in these compounds. Our findings showed that nitrogen compounds were prevalent across all cultures, however the compounds distinguishing the metabolomes are even more enriched in nitrogen. Also, the majority of compound classes separating the metabolomes are nitrogen-containing and cover a range of

chemical classes (Fig. 4). While the cultures of *P. multiseries*, *P. subpacifica*, and *P. delicatissima* were differentiated by metabolites with high compound class probabilities associated with amino acids and derivatives as well as simple amines and amides, the diatom species, *P. delicatissima*, *P. hasleana* and *P. galaxiae* metabolomes were distinguished by cyclic compounds, such as imidazoles, pyrrolidines and lactams. At this point we can only speculate what implications metabolites of these compound classes have and why they differ between the *Pseudo-nitzschia* size classes in this study. One potential explanation could be that fast-growing species have a higher proportion of compounds associated with growth and cell division, such as nucleic acids, pigments and membrane lipids, whereas resource limited strains may be relatively enriched in enzymes or may be employing amino acid salvage pathways. Lactams are prevalent in antibiotic natural products and may be important for guarding against extensive microbial colonization.

Much research has been done to establish that dissolved organic nitrogen is at least as important as or even more important than inorganic nitrogen, in regard to *Pseudo-nitzschia* growth and blooms (Hillebrand and Sommer, 1996; Howard *et al.*, 2007; Cochlan *et al.*, 2008; Loureiro *et al.*, 2009). These studies focus on certain small molecules, such as urea, glutamate, and glutamine. Here, we add evidence that diverse nitrogenous organic compounds could also serve as a nutrient source, which in turn could support species-specific microbiomes of *Pseudo-nitzschia*. Especially when inorganic nutrients are depleted, some of these organic nitrogen compounds could sustain *Pseudo-nitzschia* blooms in the form of organic nutrients or inorganic nutrients recycled by associated bacteria in the phycosphere or surrounding waters.

There is also evidence that nitrogen-containing organic molecules are preferentially utilized by bacteria and relevant compound classes can include nucleosides and amino acids. For example, a member of the *Polaribacter* genus can utilize peptides (Ferrer-González *et al.*, 2021)**.** In our study, *Polaribacter* is also highly abundant in *Pseudo-nitzschia* cultures that are differentiated

from other species, because they are enriched in alpha amino acids and derivatives and dialkylamines (Fig. 4). Some of the nitrogen compounds that are observed may also serve as precursors for secondary metabolite biosynthetic pathways in bacteria. We posit that the diverse nitrogen metabolites observed here are important in *Pseudo-nitzschia*-microbiome interactions and therefore ultimately in *Pseudo-nitzschia* ecology.

4.4 Illuminating the dark matter of the *Pseudo-nitzschia* metabolome

In metabolomic studies the gold standard for compound identification is a comparison of both chromatographic retention time and MS/MS spectra with an authentic standard, when analyzed under identical conditions. This is most commonly done as a targeted approach, where metabolites of interest are quantified, and their different abundances can provide insights into their role in experiments or the environment. Alternatively, untargeted approaches tend to annotate tens to hundreds of metabolites based on local or public libraries and/ or assign elemental composition to thousands of metabolites (Wienhausen *et al.*, 2017; Vorobev *et al.*, 2018; Heal *et al.*, 2019, 2021; Dawson *et al.*, 2020; Fiorini *et al.*, 2020; Johnson *et al.*, 2020, 2021; Longnecker and Kujawinski, 2020; Weber *et al.*, 2020; Petras *et al.*, 2021). However, even such an approach leaves thousands of detectable compounds uncharacterized as chemical dark matter (da Silva *et al.*, 2015), and these compounds might be crucial for microbial interactions, biogeochemical processes in the ocean or have potential as natural products in drug discovery.

Less than 5% of all detected features could be annotated by spectral matching with the MS/MS spectral libraries, which is typical for marine samples. Therefore, spectral library matches alone are insufficient to provide a comprehensive characterization of the metabolome, which hinders our ability to identify the chemical tokens exchanged during microbe-microbe interactions.

Spectral matches of analogs, which allow classification of compounds at the chemical family level, increased the annotation of our dataset to 12%. Since compounds with similar MS/MS spectra cluster together in subnetworks within our broader molecular network, we know they are structurally related and can infer chemical information when a spectral match is present for that subnetwork. Taking this approach and propagating the annotation provided by library ID or analog matches within molecular subnetworks, we obtained chemical information for about 25% of the features in this dataset. Nonetheless, our global molecular network demonstrates how only some subnetworks are well annotated, while the majority remain unknown (Fig. 6A). This means the chemical diversity of the metabolome is not comprehensively captured by the use of spectral libraries alone and the majority of the metabolome remains as dark matter.

Using *in-silico* tools that work independently of spectral libraries, we attempted to expand the chemical space annotated within our global *Pseudo-nitzschia*-microbiome network. We successfully illuminated some of the detectable chemical dark matter by providing MFs and compound class affiliations for 2710 features, i.e., 56% of the features produced in the cultures. If compound class affiliations are propagated throughout the subnetworks where an *in-silico* annotation is present, 75% of features could be annotated on the compound class level. Overall, *in-silico* annotations covered our global network more comprehensively (Fig. 6B). Furthermore, the tools used in this study can predict chemical information for truly unknown compounds. As an example, ZODIAC was not only able to improve the correct assignment of MFs, Ludwig *et al.* (2020) also demonstrated that two novel MFs were found in the dataset of the current study: a brominated phosphocholine and a polyhalogenated compound.

To examine the quality and level of chemical annotation provided by the *in-silico* workflow we turned to the subnetwork that contained DA (Fig. 5A). We found that the MFs and CANOPUS chemical class information including specific functional group information (e.g., high probabilities for azacyclic compounds, and more specifically pyrrolidine carboxylic acids and kainoids), which are entirely independent of library identifications, were indeed correct. The pantothenic acid

subnetwork (Fig. 5C) has probabilities for azacyclic and benzenoids close to zero, but probabilities are high for alcohols and polyols and N-acyl amides, which agrees with the known structure of pantothenic acid. CSI:FingerID predicted the correct structure as the top candidate for the DA, methyl-DA, pantothenic acid and its analog (Fig. S13A-D).

We then turned our attention to two subnetworks where no features were linked to spectral matches (Fig. 5B and 5D). These subnetworks were also interesting because their features are unique to only two *Pseudo-nitzschia* species. Furthermore, all features had relatively high m/z values, which can typically hinder accurate MF assignments and therefore, subsequent compound class and structure predictions (see Table S6 for assigned MF). In both subnetworks, the compound classes azacyclic compounds and benzenoids had high probabilities for all features. The CSI:FingerID similarity of the predicted structures is lower for these compounds (46-51%; Fig. S13E-H) than for the known compounds DA and pantothenic acid (78-86%; Fig. S13A-D), because the compounds are likely to be truly unknown and absent from structural databases. While we cannot predict the whole structure, we are able to manually extract information for putative substructures common to the features in the subnetworks (Fig. 5B and 5D) by combining the various *in-silico* annotation tools including the fragmentation trees provided by SIRIUS and further exploring molecular structures predicted by CSI:FingerID (Fig. S13) (Dührkop *et al.*, 2015, 2019, 2020). Subnetwork I contains benzene rings and the putative substructure of dimethyl-dihydropyridine-carboxaldehyde. The natural product literature does not contain reports of similar compounds and thus we cannot provide definitive identification of the substructure or hypothesize about the function of these compounds. The compounds in subnetwork II are putatively annotated as tetrapyrroles. Pigments, vitamin B12, and heme are all common metabolites containing tetrapyrrole moieties and they can originate from bacteria or microalgae and serve a variety of purposes for the cell.

Such a manual interrogation of the available data reinforced our confidence in the utility of our method and the compound class level information propagated for the global network. Furthermore, this dataset had the added benefit of serving to independently validate the approach we took to identify compounds that differentiate the interrogated samples. In this case, we could anticipate how the whole and dissolved metabolomes would differ, and the dataset and statistical approach supported this prediction. That is, the difference between the dissolved and whole metabolome was driven primarily by features identified as cell membrane compounds for example, such as phospholipids, which are expected to have low solubility (see detailed description in the Supplementary Information). Since the bioinformatic tools and novel analytical workflow produced expected results in this validation, it gave us the confidence to extend to a more open-ended question regarding the chemical nature of compounds that distinguished the metabolomes of individual *Pseudo-nitzschia* species (discussed in the previous section).

To our knowledge, this is the first study in which complex marine metabolomes have been assigned structural information at the compound class level at the scale of thousands of metabolites. The structural motifs identified for the unknown metabolites in this study could inform the choice of specific model compounds to examine in experimental systems. Since the data is publicly available and accessible for exploration and re-analysis (Jarmusch *et al.*, 2020) and MS/MS spectra can be searched for (equivalent to WebBLAST for sequencing data; (Wang *et al.*, 2020)), this study contributes to a fast-growing repository of MS/MS data which will transform our ability to use metabolomics to understand environmental systems in the future. Thanks to culture-based catalogs such as the data produced in this study, the origin of metabolites of interest in the environment may be revealed. The combination of public data sets and diverse cheminformatic tools for both annotation and analysis of mass spectrometry data are critical for illuminating dark matter and will eventually lead to a better understanding of microbial communities in the environment.

Acknowledgements

This research is funded in part by the Blasker Environmental Grant of the San Diego Foundation (BLSK201676272) to LIA, IK and MEW as well as by the NSF (OCE-1155269 to L.I.A.). Funds for facility support were also provided by the National Institute of Environmental Health Sciences/NSF Oceans and Human Health program, NIH (P01-ES021921), and NSF (OCE-1313747). Support for IK came also from various UCSD fellowships, including the Frontiers Of Innovation Scholars Program, the Edna Bailey Sussman Foundation Fellowship and the Rita L. Atkinson Fellowship. We acknowledge support from the National Science Foundation (NSF-OCE-1637632 and NSF-OCE-1756884), the National Oceanic and Atmospheric Administration (NA15OAR4320071 and NA19NOS4780181), and the Gordon and Betty Moore Foundation (GBMF3828) to AEA. Support for ZAQ came from NSF GRFP. DP, KD, ML and SB were supported by Deutsche Forschungsgemeinschaft (PE 2600/1 to DP and BO 1910/20 and 1910/23 to KD, ML and SB). P.C.D. was supported by the Gordon and Betty Moore Foundation (GBMF7622) and the U.S. National Institutes of Health (P41 GM103484, R01 GM107550). In addition, we thank the following people: Hong Zheng, Ryan Guillemette, Brandon Stephens, Irazema Islas, and Wei Ting Hua for technical assistance; and Bryce Inman for assistance with graphical design.

Competing Interests

SB, KD and ML are co-founders of Bright Giant GmbH. P.C.D is a scientific advisor to Sirenas and Cybele and co-founder and scientific advisor to Ometa and Enveda with approval by the University of California San Diego. The other authors declare no competing interests.

References

Amin, S.A., Hmelo, L.R., van Tol, H.M., Durham, B.P., Carlson, L.T., Heal, K.R., et al. (2015) Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522**: 98–101.

Amin, S.A., Parker, M.S., and Armbrust, E.V. (2012) Interactions between Diatoms and Bacteria. *Microbiol Mol Biol Rev* **76**: 667–684.

Armbrust, E.V., Berges, J.A., Bowler, C., Green, B.R., Martinez, D., Putnam, N.H., et al. (2004) The Genome of the Diatom Thalassiosira Pseudonana: Ecology, Evolution, and Metabolism. *Science* **306**: 79–86.

Bates, S.S., Douglas, D.J., Doucette, G.J., and Léger, C. (1995) Enhancement of domoic acid production by reintroducing bacteria to axenic cultures of the diatom Pseudo-nitzschia multiseries. *Nat Toxins* **3**: 428–435.

Bates, S.S., Hubbard, K.A., Lundholm, N., Montresor, M., and Leaw, C.P. (2018) Pseudo-nitzschia, Nitzschia, and domoic acid: New research since 2011. *Harmful Algae* **79**: 3–43.

Böcker, S., Letzel, M.C., Lipták, Z., and Pervukhin, A. (2009) SIRIUS: decomposing isotope patterns for metabolite identification. *Bioinforma Oxf Engl* **25**: 218–224.

Boysen, A.K., Carlson, L.T., Durham, B.P., Groussman, R.D., Aylward, F.O., Ribalet, F., et al. (2021) Particulate Metabolites and Transcripts Reflect Diel Oscillations of Microbial Activity in the Surface Ocean. *mSystems* **6**:.

Brunson, J.K., McKinnie, S.M.K., Chekan, J.R., McCrow, J.P., Miles, Z.D., Bertrand, E.M., et al. (2018) Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom. *Science* **361**: 1356–1358.

Buchan, A., LeCleir, G.R., Gulvik, C.A., and González, J.M. (2014) Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat Rev Microbiol* **12**:

686–698.

Cochlan, W.P., Herndon, J., and Kudela, R.M. (2008) Inorganic and organic nitrogen uptake by the toxigenic diatom Pseudo-nitzschia australis (Bacillariophyceae). *Harmful Algae* **8**: 111–118.

Dawson, H.M., Heal, K.R., Torstensson, A., Carlson, L.T., Ingalls, A.E., and Young, J.N. (2020) Large Diversity in Nitrogen- and Sulfur-Containing Compatible Solute Profiles in Polar and Temperate Diatoms. *Integr Comp Biol* **60**: 1401–1413.

Dittmar, T., Koch, B., Hertkorn, N., and 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.

Djoumbou Feunang, Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., et al. (2016) ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J Cheminformatics* **8**: 61.

Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A.A., Melnik, A.V., Meusel, M., et al. (2019) SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat Methods* **16**: 299–302.

Dührkop, K., Nothias, L.-F., Fleischauer, M., Reher, R., Ludwig, M., Hoffmann, M.A., et al. (2020) Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nat Biotechnol* 1–10.

Dührkop, K., Shen, H., Meusel, M., Rousu, J., and Böcker, S. (2015) Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc Natl Acad Sci* **112**: 12580–12585.

Dupont, C.L., Goepfert, T.J., Lo, P., Wei, L., and Ahner, B.A. (2004) Diurnal cycling of glutathione in marine phytoplankton: Field and culture studies. *Limnol Oceanogr* **49**: 991–996.

Fabregas, J. and Herrero, C. (1990) Vitamin content of four marine microalgae. Potential use as

source of vitamins in nutrition. *J Ind Microbiol* **5**: 259–263.

Fernandes, L.F., Hubbard, K.A., Richlen, M.L., Smith, J., Bates, S.S., Ehrman, J., et al. (2014) Diversity and toxicity of the diatom Pseudo-nitzschia Peragallo in the Gulf of Maine, Northwestern Atlantic Ocean. *Deep Sea Res Part II Top Stud Oceanogr* **103**: 139–162.

Ferrer-González, F.X., Widner, B., Holderman, N.R., Glushka, J., Edison, A.S., Kujawinski, E.B., and Moran, M.A. (2021) Resource partitioning of phytoplankton metabolites that support bacterial heterotrophy. *ISME J* **15**: 762–773.

Fiorini, F., Borgonuovo, C., Ferrante, M.I., and Brönstrup, M. (2020) A Metabolomics Exploration of the Sexual Phase in the Marine Diatom Pseudo-nitzschia multistriata. *Mar Drugs* **18**: 313.

Fu, H., Uchimiya, M., Gore, J., and Moran, M.A. (2020) Ecological drivers of bacterial community assembly in synthetic phycospheres. *Proc Natl Acad Sci* **117**: 3656–3662.

Grossart, H.-P., Levold, F., Allgaier, M., Simon, M., and Brinkhoff, T. (2005) Marine diatom species harbour distinct bacterial communities. *Environ Microbiol* **7**: 860–873.

Guannel, M.L., Horner-Devine, M.C., and Rocap, G. (2011) Bacterial community composition differs with species and toxigenicity of the diatom Pseudo-nitzschia.

Heal, K.R., Durham, B.P., Boysen, A.K., Carlson, L.T., Qin, W., Ribalet, F., et al. (2021) Marine Community Metabolomes Carry Fingerprints of Phytoplankton Community Composition. *mSystems* **6**:.

Heal, K.R., Kellogg, N.A., Carlson, L.T., Lionheart, R.M., and Ingalls, A.E. (2019) Metabolic Consequences of Cobalamin Scarcity in the Diatom Thalassiosira pseudonana as Revealed Through Metabolomics. *Protist* **170**: 328–348.

Hillebrand, H. and Sommer, U. (1996) Nitrogenous nutrition of the potentially toxic diatom Pseudonitzschia pungens f. multiseries Hasle. *J Plankton Res* **18**: 295–301.

Howard, M.D.A., Cochlan, W.P., Ladizinsky, N., and Kudela, R.M. (2007) Nitrogenous preference of toxigenic Pseudo-nitzschia australis (Bacillariophyceae) from field and

laboratory experiments. *Harmful Algae* **6**: 206–217.

Jarmusch, A.K., Wang, M., Aceves, C.M., Advani, R.S., Aguirre, S., Aksenov, A.A., et al. (2020) ReDU: a framework to find and reanalyze public mass spectrometry data. *Nat Methods* **17**: 901–904.

Johansen, J., Pinhassi, J., Blackburn, N., Zweifel, U., and Hagström, Å. (2002) Variability in motility characteristics among marine bacteria. *Aquat Microb Ecol* **28**: 229–237.

Johnson, W.M., Longnecker, K., Soule, M.C.K., Arnold, W.A., Bhatia, M.P., Hallam, S.J., et al. (2020) Metabolite composition of sinking particles differs from surface suspended particles across a latitudinal transect in the South Atlantic. *Limnol Oceanogr* **65**: 111–127.

Johnson, W.M., Soule, M.C.K., Longnecker, K., Bhatia, M.P., Hallam, S.J., Lomas, M.W., and Kujawinski, E.B. (2021) Insights into the controls on metabolite distributions along a latitudinal transect of the western Atlantic Ocean. *bioRxiv* 2021.03.09.434501.

Kaczmarska, I., Ehrman, J.M., Bates, S.S., Green, D.H., Léger, C., and Harris, J. (2005) Diversity and distribution of epibiotic bacteria on Pseudo-nitzschia multiseries (Bacillariophyceae) in culture, and comparison with those on diatoms in native seawater. *Harmful Algae* **4**: 725–741.

Lelong, A., Hégaret, H., Soudant, P., and Bates, S.S. (2012) Pseudo-nitzschia (Bacillariophyceae) species, domoic acid and amnesic shellfish poisoning: revisiting previous paradigms. *Phycologia* **51**: 168–216.

Longnecker, K. and Kujawinski, E.B. (2020) Intracellular Metabolites in Marine Microorganisms during an Experiment Evaluating Microbial Mortality. *Metabolites* **10**: 105.

Loureiro, S., Garcés, E., Fernández-Tejedor, M., Vaqué, D., and Camp, J. (2009) Pseudo-nitzschia spp. (Bacillariophyceae) and dissolved organic matter (DOM) dynamics in the Ebro Delta (Alfacs Bay, NW Mediterranean Sea). *Estuar Coast Shelf Sci* **83**: 539–549.

Ludwig, M., Nothias, L.-F., Dührkop, K., Koester, I., Fleischauer, M., Hoffmann, M.A., et al.

(2020) Database-independent molecular formula annotation using Gibbs sampling through ZODIAC. *Nat Mach Intell* **2**: 629–641.

Marek-Kozaczuk, M. and Skorupska, A. (2001) Production of B-group vitamins by plant growth-promoting Pseudomonas fluorescens strain 267 and the importance of vitamins in the colonization and nodulation of red clover. *Biol Fertil Soils* **33**: 146–151.

Morimura, Y. (1959) Synchronous culture of Chlorella: II. Changes in content of various vitamins during the course of the algal life cycle. *Plant Cell Physiol* **1**: 63–69.

Needham, D.M. and Fuhrman, J.A. (2016) Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nat Microbiol* **1**: 16005.

Nothias, L.-F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A., et al. (2020) Feature-based molecular networking in the GNPS analysis environment. *Nat Methods* **17**: 905–908.

Petras, D., Koester, I., Da Silva, R., Stephens, B.M., Haas, A.F., Nelson, C.E., et al. (2017) High-Resolution Liquid Chromatography Tandem Mass Spectrometry Enables Large Scale Molecular Characterization of Dissolved Organic Matter. *Front Mar Sci* **4**:.

Petras, D., Minich, J.J., Cancelada, L.B., Torres, R.R., Kunselman, E., Wang, M., et al. (2021) Non-targeted tandem mass spectrometry enables the visualization of organic matter chemotype shifts in coastal seawater. *Chemosphere* **271**: 129450.

Pluskal, T., Castillo, S., Villar-Briones, A., and Orešič, M. (2010) MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**: 395.

Quijano-Scheggia, S.I., Olivos-Ortiz, A., Garcia-Mendoza, E., Sánchez-Bravo, Y., Sosa-Avalos, R., Marias, N.S., and Lim, H.C. (2020) Phylogenetic relationships of Pseudo-nitzschia subpacifica (Bacillariophyceae) from the Mexican Pacific, and its production of domoic acid in culture. *PLOS ONE* **15**: e0231902.

Sapp, M., Schwaderer, A.S., Wiltshire, K.H., Hoppe, H.-G., Gerdts, G., and Wichels, A. (2007)

Species-specific bacterial communities in the phycosphere of microalgae? *Microb Ecol* **53**: 683–699.

Sapp, M., Wichels, A., and Gerdts, G. (2007) Impacts of cultivation of marine diatoms on the associated bacterial community. *Appl Environ Microbiol* **73**: 3117–3120.

Sauvey, A., Claquin, P., Le Roy, B., Le Gac, M., and Fauchot, J. (2019) Differential Influence of Life Cycle on Growth and Toxin Production of three *Pseudo-nitzschia* Species (Bacillariophyceae). *J Phycol* **55**: 1126–1139.

Shah, A.K. and Vaidya, B.S. (1977) Detection of vitamin B12 and pantothenic acid in cell exudates of blue-green algae. *Biol Plant* **19**: 426–429.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., et al. (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* **13**: 2498–2504.

Shibl, A.A., Isaac, A., Ochsenkühn, M.A., Cárdenas, A., Fei, C., Behringer, G., et al. (2020) Diatom modulation of select bacteria through use of two unique secondary metabolites. *Proc Natl Acad Sci*.

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

Sison-Mangus, M.P., Jiang, S., Kudela, R.M., and Mehic, S. (2016) Phytoplankton-Associated Bacterial Community Composition and Succession during Toxic Diatom Bloom and Non-Bloom Events. *Front Microbiol* **7**:.

Sison-Mangus, M.P., Jiang, S., Tran, K.N., and Kudela, R.M. (2014) Host-specific adaptation governs the interaction of the marine diatom, Pseudo-nitzschia and their microbiota. *ISME J* **8**: 63–76.

Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., et al. (2007) Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics Off J*

Metabolomic Soc **3**: 211–221.

Teeling, H., Fuchs, B.M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C.M., et al. (2012) Substrate-Controlled Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom. *Science* **336**: 608–611.

Tomas, C.R. and Hasle, G.R. (1997) Identifying marine phytoplankton, San Diego: Academic Press.

Trainer, V.L., Bates, S.S., Lundholm, N., Thessen, A.E., Cochlan, W.P., Adams, N.G., and Trick, C.G. (2012) Pseudo-nitzschia physiological ecology, phylogeny, toxicity, monitoring and impacts on ecosystem health. *Harmful Algae* **14**: 271–300.

Vorobev, A., Sharma, S., Yu, M., Lee, J., Washington, B.J., Whitman, W.B., et al. (2018) Identifying labile DOM components in a coastal ocean through depleted bacterial transcripts and chemical signals. *Environ Microbiol* **20**: 3012–3030.

Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., et al. (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* **34**: 828–837.

Wang, M., Jarmusch, A.K., Vargas, F., Aksenov, A.A., Gauglitz, J.M., Weldon, K., et al. (2020) Mass spectrometry searches using MASST. *Nat Biotechnol* **38**: 23–26.

Weber, L., Armenteros, M., Kido Soule, M., Longnecker, K., Kujawinski, E.B., and Apprill, A. (2020) Extracellular Reef Metabolites Across the Protected Jardines de la Reina, Cuba Reef System. *Front Mar Sci* **7**:.

Wienhausen, G., Noriega-Ortega, B.E., Niggemann, J., Dittmar, T., and Simon, M. (2017) The Exometabolome of Two Model Strains of the Roseobacter Group: A Marketplace of Microbial Metabolites. *Front Microbiol* **8**:.

Williams, R.J., Lyman, C.M., Goodyear, G.H., Truesdail, J.H., and Holaday, D. (1933) “Pantothenic Acid,” A Growth Determinant of Universal Biological Occurrence.

Fig. 1. Workflows demonstrating (A) the experimental design and sampling procedure and (B) computational tools used to analyze and annotate the mass spectra. Five different *Pseudo-nitzschia* species cultures were grown in biological duplicates and harvested at the end of the exponential growth phase. Cell pellets were used for 16S rRNA sequencing and microscopy samples for enumeration of attached and free-living bacteria were sampled during the growth phase. Metabolomic samples were taken in technical duplicates (50mL) for the whole metabolome (acidified to pH2 and sonicated) and for the dissolved metabolome (filtered and acidified to pH2). All samples were solid phase extracted and analyzed by non-targeted high-resolution liquid chromatography tandem mass spectrometry (HR LC-MS/MS) in data dependent acquisition mode. (B) MZmine 2 was used for feature extraction and calculation of feature intensities (XIC). GNPS is used for feature-based molecular networking and searching the public and commercial standard libraries for spectral matches. ClassyFire assigns chemical classification to the features with spectral matches. For each feature SIRIUS ranks molecular formula (MF) candidates based on MS1 isotope patterns and MS/MS spectra analysis; ZODIAC re-scores MF candidates of individual features by considering all features in the dataset to increase the number of correct MF annotations. The machine learning-based tool CANOPUS predicts probabilities of over 2,497 compound classes of each feature.

Fig. 2. PCoA plots showing the dissimilarities between (A) microbiomes and (B) metabolomes, color coded by the five *Pseudo-nitzschia* species. The relative abundance of 16S data was angular transformed and shows clear distinction between *Pseudo-nitzschia* species (PERMANOVA $p < 0.001$). MS1 features are TIC and chlorophyll *a* relativized and normalized (angular transformation). While filled circles correspond to the whole metabolome samples, empty circles represent dissolved metabolome samples. PC1 and PC2 reveal the separation between *Pseudo-nitzschia* species and dissolved versus whole metabolome respectively as indicated by the arrows. PERMANOVA analysis proved distinct *Pseudo-nitzschia* metabolomes as either

dissolved metabolites only, the whole metabolomes only, or the dissolved and whole metabolome together, as well as dissolved vs. whole metabolome (all $p < 0.001$). (D) depicts the distribution of the 16S microbial community on the class and family level for each *Pseudo-nitzschia* species by taking the mean abundance of culture replicates.

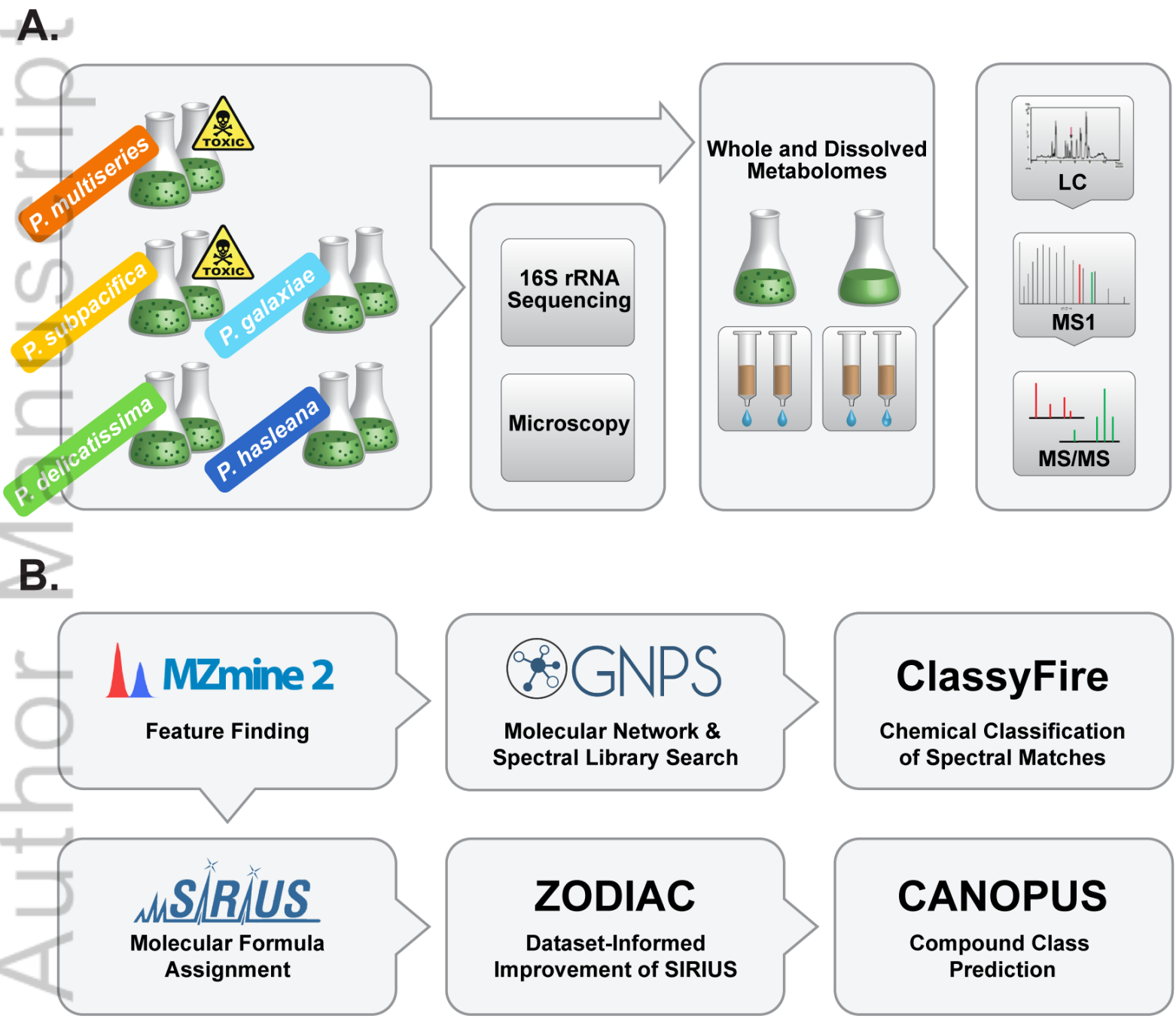
Fig. 3. Two-way cluster analysis showing the distribution of distinguishing features driving the separation between *Pseudo-nitzschia* species cultures. Area under the peak of MS1 features are TIC and chlorophyll *a* relativized and normalized (angular transformation). Z-score across dissolved (DM) and whole metabolome (WM) samples were calculated and are depicted in gray (see legend). Clusters were colored according to which *Pseudo-nitzschia* cultures the features were most abundant in. Features are labeled with putative compound names of their spectral library matches (library IDs, **bold**) and their chemical classifications (library IDs and analogs; using ClassyFire). Identification levels following the convention proposed by Sumner *et al.* (2007) are denoted by superscript following the compound name (see Supplemental Information 1.3.4 for definition). A more detailed table including adduct, measured precursor mass-to-charge ratio, the mass difference to the standard and all other gathered information about the given spectral match can be found in Table S3. Mirror plots of the spectral matches are shown in Fig. S11. Red and gray squares indicate if the annotated compound is nitrogen-containing or cyclic, respectively.

Fig. 4. Two-way cluster analysis showing the distribution of mean compound class probabilities for abundant distinguishing features (**bold**) and relative abundance of 16S data (*italic*). For compound classes dissolved metabolite samples were used and the mean of technical duplicates was calculated. Values depicted in gray were calculated z-scores for biological duplicates of the five *Pseudo-nitzschia* cultures. Clusters were colored according to the *Pseudo-nitzschia* species where values were the highest. The red and gray squares indicate if compound classes are nitrogen-containing or cyclic by definition. The colored dots indicate bacteria class.

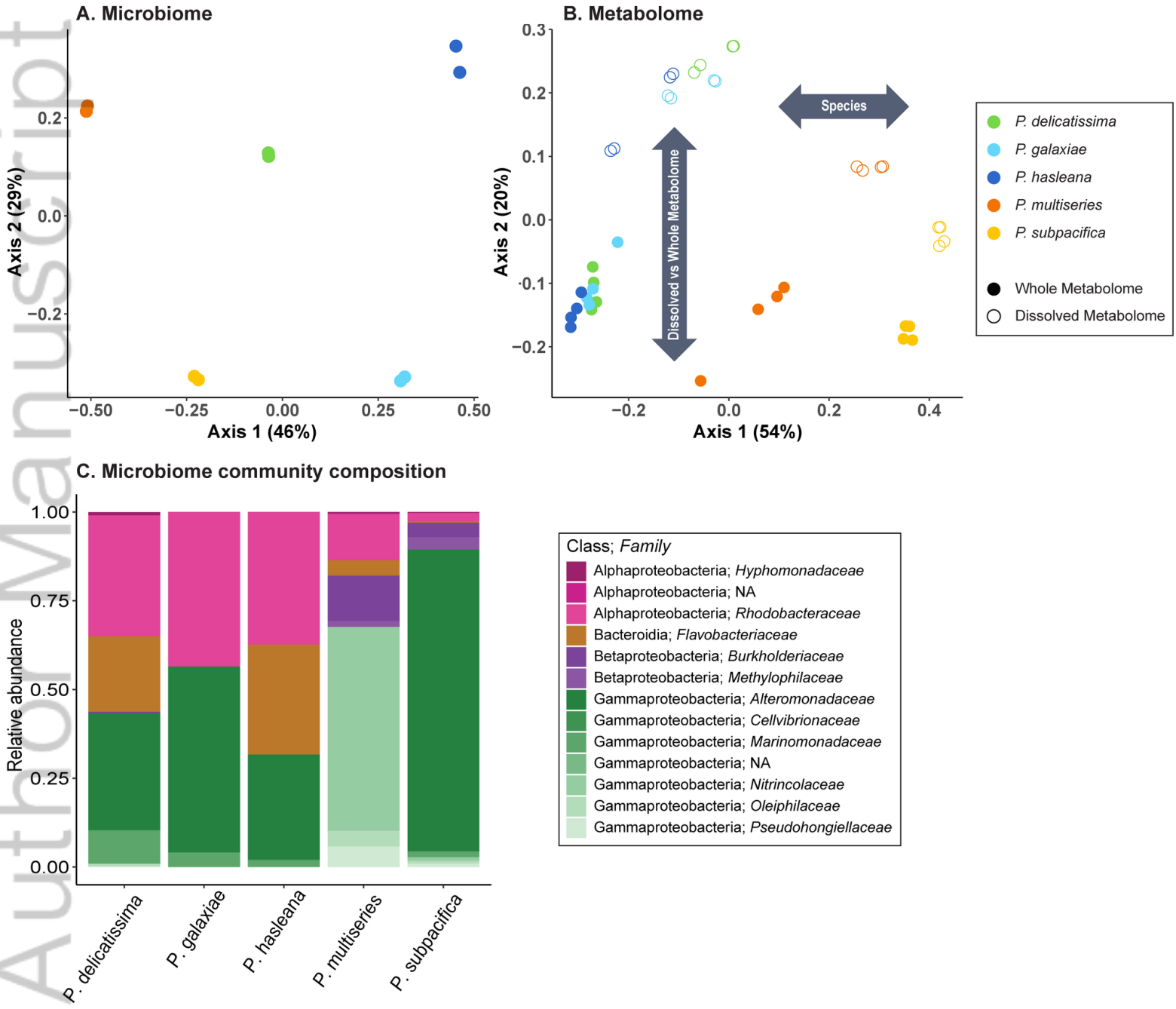
Fig. 5. Molecular subnetworks of DA, pantothenic acid and two unknown molecular families. Every node represents a feature labeled with the mass-to-charge ratio (m/z); the distribution of relative abundance across *Pseudo-nitzschia* cultures is visualized by pie charts. Features related to each other (cosine > 0.7) are connected by edges and result in a subnetwork. Subnetwork (A) has two library ID matches (DA and methyl-DA) and one analog match (methyl-DA). The other nodes represent adducts such as $[2M-H]^+$, covalently bound dimers and other DA derivatives. Compound class probabilities for the features are represented as bar graphs next to the nodes. Probabilities were high for azacyclic compounds (cyclic compounds with at least one nitrogen atom), pyrrolidine carboxylic acids and kainoids, which can all be confirmed by the known structure of DA. The pantothenic acid subnetwork (C) includes a library ID match with the adduct $[M+H-H_2O]$, an analog match, which is likely to be pantothenic acid with an acetate addition, and two other related features. The probabilities for azacyclic and benzenoids are close to zero, but probabilities are high for alcohols and polyols and N-acyl amides, which is coherent with the known structure. Subnetwork (B) and (D) represent unknown features only. While no spectral matches are available, the combination of *in-silico* annotation tools allowed the annotation of MFs (Table S6) and compound classes. Here, probabilities are high for the compound classes azacyclic compounds and benzenoids for both, but we were able to identify different substructures of the unknown subnetwork.

Fig. 6. Illuminating the dark metabolome of *Pseudo-nitzschia*-microbiome associations. Global network demonstrating the spatial distribution of (A) spectral matches (library IDs and analogs) and (B) *in-silico* annotation using the SIRIUS workflow (ZODIAC MFs, CANOPUS compound class predictions and CSI:FingerID structures). While GNPS library IDs are available for 4.7% of features, analog search increases this number to 12%. The spectral matches cover some subnetworks extensively, but most subnetworks have no annotations at all. Since subnetworks

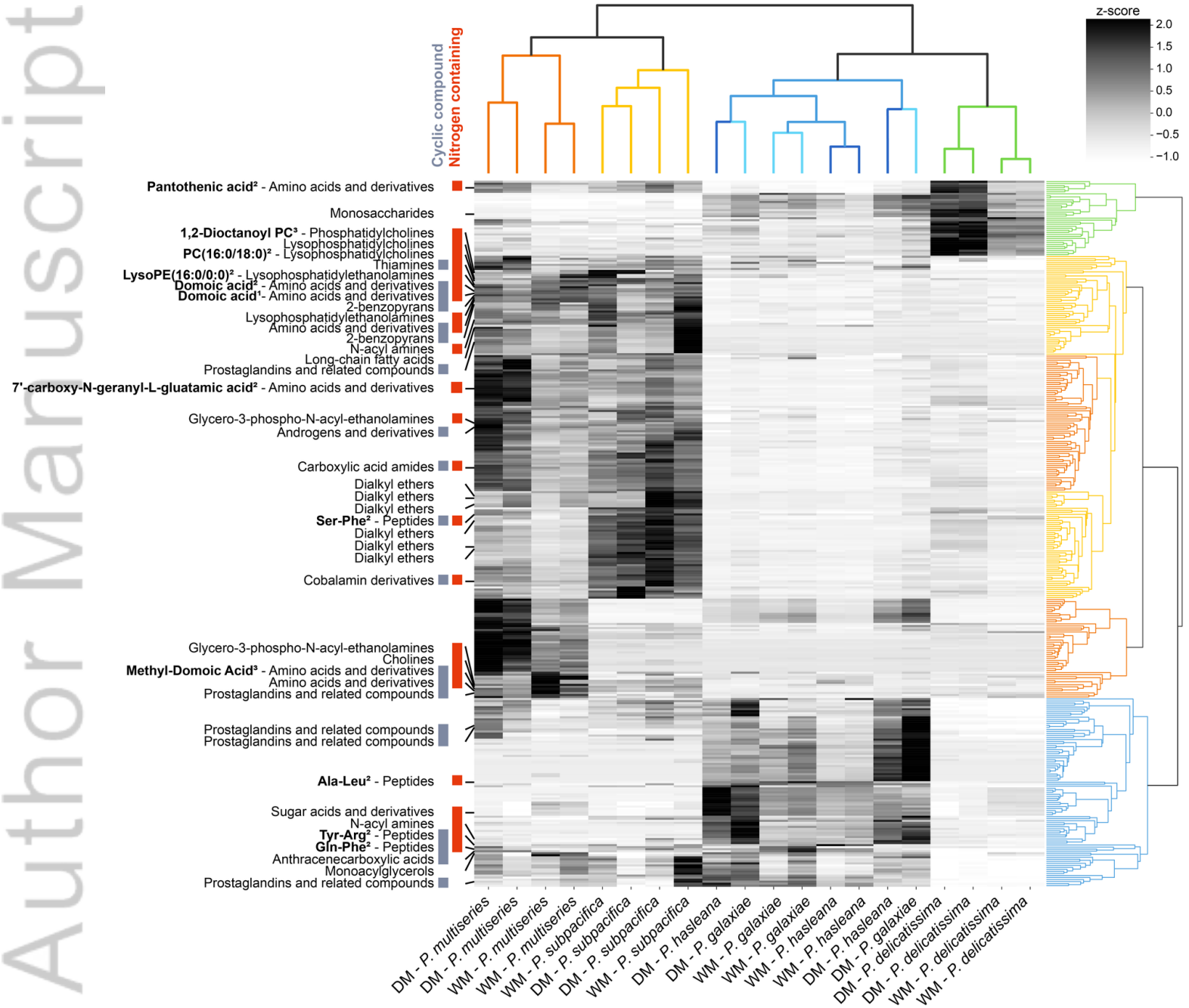
are structurally similar, we can propagate chemical classifications when spectral matches are present in a subnetwork, which means 25% of features can be classified (not shown). The SIRIUS workflow is able to annotate spectra absent from the spectral libraries and using the conservative cut-off of ZODIAC score >0.98, 56% of features were annotated. These features cover the global molecular network more comprehensively and if chemical classes are propagated within subnetworks, we have structure information about 75% of features in the dataset.



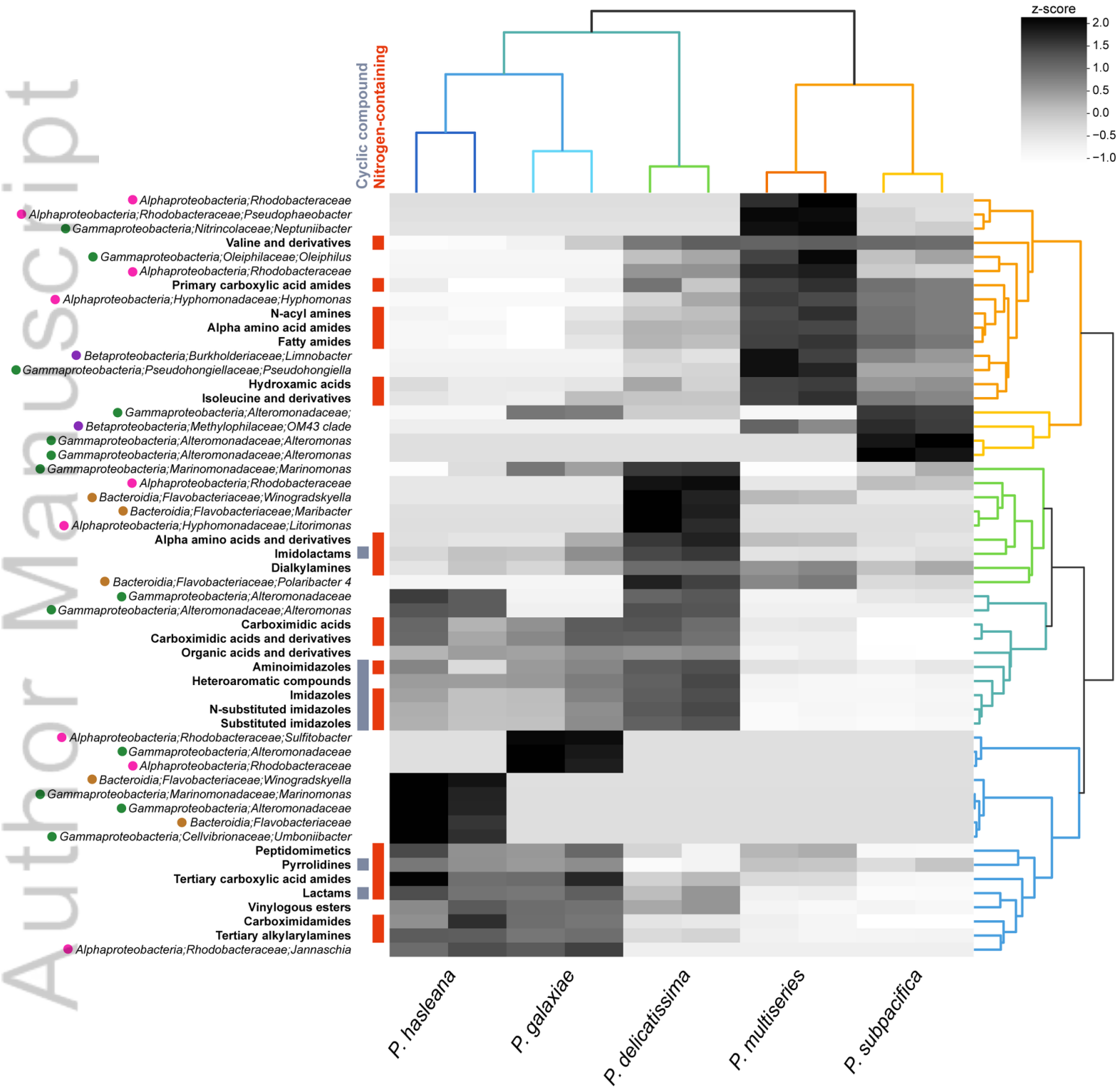
EMI_16242_Fig_1.png



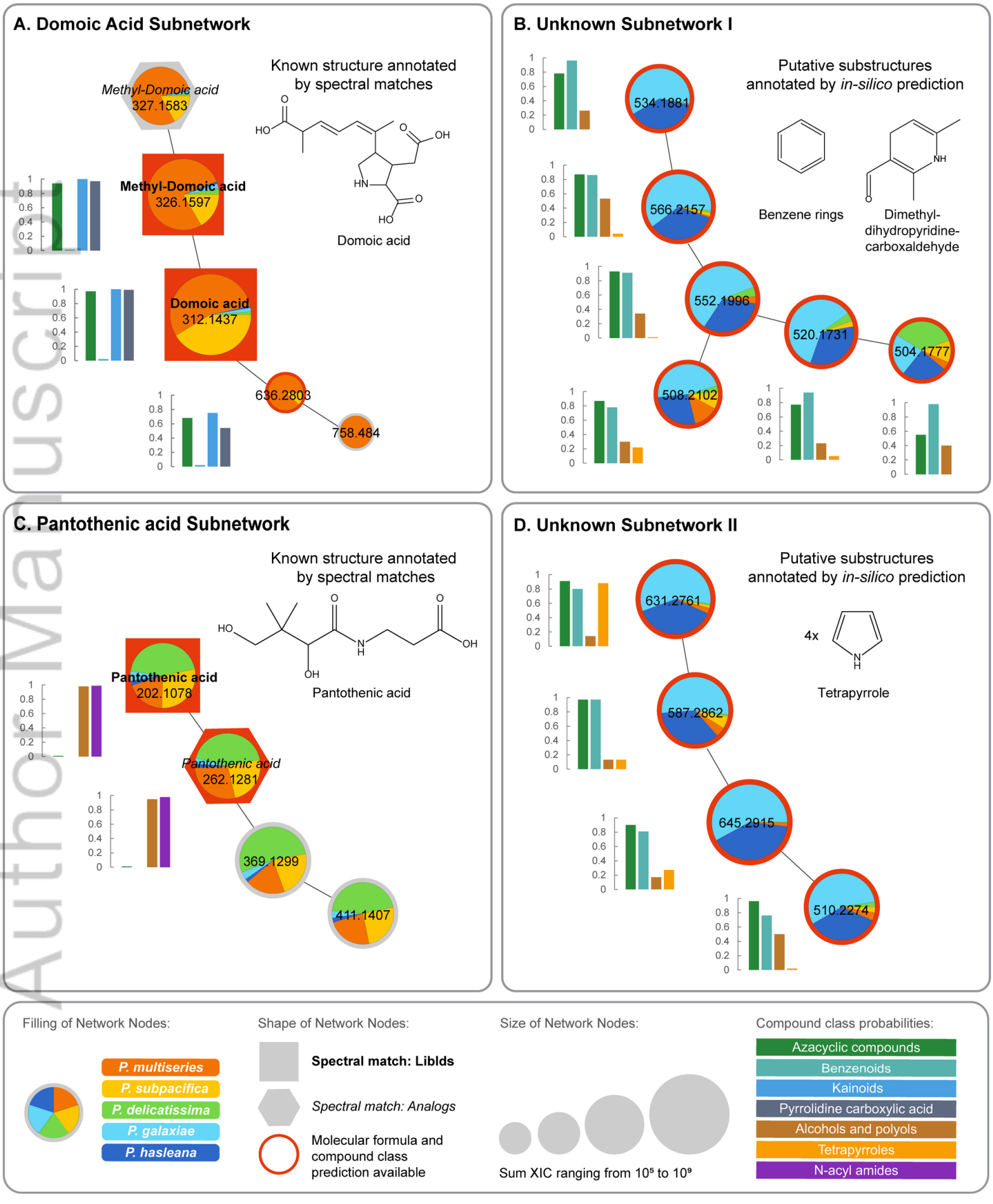
EMI_16242_Fig_2.png



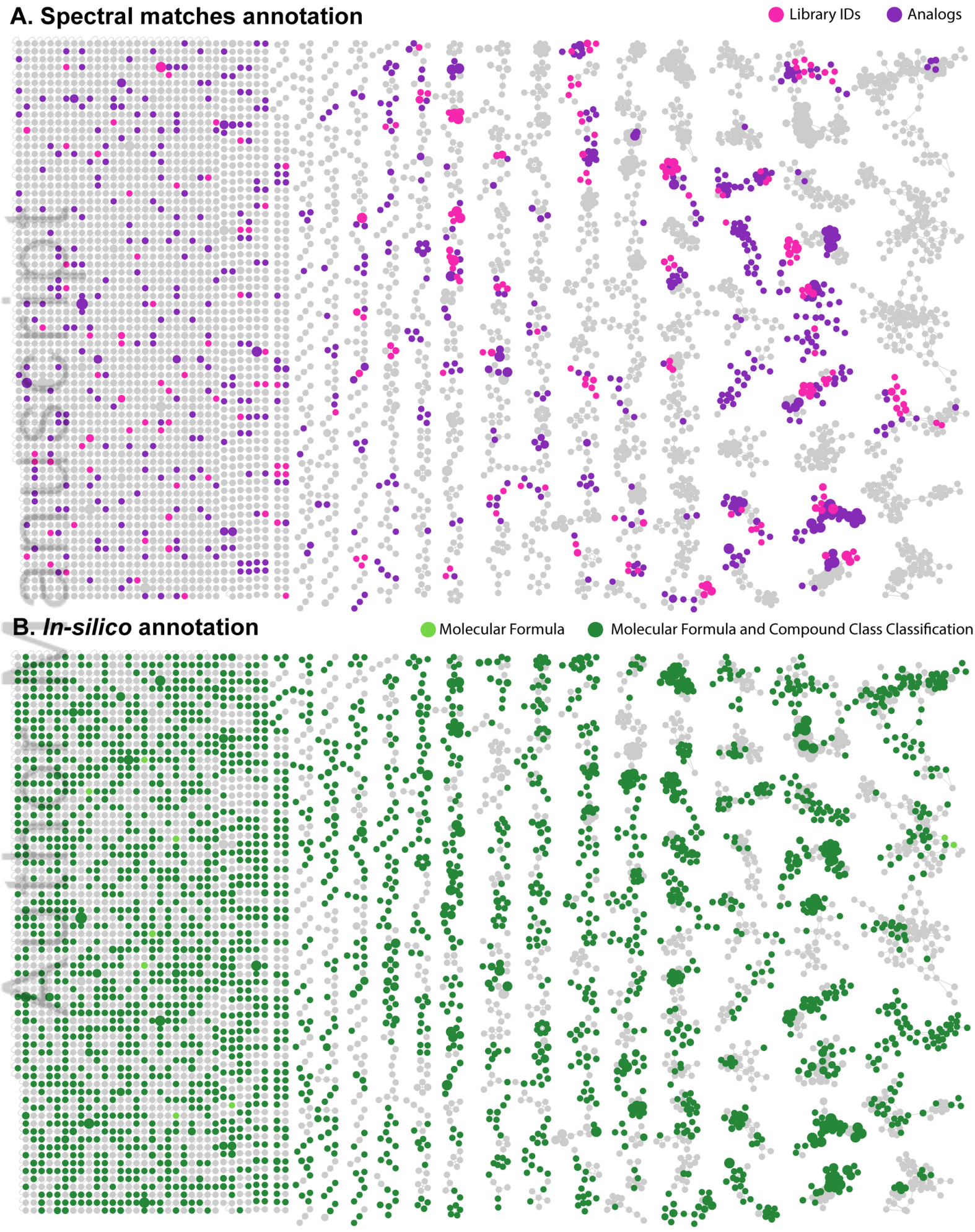
EMI_16242_Fig_3.png



EMI_16242_Fig_4.png



EMI_16242_Fig_5.png



EMI_16242_Fig_6.png