UNIVERSITY OF CALIFORNIA, SAN DIEGO

Investigating Microbial Metabolites with Novel Mass Spectrometry Tools

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Chemistry with a specialization in multi-scale biology

by

Jane Youngmi Yang

Committee in charge:

Professor Pieter C. Dorrestein, Chair
Professor Thomas Hermann
Professor Alexander Hoffmann
Professor Kit Pogliano
Professor Lei Wang

2013

The dissertation of Jane Youngmi Yang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2013

DEDICATION

I dedicate this to my family,

friends,

and Xiao Liwu the baby panda at the San Diego Zoo.

Thank you for all of the love, support, laughter, and cuteness.

TABLE OF CONTENTS

Signature Page……. iii

Dedication……..... iv

Table of Contents…… v

List of Figures……. viii

List of Tables…….. xi

Acknowledgements……. xii

Vita…….. xiv

Abstract of the Dissertation……. xvii

Chapter I

A primer on agar-based imaging mass spectrometry……. 1

Chapter II

Applications of agar-based imaging mass spectrometry…… 24

2.1 Abstract…… 25

2.2 Introduction…….. 26

2.3 Results…….. 33

2.4 Discussion…… 46

2.5 Future Directions…….. 53

2.6 Summary…….. 55

2.7 Experimental…… 57

2.7.1 MALDI-TOF MS profiling of *B. subtilis*……. 57

2.7.2 IMS profiling of *B. subtilis* mutants……. 58

2.8 Acknowledgements…….. 59

2.9 Supplemental Figures..……. 60

2.10 References……..…...……. 63

Chapter III

Molecular networking as a dereplication strategy……. 68

3.1 Abstract…… 69

3.2 Introduction…….. 70

3.3 Experimental section…… 77

3.3.1 Bacteria…… 77

3.3.2 Cyanobacteria……... 78

3.3.3 Molecular networking…….. 80

3.4 Results…….. 82

3.5 Discussion…… 95

3.6 References…… 101

3.7 Supporting Information…… 107

3.7.1 Cyanobacterial collection information……. 107

3.7.1.1 Collection that contained tumonoic acid I……... 107

3.7.1.2 Collection that contained carmabin A…….. 107

3.7.1.3 Collection containing barbamide……. 107

3.7.1.4 Collection that contained carmaphycin B…… 108

3.7.1.5 Collection of *Moorea bouillonii*…….. 108

3.7.2 Bacterial strain-specific culture conditions…….. 108

3.8 Supporting references…….. 130

Chapter IV

Future directions: mass spectrometry based networking of honey bee colony collapse disorder……. 133

4.1 Abstract (Project summary)…… 134

4.2 Background and significance…… 135

4.3 Specific aims……. 137

4.4 Research strategy…….. 139

4.5 Future directions……... 143

4.6 References……..…... 144

LIST OF FIGURES

Figure 1.1 Molecular snapshots from microbial MALDI-TOF imaging mass spectrometry, obtained on different agar surfaces…….. 3

Figure 1.2 The 5 major steps to obtain a microbial image are culturing, matrix application, dehydration of the samples, imaging mass spectrometry, and data analysis…… 4

Figure 1.3 Data interpretation challenges – flaking, black holes, and gradients…….. 5

Figure S1.1 How to minimize sample flaking inside instrument……. 20

Figure S1.2 Our hypothesis of action of dual-matrix……... 22

Figure S1.3 Multiple matrix applications to saturate agar samples with matrix…….. 22

Figure 2.1 Cell differentiation pathways in *B. subtilis*……. 32

Figure 2.2 Development of colony morphology and color of *B. subtilis* 3610 and PY79 on LB, DSM, and ISP2 media over 6 days…… 35

Figure 2.3 Gel/Stack view of 182 spectra reveals patterns of secondary metabolite production of *B. subtilis* 3610 and PY79 over time…….. 36

Figure 2.4 IMS of *B. subtilis* 3610 mutants……. 39

Figure 2.5 IMS of PY79 mutants……. 41

Figure 2.6 IMS of sigma and ECF factor mutants in *B. subtilis* PY79…… 43

Figure 2.7 Representative IMS images of unidentified mutants of SCB844 (*amyE*::*yidC2*^{6}-*lacZ*Ω*cat*) that are inducible in liquid or only on solid, and

ALB730 (sporulation deficient, swarming proficient)…… 45

Figure S2.1 Development of colony morphology and color if *B. subtilis* PY79 and AR123 on LB, DSM, and ISP2 media over 6 days…… 60

Figure S2.2 Zoomed (500-1000 Da) gel/stack view of 182 spectra reveals patterns of secondary metabolite production of *B. subtilis* 3610 and PY79 over time……. 61

Figure S2.3 Gel/Stack view of 182 spectra reveals patterns of secondary metabolite production of *B. subtilis* AR123 and PY79 over time……... 62

Figure 3.1 Three steps to implement molecular networking for dereplication… 72

Figure 3.2 Validation of nodes and annotation of MS/MS spectra…….. 74

Figure 3.3 Cyanobacterial network with a cosine similarity score cut off of 0.65…….. 83

Figure 3.4 Bacterial network with a cosine similarity score cut off of 0.65…… 86

Figure 3.5 Molecular networking can be easily implemented in traditional natural product workflows…….. 97

Figure S3.1 BioMAP and cytological profiles (CPs) of known antibiotics and *Dietzia* sp. FI-1026 extracts…… 111

Figure S3.2 Annotation of carmabin MS/MS spectra…….. 114

Figure S3.3 Annotation of lyngbyabellin MS/MS spectra and isotopic distribution of precursor ions…….. 115

Figure S3.4 Annotation of tumonoic acid I MS/MS spectra and EIC of precursor masses found to cluster with the seed…….. 116

Figure S3.5 Annotation of barbamide MS/MS spectra…… 119

Figure S3.6 Annotation of carmaphycin A and B MS/MS spectra…….. 120

Figure S3.7 Annotation of apratoxin MS/MS spectra…….. 121

Figure S3.8 Annotation of precursor mass *m/z* 694 MS/MS spectra and EIC…. 122

Figure S3.9 MS/MS spectra corresponding to phenazine and quinolone nodes.. 123

Figure S3.10 MS/MS spectra corresponding to abyssomicin nodes…… 124

Figure S3.11 MS/MS spectra corresponding to nodes in the 1+ surfactin cluster…….. 126

Figure S3.12 Annotation of surfactin-C_{13} $[M+H]^+$ MS/MS……. 128

Figure S3.13 MS/MS spectra corresponding to nodes in the 2+ surfactin cluster…….. 129

LIST OF TABLES

Table 2.1 *B. subtilis* strains used in this study…… 30

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to Pieter, my committee chair and advisor. I thank him for his endless enthusiasm for research and his willingness to allow me to forge my own path in science. I bestow special thanks to lab members past and present, especially to Kathleen and Tatiana for the surprise snacks in my drawer. It has been my privilege to be a part of the Dorrestein community and I am forever indebted for their support in both life and science.

I would also like to thank my committee members, Prof. Thomas Hermann, Prof. Alex Hoffmann, Prof. Kit Pogliano, and Prof. Lei Wang, for their valuable feedback during the pursuit of my degree.

I sincerely thank all of my collaborators for their contributions and patience during the preparation of multi-author publications and motivating scientific discussions.

Also, I acknowledge Betsy Komives for encouraging me to apply to the Interfaces Training Program, which provided support and the unique opportunity to experience interdisciplinary science first hand.

Finally, I wish to thank my parents Sungwon and Hearan, the rest of my immediate family, Deb, Dan, Marcela, and baby Yang, and my friends for their continual love and support. It takes a community.

Chapter 1, in full, is a reprint of the material as it appears in the Journal of Bacteriology 2012. Jane Y. Yang, Vanessa V. Phelan, Ryan Simkovsky, Jeramie D. Watrous, Rachelle M. Trial, Tinya C. Fleming, Roland Wenter, Bradley S. Moore,

Susan S. Golden, Kit Pogliano, and Pieter C. Dorrestein. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, has been submitted for publication of the material as it may appear in the Journal of Natural Products 2013. Jane Y. Yang, Laura M. Sanchez, Christopher M. Rath, Xueting Liu, Paul Bourdreau, Nicole Bruns, Evgenia Glukhov, Anne Wodtke, Rafael de Felicio, Amanda Fenner, Weng Ruh Wong, Roger G. Linington, Lixin Zhang, Hosana Maria Debonsi, William H. Gerwick, and Pieter C. Dorrestein. The dissertation author and LM Sanchez were the primary investigators and authors of this paper.

VITA

EDUCATION & RESEARCH EXPERIENCE

2001	Bachelor of Science in Biochemistry, University of California, Davis
2002-2004	Development Chemist, NovaCal Pharmaceuticals, Inc. (now NovaBay Pharma), Emeryville, California
2007	Master of Science in Chemistry (Biochemistry), San Francisco State University
2007-2008	Research Associate I, San Francisco State University
2013	Doctor of Philosophy in Chemistry with a specialization in multi-scale biology, University of California, San Diego

FUNDING & SELECTED AWARDS

2009-2013	Interfaces Graduate Training Program, University of California, San Diego
2009-2011	Graduate fellowship, Interfaces Graduate Training Program, University of California, San Diego
2009	Scholarship for Intensive High Throughput Screening Methods in Drug Development, University of California, San Diego Extension
2008	Distinguished Graduate Achievement Award, San Francisco State University
2001	Dean's Honor List, University of California, Davis
1998-1999	Undergraduate fellowship, Institute of Theoretical Dynamics – Research Training Group, University of California, Davis

PUBLICATIONS

1. **Yang, J.Y**., Sanchez, L.M., Rath, C.R., Liu, X., Boudreau, P., Bruns, N., Glukhov, E., Wodtke, A., de Felicio, R., Fenner, A., Debonsi, H.M., Gerwick, W.H., Dorrestein, P.C. Molecular networking as a dereplication strategy. (Submitted).

2. Moree, W.J., **Yang, J.Y.**, Zhao, X., Liu, W.-T., Aparicio, M., Atencio, L, Ballesteros, J., Sanchez, J., Gutierrez-Guevara, M., Dorrestein, P.C. Imaging mass spectrometry of a coral microbe interaction with fungi. (Submitted).

3. Rath, C.M., **Yang, J.Y.**, Alexandrov, T., Dorrestein, P.C. Data-independent microbial metabolomics with ambient ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* 2013. (In press).

4. **Yang, J.Y.**, Phelan, V.P., Simkovsky, R., Watrous, J.D., Trial, R.M., Fleming, T.C., Wenter, R., Moore, B.S., Golden, S.S., Pogliano, K., Dorrestein, P.C. A primer on agar-based microbial imaging mass spectrometry. *J. Bacteriol.* 2012, 194, 6023-6028.

Featured cover image

5. Hoefler, B.C., Gorzelnik, K.V., **Yang, J.Y.**, Hendricks, N., Dorrestein, P.C., Straight, P.D. Enzymatic resistance to the lipopeptide surfactin as identified through imaging mass spectrometry of bacterial competition. *Proc. Natl. Acad. Sci. USA.* 2012, 109, 13082-7.

6. 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., Moore, B.S., Laskin, J., Bandeira, N., Dorrestein, P.C. Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci. USA.* 2012, 109, E1743-52.

7. Udwary, D.W., Gontang, E.A., Jones, A.C., Schultz, A.W., Sorrels, C.M., Winter, J.M., **Yang, J.Y.**, Beauchemin, N., Capson, T.L., Clark, B.R., Esquenazi, E., Eustáquio, A.S., Freel, K., Gonzalez, D.J., Gerwick, L., Gerwick, W.H., Liu, W.-T., Malloy, K.L., Maloney, K.N., Nett, M., Nunnery, J.K., Penn, K., Prieto-Davo, A., Simmons, T.L., Weitz, S., Wilson, M.C., Tisa, L.S., Dorrestein, P.C. and Moore, B.S. Comparative genomic and proteomic analysis of the actinorhizal symbiont Frankia reveals significant natural product biosynthetic potential. *Appl. Environ. Microb.* 2011, 77, 3617-3625.

8. **Yang, J.Y.**, Karr, J.R., Watrous, J.D. and Dorrestein, P.C. Integrating '-omics' and natural product discovery platforms to investigate metabolic exchange in microbiomes. *Curr. Opin. Chem. Biol.* 2011, 15, 79-87.

9. Meier, J.L., Patel, A.D., Niessen, S., Meehan, M., Kersten, R., **Yang, J.Y.**, Rothmann, M., Cravatt, B.F., Dorrestein, P.C., Burkart, M.D. and Bafna, V. Practical 4'-Phosphopantetheine Active Site Discovery from Proteomic Samples. *J. Proteome Res.* 2011, 10, 320-9.

10. Liu, W.-T., Yang, Y.-L., Xu, Y., Lamsa, A., Haste, N.M., **Yang, J.Y.**, Ng, J., Gonzalez, D., Ellermeier, C.D., Straight, P.D., Pezner, P.A., Pogliano, J., Nizet, V., Pogliano, K. and Dorrestein. P.C. Imaging mass spectrometry of intraspecies metabolic exchange revealed the cannibalistic factors of *Bacillus subtilis*. *P. Natl. Acad. Sci. USA.* 2010, 107, 16286-90.

11. McDonald, C.A., **Yang, J.Y.**, Marathe, V., Yen, T.Y., Macher, B.A. Combining results from lectin affinity chromatography and glycocapture approaches substantially improves the coverage of the glycoproteome. *Mol Cell Proteomics.* 2009, 8, 287-301.

PRESENTATIONS

1. "MALDI imaging mass spectrometry of intraspecies metabolic exchange." University of California San Diego, Interfaces Graduate Training Program Symposium, 4th Annual Meeting. Dec 3, 2010.

2. Yang, J.Y., Smith, R., Pogliano, K., Dorrestein, P.C. “Imaging Mass Spectrometry of Bacteria, Swarming and Interactions.” American Society of Mass Spectrometry, 59th Annual Meeting, Denver, CO. June 6-9, 2011.

ABSTRACT OF THE DISSERTATION

Investigating Microbial Metabolites with Novel Mass Spectrometry Tools

by

Jane Youngmi Yang

Doctor of Philosophy in Chemistry with a specialization in multi-scale biology

University of California, San Diego, 2013

Professor Pieter C. Dorrestein, Chair

Microbes are everywhere. One teaspoon of soil contains an estimated 100 million to one billion bacteria. There are 100 million times more bacteria in the ocean than stars in the known universe. And microbes associated with the human body outnumber human cells ten to one. Microbes communicate with their environment through small molecules, also referred to as secondary metabolites. These microbial metabolites modulate cell to cell communication, which affects biological processes such as cellular differentiation within a colony, virulence, and the homeostatic balance between host health and disease.

Understanding these metabolites has implications in many applications, including agriculture and health. Previous studies examine one metabolite at a time, out of the context of the entire colony or community. And although one metabolite may exhibit bioactivity, a microbial community most likely produces multiple molecules simultaneously. In order to begin to understand microbial molecules and their global biological roles in the context of communities, we developed and applied novel mass spectrometry tools to single species colonies, dual species interactions, and phylogenetically distinct microbes.

The thesis begins with the introduction of matrix assisted laser desorption ionization-time of flight imaging mass spectrometry (MALDI-TOF IMS). This mass spectrometry based tool is used to visualize the two dimensional distribution of metabolites associated with microbial colonies

Chapter 2 presents the profiling of microbial metabolites by MALDI-TOF MS and the application of IMS to further understand the regulation and production of secondary metabolites in *Bacillus subtilis*, and the discussion of the biological implications.

Chapter 3 introduces mass spectrometry based molecular networking as a strategy to quickly identify the "known unknowns" within complex samples.

Chapter 4 proposes the application of molecular networking described in the dissertation to investigate honey bee colony collapse disorder.

CHAPTER I

A primer on agar-based microbial imaging mass spectrometry

JB Journals.ASM.org

MINIREVIEW

Primer on Agar-Based Microbial Imaging Mass Spectrometry

Jane Y. Yang,[a] Vanessa V. Phelan,[b] Ryan Simkovsky,[d] Jeramie D. Watrous,[a] Rachelle M. Trial,[e] Tinya C. Fleming,[e] Roland Wenter,[f] Bradley S. Moore,[b,f] Susan S. Golden,[d] Kit Pogliano,[e] and Pieter C. Dorrestein[a,b,c]

Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, California, USA[a]; Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, California, USA[b]; Department of Pharmacology, University of California at San Diego, La Jolla, California, USA[c]; Center for Chronobiology, Division of Biological Sciences, University of California at San Diego, La Jolla, California, USA[d]; Division of Biological Sciences, University of California at San Diego, La Jolla, California, USA[e]; and Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California, USA[f]

Matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) imaging mass spectrometry (IMS) applied directly to microbes on agar-based medium captures global information about microbial molecules, allowing for direct correlation of chemotypes to phenotypes. This tool was developed to investigate metabolic exchange factors of intraspecies, interspecies, and polymicrobial interactions. Based on our experience of the thousands of images we have generated in the laboratory, we present five steps of microbial IMS: culturing, matrix application, dehydration of the sample, data acquisition, and data analysis/interpretation. We also address the common challenges encountered during sample preparation, matrix selection and application, and sample adherence to the MALDI target plate. With the practical guidelines described herein, microbial IMS use can be extended to bio-based agricultural, biofuel, diagnostic, and therapeutic discovery applications.

Microbial metabolites are chemical messages that act as indicators of population density (4, 41, 61) and as cues for cellular differentiation (9, 34, 35, 51, 57) and that commonly have protective roles (6, 36, 44, 47). Their functional diversity is reflected in their structural diversity, ranging from small molecules and iron scavengers to peptides and lipids. Traditionally, individual microbial metabolites have been targeted using bioactivity-guided fractionation. More recent "omics" technologies, such as metabolomics, begin to capture molecules on a global scale, but current mass spectrometry-based approaches do not distinguish between molecules which are always present and molecules with changing spatial distributions, disregarding the spatial localization of the molecules within a phenotype and the multifaceted chemical exchange within and between microbial cell populations (18, 19, 60).

The development of imaging mass spectrometry (IMS) techniques complements metabolomic approaches by enabling the preservation of molecular localization, yielding insight into the underlying biology. Matrix-assisted laser desorption–ionization time of flight (MALDI-TOF) imaging mass spectrometry (IMS) technology has been applied to molecular pathology for the past 17 years (11, 50, 57) to visualize fundamental disease signatures from biological samples by detecting proteins, peptides (14, 58), and lipids (43) directly from tissue sections. Most MALDI mass spectrometers are capable of imaging experiments with vendor-specific software and freeware to automate data acquisition and assist in analysis (20, 28), and other IMS methods are rapidly emerging (56, 64, 71), thus reaffirming the need for these technologies.

In 2009, we applied MALDI-TOF IMS to study various molecules in the context of interacting microbial colonies (70), hereafter referred to as microbial IMS. Theoretically, any culturable microbe can be subjected to microbial IMS. To date, our lab has been able to capture chemical information from a wide range of microbes, including terrestrial (31, 34), freshwater, and marine bacteria (69) as well as amoebae, fungi, and yeast. By providing a spatial snapshot for each observed mass signal, it is possible to directly visualize the metabolic exchange within and among microbial species on various types of agar media, as shown in Fig. 1.

FIVE STEPS OF MICROBIAL IMS

Culturing— use thin agar. Microbial IMS sample preparation workflow (Fig. 2) begins with culturing one or more microbes on agar-based medium directly on top of the MALDI target plate (70) or in petri dishes (18, 34). The agar is 1 to 1.5 mm thick, equivalent to 10 to 11 ml medium in a standard 10-cm petri dish, and typically contains 1 to 2% agar. The petri dishes are either sealed with Parafilm or placed in a humidity-controlled chamber to minimize dehydration of the thin agar during incubation. Although various media can be used, a medium that we frequently use is yeast extract/malt extract ISP2 (4 g/liter yeast extract, 10 g/liter malt extract, 4 g/liter dextrose, and 1.5 to 2% agar).

The three approaches used to prepare the microbial samples for IMS are (i) pouring agar directly onto the MALDI target plate, (ii) embedding the target plate in agar medium prior to inoculation, and (iii) cutting the microbial colony and surrounding agar directly from the petri dish and transferring it to the target plate (see Video S1 in the supplemental material). Approaches one and two require vigilant sterility, are generally used to study one colony or one interaction per target plate, and occupy the target plate for days or weeks at a time. We currently use approach three, as this approach does not occupy the MALDI plates for long periods of time, increasing sample throughput and allowing numerous samples to be placed onto a single MALDI target for analysis. It is important to photograph the microbes before the matrix is applied in order to correlate the phenotype to the observed chemistry.

Published ahead of print 20 July 2012
Address correspondence to Pieter C. Dorrestein, pdorrestein@ucsd.edu.
Supplemental material for this article may be found at http://jb.asm.org/.

doi:10.1128/JB.00823-12

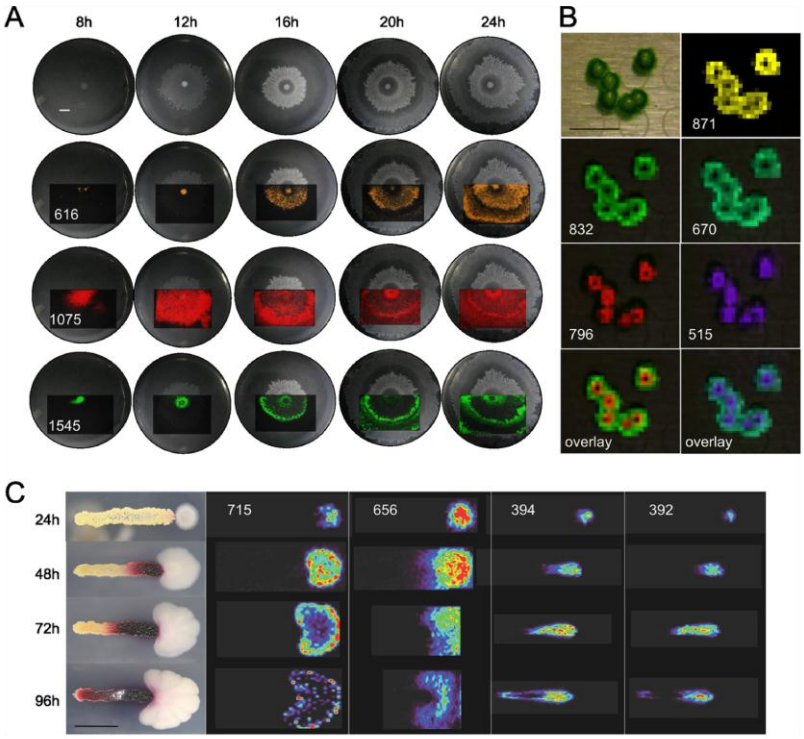


FIG 1 Molecular snapshots from microbial MALDI-TOF imaging mass spectrometry, obtained on different agar surfaces. (A) Representative ions of swarming *Bacillus subtilis* 3610 (29, 33, 34) on ISP2 (0.7% agar) medium at 30°C over time. *m/z* 616, 1,075, and 1,545 were assigned to an uncharacterized ion, surfactin-C15 [M2 K]2 , and plipastatin-C17-Val [M2 K]2 , respectively (70). Due to the size of the MALDI target plate, only the displayed regions were imaged. (B) We observed the development of previously unreported macrostructures in the growth of freshwater *Synechococcus elongatus* PCC 7942 colonies, a model photosynthetic organism for the study of prokaryotic circadian biology (37). A confluent culture was diluted 1:50,000 with BG-11 M and then spotted and grown on BG-11 M (1.5% agar) medium (10) at 30°C inside a moisture chamber under a constant light intensity of 350 E/m^2 · s for 16 days prior to IMS. The [M2 H]2 of pheophytin *a*, chlorophyll *a* without Mg^{22}, is at *m/z* 871, while *m/z* 796 and 515 putatively correspond to the observed macrostructures. (C) Interaction between *Bacillus subtilis* PY79 and *Streptomyces coelicolor* A3 (2) on ISP2 (1.5% agar) medium at 30°C over time. The color gradient is from least-intense ions (purple) to most-intense ions (red). The prodiginines streptorubin B and undecylprodigiosin [M2 H]2 are at *m/z* 392 and 394, respectively. *m/z* 656 is an uncharacterized bacillus ion, and 715 is a partially characterized polyglutamate compound (70). Scale bars, 1 cm.

Matrix application— saturate the sample. MALDI-TOF mass spectrometry requires matrix, small organic acids that cocrystallize with the sample and absorb the laser energy to assist in desorption and ionization of molecules from the sample surface (21). The spatial resolution of imaging mass spectrometry is dependent on two main factors, the diameter of the laser beam and the size of the matrix crystals. We have found that using either a 20- or 53- m stainless steel test sieve (Hogentogler & Co., Inc.) (43, 70) to apply matrix is an efficient, low-cost, and robust method to uniformly dry coat and saturate microbial samples (see Video S2 in the supplemental material) with no discernible differences in the quality of IMS data at a spatial resolution of 100 m by 100 m or greater. Microbial IMS is currently unable to resolve single cells.

A 1:1 mixture of 2,5-dihydroxybenzoic acid (DHB) and cyano-4-hydroxycinnamic acid (CHCA) matrices, sold as Universal MALDI Matrix (Sigma-Aldrich), is used for both positive- and negative-mode microbial IMS. We hypothesize that saturation of the sample with this matrix mixture is crucial for sample adherence and for yielding matrix crystals of sufficient quality to ionize molecules from the sample (see Fig. S2 and the supplemental discussion in the supplemental material). Although CHCA is commonly used for small proteins and peptides and DHB is typically used for carbohydrates and glycopeptides, the mixture has been reliable for the detection of peptides, carbohydrates, lipids, and other metabolic exchange factors from microbial samples.

Dehydration— check for flaking, sample height of about 1 mm. A critical step in microbial IMS is dehydration of the sample at 37°C, which may be followed by storage in a vacuum desiccator. This step ensures that the MALDI mass spectrometer components (the source, mass analyzer, and detector) achieve the vacuum pressure necessary for operation and that the thickness of the dehydrated microbial sample is compatible with the allowed sample height of 0.4 to 1.0 mm. In general, dehydrated microbial samples account for most of the sample height; our desiccated agar is about 100 to 200 m thick, and the contribution of matrix to sample height is at most double the thickness for many of the organisms we have tested.

Variable sample height impacts the mass calibration of a TOF

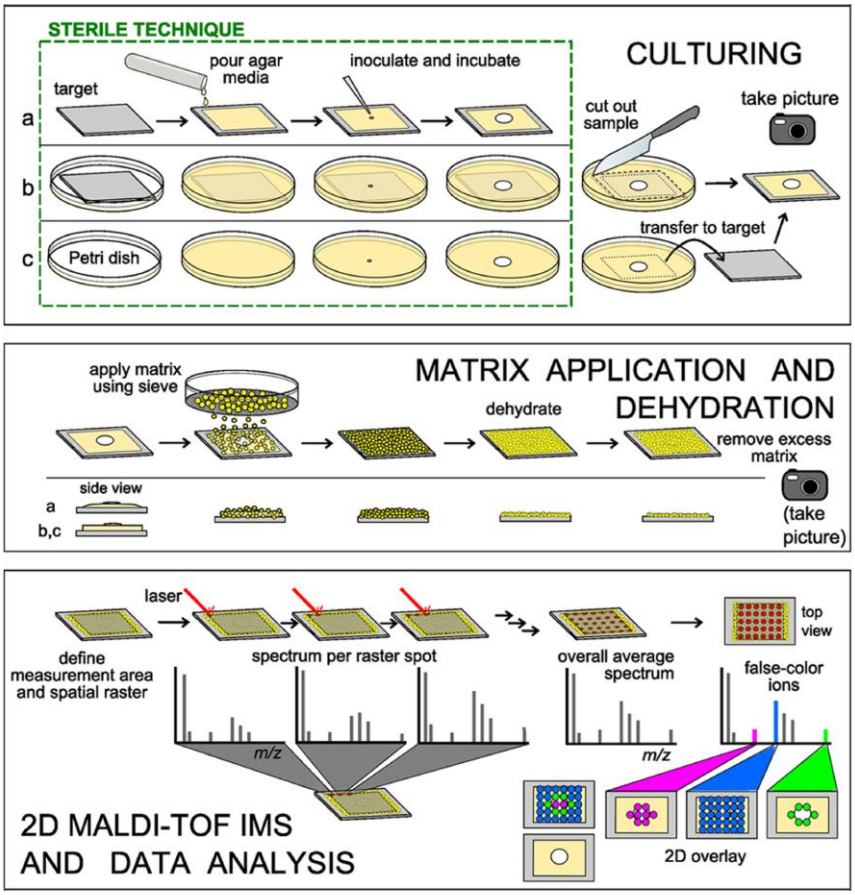


FIG 2 The 5 major steps to obtain a microbial chemotype image are culturing, matrix application, dehydration of the samples, imaging mass spectrometry, and data analysis. Culturing. Three approaches to grow microbial samples for IMS are (a) pouring agar and inoculating directly on the target, (b) embedding the target in agar prior to inoculation and incubation, and (c) culturing the microbe in a petri dish, excising the region of interest, and then transferring it to the target. Matrix application. The sample edges from culturing approach a are tapered toward the target plate, compared to approaches b and c. This will affect the matrix coverage at the edges of the sample; matrix application is depicted for samples prepared via approaches b and c. Universal matrix is applied to the sample as a dry powder using a 20- or 53- m test sieve. The hydration of the agar is sufficient for surface absorption of matrix. Sample dehydration. The sample is dehydrated at an elevated temperature (37°C), and the excess matrix is removed prior to microbial IMS. The sample height should be less than 1.0 mm. Also, the sample is examined for sufficient adherence to the target. Imaging mass spectrometry. User-defined spatial resolution and measurement area generate the *x-y* raster grid for the imaging run. A MALDI-TOF mass spectrum is collected per raster spot; spectra from the entire measurement area are averaged. Data analysis. False colors are assigned to ions in the overall average mass spectrum and displayed over a photograph of the sample. The interpretation of the ion distributions is the responsibility of the experimenter.

instrument. We typically observe 0.1- to 0.5-Da mass shifts, in agreement with the 100- to 200- m thickness of the sample. Mass measurement in a MALDI-TOF mass spectrometer relies on accurate timing of an analyte molecule traveling from the target plate through the flight tube to the detector, with longer travel paths, such as reflectron mode and longer flight tubes, resulting in larger mass deviances. When high mass accuracy is critical, an internal calibration is possible by spotting microgram amounts of calibrant on the dried agar surface.

During the drying process, sample flaking may occur. The term flaking is used to describe air bubbles, fissures, and partial or complete detachment of the dehydrated sample from the MALDI target plate (see Fig. S1 in the supplemental material), and ways to minimize flaking are discussed later and in the supplemental discussion in the supplemental material. When sample adherence is questionable, the sample should not be inserted into the mass spectrometer.

2D MALDI-TOF IMS— data acquisition. The fundamentals of data acquisition for microbial IMS are identical to those for tissue IMS (11). The user defines the desired spatial resolution and measurement area using a reference photograph of the bacterial sample on the target, determining the two-dimensional (2D) raster grid. Mass spectra generated at each raster spot are associated with coordinates.

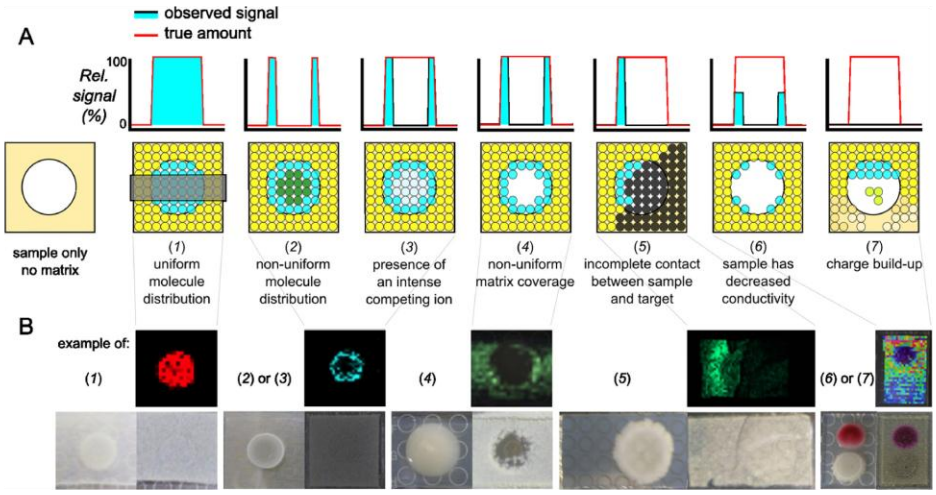


FIG 3 Data interpretation challenges—flaking, black holes, and gradients. (A) Various data interpretation challenges. (1) Uniform molecule distribution and (2) nonuniform molecule distribution. The observed signal reflects the true amount of molecule in the sample. (3) Presence of an intense competing ion. With even matrix coverage, the observed signal does not reflect the true amount of molecule in the sample due to signal suppression by a competing ion. (4) Nonuniform matrix coverage. This is common with microbes that produce lipopolysaccharide (LPS) or surfactant, sporulate, or form aerial structures that physically inhibit matrix saturation. (5) Incomplete contact between sample and target. Incomplete contact between the sample and the target, such as a bubble, a crack, or flaking, ultimately results in decreased detection of ions. The sample has (6) decreased conductivity and (7) charge buildup. The thickness of the microbial sample may act as an insulator, decreasing the conductivity and, in turn, ionization of molecules. Similarly, charge can build up on the MALDI target over time, manifesting in a gradient of higher intensity in the beginning of acquisition and little to no intensity at the end of acquisition. Charge buildup is more common in tissue IMS, which uses indium tin oxide (ITO)-coated slides. (B) Real examples of the challenges described in panel A. From left to right, *Escherichia coli*, *Staphylococcus aureus*, *Pantoea agglomerans* (Lab Environmental Strain Collection), *B. subtilis*, and *Serratia marcescens* (Lab Environmental Strain Collection) interacting with *B. subtilis* 3610. Select ions from IMS data are displayed over photographs depicting matrix coverage of the dehydrated sample. Photographs of the sample on the target prior to matrix coverage and dehydration are to the left of the matrix coverage photographs. The colors in the gradient example are from most-intense (red) to least-intense (purple) ions.

Data analysis and interpretation. After MALDI-TOF spectra are acquired, the signal intensity of a discrete mass across the sample surface can be visualized using a false color superimposed onto a photograph of the microbial sample. These IMS images are composed of pixels, similar to images on a computer or TV screen, where each pixel contains molecular information. By correlating spatial distributions of each molecule with observed phenotypes, putative functions of observed molecules can be assigned and phenotypes undetectable by eye may be revealed.

CHALLENGES IN MICROBIAL IMS

Sample adherence. The primary limitation of microbial IMS is flaking. Flaking occurs (see Fig. S1 in the supplemental material) as a result of the inability of the dried and brittle agar to remain adhered to the target plate, a result which can be due to trapped air underneath the growth medium, insufficient matrix saturation of the sample, use of smooth or polished MALDI target plates, certain medium components (e.g., high salt), and microbial molecules (1, 5, 17, 18, 22, 26, 27, 34, 38, 39, 59, 70). In some cases, sample flaking does not interfere with instrument operation (Fig. 3B, number 5); however, care must be taken to minimize flaking inside the mass spectrometer. While sample adherence is a major obstacle in the microbial IMS workflow, we have found that optimization of the sample preparation can overcome such issues. In Fig. S1 and the supplemental discussion in the supplemental material, we present guidelines and a flowchart with checkpoints to obtain quality microbial IMS images and decrease the risk of sample flaking.

Data interpretation. The challenges in data interpretation for microbial IMS are common to all MALDI-TOF-based IMS. Black holes or areas of no signal may be due to nonuniform matrix coverage or ion suppression, whereas false gradients or signals of regular increasing or decreasing intensities may be due to decreased conductivity, charge buildup or charging (25, 66), or a combination of factors (Fig. 3). Charge buildup is recognizable by an optimal signal early in the IMS run and a decreased signal over time (Fig. 3B). Analysis of multiple microbial samples in a randomized fashion will provide insight into the significance of the gradient signals. Data interpretation is the responsibility of the experimenter who evaluates the causes of patterns and whether they are real or artifacts.

Determining the origin of observed molecules. IMS of dual-species interactions yields many ions, some of which are detected only upon interaction. Oftentimes, the distribution of the ion of interest is interfacial and the producer is ambiguous. By employing analytical and biological approaches, such as microbial IMS, IMS of multiple time points (67), traditional solvent extraction, purification, and structural determination methods, such as tandem MS and nuclear magnetic resonance (NMR) (18, 30, 33, 67, 68), genetics and microbiology, 16S rRNA sequencing (54), established MALDI-TOF protocols (12, 49, 53), genome mining approaches (7, 13, 15, 45, 63) and predictive programs (2, 3, 16, 33,

40, 48, 55, 62, 65, 68), peptidogenomics (31), and literature and database searches (23, 52) (AntiMarin database, Dictionary of Natural Products, the National Institute of Standards and Technology [NIST] databases, and the SciFinder database), one can typically annotate microbial IMS data. Two main strategies to confirm the producing microbe and the resultant phenotype are genetic knockout and complementation studies or assays with purified compound, in combination with more IMS.

FINAL REMARKS

The microbial metabolic maps from 2D MALDI-TOF IMS serve as a starting point for the identification of microbial metabolites for virtually any culturable microbe. A single metabolite of low abundance can now be detected using IMS analysis due to high, localized concentrations within the preserved 2D phenotype, but annotating the molecules is still a challenge. The microbial interactions can be scaled up and extracted to isolate enough material for further mass spectrometry-guided isolation and subsequent structural characterization with mass spectrometry and NMR (69). However, this process is inefficient and costly. The challenge of annotating mass spectrometry data also applies to proteomics, metabolomics, and lipidomics. These advanced "omics" tools, even when combined with the power of genome mining, peptidogenomics, and search algorithms, annotate only a small percentage of what are presumably some of the most abundant ions, as observed in an imaged sample (32), indicating that there are still many opportunities in mass spectrometry to develop novel approaches to identify (ID) molecules (8, 24, 42, 46). We encourage the scientific community to develop systematic, integrated workflows for handling unknowns in IMS data that may capture more than 50% of the IMS signals. Finally, we hope that this article provides a starting point for researchers and that it encourages other labs to apply microbial IMS to different systems and generate tools to complement this approach to push the boundaries of spatial systems microbiology.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of the U.S. National Institutes of Health grants AI095125 and S10RR029121, the Interfaces Training Grant—a Ruth L. Kirschstein National Research Service Award, NIH 1 T32 EB009380-01, grant GM097509, and the DFG Research Fellowship (WE 4754/1-1). The freshwater cyanobacteria research was supported by DOE grant DE-EE0003373.

We acknowledge the Bruker Therapeutic Discovery Mass Spectrometry Center at the UCSD Skaggs School of Pharmacy.

REFERENCES

1. **Alaupovic P, Olson AC, Tsang J.** 1966. Studies on the characterization of lipopolysaccharides from two strains of *Serratia marcescens.* Ann. N. Y. Acad. Sci. **133**:546–565.
2. **Anand S, et al.** 2010. SBSPKS: structure based sequence analysis of polyketide synthases. Nucleic Acids Res. **38**:W487–W496.
3. **Ansari MZ, Yadav G, Gokhale RS, Mohanty D.** 2004. NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasythases. Nucleic Acids Res. **32**:W405–W413.
4. **Antunes LC, Ferreira RB, Buckner MC, Finlay BB.** 2010. Quorum sensing in bacterial virulence. Microbiology **156**:2271–2282.
5. **Arima K, Kakinuma A, Tamura G.** 1968. Surfactin, a crystalline peptidelipid surfactant produced by *Bacillus subtilis*: isolation, characterization and its inhibition of fibrin clot formation. Biochem. Biophys. Res. Commun. **31**:488–494.
6. **Bassler BL, Losick R.** 2006. Bacterially speaking. Cell **125**:237–246.
7. **Begley M, Cotter PD, Hill C, Ross RP.** 2009. Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins. Appl. Environ. Microbiol. **75**:5451–5460.
8. **Bowen BP, Northen T.** 2010. Dealing with the unknown: metabolomics and metabolite atlases. J. Am. Soc. Mass Spectrom. **21**:1471–1476.
9. **Branda SS, Gonzalez-Pastor JE, Ben-Yehuda S, Losick R, Kolter R.** 2001. Fruiting body formation by *Bacillus subtilis.* Proc. Natl. Acad. Sci. U. S. A. **98**:11621–11626.
10. **Bustos SA, Golden SS.** 1991. Expression of the *psbDII* gene in *Synechococcus* sp. strain PCC 7942 requires sequences downstream of the transcription start site. J. Bacteriol. **173**:7525–7533.
11. **Caprioli RM, Farmer TB, Gile J.** 1997. Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. Anal. Chem. **69**:4751–4760.
12. **Carbonnelle E, et al.** 2012. Robustness of two MALDI-TOF mass spectrometry systems for bacterial identification. J. Microbiol. Methods **89**: 133–136.
13. **Challis GL.** 2008. Genome mining for novel natural product discovery. J. Med. Chem. **51**:2618–2628.
14. **Chaurand P, Stoeckli M, Caprioli RM.** 1999. Direct profiling of proteins in biological tissue sections by MALDI mass spectrometry. Anal. Chem. **71**:5263–5270.
15. **Claesen J, Bibb M.** 2010. Genome mining and genetic analysis of cypemycin biosynthesis reveal an unusual class of posttranslationally modified peptides. Proc. Natl. Acad. Sci. U. S. A. **107**:16297–16302.
16. **de Jong A, van Heel AJ, Kok J, Kuipers OP.** 2010. BAGEL2: mining for bacteriocins in genomic data. Nucleic Acids Res. **38**:W647–W651.
17. **Edwards JR, Hayashi JA.** 1965. Structure of a rhamnolipid from *Pseudomonas aeruginosa.* Arch. Biochem. Biophys. **111**:415–421.
18. **Gonzalez DJ, et al.** 2011. Microbial competition between *Bacillus subtilis* and *Staphylococcus aureus* monitored by imaging mass spectrometry. Microbiology **157**:2485–2492.
19. **Goulitquer S, Potin P, Tonon T.** 2012. Mass spectrometry-based metabolomics to elucidate functions in marine organisms and ecosystems. Mar. Drugs **10**:849–880.
20. **Gustafsson JO, Oehler MK, Ruszkiewicz A, McColl SR, Hoffmann P.** 2011. MALDI imaging mass spectrometry (MALDI-IMS)—application of spatial proteomics for ovarian cancer classification and diagnosis. Int. J. Mol. Sci. **12**:773–794.
21. **Hankin JA, Barkley RM, Murphy RC.** 2007. Sublimation as a method of matrix application for mass spectrometric imaging. J. Am. Soc. Mass Spectrom. **18**:1646–1652.
22. **Hisatsuka K, Nakahara T, Minoda T, Yamada K.** 1971. Formation of rhamnolipid by *Pseudomonas aeruginosa* and its function in hydrocarbon fermentation. Agric. Biol. Chem. **35**:686–692.
23. **Horai H, et al.** 2010. MassBank: a public repository for sharing mass spectral data for life sciences. J. Mass. Spectrom. **45**:703–714.
24. **Hufsky F, Rempt M, Rasche F, Pohnert G, Böcker S.** 2012. De novo analysis of electron impact mass spectra using fragmentation trees. Anal. Chim. Acta **739**:67–76.
25. **Ibanez AJ, Muck A, Svatos A.** 2007. Dissipation of charge on MALDI-TOF polymeric chips using an electron-acceptor: analysis of proteins. J. Mass Spectrom. **42**:634–640.
26. **Inagawa H, et al.** 1992. Homeostasis as regulated by activated macrophage. II. LPS of plant origin other than wheat flour and their concomitant bacteria. Chem. Pharm. Bull. (Tokyo) **40**:994–997.
27. **Itoh S, Honda H, Tomita F, Suzuki T.** 1971. Rhamnolipid produced by *Pseudomonas aeruginosa* grown on n-paraffin. J. Antibiot. **24**:855–859.
28. **Jardin-Mathe O, et al.** 2008. MITICS (MALDI Imaging Team Imaging Computing System): a new open source mass spectrometry imaging software. J. Proteomics **71**:332–345.
29. **Jeffries CD, Rogers HE.** 1968. Enhancing effect of agar on swarming by *Proteus.* J. Bacteriol. **95**:732–733.
30. **Kearns DB.** 2010. A field guide to bacterial swarming motility. Nat. Rev. Microbiol. **8**:634–644.
31. **Kersten RD, et al.** 2011. A mass spectrometry-guided genome mining approach for natural product peptidogenomics. Nat. Chem. Biol. **7**:794–802.
32. **Kim S, Bandiera N, Pevzner PA.** 2009. Spectral profiles, a novel representation of tandem mass spectra and their applications for de novo peptide sequencing and identification. Mol. Cell. Proteomics **8**:1391–1400.
33. **Li MH, Ung PM, Zajkowski J, Garneau-Tsodikova S, Sherman DH.**

2009. Automated genome mining for natural products. BMC Bioinformatics **10**:185.
34. **Liu W-T, et al.** 2010. Imaging mass spectrometry of intraspecies metabolic exchange revealed the cannibalistic factors of *Bacillus subtilis*. Proc. Natl. Acad. Sci. U. S. A. **107**:16286–16290.
35. **Lopez D, Kolter R.** 2010. Extracellular signals that define distinct and coexisting cell fates in *Bacillus subtilis*. FEMS Microbiol. Rev. **34**:134–149.
36. **Lowery CA, Dickerson TJ, Janda KD.** 2008. Interspecies and interkingdom communication mediated by bacterial quorum sensing. Chem. Soc. Rev. **37**:1337–1346.
37. **Mackey SR, Golden SS, Ditty JL.** 2011. The itty-bitty time machine genetics of the cyanobacterial circadian clock. Adv. Genet. **74**:13–53.
38. **Matsuyama T, Sogawa M, Nakagawa Y.** 1989. Fractal spreading growth of *Serratia marcescens* which produces surface active exolipids. FEMS Microbiol. Lett. **52**:243–246.
39. **Matsuyama T, et al.** 1992. A novel extracellular cyclic lipopeptide which promotes flagellum-dependent and -independent spreading growth of *Serratia marcescens*. J. Bacteriol. **174**:1769–1776.
40. **Medema MH, et al.** 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. **39**:W339–W346.
41. **Miller MB, Bassler BL.** 2001. Quorum sensing in bacteria. Annu. Rev. Microbiol. **55**:165–199.
42. **Padliya ND, Wood TD.** 2004. A strategy to improve peptide mass fingerprinting matches through the optimization of matrix-assisted laser desorption/ionization matrix selection and formulation. Proteomics **4**:466–473.
43. **Puolitaival SM, Burnum KE, Cornett DS, Caprioli RM.** 2008. Solvent-free matrix dry-coating for MALDI imaging of phospholipids. J. Am. Soc. Mass Spectrom. **19**:882–886.
44. **Ratcliff WC, Denison RF.** 2011. Alternative actions for antibiotics. Science **332**:547–548.
45. **Rausch C, Weber T, Kohlbacher O, Wohlleben W, Huson DH.** 2005. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). Nucleic Acids Res. **33**:5799–5808.
46. **Rojas-Cherto M, et al.** 2012. Metabolite identification using automated comparison of high-resolution multistage mass spectral trees. Anal. Chem. **84**:5524–5534.
47. **Romero D, Traxler MF, Lopez D, Kolter R.** 2011. Antibiotics as signal molecules. Chem. Rev. **111**:5492–5505.
48. **Röttig M, et al.** 2011. NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res. **39**:W362–W367.
49. **Saffert RT, et al.** 2011. Comparison of Bruker Biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometer to BD Phoenix automated microbiology system for identification of gram-negative bacilli. J. Clin. Microbiol. **49**:887–892.
50. **Seeley EH, Caprioli RM.** 2011. MALDI imaging mass spectrometry of human tissue: method challenges and clinical perspectives. Trends Biotechnol. **29**:136–143.
51. **Shank EA, Kolter R.** 2011. Extracellular signaling and multicellularity in *Bacillus subtilis*. Curr. Opin. Microbiol. **14**:741–747.
52. **Smith CA, et al.** 2005. METLIN: a metabolite mass spectral database. Ther. Drug Monit. **27**:747–751.
53. **Sogawa K, et al.** 2011. Use of the MALDI BioTyper system with MALDI-TOF mass spectrometry for rapid identification of microorganisms. Anal. Bioanal. Chem. **400**:1905–1911.
54. **Stackebrandt E, Goebel BM.** 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Evol. Microbiol. **44**:846–849.
55. **Starcevic A, et al.** 2008. ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. Nucleic Acids Res. **36**:6882–6892.
56. **Steinhauser ML, et al.** 2012. Multi-isotope imaging mass spectrometry quantifies stem cell division and metabolism. Nature **481**:516–519.
57. **Stoeckli M, Farmer TB, Caprioli RM.** 1999. Automated mass spectrometry imaging with a matrix-assisted laser desorption ionization time-of-flight instrument. J. Am. Soc. Mass Spectrom. **10**:67–71.
58. **Stoeckli M, Chaurand P, Hallahan DE, Caprioli RM.** 2001. Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. Nat. Med. **7**:493–496.
59. **Straight PD, Willey JM, Kolter R.** 2006. Interactions between *Streptomyces coelicolor* and *Bacillus subtilis*: role of surfactants in raising aerial structures. J. Bacteriol. **188**:4918–4925.
60. **Straight PD, Kolter R.** 2009. Interspecies chemical communication in bacterial development. Annu. Rev. Microbiol. **63**:99–118.
61. **Swift S, et al.** 2001. Quorum sensing as a population-density-dependent determinant of bacterial physiology. Adv. Microb. Physiol. **45**:199–270.
62. **Tae H, Kong EB, Park K.** 2007. ASMPKS: an analysis system for modular polyketide synthases. BMC Bioinformatics **8**:327.
63. **Velasquez JE, van der Donk WA.** 2011. Genome mining for ribosomally synthesized natural products. Curr. Opin. Chem. Biol. **15**:11–21.
64. **Watrous JD, Dorrestein PC.** 2011. Imaging mass spectrometry in microbiology. Nat. Rev. Microbiol. **9**:683–694.
65. **Weber T, et al.** 2009. CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. J. Biotechnol. **140**:13–17.
66. **Werner HW, Morgan AE.** 1976. Charging of insulators by ion bombardment and its minimization for secondary ion mass spectrometry (SIMS) measurement. J. Appl. Phys. **47**:1232–1242.
67. **Xu Y, et al.** 2012. Bacterial biosynthesis and maturation of the didemnin anti-cancer agents. J. Am. Chem. Soc. **134**:8625–8632.
68. **Yadav G, Gokhale RS, Mohanty D.** 2003. SEARCHPKS: a program for detection and analysis of polyketide synthase domains. Nucleic Acid Res. **31**:3654–3658.
69. **Yang Y-L, et al.** 2011. Connecting chemotypes and phenotypes of cultured marine microbial assemblages by imaging mass spectrometry. Angew. Chem. Int. Ed. Engl. **50**:5839–5842.
70. **Yang Y-L, Xu Y, Straight PD, Dorrestein PC.** 2009. Translating metabolic exchange with imaging mass spectrometry. Nat. Chem. Biol. **5**:885–887.
71. **Zhang D-S, et al.** 2012. Multi-isotope imaging mass spectrometry reveals slow protein turnover in hair-cell stereocilia. Nature **481**:520–524.

A primer on agar-based microbial imaging mass spectrometry

Supplemental discussion

How to minimize flaking – full contact with non-polished target. From sample preparation through data acquisition, there are several ways to optimize adherence of the sample to the target plate and minimize flaking. A major portion of the sample preparation effort is geared toward ensuring that the dried agar is securely adhered to the MALDI target plate. We have provided a flowchart with checkpoints (Fig. S1) to maximize the ability to get consistent, reliable images and to minimize instrument down time.

During microbial IMS sample preparation, a non-polished MALDI target plate must be used. The rough surface of a non-polished target plate increases the amount of surface area to which the growth media can adhere. Polished target plates can be made "non-polished" by sanding or applying steel wool to scratch the surface, but we have found that ground steel target plates work best. An example of flaking on a polished MALDI target plate is shown in Figure S1.

The sample should remain in full contact with the target plate from pre-matrix application through data acquisition. If there are air bubbles prior to matrix application, we recommend using a spatula to gently lift up the corner of the agar sample until the bubble is reached and laying the agar back down again (Supplemental, Video 3).

Culturing – use thin agar, adjust media contents. The goal of media preparations is to minimize flaking while maintaining the desired phenotype. In most cases, the adjustment to 1-1.5 mm agar thickness in the Petri dish has had no impact on the phenotype of the microbes or their interactions. When this alteration of growth

conditions does have an undesired phenotypic impact, the amount of nutrients needs to be adjusted to achieve the equivalent phenotype. As a general rule, we use inverse linear scaling of nutrient concentration with respect to the media volume. For example, when decreasing media volume by half, the amount of nutrients is typically doubled. The optimization for the correct phenotypic response is necessary prior to IMS analysis. Also, determination of the IMS compatibility of media is necessary to ensure that background signals from media do not interfere with ion detection.

Slightly increasing the water content of the agar media is effective in minimizing flaking. The result of increasing the water content of the agar is that it becomes more gel-like, which in turn increases surface area contact to the MALDI target plate. Water content is affected by the age of the agar, incubation conditions, and the percentage of agar in the media. Be aware, the phenotype and metabolic profile can change with changes in agar percentage. For instance, swarming motility generally occurs between 0.3% and 1% agar, and is inhibited over 1% agar concentrations (Fig. 1A; 17, 18). Other factors that may affect sample adherence in IMS affect the agar media properties including the development of vegetative hyphae and molecules produced by the microbe(s).

Microbes that change agar properties or produce copious amounts of biosurfactants and/or lipopolysaccharides (LPS), such as *Bacillus subtilis* 3610 (2, 12, 22, 30, 33), *Pseudomonas aeruginosa* PAO1 and PA14 (11, 14, 16), *Pantoea agglomerans* (15), and *Serratia marcescens* (1, 23, 24), can also increase flaking. We hypothesize that the surfactants affect the interactions between the matrix, agar media, and target.

Matrix saturation hypothesis. Saturating the microbial sample with Universal MALDI matrix (Sigma-Aldrich), a 1:1 mixture of 2,5-dihydroxybenzoic acid (DHB) and α-cyano-4-hydroxycinnamic acid (CHCA) matrices, minimizes flaking and facilitates ionization of microbial molecules in microbial MALDI-TOF IMS. When each component of this mixture is applied individually, it is much more difficult to obtain quality images. DHB was absorbed into the agar resulting in samples adhering to the MALDI target with poor ionization during analysis, while CHCA remained on the sample surface resulting in ionization but with a much higher incidence of flaking, sample detachment from the target plate. Therefore, we hypothesize that DHB is acting as a weak glue to adhere the sample to the MALDI plate while the CHCA allows for better ionization of surface analyte (Fig. S2). Agar that is not saturated with matrix shrinks significantly during the drying process, supporting our hypothesis that a component of the matrix mixture anchors the sample to the target plate. The majority of the DHB in the mixture saturates the agar media, prevents shrinkage, and glues the sample to the target, whereas the majority of CHCA crystallizes on the top surface of the sample to assist in desorption and ionization of molecules. It should be possible to optimize the matrices as is done for other MALDI-TOF applications (9, 10, 13, 25) to ionize specific molecules of interest or maximize the number of microbial molecules detected while simultaneously maintaining sample adherence to the target plate.

Achieving even matrix application on mucoid colonies (Fig. 3B, *4*), samples with biosurfactants, or certain media that readily absorb matrix (Fig. S1; Fig. S3) such as YEPD (21) and Spider (15), can be challenging. This may be remedied by first coating

the target plate with matrix prior to transfer of the sample to the target or by multiple applications during the dehydration process (Fig. S3). Other media, such as MSgg (6), appear to absorb and dissolve matrix upon dehydration (Fig. S1, *2a-b*), which means that the agar is not saturated with matrix and shrinkage is apparent.

Dehydration – check for flaking, sample height of about 1 mm. During the drying process, bubbles can form or regions can detach from the target. Visual inspection of the sample for air bubbles, cracks, and edges or regions that did not fully contact the target plate can be easily recognized by being a lighter color than the surrounding sample. The areas of opaqueness, peeling, and loose edges of dehydrated samples are the regions most prone to flaking and should detach when blowing away excess matrix (Fig. S1, *4-6, 9*; Supplemental, Video 2).

2D MALDI-TOF IMS data acquisition – monitor live feed and vacuum pressure. Once the MALDI target plate containing the sample is inserted into the instrument, sample flaking will cause the live video feed to show a partially adhered or missing sample or the vacuum pressure will spike and data acquisition will pause. If flaking occurs inside the mass spectrometer, the user should abort the run, eject the target, and determine the number and size of the flakes. If the flakes are small in size (less than 1 mm x 1 mm) and few in number, the user may proceed with instrument usage at the risk of possibly damaging the machine. To minimize damage to the instrument after flaking, open up the mass spectrometer, recover the flake, and clean the instrument. As long as the sample adheres to the target plate, data acquisition is straightforward.

When flaking occurs. When a flake inside the mass spectrometer contacts the

voltage plate responsible for the acceleration of the ions into the flight tube in a time-of-flight instrument, it dampens the charge differential responsible for ion acceleration resulting in fewer ions detected. Damage can occur to the instrument if the flake causes an electrical short circuit between the target plate and the voltage plate. In general, if the sample or a portion thereof detaches from the target and remains in the instrument after the target is ejected, the instrument will need to be cleaned resulting in undesired downtime. It is therefore important to minimize flaking.

Data interpretation – black holes. A cause of black holes is uneven matrix coverage, as with mucoid, gram-negative bacteria such as *P. agglomerans* and *S. marcescens*. The center of these microbial colonies absorbs more matrix than the colony edges or surrounding agar, making it difficult to saturate the center with matrix. The uneven matrix coverage becomes more apparent post-dehydration and results in little to no signal in the center (Fig. 3B). Black holes are also observed when analyzing colonies that produce aerial hyphae or are excessively hard and dry and therefore cannot absorb matrix and physically mask the molecules underneath. One solution we have employed with filamentous fungi is to remove these structures with a swab doused in acetonitrile prior to matrix application. As previously stated, loss of signal is observed when a sample loses contact with the MALDI target and therefore a black hole could also be caused by an air bubble or lifted sample.

Data interpretation – false gradients. The cause of false gradients may be due to charging or ion suppression. The matrix-covered colony or agar may insulate the charge difference between the target plate, which serves as the ion source, and the acceleration voltage. This results in sub-par ionization where samples are thicker, often

where the colony lies. Charging of the sample also results in decreased ionization, but occurs over time as the IMS run progresses. The target plate retains charge in between spectrum acquisitions. Charging is more common with faster lasers (kHz) and indium tin oxide (ITO) coated glass slides used for tissue IMS compared to slower lasers (20 Hz) and stainless steel target plates used for microbial IMS. By decreasing the laser frequency from 1 kHz to 200 Hz or slower, enough time between raster spots is provided, allowing the electrical charge on the target plate to dissipate. These false gradients are usually irreproducible between independent IMS analyses (Fig. 3A, *6-7*) and therefore are identified when multiple samples are analyzed in, preferably, different orientations and order of acquisition when more than one sample is placed on one MALDI plate.

Molecular annotation. Microbial IMS not only detects peptides, but also small molecules, lipopeptides, metal-chelating siderophores, and other types of molecules. This approach to simultaneously detect multiple types of molecules leads to challenges in molecular annotation. In general, annotation of mass spectrometry data is an ongoing problem not just limited to IMS, and the issue of data annotation is apparent as annotation of 10-15% of an advanced “omics” dataset from a simple system, such as proteomics of a single organism with a small genome, requires a skilled informatician (20). In studies of vertebrates that contain microbiomes where host cells are outnumbered 10 to 1 by microbes (e.g. mouse or human samples), it is not atypical for a skilled mass spectrometrist and informatician to annotate only 5% of the data. Similar percentages of identifications are obtained in metabolomics, indicating a need to develop novel approaches to identify molecules in mass spectrometry datasets (5).

Determining the origin of observed molecules. If an ion of interest from a dual-species interaction is spatially localized to or diffusing away from one of the colonies in the interaction, the colony nearest to the ion is assumed to be the producer and further experimentation either confirms or negates this. In this instance, IMS should be treated as a hypothesis-generating tool with observations verified by additional experimental approaches. To initially attempt the determination of the originating organism, interacting colonies and individual control colonies are prepared and analyzed on one target plate to observe cross correlations in mass signals. Based on the IMS data, pursuable candidate ions are selected and tandem MS is attempted directly from the IMS sample to obtain information about the nature of the compound from its fragment masses. Attempts are also made to extract, purify and determine the structure or structural motifs of the molecule via other tandem MS approaches and NMR (12, 19, 22, 32, 33).

If the genomes are available, genome mining approaches (7, 26) may assist in the annotation of metabolic exchange factors observed in microbial IMS (4, 8, 31). Genome mining uses the knowledge about the biosynthetic enzymes involved to annotate and confirm the molecules produced. However, genome mining alone will not uncover which metabolites are actually produced. Peptidogenomics is another tool that can be utilized that correlates amino acid motifs annotated from tandem MS approaches with predicted metabolites from putative gene clusters (19). This tool involves searching a six-frame translation of the genome sequence for the sequence tag and is useful in identifying unannotated ribosomally encoded molecules.

If the genus is unknown or the genome sequence is unavailable, identification via

16S rRNA sequencing (29) or established MALDI-TOF protocols (27, 28) is attempted. Once the genus and/or species of the microbes are/is known, we search for known molecules associated with the masses of interest observed from IMS analysis. However, a majority of the masses from IMS do not match to characterized molecular entities when the organism has not been investigated thoroughly by previous investigators.

The most crucial step in determining which microbe produces the molecule of interest is confirmation. Two main strategies are genetic knockout and complementation studies or assays with purified compound, in combination with more IMS. In a number of cases, the molecules of interest have roles in the development and defense of the colony; rendering the colony genetically incapable of producing such important molecules impacts development and fitness. Genetic complementation or incubation with pure compound may restore normal development and fitness, thereby confirming the biological importance of the molecules observed via IMS.

References

1. **Alaupovic, P., Olson, A.C., Tsang, J.** 1966. Studies on the characterization of lipopolysaccharides from two strains of *Serratia marcescens*. Ann NY Acad Sci **133:**546-65.
2. **Arima, K., Kakinuma, A., Tamura, G.** 1968. Surfactin, a crystalline peptidelipid surfactant produced by *Bacillus subtilis*: Isolation, characterization and its inhibition of fibrin clot formation. Biochem Bioph Res Co **31:**488-494.
3. **Bassler, B.L., Losick, R.** 2006. Bacterially speaking. Cell **125:**237-246.

4. **Begley, M., Cotter, P.D., Hill, C., Ross, R.P.** 2009. Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins. Appl Environ Microbiol **75:**5451-60.

5. **Bowen, B.P., Northen, T.** 2010. Dealing with the unknown: metabolomics and metabolite atlases. J Amer Soc Mass Spectr **21:**1471-76.

6. **Branda, S.S., Gonzalez-Pastor, J.E., Ben-Yehuda, S., Losick, R., Kolter, R.** 2001. Fruiting body formation by *Bacillus subtilis*. Proc Natl Acad Sci USA **98:**11621-11626.

7. **Challis, G.L.** 2008. Genome mining for novel natural product discovery. J Med Chem **51:**2618-2628.

8. **Claesen, J., Bibb M.** 2010. Genome mining and genetic analysis of cypemycin biosynthesis reveal an unusual class of posttranslationally modified peptides. Proc Natl Acad Sci USA **107:**16297-302.

9. **Cohen, S.L., Chait, B.T.** 1996. Influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins. Anal Chem **68:**31-37.

10. **Demeure, K., Quinton, L., Gabelica, V., De Pauw, E.** 2007. Rational selection of the optimum MALDI matrix for top-down proteomics by in-source decay. Anal Chem **79:**8678-8685.

11. **Edwards, J.R., Hayashi, J.A.** 1965. Structure of a rhamnolipid from *Pseudomonas aeruginosa.* Arch Biochem Biophys **111:**415-421.

12. **Gonzalez, D.J., Haste, N.M., Hollands, A., Fleming, T., Hambya, M., Pogliano, K., Nizet, V., Dorrestein, P.C.** 2011. Microbial competition between *Bacillus subtilis* and *Staphylococcus aureus* monitored by imaging mass

spectrometry. Microbiology **57:**2485-2492.

13. **Gusev, A.I., Wilkinson, W.R., Proctor, A., Hercules, D.M.** 1995. Improvement of signal reproducibility and matrix/comatrix effects in MALDI analysis. Anal Chem **67:**1034-1041.

14. **Hisatsuka, K., Nakahara, T., Minoda, T., Yamada, K.** 1971. Formation of rhamnolipid by *Pseudomonas aeruginosa* and its function in hydrocarbon fermentation. Agric Biol Chem **35:**686-692.

15. **Inagawa, H., Nishizawa, T., Tsukioka, D., Suda, T., Chiba, Y., Okutomi, T., Morikawa, A., Soma, G.I., Mizuno, D.** 1992. Homeostasis as regulated by activated macrophage. II. LPS of plant origin other than wheat flour and their concomitant bacteria. Chem Pharm Bull (Tokyo) **40:**994-7.

16. **Itoh, S., Honda, H., Tomita, F., Suzuki, T.** 1971. Rhamnolipid produced by *Pseudomonas aeruginosa* grown on n-paraffin. J Antibiot **24:**855-859.

17. **Jeffries, C.D., Rogers, H.E.** 1968. Enhancing effect of agar on swarming by *Proteus*. J Bacteriol **95:**732-733.

18. **Kearns, D.B.** A field guide to bacterial swarming motility. 2010. Nat Rev Microbiol **8:**634-644.

19. **Kersten, R.D., Yang, Y.-L., Xu, Y., Cimermancic, P., Nam, S.J., Fenical, W., Fischbach, M.A., Moore, B.S., Dorrestein, P.C.** 2011. A mass spectrometry-guided genome mining approach for natural product peptidogenomics. Nat Chem Biol **7:**794-802.

20. **Kim, S., Bandiera, N., Pevzner, P.A.** 2009. Spectral profiles, a novel representation of tandem mass spectra and their applications for de novo peptide

sequencing and identification. Mol Cell Proteomics **8:**1391-400.

21. **Lee, K.-H., Jun, S., Hur, H.-S., Ryu, J.-J., Kim, J.** 2005. *Candida albicans* protein analysis during hyphal differentiation using an integrative HA-tagging method. Biochem Biophys Res Co **337:**784-790.

22. **Liu, W.-T., Yang, Y.-L., Xu, Y., Lamsa, A., Haste, N.M., Yang, J.Y., Ng, J., Gonzalez, D., Ellermeier, C.D., Straight, P.D., Pevzner, P.A., Pogliano, J., Nizet, V., Pogliano, K., Dorrestein, P.C.** 2010. Imaging mass spectrometry of intraspecies metabolic exchange revealed the cannibalistic factors of *Bacillus subtilis*. P Natl Acad Sci USA **107:**16286-90.

23. **Matsuyama, T., Sogawa, M., Nakagawa, Y.** 1989. Fractal spreading growth of *Serratia marcescens* with produces surface active exolipids. FEMS Microbiol Lett **52:**243-6.

24. **Matsuyama, T., Kaneda, K., Nakagawa, Y., Isa, K., Hara-Hotta, H., Yano, I.** A novel extracellular cyclic lipopeptide which promotes flagellum-dependent and –independent spreading growth of Serratia marcescens. J Bacteriol **174:**1769-76.

25. **Padliya, N.D., Wood, T.D.** 2004. A strategy to improve peptide mass fingerprinting matches through the optimization of matrix-assisted laser desorption/ionization matrix selection and formulation. Proteomics **4:**466-473.

26. **Raush, C., Weber, T., Kohlbacher, O., Wohlleben, W., Huson, D.H.** 2005. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). Nucleic Acids Res **33:**5799-5808.

27. **Saffert, R.T., Cunningham, S.A., Ihde, S.M., Jobe, K.E., Mandrekar, J., Patel,**

R. 2011. Comparison of Bruker Biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometer to BD Phoenix automated microbiology system for identification of gram-negative bacilli. J Clin Microbiol **49:**887-92.

28. **Sogawa, K., Watanabe, M., Sato, K., Segawa, S., Ishii, C., Miyabe, A., Murata, S., Saito, T., Nomura, F.** 2011. Use of the MALDI BioTyper system with MALDI-TOF mass spectrometry for rapid identification of microorganisms. Anal Bioanal Chem **400:**1905-1911.

29. **Stackebrandt, E., Goebel, B.M.** 1994. Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Evol Micr **44:**846-849.

30. **Straight, P.D., Willey, J.M., Kolter, R.** 2006. Interactions between *Streptomyces coelicolor* and *Bacillus subtilis*: Role of surfactants in raising aerial structures. J Bacteriol **188:**4918-25.

31. **Velasquez, J.E., van der Donk, W.A.** 2011. Genome mining for ribosomally synthesized natural products. Curr Opin Chem Biol **15:**11-21.

32. **Yang, Y.-L., Xu, Y., Kersten, R.D., Liu W.-T., Meehan, M.J., Moore, B.S., Bandeira, N., Dorrestein, P.C.** 2011. Connecting chemotypes and phenotypes of cultured marine microbial assemblages by imaging mass spectrometry. Angew Chem Int Edit **50:**5839-5842.

33. **Yang, Y.-L., Xu, Y., Straight, P.D., Dorrestein, P.C.** 2009. Translating metabolic exchange with imaging mass spectrometry. Nat Chem Biol **5:**885-887.

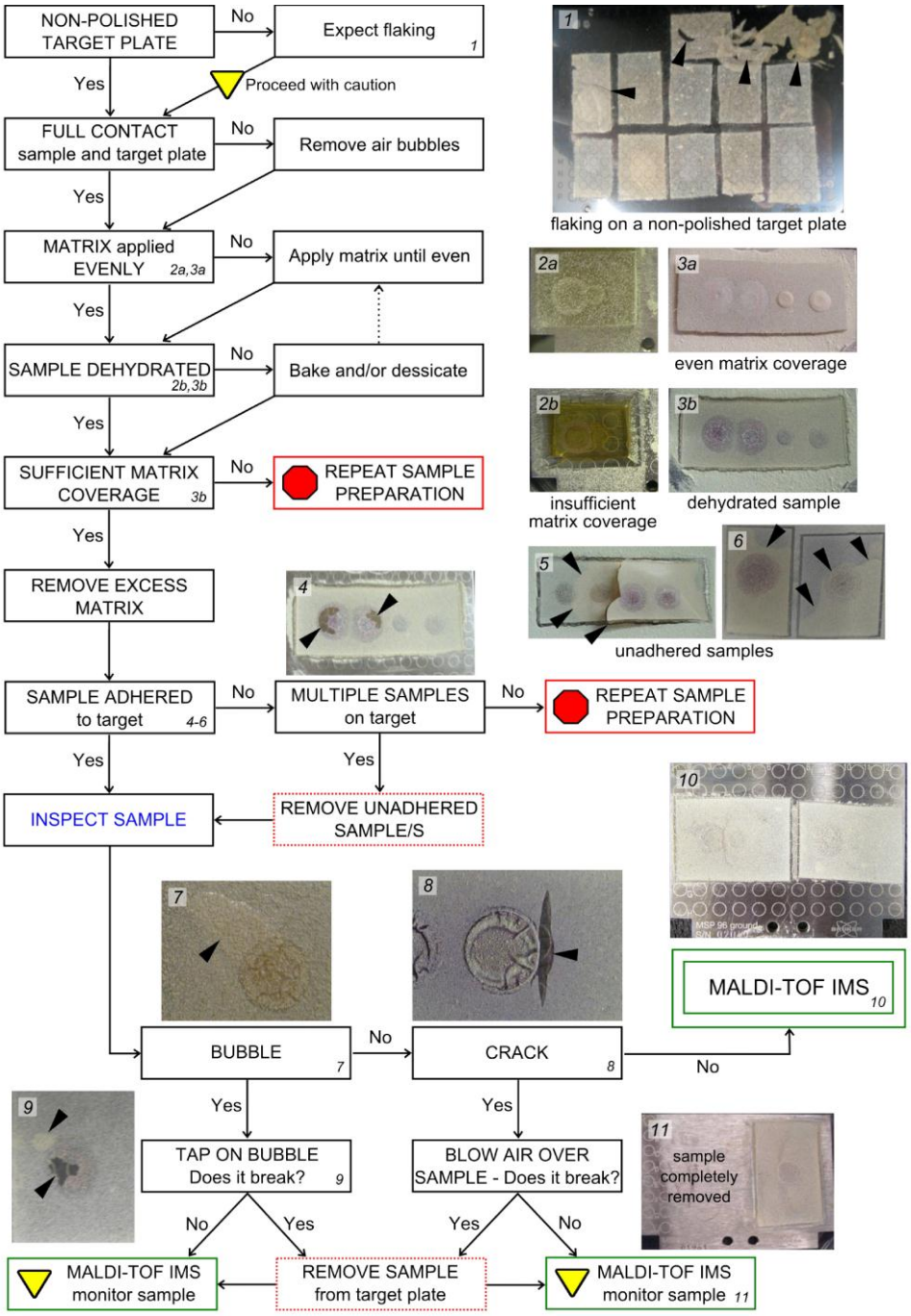


Figure S1. How to minimize sample flaking inside instrument. At the first checkpoint

we strongly recommend a non-polished MALDI target plate. If a polished target plate is used for microbial IMS, flaking will most likely occur (*1*, arrows in example pictures point to bubbles, cracks, and flakes). The second checkpoint assures that there is no air trapped between the sample and target plate. Once the sample on the target plate is bubble-free, matrix is applied until even and saturated (*3a*); then the sample is dehydrated (*3b*). There are media and microbes that may necessitate multiple rounds of matrix application during the dehydration process in order to achieve even coverage and saturation (refer to Fig. S3). If matrix coverage does not saturate the sample, the agar shrinks (*2b*) and sample preparation needs to be repeated. If there is sufficient matrix coverage after dehydration, use air to remove the loose matrix from the sample. This will also test whether or not the sample flaked during dehydration (*4-6*). If a portion of any sample flakes, repeat the sample preparation or carefully remove the flaking sample (*11*), then wipe the area with methanol and a Kimwipe, and proceed with sample inspection. Despite passing the previous checkpoints, there may be a bubble (*7*) or crack (*8*) in the dehydrated sample. The first attempts to test whether or not the sample may flake include tapping the bubble with a spatula (*9*) or blowing air over the cracked sample. If the sample remains intact, proceed to MALDI-TOF IMS and monitor the data acquisition for signs of flaking inside the instrument, such as sample fluttering or missing sample in the live feed, a spike in vacuum pressure, or a delay in the data acquisition run.

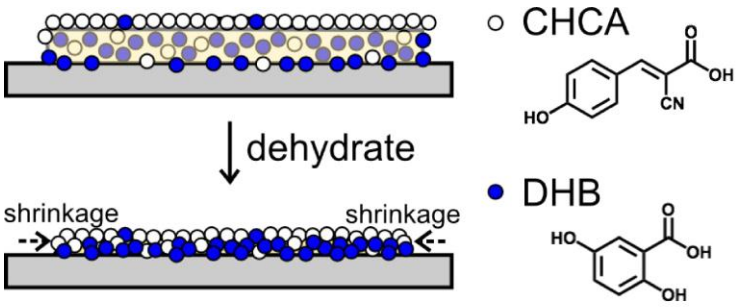
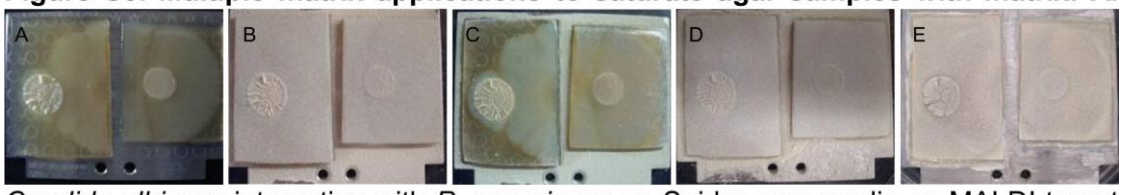


Figure S2. Our hypothesis of action of dual-matrix. Samples saturated with a 1:1 mixture of 2,5-dihydroxybenzoic acid (DHB) and alpha-cyano-4-hydroxycinnamic acid (CHCA) matrices are more likely to adhere to the target plate. We hypothesize that more DHB is absorbed into the agar and acts as an adhesive, whereas more CHCA stays on top of the sample and that saturation with matrix minimizes sample shrinkage during dehydration. The combination of these two matrices assists in adherence of the sample to the target and ionization of a variety of biomolecules from microbial samples on agar-based media.

Figure S3. Multiple matrix applications to saturate agar samples with matrix. A.



Candida albicans interacting with *P. aeruginosa* on Spider agar media on MALDI target plate. **B**. First matrix application shows even coverage and what appears to be saturation. **C**. After 1 h dehydration, the agar has absorbed much of the matrix and the sample is not saturated with matrix. **D**. Second matrix application shows even matrix coverage and saturation. **E**. The sample is completely dehydrated, matrix coverage is even, and the excess matrix has been removed.

Chapter 1, in full, is a reprint of the material as it appears in the Journal of Bacteriology, July 2012. Jane Y. Yang, Vanessa V. Phelan, Ryan Simkovsky, Jeramie D. Watrous, Rachelle M. Trial, Tinya C. Fleming, Roland Wenter, Bradley S. Moore, Susan S. Golden, Kit Pogliano, and Pieter C. Dorrestein. The dissertation author was the primary investigator and author of this paper.

CHAPTER II

Application of MALDI-TOF MS and agar-based microbial imaging mass spectrometry in the study of *Bacillus subtilis*

2.1 Abstract

Microbes communicate within colonies and with their surroundings through secondary metabolites. *Bacillus subtilis*, a sporulating Gram-positive microorganism produces molecules with antibacterial and antifungal properties that serve as communication factors to drive cell differentiation within its colony. Signals from the environment activate pathways that upregulate sigma factors and/or extracytoplasmic function (ECF) sigma factors, which regulate gene transcription, leading to differential gene expression, cell differentiation, and ultimately, secondary metabolite production. We hypothesize that mutations affecting sigma factors and extracytoplasmic function (ECF) sigma factors affect secondary metabolite production in *B. subtilis*. Matrix assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry (MS) and agar-based MALDI-TOF imaging mass spectrometry (IMS) were used to evaluate the production and distribution of metabolites in colonies of *B. subtilis* NCIB3610 and PY79 and regulatory factor mutants.

2.2 Introduction

Planet earth is teeming with microbes. One tablespoon of soil is estimated to contain 10 million bacteria and a total of 50 billion microbes.[1-3] In nature, microorganisms interact with other microbes and their environments via molecules referred to as secondary metabolites or natural products.[4-6] These metabolites are not necessarily involved in primary metabolism pathways, can act within a colony as quorum sensors,[7] and have activity against other bacteria or fungi.[8, 9] Due to the production of specialized metabolites, the bacterial genus *Bacillus* has been used as probiotics,[10, 11] in agriculture,[12] and in industrial applications[13] to produce surfactants and amylase.[14-18]

Bacillus subtilis, a well-characterized sporulating Gram-positive microbe is found in soil, the human gut,[19-21] and contaminated wounds.[22] When cultured on solid media in the laboratory, *B. subtilis* differentiates into subpopulations, which produce lipopeptide antibiotic biosurfactants surfactin and plipastatin, an antimicrobial peptide subtilosin, and peptidic cannibalistic factors sporulation killing factor (Skf) and sporulation delaying protein (Sdp). These secondary metabolites are produced in response to intracellular, self-produced, and extracellular signals, such as pH, extracellular salts, and starvation.[7, 23, 24]

Multicellularity occurs in response to signals that are sensed by kinases within *B. subtilis* cells, which activate one of three master regulators – DegU, ComA, and Spo0A – each of which, in turn, activates pathways that result in different cell types.[7, 25] Following the activation of a master regulator, different sigma factors, which help

initiate bacterial transcription, bind to specific gene promoters indicating to RNA polymerase where to begin transcription. Thus, different genetic programs are transcribed depending on the input signals, ultimately leading to differentiation[26, 27] and production of secondary metabolites.

The default cell state of *B. subtilis* prior to differentiation has been speculated to be motile or flagellated cells,[25] which are a majority of the colony population before 24 h on biofilm-inducing media,[24] because this cell state does not depend on the activation of any of the three master regulators. Activation of DegU results in miner cells that secrete proteases, whereas activation of ComA yields surfactin producers and competent cells, and activation of Spo0A leads to extracellular matrix producers, cannibals, and spores. The markers for the activation of one of the three master regulators are the secondary metabolites produced by the different cell populations. Surfactin producers secrete a biosurfactant which alters surface properties, allows flagellated cells to swarm, and the colony to expand in search of nutrients or space.[28] Cannibals produce two cyclic peptide antibiotics, sporulation killing factor (Skf) and sporulation delaying protein (Sdp) in a self-sacrificial effort to generate nutrients to delay the activation of the terminal sporulation pathway. Additionally, the different cell types contribute towards survival by induction of competence to increase genetic variability, generation of nutrients via protease secretion, production of extracellular matrix for structure or protection, and formation of dormant spores to preserve genetic material until conditions become more favorable.[29-31]

A genetically homogeneous colony of *B. subtilis* is able to differentiate into subpopulations based on the set of genes transcribed and expressed within each cell.

Transcriptional regulation in *B. subtilis* involves 10 sigma factors and 7 extracytoplasmic function (ECF) sigma factors.[32, 33] Sigma factors are involved in housekeeping, stress factors, chemotaxis, motility, and different stages of sporulation.[34] Transcriptomics studies showed overlap between sigma M, W, and X in cell wall stress and biofilm formation and suggested the hypothesized the involvement of sigma factors V, Y, Z, and ylaC in exopolysaccharide production,[32, 35, 36] however, not much is known about these ECF sigma factors.

In this study, we used two progenitor strains: *B. subtilis* NCIB3610, an undomesticated wild strain, hereafter referred to as 3610, and PY79,[37-40] a laboratory domesticated strain that has a frameshift mutation in the *sfp* gene,[41] which encodes for the phosphopantetheinyl transferase involved in non-ribosomal peptide synthetase (NRPS) megaenzymes responsible for the production of certain secondary metabolites, such as surfactin and plipastatin.[42, 43] Hence, PY79 is deficient in surfactin and plipastatin production. Looking at strains deficient in sigma factors in both may provide insight into understanding cellular differentiation and transcriptional regulation through the production of secondary metabolites from *B. subtilis*. We hypothesize that transcriptional regulatory factors, such as sigma factors and ECF sigma factors, affect global secondary metabolite production in *B. subtilis*.

First, MALDI-TOF MS was used to profile the metabolic output of *B. subtilis* 3610 and PY79 colonies grown on three types of solid media over multiple time points. Media type limits the nutrients, carbon source, and salts available to the growing colony, which may affect the timing and quantity of secondary metabolite production by these different celltypes.[7, 25] These metabolomic profiles assisted in

establishing the molecules detectable by MALDI-TOF and the times at which the molecules were produced, and also served as a preview of the ions that would be detected in IMS analysis.

Based on the timing of Skf production, an indication of the induction of the sporulation pathway, the 48 h time point on yeast extract-malt extract (ISP2) agar media was selected for IMS analysis. This time point captured the production of the hallmark *B. subtilis* secondary metabolites,[44] but was still prior to the detection of the Sdp, produced later in the sporulation pathway than Skf.[45] This ensured that the secondary metabolites detected in IMS were not suppressed by the presence of spores.

To test the hypothesis that transcriptional regulators affect the production of secondary metabolites of *B. subtilis*, IMS was conducted on 23 mutants in both 3610 and PY79 that directly or indirectly affect 5 sigma factors (B, D, F, H, and L), 7 ECF factors (M, W, X, V, Y, Z, ylaC), and other genes involved in secretion or matrix production (Table 2.1) to identify global patterns of molecules associated with cellular differentiation (Figure 2.1).

We also evaluated the use of IMS as a screening tool by analyzing unidentified mutants and comparing their IMS results to the IMS of the genetically characterized progenitor and mutant strains. The unidentified mutants included one potential *yidC2* mutant that was sporulation deficient and swarming proficient, as well as 11 uncharacterized mutants of SCB844 (*amyE*::*yidC2*^{6}*-lacZΩcat*) that were evaluated to either be inducible by isopropyl β-D-1-thiogalactopyranoside (IPTG) in liquid media or only on solid media. IPTG induction results in YidC2^{6}-LacZ expression, or β-galactosidase activity.

Table 2.1 *B. subtilis* strains used in this study

Strain	Genotype and description
NCIB3610 (referred to as 3610)	Undomesticated wild strain; prototroph
PY79	Lab domesticated strain; prototroph derivative of *B. subtilis* 168
ALB1033A	3610 Δ*yidC2*::*mls* *yidC2* – Sec-independent membrane protein translocase; membrane insertion of proteins and protein secretion
ALB975A	3610 Δ*spoIIIJ*::*cm*::*spec* *spoIIIJ* – membrane protein translocase; membrane insertion of proteins and protein secretion; constitutively expressed
KP1032	3610 *sigD*::*tet* *sigD* – induces motility and swarming genes
KP535	3610 Δ*spoVS*::*spec* (deletion increases sigma D activity)
ALB1035A	3610 Δ*spo0A*::*erm* *spo0A* – induction of matrix production and sporulation
AR123	PY79 Δ*yidC2*::*mls*
XJ149	PY79 Δ*spoIIIJ*::*cm*::*spec*
ALB1006A	PY79 *sigD*::*tet*, *amyE*::*yidC2*^{6}-*lacZΩcat*
KP789	PY79 Δ*abrB*::*cm* *abrB* – represses matrix production, induces competence
KP798	PY79 Δ*sinR*::*neo* *sinR* – represses matrix production
ALB964A	PY79 Δ*spo0A*, *amyE*::*yidC2*^{6}-*lacZΩcat* (to verify that Δ*spo0A* caused elevated surfactin production[45])
ALB1045A	PY79 *sinI*::*kan* sinI 1S98 represses sinR, induces matrix production
ALB1036A	PY79 *sigB*::*cat* sigB 1A675 general stress response
KP324	PY79 Δ*spoIIAC*::*kan* (sigma F) *spoIIAC*, sigma F, early forespore sigma factor
KP427	PY79 pSPAC-*spo0HΩerm* (sigma H) *sigH* – transition from exponential to stationary phase, induces matrix production by inducing Spo0A
ALB1044A	PY79 *sigL*::*kan* sigL JH642 1A914 cold shock response
ALB1040A	PY79 *sigM*::*kan* sigM 168 1A906 stress response, ECF sigma factor

Table 2.1 *B. subtilis* strains used in this study, continued

Strain	Genotype and description
ALB1039A	PY79 *sigW*::*mls* sigW 168 1A905 antibiotic resistance, ECF sigma factor
ALB1037A	PY79 *sigX*::*spec* sigX 1A901 cell wall metabolism, ECF sigma factor, together w/ SigM involved in teichoic acid biosynthesis and septum formation
ALB1041A	PY79 *sigV*::*kan* sigV 168 1A907 overlaps sigM, sigW, and sigX, ECF sigma factor
ALB1043A	PY79 *sigY*::*mls* sigY 168 1A909 regulating a toxic peptide and its immunity gene, ECF sigma factor
ALB1038A	PY79 *sigZ*::*kan*
ALB1042A	PY79 *ylaC*::*kan* ylaC 168 1A908 hydrogen peroxide resistance, ECF sigma factor
ALB244 ALB247 ALB402 ALB456 ALB608	Unidentified mutants of SCB844 (*amyE::yidC2*^{6}*-lacZΩcat*) that are induced in liquid media.
ALB261 ALB554 ALB607 ALB609 ALB725 AL729	Unidentified mutants of SCB844 that are only induced on solid media.
ALB730	possible *yidC2* mutant sporulation deficient and swarming proficient

Δ = deletion
- = fusion
Ω = genetic contruct introduced from a two-point crossover
:: = insertion
mls = macrolide-lincosamide-streptogramin, cm = chloramphenicol, spec = spectinomycin, *tet* = tetracycline, *erm* = erythromycin, neo = neomycin, kan = kanamycin

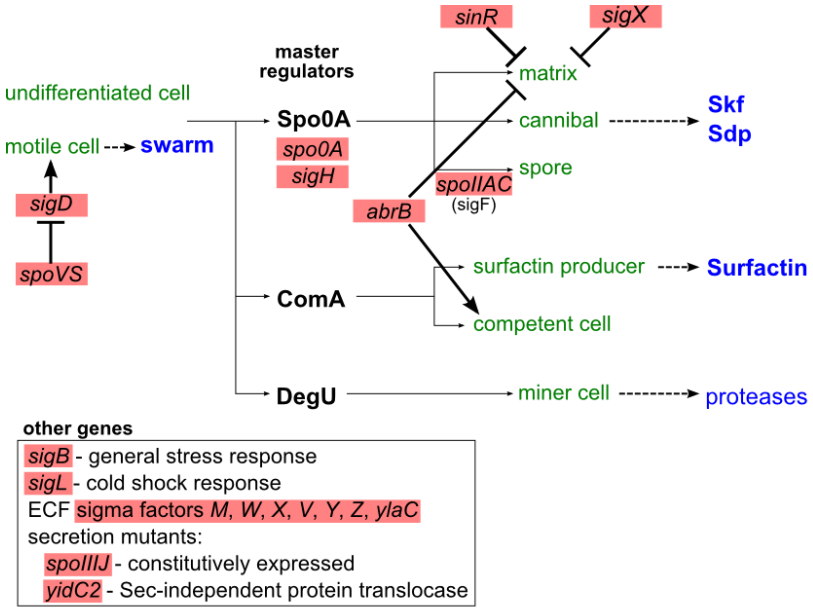


Figure 2.1 Cell differentiation pathways in *B. subtilis*. Cell differentiation pathways are indicated by the thin arrows. Cell types are in green font. The three master regulators are in bold font. The biological output, phenotype or secreted molecules from the cells, are indicated by the dashed arrows and blue font. The genes mutated in this study are italicized and highlighted with a pink box. This figure was modified from a schematic in Lopez and Kolter 2009.[7, 46]

2.3 Results

B. subtilis 3610 and PY79 colonies were distinguishable on the different growth media by the color of the underside of the colony by Day 3 and the shape of the colonies beginning on Day 4 (Figure 2.2). Beyond Day 3, the color ranged from white or pale yellow to slightly red or reddish-brown in the colony center, depending on the media type. The colonies were round with smooth edges, except for 3610 colonies on Difco sporulation media (DSM), which induces *B. subtilis* sporulation,[47] and PY79 on Luria Broth (LB) agar, which swarmed as indicated by the uneven growth along the edges of the colony. *B.subtilis* AR123 (Δ*yidC2*::*mls*), deficient in Sec-independent membrane insertion of proteins but still capable of protein secretion through *spoIIIJ*,[48] phenotypically resembled progenitor strain PY79 (Supplemental Figure S2.1).

MALDI-TOF analysis of *B. subtilis* over the course of six days (Figure 2.3) revealed that 3610 and PY79 shared the genetic capacity to produce secondary metabolites. In MALDI-TOF MS, the ions typically detected are $[M+H]^+$, reported as a mass to charge ratio (*m/z*) or Da after subtracting the mass of a proton; these values are used interchangeably. On ISP2, strain 3610 is known to produce surfactin (1008.6597-1074.6468 calc. $[M+H]^+$-$[M+K]^+$ for varying acyl chain lengths), plipastatin (1463.8083-1529.7909 calc. $[M+H]^+$ - $[M+K]^+$ for varying acyl chain lengths and amino acid variants), subtilosin (1134.1972 calc. $[M+H]^+$; 3400.4 $[M+H]^+$ in MALDI), a partially characterized polyglutamate (*m/z* 715), and unknown molecule (*m/z* 655) on ISP2.[44] And strain PY79 is known to produce Skf (2781.302 Da) and Sdp

(4311.209 $[M+H]^+$).[45] These *B. subtilis* molecules were detected via MALDI-TOF profiling. In addition, an uncharacterized molecule detected ~2100 Da was present at detectable levels in PY79 on ISP2, whereas the molecule ~3850 Da was present in 3610.

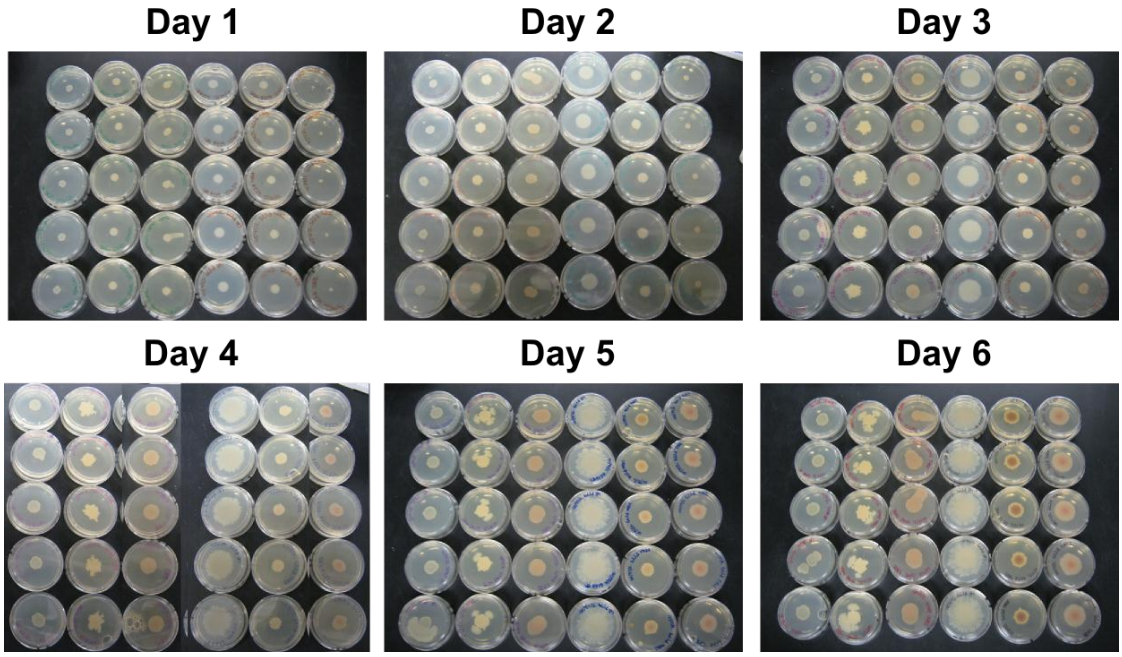


Figure 2.2 Development of colony morphology and color of *B. subtilis* 3610 and PY79 on LB, DSM, and ISP2 media over 6 days. The bottoms of the Petri dishes from each time point were photographed. The actual time of incubation is indicated in parentheses. From L to R for each time point (in replicates of 5): 3610 on LB, DSM, ISP2 and PY79 on LB, DSM, and ISP2. The photograph for Day 4 was corrected to maintain consistent order.

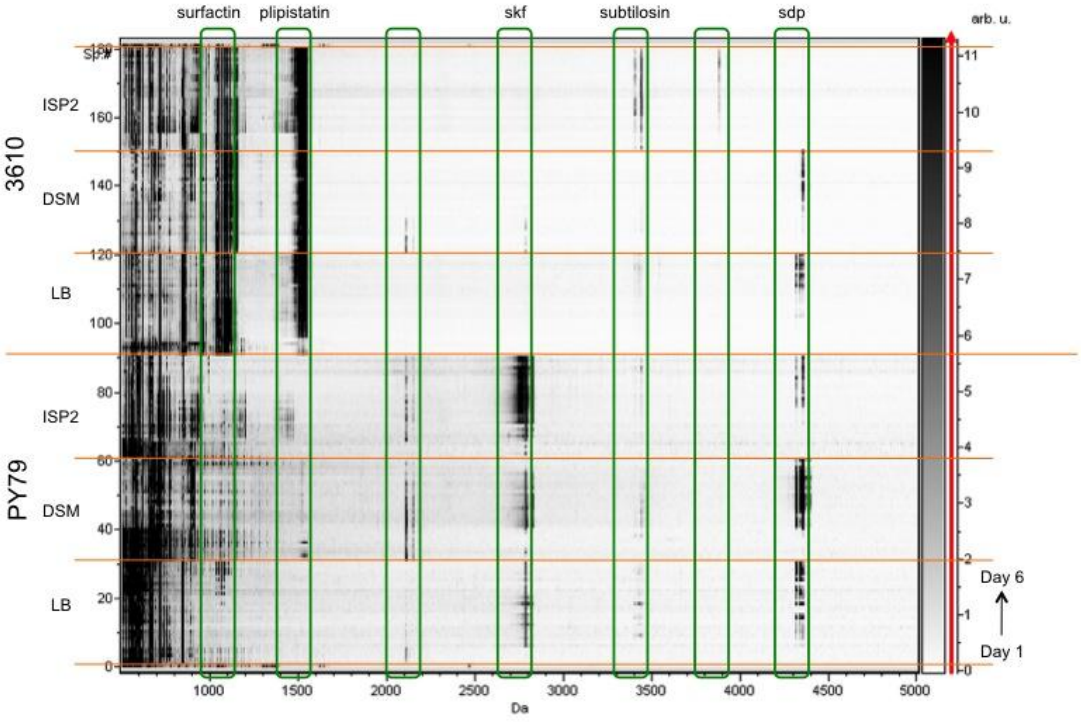


Figure 2.3 Gel/Stack view of 182 spectra reveals patterns of secondary metabolite production of *B. subtilis* 3610 and PY79 over time. Metabolites were detected from 500-5000 Daltons (Da; x-axis). The spectra were numbered 0-181 (left y-axis), and the relative intensity was adjusted to 0-11.5 (right y-axis). The spectra on the very top and bottom are of the Peptide Calibration Standard. The top half corresponds to spectra for 3610, the bottom half, PY79, indicated by the longest horizontal orange line. These halves are split into three sections (shorter horizontal orange lines), corresponding to growth on three different media, ISP2, DSM, and LB (from top to bottom), from Day 1-6 (from bottom to top). Green rectangles highlight known and unknown molecules that are detected by MALDI-TOF MS.

3610 produced surfactin on all three media by Day 1, whereas PY79 produced surfactin lipopeptides 1000-1100 Da by Day 5 on LB. Plipastatin 1450-1550 Da was detected in 3610 by Day 1 on DSM and ISP2, but wasn't detected until Day 2 on LB. Comparatively, PY79 did not produce much plipastatin. Unknown metabolite ~ 2100 Da was detected on Day 1 from 3610 on DSM and PY79 on LB, Days 1-5 from PY79 on DSM, and Days 2-6 from PY79 on ISP2. Subtilosin 3400 Da was detected from 3610 on ISP2 and a little from PY79. The detection of Skf from PY79, unknown metabolite ~3900 Da from 3610 on ISP2, and Sdp from both 3610 and PY79 occurred at Day 2 or later. Metabolites between 500-1000 Da included uncharacterized 551, 616, 655, 886, and the partially characterized polyglutamate at 715 Da (Figure S2.2). MALDI-TOF MS analysis of the metabolic output of PY79 and AR123 was similar on all three media across 6 days (Figure S2.3). The main difference was the timing of Sdp detection on LB; PY79 produced detectable Sdp Days 2-6, whereas Sdp was detected from AR123 on Days 3-6.

MALDI-TOF MS profiles and previously published IMS analyses of strains 3610 and PY79 on ISP2[44, 45] served as the baseline of secondary metabolic output and highlighted *B.subtilis* ions produced at the Day 2 time point. The expected ions from 3610 were *m/z* 616, 655, 715 (the partially characterized polyglutamate), 886, 1075 (the representative surfactin $[M+K]^+$), 1545 (the representative plipastatin $[M+K]^+$), and 3438 (subtilosin); PY79 ions included *m/z* 551, 655, 715, 2785 (Skf), and 4350 (Sdp). Due to the number of mutant strains tested, select ions were displayed in the figures.

B. subtilis mutants in a 3610 background, ALB1033A, ALB975A, KP1032, ALB1035A, and KP535 (*yidC2*, *spoIIIJ*, *sigD*, *spo0A*, and *spoVS* mutants, respectively) produced *m/z* 616, 715, and 1075 (Figure 2.4). Two strains, KP535 and ALB1035A, differ from the expected 3610 metabolic output. KP535 appeared to have two IMS metabolomic profiles. In one instance, KP535 produced *m/z* 551, 886, and Skf, but did not produce surfactin. In the other case, KP535 was more 3610-like. ALB1035A did not produce *m/z* 655, produced *m/z* 551 and 886, and 1075 was colony associated, whereas for the other strains *m/z* 1075 appeared to be secreted.

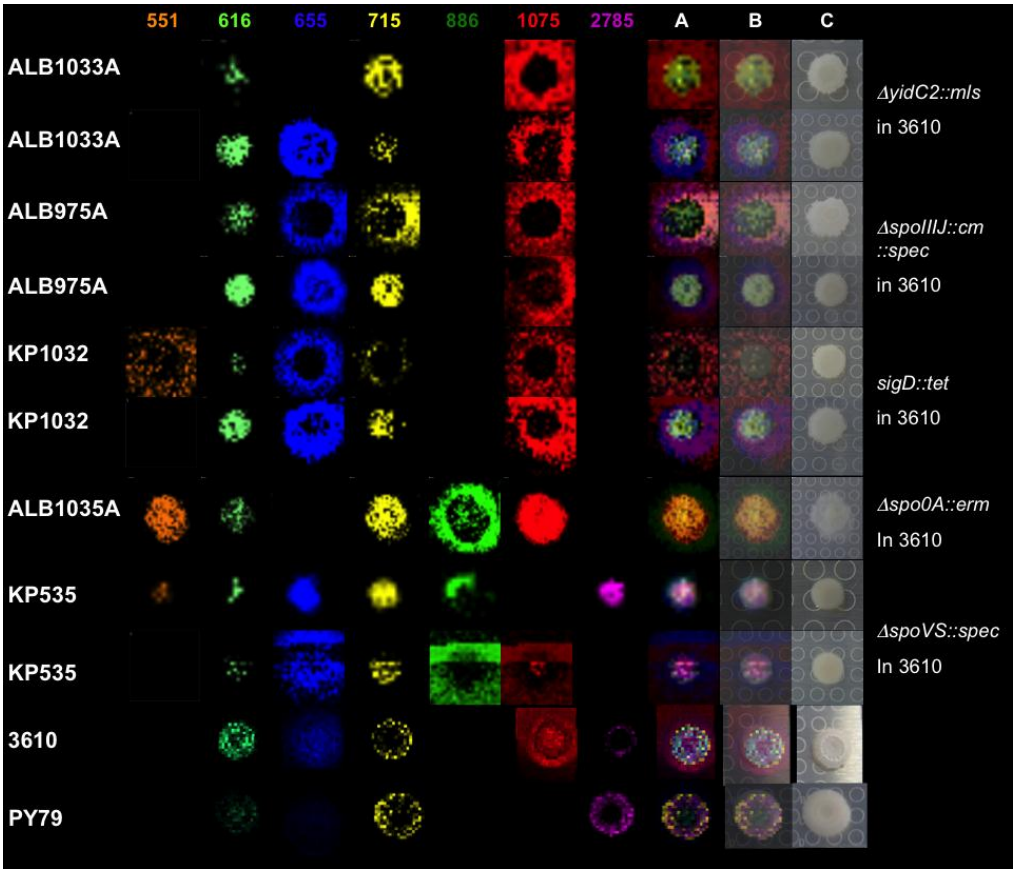


Figure 2.4 IMS of *B. subtilis* 3610 mutants. A. Overlay of false-colored images. B. Overlay of false-colored images on photograph of colony. C. Photograph of colony on MALDI target plate. Circles on the target have a diameter of 3 mm.

IMS of *B. subtilis* mutant AR123 (*yidC2*) in a PY79 background resembled IMS analysis of PY79 without Skf (Figure 2.5). Strains XJ149, ALB1006A, KP789, and KP798, (*spoIIIJ*, *sigD*, *abrB*, and *sinR* mutants, respectively), produced *m/z* 616, 655, 715, 886, and Skf in amounts detectable by IMS (Figure 2.5). These mutants did not produce surfactin or *m/z* 551. ALB964A (*spo0A)* was more 3610-like, with *m/z* 616, 655, 715 and surfactin. However, the surfactin distribution was flowery rather than evenly diffuse around the perimeter of the colony, and ALB964A also produced *m/z* 551. Unlike the other mutants, ALB1045A (*sinI*) did not produce Skf.

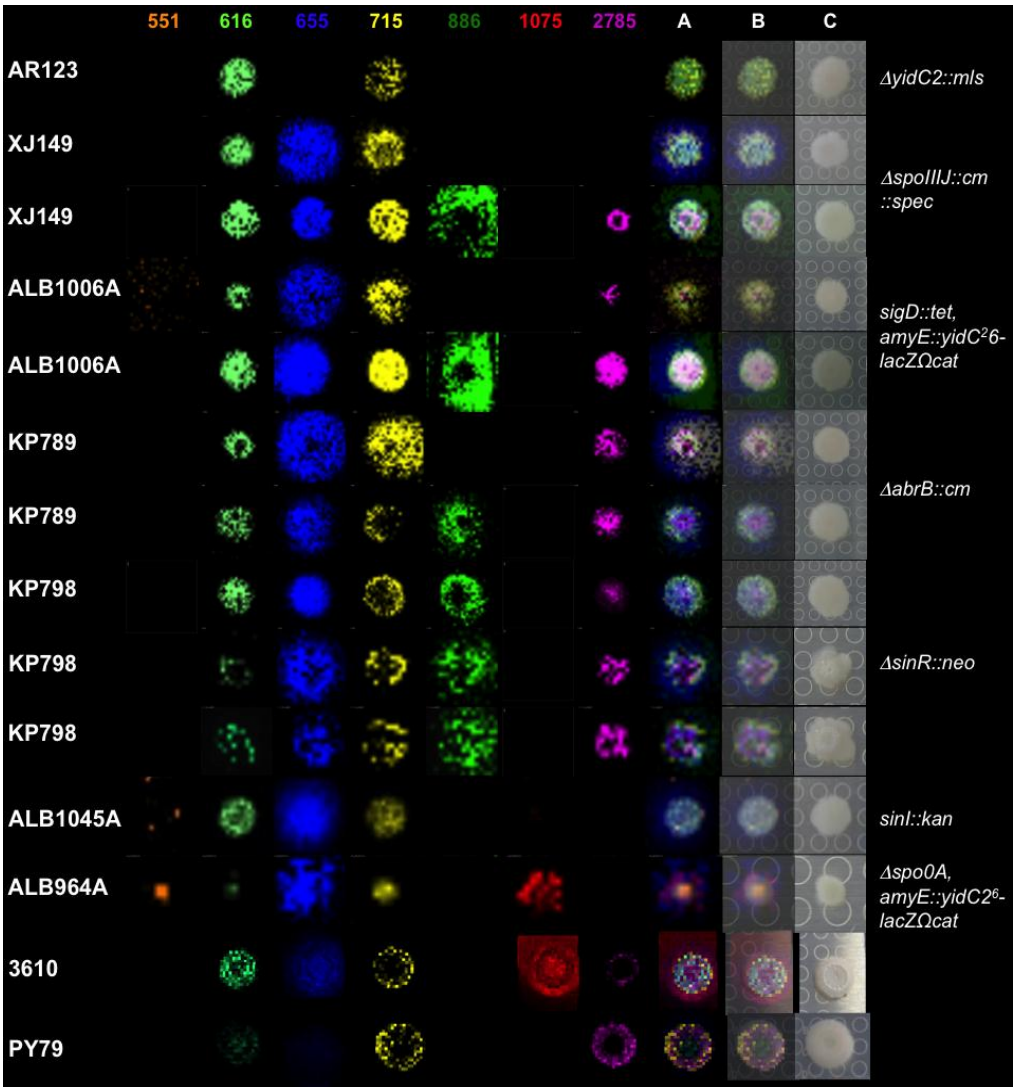


Figure 2.5 IMS of PY79 mutants. A. Overlay of false-colored images. B. Overlay of false-colored images on photograph of colony. C. Photograph of colony on MALDI target plate. Circles on the target have a diameter of 3 mm.

The metabolic output of mutants that directly affect sigma and ECF sigma factors, KP324, KP427, ALB1036A through ALB1044A (*spoIIAC* (sigma F), *spo0H* (sigma H), *sigB*, *sigX*, *sigZ*, *sigW*, *sigM*, *sigV*, *ylaC*, *sigY*, and *sigL*, respectively), all produced *m/z* 616, 655, and 715 at the 48 h time point (Figure 2.6). ALB1040A, ALB1043A, and ALB1044A (*sigM*, *sigY*, and *sigL*) may have produced *m/z* 551. KP324 (*spoIIAC*, affects sigma F) produced *m/z* 886. All but KP324, ALB1040A (*sigM)*, and ALB1041A (*sigV)* produced surfactin, indicated by the representative *m/z* 1075 puncta in the center of the colonies. KP324, ALB1044A, ALB1040A, ALB1039A (*sigW*), ALB1037A (*sigX*), ALB1043A (*sigY)*, ALB1038A (*sigZ)*, and ALB1042A (*ylaC)* produced Skf, mostly punctate in the center of the colonies.

The *B. subtilis* mutants affected by the master regulator Spo0A are: KP324 (*spoIIAC* in PY79), KP535 (*spoVS* in 3610), ALB964A (*spo0A* in PY79), and KP427 (pSPAC-*spo0H*Ω*erm* in PY79; Figures 2.4-2.6). These mutants all produced *m/z* 616, 655, and 715. *M/z* 551 was detected in KP535 and ALB964A. KP324 and KP535 produced *m/z* 886 and Skf, but not surfactin. ALB964A and KP427 did not produce *m/z* 886 or Skf, but did produce colony-associated surfactin.

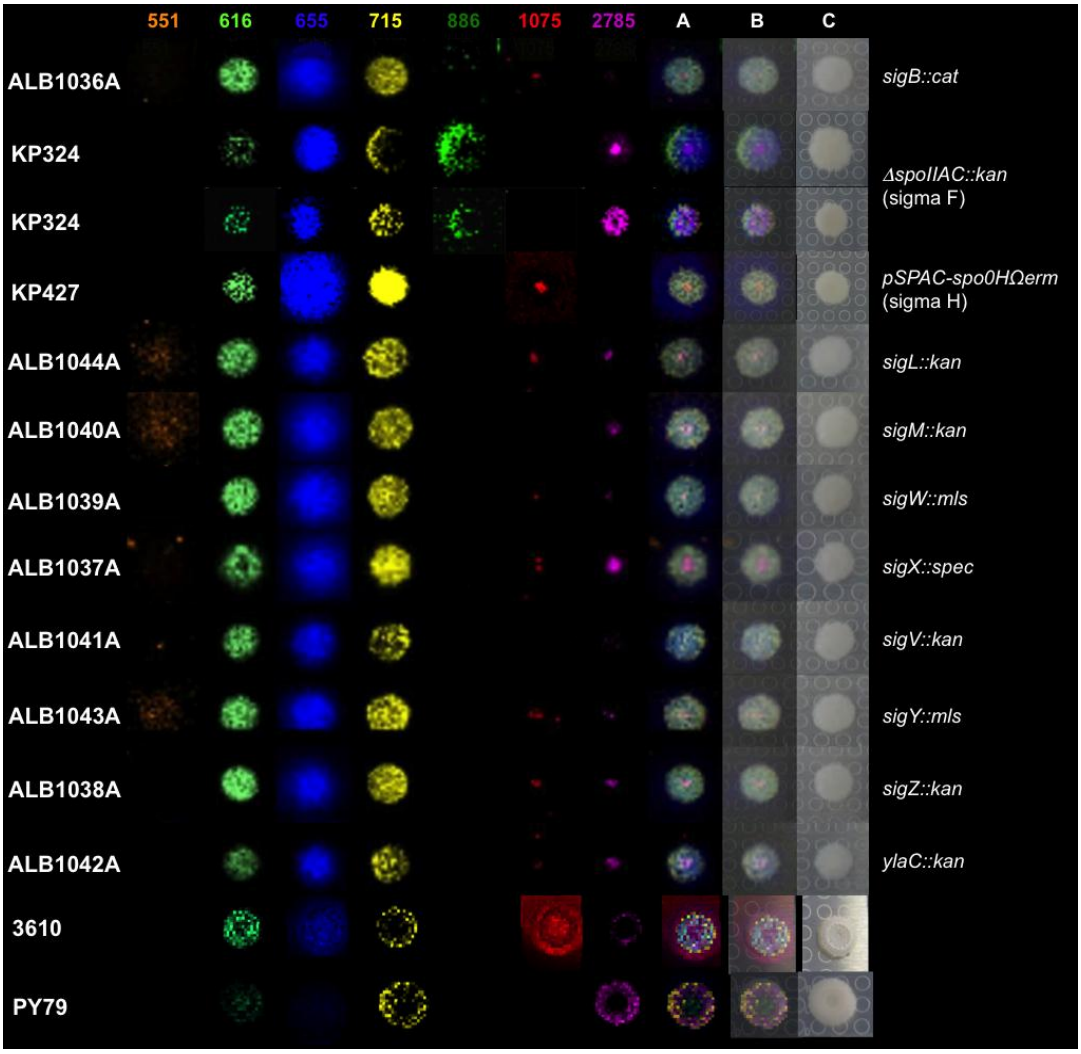


Figure 2.6 IMS of sigma and ECF factor mutants in *B. subtilis* PY79. A. Overlay of false-colored images. B. Overlay of false-colored images on photograph of colony. C. Photograph of colony on MALDI target plate. Circles on the target have a diameter of 3 mm.

IMS metabolite profiles of unidentified mutants of SCB844 (*amyE*::*yidC2*^{6}-*lacZΩcat)* in a PY79 background were varied, but they all produced the partially characterized polyglutamate *m/z* 715 (Figure 2.7). Among the group of mutants that were inducible in liquid media, ALB244 uniquely produced m/z 551, low levels of 886, and colony-associated surfactin. The only detectable ion from ALB247 and ALB402 was *m/z* 715. The colony size of ALB402 was small compared to the other mutants, 3610, and PY79. ALB456 and ALB608 resembled the sigma and ECF factor mutants in IMS.

In the group of unidentified mutants only inducible on solid media, ALB261 looked the most PY79-like by producing *m/z* 616, 655, 715, and Skf (Figure 2.7). The colonies of ALB554, ALB607, ALB609, ALB725, and ALB729 were small; the ions detected were *m/z* 551, 616, and 715, and not *m/z* 886 or Skf. ALB554 and ALB609 produced a ring of surfactin, ALB607 and ALB729 did not produce surfactin, while ALB725 potentially produced low levels of surfactin.

The potential *yidC2* mutant, ALB730 (sporulation deficient, swarming proficient), produced *m/z* 551, 616, no 655, 715, no 886, no surfactin, and no Skf, most closely resembling ALB607 and ALB729 (Figure 2.7). The colony did not swarm, because swarming is defined as movement across a semi-solid surface (0.7% agar), so no conclusions about swarming can be made under these experimental conditions.

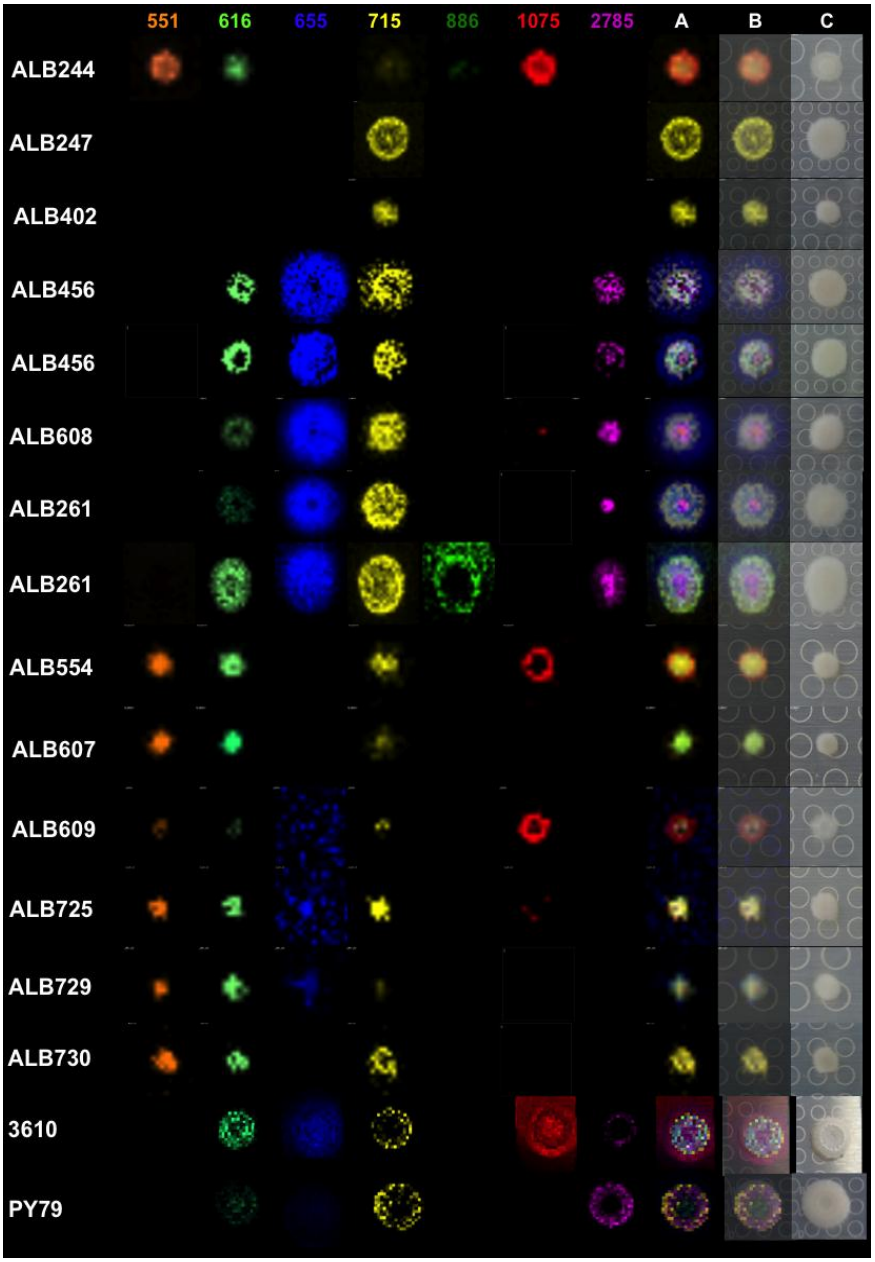


Figure 2.7 Representative IMS images of unidentified mutants of SCB844 (*amyE*::*yidC2*^{6}-*lacZ*Ω*cat*) that are inducible in liquid or only on solid, and ALB730 (sporulation deficient, swarming proficient). A. Overlay of false-colored images. B. Overlay of false-colored images on photograph of colony. C. Photograph of colony on MALDI target plate. Circles on the target have a diameter of 3 mm. From the top, ALB244-ALB608 were induced in liquid and ALB261-ALB729 were induced only on solid. ALB730 resembled the unidentified mutants induced only on solid.

2.4 Discussion

Intra- and extra-cellular signals activate one of three master regulators that lead to differentiation into the different subpopulations of *B. subtilis.* Matrix producers, spores, and cannibal cells are under the activation of master regulator Spo0A,[24] whereas surfactin producers and competent cells are under the regulation of ComA.[7] DegU regulates the miner cells that secrete proteases,[7] which were not the focus of this study. Depending on the nature of the signals, the master regulators activate different sigma factors, which result in different gene transcription programs. MALDI-TOF MS and IMS were used to analyze the secondary metabolites from strains deficient in transcriptional regulation via sigma factors and the ECF sigma factors, in membrane secretion proteins, or the master regulators themselves, to further understand the regulation of secondary metabolite production in *B. subtilis*.

MALDI-TOF MS time course profiles of progenitor strains 3610 and PY79 showed that different nutrients affect the metabolic output, either overall or the timing of production. The data suggests that 3610 produces plipastatin after surfactin. This corroborates previous studies detecting surfactin production as early as 2 h and plipastatin at 22 h growth from 3610 on ISP2.[44] Surfactin has been exploited as a biosurfactant[49] and has stronger antibacterial properties,[50] whereas plipastatin inhibits filamentous fungi[51, 52] and together with subtilosin production indicates a shift in the secondary metabolism of *B. subtilis* towards sporulation.[44] The production of surfactin prior to plipastatin implies that the signal peptide ComX is expressed early in colony development, and activates the master regulator ComA, resulting in differentiation of

surfactin producer cells, which are present in a ratio of one surfactin producer cell to hundreds of non-surfactin producers.[7] Surfactin protects the *B. subtilis* colony through its antibacterial properties and enables colony expansion through swarming, while serving as a signaling molecule to certain cells within the colony to differentiate into matrix producers.[7, 28, 53] ComX was most likely not observed because it is produced in localized, undetectable quantities, under the limit of detection and resolution of the MALDI-TOF mass spectrometer.

The time course profile of PY79 suggested that Skf was produced prior to Sdp. These two cannibalistic factors are produced under nutrient limited conditions in Spo0A activated cells to provide nutrients via self-sacrifice, thereby delaying sporulation.[30] Skf has been shown to inhibit *B. subtilis* growth and mediate killing activity. However, Sdp, due to higher bioactivity, has been implicated as the toxin responsible for the cannibalistic activity of *B. subtilis* as well as bioactivity against other bacteria.[45] The lower killing ability and earlier production of Skf, compared to Sdp, may provide a means for *B. subtilis* to minimize its losses by initially sacrificing as few cells as necessary. PY79 also produced surfactin at Days 5-6 on LB, despite the frameshift mutation in *sfp* which affects the ability of PY79 to produce of surfactin and plipastatin.[40] This result supports the previously proposed existence of a lower efficiency phosphopantethenyl transferase that gets upregulated when spo0A is absent.[45]

IMS of the Spo0A mutants were expected to be deficient in matrix production, cannibalism, and sporulation. ALB1035A (Δ*spo0A* in 3610) maintained its capability to produce but not secrete surfactin. Surfactin remained colony-associated at the 48 h

time point, suggesting that signals that lead to or result from Spo0A have a role in surfactin secretion. Colony-associated surfactin may have led to a more transparent colony compared to 3610 and other 3610 mutant strains. The PY79 counterpart ALB964A verified that the deficiency in Spo0A caused elevated surfactin production as reported previously.[45] Surfactin was secreted, but ALB964A growth was impaired, alluding to a complex signaling network that drives *B. subtilis* colony growth and development.

PY79 mutants downstream of Spo0A activation expected to affect matrix production were ALB1045A (*sinI*), KP798 (*sinR*), and KP789 (*abrB*). Differences between ALB1045A and KP798 were expected because the *sinI* mutant represses *sinR* and induces matrix production, and the *sinR* mutant represses matrix production. KP789 represses matrix production and induces competence. The IMS profiles of KP798 and KP789 looked similar; both produced *m/z* 616, 655, and 886, polyglutamate at *m/z* 715, and Skf. ALB1045A did not produce polyglutamate or Skf. The IMS image for *m/z* 551 of ALB1045A may have been hot spots, a known artifact of IMS, and hence, was disregarded. These results implicate *m/z* 886 as a molecule related to matrix repression.

Sigma factor F is a sporulation-specific factor expressed in early forespore gene expression[33] after Spo0A activation and possibly after the matrix producers begin to shift to sporulation.[24] Hence, KP324 in PY79 (*spoIIAC*) may metabolically resemble the *spo0A*, *sinR*, or *abrB* mutants. KP324 resembled KP789 (*abrB*) and KP798 (*sinR*), both mutants of which repress matrix production. Since sporulation is a terminal pathway, there must be stringency to first attempt cannibalism to provide

nutrients to the colony and then even stricter stringency to ensure that the sporulation pathway is activated so colony DNA is preserved in spore form. KP324 also resembled XJ149 (*spoIIIJ*). This mutation in a constitutively expressed membrane protein translocase may affect the inter-cellular stress, which may in turn trigger Spo0A activation.

IMS of secretion mutants, such as insertional mutations of *yidC2* (ALB1033A in 3610; AR123 in PY79), *spoIIIJ* (ALB975 in 3610; XJ149 in PY79) resembled the corresponding progenitor strains. AR123 and XJ149 were in a PY79 background and did not produce detectable amounts of surfactin at 48 h. Surfactin secretion of ALB1033A and ALB975A was not affected. Hence, surfactin is secreted by different means and these two mutations did not affect the detected secondary metabolites from *B. subtilis*. AR123 looked like ALB1033A without surfactin; XJ149 looked like ALB975A without surfactin. The time course of AR123 on ISP2 resembled that of PY79 (Figure 2.5). Skf was detected from AR123 at 48 h via MALDI-TOF profiling, but not by IMS. This may have been due to differences in cell number to agar volume, instrument settings, or variance in microbiology or environment.

Sigma B is involved in general stress response.[33] The IMS of ALB1036A looked like the progenitor strain PY79, except that ALB1036A was deficient in Skf production and may have produced a low amount of surfactin in the center of the colony. Skf production may be linked to the general stress response in PY79 through Spo0A activation and subsequent expression of sigma factor B.

Sigma factor D induces motility via flagellar and swarming gene expression.[33] KP1032 in 3610 (*sigD*) had a similar metabolic profile to 3610, sans Skf production.

KP535 in 3610 (*spoVS*) increases sigma factor D activity. KP535 should differ from KP1032; the metabolic profile may mirror other strains that are swarming proficient. However, *B. subtilis* is not likely to exhibit a pronounced swarming phenotype on 2% agar.[53] IMS revealed that KP535 was capable of producing all representative ions. In one instance, KP535 produced all representative ions but surfactin; in the other, KP535 produced all but m/z 551 and Skf. KP1032, on the other hand, did not produce *m/z* 551, 886, or Skf. ALB1006A in PY79 (*sigD*) did not produce *m/z* 551 or surfactin, but did produce *m/z* 886 and Skf. Sigma factor D did not affect surfactin production or Skf production in 3610 and PY79, respectively. The common effect of sigma factor D deficiency was the lack of *m/z* 551; an increase in sigma D activity resulted in production of molecule at *m/z* 551 at 48 h on ISP2. *M/z* 551 may be affected by sigma factor D, or otherwise is affected by another sigma factor expressed at the same time as D.

Sigma H plays a role in the transition from exponential to stationary phase as well as competence and early sporulation genes.[33] KP427 (pSPAC-*spo0H□erm*) has similar metabolic output as progenitor PY79 but without Skf and with punctate surfactin in the center of the colony. The IPTG inducible promoter pSPAC[54] was not induced.

Sigma factor L does not affect mobility, competence, or sporulation.[55] The IMS data of ALB1044A (*sigL*) resembles the profiles of the ECF sigma factors. The ECF sigma factors M, W, and X (ALB1040A, ALB1039A, and ALB1037A, respectively) and possibly V, Y, Z, and YlaC (ALB1041A, ALB1043A, ALB1038A, and ALB1042A, respectively) have been implicated in exopolysaccharide

production,[35] a component of biofilm or matrix. These sigma factor mutants may resemble the *spo0A* mutant in PY79 because matrix producers are under the regulation of Spo0A activation. However, based on IMS, this was not the case. All seven of the ECF sigma factor mutants ALB1037A-ALB1043A had similar metabolic profiles. ALB1040A (*sigM*) did not produce surfactin and ALB1041 (*sigV*) did not produce surfactin or Skf. The IMS results supported the overlap of ECF sigma factors seen from transcriptomics studies.[35]

The utility of IMS as a screen was determined to be useful but not conducive to high-throughput, based on analysis of unidentified mutants of SCB844 in a PY79 background inducible in liquid or only on solid. *B. subtilis* metabolic output was fairly consistent between PY79 mutants. However, the strain with a unique pattern, such as ALB244, which produced surfactin in a PY79 background, can be prioritized for further experiments.

Based on MALDI-TOF IMS of *B. subtilis* mutant strains, it would be difficult to characterize ALB730 in terms of sigma factor mutants without assumptions. ALB730, deficient in sporulation, but proficient in swarming may be a *yidC2* mutant. IMS showed that ALB730 most closely resembles ALB607 and ALB729, two unidentified mutants of SCB844 only induced on solid media. Interestingly, ALB730 was swarming proficient, but did not produce detectable surfactin. Perhaps development or growth of the colony was delayed and the 48 h time point did not capture surfactin production.

As stated in Chapter 1, intensities within IMS images are relative. Thus, quantitative comparisons within one acquisition may be reasonable, whereas

comparisons across acquisitions may not be. In order to obtain more quantitative data or even identifications corresponding to *m/z*, orthogonal approaches are necessary.

2.5 Future Directions

This study profiled baseline metabolic output of *B. subtilis* strains using MALDI-TOF MS and IMS. As a follow up study, it would be interesting to test the hypothesis that the induction of a response from these strains will alter the secondary metabolite profile from the baseline over time. The strains available to induce are: the unidentified mutants in SCB844 inducible in liquid or only on solid media, KP427 (pSPAC-*spo0HΩerm*), and the sigma factor mutants.

The induction of the unidentified mutants may assist in characterization. The induction of KP427 (pSPAC-*spo0HΩerm*) may cause a swarming phenotype in the PY79 background, resulting in a secondary metabolite profile like 3610. Or induction may drive the colony towards sporulation due to Spo0A activation and deficient matrix production. Inducing a sigma factor response such as exposure to hydrogen peroxide (*ylaC*), cell wall inhibiting antibiotics (*sigW*, *sigX*, *sigM* and *sigV*), and cold (*sigL*) may result in similar end point secondary metabolic profiles, but differ in the timing of production and detection. In parallel, it would be interesting to conduct RNA-Seq analyses to see how well transcriptomics data correlates to secondary metabolite expression.

Other interesting projects include monitoring cellular differentiation via fluorescence. There are reporter strains for matrix producers, surfactin producers, motile cells, and sporulating cells.[7, 24] The pattern of surfactin production from ALB964A (*spo0A* in PY79) was unique. ALB1035A (*spo0A* in 3610), ALB244, ALB554, and ALB609 (unidentified mutants) produced surfactin, but it remained

colony associated. I hypothesize that these colonies are deficient in matrix producers, and hence do not produce Skf. By creating the same mutation as ALB1035A (spo0A) in the reporter strain for matrix producers and comparing with the wild type reporter strain, fluorescence microscopy can reveal whether or not these colonies are deficient in matrix production.

2.6 Summary

The time course profiling of metabolic output via MALDI-TOF MS is an efficient and reproducible approach to evaluate and determine the media selection and time point for optimal detection of *B. subtilis* metabolite production. This approach extends to other bacterial strains that are culturable on agar media, with the optimization of solvents, matrix selection, and MALDI sample preparation.

The global patterns of the metabolic output from 3610 and PY79 differed; 3610 made surfactin and plipastatin, and PY79 made Skf and Sdp. Over time, 3610 is capable of producing Skf and Sdp and PY79 is capable of producing surfactin. The MALDI-TOF MS and IMS results support previous studies of *B. subtilis* lipopeptides, the cannibalistic factors Skf and Sdp, and ECF sigma factors.

Transcriptional regulatory mutants affect the secondary metabolic profiles of *B. subtilis*. However, the differences detectable by MALDI-TOF MS and IMS were few. Overall, the metabolic output of the different *B. subtilis* mutants was similar to the progenitor strains, with some exceptions. This suggests that the regulation of the specialized metabolites is a complex web, not linearly controlled by one path. Instead, there are redundancies in sigma factor regulation,[27] regulatory factors, post-translational regulation of the sigma factors via anti-sigma factors, leading to crosstalk within a *B. subtilis* cell on a transcriptional level to drive multi-cellularity within the community.

The location of secondary metabolites within the colony was studied via IMS. Prior to investing time and effort in preparing IMS samples, acquiring data, and

analyzing data, evaluation of agar media type and time point to best study the system of interest is recommended. Orthogonal methods to validate *m/z* identities are necessary.

2.7 Experimental

2.7.1 MALDI-TOF MS profiling of *B. subtilis*

From a freshly streaked LB plate (bacterial glycerol stock streaked on LB and incubated overnight at 30˚C), cells were picked from one colony of *B. subtilis* 3610, PY79, and AR123 (Δ*yidC2*::*mls*) into 1 mL of 3 different liquid media – Luria Broth (LB), Difco Sporulation Medium (DSM), and yeast extract-malt extract (ISP2) – and incubated with shaking at 30˚C overnight. The OD_{600} was adjusted to 0.05-0.1 and the diluted culture was incubated at 30˚C for approximately 3 h, to ensure that the cells were in logarithmic phase. One μL was spotted in the center of a 3.5 mm Petri dish per media type in replicates of 5. This was repeated for 6 consecutive days. On Day 7, the cells of one colony were scraped off the agar and placed into 100 μL HPLC grade water (Fisher) and mixed well. One hour later, 1 μL of the cell suspension was mixed with 1 μL of matrix solution, 20 mg Universal MALDI matrix (Sigma) in 78% acetonitrile, 0.1 % trifluoroacetic acid. Then 1 μL was spotted onto the MALDI target plate and air-dried.

MALDI-TOF MS was conducted on a Microflex (Bruker) immediately after workup of all samples. The reflectron or linear positive method in FlexControl 2.0 was calibrated to at least 4 points using the Peptide Calibration Standard (Bruker) with a quadratic fit. Ions *m/z* 500-5000 were detected as an average of 300-500 laser shots per spectrum, sampled with random walk. Spectra were displayed as a heatmap in Gel/Stack view in ClinProTools version 2.0. The bands correspond to *m/z* peaks,

where the darkness of the band corresponds to the area under the curve or the intensity of the peak.

2.7.2 IMS profiling of *B. subtilis* mutants

Two mL ISP2 liquid media were inoculated with a scrape of frozen stock stored at -80˚C in 20% glycerol, and incubated overnight at 28˚C with shaking. A 1 μL aliquot was spotted on 10mL ISP2 agar per 10 cm diameter Petri dish. Once dry, the Petri dish was parafilmed and incubated at 30˚C for 48 h. Samples were prepared for IMS as previously described in Chapter 1.

MALDI-TOF IMS analysis was conducted on a Microflex (Bruker). The reflectron or linear positive method in FlexControl 2.0 was calibrated to at least 4 points using the Peptide Calibration Standard (Bruker) with a quadratic fit. In FlexImaging version 2.0, the x-y raster was defined between 600-800 μm. At each raster point, ions *m/z* 500-5000 were detected as an average of 300-500 laser shots per spectrum, sampled with random walk. Ions were false-colored and overlaid on a photograph as a 2-D image in FlexImaging. The intensity and distribution of the false-colored *m/z* ions correspond to the average intensity over the area imaged.

For a phenotypic evaluation of the *B. subtilis* mutants, *Streptomyces coelicolor* A3(2) spores were inoculated in a 1.5 cm line onto ISP2 agar and incubated overnight at 30˚C, as previously described.[44] Then 1 μL of *B. subtilis* mutant strain (OD_{600} ~0.1) was spotted adjacent to one tip of the line and incubated for 48 h at 30˚C. Photographs were taken at 48 h.

2.8 Acknowledgements

Dr. Anne Lamsa inoculated all of the colonies for the time course studies and provided her *B. subtilis* expertise. *B. subtilis* mutant strains and PY79 were generously gifted by Dr. Kit Pogliano at the University of California San Diego. *S. coelicolor* A3(2) was generously gifted by Michael Burkhart at the University of California San Diego. And *B. subtilis* NCIB3610 was generously gifted by Paul Straight at Texas A&M University.

2.9 Supplemental Figures

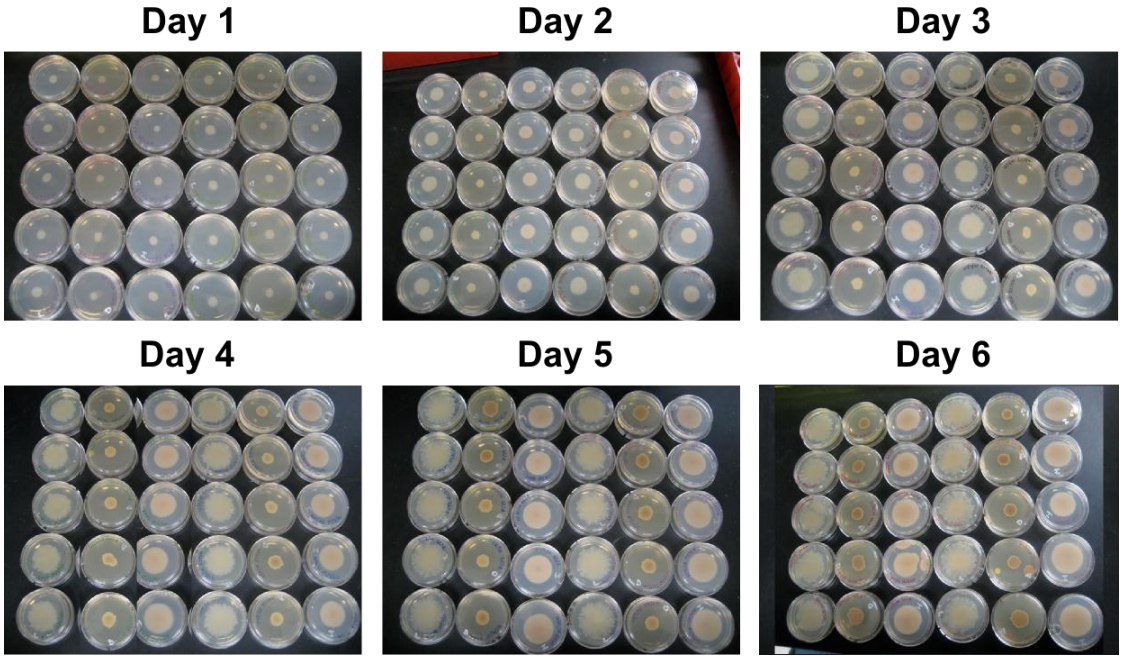


Figure S2.1 Development of colony morphology and color of *B. subtilis* PY79 and AR123 on LB, DSM, and ISP2 media over 6 days. The bottoms of the Petri dishes were photographed from each time point. The actual time of incubation is indicated in parentheses. From L to R for each time point (in replicates of 5): PY79 on LB, DSM, ISP2 and AR123 on LB, DSM, and ISP2. The photograph for Day 4 was corrected to maintain consistent order.

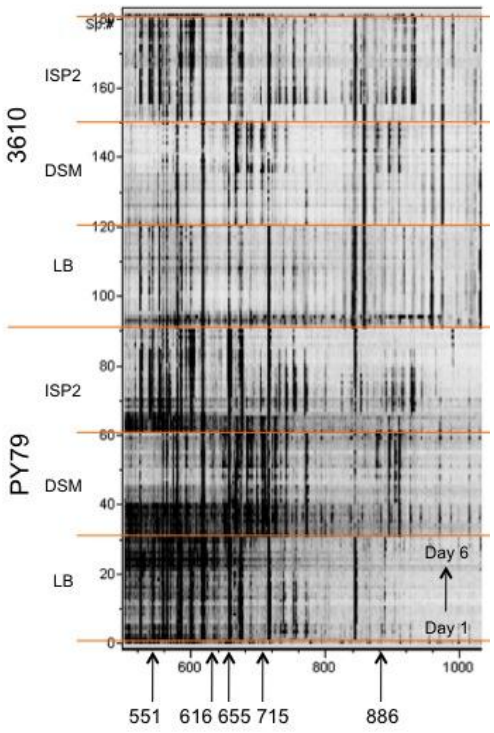


Figure S2.2 Zoomed (500-1000 Da) gel/stack view of 182 spectra reveals patterns of secondary metabolite production of *B. subtilis* 3610 and PY79 over time. Uncharacterized metabolites were detected at 551, 616, 655, and 886 Da. The metabolite at 715 Da corresponded to a partially characterized polyglutamate.[44]

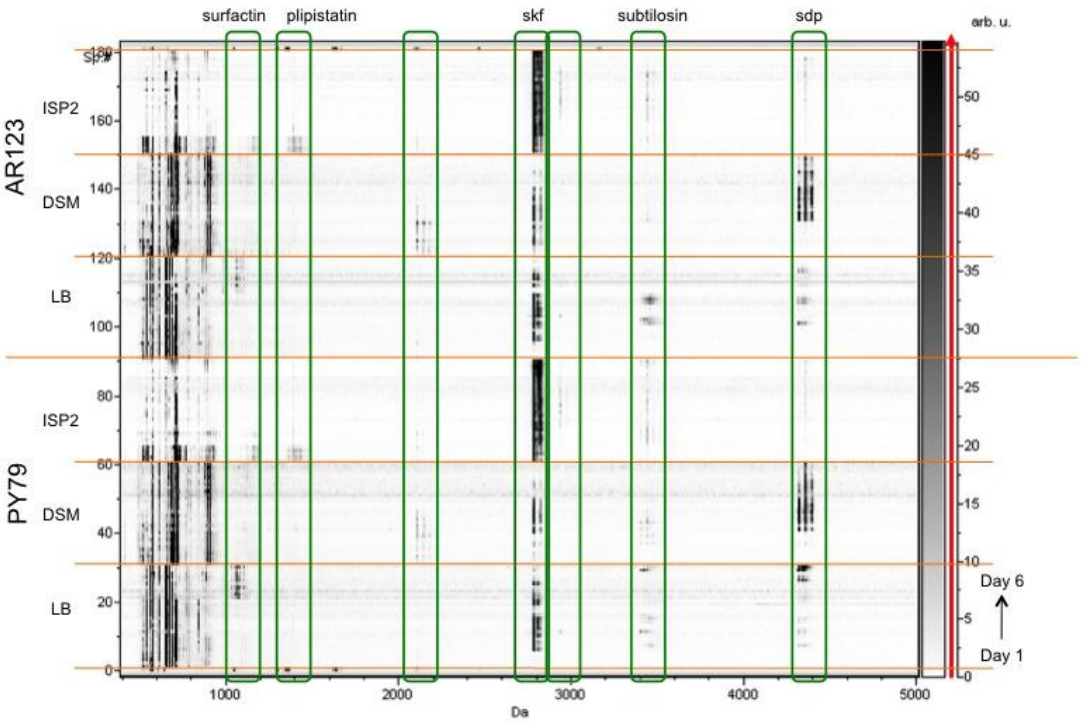


Figure S2.3 Gel/Stack view of 182 spectra reveals patterns of secondary metabolite production of *B. subtilis* AR123 and PY79 over time. The x-axis is *m/z* 500-5000 in Da, the left y-axis is spectrum number from 0-181, and the right y-axis is the relative intensity. The spectra on the very top and bottom are of the Peptide Calibration Standard. The top half is AR123, the bottom half is PY79, indicated by the longest horizontal orange line. These halves are split into three sections (shorter horizontal orange lines), corresponding to growth on three different media, ISP2, DSM, and LB (from top to bottom), from Day 1-6 (from bottom to top). Green rectangles highlight known and unknown molecules that are detected by MALDI-TOF MS.

2.9 References

1. Microbiology by numbers. *Nature reviews. Microbiology* 2011, 9, 628-628.

2. Schloss, P. D.; Handelsman, J., Toward a census of bacteria in soil. *Plos Computational Biology* 2006, 2, 786-793.

3. Whitman, W. B.; Coleman, D. C.; Wiebe, W. J., Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences of the United States of America* 1998, 95, 6578-6583.

4. Hulcr, J.; Latimer, A. M.; Henley, J. B.; Rountree, N. R.; Fierer, N.; Lucky, A.; Lowman, M. D.; Dunn, R. R., A Jungle in There: Bacteria in Belly Buttons are Highly Diverse, but Predictable. *Plos One* 2012, 7.

5. Kallus, S. J.; Brandt, L. J., The Intestinal Microbiota and Obesity. *Journal of Clinical Gastroenterology* 2012, 46, 16-24.

6. Schoenian, I.; Spiteller, M.; Ghaste, M.; Wirth, R.; Herz, H.; Spiteller, D., Chemical basis of the synergism and antagonism in microbial communities in the nests of leaf-cutting ants. *Proceedings of the National Academy of Sciences of the United States of America* 2011, 108, 1955-1960.

7. Lopez, D.; Kolter, R., Extracellular signals that define distinct and coexisting cell fates in Bacillus subtilis. *Fems Microbiology Reviews* 2010, 34, 134-149.

8. Lowery, C. A.; Dickerson, T. J.; Janda, K. D., Interspecies and interkingdom communication mediated by bacterial quorum sensing. *Chemical Society Reviews* 2008, 37, 1337-1346.

9. Antunes, L. C. M.; Ferreira, R. B. R.; Buckner, M. M. C.; Finlay, B. B., Quorum sensing in bacterial virulence. *Microbiology-Sgm* 2010, 156, 2271-2282.

10. Paik, H.; Lee, K.; Kim, J.; Jun, S.; Kim, W., Optimization for culture medium and fermentation process of a Bacillus probiotics. *Abstracts of the General Meeting of the American Society for Microbiology* 2001, 101, 549-550.

11. Spinosa, M. R.; Braccini, T.; Ricca, E.; De Felice, M.; Morelli, L.; Pozzi, G.; Oggioni, M. R., On the fate of ingested Bacillus spores. *Research in Microbiology* 2000, 151, 361-368.

12. Kloepper, J. W.; Ryu, C. M.; Zhang, S. A., Induced systemic resistance and promotion of plant growth by Bacillus spp. *Phytopathology* 2004, 94, 1259-1266.

13. Schallmey, M.; Singh, A.; Ward, O. P., Developments in the use of Bacillus species for industrial production. *Canadian Journal of Microbiology* 2004, 50, 1-17.

14. Goes, A. P.; Sheppard, J. D., Effect of surfactants on alpha-amylase production in a solid substrate fermentation process. *Journal of Chemical Technology and Biotechnology* 1999, 74, 709-712.

15. Sen, R., Surfactin: Biosynthesis, Genetics and Potential Applications. In *Biosurfactants*, Sen, R., Ed. Springer-Verlag Berlin: Berlin, 2010; Vol. 672, pp 316-323.

16. Banat, I. M.; Makkar, R. S.; Cameotra, S. S., Potential commercial applications of microbial surfactants. *Applied Microbiology and Biotechnology* 2000, 53, 495-508.

17. Hasan, F.; Shah, A. A.; Hameed, A., Industrial applications of microbial lipases. *Enzyme and Microbial Technology* 2006, 39, 235-251.

18. Yeh, M. S.; Wei, Y. H.; Chang, J. S., Enhanced production of surfactin from Bacillus subtilis by addition of solid carriers. *Biotechnology Progress* 2005, 21, 1329-1334.

19. Hong, H. A.; Khaneja, R.; Tam, N. M. K.; Cazzato, A.; Tan, S.; Urdaci, M.; Brisson, A.; Gasbarrini, A.; Barnes, I.; Cutting, S. M., Bacillus subtilis isolated from the human gastrointestinal tract. *Research in Microbiology* 2009, 160, 134-143.

20. Hoa, N. T.; Baccigalupi, L.; Huxham, A.; Smertenko, A.; Van, P. H.; Ammendola, S.; Ricca, E.; Cutting, S. M., Characterization of Bacillus species used for oral bacteriotherapy and bacterioprophylaxis of gastrointestinal disorders. *Applied and Environmental Microbiology* 2000, 66, 5241-5247.

21. Tam, N. K. M.; Uyen, N. Q.; Hong, H. A.; Duc, L. H.; Hoa, T. T.; Serra, C. R.; Henriques, A. O.; Cutting, S. M., The intestinal life cycle of Bacillus subtilis and close relatives. *Journal of Bacteriology* 2006, 188, 2692-2700.

22. Hong, H. A.; To, E.; Fakhry, S.; Baccigalupi, L.; Ricca, E.; Cutting, S. M., Defining the natural habitat of Bacillus spore-formers. *Research in Microbiology* 2009, 160, 375-379.

23. Aguilar, C.; Vlamakis, H.; Losick, R.; Kolter, R., Thinking about Bacillus subtilis as a multicellular organism. *Current Opinion in Microbiology* 2007, 10, 638-643.

24. Vlamakis, H.; Aguilar, C.; Losick, R.; Kolter, R., Control of cell fate by the formation of an architecturally complex bacterial community. *Genes & Development* 2008, 22, 945-953.

25. Lopez, D.; Vlamakis, H.; Losick, R.; Kolter, R., Paracrine signaling in a bacterium. *Genes & Development* 2009, 23, 1631-1638.

26. Burbulys, D.; Trach, K. A.; Hoch, J. A., INITIATION OF SPORULATION IN BACILLUS-SUBTILIS IS CONTROLLED BY A MULTICOMPONENT PHOSPHORELAY. *Cell* 1991, 64, 545-552.

27. Hamoen, L. W.; Venema, G.; Kuipers, O. P., Controlling competence in Bacillus subtilis: shared use of regulators. *Microbiology-Sgm* 2003, 149, 9-17.

28. Kearns, D. B.; Losick, R., Swarming motility in undomesticated Bacillus subtilis. *Molecular Microbiology* 2003, 49, 581-590.

29. Gonzalez-Pastor, J. E.; Hobbs, E. C.; Losick, R., Cannibalism by sporulating bacteria. *Science* 2003, 301, 510-513.

30. Gonzalez-Pastor, J. E., Cannibalism: a social behavior in sporulating Bacillus subtilis. *Fems Microbiology Reviews* 2011, 35, 415-424.

31. Nandy, S. K.; Bapat, P. M.; Venkatesh, K. V., Sporulating bacteria prefers predation to cannibalism in mixed cultures. *Febs Letters* 2007, 581, 151-156.

32. Helmann, J. D., The extracytoplasmic function (ECF) sigma factors. *Advances in Microbial Physiology, Vol 46* 2002, 46, 47-110.

33. Haldenwang, W. G., THE SIGMA-FACTORS OF BACILLUS-SUBTILIS. *Microbiological Reviews* 1995, 59, 1-30.

34. SubtiWiki. http://www.subtiwiki.uni-goettingen.de/wiki/index.php/Main_Page.

35. Luo, Y.; Asai, K.; Sadaie, Y.; Helmann, J. D., Transcriptomic and Phenotypic Characterization of a Bacillus subtilis Strain without Extracytoplasmic Function sigma Factors. *Journal of Bacteriology* 2010, 192, 5736-5745.

36. Earl, A. M.; Losick, R.; Kolter, R., Ecology and genomics of Bacillus subtilis. *Trends in Microbiology* 2008, 16, 269-275.

37. Zeigler, D. R.; Pragai, Z.; Rodriguez, S.; Chevreux, B.; Muffler, A.; Albert, T.; Bai, R.; Wyss, M.; Perkins, J. B., The Origins of 168, W23, and Other Bacillus subtilis Legacy Strains. *Journal of Bacteriology* 2008, 190, 6983-6995.

38. Branda, S. S.; Gonzalez-Pastor, J. E.; Ben-Yehuda, S.; Losick, R.; Kolter, R., Fruiting body formation by Bacillus subtilis. *Proceedings of the National Academy of Sciences of the United States of America* 2001, 98, 11621-11626.

39. Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A. M.; Alloni, G.; Azevedo, V.; Bertero, M. G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Borriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S. C.; Bron, S.; Brouillet, S.; Bruschi, C. V.; Caldwell, B.; Capuano, V.; Carter, N. M.; Choi, S. K.; Codani, J. J.; Connerton, I. F.;

Cummings, N. J.; Daniel, R. A.; Denizot, F.; Devine, K. M.; Dusterhoft, A.; Ehrlich, S. D.; Emmerson, P. T.; Entian, K. D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S. Y.; Glaser, P.; Goffeau, A.; Golightly, E. J.; Grandi, G.; Guiseppi, G.; Guy, B. J.; Haga, K.; Haiech, J.; Harwood, C. R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M. F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; KlaerrBlanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S. M.; Levine, A.; Liu, H.; Masuda, S.; Mauel, C.; Medigue, C.; Medina, N.; Mellado, R. P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; Oreilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S. H.; Parro, V.; Pohl, T. M.; Portetelle, D.; Porwollik, S.; Prescott, A. M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlan, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S. J.; Serror, P.; Shin, B. S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H. F.; Zumstein, E.; Yoshikawa, H.; Danchin, A., The complete genome sequence of the Gram-positive bacterium Bacillus subtilis. *Nature* 1997, 390, 249-256.

40. Earl, A. M.; Eppinger, M.; Fricke, W. F.; Rosovitz, M. J.; Rasko, D. A.; Daugherty, S.; Losick, R.; Kolter, R.; Ravel, J., Whole-Genome Sequences of Bacillus subtilis and Close Relatives. *Journal of Bacteriology* 2012, 194, 2378-2379.

41. Quadri, L. E. N.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T., Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* 1998, 37, 1585-1595.

42. Straight, P. D.; Fischbach, M. A.; Walsh, C. T.; Rudner, D. Z.; Kolter, R., A singular enzymatic megacomplex from Bacillus subtilis. *Proceedings of the National Academy of Sciences of the United States of America* 2007, 104, 305-310.

43. Volpon, L.; Besson, F.; Lancelin, J. M., NMR structure of antibiotics plipastatins A and B from Bacillus subtilis inhibitors of phospholipase A(2). *Febs Letters* 2000, 485, 76-80.

44. Yang, Y.-L.; Xu, Y.; Straight, P.; Dorrestein, P. C., Translating metabolic exchange with imaging mass spectrometry. *Nature Chemical Biology* 2009, 5, 885-887.

45. Liu, W.-T.; Yang, Y.-L.; Xu, Y.; Lamsa, A.; Haste, N. M.; Yang, J. Y.; Ng, J.; Gonzalez, D.; Ellermeier, C. D.; Straight, P. D.; Pevzner, P. A.; Pogliano, J.; Nizet, V.;

Pogliano, K.; Dorrestein, P. C., Imaging mass spectrometry of intraspecies metabolic exchange revealed the cannibalistic factors of Bacillus subtilis. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107, 16286-16290.

46. Phelan, V. V.; Liu, W.-T.; Pogliano, K.; Dorrestein, P. C., Microbial metabolic exchange-the chemotype-to-phenotype link. *Nature Chemical Biology* 2012, 8, 26-35.

47. Monteiro, S. M.; Clemente, J. J.; Henriques, A. O.; Gomes, R. J.; Carrondo, M. J.; Cunha, A. E., A procedure for high-yield spore production by Bacillus subtilis. *Biotechnology Progress* 2005, 21, 1026-1031.

48. Chiba, S.; Lamsa, A.; Pogliano, K., A ribosome-nascent chain sensor of membrane protein biogenesis in Bacillus subtilis. *Embo Journal* 2009, 28, 3461-3475.

49. Altmajer Vaz, D.; Gudina, E. J.; Jurado Alameda, E.; Teixeira, J. A.; Rodrigues, L. R., Performance of a biosurfactant produced by a Bacillus subtilis strain isolated from crude oil samples as compared to commercial chemical surfactants. *Colloids and Surfaces B-Biointerfaces* 2012, 89, 167-174.

50. Gonzalez, D. J.; Haste, N. M.; Hollands, A.; Fleming, T. C.; Hamby, M.; Pogliano, K.; Nizet, V.; Dorrestein, P. C., Microbial competition between Bacillus subtilis and Staphylococcus aureus monitored by imaging mass spectrometry. *Microbiology-Sgm* 2011, 157, 2485-2492.

51. Vanittanakom, N.; Loeffler, W.; Koch, U.; Jung, G., FENGYCIN - A NOVEL ANTIFUNGAL LIPOPEPTIDE ANTIBIOTIC PRODUCED BY BACILLUS-SUBTILIS F-29-3. *Journal of Antibiotics* 1986, 39, 888-901.

52. Honma, M.; Tanaka, K.; Konno, K.; Tsuge, K.; Okuno, T.; Hashimoto, M., Termination of the structural confusion between plipastatin A1 and fengycin IX. *Bioorganic & Medicinal Chemistry* 2012, 20, 3793-3798.

53. Kearns, D. B., A field guide to bacterial swarming motility. *Nature Reviews Microbiology* 2010, 8, 634-644.

54. Youngman, P.; Perkins, J. B.; Losick, R., CONSTRUCTION OF A CLONING SITE NEAR ONE END OF TN917 INTO WHICH FOREIGN DNA MAY BE INSERTED WITHOUT AFFECTING TRANSPOSITION IN BACILLUS-SUBTILIS OR EXPRESSION OF THE TRANSPOSON-BORNE ERM GENE. *Plasmid* 1984, 12, 1-9.

55. Debarbouille, M.; Martinverstraete, I.; Kunst, F.; Rapoport, G., THE BACILLUS-SUBTILIS-SIGL GENE ENCODES AN EQUIVALENT OF SIGMA-54 FROM GRAM-NEGATIVE BACTERIA. *Proceedings of the National Academy of Sciences of the United States of America* 1991, 88, 9092-9096.

CHAPTER III

Molecular networking as a dereplication strategy

3.1 Abstract

A major goal in natural product discovery programs is to rapidly dereplicate known entities from complex biological extracts. We demonstrate here that molecular networking, an approach that organizes MS/MS data based on chemical similarity, is a powerful complement to traditional dereplication strategies. Successful dereplication with molecular networks requires MS/MS spectra of the natural product mixture along with MS/MS spectra of known standards, synthetic compounds, or well-characterized organisms, preferably organized into robust databases. This approach can accommodate different ionization platforms, enabling cross correlations of MS/MS data from ambient ionization, direct infusion, and LC-based methods. Molecular networking not only dereplicates known molecules from complex mixtures, it also captures related analogs, a challenge for many other dereplication strategies. To illustrate its utility as a dereplication tool, we apply mass spectrometry-based molecular networking to a diverse array of marine and terrestrial microbial samples, illustrating the dereplication of 58 molecules including analogs.

3.2 Introduction

Natural products (NPs) are produced by various life forms as specialized metabolites that control cellular processes and drive biology.[1] They are also some of the most prolific sources of therapeutics.[2] Despite the overwhelming number of NPs and NP derivatives approved by the FDA, the majority of the pharmaceutical industry has eliminated NP-based drug discovery programs,[3] in part due to the costs behind high rates of rediscovery in the late stages of the isolation process. However, academic scientists continue to explore the natural world for molecules with interesting chemistries, for they are the drivers of biology and are important to biotechnology, biofuels, agriculture, and medicine.[1]

Dereplication, or the identification of known molecules, early in the NP workflow minimizes time, effort, and cost.[3] Current dereplication strategies include hyphenated techniques, such as HPLC-MS, HPLC-NMR, HPLC-NMR-MS, and HPLC-SPE-NMR,[4-6] or bioactivity fingerprints, such as cytological profiling or BioMAP.[7,8] The common denominator of these strategies is the attempt to exploit the fact that structurally similar or identical molecules share similar physical characteristics, such as UV-vis profiles, chromatographic retention times, MS, NMR chemical shifts, or biological properties. However, even when multiple characteristics are experimentally verified, a full elucidation either occurs late in the workflow or otherwise remains elusive. Hence, alternative and orthogonal dereplication strategies need to be developed.

Mass spectrometry based dereplication is critical to modern natural product dereplication pipelines.[9-13] MS is more sensitive than NMR, however, it is often difficult to reliably dereplicate a given compound when solely using precursor or parent masses because of the sheer number of results returned when searching databases such as AntiBase,[14] MarinLit,[15] Beilstein Dictionary of Natural Products,[16] and Scifinder. However, MS/MS fragmentation has yet to be routinely exploited in NP dereplication. In principle, MS/MS data are definitive characteristics of a molecule, and thus could be effectively used in dereplication. The underlying assumption of MS together with MS/MS-based dereplication is that structural architecture, chemical stability, and functional groups combine to dictate reactivity by collision-induced dissociation in the gas phase, and therefore, similarities in MS/MS fragmentation patterns can be used as proxies for chemical similarity. Thus, these MS/MS patterns are more discriminatory than parent mass alone. Because natural product workflows typically incorporate LC-MS analysis, and MS/MS data can be simultaneously acquired on most mass spectrometers, introduction of molecular networking does not require significant adjustment of the current NP discovery process. Here we present molecular networking as a complement to current dereplication strategies.

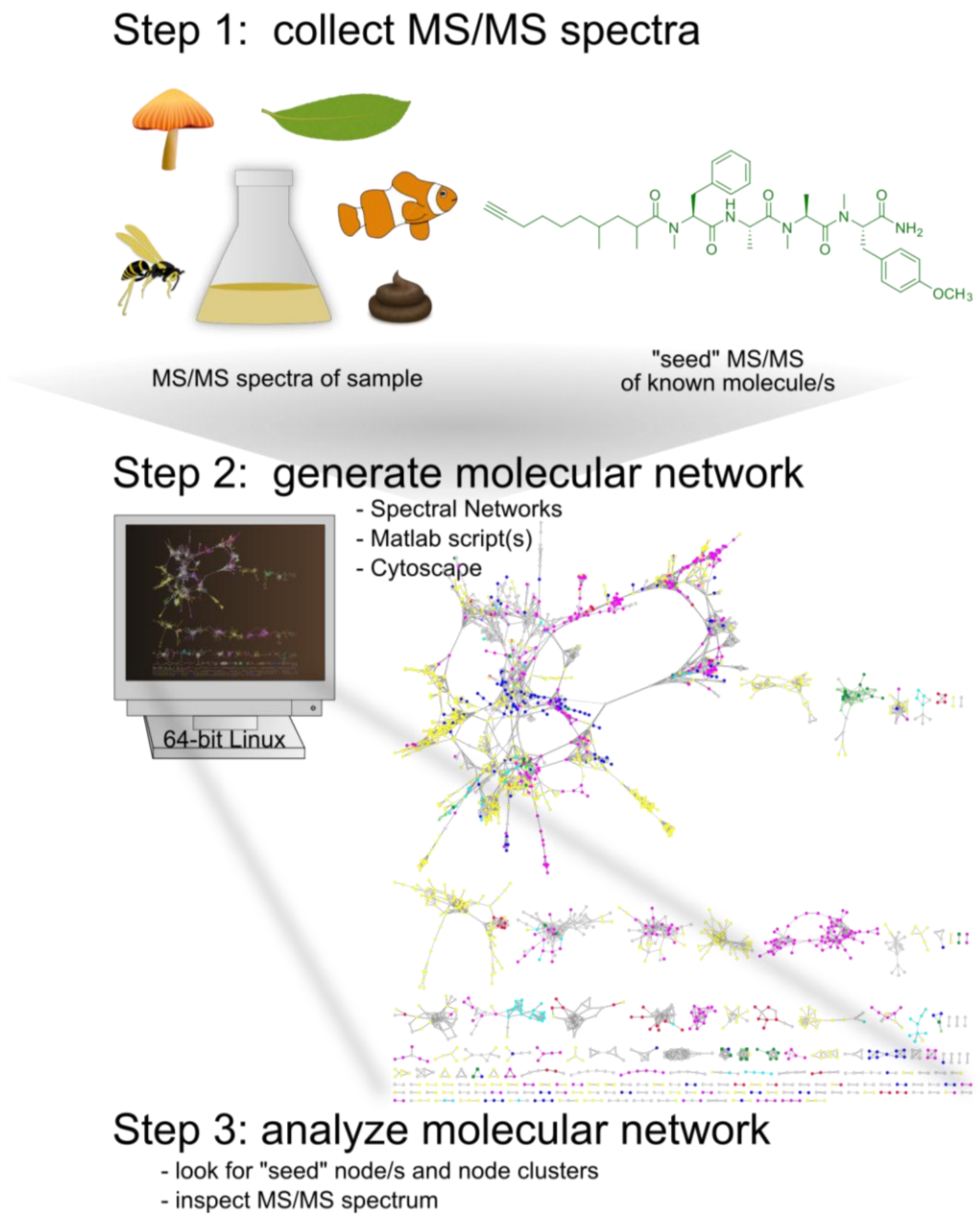


Figure 3.1 Three steps to implement molecular networking for dereplication. Experimental MS/MS spectra of samples, which can be of a high degree of complexity and heterogeneity, and MS/MS spectra of known molecules, dubbed "seed" spectra, are analyzed on a 64-bit Linux system. The resultant network is visualized in Cytoscape where one node represents one consensus MS/MS spectrum and is labeled with the precursor mass, and an edge represents relatedness, where edge thickness indicates cosine similarity.

Mass spectrometry-based molecular networking relies on the observation that structurally similar molecules share similar MS/MS fragmentation patterns. Molecular networking is implemented in three fundamental steps (Figure 1). First, MS/MS spectra are collected. Second, a molecular network is generated using ‘cosine scores’ which measure relatedness in MS/MS spectra and can be visualized using Cytoscape, a tool designed to visualize correlations of large datasets.[17,18,32,33] Finally, the molecular network is analyzed as described in Figure 2. Thus, a molecular network is a visual representation of molecular relatedness (chemical similarity) of any given set of compounds.

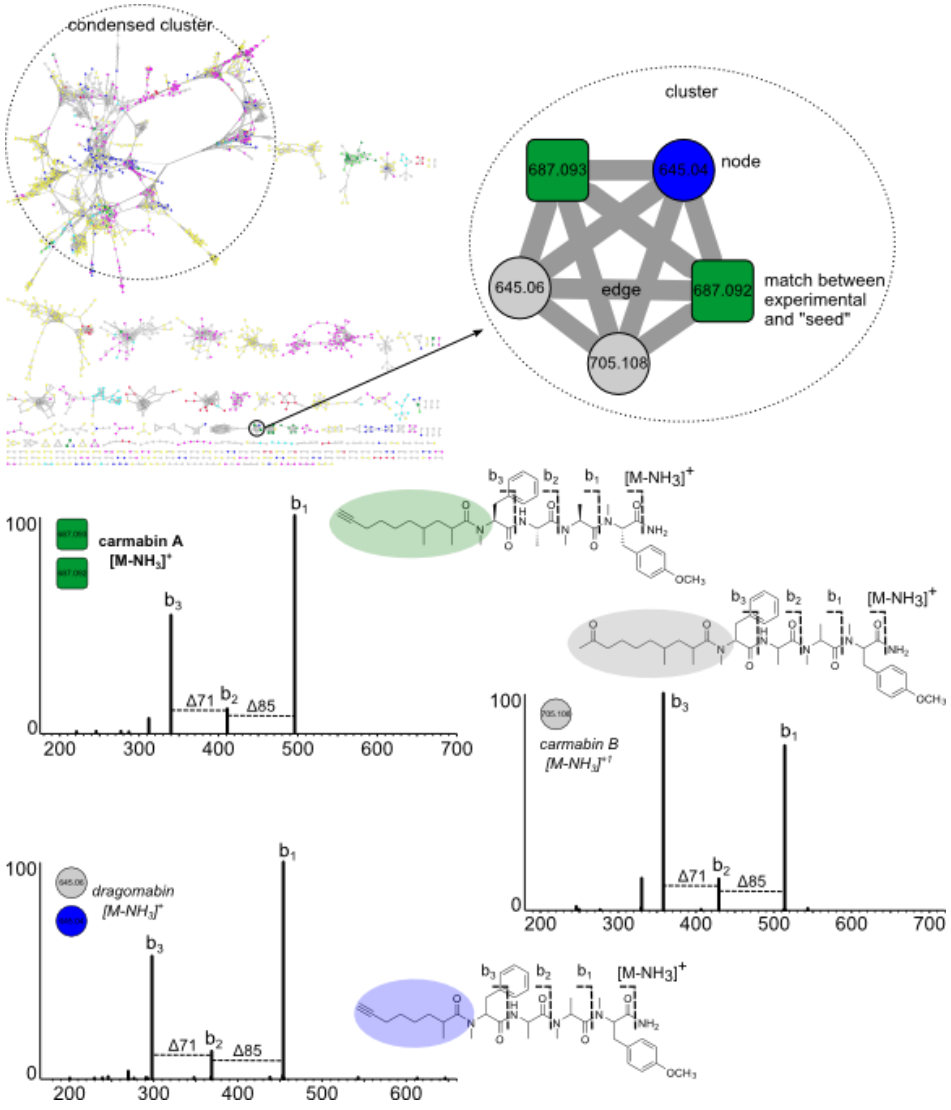


Figure 3.2 Validation of nodes and annotation of MS/MS spectra. Using the carmabin A "seed" as a starting node, molecular networking revealed that carmabin A was present in two different cyanobacterial collections (also in Figure 3). In the same cluster, there were three additional nodes of two precursor masses, *m/z* 645 and 705. Putative assignment of modifications based on the precursor mass difference – sodium, potassium, alkylated, halogenated, oxidized, etc. – and mass differences between peaks in the MS/MS spectra also provide insight to the structures of the related NPs. Manual inspection of the carmabin A MS/MS spectrum showed a sequence tag consistent with *b*-type ions. Inspection of the MS/MS spectra of the connected nodes, *m/z* 645 and 705, matched to known lipopeptides in the same molecular family as carmabin A, but with differing lengths and degrees of unsaturation of the lipophilic tail. With one seed, carmabin A, molecular networking was able to dereplicate carmabin A and two MS/MS spectra that match to the structurally related analogs, carmabin B and dragomabin.

Because structurally similar NPs share similar MS/MS fragmentation patterns, molecular families tend to cluster together within a network.[18] These networks then allow for the simultaneous visual exploration of identical molecules, analogs, or compound families, within single or multiple datasets and from a wide variety of biological sources.[17, 18] Within the network, one node corresponds to one consensus MS/MS spectrum, a mathematical merging of MS/MS spectra with nearly identical precursor mass and peak patterns, and is typically labeled with the precursor mass. Edges (lines) connect nodes with related consensus MS/MS spectra. In the networks depicted in this manuscript, we have added node and edge attributes so that the color of the node corresponds to the origin of the sample and the thickness of the edge reflects the similarity as defined by the cosine score.

In the current study, we generated two networks of MS/MS data from phylogenetically distinct sample sets to demonstrate how molecular networking can be used to dereplicate known compounds and identify close analogs. One network contained LC-MS/MS spectra of fractionated extracts and purified NPs from cyanobacterial collections. The other network combined MS/MS spectra from direct analysis of microbial colonies using ambient ionization nanoDESI[18] and direct infusion nanoESI of purified NPs and crude extracts of well-studied bacteria. The MS/MS data of purified NPs act as "seed" spectra, which serve as the initial focal points in the network when processed with the MS/MS spectra from a mixture of unknown compounds. Implementation of molecular networking to six cyanobacterial collections and eight bacterial strains resulted in the concurrent dereplication of four

NPs and 38 possible analogs, and eight NPs and eight analogs, respectively, encompassing NPs of different structural classes, such as polyketides, alkaloids, and peptides. The identification of known molecules is most rapid when standards are included, but the utility of molecular networking does not depend on inclusion of a database. Networking identifies potential analogs and guides the user to molecules that may be unknown, thereby decreasing the time to the decision as to pursue or not pursue the isolation of individual metabolites.

3.3 Experimental section

3.3.1 Bacteria

Bacteria (*Pseudomonas aeruginosa* PAO1 and PA14, *Serratia marcescens* sp. ES129, *Verrucosispora* sp. MS100128 and SRM7) were sampled directly from colonies grown on agar media via nanoDESI using a modified Prosolia Omnispray DESI source coupled to a 6.42 T Thermo LTQ-FT-ICR mass spectrometer.[18,19] Additional bacteria (*Bacillus subtilis* 3610 and PY79 and *Dietzia* sp. FI-1026) were extracted then directly infused into the mass spectrometer using a Triversa nanomate-electrospray ionization source (Advion Biosystems, Ithaca, NY) coupled to a 6.42 T Thermo LTQ-FT-ICR mass spectrometer. Strain-specific growth conditions are described in the Supporting Information. For nanomate, samples were diluted in 50:50 MeOH:H_2O and then directly infused using a back pressure of 0.35-0.5 psi and a spray voltage of 1.3-1.45 kV. FT-MS and ion trap MS/MS spectra were acquired using Tune Plus software version 1.0 and Xcalibur software version 1.4 SR1. The instrument was tuned on *m/z* 816, the 15+ charge state of cytochrome C. The instrument scan cycle consisted of one 10 min segment, during which a profile FT scan with a resolution of 25,000 was cycled with four data-dependent scans in the ion trap. The data-dependent scan iteratively cycled through the top four most intense ions from the FT scan, after which they were placed on an exclusion list for 600 s. Purified compounds hydroxyphenazine, methoxyphenazine, *N*-methylphenazine, PQS (pseudomonas quinolone signal), pyocyanin, abyssomicins B, C, D, H, J, and L, proximicin B, and commercially available surfactin from *B. subtilis* (Sigma S3523), were also directly

infused into the mass spectrometer. MS/MS data were collected in a data dependent manner, during which profile mode FT-MS scans cycled with five MS/MS scans in the ion trap. The five most abundant peaks were fragmented and then added to an exclusion list. Data were acquired for 10 min.

3.3.2 Cyanobacteria

Five cyanobacterial collections (*Moorea bouilloni* PNG05-198, GenBank collection number FJ041298), containing apratoxins A and B as well as lyngbyabellin A) and a mixture of *Lyngbya* sp. and *Schizothrix* sp. (containing tumonoic acid I) were collected from Papua New Guinea, while a red filamentous cyanobacterium (containing barbamide), a red *Moorea* sp. (containing carmabin A), and a brown *Schizothrix* (containing carmaphycin B) were collected from Panamá. Specific collection information is described in the Supporting Information. Each of the five cyanobacterial collections was extracted with 2:1 CH_2Cl_2/MeOH, fractionated into nine sub-fractions using silica gel Vacuum Liquid Chromatography (hexanes, EtOAc, MeOH), and analyzed via LC-MS/MS on a Thermo Finnigan Surveyor Autosampler-Plus/LC-Pump-Plus/PDA-Plus system coupled to a Thermo Finnigan LCQ Advantage Max mass spectrometer fitted with a Phenomenex Kinetex C-18 100Å 100 x 60 mm column. Using a flow rate of 700 μL/min and a 45 min gradient of CH_3CN and 0.1% HCO_2H acidified H_2O, the elution began at 50% CH_3CN to H_2O, held there for 5 min, then a linear gradient to 100% CH_3CN at 30 min, held there until 35 min, and then a linear gradient back to 50% CH_3CN to H_2O at 40 min which was maintained until the end of the run. The divert valve was set to waste for the first 2.5 min. ESI conditions

set with the capillary temperature at 325°C, source voltage at 5 kV, and a sheath gas flow rate of 69 L/min. Four scan events: positive MS, window from *m/z* 300-2000; then three data dependent MS/MS scans of the first, second, and third most intense ions from the first scan event. MS/MS settings of 35% normalized collision energy, default charge of 1, minimum intensity of 10^5 counts, isolation width of *m/z* 2, dynamic exclusion count of 5, repeat duration of 1 min, exclusion list size of 25, and an exclusion duration of 3 min.

In alternative preparation, crude cyanobacterial extracts were first separated by vacuum liquid chromatography to create nine fractions of distinct polarity, which were then analyzed by LC-MS/MS to generate the molecular network. Using the same instrument and the same column with a flow rate of 700 μL/min and a 39 min gradient of CH_3CN and 0.1% HCO_2H acidified H_2O, the elution began at 40% CH_3CN to H_2O, held there for 5 min, then a linear gradient reaching 100% CH_3CN at 25 min, held there until 35 min, and then a linear gradient back to 40% CN_3CN to H_2O at 36 min which was maintained until the end of the run. Three scan events: first scan negative MS, window from *m/z* 80-2000; second scan positive MS, window from *m/z* 100-2000; and third scan MS/MS data dependent most intense from second scan, using the same ESI and MS/MS settings but without dynamic exclusion. Characterized NPs (carmaphycins A and B, barbamide, carmabin A, tumonoic acid I), either purified or synthesized, and an injection consisting only of solvent, were also analyzed via LC-MS/MS. The retention times and MS/MS spectra for the standards were manually

verified against the MarinLit database[15] and previously published data, respectively.[20-24]

3.3.3 Molecular networking

The data was converted to mzXML format, a text-based format used to represent mass spectrometry data describing the scan number, precursor *m/z*, and the *m/z* and intensity of each ion observed in MS/MS, using ReadW from Thermo or msconvert, part of the ProteoWizard package.[25] Molecular networks were generated as previously described.[18] The bacterial network also included MS/MS data of ions below *m/z* 2000 from HMDB,[26] LipidMaps,[27] MassBank,[28] Metlin,[29] and NIST[30] databases in mgf format, a tab delimited text-based representation of the mass spectrometry data. Once all the data was in a text format (mzXML or mgf), the data was subjected to Spectral Networks, which includes MS-Clustering,[31] followed by generating text files with attributes using MATLAB.

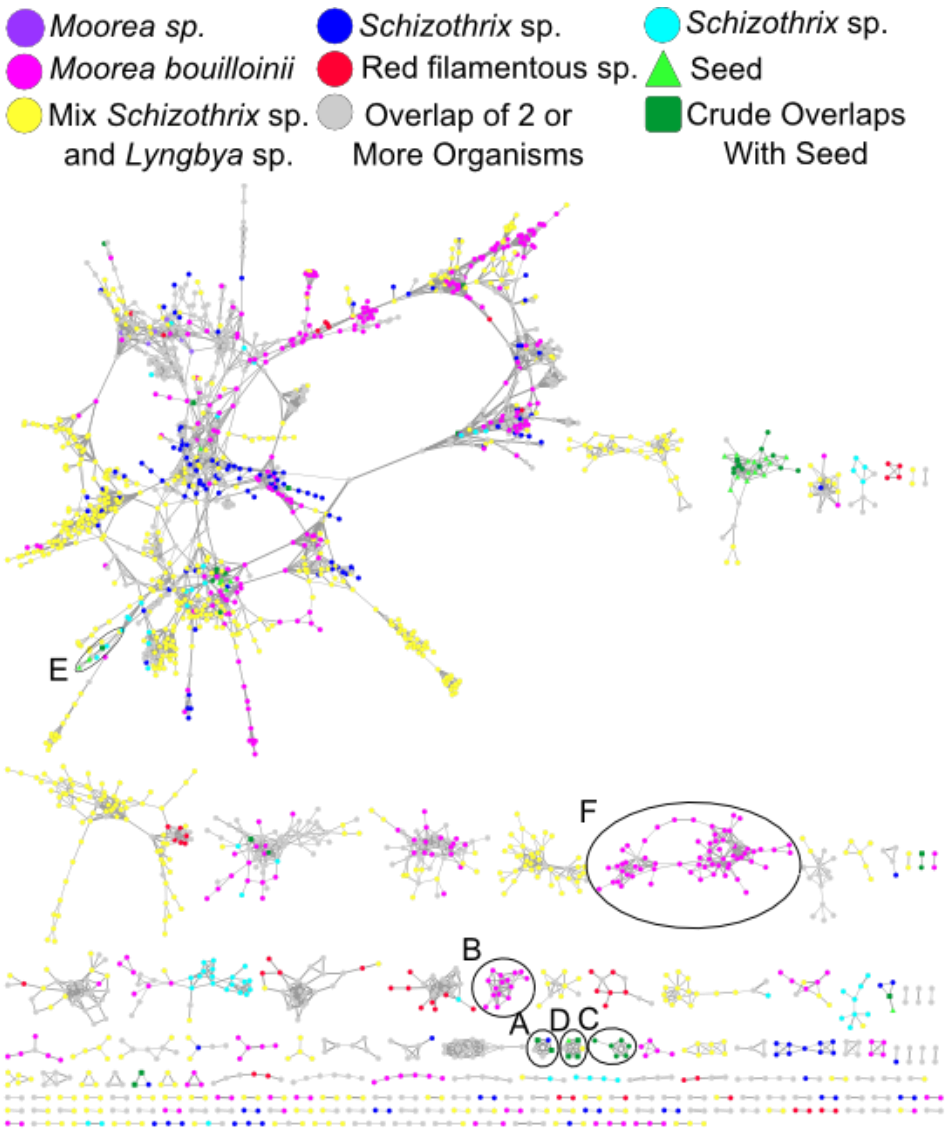
The MS-Clustering algorithm[31] first combined identical spectra into consensus spectra, and then Spectral Networks compared all possible pairs of consensus MS/MS spectra and assigned a vector-based cosine similarity score (dot product) to each pair that ranges from 0 to 1, where 1 represents identical spectra. The higher the cosine score between two spectra, the more similar the MS/MS spectra, and by extension, the more similar the corresponding molecules. Subsequent MATLAB scripts reported data with cosine scores of and above the user-defined cosine threshold, herein chosen to be 0.65, which were then imported into Cytoscape and displayed as a network of nodes and edges.[32] To simplify the network, the background nodes from solvent and the

database nodes that were not directly connected to a sample node via an edge were removed from the network. Remaining nodes within the network were organized with the FM3 layout plug-in,[33] node colors were mapped based on the source files of the MS/MS, and the edge thickness attribute was defined to reflect cosine similarity scores with thicker lines indicating higher similarity. Sub-networks were generated in Cytoscape from isolated portions of the larger network in order to improve visibility of node connectivity.

3.4 Results

The cyanobacterial network contained MS/MS data from six different marine collections (Figure 3) and five pure NPs which had been previously characterized by NMR, UV-vis, HRMS and comparison to MarinLit and/or chromatographic retention time.[15] The bacterial network (Figure 4) was generated from direct infusion of extracts or direct sampling via nanoDESI[18] of eight bacterial strains from three phyla (Firmicutes, Actinobacteria, and Proteobacteria), 13 NP standard samples, and 104,228 MS/MS spectra from five metabolomics databases. However, the actual number of unique compounds in these databases is much less as there are multiple spectra for many compounds. In the networks, each cyanobacterial collection or bacterial strain is represented by a different color; overlapping experimental nodes with data from more than one organism are grey circles, nodes that are a consensus of experimental and seed spectra are dark green squares, and seed nodes that do not overlap with experimental data are light green triangles. The dereplicated nodes were manually inspected and MS/MS fragmentation patterns were annotated to verify that the fragmentation patterns were related (Supporting Information).

Figure 3.3 Cyanobacterial network with a cosine similarity score cut off of 0.65. This network was generated when crude extracts of cyanobacteria (six distinct collections) were separated into nine fractions based on polarity and then analyzed via LC-MS/MS with five known compounds. The inlaid portions of the network were rearranged in Cytoscape for easier visualization of node connectivity. Italicized labels indicate putative assignments. **A.** Standard "seed" spectra of carmabin A overlapped with three different cyanobacterial collections (bold outlined dark green squares). Additionally carmabin B and dragomabin were identified from the same collection and two different genera, respectively. **B.** A cluster of nodes from a collection of *Moorea bouillonii*, a known producer of the lyngbyabellins, was highly suggestive of lyngbyabellin A and potential novel analogs, including a monochlorinated species. The presence of M+2 nodes within this cluster provided further evidence for these analogs. **C.** Tumonoic acid I clustered closely with novel analogs present in both the crude cyanobacterial sample and the pure seed compound. **D.** Barbamide and the recently described 4-*O*-demethylbarbamide analog were present in multiple cyanobacterial samples. Additionally, a dimethyl analog was present in two samples. **E.** Carmaphycins A and B were included in the network, but only carmaphycin B was in one *Schizothrix* sp. collection. Also, the precursor ion of a putative non-oxygenated methionine analog clustered closely, as well as an analog with a precursor mass of *m/z* 511. **F.** A cluster of nodes from *Moorea bouillonii*, a known producer of the apratoxins, putatively contained five known and 13 novel apratoxin analogs.


Moorea sp.
Moorea bouilloinii
Mix Schizothrix sp. and Lyngbya sp.
Schizothrix sp.
Red filamentous sp.
Overlap of 2 or More Organisms
Schizothrix sp.
Seed
Crude Overlaps With Seed
E
F
B
A
D
C

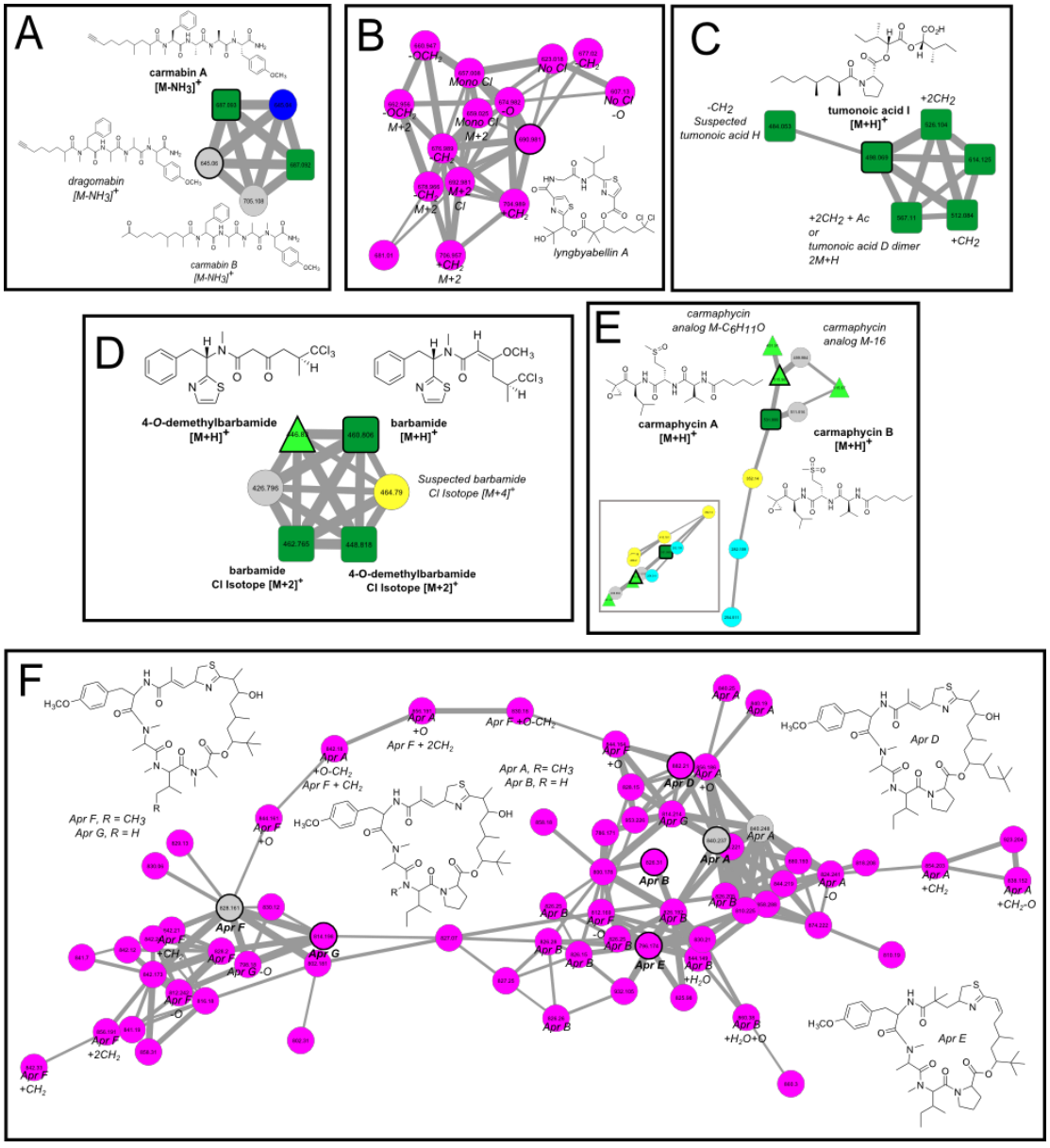
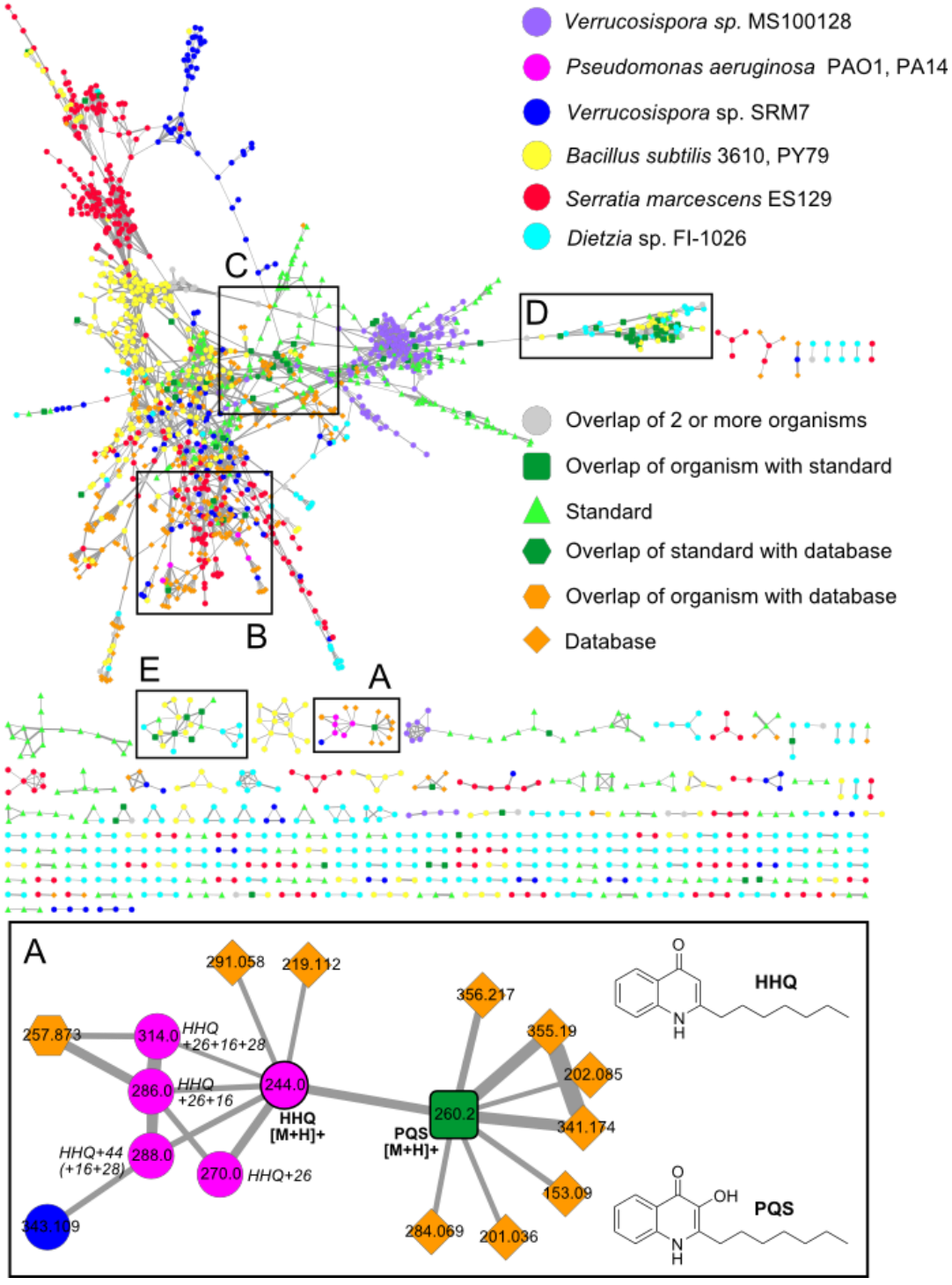


Figure 3.3 Cyanobacterial network, continued

Figure 3.4 Bacterial network with a cosine similarity score cut off of 0.65. This network was generated from direct infusion of extracts or direct nanoDESI-MS/MS sampling of 8 bacteria, 14 known compounds, and 5 databases. The inlaid portions of the network were rearranged in Cytoscape for easier visualization of node connectivity and annotation. Italicized labels indicate putative assignments. **A.** 4-Hydroxy-2-heptylquinoline (HHQ) and *Pseudomonas* quinolone signal (PQS) fragment similarly. The structures of the connected database nodes do contain conjugated cyclic moieties, but are not of the same molecular family as HHQ and PQS. **B.** Small molecules <350 *m/z* in the condensed cluster. Commercial *Pseudomonas aeruginosa* molecules pyocyanin and methoxyphenazine overlap with sample spectra, whereas *N*-methylphenazine does not. Phenazine fragments differently. *S. marcescens* sp. ES129 yields a node with a precursor mass that matches to prodigiosin, which is supported by comparison of the MS/MS sample spectra to published spectra. **C.** MS/MS spectra of abyssomicins (Aby) B, C, D, H, and L purified from *Verrucosispora* sp. MS100128 and proximicin B isolated from sp. SRM7 were incorporated into the bacterial network. Aby B and H "seed" spectra overlapped with data directly sampled from *Verrucosispora* sp. VM37 (bold outlined dark green squares). Aby C, D, and L were incorporated into the network, but do not match to sample data (bold outline light green triangles). Aby J and proximicin B were not incorporated into the network. **D.** The surfactins (Srf) 1+ cluster includes protonated (H), sodiated (Na), and potassiated (K) forms of surfactin $C_{13\text{-}15}$. Protonated precursor nodes are labeled with bold solid outlines. This cluster reveals nodes for precursor mass *m/z* 994.6, 14 less than Srf-C_{13}, suggesting the presence of a C_{12} surfactin and was instrumental in dereplication of the bioactive molecule/s from *Dietzia* sp. FI-1026 (cyan). **E.** Putatively assigned surfactins 2+ cluster.


Verrucosispora sp. MS100128
Pseudomonas aeruginosa PAO1, PA14
Verrucosispora sp. SRM7
Bacillus subtilis 3610, PY79
Serratia marcescens ES129
Dietzia sp. FI-1026
Overlap of 2 or more organisms
Overlap of organism with standard
Standard
Overlap of standard with database
Overlap of organism with database
Database
A
B
C
D
E
A
291.058
219.112
257.873
314.0
HHQ +26+16+28
286.0
HHQ +26+16
244.0
HHQ [M+H]+
HHQ+44 (+16+28)
288.0
270.0
HHQ+26
243.109
356.217
355.19
202.085
260.2
PQS [M+H]+
341.174
153.09
284.069
201.036
HHQ
PQS
OH
N
H
O

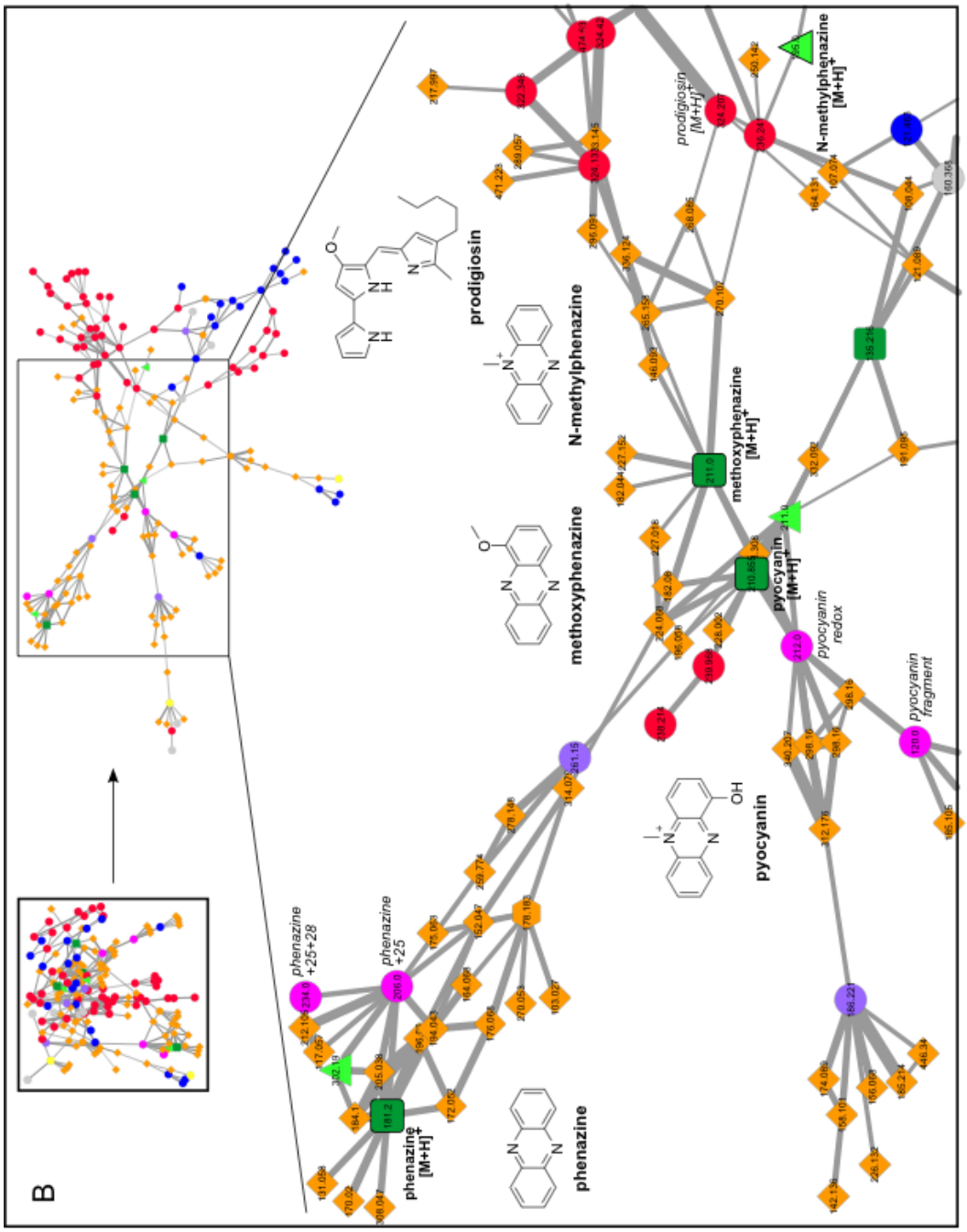


Figure 3.4 Bacterial network, continued

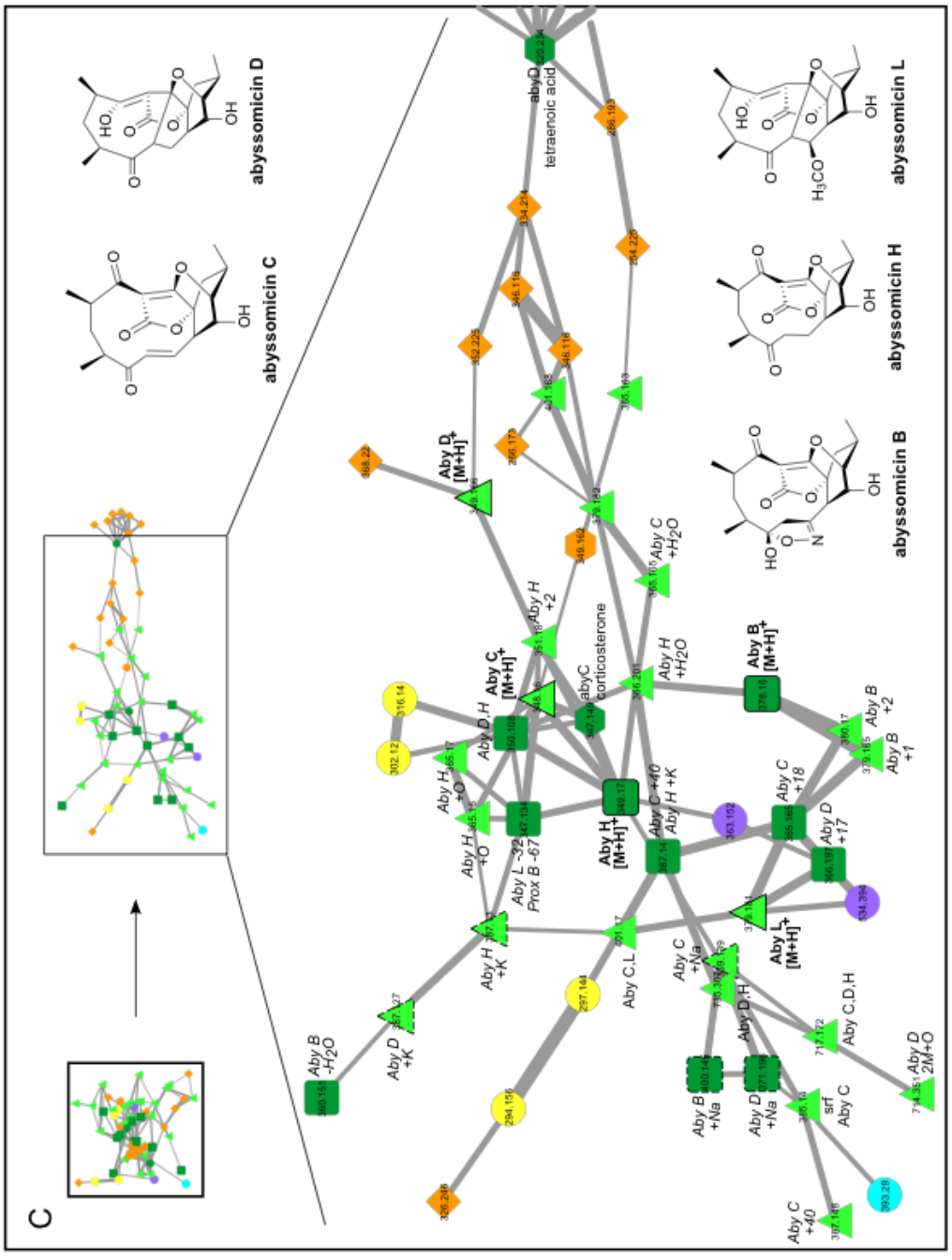


Figure 3.4 Bacterial network, continued

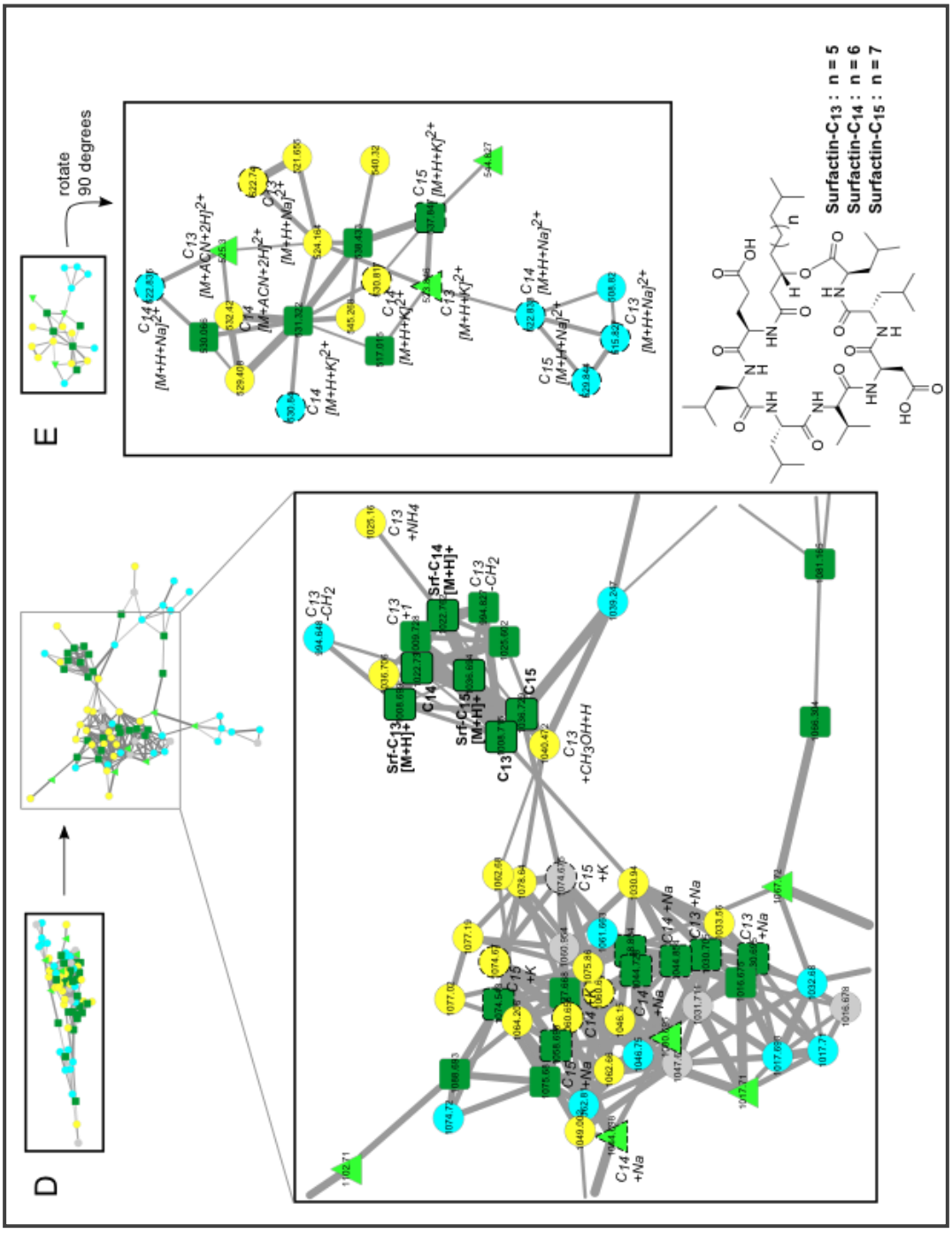


Figure 3.4 Bacterial network, continued

Molecular networking of the cyanobacterial data resulted in dereplication of carmabin A,[21] tumonoic acid I,[20] barbamide,[23] and carmaphycin B;[24] interestingly, carmaphycin A was not present in the network. Carmabin A was present in crude fractions of a *Moorea* sp. extract (Figure 2) and the cluster of nodes around carmabin A included precursor masses and corresponding MS/MS spectra (Figure S2, Supporting Information) that agreed with analogs carmabin B and dragomabin.[34]

A cluster of six dark green square nodes revealed that a mixed collection of two cyanobacteria, *Lyngbya* sp. and *Schizothrix* sp. from Papua New Guinea, contained tumonoic acid I (Figure 3C).[20] This cluster suggested the presence of five analogs within the seed LC-MS/MS run for tumonoic acid I, which was supported by MS/MS spectra and extracted ion chromatograms (EICs) (Figure S4, Supporting Information).

From the network, it was determined that barbamide and its characteristic complex chlorine isotope pattern were present in multiple collections (Figures 3D and S5, Supporting Information).[23] In addition to this "seed," the recently described 4-*O*-demethylbarbamide along with its chlorine isotope pattern were also present,[22] as well as a putative analog of barbamide. The MS/MS fragmentation pattern for 4-*O*-demethylbarbamide contained an intact *y* ion, thus supporting its assignment. The MS/MS for the putative barbamide analog contained *b* and *y* ions consistent with dechlorobarbamide.[35]

Carmaphycin B and two nodes suggestive of novel carmaphycin analogs were dereplicated from the cyanobacterial network. Carmaphycin B was present in the

crude sample from one collection from Panama, confirmed by matching MS/MS fragmentation patterns and retention times. The two novel carmaphycin analogs, one consistent with a non-oxidized methionine analog and a second with an *m/z* 511, were supported by the presence of *b* and *y* ions in the MS/MS spectra (Figure S6, Supporting Information). These predictions are being pursued with a full structure elucidation of each following isolation and NMR analysis.

The cyanobacterial network also contained two clusters from *Moorea bouillonii* that contain precursor masses that correspond to known lyngbyabellins and apratoxins. *M. bouillonii* is well known to produce both lyngbyabellins and apratoxins, and these clusters could be identified based on MS/MS fragmentation, precursor masses, and isotopic patterns despite the lack of seed spectra for dereplication. The lyngbyabellin A[36] cluster had masses which suggested novel analogs including a rare monochlorinated species, a dehydroxylated species, a demethylated species, a demethylated and dehydroxylated species, and a methylated species. Additionally, the precursor masses for these analogs had distinct isotopic patterns characteristic of chlorinated species, thus providing orthogonal evidence for the presence of lyngbyabellin analogs in this cyanobacterial sample. The other cluster contained 67 nodes and likely contained apratoxins A and B[37,38] as well as precursor masses suggestive of the analogs apratoxins D,[39] E,[40] F, and G[41] (Figure S7, Supporting Information). However, the molecular formulas of these compounds could not be definitively assigned due to high ppm error values and the lack of standard MS/MS spectra.

In the bacterial and database network, molecular networking enabled the dereplication of polycyclic polyketide-type antibiotics termed abyssomicins,[42-44] phenazines,[45] quinolones,[46] and surfactin lipopeptides.[47] MS/MS spectra of two of the NMR-pure abyssomicin standards formed consensus nodes from data directly collected from *Verrucosispora* sp. MS100128, thereby dereplicating abyssomicins B and H as well as revealed abyssomicin H (+OH) as previously described[44] (Figures 4C and S11, Supporting Information). Nodes for abyssomicins C, D, and L were also present in the network, but these nodes originated from standards and were not present in the sample. Similarily, the node corresponding to proximicin B, previously determined to be produced by *Verrucosispora* sp. SRM7, was present in the network but did not match or connect to any experimental SRM7 node.[48]

Molecular networking analysis of data-independent nanoDESI-MS/MS of *Pseudomonas aeruginosa* PAO1 and PA14[19] with five reference spectra – hydroxyphenazine, methoxyphenazine, *N*-methylphenazine, *Pseudomonas* quinolone signal (PQS), and pyocyanin – resulted in dereplication of methoxyphenazine,[19] pyocyanin, PQS, a reduced pyocyanin analog, phenazine, and 2-heptyl-4-quinolone (HHQ). Pyocyanin (*m/z* 211 $[M+H]^+$ calc.) was immediately linked to a putative analog 2 amu higher, corresponding to a reduced analog. PQS exhibited a perfect match between experimental and reference data, resulting in an overlapping node. Although when manually inspected, methoxyphenazine and *N*-methylphenazine appeared to fragment differently from phenazine (Figure S10, Supporting

Information), their nodes were still found in the same region of the network, indicating an underlying fragmentation similarity between these compounds (Figure 4B).

Molecular networking dereplicated prodigiosin from *Serratia marcescens* sp. Environmental Strain 129 (Lab Environmental Strain Collection), without a "seed" spectrum or a full characterization of the organism. *S. marcescens* sp. ES129 was isolated from soil, produced a red pigment, and was sampled directly via nanoDESI.[18] *S. marcescens* strains are known to produce the red-pigmented antibiotic prodigiosin (323.1997 calc.), a tripyrrole alkaloid, which has a published MS/MS spectrum. Comparison of the sample MS/MS spectrum to the published verified that the fragmentation patterns matched. Thus, *m/z* 324 from *S. marcescens* sp. ES129 matches to prodigiosin.

Nodes from *Dietzia* sp. FI-1026 extracts clustered with nodes from *B. subtilis* and surfactin MS/MS datasets in two clusters that were assigned to singly charged and doubly charged surfactin (Figure 4E and F, respectively). The consensus nodes (dark green squares) in both clusters contained MS/MS spectra from all three sources, and contained protonated as well as sodium and potassium adducts of surfactin $C_{13\text{-}15}$. The cluster for singly charged surfactin (Figure 4E) revealed nodes with a precursor mass *m/z* 994.6, a difference of 14 amu less than surfactin-C_{13}, suggesting the presence of a surfactin-C_{12}.

3.5 Discussion

Mass spectrometry-based molecular networking is a useful tool when applied to cyanobacterial and bacterial MS/MS datasets. In the current study it was used to dereplicate 58 molecules including analogs. Three advantages of molecular networking as a dereplication strategy are highlighted here. First, it allows for the simultaneous identification of known NPs and analogs from complex mixtures. Second, molecular networking is compatible with any mass spectrometry ionization platform. And lastly, molecular networking is easily incorporated into standard NP discovery workflows.

Molecular networking based dereplication can be used to match identical as well as related molecules, an asset that was quite apparent with the cyanobacterial examples. The carmabin A consensus node clustered tightly with 4 other nodes (Figures 2 and 3A). The carmabin A node was found in one *Moorea* sp. sample. This node strongly clustered with *m/z* 705.1 and 645.0, found in the same cyanobacterial sample; *m/z* 705.1 was also found in a collection of *Schizothrix* sp. After examination of the MS/MS spectra and comparison to the seed NP, these compounds were dereplicated as carmabin B and dragomabin, respectively. Thus, one identical match in the network resulted in the dereplication of two additional NPs.

In the case of carmaphycin, the network contained a node for carmaphycin B from a collection of *Schizothrix* sp. when both A and B were used as seeds. Moreover, there were nodes for previously uncharacterized carmaphycin analogs of *m/z* 499.9 and 511.0. Inspection of the MS/MS spectra revealed an intact *b* ion for both analogs

suggesting the valine and lipid chain were intact. For *m/z* 499.0, the mass difference and corresponding *y* ion suggested the presence of a non-oxidized methionine residue, while the modification for *m/z* 511.0 is uncertain and would require a full structure elucidation via NMR to identify the location and modified moiety.

Tumonoic acid I was dereplicated from a mixed *Schizothrix* sp. and *Lyngbya* sp. collection by retention time comparison with a verified standard in addition to examination of the MS/MS spectra. Novel tumonoic acid analogs were present in the collection as well as the seed sample. EICs for the seed suggested that the analogs were present in low titer, as their UV signal/s were not detected in the chromatogram. This result highlights the exceptional sensitivity of mass spectrometry and molecular networking which thus allows for the identification of even very low abundance NPs.

Any mass spectrometry ionization platform is compatible with molecular networking at multiple stages of the workflow (Figure 5), as long as MS/MS spectra are acquired. Advances in ionization, including ambient ionization, methods in mass spectrometry remove the time and labor for sample work-up, allowing for molecular networking to be implemented at early-stages of natural product discovery and can be used for networking. For example, a variety of ionization sources, such as ESI, nanoDESI, and direct infusion nanoESI, were used to generate the MS/MS spectra that were networked together in the current study. The utility of molecular networking extends to low molecular weight molecules, such as less than *m/z* 400 as exemplified by the phenazines and quinolones produced by *Pseudomonas aeruginosa* (Fig. 4A and B). These NPs have been implicated in virulence and competitive fitness.[49]

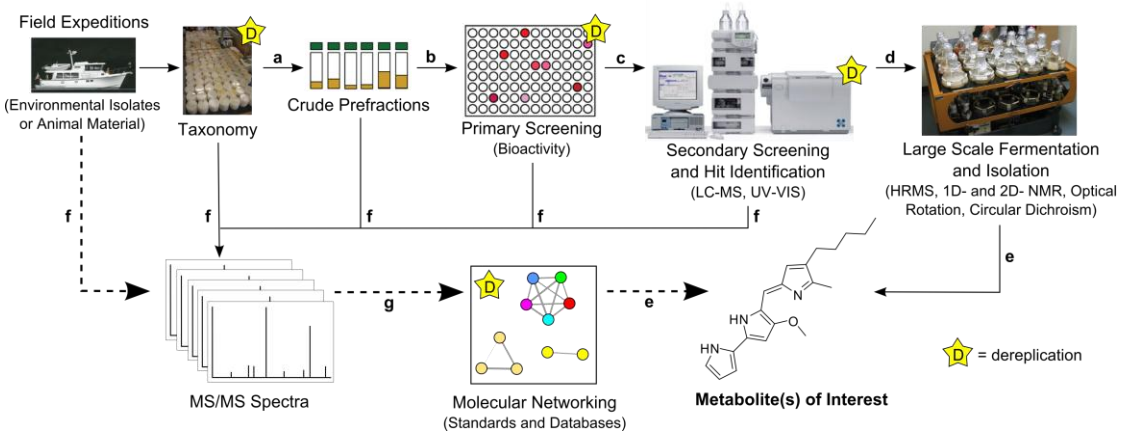


Figure 3.5 Molecular networking can be easily implemented in traditional natural product workflows.

Molecular networking was used to dereplicate abyssomicins B and H, two members of a newer class of polyketide antibiotics. There was no node for abyssomicin J; the proposed structure of abyssomicin J consists of a dimer of abyssomicin K connected by a thioether. Abyssomicins C, D, and L nodes were present as seed nodes only. Abyssomicin C is one of a handful of new potential antibiotics with activity against gram-positive bacteria such as methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *S. aureus* (VRSA) through inhibition of p-aminobenzoate (*p*ABA) biosynthesis.[42] Thus, identification of analogs via molecular networking may lead to more candidate antibiotics.

It is important to note that the incorporation of seed spectra is not necessary for molecular networking to be highly useful. For example, in the cyanobacterial network, there were no seed spectra for the lyngbyabellins or apratoxins, yet these two well-known compound families formed two distinct clusters and were easily identified. Detection of new analogs in these compound families will be helpful for developing an understanding of structure-activity relationships (SAR) in the future.

Molecular networking can be easily integrated into existing NP workflows to provide useful guidance in the discovery of new bioactive substances. To highlight one example, fractionated extracts of *Dietzia* sp. FI-1026 isolated from fish intestines[50] were previously screened against 15 bacterial strains via BioMAP analysis.[8] The BioMAP profile clustered to valinomycin and cytological profiling using HeLa cells showed cell death (Supporting Information). The MS/MS spectra of these extracts were included in the bacterial network and clustered with surfactin and *B. subtilis*

(Figure 4D and E). The presence of FI-1026 nodes in both the singly charged and doubly charged surfactin clusters increased confidence that the bioactive component of the FI-1026 samples was surfactin. Because this is a well-studied molecule, it was likely introduced as a contaminant during the preparation of the sample. Hence, the fractions from this particular experiment were eliminated from further analysis.

At a cosine score cutoff of 0.65, the bacterial and database network (Figure 4) contained 1948 total nodes, of which 296 (15.2%) were database MS/MS nodes or overlapping nodes with a database MS/MS spectrum. Most of these database nodes were less than *m/z* 400 and were located in the condensed cluster, the portion of the network concentrated with nodes and edges. A majority of these nodes 276 (14.2%) were nodes with MS/MS spectra from databases only, three were nodes (0.15%) for consensus between a database and standard, and 17 (0.87%) were nodes for consensus between a database and a bacterial sample. Of the 17 consensus nodes of interest, none were able to be dereplicated, highlighting the need for an integrated use of relevant MS/MS libraries, databases, and de novo structure elucidations.

Developing relevant MS/MS libraries for microorganisms will significantly benefit NP drug discovery efforts by improving the results of the dereplication strategy presented herein. In the next few years, we aim to create an open access database of MS/MS spectra of known molecules, with data repository and search functionality for public use. Once data are collected and deposited, there will be less need for others to repeat data collection for that sample, saving time and resources. As this collection grows, the database will become ever more useful, and may transform

first pass natural product analysis in a fashion similar to that of gene annotations in Genbank.[51,52]

There are many conceivable creative applications of molecular networking, such as taxomonic dereplication of mixed samples in parallel to metagenomics studies,[53] inclusion of geneology to successfully dereplicate molecules, SAR and biosynthetic pathway prediction studies, and the creation of an all-encompassing database for metabolomics studies of host-pathogen systems as diagnostic tools for human health.

In conclusion, molecular MS/MS networking is an alternative approach that can be used for dereplication. The workflow can be incorporated at any stage of a natural product discovery program. Additionally, it is compatible with workflows currently in use in many natural product laboratories that use MS to characterize their samples simply by acquiring MS/MS at the same time. This will be useful to capture a wide range of structural classes and analogues even when the retention times are very different or if different ionization platforms are used; the only requirement is to obtain MS/MS. Molecular networking should find utility in efficiently dereplicating mixtures and pure compounds alike, especially for finding related analogs in our efforts to gain insight into the functional roles of these fascinating specialized metabolites.

3.6 References

1. Davies, J.; Ryan, K. S., Introducing the Parvome: Bioactive Compounds in the Microbial World. *ACS Chem. Biol.* **2012,** 7, (2), 252-259.

2. Newman, D. J.; Cragg, G. M., Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *J. Nat. Prod.* **2012,** 75, (3), 311-335.

3. Gerwick, W. H.; Moore, B. S., Lessons from the Past and Charting the Future of Marine Natural Products Drug Discovery and Chemical Biology. *Chem. Biol.* **2012,** 19, (1), 85-98.

4. Johansen, K. T.; Wubshet, S. G.; Nyberg, N. T.; Jaroszewski, J. W., From Retrospective Assessment to Prospective Decisions in Natural Product Isolation: HPLC-SPE-NMR Analysis of Carthamus oxyacantha. *J. Nat. Prod.* **2011,** 74, (11), 2454-2461.

5. Mansson, M.; Phipps, R. K.; Gram, L.; Munro, M. H. G.; Larsen, T. O.; Nielsen, K. F., Explorative Solid-Phase Extraction (E-SPE) for Accelerated Microbial Natural Product Discovery, Dereplication, and Purification. *J. Nat.Prod.* **2010,** 73, (6), 1126-1132.

6. Wubshet, S. G.; Johansen, K. T.; Nyberg, N. T.; Jaroszewski, J. W., Direct C-13 NMR Detection in HPLC Hyphenation Mode: Analysis of Ganoderma lucidum Terpenoids. *J. Nat.Prod.* **2012,** 75, (5), 876-882.

7. Perlman, Z. E.; Slack, M. D.; Feng, Y.; Mitchison, T. J.; Wu, L. F.; Altschuler, S. J., Multidimensional drug profiling by automated microscopy. *Science* **2004,** 306, (5699), 1194-8.

8. Wong, W. R.; Oliver, A. G.; Linington, R. G., Development of Antibiotic Activity Profile Screening for the Classification and Discovery of Natural Product Antibiotics. *Chem. Biol.* **2012,** 19, (11), 1483-1495.

9. Gu, Z.-M.; Wang, L.-Q.; Wu, J., Mass defect filter - a new tool to expedite screening and dereplication of natural products and generate natural product profiles. *Nat. Prod. J.* **2011,** 1, (2), 135-145.

10. Hou, Y.; Braun, D. R.; Michel, C. R.; Klassen, J. L.; Adnani, N.; Wyche, T. P.; Bugni, T. S., Microbial Strain Prioritization Using Metabolomics Tools for the Discovery of Natural Products. *Anal. Chem.* **2012,** 84, (10), 4277-4283.

11. Nielsen, K. F.; Maansson, M.; Rank, C.; Frisvad, J. C.; Larsen, T. O., Dereplication of Microbial Natural Products by LC-DAD-TOFMS. *J. Nat. Prod.* **2011,** 74, (11), 2338-2348.

12. Kasper, P. T.; Rojas-Cherto, M.; Mistrik, R.; Reijmers, T.; Hankemeier, T.; Vreeken, R. J., Fragmentation trees for the structural characterisation of metabolites. *Rapid Commun. Mass Sp.* **2012,** 26, (19), 2275-2286.

13. Stafsnes, M. H.; Dybwad, M.; Brunsvik, A.; Bruheim, P., Large scale MALDI-TOF MS based taxa identification to identify novel pigment producers in a marine bacterial culture collection. *Anton. Leeuw. Int. J. G.* **2013,** 103, (3), 603-615.

14. Laatsch, H., Antibase, A Data Base for Rapid Dereplication and Strcuture Determination of Microbial Natural Products. In *Antibase, A Data Base for Rapid Dereplication and Strcuture Determination of Microbial Natural Products*, Wiley-VCH: Weinheim, Germany: 2013.

15. In *MarinLit*, Department of Chemistry, University of Canterbury: http://www.chem.canterbury.ac.nz/marinlit/marinlit.shtml.: 2013.

16. In *Dictionary of Natural Products*, Chapman & Hall Chemical Database: 2013.

17. Nguyen, D. D.; Wu, C. H.; Moree, W. J.; Lamsa, A.; Medema, M. H.; Zhao, X.; Gavilan, R. G.; Aparicio, M.; Atencio, L.; Jackson, C.; Ballesteros, J.; Sanchez, J.; Watrous, J. D.; Phelan, V. V.; van de Wiel, C.; Kersten, R. D.; Mehnaz, S.; De Mot, R.; Shank, E. A.; Charusanti, P.; Nagarajan, H.; Duggan, B. M.; Moore, B. S.; Bandeira, N.; Palsson, B. O.; Pogliano, K.; Gutierrez, M.; Dorrestein, P. C., MS/MS Networking Guided Analysis of Molecule and Gene Cluster Families. *Submitted* **2013**.

18. 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.; Moore, B. S.; Laskin, J.; Bandeira, N.; Dorrestein, P. C., Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci. USA* **2012,** 109, e1743–e1752.

19. Rath, C. M.; Yang, J. Y.; Alexandrov, T.; Dorrestein, P. C., Data-Independent Microbial Metabolomics with Ambient Ionization Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2013,** Accepted.

20. Clark, B. R.; Engene, N.; Teasdale, M. E.; Rowley, D. C.; Matainaho, T.; Valeriote, F. A.; Gerwick, W. H., Natural Products Chemistry and Taxonomy of the Marine Cyanobacterium Blennothrix cantharidosmum. *J. Nat. Prod.* **2008,** 71, (9), 1530-1537.

21. Hooper, G. J.; Orjala, J.; Schatzman, R. C.; Gerwick, W. H., Carmabins A and B, New Lipopeptides from the Caribbean Cyanobacterium Lyngbya majuscula. *J. Nat. Prod.* **1998,** 61, (4), 529-533.

22. Kim, E. J.; Lee, J. H.; Choi, H.; Pereira, A. R.; Ban, Y. H.; Yoo, Y. J.; Kim, E.; Park, J. W.; Sherman, D. H.; Gerwick, W. H.; Yoon, Y. J., Heterologous Production

of 4-O-Demethylbarbamide, a Marine Cyanobacterial Natural Product. *Org. Lett.* **2012,** 14, (23), 5824-5827.

23. Orjala, J.; Gerwick, W. H., Barbamide, a Chlorinated Metabolite with Molluscicidal Activity from the Caribbean Cyanobacterium Lyngbya majuscula. *J. Nat. Prod.* **1996,** 59, (4), 427-430.

24. Pereira, A. R.; Kale, A. J.; Fenley, A. T.; Byrum, T.; Debonsi, H. M.; Gilson, M. K.; Valeriote, F. A.; Moore, B. S.; Gerwick, W. H., The Carmaphycins: New Proteasome Inhibitors Exhibiting an α,β-Epoxyketone Warhead from a Marine Cyanobacterium. *ChemBioChem* **2012,** 13, (6), 810-817.

25. ProteoWizard Software Foundation: 2013.

26. Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y. F.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J. G.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A., HMDB 3.0-The Human Metabolome Database in 2013. *Nucleic Acids Res.* **2013,** 41, D801-D807.

27. Sud, M.; Fahy, E.; Cotter, D.; Brown, A.; Dennis, E. A.; Glass, C. K.; Merrill, A. H.; Murphy, R. C.; Raetz, C. R. H.; Russell, D. W.; Subramaniam, S., LMSD: LIPID MAPS structure database. *Nucleic Acids Res.* **2007,** 35, D527-D532.

28. Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T., MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **2010,** 45, (7), 703-714.

29. Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G., METLIN - A metabolite mass spectral database. *Ther. Drug Monit.* **2005,** 27, (6), 747-751.

30. In *http://www.nist.gov/index.html*, The National Institute of Standards and Technology (NIST): 2012.

31. Frank, A. M.; Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A.; Pevzner, P. A., De novo peptide sequencing and identification with precision mass spectrometry. *J. Proteome Res.* **2007,** 6, (1), 114-123.

32. Smoot, M. E.; Ono, K.; Ruscheinski, J.; Wang, P. L.; Ideker, T., Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **2011,** 27, (3), 431-432.

33. In *http://apps.cytoscape.org/apps/fm3*, 2013.

34. McPhail, K. L.; Correa, J.; Linington, R. G.; GonzÃ¡lez, J.; Ortega-BarrÃa, E.; Capson, T. L.; Gerwick, W. H., Antimalarial Linear Lipopeptides from a Panamanian Strain of the Marine Cyanobacterium Lyngbya majuscula. *J. Nat. Prod.* **2007,** 70, (6), 984-988.

35. Sitachitta, N.; Marquez, B. L.; Williamson, R. T.; Rossi, J.; Roberts, M. A.; Gerwick, W. H.; Nguyen, V. A.; Willis, C. L., Biosynthetic pathway and origin of the chlorinated methyl group in barbamide and dechlorobarbamide, metabolites from the marine cyanobacterium Lyngbya majuscula. *Tetrahedron* **2000,** 56, (46), 9103-9113.

36. Luesch, H.; Yoshida, W. Y.; Moore, R. E.; Paul, V. J.; Mooberry, S. L., Isolation, Structure Determination, and Biological Activity of Lyngbyabellin A from the Marine Cyanobacterium Lyngbya majuscula. *J. Nat. Prod.* **2000,** 63, (5), 611-615.

37. Luesch, H.; Yoshida, W. Y.; Moore, R. E.; Paul, V. J.; Corbett, T. H., Total Structure Determination of Apratoxin A, a Potent Novel Cytotoxin from the Marine Cyanobacterium Lyngbya majuscula. *J. Am. Chem. Soc.* **2001,** 123, (23), 5418-5423.

38. Luesch, H.; Yoshida, W. Y.; Moore, R. E.; Paul, V. J. C. P. s., New apratoxins of marine cyanobacterial origin from Guam and Palau. *Bioorg. Med. Chem.* **2002,** 10, 1973-1978.

39. Gutierrez, M.; Suyama, T. L.; Engene, N.; Wingerd, J. S.; Matainaho, T.; Gerwick, W. H., Apratoxin D, a potent cytotoxic cyclodepsipeptide from Papua New Guinea collections of the marine cyanobacteria Lyngbya majuscula and Lyngbya sordida. *J. Nat.Prod.* **2008,** 71, 1099-1103.

40. Matthew, S.; Schupp, P. J.; Luesch, H., Apratoxin E, a cytotoxic peptolide from a guamanian collection of the marine cyanobacterium Lyngbya bouillonii. *J. Nat. Prod.* **2008,** 71, 1113-1116.

41. Tidgewell, K.; Engene, N.; Byrum, T.; Media, J.; Doi, T.; Valeriote, F. A.; Gerwick, W. H., Evolved Diversification of a Modular Natural Product Pathway: Apratoxins F and G, Two Cytotoxic Cyclic Depsipeptides from a Palmyra Collection of Lyngbya bouillonii. *ChemBioChem* **2010,** 11, (10), 1458-1466.

42. Bister, B.; Bischoff, D.; Strobele, M.; Riedlinger, J.; Reicke, A.; Wolter, F.; Bull, A. T.; Zahner, H.; Fiedler, H. P.; Sussmuth, R. D., Abyssomicin C - A polycyclic antibiotic from a marine Verrucosispora strain as an inhibitor of the p-aminobenzoic acid/tetrahydrofolate biosynthesis pathway. *Angew. Chem. Int. Ed.* **2004,** 43, 2574-2576.

43. Keller, S.; Nicholson, G.; Drahl, C.; Sorensen, E.; Fiedler, H.-P.; Suessmuth, R. D., Abyssomicins G and H and atrop-abyssomicin C from the marine Verrucosispora strain AB-18-032. *J. Antibiot.* **2007,** 60, 391-394.

44. Wang, Q.; Song, F.; Xiao, X.; Huang, P.; Li, L.; Monte, A.; Abdel-Mageed, W. M.; Wang, J.; Guo, H.; He, W.; Xie, F.; Dai, H.; Liu, M.; Chen, C.; Xu, H.; Liu, M.; Piggott, A. M.; Liu, X.; Capon, R. J.; Zhang, L., Abyssomicins from the South China Sea Deep-Sea Sediment Verrucosispora sp.: Natural Thioether Michael Addition Adducts as Antitubercular Prodrugs. *Angew. Chem. Int. Ed.* **2013,** 52, 1231-1234.

45. Moree, W. J.; Phelan, V. V.; Wu, C. H.; Bandeira, N.; Cornett, D. S.; Duggan, B. M.; Dorrestein, P. C., Interkingdom metabolic transformations captured by microbial imaging mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** 109, (34), 13811-13816.

46. Rath, C. M.; Alexandrov, T.; Higginbottom, S. K.; Song, J.; Milla, M. E.; Fischbach, M. A.; Sonnenburg, J. L.; Dorrestein, P. C., Molecular Analysis of Model Gut Microbiotas by Imaging Mass Spectrometry and Nanodesorption Electrospray Ionization Reveals Dietary Metabolite Transformations. *Anal. Chem.* **2012,** 84, (21), 9259-9267.

47. Arima, K.; Kakinuma, A.; Tamura, G., SURFACTIN A CRYSTALLINE PEPTIDELIPID SURFACTANT PRODUCED BY BACILLUS SUBTILIS - ISOLATION CHARACTERIZATION AND ITS INHIBITION OF FIBRIN CLOT FORMATION. *Biochem. Biophys. Res. Commun.* **1968,** 31, (3), 488-&.

48. Fiedler, H.-P.; Bruntner, C.; Riedlinger, J.; Bull, A. T.; Knutsen, G.; Goodfellow, M.; Jones, A.; Maldonado, L.; Pathom-aree, W.; Beil, W.; Schneider, K.; Keller, S.; Sussmuth, R. D., Proximicin A, B and C, novel aminofuran antibiotic and anticancer compounds isolated from marine strains of the actinomycete Verrucosispora. *J. Antibiot.* **2008,** 61, 158-163.

49. Pesci, E. C.; Milbank, J. B. J.; Pearson, J. P.; McKnight, S.; Kende, A. S.; Greenberg, E. P.; Iglewski, B. H., Quinolone signaling in the cell-to-cell communication system of Pseudomonas aeruginosa. *Proc. Natl. Acad. Sci. U. S. A.* **1999,** 96, (20), 11229-11234.

50. Sanchez, L. M.; Wong, W. R.; Riener, R. M.; Schulze, C. J.; Linington, R. G., Examining the Fish Microbiome: Vertebrate-Derived Bacteria as an Environmental Niche for the Discovery of Unique Marine Natural Products. *Plos One* **2012,** 7, (5), e35398.

51. Benson, D. A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostell, J.; Sayers, E. W., GenBank. *Nucleic Acids Res.* **2013,** 41, (D1), D36-D42.

52. Hooft, J. J.; Vos, R. H.; Ridder, L.; Vervoort, J.; Bino, R., Structural elucidation of low abundant metabolites in complex sample matrices. *Metabolomics* **2013,** ASAP.

53. Hakvaag, S.; Fjaervik, E.; Josefsen, K. D.; Ian, E.; Ellingsen, T. E.; Zotchev, S. B., Characterization of Streptomyces spp. isolated from the sea surface microlayer in the Trondheim fjord, Norway. *Mar. Drugs* **2008,** 6, (4), 620-635.

3.7 Supporting Information

3.7.1 Cyanobacterial collection information.

3.7.1.1 Collection that contained tumonoic acid I.

An assemblage of *Moorea producens* (previously *Lyngbya* sp.) and *Schizothrix* (PNG-22/APR/06-2) was collected in April 2006 at a depth of 10-30 ft near Nuakata Island in Dudawali Bay in the Commonwealth realm of Papua New Guinea with GPS coordinates of 10° 17’ 274’’ S and 151° 00’ 390’’ E. The sample, measuring 2 L in total biomass, was preserved in a 1:1 mixture of isopropanol and seawater, transported to San Diego, and stored at -20 °C until extraction. A small sample of the cyanobacterial biomass was also preserved in RNAlater (Qiagen) for subsequent 16S rRNA sequencing and analysis. A voucher sample is in our laboratory at Scripps (PNG-22/APR/06-2).

3.7.1.2 Collection that contained carmabin A.

A red-colored sample of *Moorea producens* was collected in Portobelo (Cacique) in Panama. The sample (750 mL) was preserved in a 1:1 isopropanol-seawater solution and stored at -20°C. A small sample of cyanobacterial biomass was also preserved in RNAlater (Qiagen) for subsequent 16S rRNA sequencing and analysis. A voucher sample is in our laboratory at Scripps (PAP-25/Jun/12-2).

3.7.1.3 Collection containing barbamide.

The sample of red filamentous cyanobacteria (PAP-24/Mar/12-1) was collected in March 2012 at a depth of 8-12 feet by scuba diving in Tres Hermanas, Portobelo, Panama. The sample (500 mL) was preserved in 1:1 isopropanol:seawater

and stored at -20°C until extraction. A voucher sample is in our laboratory at Scripps (PAP-24/Mar/12-1).

3.7.1.4 Collection that contained carmaphycin B.

Growing as small colonies (1-2 cm) with a yellow-green base and puffy red tips, this sample was collected in June 2012 from coral at various shallow water depths. It was found in three different sites (#1 Isla Mamae, 20-54 ft; #2 Cacique, 20-55 ft; #3 Island across from Bananas Resort, 20-40 ft) near Portobelo (Panama). The sample was field identified as *Schizothrix* sp. A 500 mL sample was preserved in a 1:1 isopropanol:seawater solution, transported to San Diego, and stored at -20°C until it was extracted. A small sample of cyanobacterial biomass was also preserved in RNAlater (Qiagen) for subsequent 16S rRNA sequencing and analysis. A voucher sample is in our laboratory at Scripps (PAP-25/JUN/12-1).

3.7.1.5 Collection of *Moorea bouillonii.*

Moorea bouillonii PNG05-198^{T} was originally collected off of the island of New Ireland, Papua New Guinea and since then has been grown in culture in the Gerwick laboratory. It was defined as the type strain of this species in prior work to describe the genus *Moorea*.[1] Work presented in this manuscript was conducted with extractions of the cultured material. The Genbank codes for the 16s rRNA sequencing and analysis for this strain are FJ041298/9.

3.7.2 Bacterial strain-specific culture conditions.

Strain MS100128 was isolated using oatmeal agar from a sediment sample collected in April 2010 from the South China Sea (20° 9.795′ N, 118° 18.124′ E) at

2733 m below sea level and was identified as a *Verrucosispora* sp. using 16S rRNA gene sequence analysis (GenBank accession no. JQ724543). The strain has been preserved at the China General Microbiological Culture Collection Center (accession no. 5847). Strain MS100128 was cultivated on VER01 agar plate (starch 1%, glucose 1%, glycerol 1%, corn steep powder (Sigma) 0.25%, peptone (Difco) 0.5%, yeast extract 0.2%, NaCl 0.1%, $CaCO_3$ 0.3%, agar 2%; pH 7.0) at 28°C for 7 days.

The strain SRM7 was isolated using oatmeal agar from a sediment sample collected in the South China Sea at 20°9.795′ N and 118°18.124′ E, 2733m below sea level in April 2010. Spore morphology and ornamentation were determined by scanning electron microscopy. The organism forms a well-developed, branched, red colored substrate mycelium, which carries single spores, which have a warty ornamentation. It was identified as *Verrucosispora* sp. using 16S rRNA gene sequence analysis. The bacteria has been assigned the accession number SRM7 in the culture collection at the Institute of Microbiology, Chinese Academy of Sciences, Beijing. SRM7 was cultivated under the same conditions and procedure as MS100128.

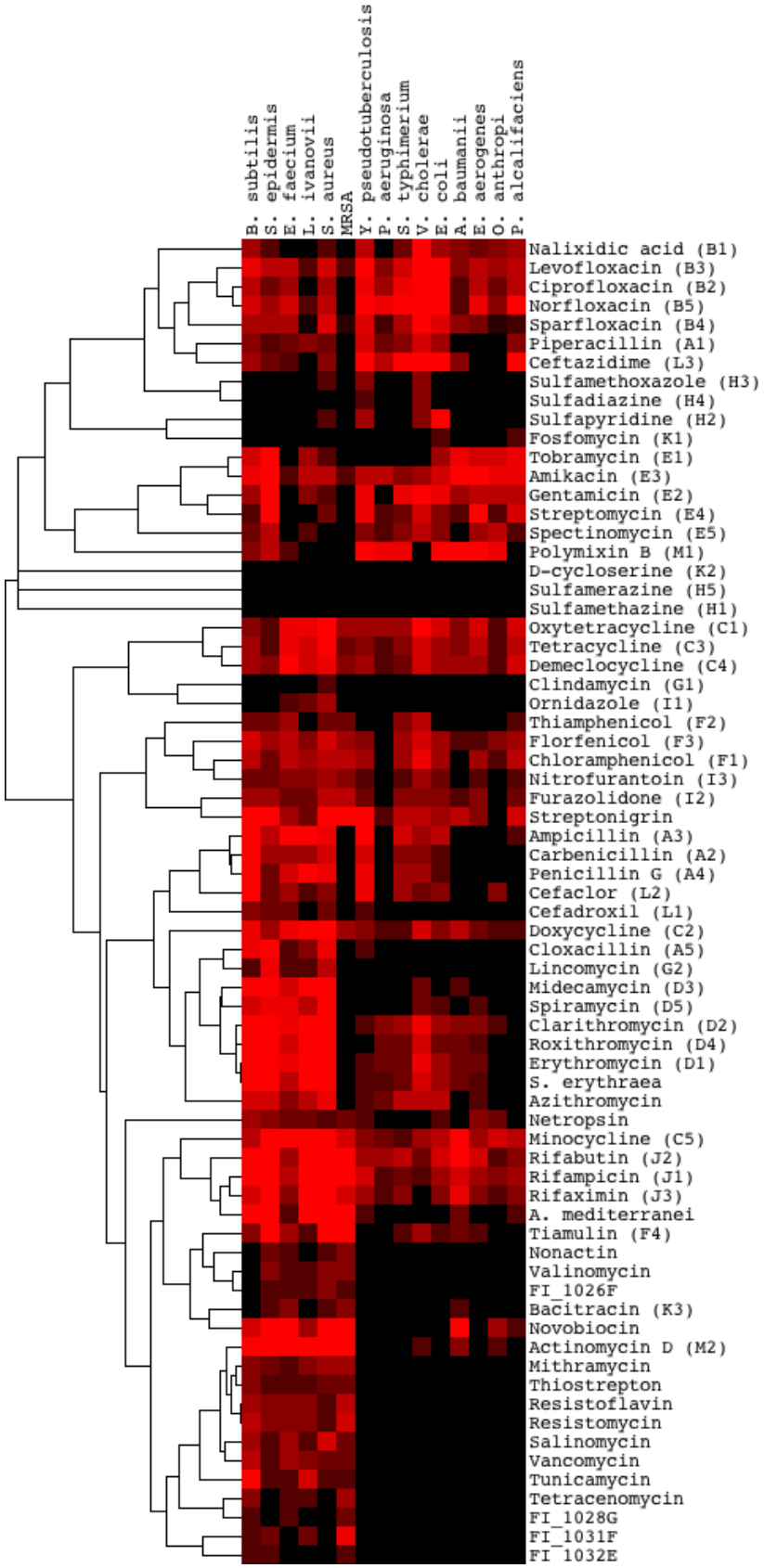
Purified FI-1026, which was identified as a *Dietzia* sp. by 16S rDNA analysis (GenBank accession no. JQ691548), bacterial colonies were grown in 1 L of modified SYP broth (1 L MilliQ water, 32.1 g Instant Ocean™, 10 g starch, 4 g peptone, 2 g yeast) with 20 g of Amberlite XAD-16 resin for 10 days at 27°C. Culture broth and resin slurries were filtered through glass microfiber filters, washed with water (3 x 200 mL) and the cells, resin, and filter paper extracted with 1:1 methanol/dichloromethane (250 mL). Organic fractions were dried *in vacuo* and subjected to solid phase extraction (SPE) using Supelco-Discovery C_{18} cartridges (5 g) eluting with a step

gradient of 40 mL of MeOH/H_2O solvent mixtures (10% MeOH, 20% MeOH, 40% MeOH, 60% MeOH, 80% MeOH, 100% MeOH) and finally with EtOAc to afford seven fractions.

Pseudomonas aeruginosa PAO1[2] and PA14[2] were first streaked on LB agar plates from frozen glycerol stocks. Single colonies were picked and used to inoculate LB liquid medium at 37°C. Then 1 mL inoculum was applied to ISP2 agar plates that were incubated at 37°C.

Serratia marcescens sp. ES129 (Lab Environmental Strain Collection) and *Bacillus subtilis* 3610 and PY79[2,3] each was streaked on ISP2 agar plates from frozen glycerol stocks. One colony of each was grown in 2 mL of ISP2 at 28°C under aerobic conditions. One mL inoculum (OD_{600}~0.1) of each bacteria strain was applied to ISP2 agar media and incubated at 30°C for 48 h.

Figure S3.1 BioMAP[4] and cytological profiles[5] (CPs) of known antibiotics and *Dietzia* sp. FI-1026 extracts. In BioMAP, the extract clusters closely with valinomycin, indicative of the structure being a large cyclic peptide. In CP the extract causes death for HeLa cells.


B. subtilis
S. epidermis
E. faecium
L. ivanovii
S. aureus
MRSA
Y. pseudotuberculosis
P. aeruginosa
S. typhimerium
V. cholerae
E. coli
A. baumanii
E. aerogenes
O. anthropi
P. alcalifaciens
Nalixidic acid (B1)
Levofloxacin (B3)
Ciprofloxacin (B2)
Norfloxacin (B5)
Sparfloxacin (B4)
Piperacillin (A1)
Ceftazidime (L3)
Sulfamethoxazole (H3)
Sulfadiazine (H4)
Sulfapyridine (H2)
Fosfomycin (K1)
Tobramycin (E1)
Amikacin (E3)
Gentamicin (E2)
Streptomycin (E4)
Spectinomycin (E5)
Polymixin B (M1)
D-cycloserine (K2)
Sulfamerazine (H5)
Sulfamethazine (H1)
Oxytetracycline (C1)
Tetracycline (C3)
Demeclocycline (C4)
Clindamycin (G1)
Ornidazole (I1)
Thiamphenicol (F2)
Florfenicol (F3)
Chloramphenicol (F1)
Nitrofurantoin (I3)
Furazolidone (I2)
Streptonigrin
Ampicillin (A3)
Carbenicillin (A2)
Penicillin G (A4)
Cefaclor (L2)
Cefadroxil (L1)
Doxycycline (C2)
Cloxacillin (A5)
Lincomycin (G2)
Midecamycin (D3)
Spiramycin (D5)
Clarithromycin (D2)
Roxithromycin (D4)
Erythromycin (D1)
S. erythraea
Azithromycin
Netropsin
Minocycline (C5)
Rifabutin (J2)
Rifampicin (J1)
Rifaximin (J3)
A. mediterranei
Tiamulin (F4)
Nonactin
Valinomycin
FI_1026F
Bacitracin (K3)
Novobiocin
Actinomycin D (M2)
Mithramycin
Thiostrepton
Resistoflavin
Resistomycin
Salinomycin
Vancomycin
Tunicamycin
Tetracenomycin
FI_1028G
FI_1031F
FI_1032E

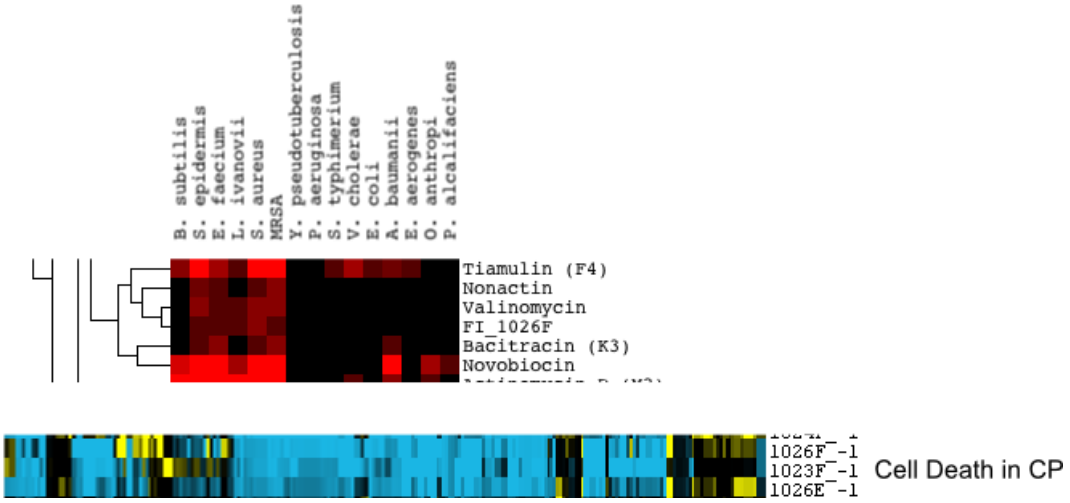


Figure S3.1 BioMAP and cytological profiles, continued

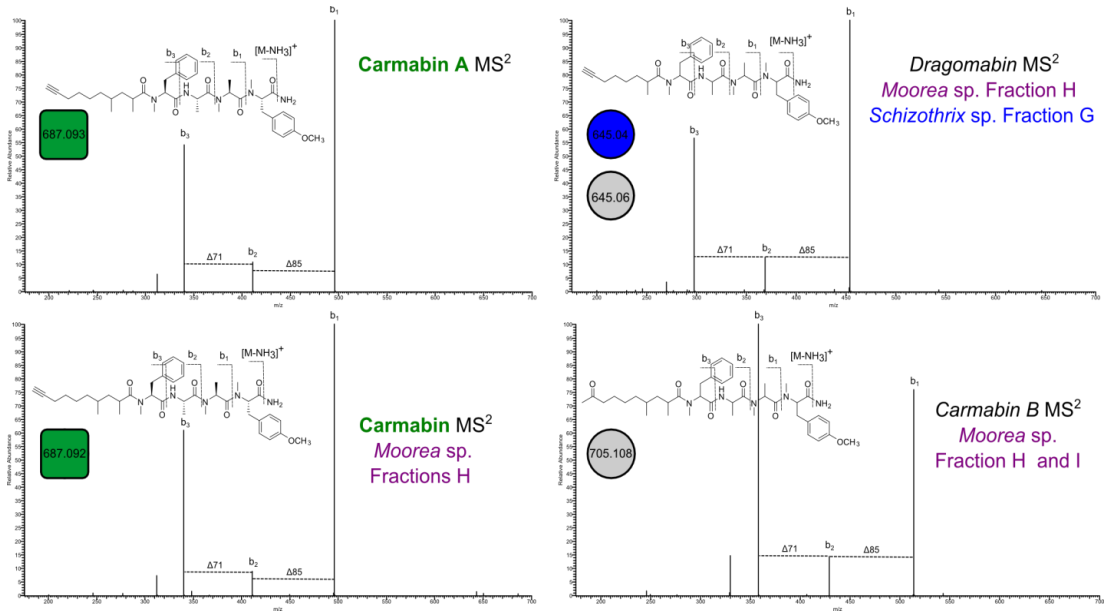


Figure S3.2 Annotation of carmabin MS/MS spectra. Manual inspection of the carmabin A cluster showed that the crude *Moorea* sp. fraction and the carmabin A seed have similar retention times (within five milliseconds).

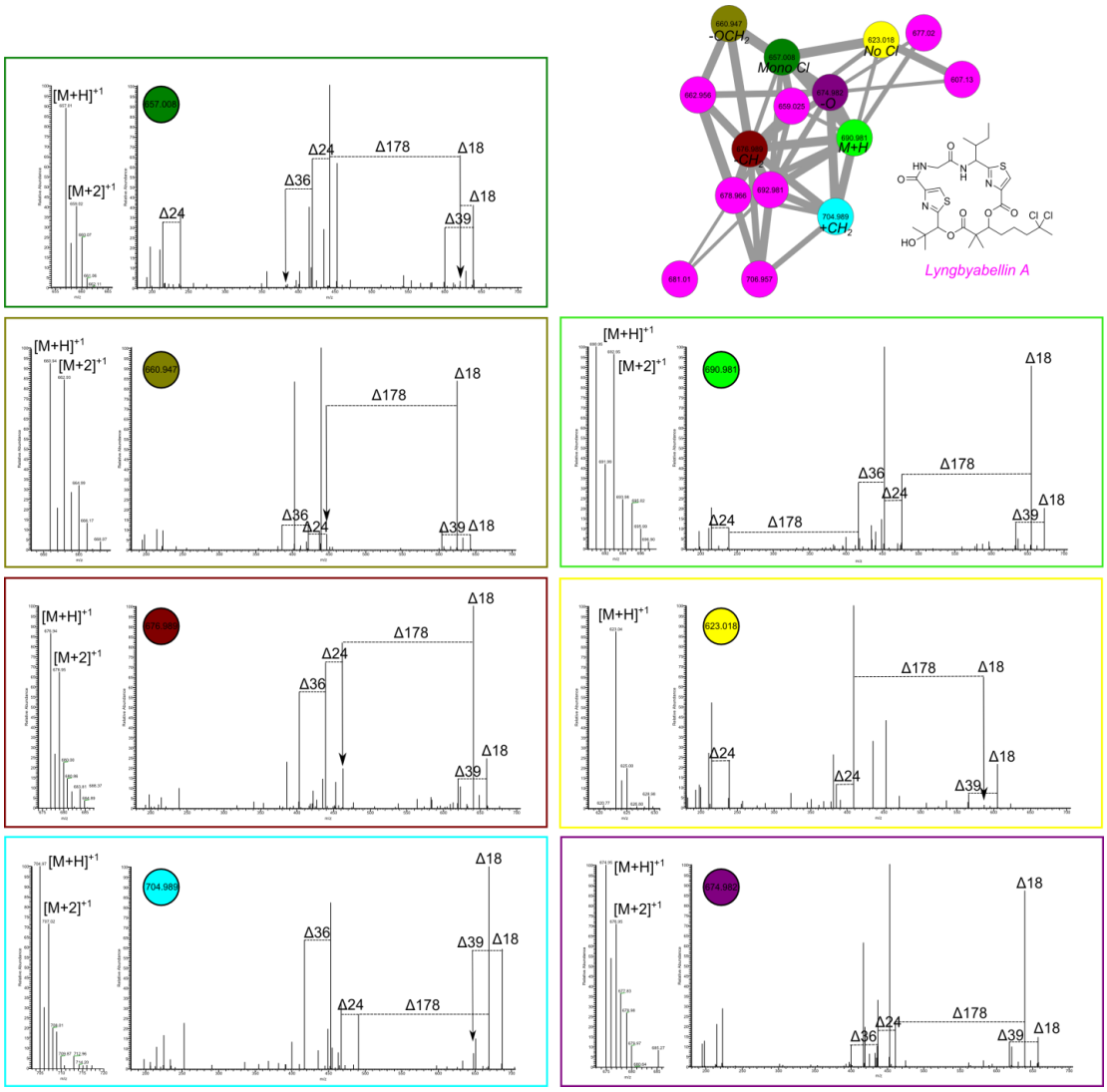
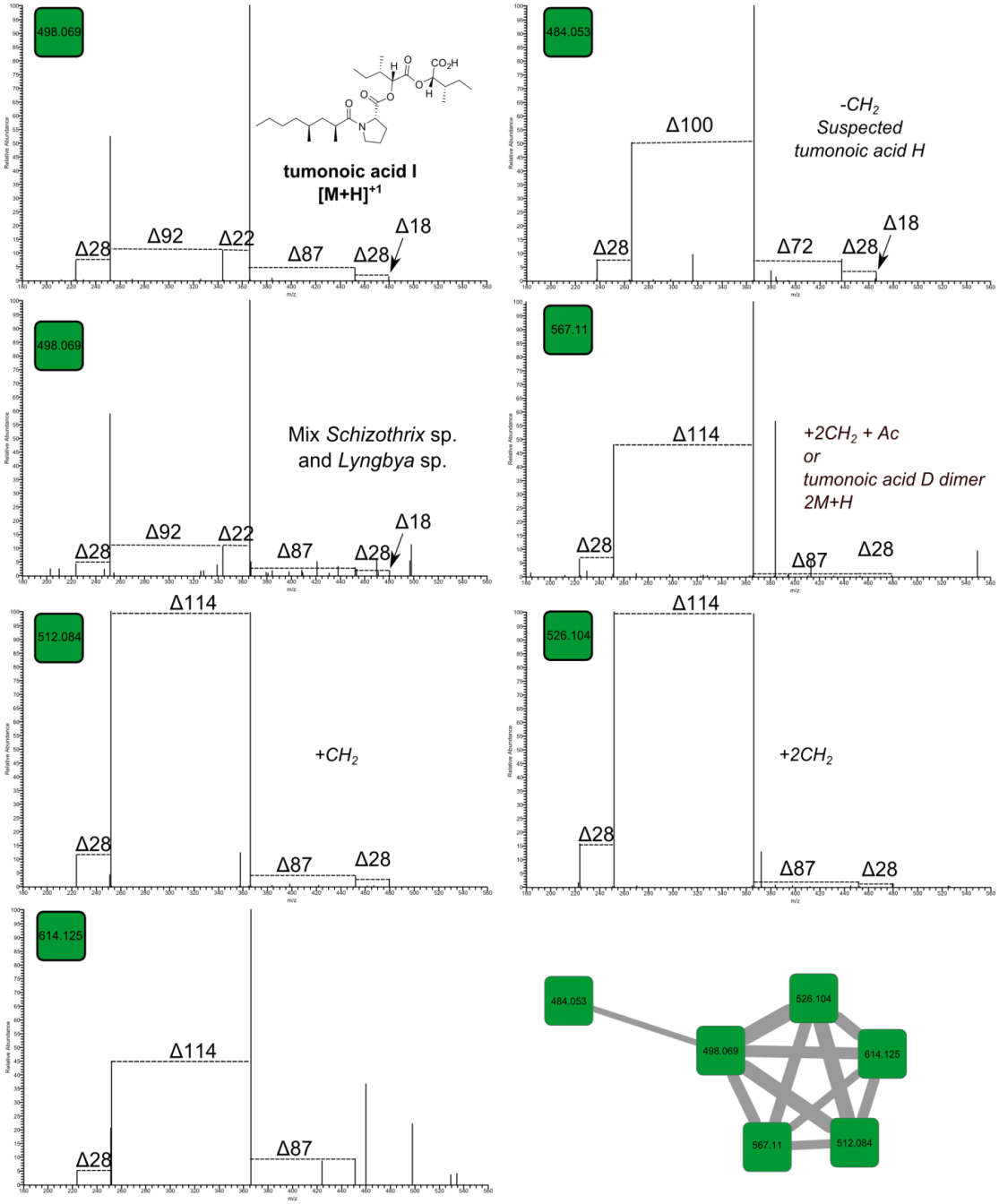


Figure S3.3 Annotation of lyngbyabellin MS/MS spectra and isotopic distribution of precursor ions. Two clusters from *Moorea bouillonii*, known to produce lyngbyabellins and apratoxins, did not contain seed compounds. The MS/MS fragmentation, precursor masses, and isotopic pattern for one cluster were highly suggestive of the known metabolite lyngbyabellin A.[6] This cluster suggested novel analogs, including a rare monochlorinated species, a dehydroxylated species, a demethylated species, a demethylated and dehydroxylated species, and a methylated species. Additionally, the precursor masses for these analogs had complex chlorine isotopic patterns (M+2), which were highly suggestive of their presence in this cyanobacterial sample.

Figure S3.4 Annotation of tumonoic acid I MS/MS spectra and EIC of precursor masses found to cluster with the seed. Tumonoic acid I was found in a mixed collection of two cyanobacteria, *Lyngbya* sp. and *Schizothrix* sp., from Papua New Guinea.[7] Interestingly, within the seed LC-MS/MS run, there appeared to be other analogs of tumonoic acid I as well as the known compound. Inspection of the extracted ion chromatogram (EIC) of these precursor ions as well as the UV signal indicated the seed as the most abundant compound. Based on the EIC, the retention times of the other analogs were consistent with the analogs identified from networking. However, the analogs were in very low titer, as illustrated by the absence of their UV signals. For example, based on networking, there appeared to be an analog that contained two extra methyl groups (*m/z* 526).This analog eluted 4 minutes after tumonoic acid I indicating it was more non-polar than the seed, consistent with the presence of two extra methyl groups.


498.069
tumonoic acid I
[M+H]+1
Δ28
Δ92
Δ22
Δ87
Δ28
Δ18
484.053
-CH2
Suspected
tumonoic acid H
Δ100
Δ28
Δ72
Δ28
Δ18
498.069
Mix Schizothrix sp.
and Lyngbya sp.
Δ28
Δ92
Δ22
Δ87
Δ28
Δ18
567.11
+2CH2 + Ac
or
tumonoic acid D dimer
2M+H
Δ114
Δ28
Δ87
Δ28
512.084
+CH2
Δ114
Δ28
Δ87
Δ28
526.104
+2CH2
Δ114
Δ28
Δ87
Δ28
614.125
Δ114
Δ28
Δ87
484.053
526.104
498.069
614.125
567.11
512.084

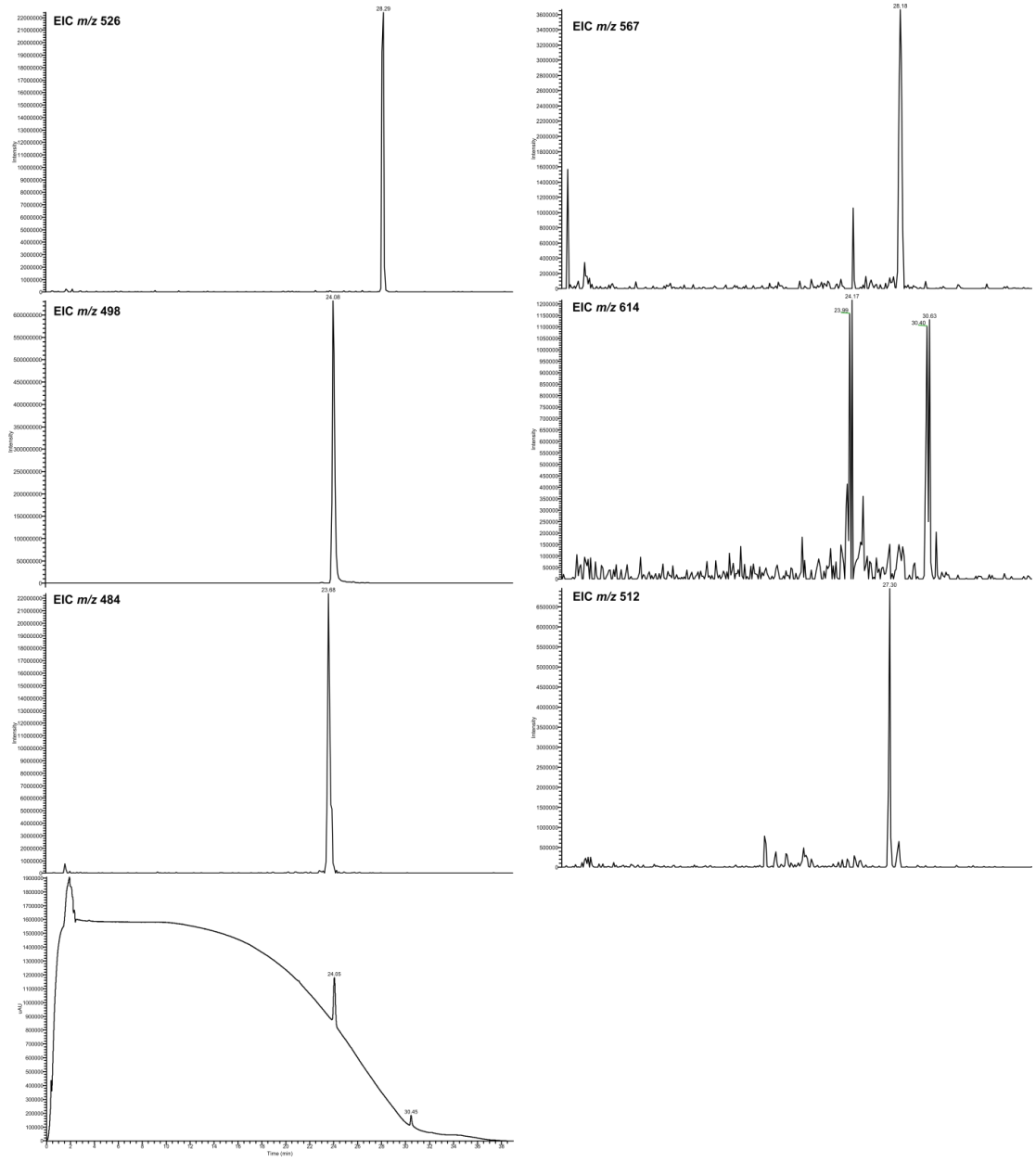


Figure S3.4 Annotation of tumonoic acid I, continued

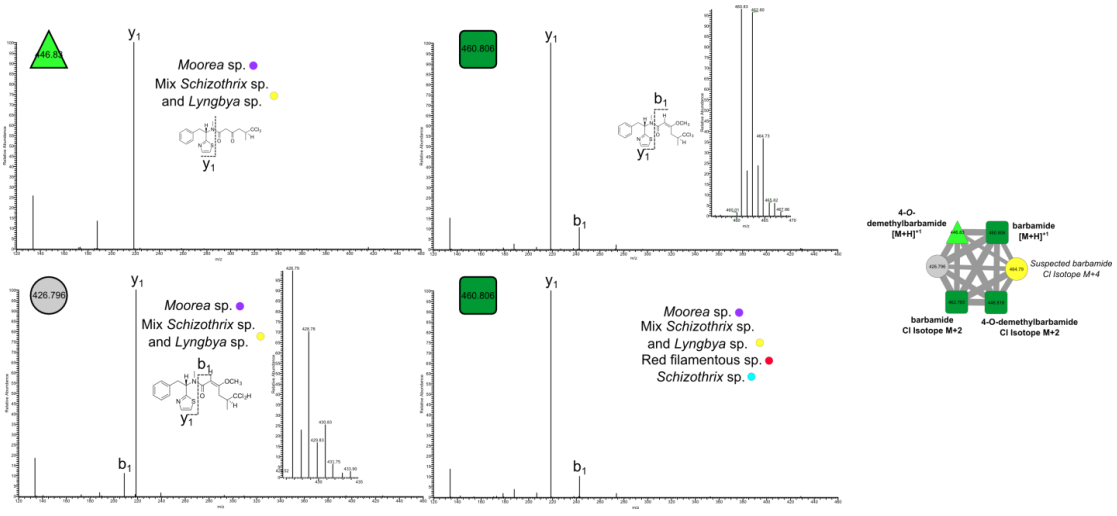


Figure S3.5 Annotation of barbamide MS/MS spectra. Barbamide was found in multiple collections of cyanobacteria.[8] In addition to this seed, the chlorine isotope was detected as well the recently described 4-*O*-demethylbarbamide and its complex chlorine isotopic pattern.[9] While this analog was not used as a seed, the fragmentation pattern agreed well with 4-*O*-demethylbarbamide since the y ion was intact. Additionally, there was a compound seen in two of the cyanobacterial collections, which displayed high similarity to the spectrum of the barbamide seed. The mass difference and precursor isotopic pattern was consistent with dechlorobarbamide and verified by the presence of *y* and *b* ions.[10]

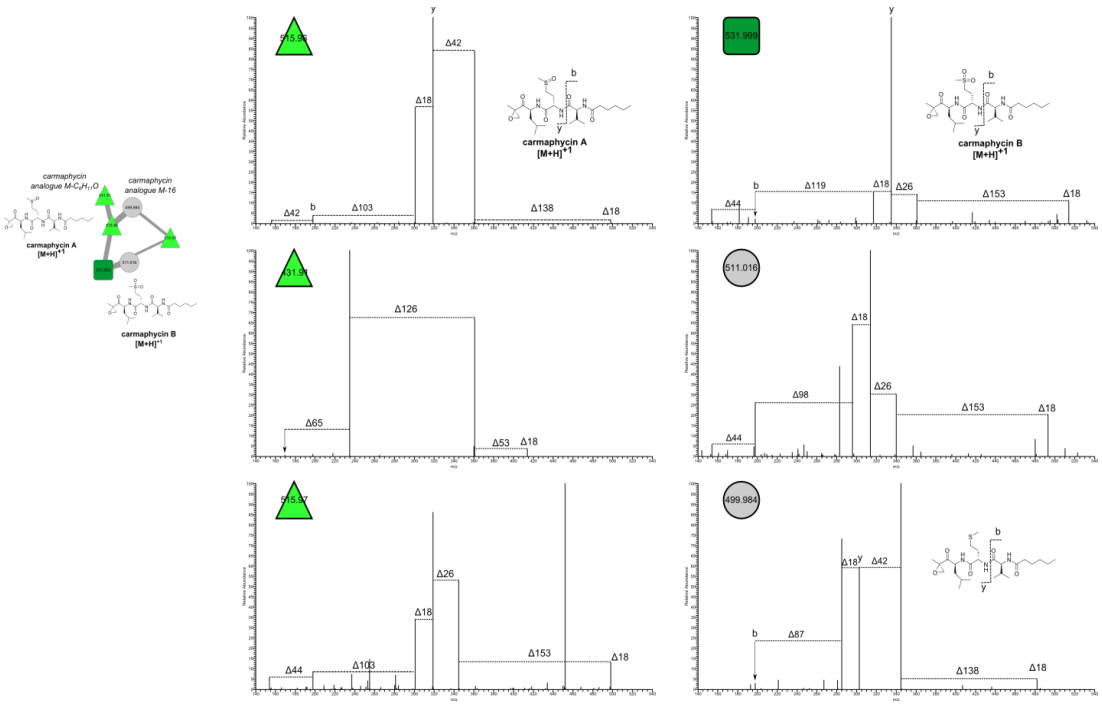


Figure S3.6 Annotation of carmaphycin A and B MS/MS spectra. Carmaphycin A and B were used as seeds and only carmaphycin B was found in a one collection of cyanobacteria from Panama.[11] Through networking and comparison of retention times, the presence of carmaphycin B was confirmed in the crude sample. Additionally, there were two nodes suggestive of novel carmaphycin analogs. One node was suggestive of non-oxidized methionine analog, with *y* and *b* ions shown which agreed with this modification, and an analog at *m/z* 511, which has an intact *b* ion that matched carmaphycin B. A full structure elucidation would be required to identify this modification.

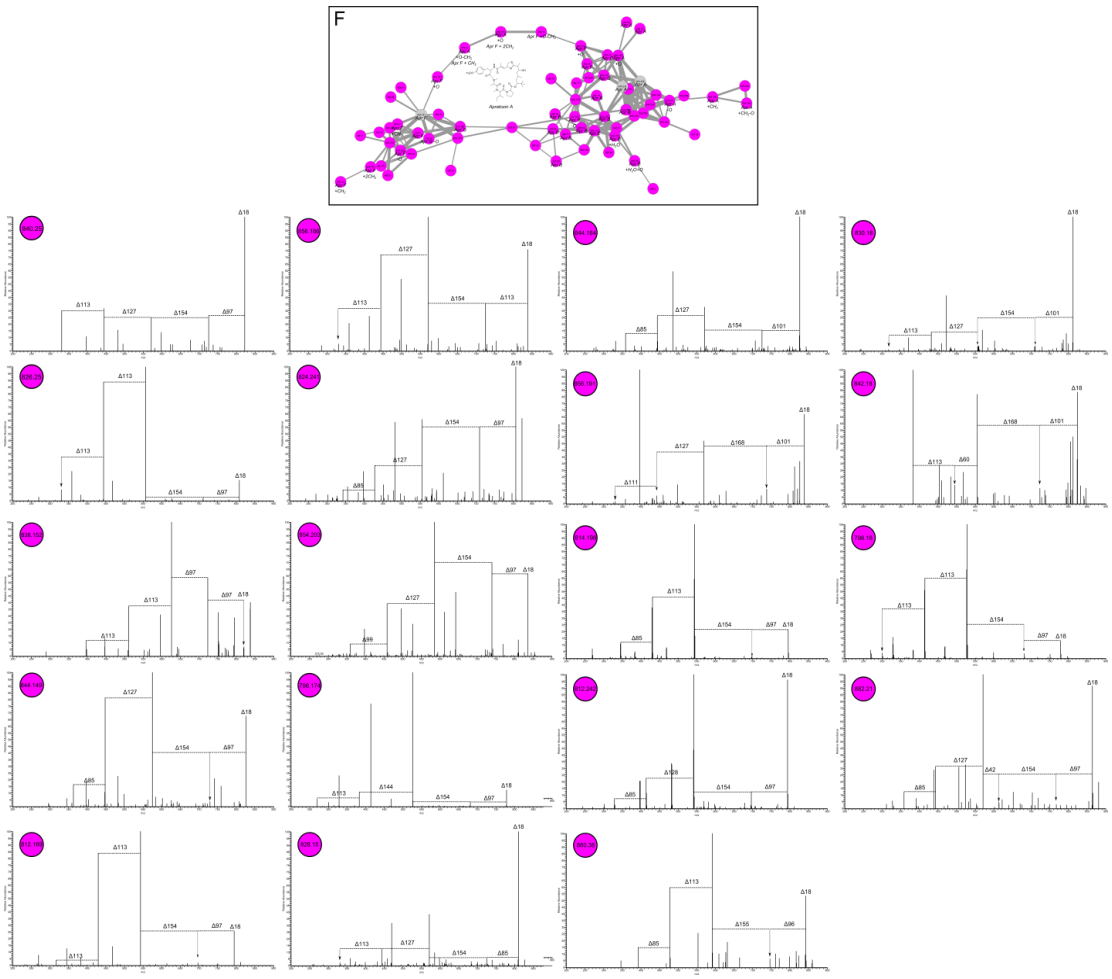


Figure S3.7 Annotation of apratoxin MS/MS spectra. Two clusters from *Moorea bouillonii*, known to produce lyngbyabellins and apratoxins, did not contain seed compounds. The other cluster was large and likely contained apratoxins A and B as well as masses suggesting the analog structures of apratoxins D, E, F, and G.[12-17] However, molecular formulas of these compounds could not be assigned definitively because ppm error values were too high, and without standards available to compare to these assignments were as such not definitive.

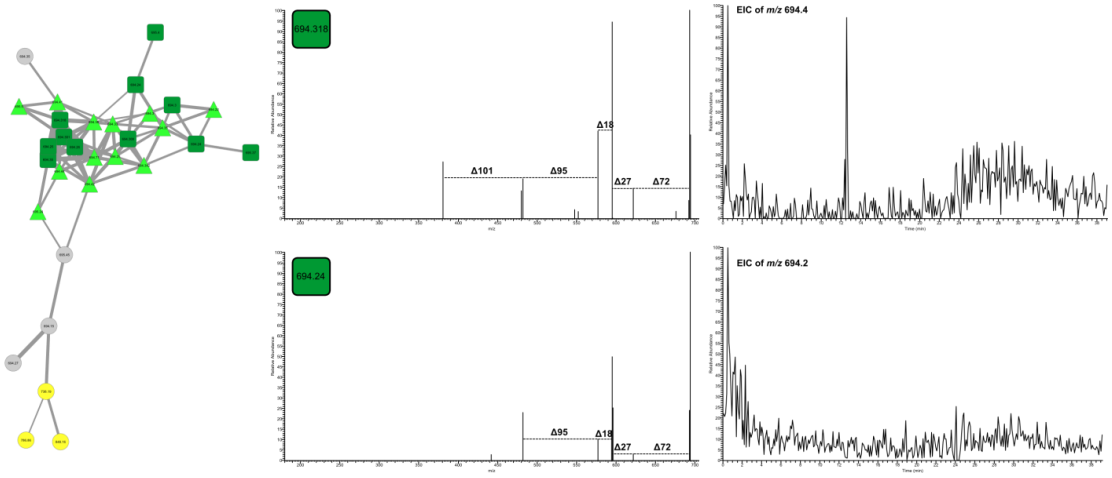


Figure S3.8 Annotation of precursor mass *m/z* 694 MS/MS spectra and EIC.

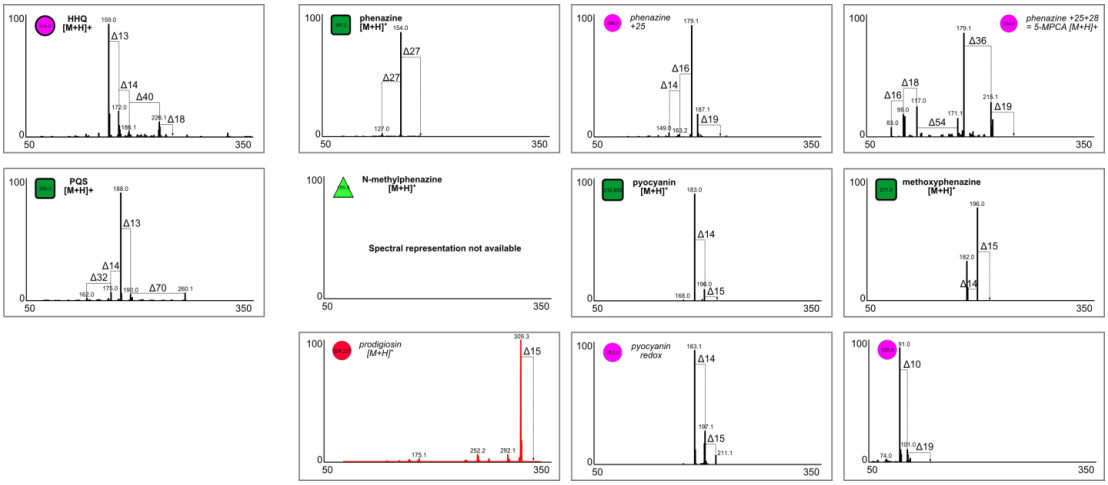


Figure S3.9 MS/MS spectra corresponding to phenazine and quinolone nodes. Molecular networking analysis of data-independent nanoDESI-MS/MS of *Pseudomonas aeruginosa* PAO1 and PA14[2] with six reference spectra – hydroxyphenazine, methoxyphenazine, *N*-methylphenazine, phenazine, *Pseudomonas* quinolone signal (PQS), and pyocyanin – yielded 3 perfect matches to methoxyphenazine, pyocyanin, PQS, and phenazine. Pyocyanin (211.0866 $[M+H]^+$ calc.) was immediately linked to a putative analog 2 *m/z* higher, corresponding to a reduced analog. Methoxyphenazine was recently identified by our lab as a fungal transformation product of a *Pseudomonas* metabolite, and hence, was not expected in our sample. However, methoxyphenazine did cluster to the expected molecular precursor.[18] Our reference spectrum for phenazine did not cluster with the other phenazines, but it did cluster in the same general region of the condensed cluster. PQS exhibited a perfect match between experimental and reference data, resulting in an overlapping node. Due to the displayed similarity in the molecular networking approach, one analog 2-heptyl-4-quinolone (HHQ) could readily be identified.

A microbe isolated from soil, Environmental Strain 129 (Lab Environmental Strain Collection), produced a red pigment and claded with *Serratia marscesens*via 16S rRNA. Known *Serratia marcescens* strains produce the red-pigmented antibiotic prodigiosin (323.1997 calc.), a tripyrrole alkaloid. In the bacteria network, there was a node with a precursor mass that matches prodigiosin, but there was no standard/purified compound "seed" MS/MS spectrum. Examination of the experimental MS/MS spectrum and comparison to published MS/MS,[19] confirms that the fragmentation patterns match. This was sufficient to rule out the molecule as a new structure and in traditional natural product discovery workflows, efforts can then be directed elsewhere. If absolute dereplication is desired, purification and isolation is required.

Figure S3.10 MS/MS spectra corresponding to abyssomicin nodes. The bacteria network included MS/MS spectra directly collected from a *Verrucosispora* sp. MS100128 colony grown on agar media, which is known to produce abyssomicins. The MS/MS spectra for each of the NMR-pure standards were not parsed out from the data files, and hence the network included MS/MS spectra for impurities present in the standard samples. This, in turn, increased the number of nodes corresponding to the standards. The abyssomicin standards contained impurities, as indicated per the nodes containing spectra from multiple abyssomicin samples. This was also the case for proximicin B. The uncharacterized strain SRM7 is known to produce proximicin B (data not shown). Within the network, the node corresponding to the proximicin B standard was present (zoomed data not shown) but did not match or connect to any experimental SRM7 node. Molecular networking does not distinguish between polarities. Database nodes from negative polarity $[M-H]^-$ spectra are incorporated into the network. Database nodes correspond to molecular structures that are not related to the abyssomicins or to proximicin.

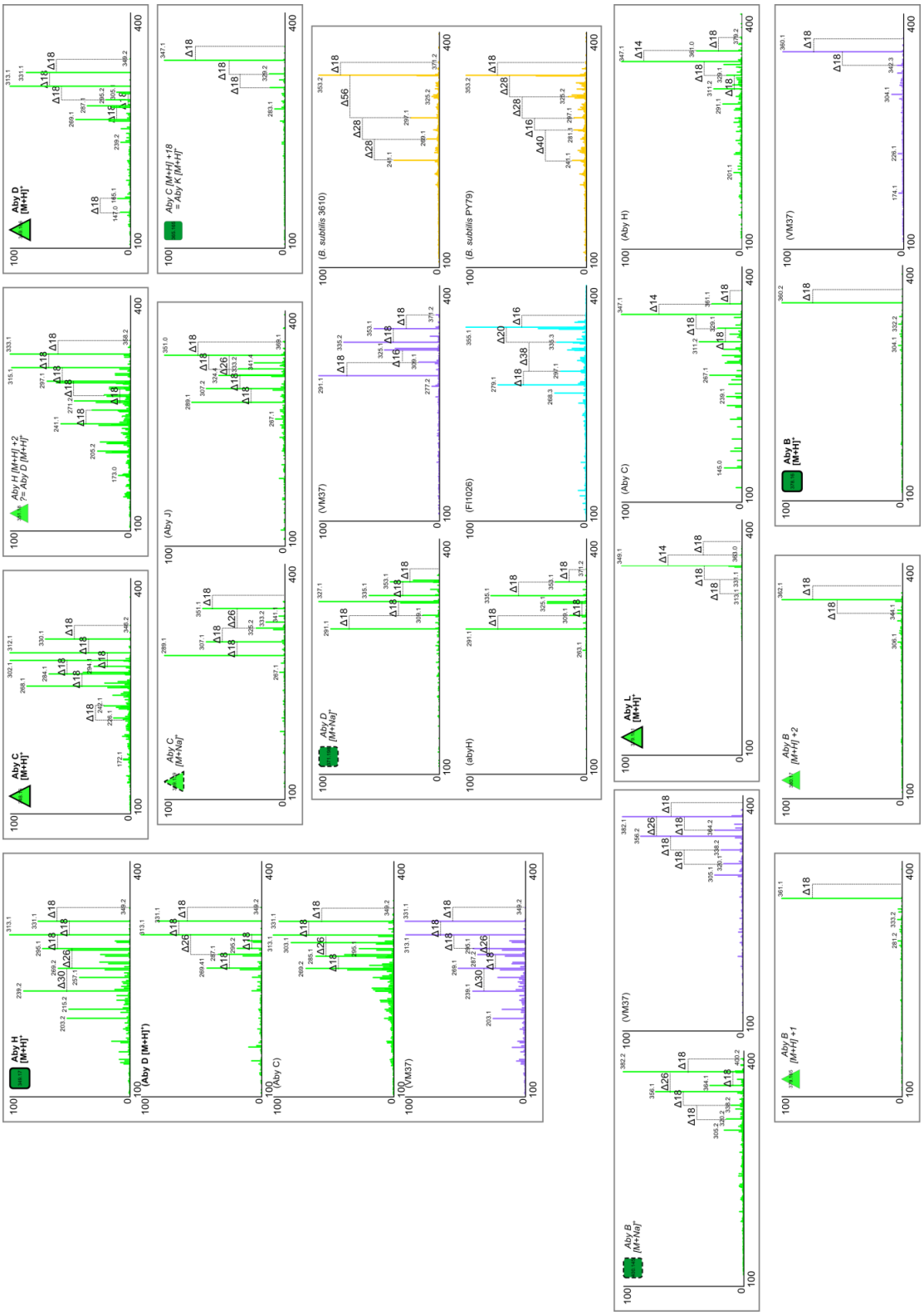
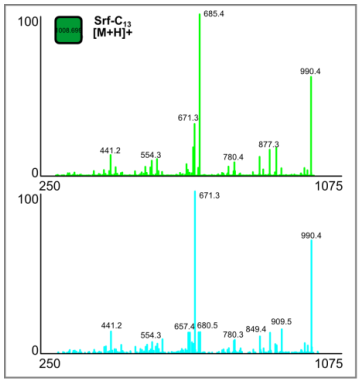
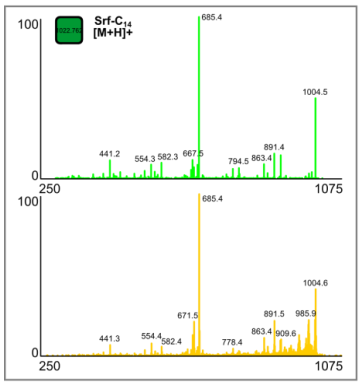
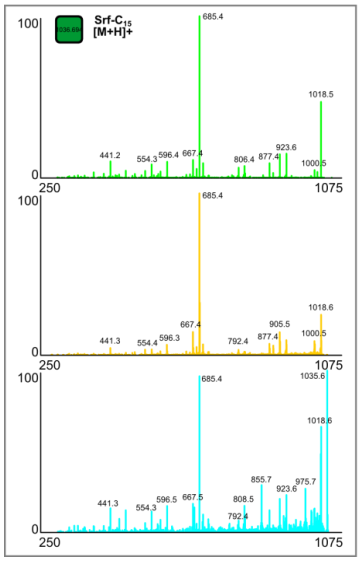
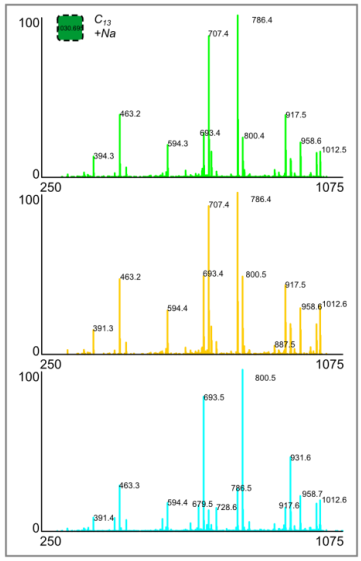

Aby H [M+H]+
Aby D [M+H]+
Aby C [M+H]+
Aby H [M+H] +2 ?= Aby D [M+H]+
Aby D [M+H]+
Aby C [M+Na]+
(Aby J)
Aby C [M+H] +18 = Aby K [M+H]+
(B. subtilis 3610)
(B. subtilis PY79)
(VM37)
(F11026)
(abyH)
Aby L [M+H]+
(Aby H)
(Aby C)
Aby B [M+H] +2
(VM37)
Aby B [M+Na]+
Aby B [M+H] +1

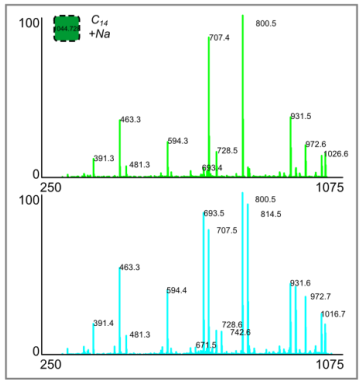
Figure S3.11 MS/MS spectra corresponding to nodes in the 1+ surfactin cluster. A noticeable computational limitation meriting further development was multiple nodes for precursor masses that are within 0.5 Da, as seen by multiple nodes for 1008.6597, 1022.6753, and 1036.6910 [M+H]+ calc. (Fig 4D). 994.827 / 994.648 is -14 from surfactin-C_{13} $[M+H]^+$. These nodes do not merge into consensus spectra, despite meeting the precursor mass tolerance set at 0.5 Da. However, the presence of dual nodes in one cluster in the network verified the presence of each of the surfactins and respective analogs.

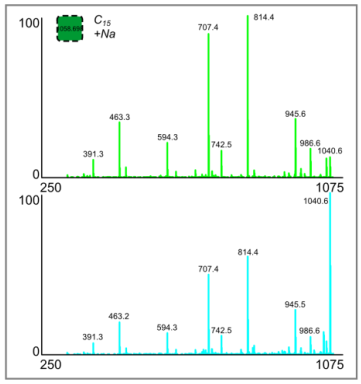

Srf-C13
[M+H]+
685.4
990.4
671.3
441.2
554.3
780.4
877.3
671.3
990.4
441.2
554.3
657.4
680.5
780.3
849.4
909.5
100
0
250
1075

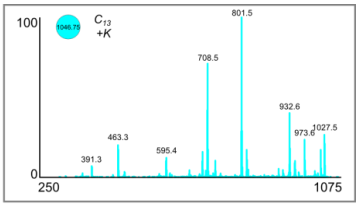

Srf-C14
[M+H]+
685.4
1004.5
441.2
554.3
582.3
667.5
794.5
863.4
891.4
685.4
1004.6
671.5
441.3
554.4
582.4
778.4
863.4
891.5
909.6
985.9
100
0
250
1075

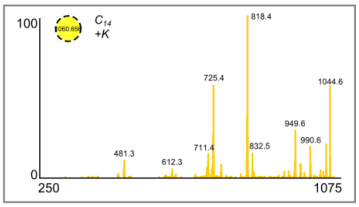

Srf-C15
[M+H]+
685.4
1018.5
441.2
554.3
596.4
667.4
806.4
877.4
923.6
1000.5
685.4
1018.6
441.3
554.4
596.3
667.4
792.4
877.4
905.5
1000.5
685.4
1035.6
1018.6
441.3
554.3
596.5
667.5
792.4
808.5
855.7
923.6
975.7
100
0
250
1075

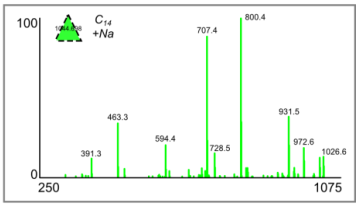

C13
+Na
786.4
707.4
463.2
394.3
594.3
693.4
800.4
917.5
958.6
1012.5
707.4
786.4
463.2
693.4
800.5
917.5
391.3
594.4
958.6
1012.6
887.5
800.5
693.5
931.6
463.3
786.5
391.4
594.4
679.5
728.6
958.7
917.6
1012.6
100
0
250
1075

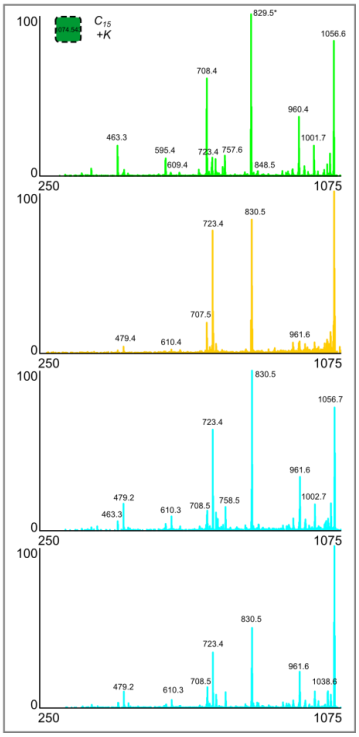

C14
+Na
800.5
707.4
463.3
594.3
931.5
391.3
481.3
728.5
972.6
1026.6
693.4
800.5
693.5
814.5
707.5
463.3
594.4
931.6
972.7
391.4
481.3
728.6
742.6
1016.7
671.5
100
0
250
1075


C15
+Na
814.4
707.4
463.3
945.6
594.3
742.5
986.6
1040.6
391.3
1040.6
814.4
707.4
945.5
463.2
594.3
742.5
986.6
391.3
100
0
250
1075


C13
+K
801.5
708.5
932.6
463.3
595.4
973.6
1027.5
391.3
100
0
250
1075


C14
+K
818.4
725.4
1044.6
949.6
990.6
481.3
612.3
711.4
832.5
100
0
250
1075


C14
+Na
800.4
707.4
463.3
931.5
594.4
728.5
972.6
1026.6
391.3
100
0
250
1075


C15
+K
829.5*
1056.6
708.4
960.4
463.3
595.4
723.4
757.6
1001.7
609.4
848.5
723.4
830.5
707.5
479.4
610.4
961.6
830.5
1056.7
723.4
961.6
479.2
610.3
708.5
758.5
1002.7
463.3
830.5
723.4
961.6
479.2
610.3
708.5
1038.6
100
0
250
1075

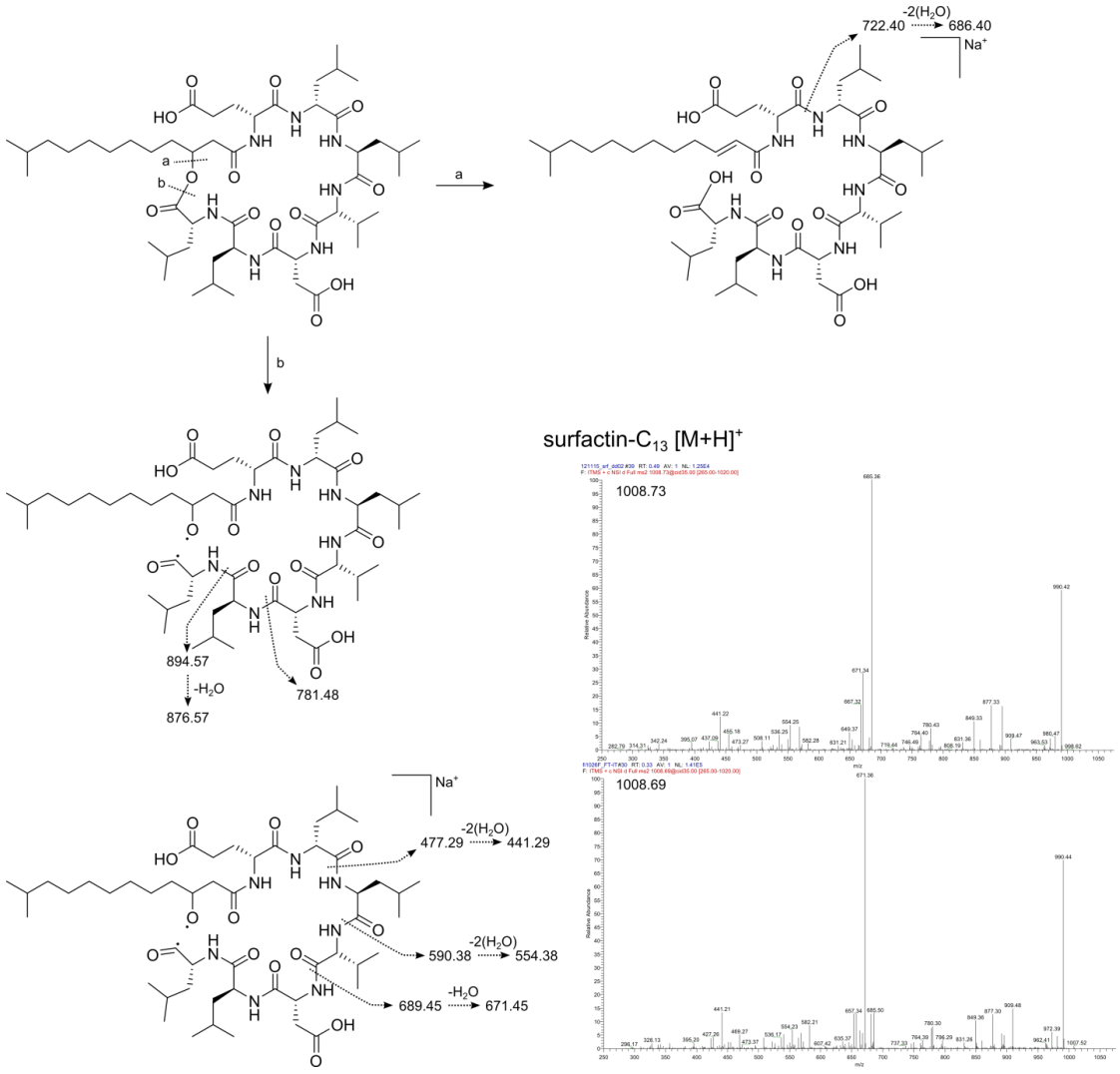


Figure S3.12 Annotation of surfactin-C_{13} $[M+H]^+$ MS/MS.

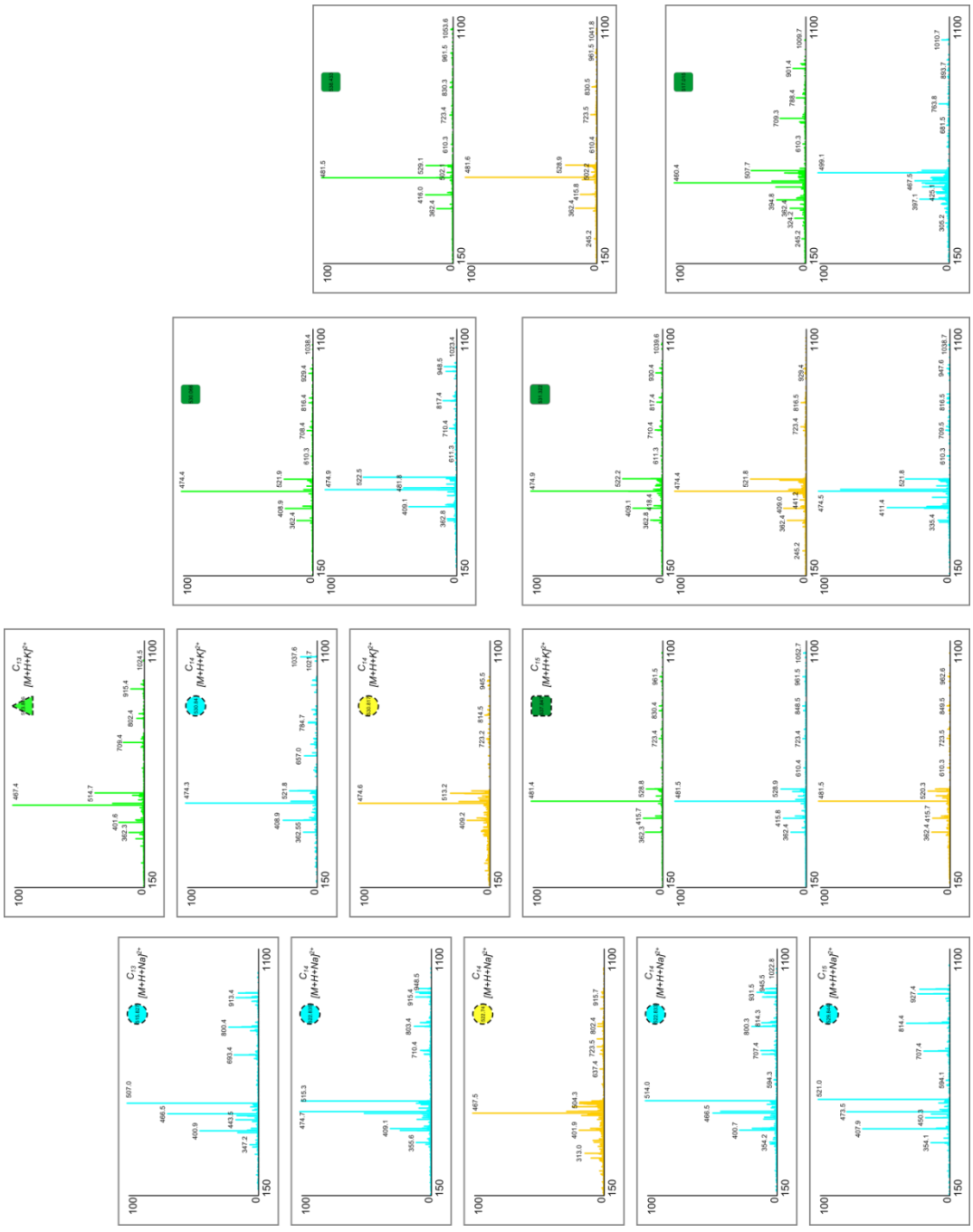


Figure S3.13 MS/MS spectra corresponding to nodes in the 2+ surfactin cluster.

3.8 Supporting References

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

2. Rath, C. M.; Yang, J. Y.; Alexandrov, T.; Dorrestein, P. C., Data-Independent Microbial Metabolomics with Ambient Ionization Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2013,** Accepted.

3. Yang, Y. L.; Xu, Y. Q.; Straight, P.; Dorrestein, P. C., Translating metabolic exchange with imaging mass spectrometry. *Nat. Chem. Biol.* **2009,** 5, (12), 885-887.

4. Wong, W. R.; Oliver, A. G.; Linington, R. G., Development of Antibiotic Activity Profile Screening for the Classification and Discovery of Natural Product Antibiotics. *Chem. Biol.* **2012,** 19, (11), 1483-1495.

5. Schulze, C. J.; Bray, W. M.; Woerhmann, M. H.; Stuart, J.; Lokey, R. S.; Linington, R. G., "Function-First" Lead Discovery: Mode of Action Profiling of Natural Product Libraries Using Image-Based Screening. *Chem. Biol.* **2013,** 20, (2), 285-295.

6. Luesch, H.; Yoshida, W. Y.; Moore, R. E.; Paul, V. J.; Mooberry, S. L., Isolation, Structure Determination, and Biological Activity of Lyngbyabellin A from the Marine Cyanobacterium Lyngbya majuscula. *J. Nat. Prod.* **2000,** 63, (5), 611-615.

7. Clark, B. R.; Engene, N.; Teasdale, M. E.; Rowley, D. C.; Matainaho, T.; Valeriote, F. A.; Gerwick, W. H., Natural Products Chemistry and Taxonomy of the Marine Cyanobacterium Blennothrix cantharidosmum. *J. Nat. Prod.* **2008,** 71, (9), 1530-1537.

8. Orjala, J.; Gerwick, W. H., Barbamide, a Chlorinated Metabolite with Molluscicidal Activity from the Caribbean Cyanobacterium Lyngbya majuscula. *J. Nat. Prod.* **1996,** 59, (4), 427-430.

9. Kim, E. J.; Lee, J. H.; Choi, H.; Pereira, A. R.; Ban, Y. H.; Yoo, Y. J.; Kim, E.; Park, J. W.; Sherman, D. H.; Gerwick, W. H.; Yoon, Y. J., Heterologous Production of 4-O-Demethylbarbamide, a Marine Cyanobacterial Natural Product. *Org. Lett.* **2012,** 14, (23), 5824-5827.

10. Sitachitta, N.; Marquez, B. L.; Williamson, R. T.; Rossi, J.; Roberts, M. A.; Gerwick, W. H.; Nguyen, V. A.; Willis, C. L., Biosynthetic pathway and origin of the chlorinated methyl group in barbamide and dechlorobarbamide, metabolites from the marine cyanobacterium Lyngbya majuscula. *Tetrahedron* **2000,** 56, (46), 9103-9113.

11. Pereira, A. R.; Kale, A. J.; Fenley, A. T.; Byrum, T.; Debonsi, H. M.; Gilson, M. K.; Valeriote, F. A.; Moore, B. S.; Gerwick, W. H., The Carmaphycins: New Proteasome Inhibitors Exhibiting an α,β-Epoxyketone Warhead from a Marine Cyanobacterium. *ChemBioChem* **2012,** 13, (6), 810-817.

12. Gu, Z.-M.; Wang, L.-Q.; Wu, J., Mass defect filter - a new tool to expedite screening and dereplication of natural products and generate natural product profiles. *Nat. Prod. J.* **2011,** 1, (2), 135-145.

13. Luesch, H.; Yoshida, W. Y.; Moore, R. E.; Paul, V. J.; Corbett, T. H., Total Structure Determination of Apratoxin A, a Potent Novel Cytotoxin from the Marine Cyanobacterium Lyngbya majuscula. *J. Am. Chem. Soc.* **2001,** 123, (23), 5418-5423.

14. Luesch, H.; Yoshida, W. Y.; Moore, R. E.; Paul, V. J. C. P. s., New apratoxins of marine cyanobacterial origin from Guam and Palau. *Bioorg. Med. Chem.* **2002,** 10, 1973-1978.

15. Matthew, S.; Schupp, P. J.; Luesch, H., Apratoxin E, a cytotoxic peptolide from a guamanian collection of the marine cyanobacterium Lyngbya bouillonii. *J. Nat. Prod.* **2008,** 71, 1113-1116.

16. Tidgewell, K.; Engene, N.; Byrum, T.; Media, J.; Doi, T.; Valeriote, F. A.; Gerwick, W. H., Evolved Diversification of a Modular Natural Product Pathway: Apratoxins F and G, Two Cytotoxic Cyclic Depsipeptides from a Palmyra Collection of Lyngbya bouillonii. *ChemBioChem* **2010,** 11, (10), 1458-1466.

17. Gutierrez, M.; Suyama, T. L.; Engene, N.; Wingerd, J. S.; Matainaho, T.; Gerwick, W. H., Apratoxin D, a potent cytotoxic cyclodepsipeptide from Papua New Guinea collections of the marine cyanobacteria Lyngbya majuscula and Lyngbya sordida. *J. Nat.Prod.* **2008,** 71, 1099-1103.

18. Moree, W. J.; Phelan, V. V.; Wu, C. H.; Bandeira, N.; Cornett, D. S.; Duggan, B. M.; Dorrestein, P. C., Interkingdom metabolic transformations captured by microbial imaging mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** 109, (34), 13811-13816.

19. Chen, K.; Rannulu, N. S.; Cai, Y.; Lane, P.; Liebl, A. L.; Rees, B. B.; Corre, C.; Challis, G. L.; Cole, R. B., Unusual Odd-Electron Fragments from Even-Electron Protonated Prodiginine Precursors Using Positive-Ion Electrospray Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2008,** 19, 1856-1866.

Chapter 3, in full, has been submitted to the Journal of Natural Products, June 2013. Jane Y. Yang, Laura M. Sanchez, Christopher M. Rath, Xueting Liu, Paul Boudreau, Nicole Bruns, Evgenia Glukhov, Anne Wodtke, Rafael de Felicio, Amanda Fenner, Weng Ruh Wong, Roger G. Linington, Lixing Zhang, Hosana M. Debonsi, William H. Gerwick, and Pieter C. Dorrestein. The dissertation author and Laura M. Sanchez were the primary investigators and authors of this paper.

CHAPTER IV

Future directions: mass spectrometry based networking of honey bee colony collapse disorder

4.1 Abstract (Project Summary)

Colony collapse disorder (CCD) has been responsible for the annual loss of approximately one-third of the honey bee population since 2006, threatening commercial pollination of crops worldwide. CCD is characterized by the unexplained disappearance of adult worker bees, the presence of sufficient food, unhatched brood, and an abandoned queen. Despite many implicated causes, which include extended exposure to insecticides, parasitic mites, and viruses, studies have focused on one culprit at a time. More recent metagenomic and transcriptomic studies have identified associated microbes and the corresponding transcribed genes, which have provided valuable insight into the genes potentially involved in CCD. However, physiology is dictated and affected by small molecules. Surprisingly, there have been little to no metabolomics studies of CCD. In this study, we aim to use metabolomics approach in conjunction with mass spectrometry based molecular networking to uncover the chemical differences of healthy and CCD hives. In parallel, we aim to use metagenomics to investigate the microbiota present. Microbiota has been shown to have a role in modulating host health and disease. Through the combination of these data, we hope to guide future studies of CCD to improve honey bee survival. In addition, this study will establish a workflow, to identify chemical differences and associated microbiota, adaptable to any host-microbiota system.

4.2 Background and Significance

The total honey bee population has decreased by approximately one-third each year since 2006, affecting the commercial pollination of $15 billion in crop value in the US[1-4] as well as food crops worldwide.[5] This decline is attributed to colony collapse disorder (CCD), which describes the abrupt disappearance of adult bees from hives, oftentimes with adequate food supplies, capped or an unhatched brood, and a queen left behind. There are a number of proposed causes – pathogens, mites, fungi, pesticides, antibiotics, stress due to the environment or long-distance transport, and nutrition – or a combination thereof has been implicated in immunodeficiencies related to CCD.[6-15] However, the causes of CCD remain elusive.[16-18]

Since the publication of the honey bee (*Apis mellifera*) genome,[19, 20] the focus has shifted from the exposure to sub-lethal levels of neonicotinoid insecticides[13, 21-24] to metagenomics[14, 25] and transcriptomics.[12, 26] These 'omics' analyses have shown that pesticides may not have as large a role in CCD as previously thought, and instead implicated an increased pathogenic or viral load.[9, 25, 26] In addition, the metagenomic profiles included 81 distinct fungi, bacteria not previously found in bees, and a shift in healthy honey bee gut microbiota[27] associated with disease.[14, 25] Interestingly, there has been little emphasis on proteomics[15, 18, 28] or metabolomics[11] of colonies suffering from CCD.

Honey bees and their microbiota are exposed to many insecticides, pathogens, and stresses, all which affect honey bees on a molecular or metabolic level. **We hypothesize that there are chemical differences between healthy and CCD**

affected colonies and that these differences are a direct result of the associated microbiota. These chemical differences (the effects), once identified, will implicate or preclude (the cause/s), especially with metagenomic analysis in parallel, thereby contributing to a deeper understanding of CCD.

To address this hypothesis, we propose to directly compare the metabolomic profiles of healthy and abandoned CCD hives and corresponding bees, when possible, utilizing mass spectrometry based molecular networking.[29, 30] We expect to identify the chemical differences and similarities with mass spectrometry based tools and associate them with the microbial communities identified by metagenomics. These molecules and the associated microbial profiles may facilitate the implementation of preventative measures to improve honey bee colony survivorship.

4.3 Specific Aims

We seek to identify important molecules associated with CCD by comparing healthy and CCD samples in order to begin understanding the chemical differences and the significance of these chemical cues in honey bee hive health. These differences, when combined with the metagenomic microbial profiles, may implicate or eliminate proposed causes of CCD. The three specific aims of this proposal are:

Aim 1: Establish the chemical differences and similarities present in healthy and CCD hives and bees.

1a. Samples from healthy and diseased hives and bees will be collected, prepared, and subjected to tandem mass spectrometry based metabolomics analysis.

1b. The data generated in Aim 1a will be used to generate MS/MS-based molecular networks that enable simultaneous visualization of multiple chemical differences.

Aim 2: Establish the microorganisms present in healthy and CCD hives and bees.

2a. In parallel to Aim 1, samples will be collected, prepared, and subjected to metagenomics sequencing.

2b. Results from Aim 2a will be analyzed to reveal the differences in microbiota of healthy versus CCD samples.

Aim 3: Identify and verify the molecules and correlate them to the microbial source. Based on the results from Aim 1, we will select molecules of interest, characterize them, and correlate them to the microbial differences from Aim 2.

3a. Identify and characterize the molecules that differ between healthy and CCD affected samples.

3b. Generate a model for the microbial contribution to honey bee health.

4.4 Research Strategy

Approach Summary. The goal of this proposal is to identify molecules associated with CCD. There are three key steps in this investigation. First, metabolomics and MS/MS based molecular networking analysis will be used to reveal the molecules that differ between healthy and CCD hives and bees. Second, metagenomics will be used to identify the microbes present. Both metabolomic and genomic data will be collected in parallel, to associate the molecules with the microbes. Third, with the data from steps 1 and 2, candidate molecules of interest will be isolated, characterized, and correlated to microbial differences.

Aim 1: Establish the chemical differences and similarities present in healthy and CCD hives and bees. Within this aim, we will recruit beekeepers from multiple locations within the United States who manage hives that are both healthy and showing signs of CCD. Power calculations will dictate the sample size, or number of bees per group, required to achieve statistical significance. Metadata, such as location, climate conditions, colony status, etc., will be reported by the beekeepers. Samples will be subjected to solvent extraction, LC-MS/MS, and mass spectrometry based molecular networking to identify the molecules unique to healthy or CCD samples. These patterns will also be compared between different locations.

1a. Determine the molecular differences. Samples from healthy and diseased hives and bees will be collected, immediately prepared, and subjected to LC-MS/MS analysis. Guts may be isolated as a part of the preparation, in order to partition out the molecules from the gut microbiota. Samples will be ground up and treated with

chemical solvents to extract metabolites of differing polarities. These extracts will be dried and re-dissolved in mass spectrometry-compatible solvents, and injected onto a reverse phase HPLC system coupled to a high-resolution mass spectrometer capable of tandem mass spectrometry. Data will be acquired using data dependent fragmentation.

1b. Generate and analyze MS/MS-based molecular networks. Results from Aim 1a will be incorporated into a molecular network, which enables visualization of molecular similarity based on MS/MS fragmentation patterns. This approach allows for the cross correlation of multiple, large datasets. Similar molecules will cluster or overlap within the network; dissimilar molecules will be located away from each other within the network. Dissimilar molecules will be easily identified, decreasing the time to determine which candidate molecules to focus upon.

Expected results. Within this specific aim, we expect to address an important issue for which our current understanding is incomplete, the chemical differences between healthy and CCD hives and bees. The results from Aim 1 will define the key metabolites associated with health and CCD and would provide a foundation for the investigation of metabolic pathways, gene expression, or microbiota within honey bees. If, perchance, there are no explicit differences detected within the molecular network, we will utilize iterations of XCMS software[31, 32] to determine relative quantities of molecules present.

Aim 2: Establish the microorganisms present in healthy and CCD hives and bees. Microbes interact via metabolic exchange and also play a role in host health.[33-39] Previous metagenomics studies implicated an increased viral load, but also

identified fungi and a shift in the gut microbiota of CCD honey bees.[14] Differences in microbiota of healthy versus CCD samples may explain the chemical differences resultant from Aim 1.

2a. Metagenomic sequencing will reveal the microbes associated with honey bees and hives. In parallel to Aim 1, each sample will be processed and subjected to 16s rDNA sequencing to determine the microbes present.

2b. Comparison of microbes associated with health versus CCD. Results from Aim 2a will be analyzed to reveal the differences in microbial profiles between healthy versus CCD samples.

Expected results. The results of this aim, at the very least, will corroborate the findings of the previous study.[14] Multiple datasets will enable the correlation of differences in microbiota to chemical differences determined in Aim 1. This will provide insight into what may cause CCD.

Aim 3: Identify and verify the molecules and correlate them to the microbial source. Within this aim, we will attempt to identify and verify the molecules of interest determined from Aim 1 and incorporate the microbial differences from Aim 2 to generate a model for the microbial contribution to honey bee health.

3a. Identify and characterize the molecules that differ between healthy and CCD samples. Literature and database searches of known insecticides, metabolites resulting from insecticides, parasitic mites, fungicides, and nutrient sources will be utilized in matching the unknown molecules to an identity. These will

then be cataloged into a ‘honey bee’ database, which can be accessed for future studies. Manual inspection and annotation of MS/MS spectra may provide additional structural information about the molecules of interest, as will MS^n follow-up experiments. If the molecule is present in sufficient quantities, purification and subsequent NMR structure elucidation studies will be conducted.

3b. Generate a model. We will combine data from Aims 1 and 2 to generate a working model for the microbial contribution to honey bee health. Addition of all other data generated will assist in pointing to the causes of CCD.

Expected results. Through identification of the molecules and microbes associated with CCD, we will be set up to propose a model for the possible causes of CCD. There is the possibility that the differences in microbiota of healthy and CCD are an effect of the actual cause. These results will help researchers retrace one more step back from chemical and microbial differences in order to more fully understand CCD.

4.5 Future Directions

The long-term goal of this study is to help prevent or decrease the incidence of honey bee losses due to CCD by determining the contributing chemical and microbial factors. In order to begin to identify the causes contributing to CCD, we first seek to determine the chemical differences between healthy and CCD honey bee colonies by identifying small molecules present in the different hive states. We will collaborate with experts in metagenomics in order to identify the microbes, such as fungi, bacteria, and viruses, present in these samples, and consult with statisticians to evaluate the relevance of the correlations between the molecules and microbiota. Then we will incorporate the data and knowledge from CCD-related studies of the past seven years to generate a CCD model in the attempt to determine the potential cause/s of CCD. We hope that this model will assist in the implementation of the appropriate preventative measures, perhaps a probiotic-type approach, to improve honey bee colony survivorship. This approach can be applied to other host-microbiota communities, biofilm communities, and potentially multi-factorial epidemics.

4.6 References

1. Committee, C. S. *Colony Collapse Disorder Action Plan*; 2007; p 27.

2. Calderone, N. W., Insect Pollinated Crops, Insect Pollinators and US Agriculture: Trend Analysis of Aggregate Data for the Period 1992-2009. *Plos One* 2012, 7.

3. vanEngelsdorp, D.; Caron, D.; Hayes, J.; Underwood, R.; Henson, M.; Rennich, K.; Spleen, A.; Andree, M.; Snyder, R.; Lee, K.; Roccasecca, K.; Wilson, M.; Wilkes, J.; Lengerich, E.; Pettis, J.; Bee Informed, P., A national survey of managed honey bee 2010-11 winter colony losses in the USA: results from the Bee Informed Partnership. *Journal of Apicultural Research* 2012, 51, 115-124.

4. Spleen, A. M.; Lengerich, E. J.; Rennich, K.; Caron, D.; Rose, R.; Pettis, J. S.; Henson, M.; Wilkes, J. T.; Wilson, M.; Stitzinger, J.; Lee, K.; Andree, M.; Snyder, R.; vanEngelsdorp, D.; Bee Informed, P., A national survey of managed honey bee 2011-12 winter colony losses in the United States: results from the Bee Informed Partnership. *Journal of Apicultural Research* 2013, 52.

5. Committee, N. H. B. H. S. C. S. *Report on the National Stakeholders Conference on Honey Bee Health*; 2012; p 72.

6. Pettis, J.; Vanengelsdorp, D.; Cox-Foster, D., Colony collapse disorder working group pathogen sub-group progress report. *American Bee Journal* 2007, 147, 595-597.

7. Boncristiani, H.; Underwood, R.; Schwarz, R.; Evans, J. D.; Pettis, J.; vanEngelsdorp, D., Direct effect of acaricides on pathogen loads and gene expression levels in honey bees Apis mellifera. *Journal of Insect Physiology* 2012, 58, 613-620.

8. Higes, M.; Meana, A.; Bartolome, C.; Botias, C.; Martin-Hernandez, R., Nosema ceranae (Microsporidia), a controversial 21st century honey bee pathogen. *Environmental Microbiology Reports* 2013, 5, 17-29.

9. Cornman, R. S.; Tarpy, D. R.; Chen, Y.; Jeffreys, L.; Lopez, D.; Pettis, J. S.; vanEngelsdorp, D.; Evans, D., Pathogen Webs in Collapsing Honey Bee Colonies. *Plos One* 2012, 7.

10. Pettis, J. S.; vanEngelsdorp, D.; Johnson, J.; Dively, G., Pesticide exposure in honey bees results in increased levels of the gut pathogen Nosema. *Naturwissenschaften* 2012, 99, 153-158.

11. Aliferis, K. A.; Copley, T.; Jabaji, S., Gas chromatography-mass spectrometry metabolite profiling of worker honey bee (Apis mellifera L.) hemolymph for the study of Nosema ceranae infection. *Journal of Insect Physiology* 2012, 58, 1349-1359.

12. Kunieda, T.; Fujiyuki, T.; Kucharski, R.; Foret, S.; Ament, S. A.; Toth, A. L.; Ohashi, K.; Takeuchi, H.; Kamikouchi, A.; Kage, E.; Morioka, M.; Beye, M.; Kubo, T.; Robinson, G. E.; Maleszka, R., Carbohydrate metabolism genes and pathways in insects: insights from the honey bee genome. *Insect Molecular Biology* 2006, 15, 563-576.

13. Kamel, A., Refined Methodology for the Determination of Neon icotinoid Pesticides and Their Metabolites in Honey Bees and Bee Products by Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS). *Journal of Agricultural and Food Chemistry* 2010, 58, 5926-5931.

14. Cox-Foster, D. L.; Conlan, S.; Holmes, E. C.; Palacios, G.; Evans, J. D.; Moran, N. A.; Quan, P.-L.; Briese, T.; Hornig, M.; Geiser, D. M.; Martinson, V.; vanEngelsdorp, D.; Kalkstein, A. L.; Drysdale, A.; Hui, J.; Zhai, J.; Cui, L.; Hutchison, S. K.; Simons, J. F.; Egholm, M.; Pettis, J. S.; Lipkin, W. I., A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* 2007, 318, 283-287.

15. Bromenshenk, J. J.; Henderson, C. B.; Wick, C. H.; Stanford, M. F.; Zulich, A. W.; Jabbour, R. E.; Deshpande, S. V.; McCubbin, P. E.; Seccomb, R. A.; Welch, P. M.; Williams, T.; Firth, D. R.; Skowronski, E.; Lehmann, M. M.; Bilimoria, S. L.; Gress, J.; Wanner, K. W.; Cramer, R. A., Jr., Iridovirus and Microsporidian Linked to Honey Bee Colony Decline. *PLoS ONE* 2010, 5, 1-11.

16. Watanabe, M. E., Colony collapse disorder: Many suspects, no smoking gun. *Bioscience* 2008, 58, 384-388.

17. Dussaubat, C.; Maisonnasse, A.; Crauser, D.; Beslay, D.; Costagliola, G.; Soubeyrand, S.; Kretzchmar, A.; Le Conte, Y., Flight behavior and pheromone changes associated to Nosema ceranae infection of honey bee workers (Apis mellifera) in field conditions. *Journal of Invertebrate Pathology* 2013, 113, 42-51.

18. Foster, L. J., Interpretation of Data Underlying the Link Between Colony Collapse Disorder (CCD) and an Invertebrate Iridescent Virus. *Molecular & Cellular Proteomics* 2011, 10.

19. Honeybee Genome Sequencing, C., Insights into social insects from the genome of the honeybee Apis mellifera. *Nature* 2006, 443, 931-49.

20. Rothberg, J. M.; Leamon, J. H., The development and impact of 454 sequencing. *Nature Biotechnology* 2008, 26, 1117-1124.

21. Cresswell, J. E.; Desneux, N.; vanEngelsdorp, D., Dietary traces of neonicotinoid pesticides as a cause of population declines in honey bees: an evaluation by Hill's epidemiological criteria. *Pest Management Science* 2012, 68, 819-827.

22. Tapparo, A.; Giorio, C.; Solda, L.; Bogialli, S.; Marton, D.; Marzaro, M.; Girolami, V., UHPLC-DAD method for the determination of neonicotinoid insecticides in single bees and its relevance in honeybee colony loss investigations. *Analytical and Bioanalytical Chemistry* 2013, 405, 1007-1014.

23. Garcia-Chao, M.; Jesus Agruna, M.; Flores Calvete, G.; Sakkas, V.; Llompart, M.; Dagnac, T., Validation of an off line solid phase extraction liquid chromatography-tandem mass spectrometry method for the determination of systemic insecticide residues in honey and pollen samples collected in apiaries from NW Spain. *Analytica Chimica Acta* 2010, 672, 107-113.

24. Bacandritsos, N.; Granato, A.; Budge, G.; Papanastasiou, I.; Roinioti, E.; Caldon, M.; Falcaro, C.; Gallina, A.; Mutinelli, F., Sudden deaths and colony population decline in Greek honey bee colonies. *Journal of Invertebrate Pathology* 2010, 105, 335-340.

25. Runckel, C.; Flenniken, M. L.; Engel, J. C.; Ruby, J. G.; Ganem, D.; Andino, R.; DeRisi, J. L., Temporal Analysis of the Honey Bee Microbiome Reveals Four Novel Viruses and Seasonal Prevalence of Known Viruses, Nosema, and Crithidia. *Plos One* 2011, 6.

26. Johnson, R. M.; Evans, J. D.; Robinson, G. E.; Berenbaum, M. R., Changes in transcript abundance relating to colony collapse disorder in honey bees (Apis mellifera). *Proceedings of the National Academy of Sciences of the United States of America* 2009, 106, 14790-14795.

27. Engel, P.; Martinson, V. G.; Moran, N. A., Functional diversity within the simple gut microbiota of the honey bee. *Proceedings of the National Academy of Sciences of the United States of America* 2012, 109, 11002-11007.

28. Cardoen, D.; Ernst, U. R.; Van Vaerenbergh, M.; Boerjan, B.; de Graaf, D. C.; Wenseleers, T.; Schoofs, L.; Verleyen, P., Differential Proteomics in Dequeened Honeybee Colonies Reveals Lower Viral Load in Hemolymph of Fertile Worker Bees. *Plos One* 2011, 6.

29. 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.; Moore, B. S.; Laskin, J.; Bandeira, N.; Dorrestein, P. C., Mass spectral molecular networking of living microbial colonies. *Proceedings of the National Academy of Sciences of the United States of America* 2012, 109.

30. Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P.; Bruns, N.; Glukhov, E.; Wodtke, A.; de Felicio, R.; Fenner, A.; Wong, W. R.; Linington, R. G.; Zhang, L.; Debonsi, H. M.; Gerwick, W. H.; Dorrestein, P. C., Molecular networking as a dereplication strategy. *Journal of Natural Products* 2013, submitted.

31. Tautenhahn, R.; Patti, G. J.; Kalisiak, E.; Miyamoto, T.; Schmidt, M.; Lo, F. Y.; McBee, J.; Baliga, N. S.; Siuzdak, G., metaXCMS: Second-Order Analysis of Untargeted Metabolomics Data. *Analytical Chemistry* 2011, 83, 696-700.

32. Benton, H. P.; Wong, D. M.; Trauger, S. A.; Siuzdak, G., XCMS2: Processing tandem mass spectrometry data for metabolite identification and structural characterization. *Analytical Chemistry* 2008, 80, 6382-6389.

33. Rath, C. M.; Dorrestein, P. C., The bacterial chemical repertoire mediates metabolic exchange within gut microbiomes. *Current Opinion in Microbiology* 2012, 15, 147-154.

34. Schoenian, I.; Spiteller, M.; Ghaste, M.; Wirth, R.; Herz, H.; Spiteller, D., Chemical basis of the synergism and antagonism in microbial communities in the nests of leaf-cutting ants. *Proceedings of the National Academy of Sciences of the United States of America* 2011, 108, 1955-1960.

35. Nicholson, J. K.; Holmes, E.; Wilson, I. D., Gut microorganisms, mammalian metabolism and personalized health care. *Nature Reviews Microbiology* 2005, 3, 431-438.

36. Rickard, A. H.; Underwood, A. J.; Nance, W., A Holistic View of Interspecies Bacterial Interactions Within Human Dental Plaque. *Oral Microbial Ecology: Current Research and New Perspectives* 2013, 97-110.

37. Watrous, J. D.; Phelan, V. V.; Hsu, C.-C.; Moree, W. J.; Duggan, B. M.; Alexandrov, T.; Dorrestein, P. C., Microbial metabolic exchange in 3D. *Isme Journal* 2013, 7, 770-780.

38. Phelan, V. V.; Liu, W.-T.; Pogliano, K.; Dorrestein, P. C., Microbial metabolic exchange-the chemotype-to-phenotype link. *Nature Chemical Biology* 2012, 8, 26-35.

39. Morris, B. E. L.; Henneberger, R.; Huber, H.; Moissl-Eichinger, C., Microbial syntrophy: interaction for the common good. *Fems Microbiology Reviews* 2013, 37, 384-406.