SUPPORTING INFORMATION

Molecular Networking as a Dereplication Strategy

Jane Y. Yang,[†°] *Laura M. Sanchez,*[‡°] *Christopher M. Rath,*[‡] *Xueting Liu,*[§] *Paul D. Boudreau,*[⊥] *Nicole Bruns,*[⊥] *Evgenia Glukhov,*[⊥] *Anne Wodtke,*[⊥] *Rafael de Felicio,*[⊥∥] *Amanda Fenner,*[⊥] *Weng Ruh Wong,*[▽] *Roger G. Linington,*[▽] *Lixin Zhang*[§], *Hosana M. Debonsi,*[∥] *William H. Gerwick,*[⊥] *Pieter C. Dorrestein*[*†‡]

[†]Department of Chemistry & Biochemistry, University of California San Diego, La Jolla, California 92093, United States

[‡]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California 92093, United States

[§]Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology Chinese Academy of Sciences, Beijing, 100190, China

[⊥]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093, United States

[∥]Nucleo de Pesquisaem Produtos Naturais e Sinteticos - Departamento de Fisica e Quimica - Faculdade de Ciencias Farmaceuticas de Ribeirao Preto, Universidade de Sao Paulo, Av. Do Café, s/n, Campus Universitario, CEP 14040-903, Ribeirao Preto, Sao Paulo, Brazil

▽Department of Chemistry and Biochemistry, University of California Santa Cruz, Santa Cruz, California 95064, United States

°Equal Contributions

Cyanobacterial collection information.

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).

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).

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).

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).

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.

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.

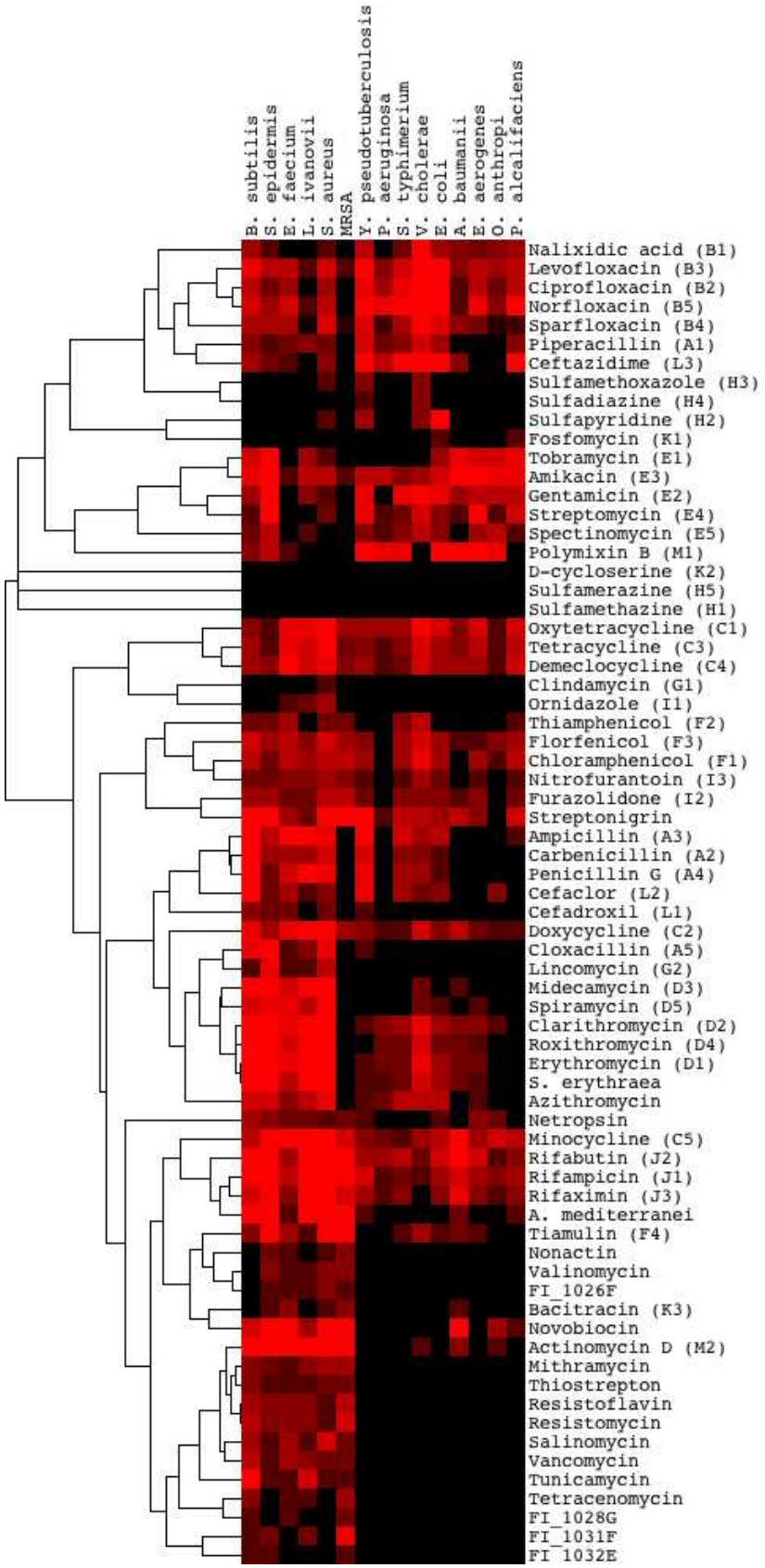
The strain SRM7 was isolated using oatmeal agar from a sediment sample collected in the South China Sea at 20°9.795′ N and 118°18.124′ E, 2733m below sea level in April 2010. Spore morphology and ornamentation were determined by scanning electron microscopy. The organism forms a well-developed, branched, 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.


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

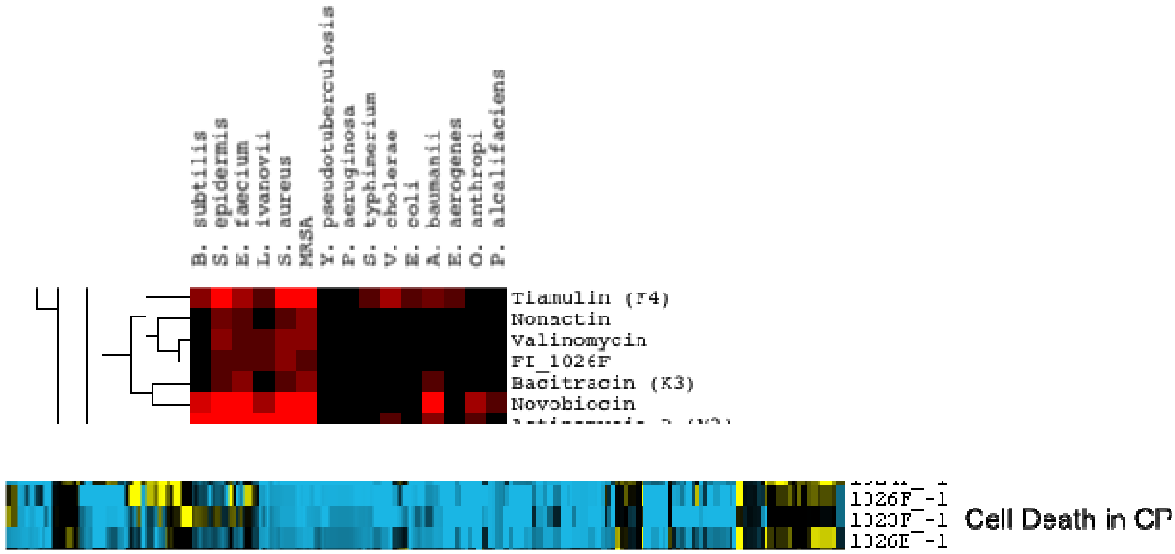


Figure S1. BioMAP[4] and cytological profiles[5] (CPs) of known antibiotics and *Dietzia* sp. FI-1026 extracts. In BioMAP, the extract clusters closely with valinomycin, indicative of the structure being a large cyclic peptide. In CP the extract causes death for HeLa cells.

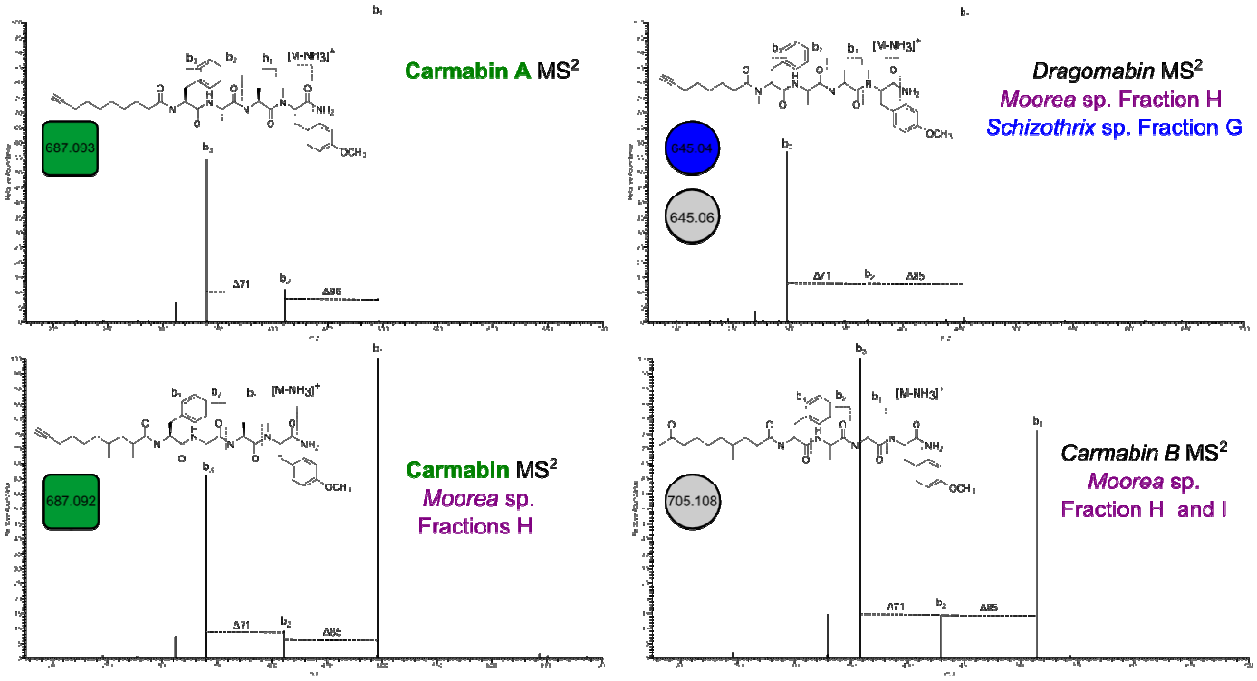


Figure S2. Annotation of carmabin MS/MS spectra. Manual inspection of the carmabin A cluster showed that the crude *Moorea* sp. fraction and the carmabin A seed have similar retention times (within five milliseconds).

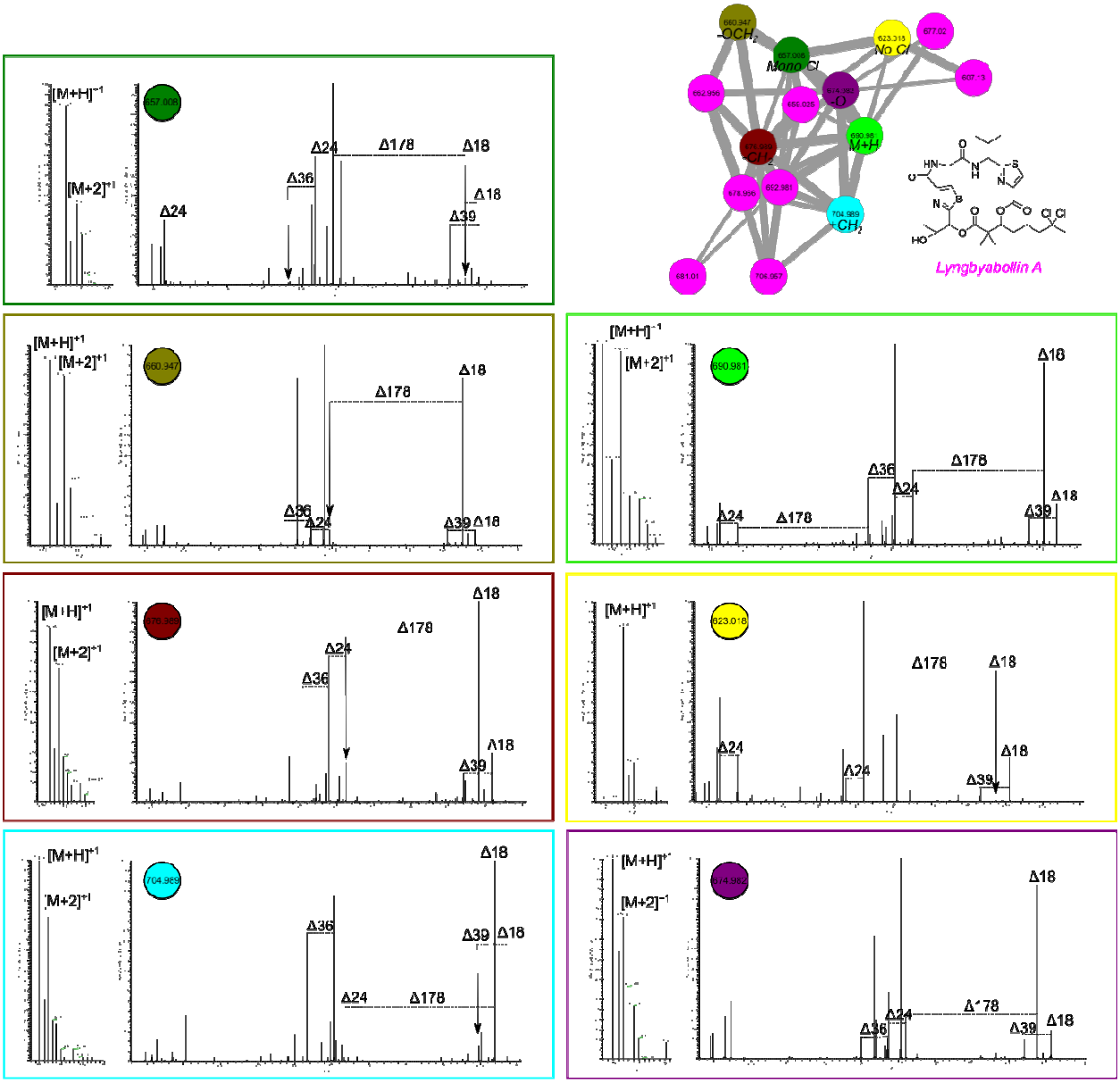
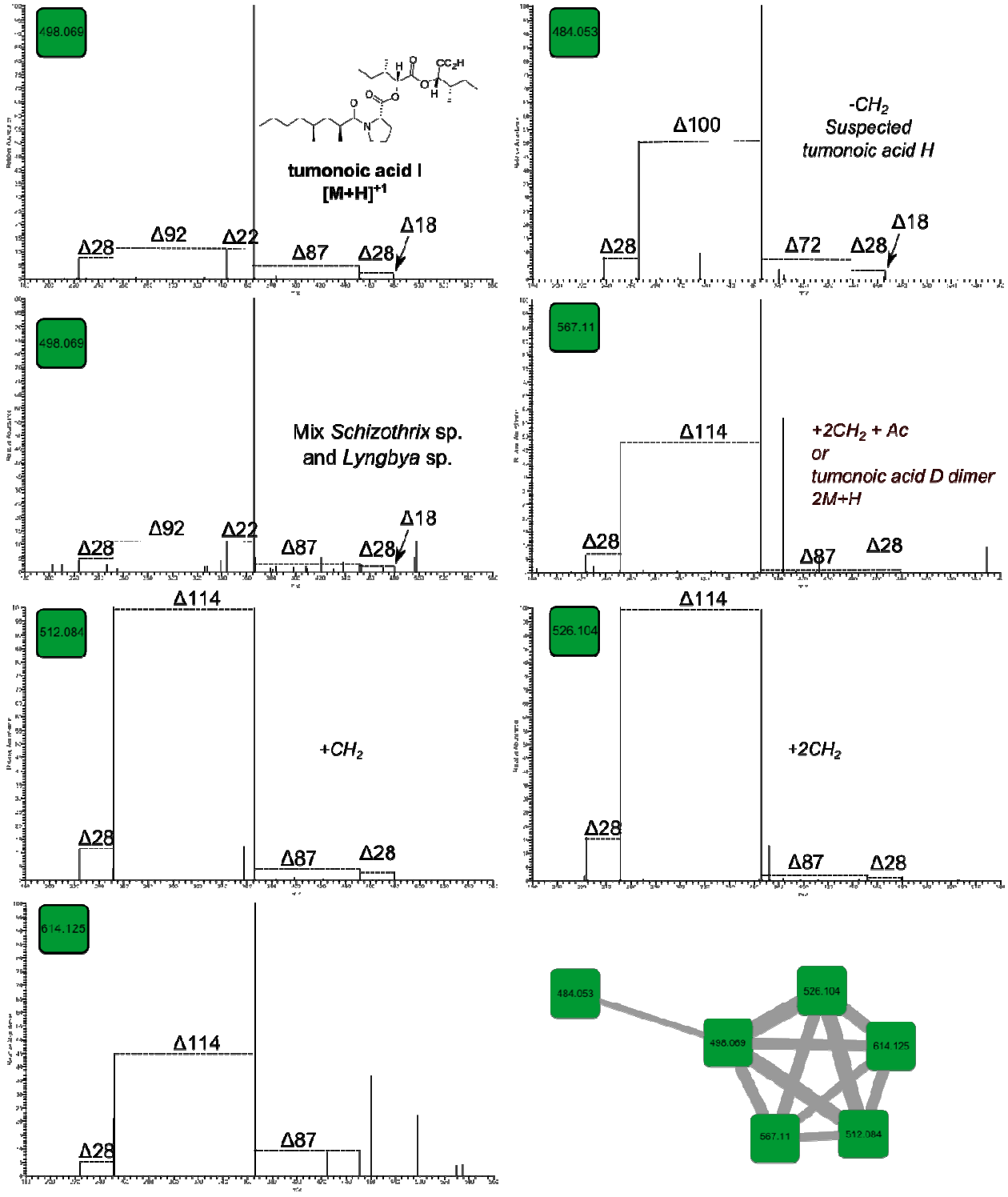


Figure S3. Annotation of lyngbyabellin MS/MS spectra and isotopic distribution of precursor ions. Two clusters from *Moorea bouillonii*, known to produce lyngbyabellins and apratoxins, did not contain seed compounds. The MS/MS fragmentation, precursor masses, and isotopic pattern for one cluster were highly suggestive of the known metabolite lyngbyabellin A.[6] This cluster suggested novel analogs, including a rare monochlorinated species, a deoxygenated species, a demethylated species, a demethylated and deoxygenated 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.


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

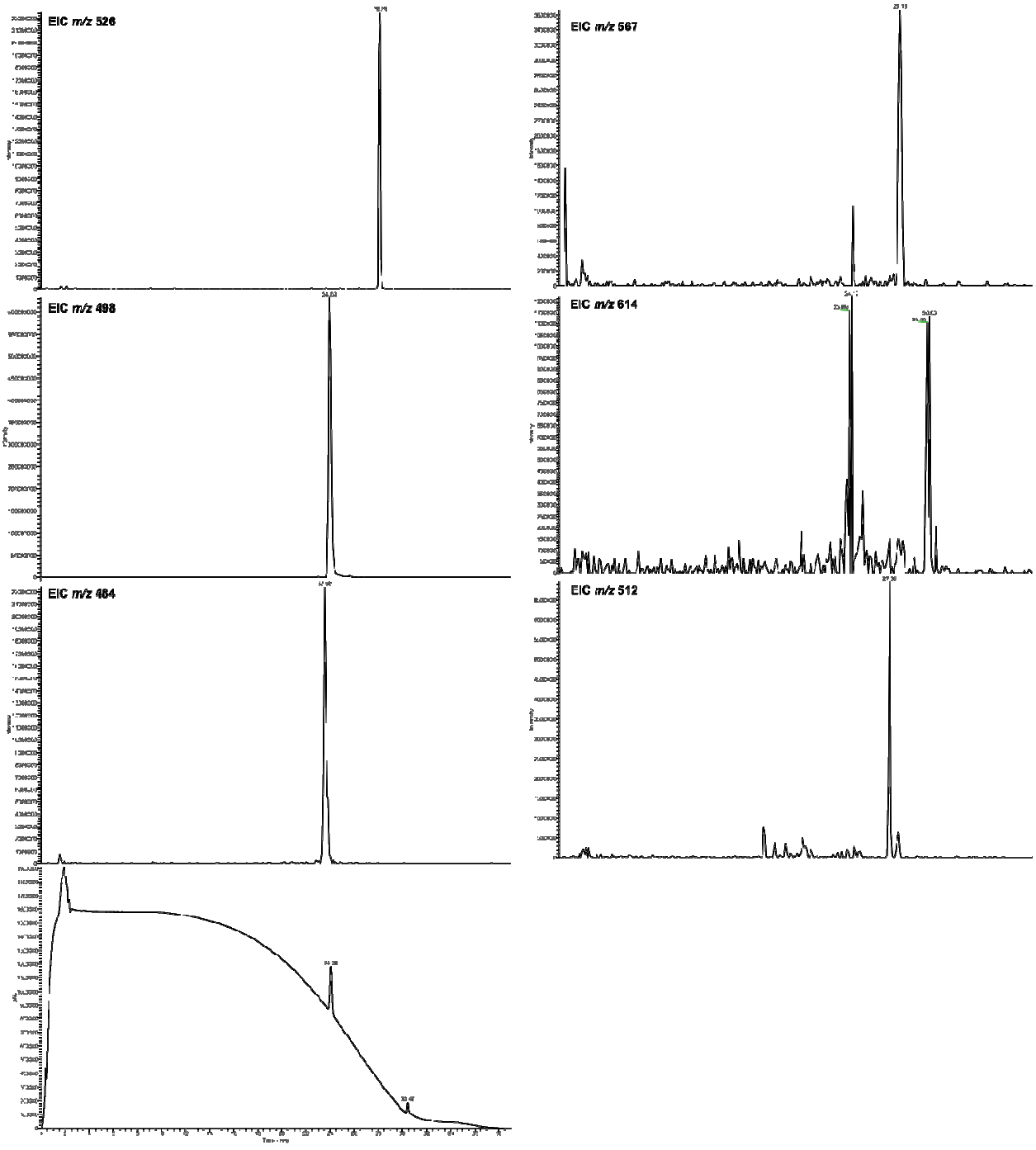


Figure S4. Annotation of tumonoic acid I MS/MS spectra and EIC of precursor masses found to cluster with the seed. Tumonoic acid I was found in a mixed collection of two cyanobacteria, *Lyngbya* sp. and *Schizothrix* sp., from Papua New Guinea.[7] Interestingly, within the seed LC-MS/MS run, there appeared 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.

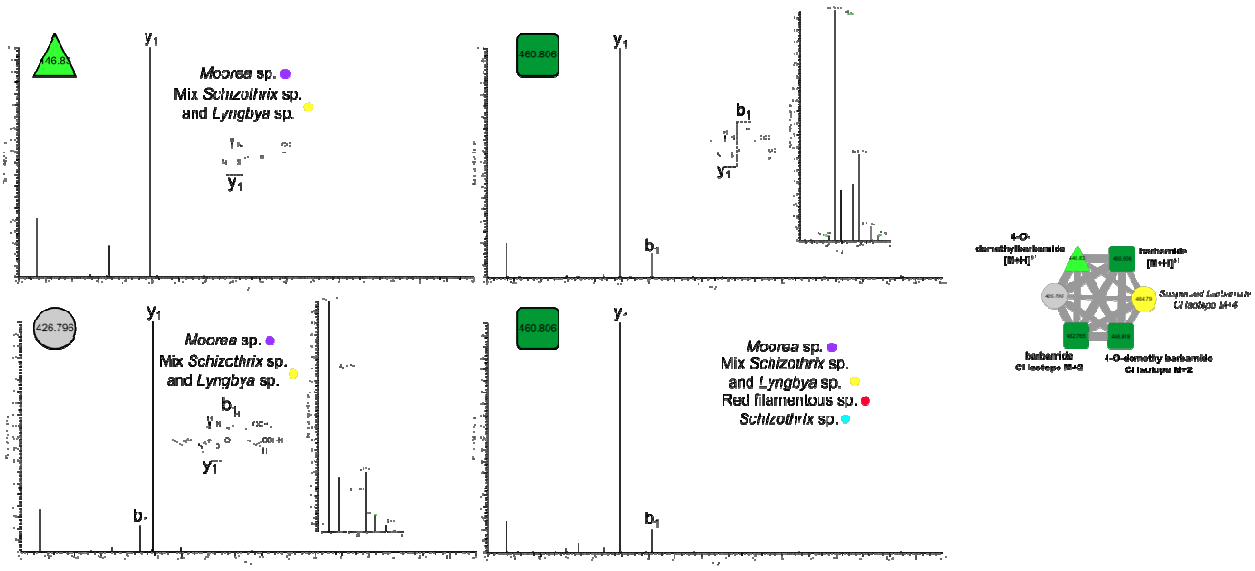


Figure S5. Annotation of barbamide MS/MS spectra. Barbamide was found in multiple collections of cyanobacteria.[8] In addition to this seed, the chlorine isotope was detected as well the recently described 4-*O*-demethylbarbamide and its complex chlorine isotopic pattern.[9] While this analog was not 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 *b* and *y* ions.[10]

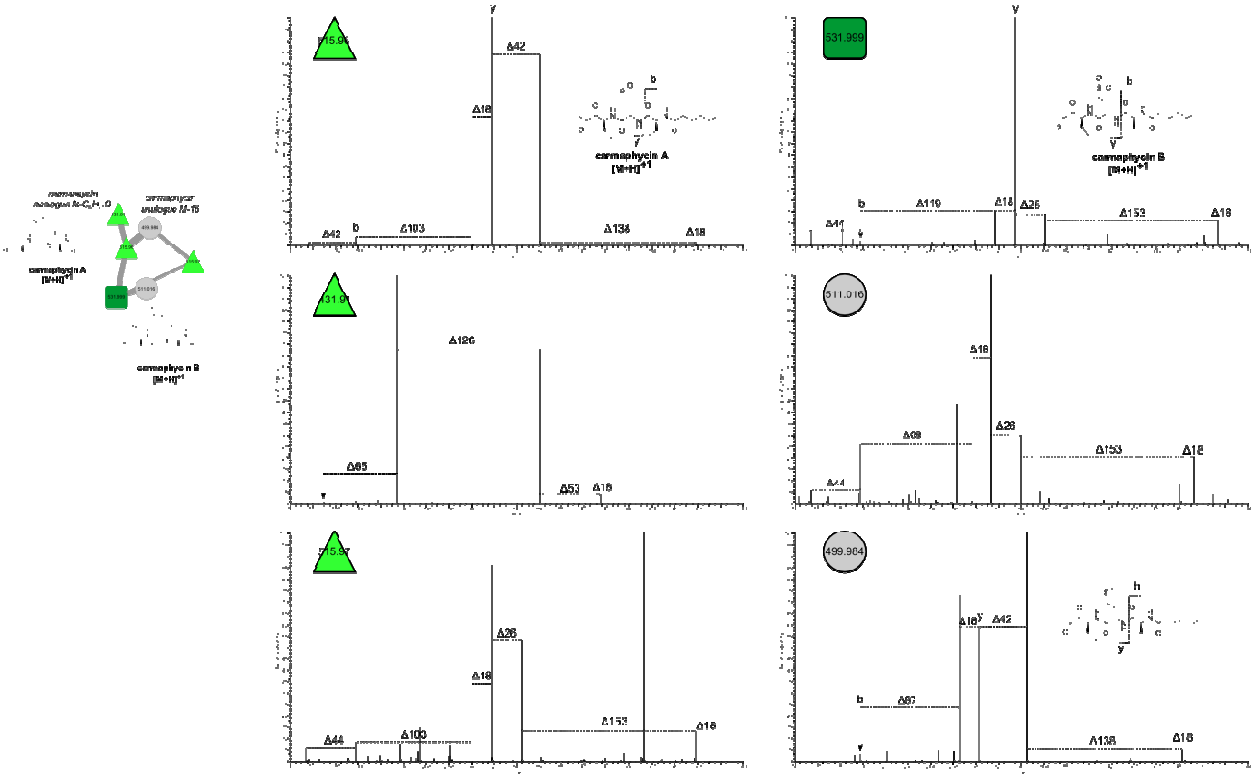


Figure S6. Annotation of carmaphycin A and B MS/MS spectra. Carmaphycin A and B were used as seeds and only carmaphycin B was found in a one collection of cyanobacteria from Panama.[11] Through networking and comparison of retention times, the presence of carmaphycin B was confirmed in the crude sample. Additionally, there were two nodes suggestive of novel carmaphycin analogs. One node was suggestive of non-oxidized methionine analog, with *b* and *y* 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.

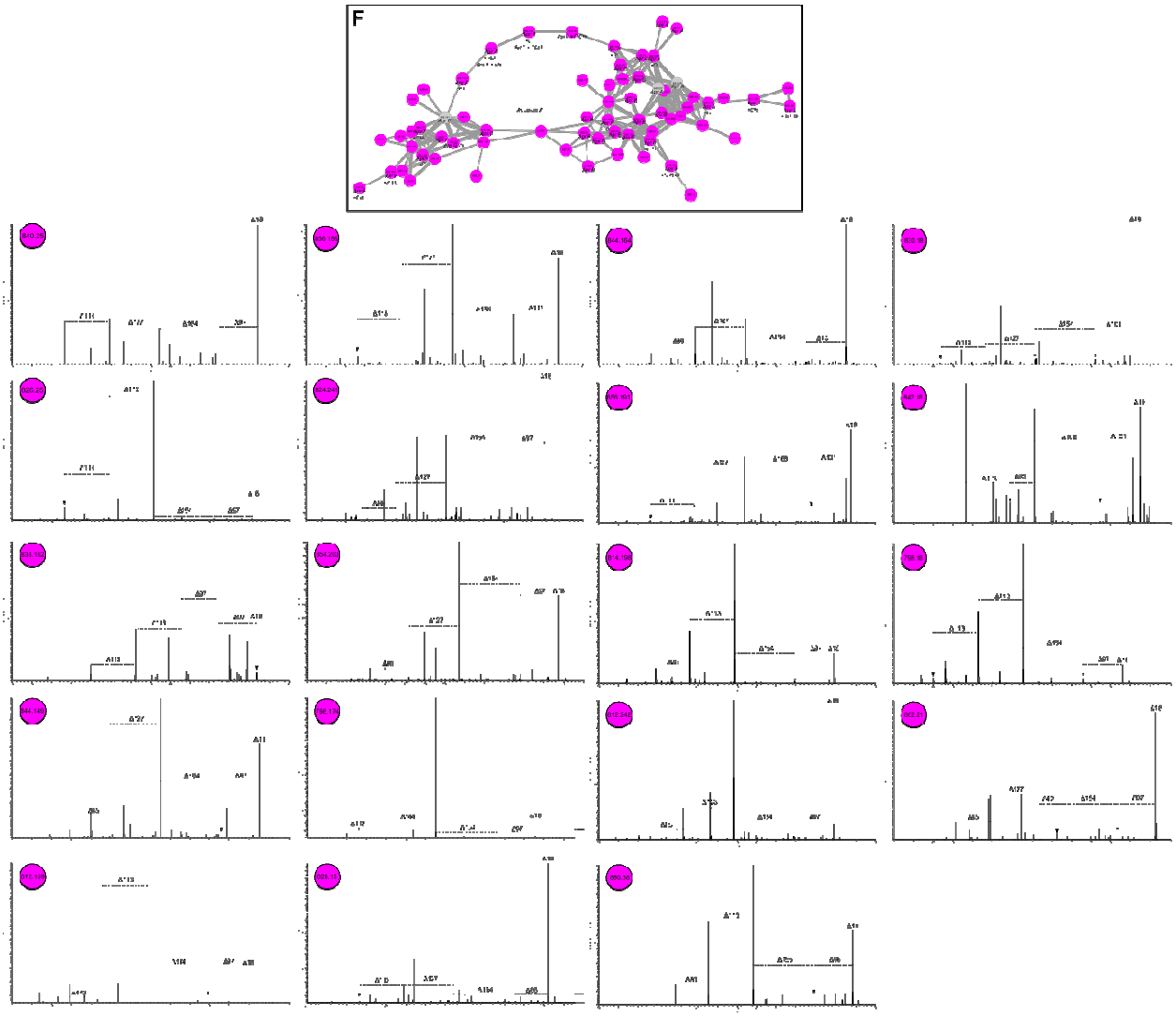


Figure S7. Annotation of apratoxin MS/MS spectra. Two clusters from *Moorea bouillonii*, known to produce lyngbyabellins and apratoxins, did not contain seed compounds. The other cluster was large and likely contained apratoxins A and B as well as masses suggesting the analog structures of apratoxins 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.

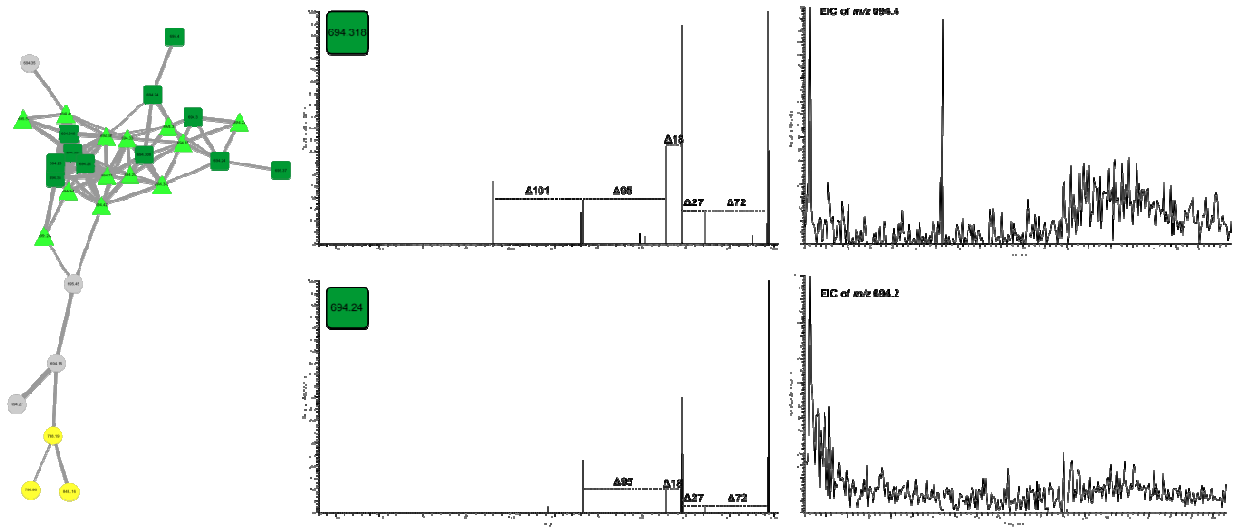


Figure S8. Annotation of precursor mass *m/z* 694 MS/MS spectra and EIC.

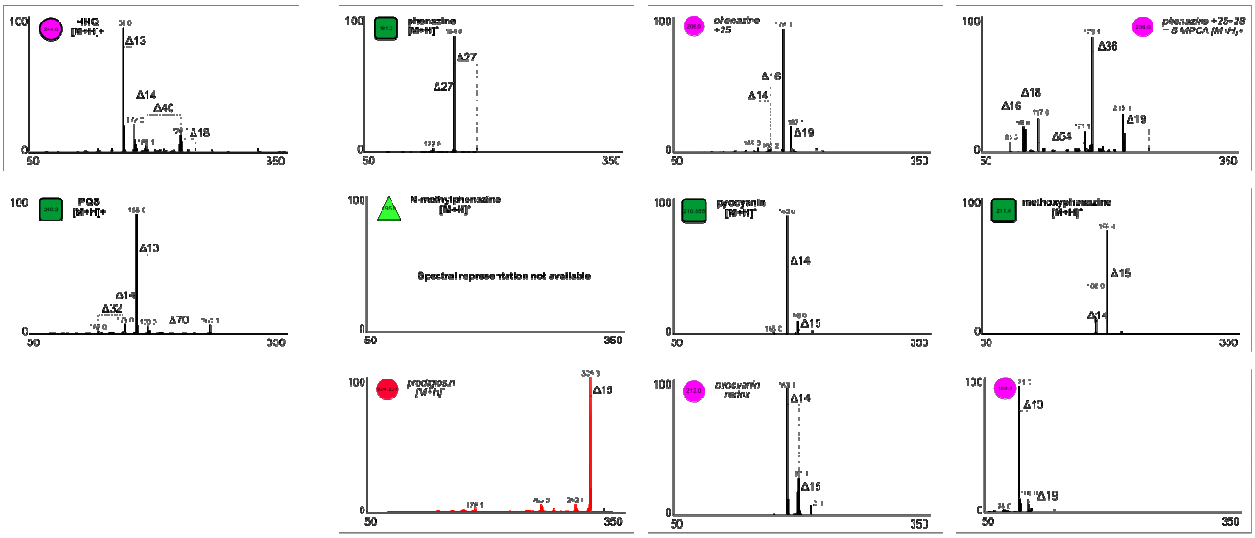


Figure S9. MS/MS spectra corresponding to phenazine and quinolone nodes. Molecular networking analysis of data-independent nanoDESI-MS/MS of *Pseudomonas aeruginosa* PAO1 and PA14[2] with six reference spectra – hydroxyphenazine, methoxyphenazine, *N*-methylphenazine, phenazine, *Pseudomonas* quinolone signal (PQS), and pyocyanin – yielded 3 perfect matches to methoxyphenazine, pyocyanin, PQS, and phenazine. Pyocyanin (211.0866 $[M+H]^+$ calcd) 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 $[M+H]^+$ calcd), 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.

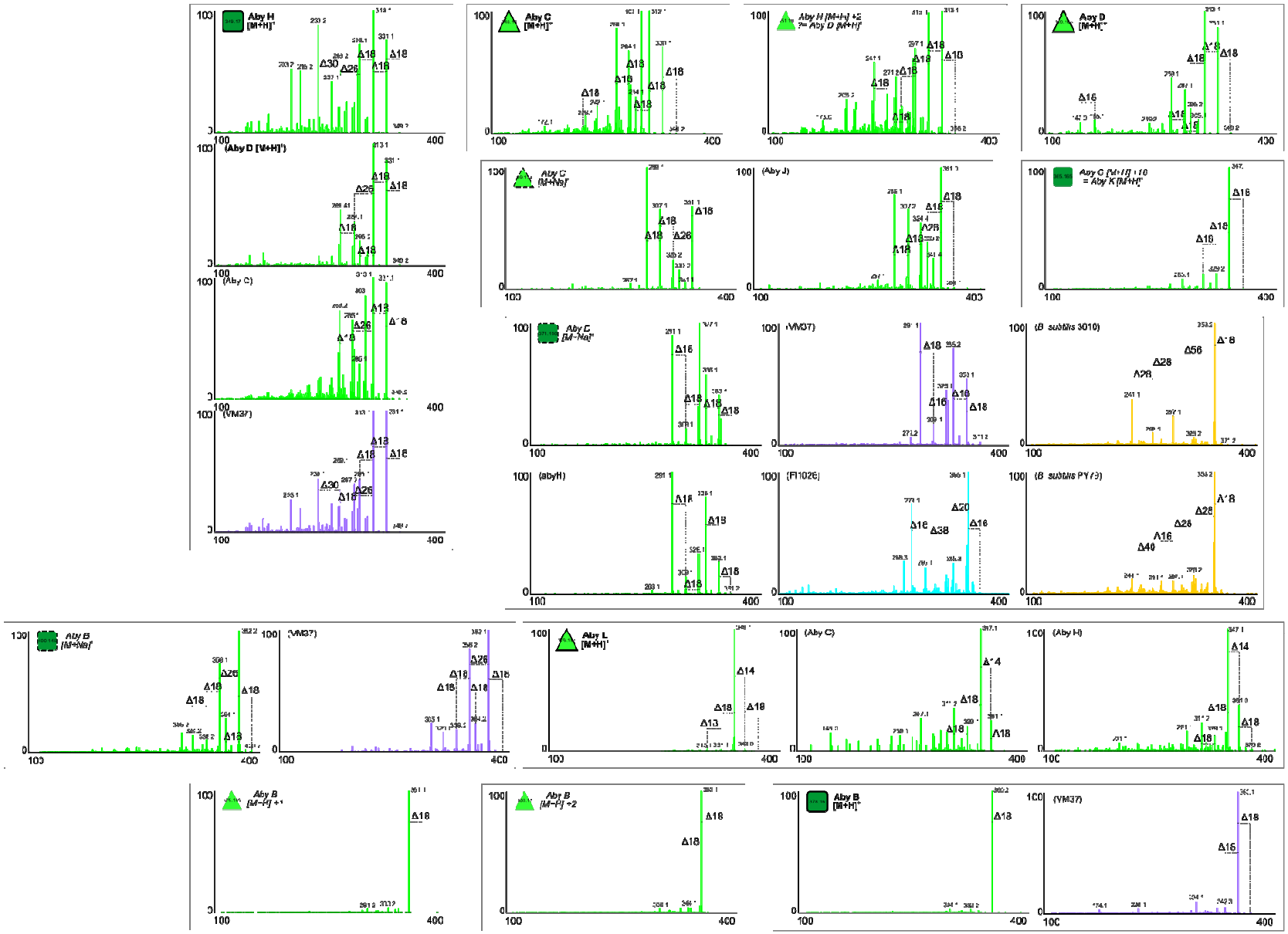


Figure S10. MS/MS spectra corresponding to abyssomicin nodes. The bacteria network included MS/MS spectra directly collected from a *Verrucosispora* sp. MS100128 colony grown on agar media, which is known to produce abyssomicins. The MS/MS spectra for each of the NMR-pure standards were not parsed out 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.

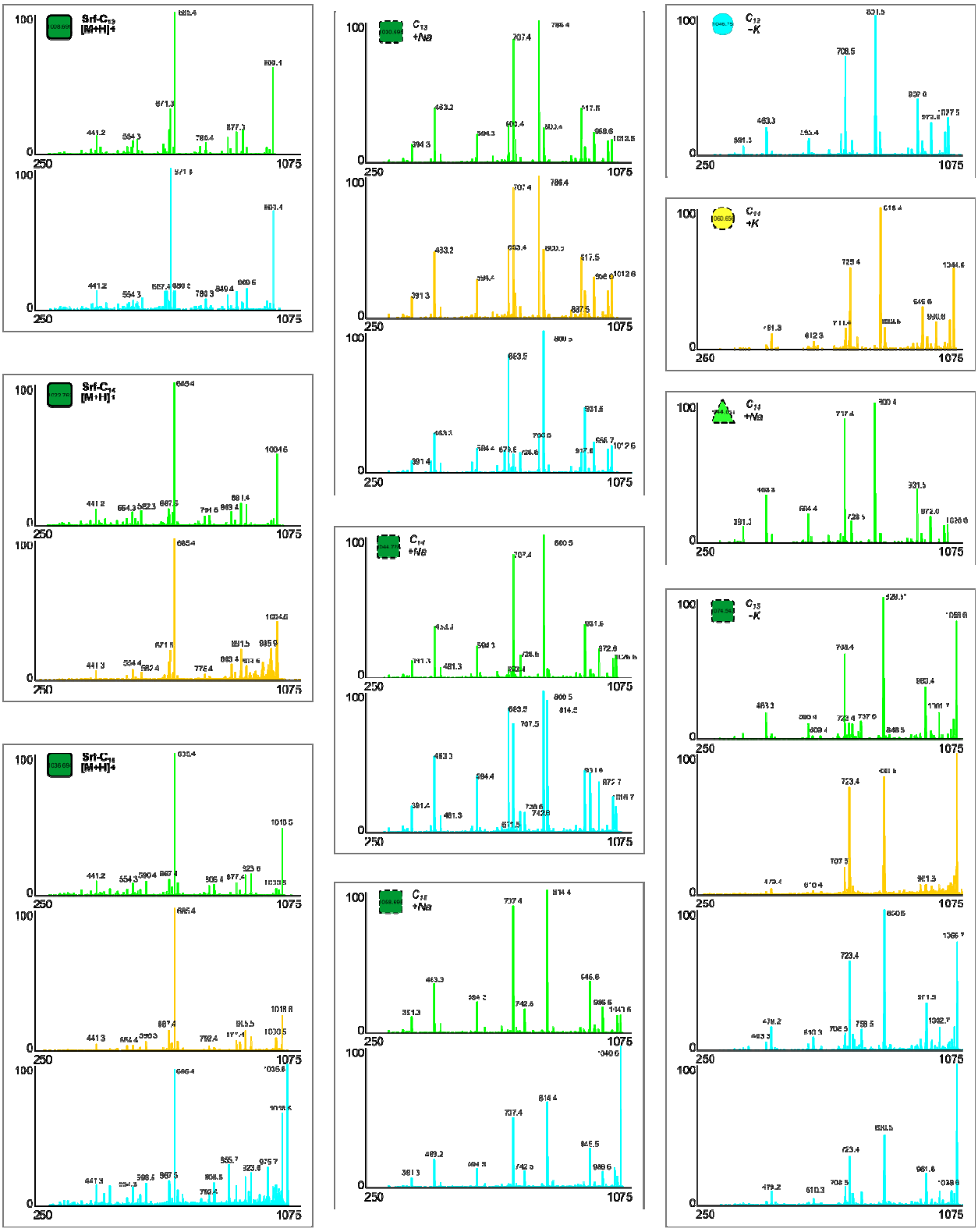


Figure S11. MS/MS spectra corresponding to nodes in the 1+ surfactin cluster. A noticeable computational limitation meriting further development was multiple nodes for precursor masses

that are within 0.5 Da, as seen by multiple nodes for 1008.6597, 1022.6753, and 1036.6910 [M+H]+ calcd (Figure 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.

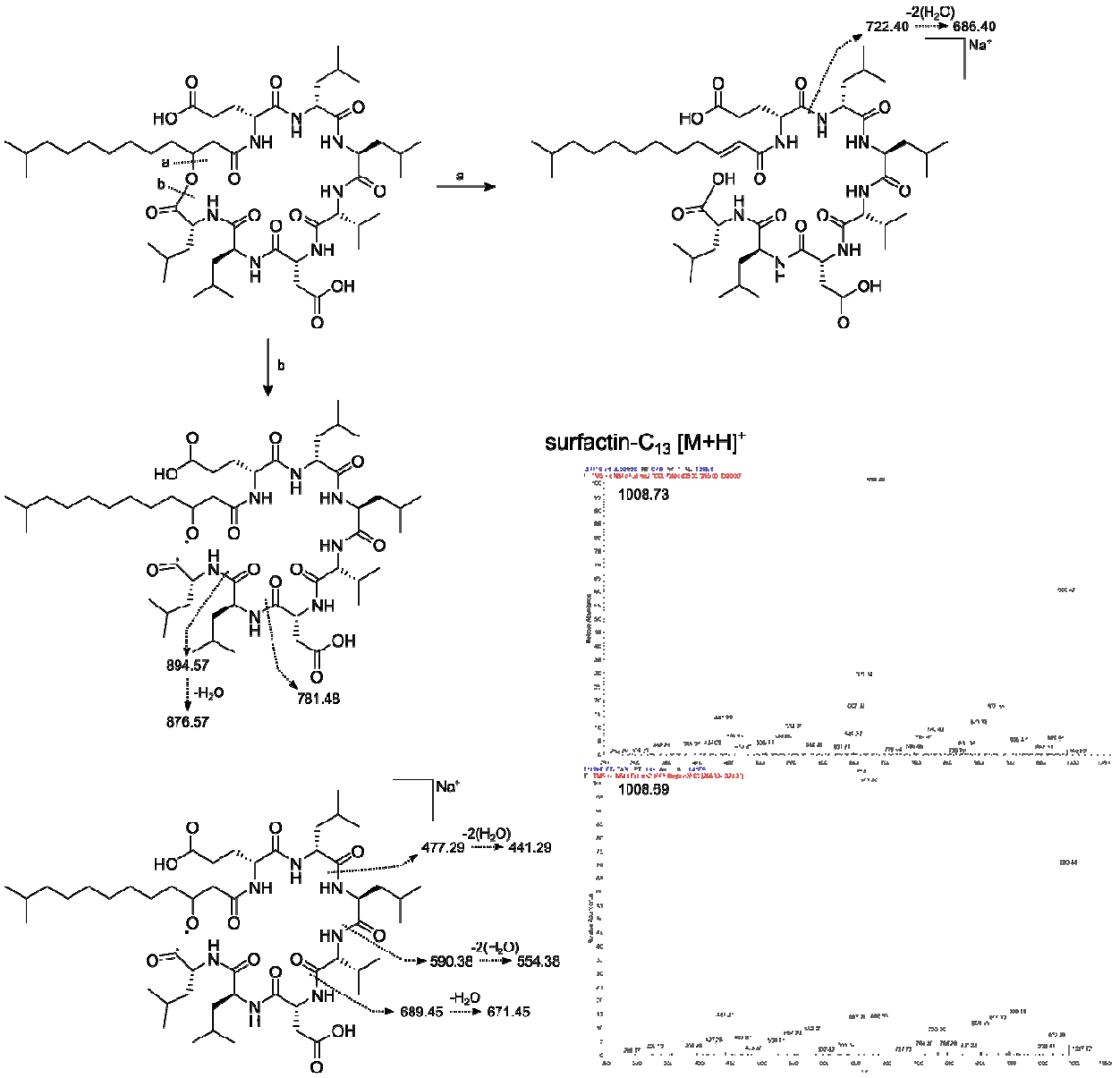


Figure S12. Annotation of surfactin-C_{13} $[M+H]^+$ MS/MS.

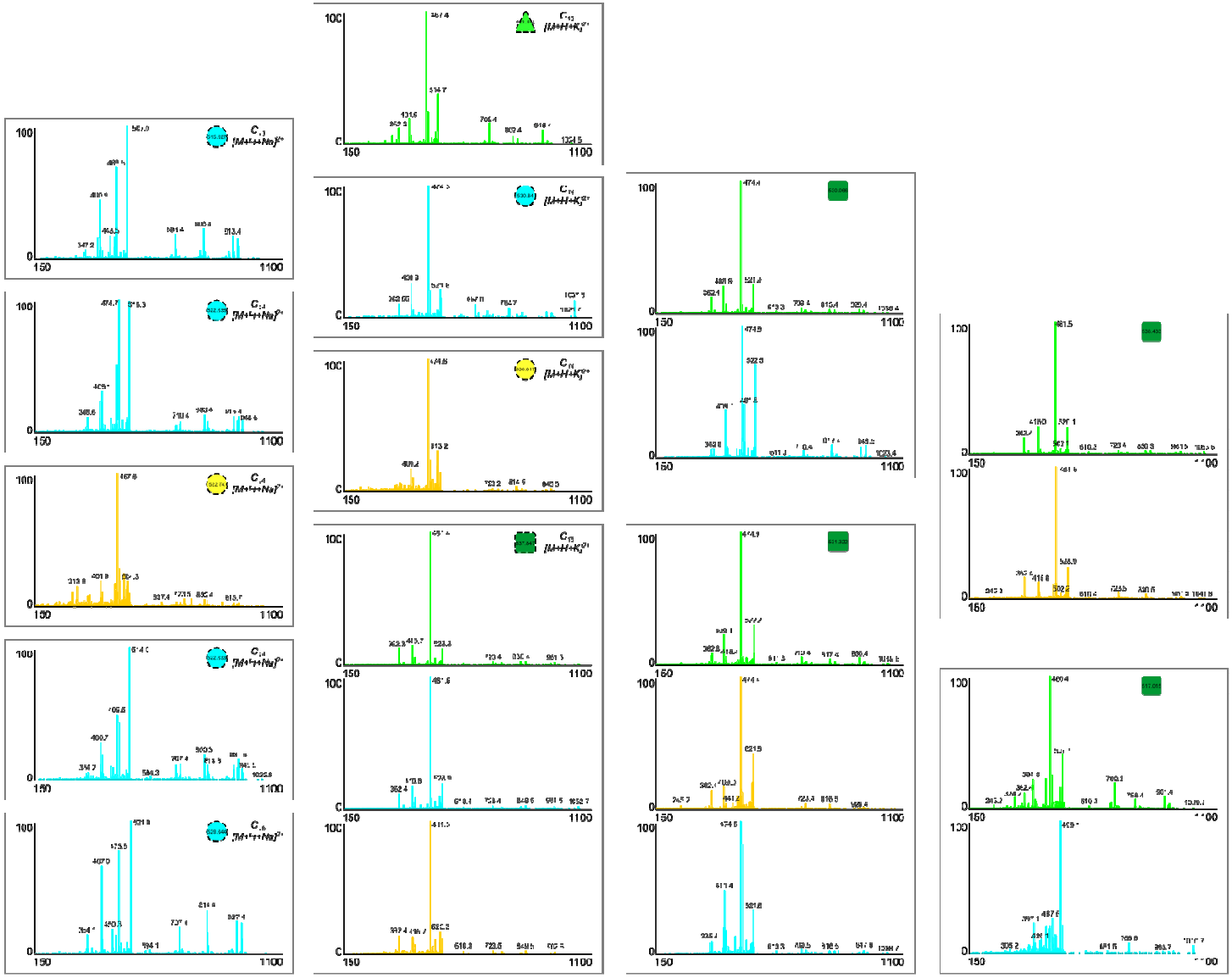


Figure S13. MS/MS spectra corresponding to nodes in the 2+ surfactin cluster.

References

(1)Engene, N.; Rottacker, E. C.; Kastovsky, J.; Byrum, T.; Choi, H.; Ellisman, M. H.; Komarek, J.; Gerwick, W. H. *Int. J. Syst. Evol. Microbiol.* **2012**, *62*, 1171-1178.

(2)Rath, C. M.; Yang, J. Y.; Alexandrov, T.; Dorrestein, P. C. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1167-1176.

(3)Yang, Y. L.; Xu, Y. Q.; Straight, P.; Dorrestein, P. C. *Nat. Chem. Biol.* **2009**, *5*, 885-887.

(4)Wong, W. R.; Oliver, A. G.; Linington, R. G. *Chem. Biol.* **2012**, *19*, 1483-1495.

(5)Schulze, C. J.; Bray, W. M.; Woerhmann, M. H.; Stuart, J.; Lokey, R. S.; Linington, R. G. *Chem. Biol.* **2013**, *20*, 285-295.

(6)Luesch, H.; Yoshida, W. Y.; Moore, R. E.; Paul, V. J.; Mooberry, S. L. *J. Nat. Prod.* **2000**, *63*, 611-615.

(7)Clark, B. R.; Engene, N.; Teasdale, M. E.; Rowley, D. C.; Matainaho, T.; Valeriote, F. A.; Gerwick, W. H. *J. Nat. Prod.* **2008**, *71*, 1530-1537.

(8)Orjala, J.; Gerwick, W. H. *J. Nat. Prod.* **1996**, *59*, 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. *Org. Lett.* **2012**, *14*, 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. *Tetrahedron* **2000**, *56*, 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. *ChemBioChem* **2012**, *13*, 810-817.

(12)Gu, Z.-M.; Wang, L.-Q.; Wu, J. *Nat. Prod. J.* **2011**, *1*, 135-145.

(13)Luesch, H.; Yoshida, W. Y.; Moore, R. E.; Paul, V. J.; Corbett, T. H. *J. Am. Chem. Soc.* **2001**, *123*, 5418-5423.

(14)Luesch, H.; Yoshida, W. Y.; Moore, R. E.; Paul, V. J. C. P. s. *Bioorg. Med. Chem.* **2002**, *10*, 1973-1978.

(15)Matthew, S.; Schupp, P. J.; Luesch, H. *J. Nat. Prod.* **2008**, *71*, 1113-1116.

(16)Tidgewell, K.; Engene, N.; Byrum, T.; Media, J.; Doi, T.; Valeriote, F. A.; Gerwick, W. H. *ChemBioChem* **2010**, *11*, 1458-1466.

(17)Gutierrez, M.; Suyama, T. L.; Engene, N.; Wingerd, J. S.; Matainaho, T.; Gerwick, W. H. *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. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 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. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1856-1866.