Title: Coupling MALDI-TOF mass spectrometry protein and specialized metabolite analyses to rapidly discriminate bacterial function.

Authors: Chase M. Clark[1], Maria S. Costa[2], Laura M. Sanchez[1*], Brian T. Murphy[1*]

Affiliations:

[1]Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, 833 South Wood Street (MC 781), Room 539, Chicago, IL, United States; [2]Faculty of Pharmaceutical Sciences, University of Iceland, Hagi, Hofsvallagata 53, IS-107 Reykjavík, Iceland

Supplementary Information

Table of Contents

Table S4:

Annotated peak list of data presented in Fig. 1B

Text S1:

Coding and logic of Metabolite Association Networks (MANs).

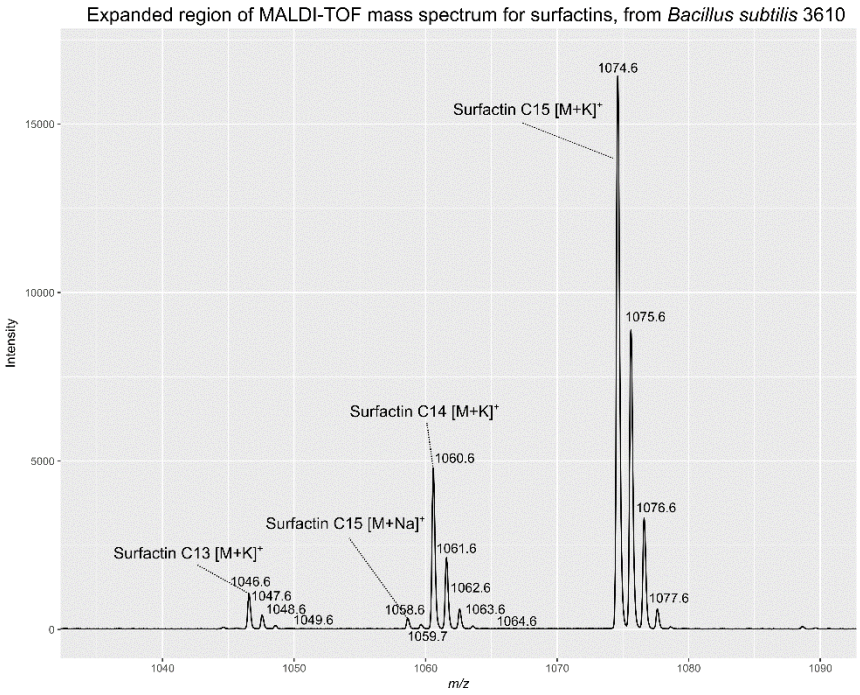
Text S2:

Expanded methods section.

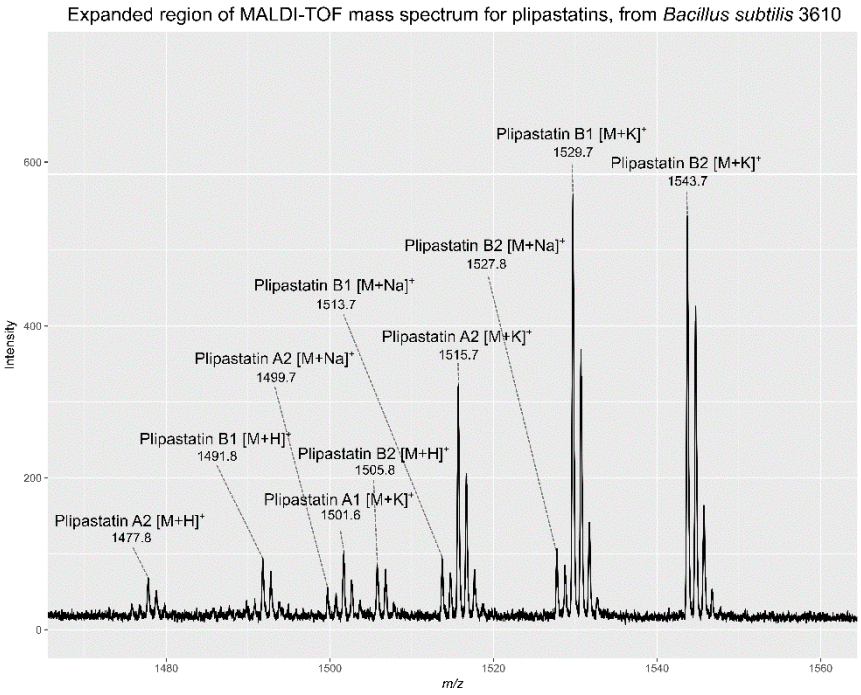
Supplementary References

Figure S1: Annotation of surfactin and plipastatin analogs in *Bacillus subtilis* 3610 MS spectrum. Figure S1A and S1B show the averaged mass spectrum for ten *B. subtilis* 3610 MALDI-MS replicates.

A



B



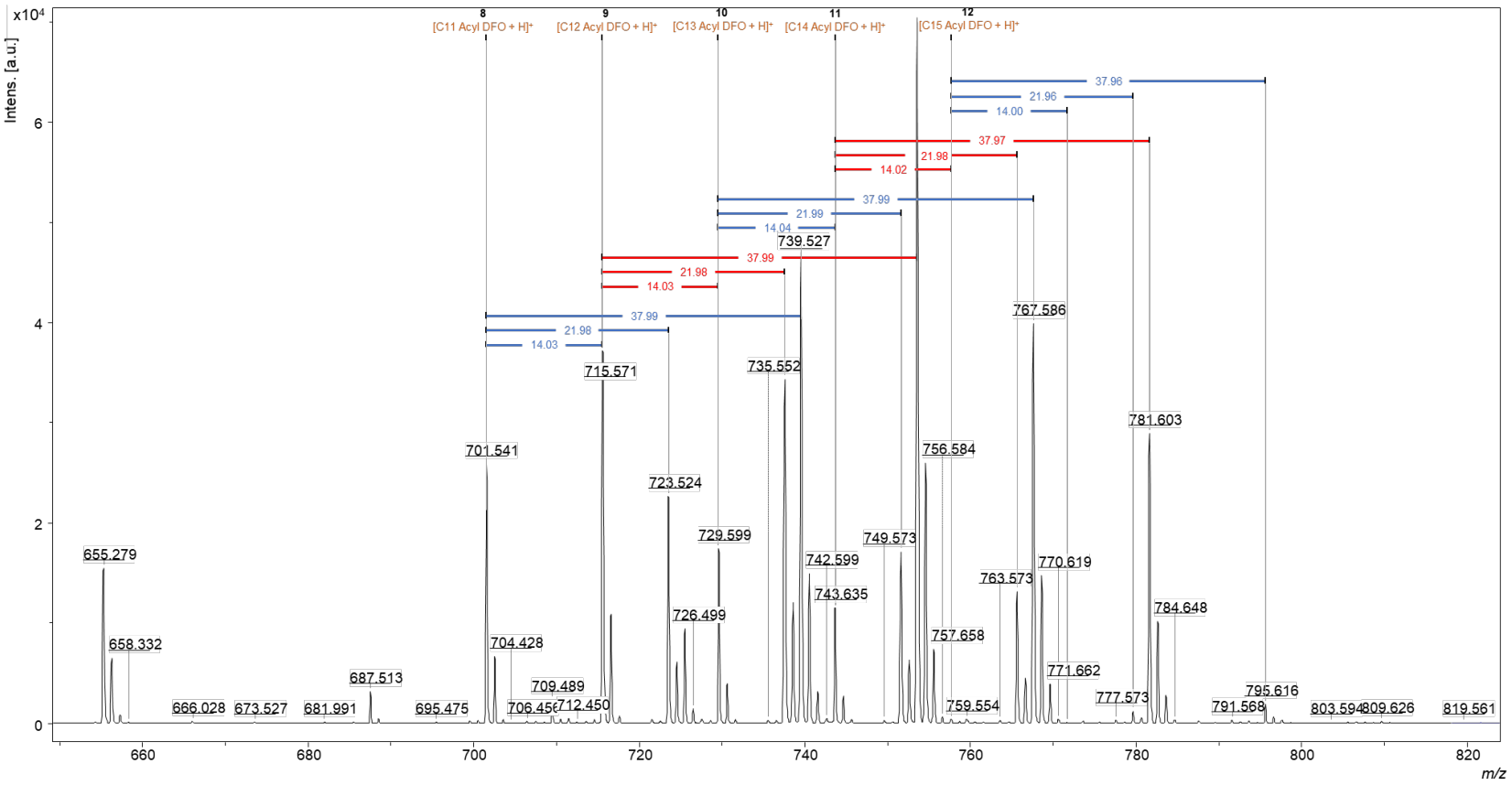
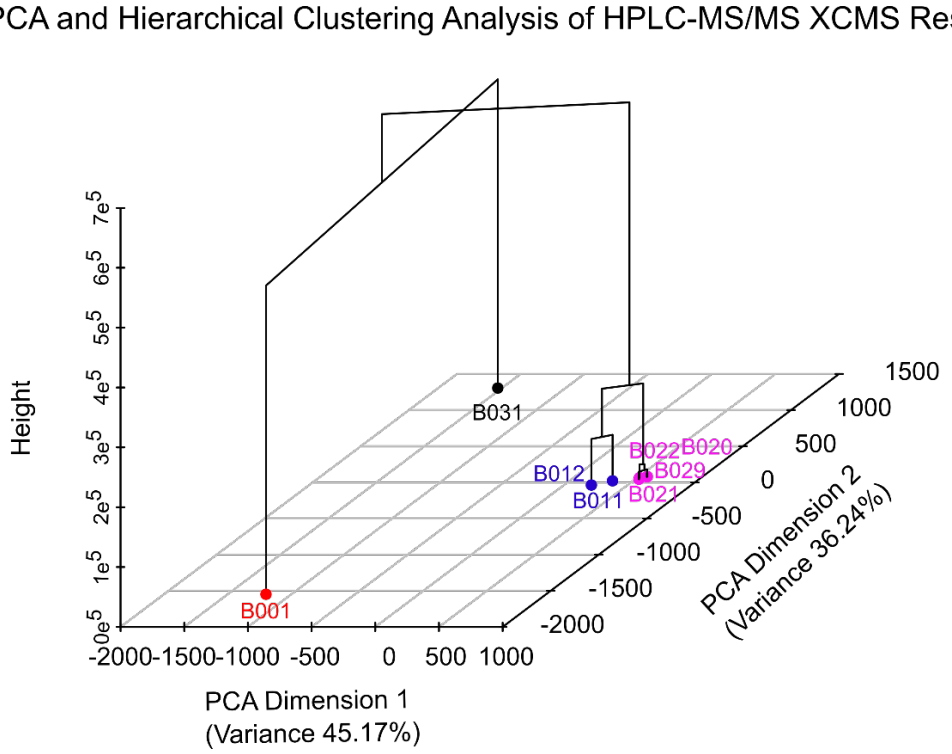


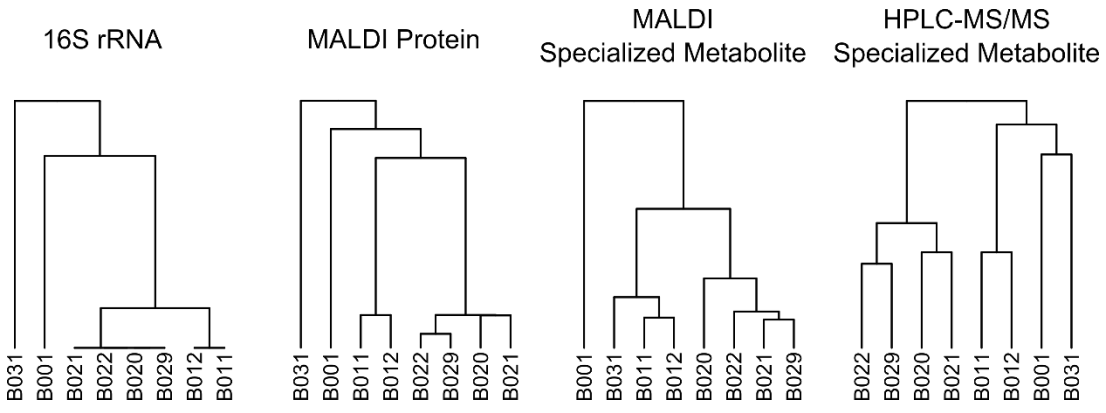
Figure S2: Annotation of *Micromonospora* B031 MALDI spectrum showing intact acyl-desferrioxamine analogs. Annotation of the MALDI spectrum of a single technical replicate of *Micromonospora* isolate B031. Analysis and annotation was performed using Bruker Daltonics' flexAnalysis software v3.4 (build 50). All labeled peaks had a signal-to-noise resolution greater than 3:1 and represent the non-chelated desferrioxamine form. All intact molecules shown above represent $[M+H]^+$, $[M+Na]^+$, and $[M+K]^+$ adducts. The core structure may be found in Fig. S6A and a molecular weight summary in Table S1. Note: In the figure, "desferrioxamine" has been abbreviated as "DFO".

Figure S3: MALDI-MS protein profiling of eight *Micromonospora* isolates provided similar groupings as both HPLC-MS/MS and 16S rRNA gene sequencing analysis. (**a**) Principle component analysis was performed on the log-transformed HPLC-MS/MS dataset, and hierarchical clustering results were overlaid. XCMS settings and output data can be found at https://doi.org/10.17632/ysrtr9c5s7.1 (**b**) Hierarchical clustering of MALDI specialized metabolite and HPLC-MS/MS data provided similar groupings as 16S rRNA gene sequencing analysis, however minor differences in metabolism are masked. Only by manual inspection or MAN analysis, with inverse weighting of common features, were we able to realize the differential production of desferrioxamines between strains that didn't follow phylogenetic patterns. (Fig 2B).

A



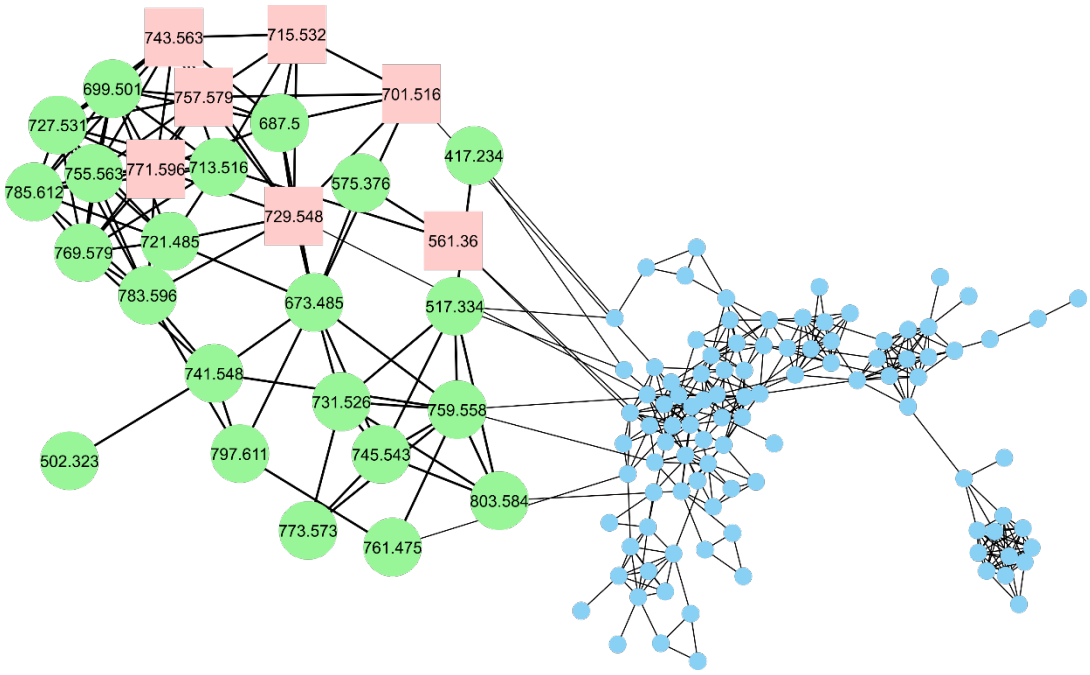
B



Hierarchical clustering of MALDI and HPLC data provides similar groupings as 16S rRNA but isn't sufficient to highlight differences in specialized metabolism.

Figure S4: GNPS Molecular Network displays observed desferrioxamine precursor ions connected by fragmentation similarity. To confirm the presence of desferrioxamines and the pattern of production among the *Micromonospora* isolates, we created a molecular network from ESI Q-Exactive HPLC-MS/MS data (method in main text) for the eight *Micromonospora* isolates, using the GNPS online workflow. The data were filtered by removing all MS/MS peaks within +/- 17 Da of the precursor *m/z*. MS/MS spectra were window filtered by choosing only the top 6 peaks in the +/- 50 Da window throughout the spectrum. The data were then clustered with MS-Cluster with a parent mass tolerance of 0.005 Da and a MS/MS fragment ion tolerance of 0.005 Da to create consensus spectra. Further, consensus spectra that contained fewer than two spectra were discarded. A network was then created where edges were filtered to have a cosine score above 0.7 and more than three matched peaks. The molecular network was trimmed so that two nodes were only connected if both nodes appeared in each other's top 10 most similar nodes.

The desferrioxamine analogs were found in a tight grouping contained within the largest network formed. We also noticed other similar masses within the network, which we ascribed to non-uniform fragmentation across eluting peaks, an artifact of the HPLC-MS/MS acquisition settings not being set for active exclusion. Please refer to Table S1 for a list of observed desferrioxamine and corresponding ferrioxamine masses and GNPS hyperlinks to their respective spectra. Desferrioxamine B and five acyl-desferrioxamines were chosen for further dereplication. These are highlighted as red squares in the network below.



GNPS results and centroided chromatograms may be accessed at the following link: http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=42b3fa45fd9c409fa447b020a6b8f08c

Raw LC-MS/MS spectra and the converted, peak-picked spectra may be accessed via the following link:
https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=7e361fe856424ef1bc2a5d2c4b5d5f3b

Figure S5 (A-F): Extracted ion chromatograms from eight *Micromonospora* isolates reveal the absence of acylated desferrioxamine production in isolate B001.

Figure S5A: Desferrioxamine B.

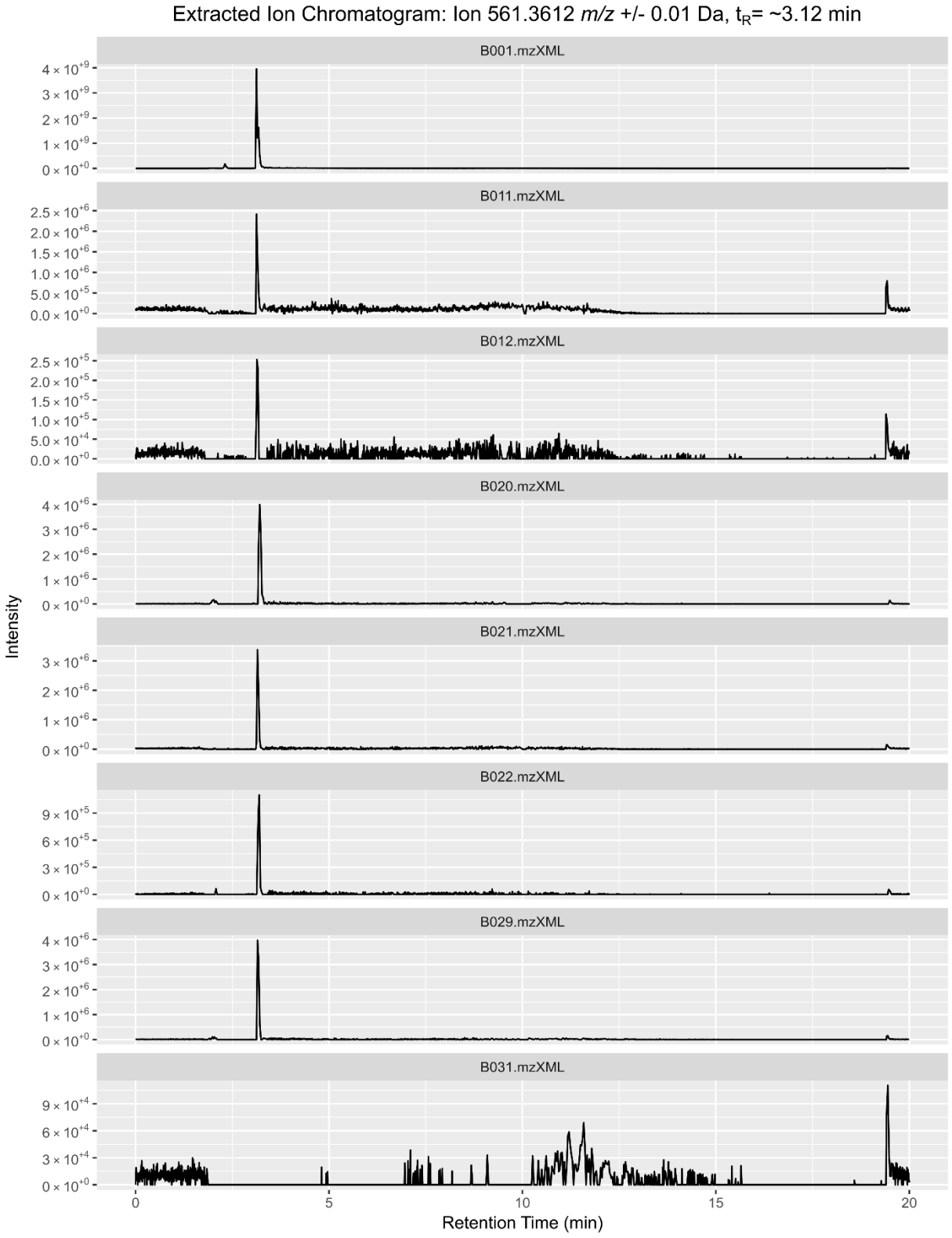


Figure S5B: C11 acyl-desferrioxamine (compound 8 in manuscript Fig 2B).

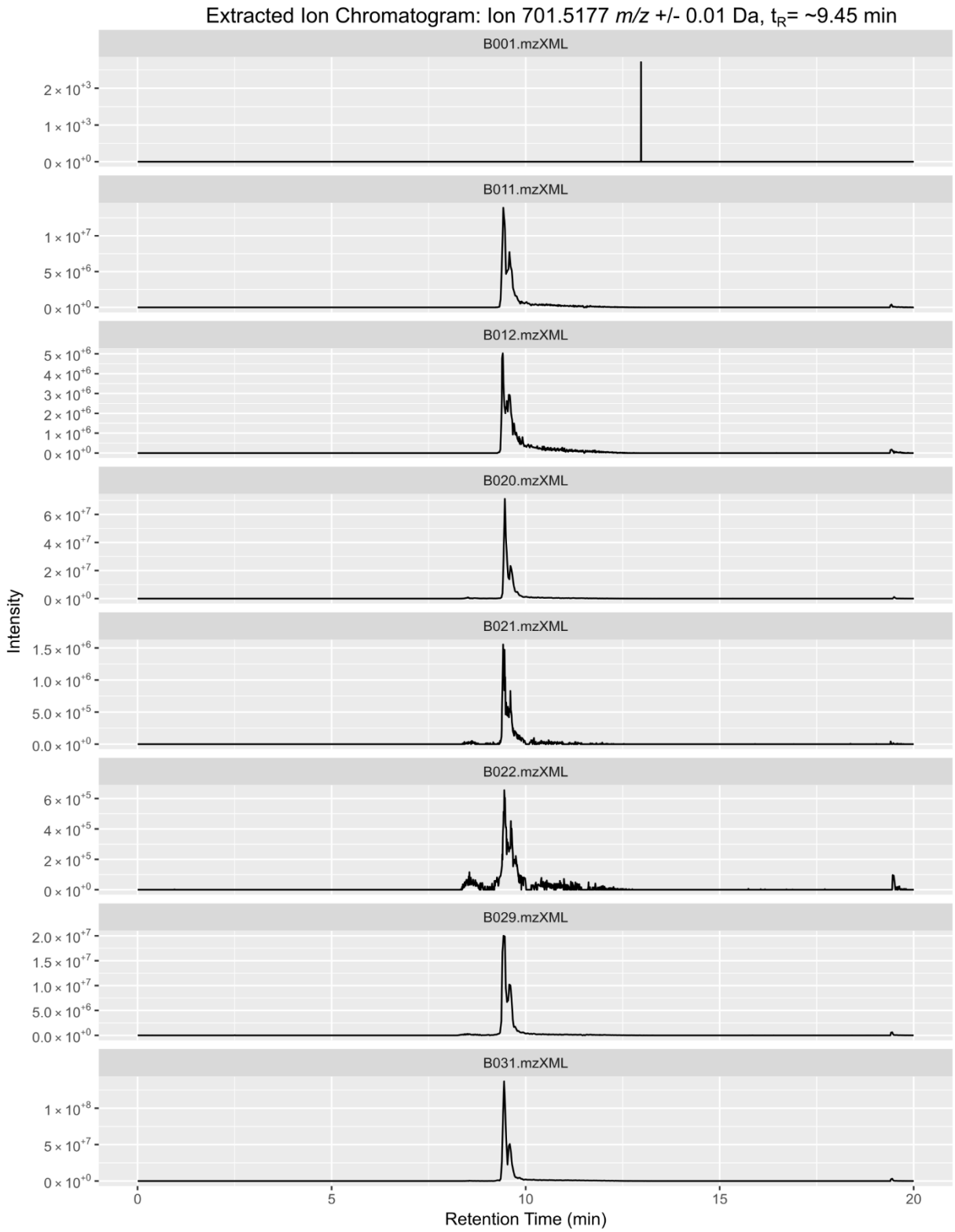


Figure S5C: C12 acyl-desferrioxamine (compound 9 in manuscript Fig 2B).

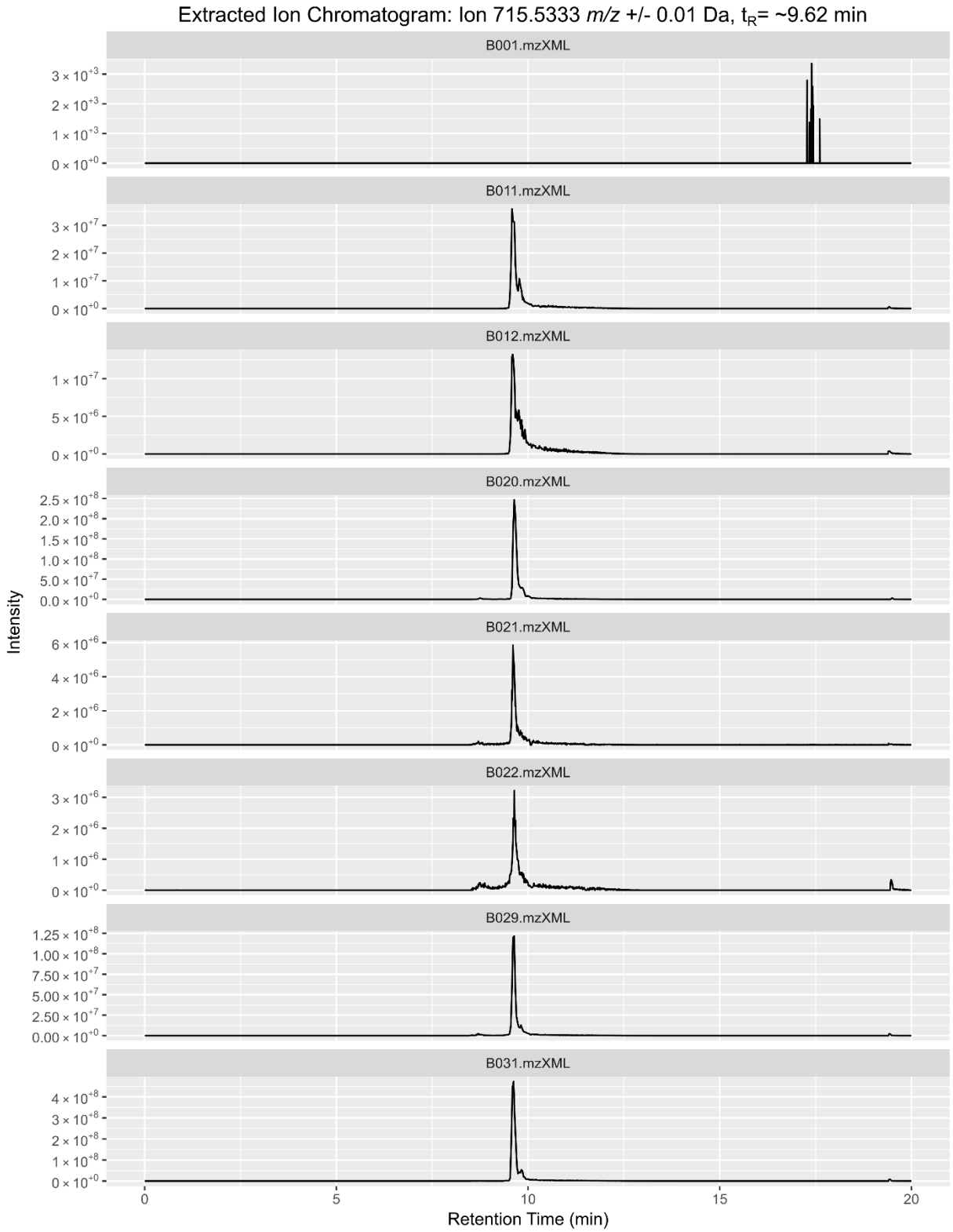


Figure S5D: C13 acyl-desferrioxamine (compound 10 in manuscript Fig 2B).

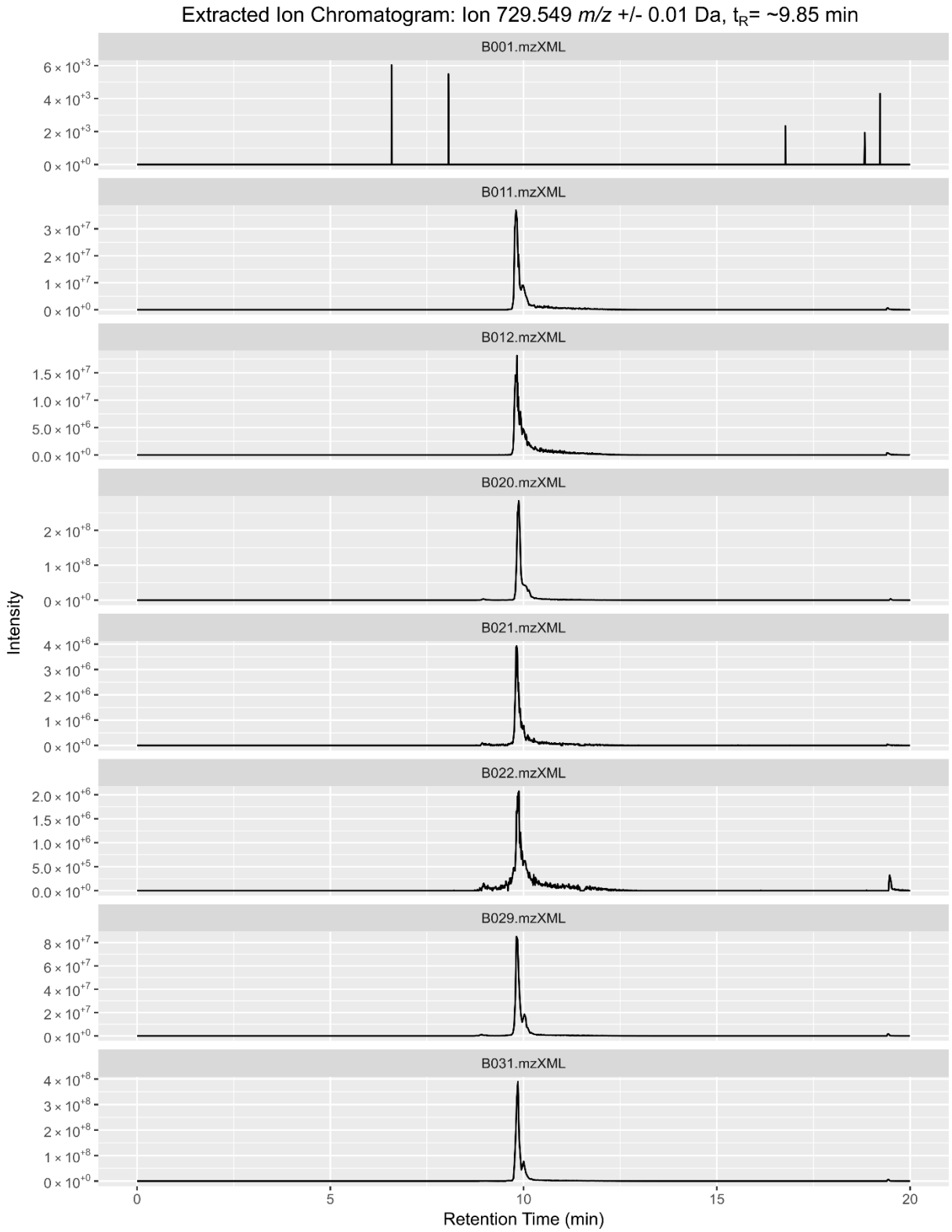


Figure S5E: C14 acyl-desferrioxamine (compound 11 in manuscript Fig 2B).

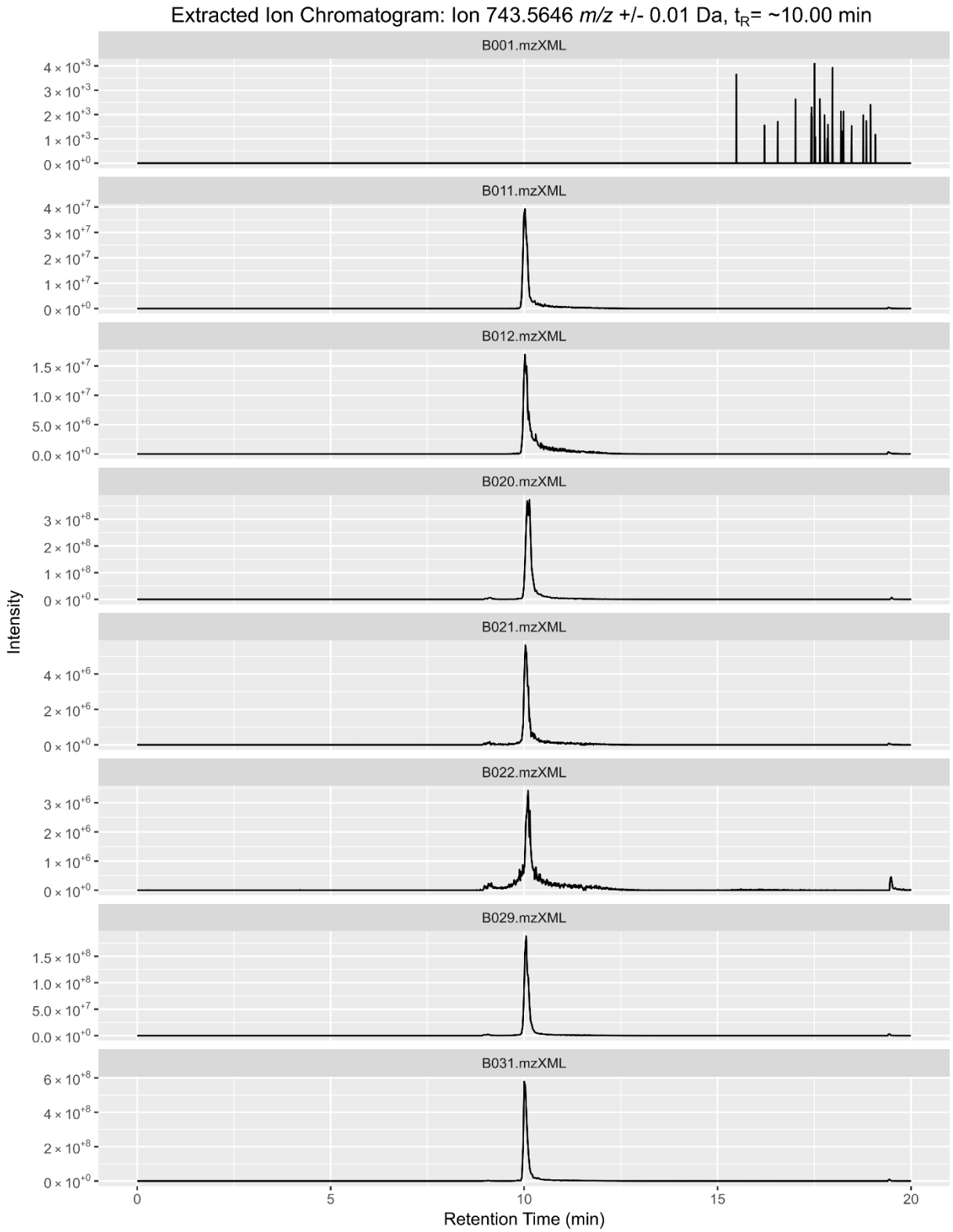


Figure S5F: C15 acyl-desferrioxamine (compound 12 in manuscript Fig 2B).

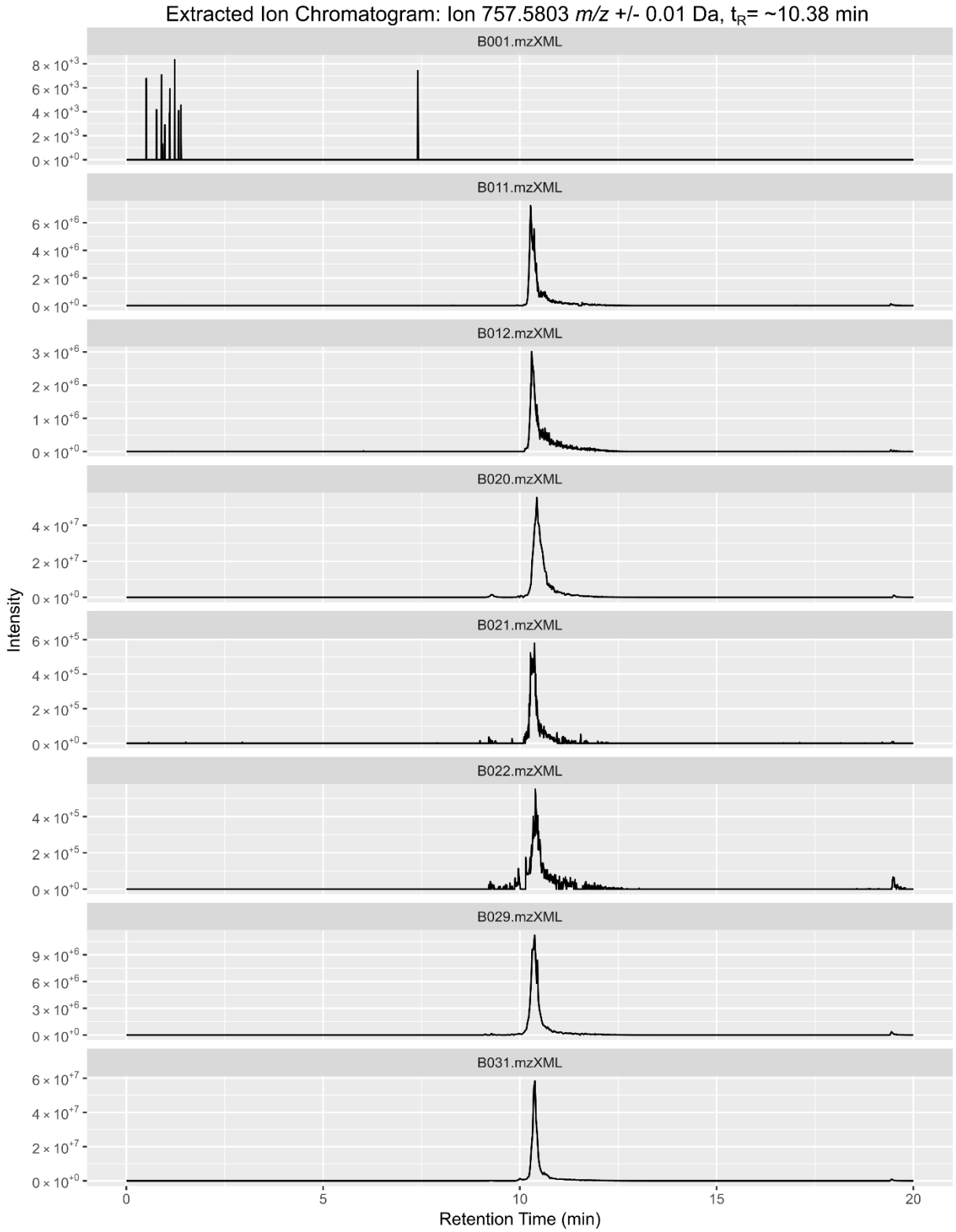


Figure S6 (A-F): HPLC-MS/MS dereplication of desferrioxamine analogs. Figure S6A shows the calculated masses for fragments common to *all* analogs discussed below. It also displays the five fragments whose masses vary as a function of the differing R group (compounds **8**-**12**). The calculated masses for the five variable fragments are provided in the table below. Figs. 6B-F display MS/MS fragmentation spectra for desferrioxamine standards (top panes) deposited into GNPS by Traxler et al (1). The lower panes display eight spectra, one for each of the *Micromonospora* isolates described in our study.

All desferrioxamine metabolites detected in this study share the common backbone shown below, with variable R-groups. The desferrioxamine biosynthetic pathway is promiscuous in its incorporation of these acyl R groups (1), and we observed, via LC-MS/MS, varying-length acyl chains that were fully saturated, mono-unsaturated, and/or mono-hydroxylated. These analogs have been previously described by Traxler (1) and Sidebottom et al (2). MS/MS fragmentation of these compounds led to similar MS fingerprints due to the peptide-like fragmentation of the shared hydroxamate core. We consistently observed the characteristic b/y fragments as displayed below; however, likely due to the use of higher energy collision dissociation (HCD), we observed higher molecular weight fragments less frequently (for example, high molecular weight fragments 1 and 2 appear in significantly lower abundances in 6B-F).

Figure S6A.

Calculated masses for observed b, y, and z ions are displayed in Fig. S6A. Calculated masses for fragment ions (b1, y4, b3, y2, b5) varied according to the acyl-chain length of an analogue. These variable fragment ions are summarized in tables above Fig. S6B-G. It should be noted that while we used a non-canonical peptide-annotation schema as desferrioxamine is actually formed from the successive condensation of modified 1,5-diaminopentane monomers.

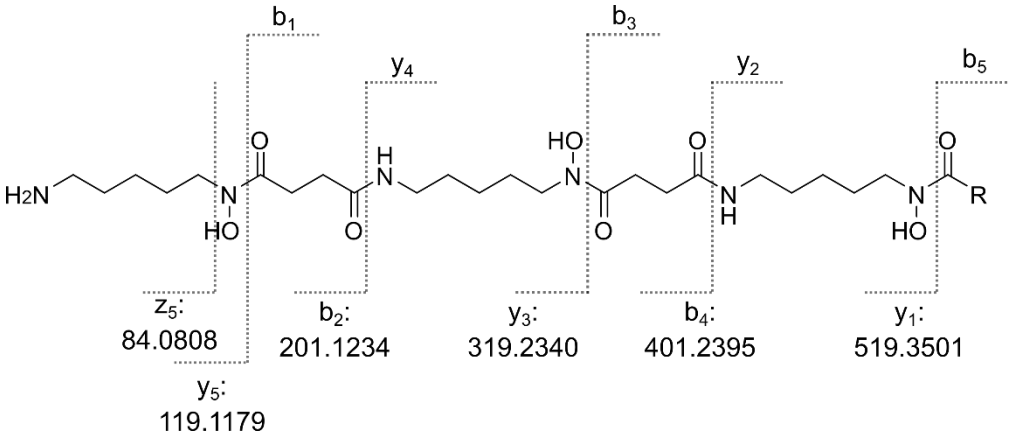
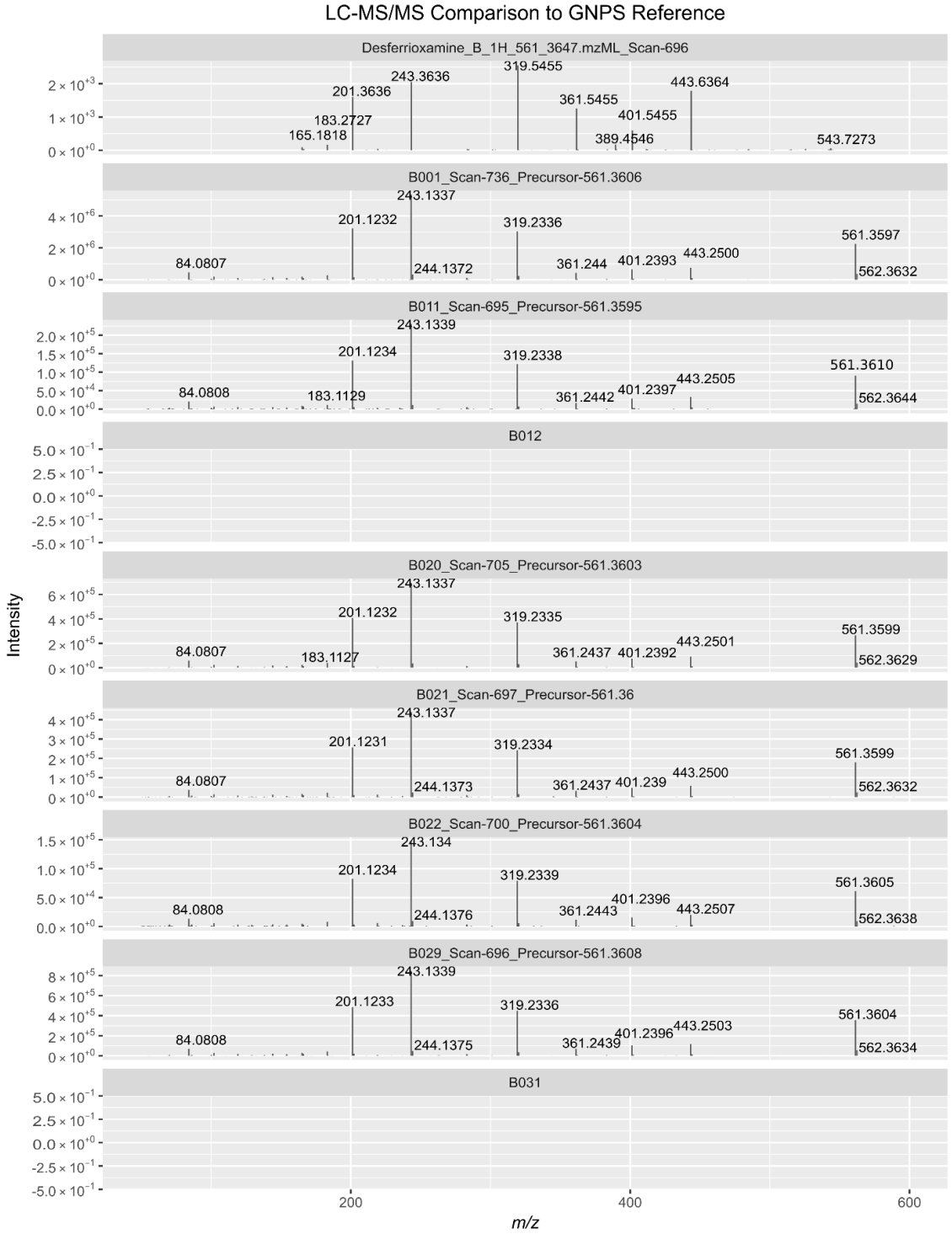


Figure S6B: Dereplication of desferrioxamine B.

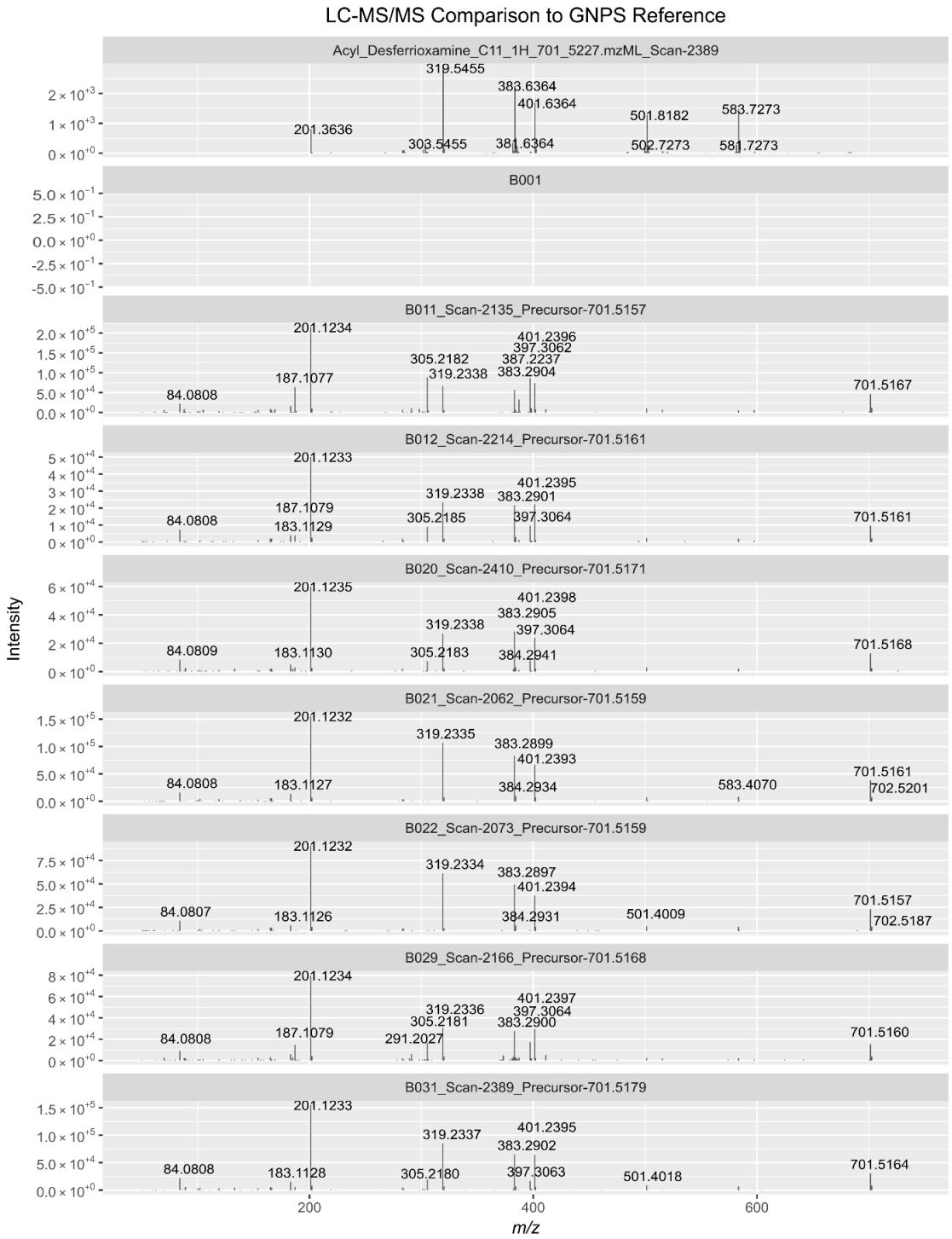
			Calculated Fragment Mass				
	Molecular Formula $[M+H]^+$	Calculated Intact Mass	b_1	y_4	b_3	y_2	b_5
Desferrioxamine B	$C_{25}H_{49}N_6O_8$	561.3612	443.2500	361.2445	243.1339	161.1285	NA



No precursor ion observed for B012 or B031.

Figure S6C: Dereplication of C11 acyl-desferrioxamine (compound 8).

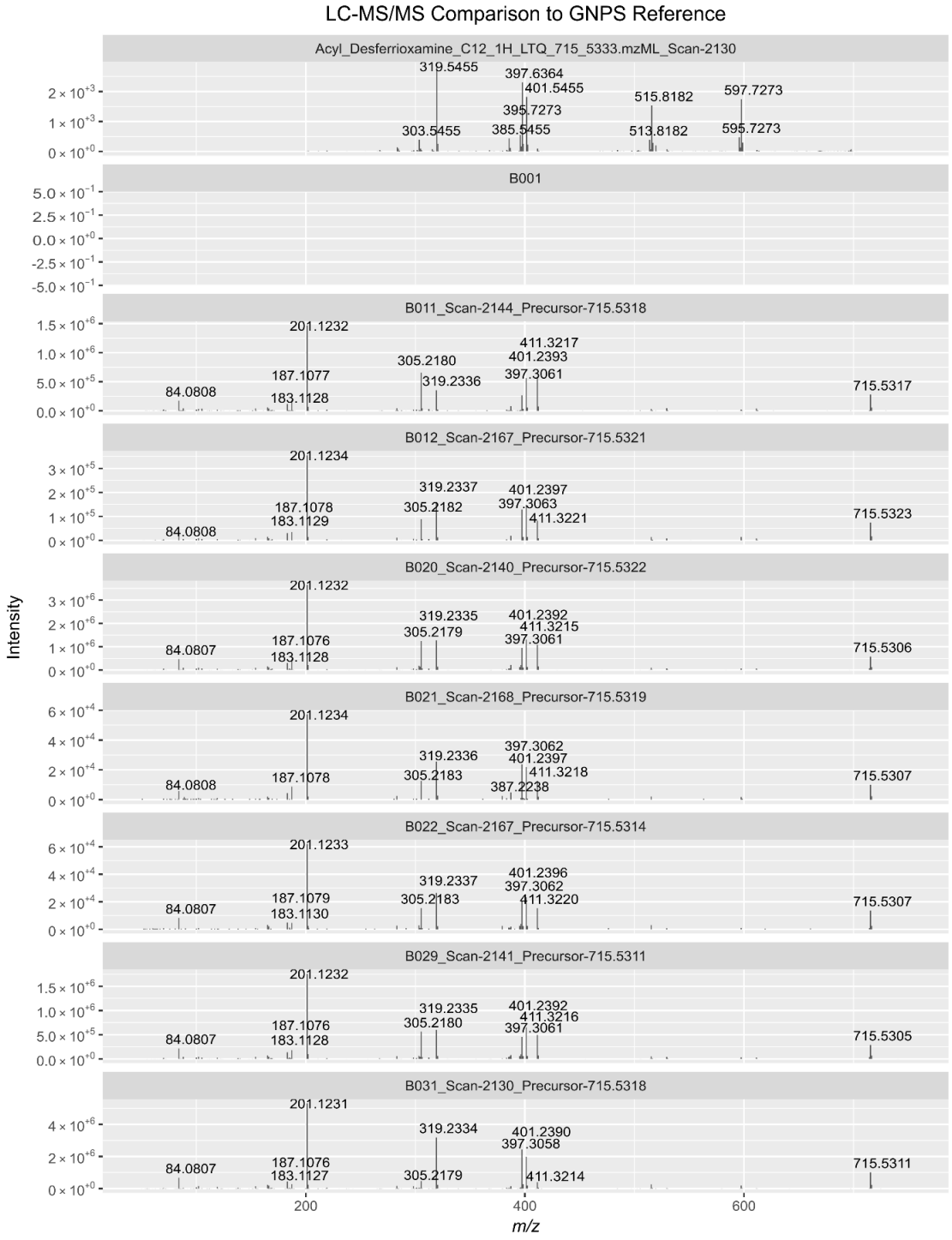
			Calculated Fragment Mass				
	Molecular Formula $[M+H]^+$	Calculated Intact Mass	b_1	y_4	b_3	y_2	b_5
C11 acyl-desferrioxamine (8)	$C_{35}H_{69}N_6O_8$	701.5177	583.4065	501.4010	383.2904	301.2850	183.1743



No precursor ion observed for B001.

Figure S6D: Dereplication of C12 acyl-desferrioxamine (compound 9).

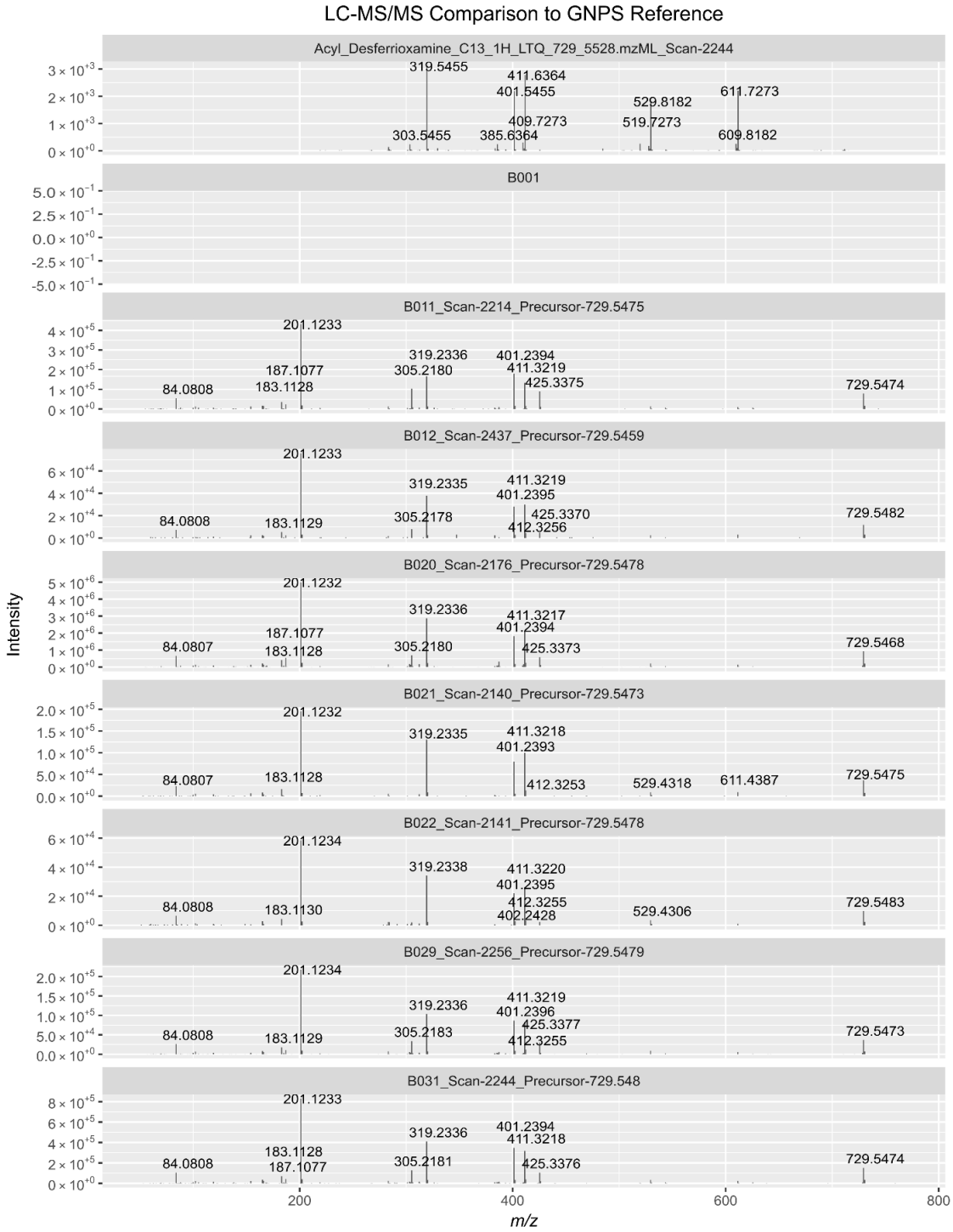
			Calculated Fragment Mass				
	Molecular Formula $[M+H]^+$	Calculated Intact Mass	b_1	y_4	b_3	y_2	b_5
C12 acyl-desferrioxamine (9)	$C_{36}H_{71}N_6O_8$	715.5333	597.4222	515.4167	397.3061	315.3007	197.1900



No precursor ion observed for B001.

Figure S6E: Dereplication of C13 acyl-desferrioxamine (compound 10).

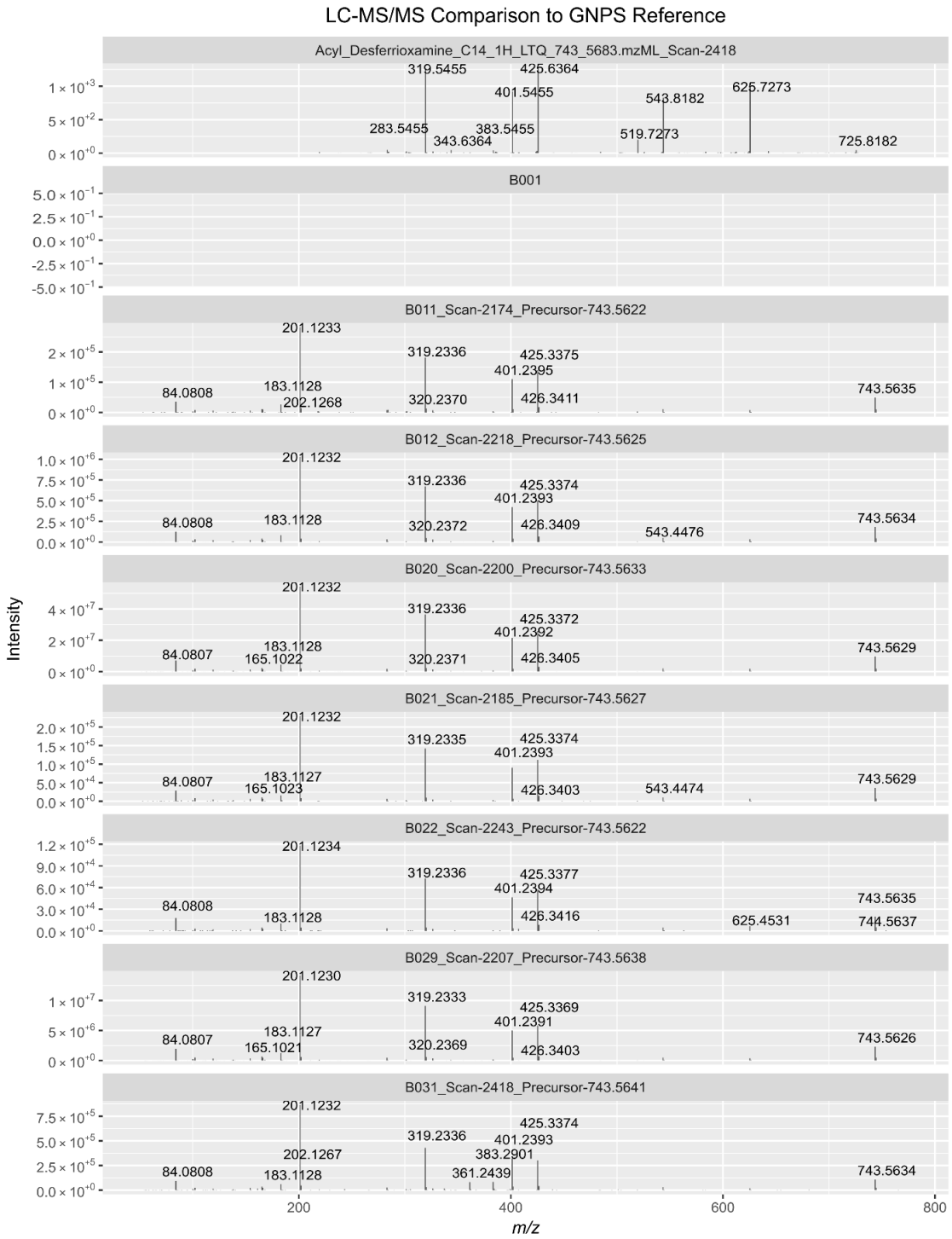
			Calculated Fragment Mass				
	Molecular Formula $[M+H]^+$	Calculated Intact Mass	b_1	y_4	b_3	y_2	b_5
C13 acyl-desferrioxamine (10)	$C_{37}H_{73}N_6O_8$	729.5490	611.4378	529.4323	411.3217	329.3163	211.2056



No precursor ion observed for B001.

Figure S6F: Dereplication of C14 acyl-desferrioxamine (compound 11).

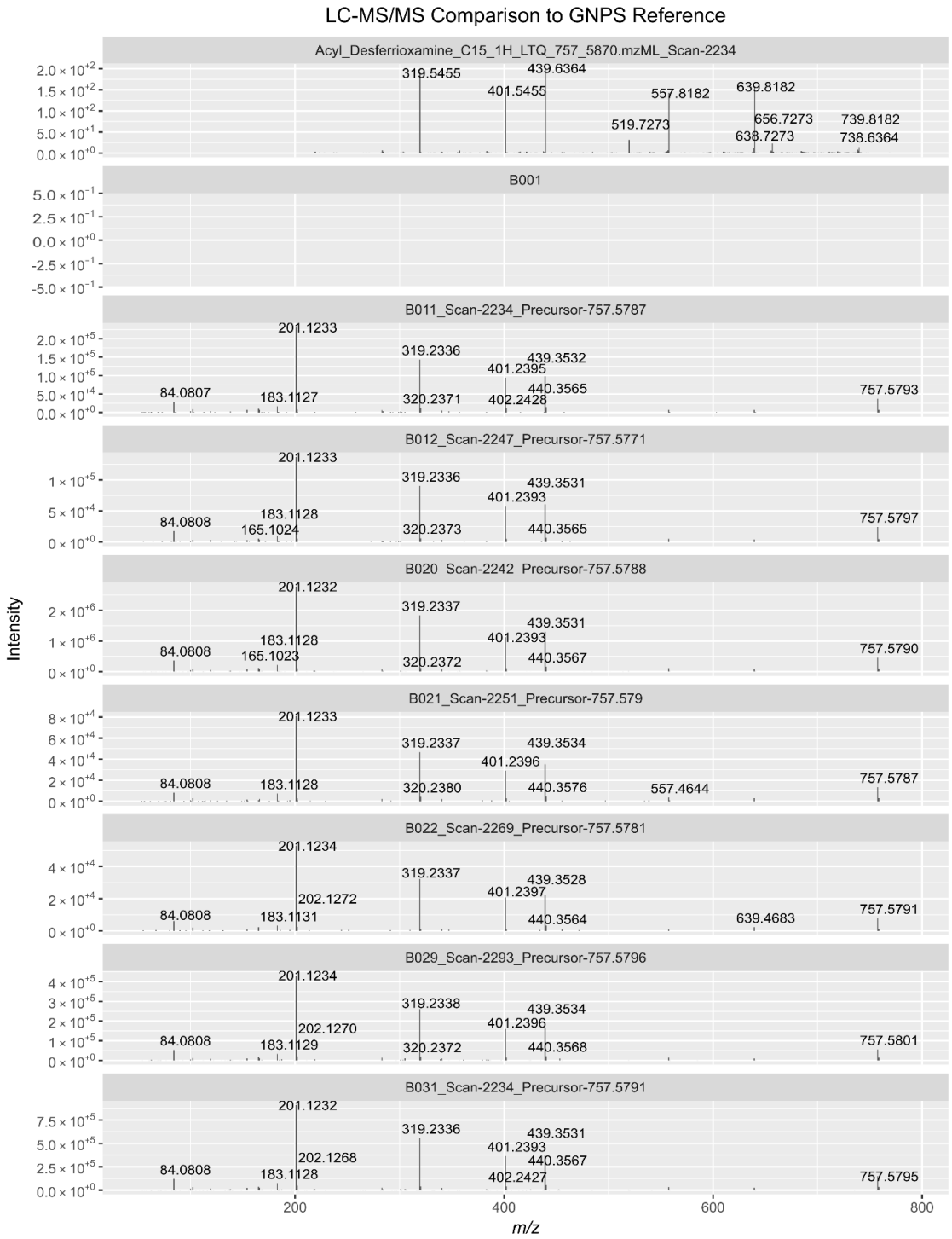
			Calculated Fragment Mass				
	Molecular Formula $[M+H]^+$	Calculated Intact Mass	b_1	y_4	b_3	y_2	b_5
C14 acyl-desferrioxamine (11)	$C_{38}H_{75}N_6O_8$	743.5646	625.4535	543.4480	425.3374	343.3320	225.2213



No precursor ion observed for B001.

Figure S6G: Dereplication of C15 acyl-desferrioxamine (compound 12).

			Calculated Fragment Mass				
	Molecular Formula $[M+H]^+$	Calculated Intact Mass	b_1	y_4	b_3	y_2	b_5
C15 acyl-desferrioxamine (12)	$C_{39}H_{77}N_6O_8$	757.5803	639.4691	557.4636	439.3530	357.3476	239.2369



No precursor ion observed for B001.

Figure 7 (A-F): Observation of the iron-chelated $[M-2H+Fe]^+$ ferrioxamine compounds by HPLC-MS/MS. As further evidence that our *Micromonospora* isolates produced desferrioxamines, we looked for the presence of the ferrioxamine (iron-bound desferrioxamine) species in the HPLC-MS/MS spectra (ferrioxamine analogues were not observed in MALDI spectra). To do so we analyzed the isotopologue distribution of corresponding MS1 features from the HPLC-MS/MS data. Below, eight centroided HPLC-MS/MS MS1 spectra are displayed as solid bars, while theoretical profile spectra generated in mMass (3) are presented as dashed lines. The spectra were normalized to the intact acyl-ferrioxamine ^{56}Fe isotopologue peak. The theoretical $[M-2H+Fe]^+$ isotopic ratio of ^{54}Fe to ^{56}Fe yields a 6.3703 to 100 intensity difference (4).

Figure S7A

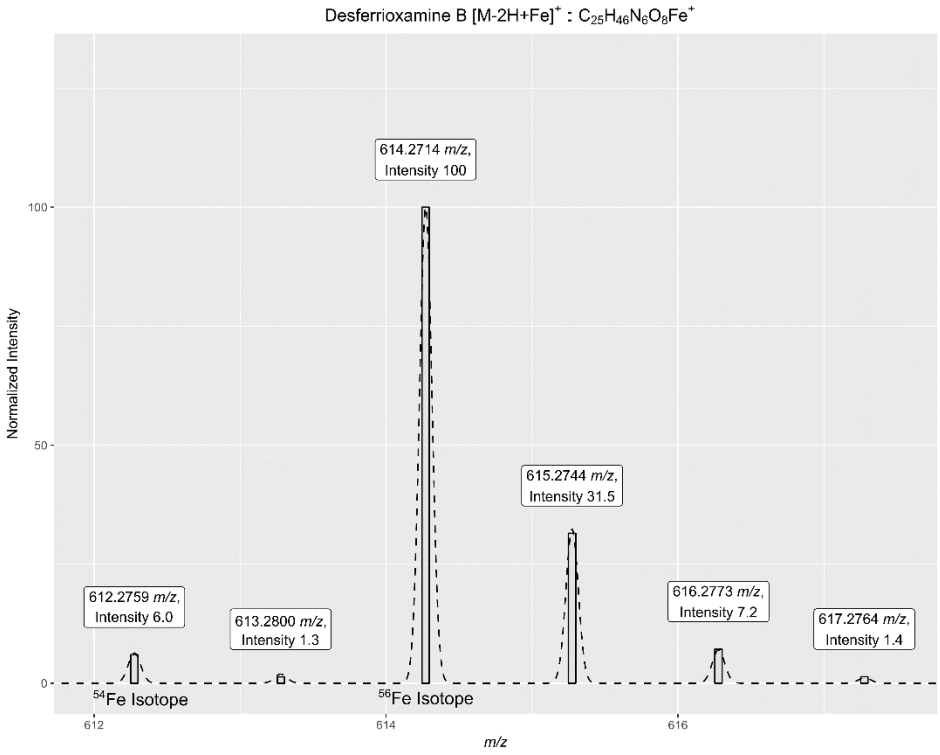


Figure S7B

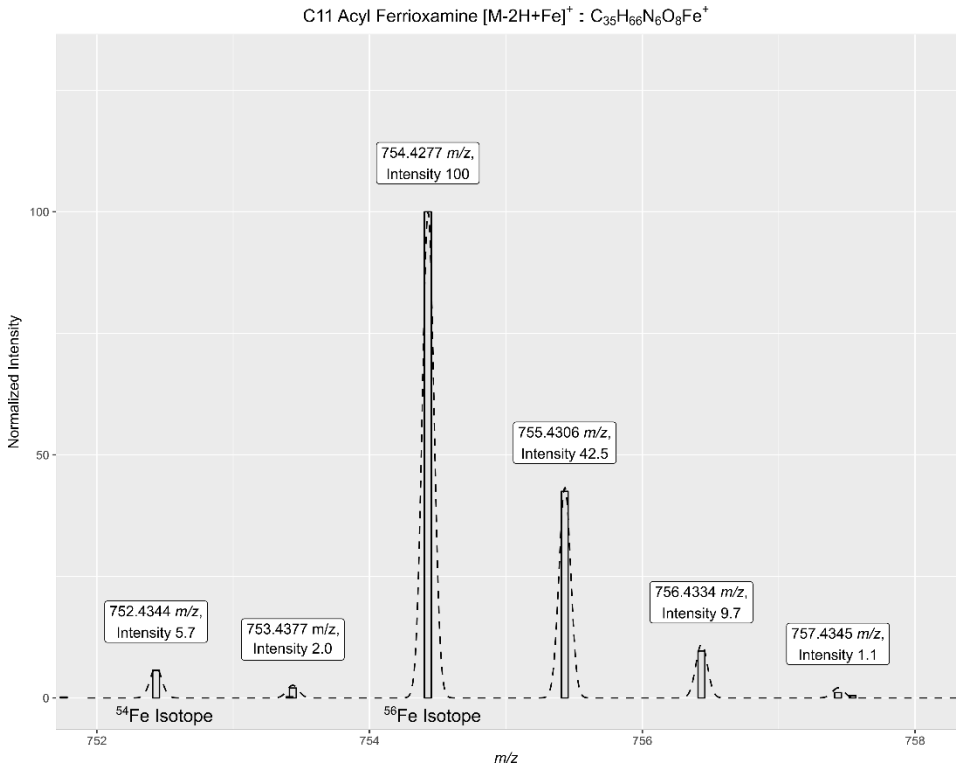


Figure S7C

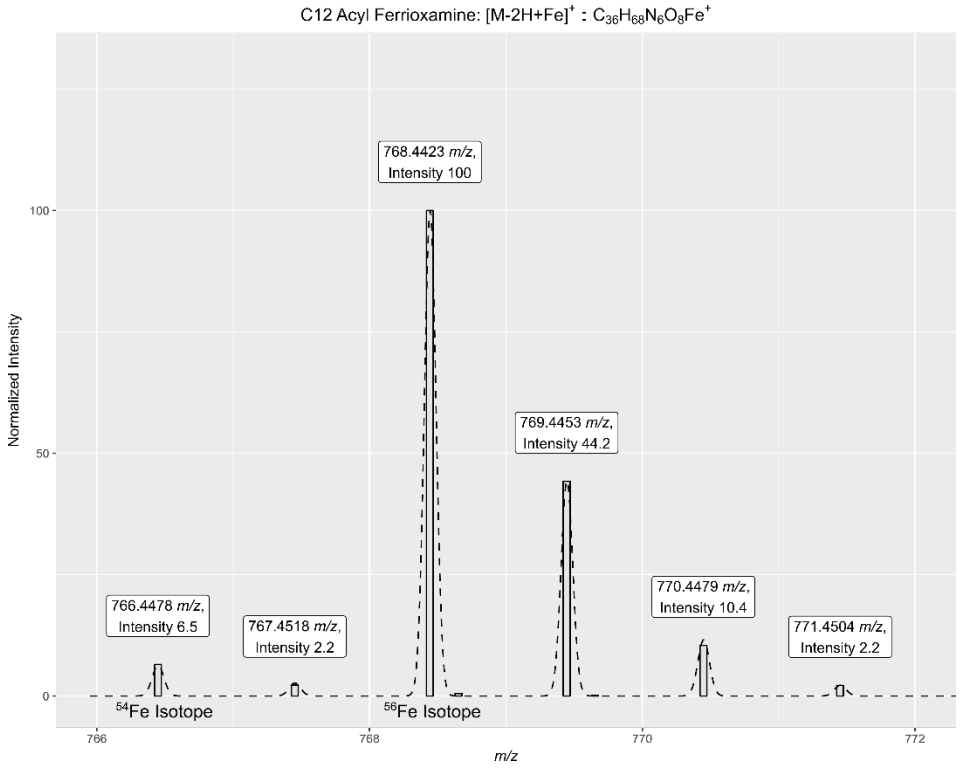


Figure S7D

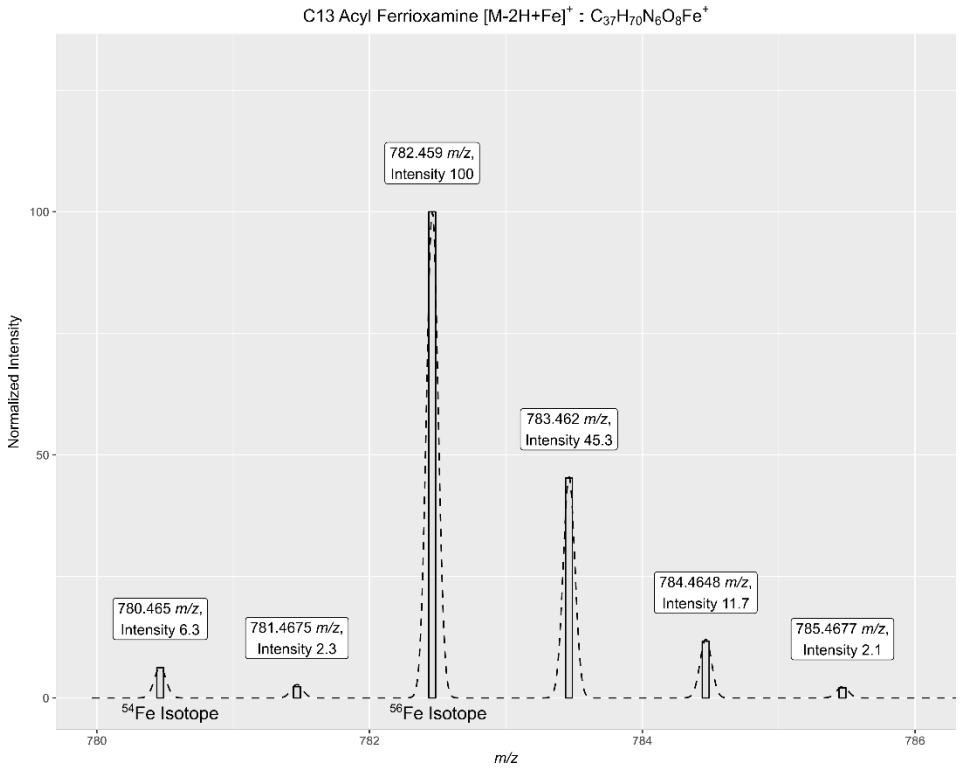


Figure S7E

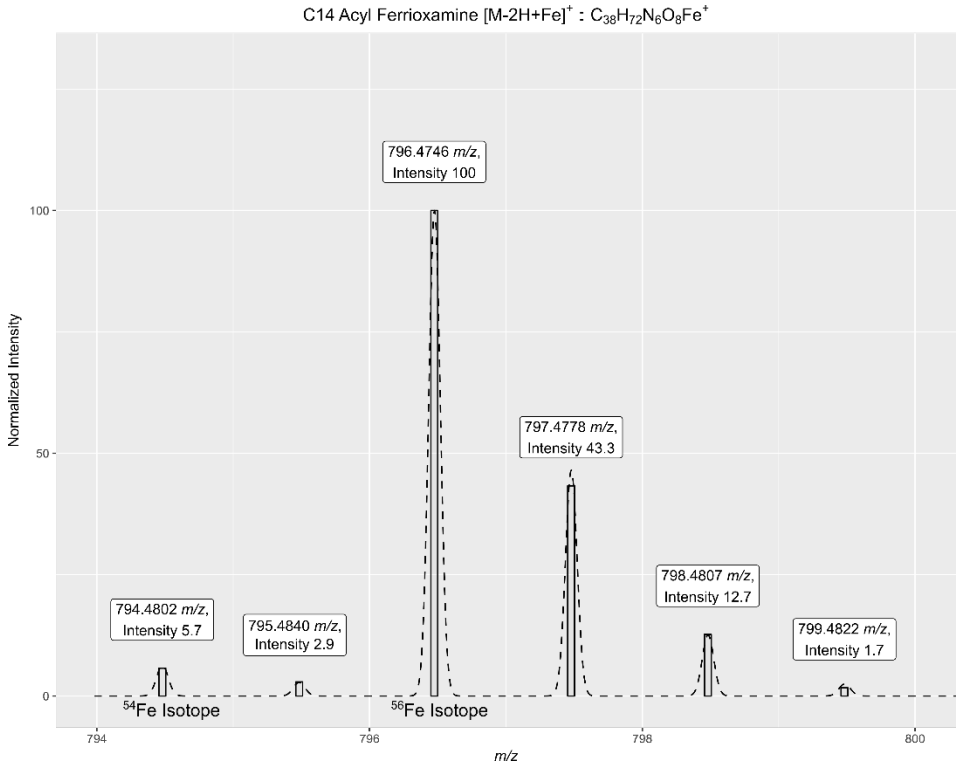


Figure S7F

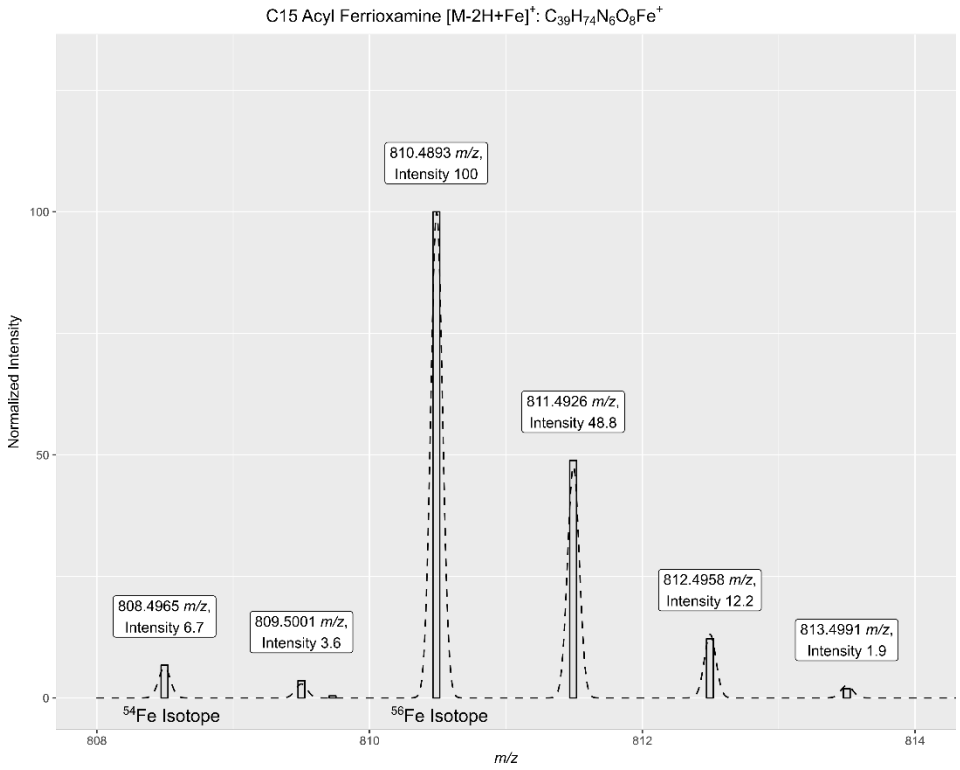

C15 Acyl Ferrioxamine [M-2H+Fe]+: C39H74N6O8Fe+
810.4893 m/z, Intensity 100
811.4926 m/z, Intensity 48.8
812.4958 m/z, Intensity 12.2
813.4991 m/z, Intensity 1.9
808.4965 m/z, Intensity 6.7
809.5001 m/z, Intensity 3.6
54Fe Isotope
56Fe Isotope
Normalized Intensity
100
50
0
808
810
812
814
m/z

Figure S8: **Neighbor-joining phylogenetic tree of diversity plate 172.** 16S rRNA gene sequence contigs were assembled using the Geneious alignment tool in the Geneious software package (Biomatters limited). SILVA Incremental Aligner (5) was used to align the sequences and the tree constructed using Jukes-Cantor (6). Node support was estimated using 100 bootstrap replicates.

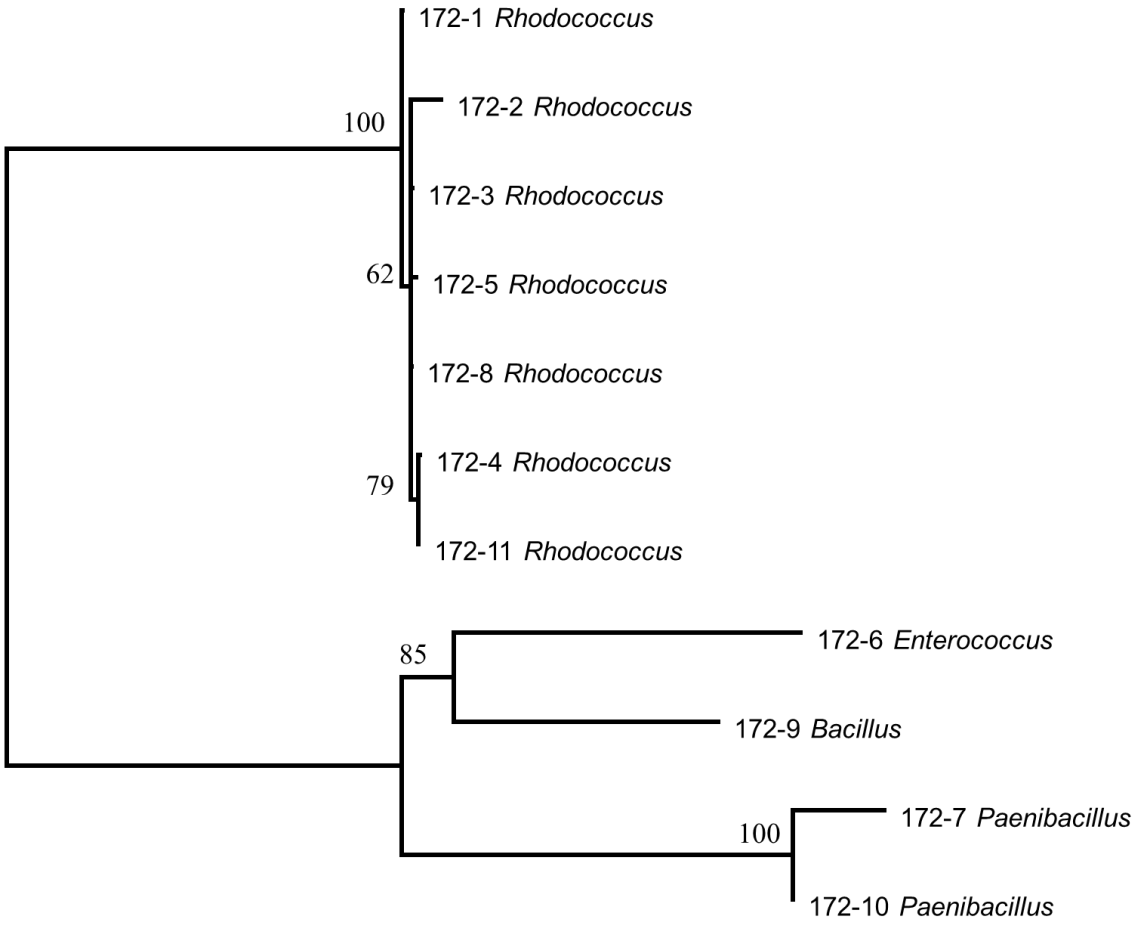


Figure S9: Comparison between 16S rRNA gene sequencing and MALDI protein groupings using Euclidean distance and cosine distance. Figure S9 shows the similarity groupings by 16S rRNA gene sequencing (see Fig. S8) and MALDI protein profiling. The IDBac workflow allows the user to choose different distance measures and clustering methods, depending on their preference. For this study we employed both Euclidean distance and ward.D2, or cosine distance and single/complete/average linkage clustering.

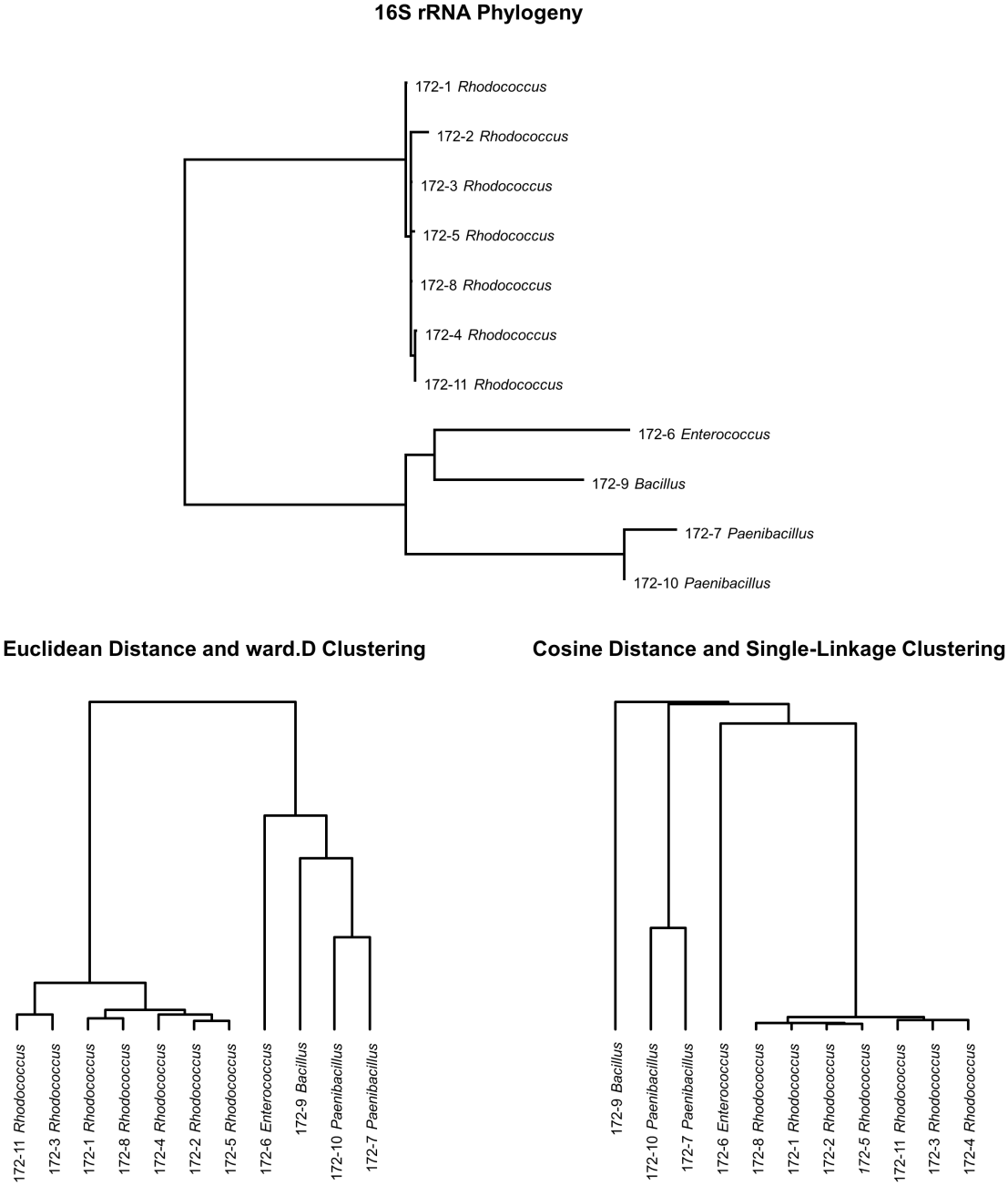


Figure S10: MAN of colonies from diversity plate vs MAN of colonies from pure bacterial isolates. In the case of analyzing bacteria directly from environmental diversity plates, we were concerned that the unequal distribution of colonies on a plate might create microenvironments that would lead to differing specialized metabolite profiles, *e.g.*, two genetically identical isolates that exhibit slightly different specialized metabolite production fingerprints. To test the latter, we compared the MAN of MALDI data acquired from colonies from our environmental isolation plate 172 (colonies as they exist on the diversity plate, 3 technical replicates each), with MAN profiles acquired from each colony when grown in isolation (10 technical replicates each). Aside from minor changes to 172-1 (*Rhodococcus* sp.), the only colony exhibiting significant metabolite differences was 172-9 (*Bacillus* sp.). Importantly, the differences in specialized metabolite production between 172-7 and 172-10 (both *Paenibacillus* sp.) that were observed in both MANs remained consistent. Thus, IDBac analysis successfully detected subtle intra-species differences in *Paenibacillus* specialized metabolite production, and these were not simply due to other environmental factors such as proximity toward or chemical crosstalk with other colonies on the plate. Reproducibility: The two networks represent data from 143 individual sampling and data acquisition events.

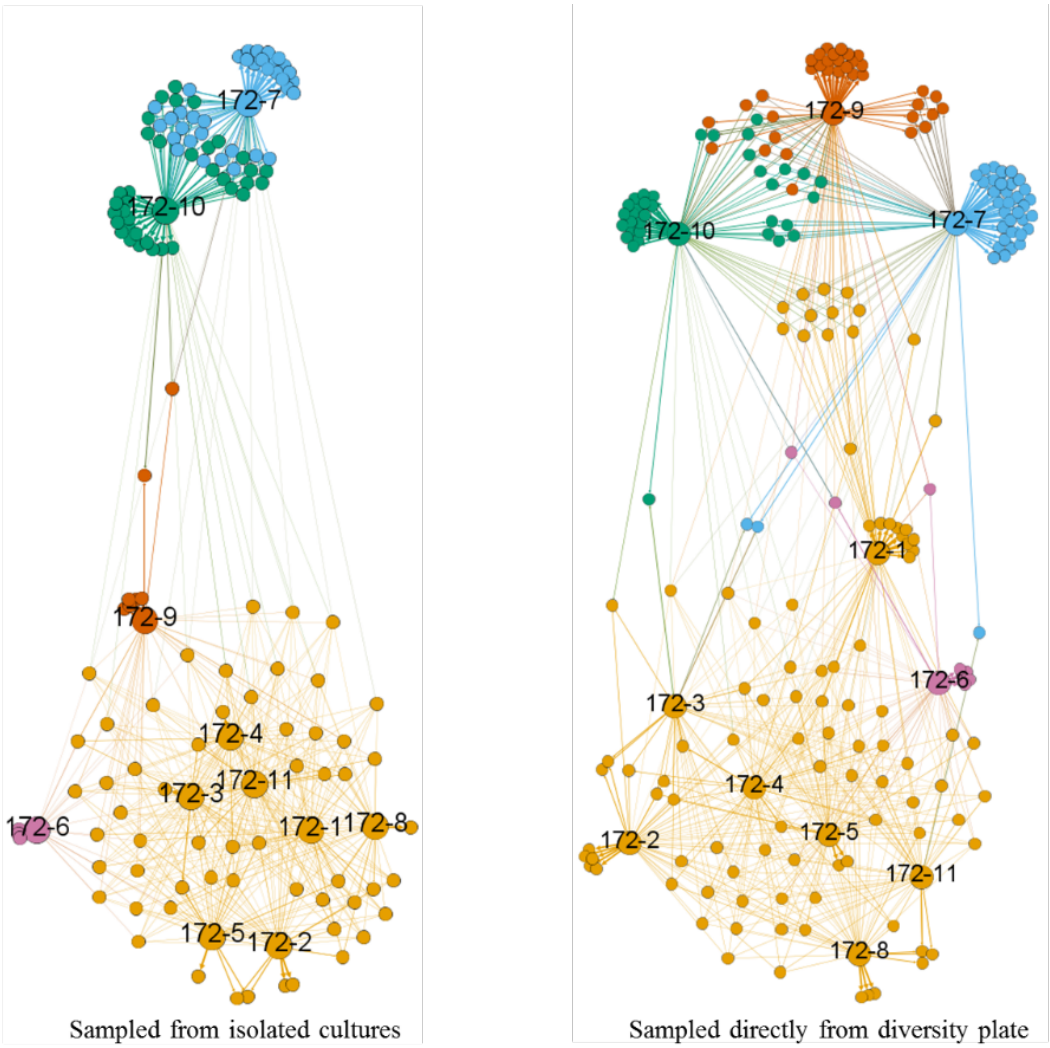
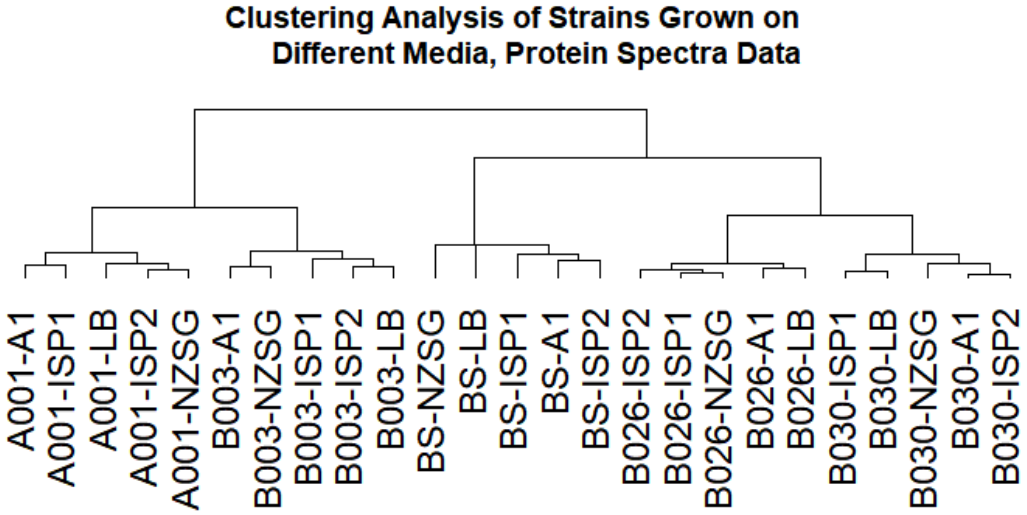


Figure S11: Effect of altering culture media on MALDI-TOF MS protein groupings of bacterial isolates. Because our study involved directly comparing protein MS profiles of strains acquired on different nutrient media, we tested the effect of altering these media on the resulting protein MS profiles of five strains cultivated on five media types (A1, ISP1, ISP2, NZSG and LB). We acquired three technical replicates of each sample. While it is true that altering nutrient media introduces variance in spectral groupings, we concluded, for these five strains, intra-strain variance was smaller than the inter-strain variance. Of course, standardized methodology, including growing bacteria on a single media type, will often provide the most consistent and robust data.



Strain ID	Genus species	Accession
A001	*Streptomyces tendae*	MG188670
B003	*Streptomyces koyangensis*	MG188671
BS	*Bacillus subtilis* 3610	CP020102
B026	*Micromonospora tulbaghiae*	KP009553
B030	*Micromonospora chokoriensis*	KY858241

Media Ingredient / 1 L	LB	NZSG	A1	ISP1	ISP2
Malt Extract	0 g	0 g	0 g	0 g	10 g
Yeast Extract	5 g	5 g	2 g	3 g	4 g
Soluble Starch	0 g	20 g	10 g	0 g	0 g
Glucose	0 g	10 g	0 g	0 g	4 g
N-Z Amine A	0 g	5 g	0 g	0 g	0 g
Calcium Carbonate	0 g	3 g	0 g	0 g	0 g
Casitone	0 g	0 g	0 g	5 g	0 g
Peptone	0 g	0 g	1 g	0 g	0 g
Tryptone	10 g	0 g	0 g	0 g	0 g
NaCl	10 g	0 g	0 g	0 g	0 g

Table S1: Detected desferrioxamines from *Micromonospora* isolates. Five desferrioxamine analogs are highlighted in Figure 2B (C11 acyl-DFO, C12 acyl-DFO, C13 acyl-DFO, C14 acyl-DFO, C15 acyl-DFO. HPLC-MS/MS facilitated chromatographic resolution and dereplication of a greater number of analogs, which may be accessed at the following link or using hyperlinks provided in the table below:
http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=42b3fa45fd9c409fa447b020a6b8f08c

Compound Number in Main Text	As reported in Sidebottom et al. (2)	As reported in Traxler et al. (1)	Formed Ion Formula	Ion Species	Calculated Monoisotopic Mass	GNPS Hyperlink
		Desferrioxamine B	$C_{25}H_{49}N_6O_8$	**[M+H]$^+$**	561.3612	2511
		Desferrioxamine E	$C_{27}H_{49}N_6O_9$	**[M+H]$^+$**	601.3561	2789
		C7 Acyl-DFO	$C_{31}H_{61}N_6O_8$	**[M+H]$^+$**	645.4551	3101
		C9 Acyl-DFO	$C_{33}H_{65}N_6O_8$	**[M+H]$^+$**	673.4864	3264
	Amphiphilic ferrioxamine 1	C10 Acyl-DFO	$C_{34}H_{67}N_6O_8$	**[M+H]$^+$**	687.5020	3388
8	Amphiphilic ferrioxamine 4	C11 Acyl-DFO	$C_{35}H_{69}N_6O_8$	**[M+H]$^+$**	701.5177	3484
9	Amphiphilic ferrioxamine 7	C12 Acyl-DFO	$C_{36}H_{71}N_6O_8$	**[M+H]$^+$**	715.5333	3554
10	Amphiphilic ferrioxamine 10	C13 Acyl-DFO	$C_{37}H_{73}N_6O_8$	**[M+H]$^+$**	729.5490	3649
11	Amphiphilic ferrioxamine 13	C14 Acyl-DFO	$C_{38}H_{75}N_6O_8$	**[M+H]$^+$**	743.5646	3724
12		C15 Acyl-DFO	$C_{39}H_{77}N_6O_8$	**[M+H]$^+$**	757.5803	3812
		C16 Acyl-DFO	$C_{40}H_{79}N_6O_8$	**[M+H]$^+$**	771.5959	3907
		C17 Acyl-DFO	$C_{41}H_{81}N_6O_8$	**[M+H]$^+$**	785.6116	3982
	Amphiphilic ferrioxamine 2		$C_{34}H_{65}N_6O_8$	**[M+H]$^+$**	685.4864	3383
	Amphiphilic ferrioxamine 3		$C_{34}H_{67}N_6O_9$	**[M+H]$^+$**	703.4970	3489
	Amphiphilic ferrioxamine 5		$C_{35}H_{67}N_6O_8$	**[M+H]$^+$**	699.5021	3469
	Amphiphilic ferrioxamine 8		C36H69N6O8	**[M+H]$^+$**	713.5177	3549
	Amphiphilic ferrioxamine 9		$C_{36}H_{71}N_6O_9$	**[M+H]$^+$**	731.5283	3651
	Amphiphilic ferrioxamine 11		$C_{37}H_{71}N_6O_8$	**[M+H]$^+$**	727.5334	3644
	Amphiphilic ferrioxamine 12		$C_{37}H_{73}N_6O_9$	**[M+H]$^+$**	745.5439	3729
	Amphiphilic ferrioxamine 14		$C_{38}H_{73}N_6O_8$	**[M+H]$^+$**	741.5490	3715
	Amphiphilic ferrioxamine 15		$C_{38}H_{75}N_6O_9$	**[M+H]$^+$**	759.5596	3813

Table S2: **Collection coordinates and GenBank accession numbers of strains used in this work.** Strains used in this study were identified based on sequencing their 16S rRNA genes. The corresponding sequences were uploaded to NCBI GenBank and can be accessed using the listed accession numbers. Strains were assigned a genus using BLASTn and SILVA Incremental Aligner (5).

Strain ID	Latitude	Longitude	Genus	Accession	Length (bp)
B020	43°13'27.63"N	87°34'10.62"W	*Micromonospora*	KY858243	1,476
B022	43°13'27.63"N	87°34'10.62"W	*Micromonospora*	KY858245	1,474
B021	43°13'27.63"N	87°34'10.62"W	*Micromonospora*	KY858244	1,474
B012	43°13'27.63"N	87°34'10.62"W	*Micromonospora*	KY858247	1,478
B031	41°45'42.60"N	86°49'28.12"W	*Micromonospora*	KY858239	1,478
B029	43°13'27.63"N	87°34'10.62"W	*Micromonospora*	KY858242	1,474
B011	43°13'27.63"N	87°34'10.62"W	*Micromonospora*	KY858246	1,476
B001	43°13'27.63"N	87°34'10.62"W	*Micromonospora*	KY858238	1,475
172-1	45º5'16.012"N	87º35'10.468"W	*Rhodococcus*	KY858237	1,188
172-2	45º5'16.012"N	87º35'10.468"W	*Rhodococcus*	KY858236	1,189
172-3	45º5'16.012"N	87º35'10.468"W	*Rhodococcus*	KY858235	1,177
172-4	45º5'16.012"N	87º35'10.468"W	*Rhodococcus*	KY858234	1,257
172-5	45º5'16.012"N	87º35'10.468"W	*Rhodococcus*	KY858233	1,286
172-6	45º5'16.012"N	87º35'10.468"W	*Enterococcus*	KY858232	1,289
172-7	45º5'16.012"N	87º35'10.468"W	*Paenibacillus*	KY858231	1,498
172-8	45º5'16.012"N	87º35'10.468"W	*Rhodococcus*	KY858230	1,171
172-9	45º5'16.012"N	87º35'10.468"W	*Bacillus*	KY858229	1,230
172-10	45º5'16.012"N	87º35'10.468"W	*Paenibacillus*	KY858228	1,498
172-11	45º5'16.012"N	87º35'10.468"W	*Rhodococcus*	KY858227	1,230
A001	42°21'57.18"N	70º58'16.74"W	*Streptomyces*	MG188670	1,488
B003	43°13'27.63"N	87°34'10.62"W	*Streptomyces*	MG188671	1,489
3610	No information	No information	*Bacillus subtilis*	CP020102	4,215,607
B026	43°13'27.63"N	87°34'10.62"W	*Micromonospora*	KP009553	1,476
B030	43°13'27.63"N	87°34'10.62"W	*Micromonospora*	KY858241	1,476

Table S3: **Detailed MALDI-TOF MS acquisition parameters.**

Detailed MALDI-TOF MS Parameters		
Parameter	**Protein Spectra**	**Small Molecule Spectra**
Instrument type	autoflex	autoflex
flexControl version	flexControl 3.4.135.0	flexControl 3.4.135.0
Type of digitizer	LeCroy LSA2000	LeCroy LSA2000
Number of shots	1200	5000
Retention time for Warp-LC	0 s	0 s
Digitizer bit depth	8	8
AIDA version number	AIDA4.7.373.7	AIDA4.7.373.7
Target type	0219793	0219793
Spectrum delay	29793 ns	9297 ns
SampleRate reciprocal	1.6 ns	0.2 ns
Spectrum size	42154 pts	253781 pts
Himass turbo mode	false	false
Laser repetition rate	1000 Hz	2000 Hz
Linear detector voltage	2.919 kV	2.919 kV
Reflector detector voltage	1.883 kV	1.883 kV
Voltage of high mass detector	0 kV	0 kV
Realtime smooth	high	off
AutoXecute method	IDBac_Protein_AutoX	IDBac_Small-Molecule_autoX
A flat line is created if the acceptance criteria of AutoXecute are not met	true	true
Flag indicating In Source Decay measurement	false	false
Flag indicating HPC usage	false	false
Calibration mass control list used	MBT_Standard	IDBac_Small_Molecule_Calibration
Sensitivity of digitizer	100 mV/fullscale	100 mV/fullscale
Analog Offset	1.4 mV	2 mV
Deflection pulser cal	0	0
PIE delay	220 ns	120 ns
Positive voltage polarity	POS	POS
Reflector voltage 2	0 kV	9.7 kV
Ion source voltage 1	19.5 kV	19 kV
Ion source voltage 2	18.2 kV	16.55 kV
Lens voltage	7.5 kV	8.3 kV
Reflector voltage 1	0 kV	21 kV
Matrix suppression mode	deflection	deflection
Matrix suppression cut off mass	1500	50

Table S4: Annotated peak list of data presented in Fig. 1B.

Strain	*m/z*	ID	Abbreviation
BS3610	266.9	266.9 *m/z*	
BS3610	440.1	440.1 *m/z*	
BS3610	458.1	458.1 *m/z*	
BS3610	616.2	616.2 *m/z*	
BS3610	617.2	617.2 *m/z*	
BS3610	714.3	714.3 *m/z* polyglutamate	
BS3610	715.3	714.3 *m/z* polyglutamate + 1	
BS3610	716.3	714.3 *m/z* polyglutamate + 2	
BS3610	843.3	843.3 *m/z* polyglutamate	
BS3610	844.3	844.3 *m/z* polyglutamate + 1	
BS3610	1046.5	Surfactin-C13 + K	S13
BS3610	1047.6	Surfactin-C13 + K + 1	S13
BS3610	1058.7	Surfactin-C15 + Na	S15
BS3610	1060.6	Surfactin-C14 + K	S14
BS3610	1061.6	Surfactin-C14 + K + 1	S14
BS3610	1074.6	Surfactin-C15 + K	S15
BS3610	1075.6	Surfactin-C15 + K + 1	S15
BS3610	1076.6	Surfactin-C15 + K + 2	S15
BS3610	1077.6	1077.6 *m/z*	
BS3610	1112.6	1112.6 *m/z*	
BS3610	1477.8	Plipastatin-C17-Ala + H	PA2
BS3610	1491.8	Plipastatin-C16-Val + H	PB1
BS3610	1492.8	Plipastatin-C16-Val + H + 1	PB1
BS3610	1501.7	Plipastatin-C16-Ala + K	PA1
BS3610	1502.7	Plipastatin-C16-Ala + K + 1	PA1
BS3610	1505.8	Plipastatin-C17-Val + H	PB2
BS3610	1506.8	Plipastatin-C17-Val + H + 1	PB2
BS3610	1513.7	Plipastatin-C16-Val + Na	PB1
BS3610	1514.7	Plipastatin-C16-Val + Na + 1	PB1
BS3610	1515.7	Plipastatin-C17-Ala + Na	PA2
BS3610	1516.7	Plipastatin-C17-Ala + Na +1	PA2
BS3610	1517.8	Plipastatin-C17-Ala + Na+ 2	PA2
BS3610	1527.8	Plipastatin-C17-Val + Na	PB2
BS3610	1528.8	Plipastatin-C17-Val + Na + 1	PB2
BS3610	1529.7	Plipastatin-C16-Val + K	PB1
BS3610	1530.7	Plipastatin-C16-Val + K + 1	PB1
BS3610	1531.7	Plipastatin-C16-Val + K + 2	PB1
BS3610	1543.8	Plipastatin-C17-Val + K	PB2
BS3610	1544.8	Plipastatin-C17-Val + K + 1	PB2
BS3610	1545.8	Plipastatin-C17-Val + K + 2	PB2
PY79	266.9	266.9 *m/z*	
PY79	273.0	273.0 *m/z*	
PY79	458.1	458.1 *m/z*	
PY79	493.0	493.0 *m/z*	
PY79	714.3	714.3 *m/z* polyglutamate	
PY79	715.3	715.3 *m/z* polyglutamate + 1	
PY79	716.3	716.3 *m/z* polyglutamate + 2	
PY79	843.3	843.3 *m/z* polyglutamate	
PY79	844.3	844.3 *m/z* polyglutamate + 1	

Text S1: Coding and logic of Metabolite Association Networks (MANs). While dimensionality reduction methods such as PCA allow for visualizing complex datasets such as MALDI MS spectra, the variables (*m/z* peaks) contributing to visual groupings are not readily apparent. We turned instead to network analysis for simultaneously visualizing strain groupings and the contributing *m/z* peaks. To generate the MAN networks, we created a data matrix, composed of rows representing strains and columns representing the presence or absence of binned *m/z* peaks, we then converted this matrix from wide to long format. This created a two-column matrix:

"Wide" Format

	Peak_1	**Peak_2**	**Peak_3**	**Peak_4**	**Peak_5**
Sample_1	Present	Present	Absent	Present	Absent
Sample_2	Absent	Present	Present	Absent	Present

"Long" Format

Source	**Target**
Sample_1	Peak_1
Sample_1	Peak_2
Sample_2	Peak_2
Sample_2	Peak_3
Sample_1	Peak_4
Sample_2	Peak_5

This created an edge-node relationship when rows representing peak absence were removed. A network of nodes representing strains and peaks remained, with the former connected to the latter only if the strain contained that peak within its spectra(um). A simple example of how MANs were created in the R-language is shown below. This concept can be applied to multiple analytical platforms and has thus far been successfully utilized to analyze MALDI-TOF-MS, FTIR, LC-UV, LCMS, and GNPS-created molecular networking datasets, but needn't be limited to these.

An example in R follows:

```
library("networkD3")
library("reshape2")
peakMatrix<-
rbind(c("Present","Present","Absent","Present","Absent"),c("Absent","Present","Present","Absent","Present"))
rownames(peakMatrix)<-c("Bacteria_1","Bacteria_2")
colnames(peakMatrix)<-c("Peak_1","Peak_2","Peak_3","Peak_4","Peak_5")
print(peakMatrix)
cpn<-melt(peakMatrix)
cpn<-subset(cpn, value!="Absent")
colnames(cpn)<-c("Source","Target","Present/Absent")
print(cpn)
simpleNetwork(cpn,zoom=TRUE)
```

We also used the mathematical function below (the inverse of inter-sample peak occurrence) to weight *m/z* values according to their frequency of appearance among analyzed strains. This functions to draw “biomarker” peaks closer to their source sample while reducing the influence of the most frequently occurring peaks.

```
bool[,colnames(bool)] <- sapply(bool[,colnames(bool)],function(x) ifelse(x==1,1/sum(x),x))
```

Note: “bool” is a data matrix where rows represent samples and columns contain *m/z* peaks. A value of 1 denominates peak presence while 0 represents no peak occurring in that sample.

Text S2: Expanded methods section.

Sponge collection and processing. A freshwater sponge sample was collected June 6th, 2016 from Marinette, Wisconsin (45º5'16.012"N, 87º35'10.468"W), from pilings near Red Arrow Beach at a depth of 8 ft using SCUBA. The sponge was separated from associated macro-organisms, rinsed with filter sterilized Lake Michigan water five times to remove bacteria from surrounding lake water, and most of the water expelled from the sponge by applying gentle pressure. A 1 cm^3 section of tissue and 10 mL of sterile 20% glycerol solution were ground for two minutes using an autoclaved mortar and pestle. A 60 ºC dry bath was utilized to pretreat a 500 µL aliquot for 9 minutes. The sample was then diluted 1:10 with 20% sterile glycerol solution. To an agar plate containing A1 nutrient media (5 g of soluble starch, 2 g of yeast extract, 1 g of peptone, 250 mL of filter-sterilized Lake Michigan water and 250 mL of distilled water), 50 µL of sample was added and spread across the surface. The plate was sealed with Parafilm and left at 27 ºC for 90 days.

MALDI-TOF MS sample preparation. For MALDI-TOF MS analysis, proteins were extracted using an extended direct transfer method that included a formic acid overlay (7, 8). Using a sterile toothpick, bacterial colonies that grew on nutrient agar were applied as a thin film onto a MALDI ground-steel target plate (Bruker Daltonics, Billerica, MA). Over each bacterial smear, 1 µL of 70% LC-MS grade formic acid (Optima, Fisher Chemical) was added and allowed to evaporate, followed by the addition and subsequent evaporation of 1 µL of 10 mg/mL α-cyano-4-hydroxycinnamic acid (CHCA; recrystallized from the 98% pure Sigma-Aldrich) solubilized in 50% acetonitrile, 2.5% trifluoroacetic acid and 47.5% water (7, 8). All solvents were HPLC or MS grade.

MALDI-TOF data acquisition. Measurements were performed using an Autoflex Speed LRF mass spectrometer (Bruker Daltonics) equipped with a smartbeam™-II laser (355 nm). Detailed instrument settings are available in the Supplementary Information. Specialized metabolite spectra were recorded in positive reflectron mode (5000 shots; RepRate: 2000 Hz; delay: 9297 ns; ion source 1 voltage: 19 kV; ion source 2 voltage:16.55 kV; lens voltage: 8.3 kV; mass range: 50 Da to 2,700 Da, matrix suppression cutoff: 50 Da). Protein spectra were recorded in positive linear mode (1200 shots; RepRate: 1000; delay: 29793 ns; ion source 1 voltage: 19.5 kV; ion source 2 voltage: 18.2 kV; lens voltage: 7.5 kV; mass range: 1.9 kDa to 2.1 kDa, matrix suppression cutoff: 1.5 kDa). Protein spectra were corrected with external Bruker

Daltonics bacterial test standard (BTS). Specialized metabolite spectra were corrected with external Bruker Daltonics peptide calibration standard and CHCA $[2M+H]^+$ (379.0930 Da).

Automated data acquisitions were performed using flexControl software v. 3.4.135.0 (Bruker Daltonics) and flexAnalysis software v. 3.4. Spectra were automatically evaluated during acquisition to determine whether a spectrum was of high enough quality to retain and add to the sum of the sample acquisition. The number of added spectra, quality requirements and other detailed acquisition settings are available in Table S3; for flexControl and flexAnalysis scripts, see the "Publication Code and Data Availability" section in the main manuscript.

16S rRNA gene sequence analysis and data workup. Bacterial isolates were characterized by analysis of the 16S rRNA gene. Total genomic DNA was extracted using the Ultra Clean Microbial DNA Isolation kit (MOBIO Laboratories) according to the manufacturer's instructions. The primers 8F, FC27, 1100F, RC1492 and 519R were used for amplifying the 16S rRNA gene. The polymerase chain reaction (PCR) conditions were as follows: initial denaturation at 95°C for 5 minutes, followed by 35 cycles of denaturation at 95°C for 15 seconds, annealing at 60°C for 15 seconds, and extension at 72°C for 15 seconds, and a final extension step at 72°C for 2 minutes. PCR products were purified using QIAquick PCR Purification kit from Qiagen and the amplicons sequenced by Sanger sequencing. Geneious V10.0.9 software was used to produce a consensus sequence of the 16S rRNA gene amplified for each bacterial isolate and the identity of the isolates determined using BLASTn. All sequences were submitted to GenBank and accession numbers along with information regarding each of the isolates' origin is available in Table S2. Phylogenetic trees were created by trimming aligned sequences (SILVA Incremental Aligner) (5) to equal length and using Geneious' "Tree Builder" with Jukes-Cantor and Neighbor Joining algorithms with 100 bootstrap replicates using a support threshold of 80%.

Extraction of *Micromonospora* isolates for LC-MS/MS analysis. Extractions were performed from bacterial cultures growing on solid A1 agar media following the protocol of Bligh, E. G. and Dyer, W. J. (9). Agar cultures were divided into 1 cm^3 pieces and 3 mm glass beads were added. Extraction solvent was added in three steps with vigorous vortexing between steps 1) 1:2 (v/v) $CHCl_3$:MeOH, 2) $CHCl_3$ in 1/3 the added volume of step one, 3) H_2O in 1/3 the added volume of step one. From the resulting two-layer liquid partition, the organic layer was retained for further analysis.

LC-MS/MS analysis of bacterial extracts. *Micromonospora* extracts were analyzed via LC-MS/MS with a method adapted from that described by Goering et al (10). Experiments were

performed on an Agilent 1200 workstation connected to a Thermo Fisher Scientific Q-Exactive mass spectrometer with an electrospray ionization source. Reversed-phase chromatography was performed by injection of 20 µL of 0.1 mg/mL of extract at a 0.3 mL/min flow rate across a Phenomenex Kinetex C18 RPLC column (150 mm x 2.1 mm i.d., 2 µm particle size). Mobile phase A was water with 0.1% formic acid and mobile phase B was acetonitrile with 0.1% formic acid. Mobile phase B was held at 15% for 1 minute, then adjusted to 95% over 12 minutes, where it was held for 2 minutes, and the system re-equilibrated for 5 minutes. The mass spectrometry parameters were as follows: scan range 200-2000 *m/z*, resolution 35,000, scan rate ~3.7 per second. Data were gathered in profile and the top 5 most intense peaks in each full spectrum were targeted for fragmentation that employed a collision energy setting of 25 eV for Higher-energy Collisional Dissociation (HCD) and isolation window of 2.0 *m/z*. Data were converted to mzXML and uploaded to the GNPS: Global Natural Products Social Molecular Networking platform for dereplication (11) and XCMS (12) for comparative metabolomics. For analysis in R, GNPS library matches were downloaded and converted to mzML, where necessary, with MSConvert. “Amphiphilic ferrioxamine” mzML files were manually edited in Notepad++ to the correct XML schema.

Supplementary References

1. Traxler MF, Watrous JD, Alexandrov T, Dorrestein PC, Kolter R (2013) Interspecies Interactions Stimulate Diversification of the *Streptomyces coelicolor* Secreted Metabolome. *MBio* 4(4):e00459-13-e00459-13.

2. Sidebottom AM, Johnson AR, Karty JA, Trader DJ, Carlson EE (2013) Integrated Metabolomics Approach Facilitates Discovery of an Unpredicted Natural Product Suite from *Streptomyces coelicolor* M145. *ACS Chem Biol* 8(9):2009–2016.

3. Strohalm M, Kavan D, Novák P, Volný M, Havlíček V (2010) *mMass* 3: A Cross-Platform Software Environment for Precise Analysis of Mass Spectrometric Data. *Anal Chem* 82(11):4648–4651.

4. Meija J, et al. (2016) Isotopic Compositions of the Elements 2013 (IUPAC Technical Report). *Pure Appl Chem* 88(3):293–306.

5. Pruesse E, Peplies J, Glöckner FO (2012) SINA: Accurate High-Throughput Multiple Sequence Alignment of Ribosomal RNA Genes. *Bioinformatics* 28(14):1823–1829.

6. Erickson K (2010) The Jukes-Cantor Model of Molecular Evolution. *PRIMUS* 20(5):438–445.

7. Freiwald A, Sauer S (2009) Phylogenetic Classification and Identification of Bacteria by Mass Spectrometry. *Nat Protoc* 4(5):732–742.

8. Schumann P, Maier T (2014) MALDI-TOF Mass Spectrometry Applied to Classification and Identification of Bacteria. *Methods Microbiol* 41:275–306.

9. Bligh EG, Dyer WJ (1959) A Rapid Method of Total Lipid Extraction and Purification. *Can J Biochem Physiol* 37(8):911–917.

10. Goering AW, et al. (2016) Metabologenomics: Correlation of Microbial Gene Clusters with Metabolites Drives Discovery of a Nonribosomal Peptide with an Unusual Amino Acid Monomer. *ACS Cent Sci* 2(2):99–108.

11. Wang M, et al. (2016) Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* 34(8):828–837.

12. Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G (2012) XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Anal Chem* 84(11):5035–5039.