

# IDBac

Bioinformatics Software for Microbial  
Drug Discovery Prioritization and  
Culturomics Characterization

Chase Clark  
PhD Candidate  
Murphy Lab

Dept. of Pharmaceutical Sciences  
Center for Biomolecular Sciences  
University of Illinois at Chicago



@ChasingMicrobes



Photos of slides OK



# Outline

- Introduction to IDBac
- Applying IDBac to prioritize bacteria isolates from freshwater sponges
- Using IDBac beyond prioritizing microbial libraries



# We Scour the Globe for New Bacteria, New Drug Leads



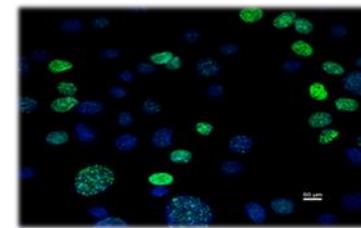
# Discovery Pipeline (Simplified)



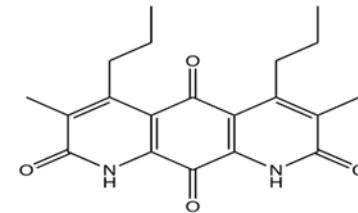
Collect  
Samples



Grow  
Bacteria



Antibiotic  
Activity?



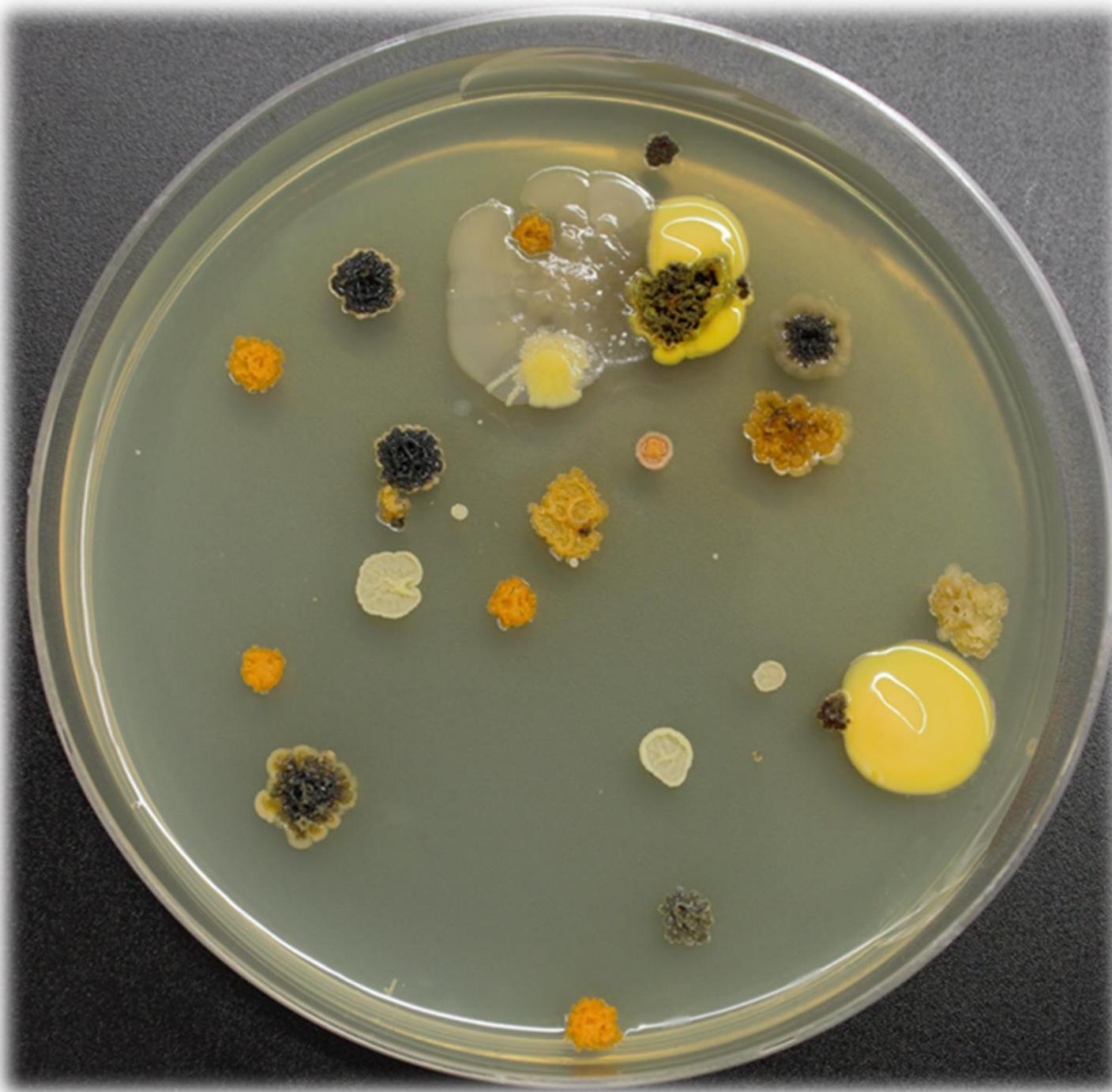
Elucidate  
Antibiotic



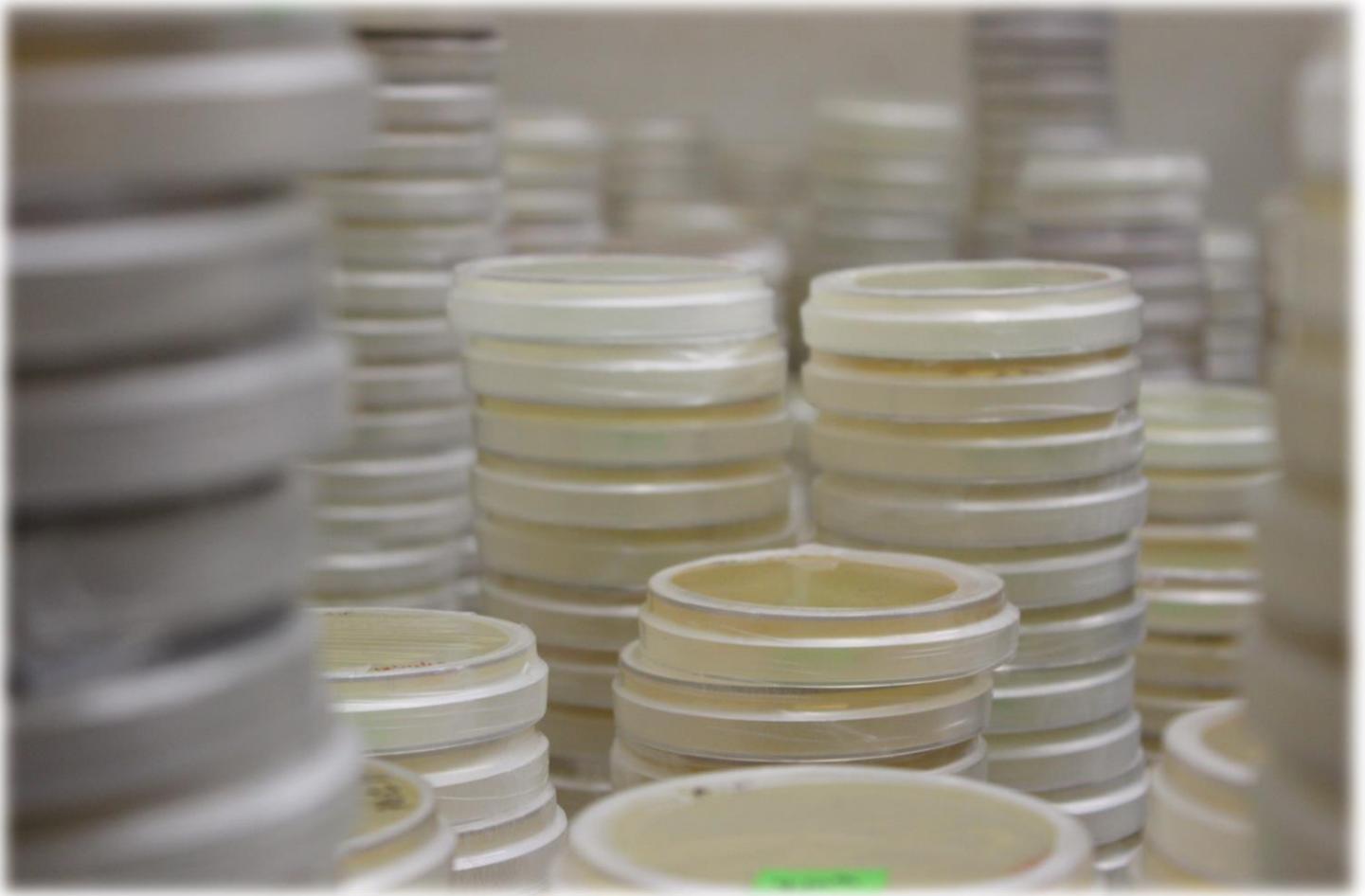
Drug

Expensive!  
Can take weeks to months

Labs have trouble prioritizing which bacteria to study.

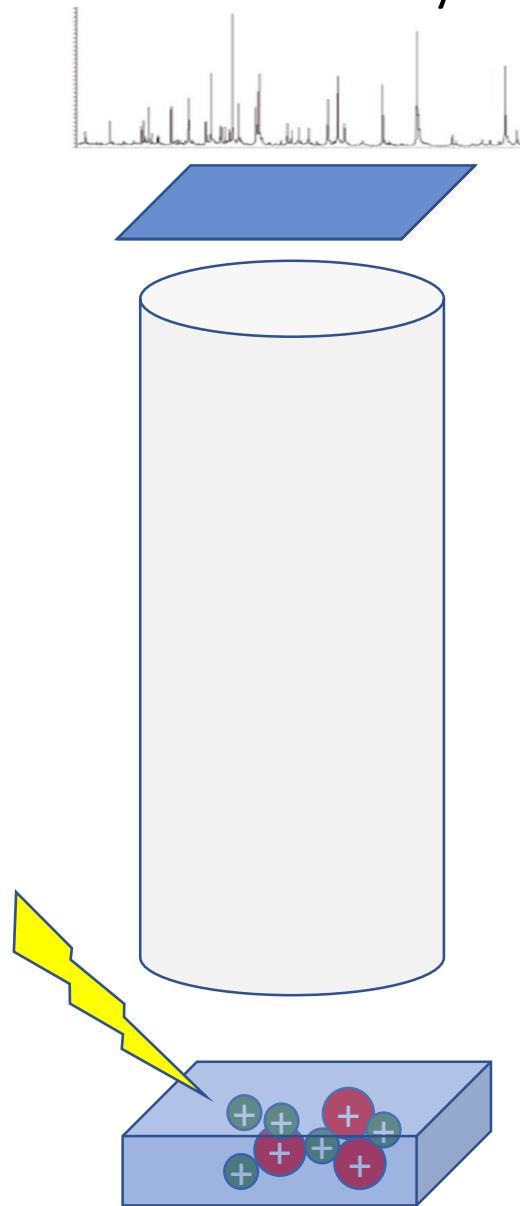


Labs have trouble prioritizing which bacteria to study.



# MALDI-TOF Mass Spectrometry is Fast, Cheap, and Easy

(We can collect data on every isolatable colony)



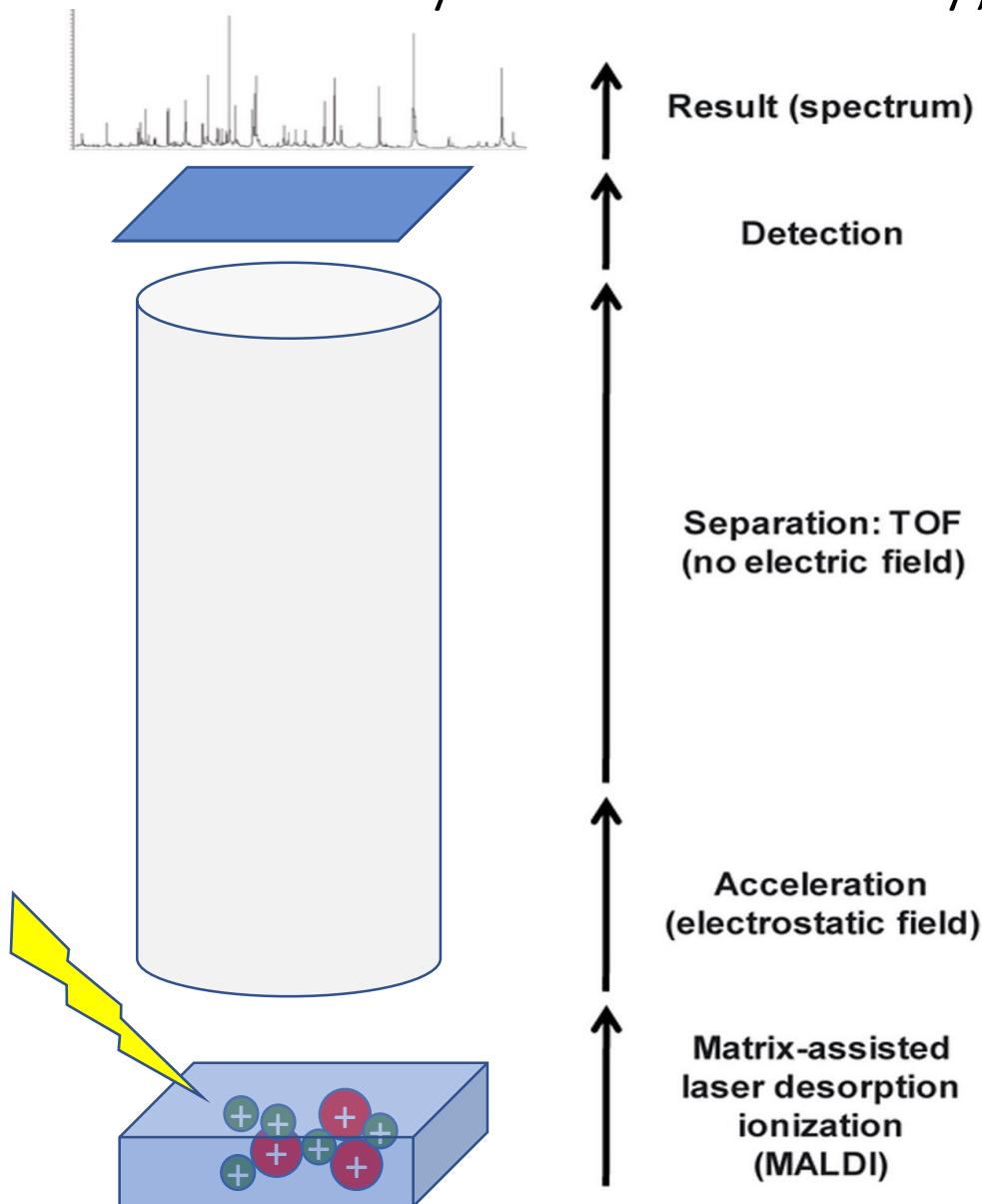
# MALDI-TOF Mass Spectrometry is Fast, Cheap, and Easy

(We can collect data on every isolatable colony)



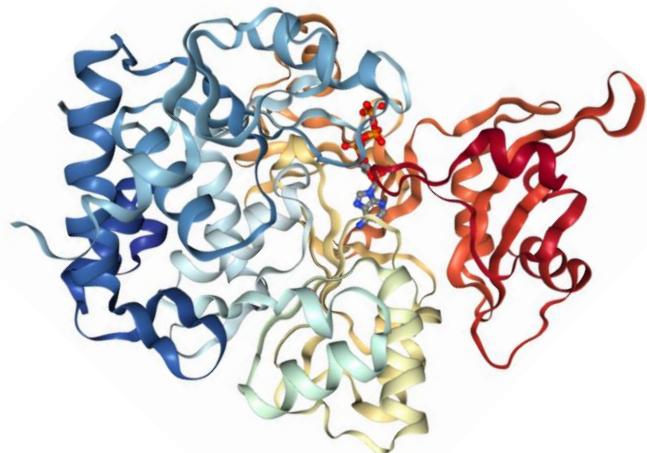
# MALDI-TOF Mass Spectrometry is Fast, Cheap, and Easy

(We can collect data on every isolatable colony)

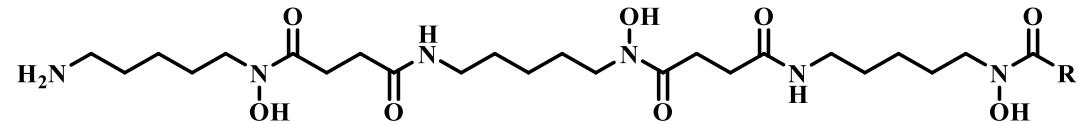


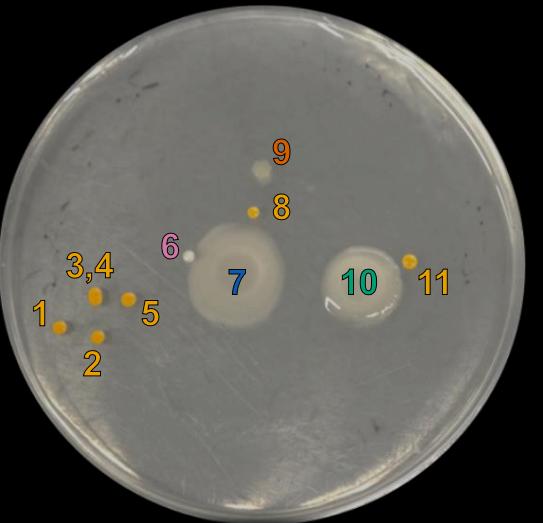
# We Collect Two Fingerprints For Every Colony

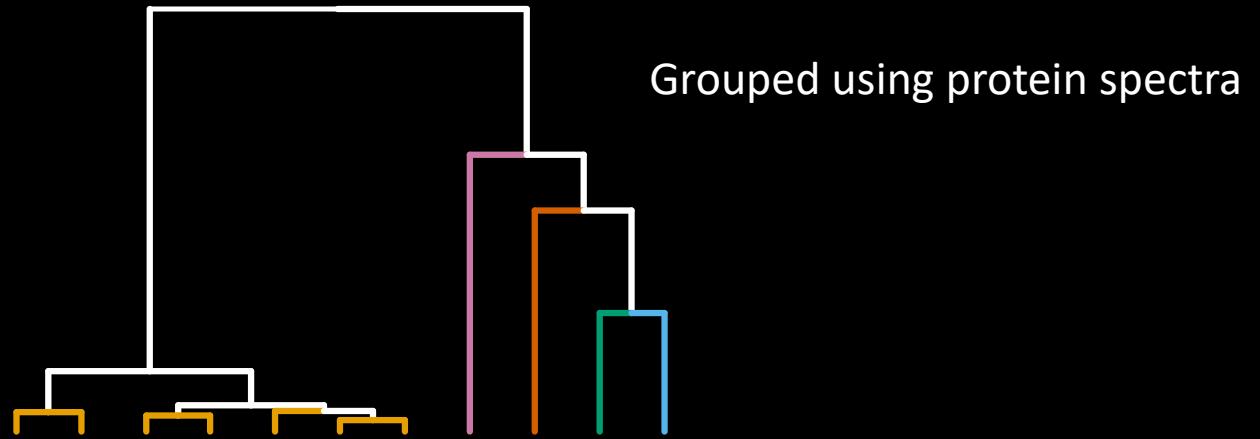
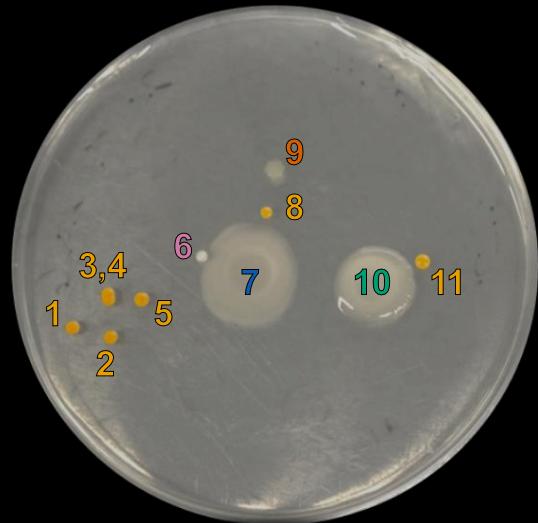
Protein Spectra  
2-20 kDa  
Pseudo-Phylogenetic Groupings



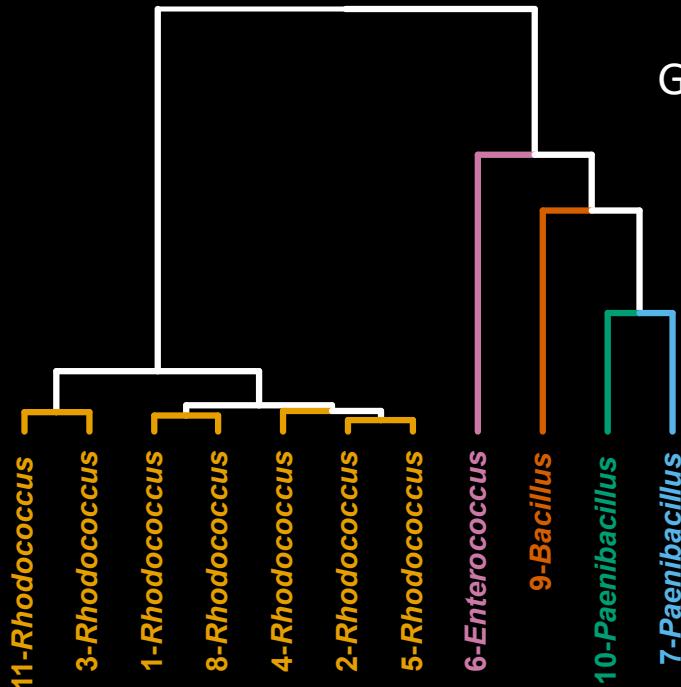
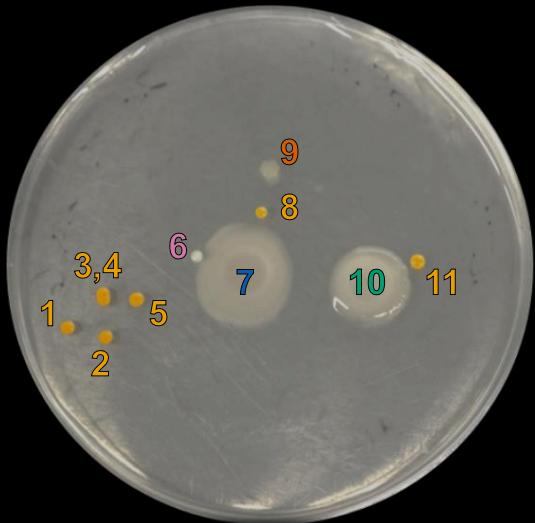
Small Molecule Spectra  
0.2-2 kDa  
Measuring Produced Metabolites



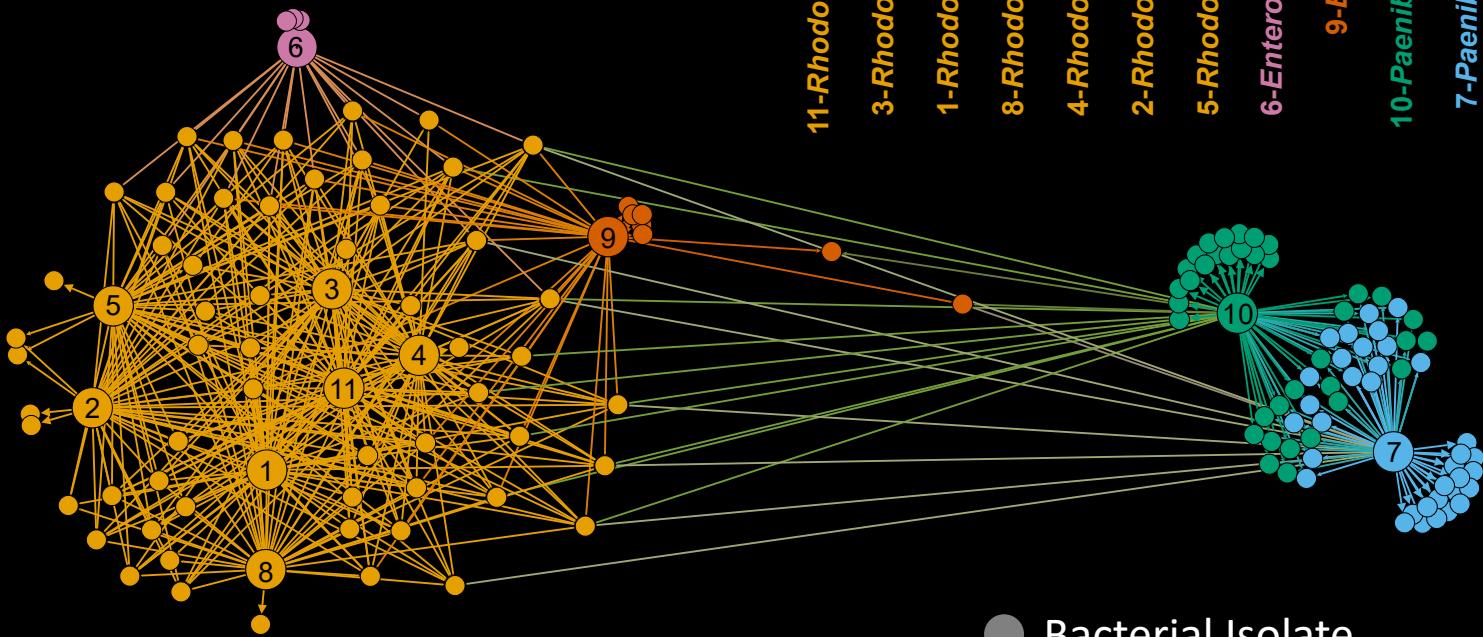
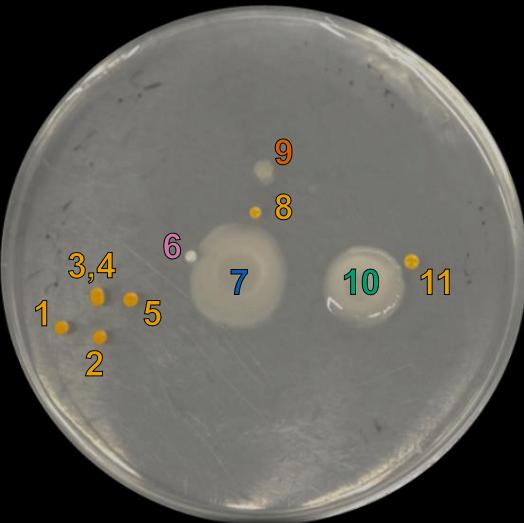




Grouped using protein spectra

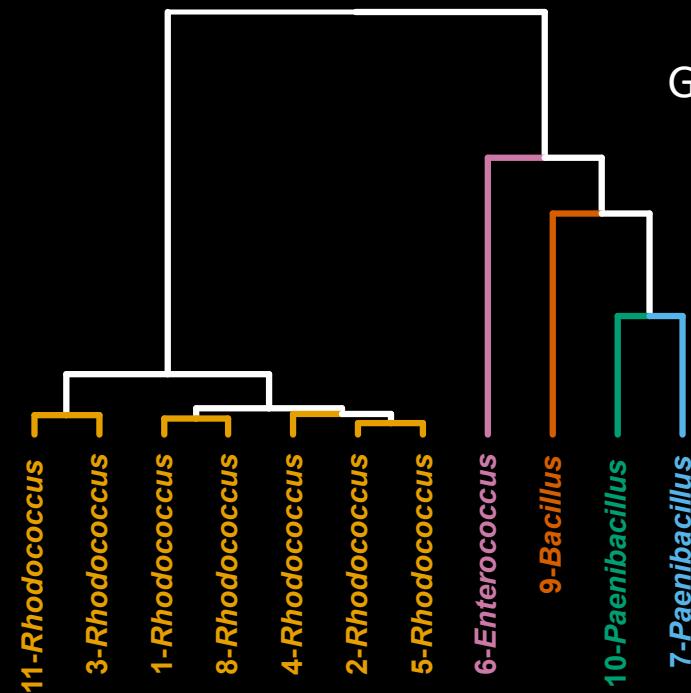


Grouped using protein spectra



Visualized small molecule data

● Bacterial Isolate  
●  $m/z$  Peak



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

Chase M. Clark<sup>a,1</sup>, Maria S. Costa<sup>b,1</sup>, Laura M. Sanchez<sup>c,2</sup>, and Brian T. Murphy<sup>a,2</sup><sup>a</sup>Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL; and <sup>b</sup>Faculty of Pharmaceutical Sciences, University of Iceland, Hagi, IS-107 Reykjavik, Iceland

Edited by Jerryd Meinwald, Cornell University, Ithaca, NY, and approved April 5, 2018 (received for review January 22, 2018)

For decades, researchers have used the rapid and accurate identification of microbial identity with bacterial metabolism. Since specialized metabolites are critical to bacterial function and survival in the environment, we designed a data acquisition and bioinformatics technique (IDBac) that utilizes in situ matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) to analyze protein and specialized metabolite spectra recorded from single bacterial colonies picked from agar plates. We demonstrated the power of our approach by discriminating between two *Bacillus subtilis* strains in ~30 min using MALDI-TOF MS and IDBac.

Specialized metabolites represent functional traits in bacteria, and these molecules are useful for defining “what a bacterium does,” rather than “who it is” in the context of its immediate environment. Clark et al. (3) analyzed 16S rRNA genes from within a group of three closely related species that shared 99% 16S rRNA gene sequence identity: *Salinipora arenicola*, *Salinipora pacifica*, and *Salinipora tropica*. The isolates were predicted to contain a surprising 229 distinct specialized metabolite biosynthetic clusters, many of which were active through secret horizontal gene transfer events and occurred in only one or two isolates in the group. Given this large potential chemical diversity harbored within *Salinipora* genomes, these findings highlight the limitations of using phylogenetic approaches based on 16S rRNA genes to infer bacterial function and stress the need for alternative approaches to rapidly assess chemical differences between closely related bacterial isolates.

## Significance

Mass spectrometry is a powerful technique that has been used to identify bacteria by their protein content and to assess bacterial function by the analysis of their specialized metabolites. However, until now these analyses have operated independently, which has resulted in the inability to rapidly connect bacterial phylogenetic identity with potential environmental function. To bridge this gap, we designed a MALDI-TOF mass spectrometry-based bioinformatics pipeline (IDBac) to integrate data from both intact protein and specialized metabolite spectra directly from bacterial cells grown on agar. This technique organizes bacteria into highly similar phylogenetic groups and allows for comparison of metabolic differences of hundreds of isolates in just a few hours.

natural products | specialized metabolites | mass spectrometry | bioinformatics | metabolomics

**F**or nearly two centuries researchers have studied bacteria to diagnose and treat diseases, elucidate intricate interspecies and intraspecies evolutionary processes, manage and develop agricultural biotechnological practices, and, broadly speaking, learn about the complex roles of microorganisms in the environment.

<https://doi.org/10.1073/pnas.1801247115>

**JOURNAL OF NATURAL PRODUCTS** Article pubs.acs.org/jnp

**Minimizing Taxonomic and Natural Product Redundancy in Microbial Libraries Using MALDI-TOF MS and the Bioinformatics Pipeline IDBac**

Maria S. Costa,<sup>1,2</sup> Chase M. Clark,<sup>2,3</sup> Sesselia Ómarsdóttir,<sup>1</sup> Laura M. Sanchez,<sup>2,3</sup> and Brian T. Murphy<sup>3,4</sup>

<sup>1</sup>Faculty of Pharmaceutical Sciences, University of Iceland, Hagi, Hofsvallagata 53, IS-107 Reykjavik, Iceland  
<sup>2</sup>Department of Pharmaceutical Sciences, College of Pharmacy, University of Illinois at Chicago, 833 South Wood Street (MC 781), Room 539, Chicago, Illinois 60607, United States  
<sup>3</sup>Supporting Information

**ABSTRACT:** Libraries of microorganisms have been a cornerstone of drug discovery efforts since the mid-1950s, but strain duplication in some libraries has resulted in unwanted natural product redundancy. In the current study, we implemented a workflow that minimizes both the natural product overlap and the total number of bacterial isolates in a library. Using a collection of 86 environmental samples, we purified over 800 distinct bacterial colony isolates derived from 86 environmental samples. We employed our mass spectrometry (MS)-based IDBac workflow on these isolates to form groups of taxa based on protein MS fingerprints (3–15 kDa) and further distinguished taxa subgroups based on their degree of overlap within corresponding natural product spectra (0.2–2 kDa). This informed the decision to create a library of 301 isolates spanning 54 genera. This process required over 25 h of data acquisition and 2 h of analysis. In a separate experiment, we reduced the size of the library by 60% by using the IDBac pipeline to generate protein MS spectra of bacterial colonies (from 833 to 233 isolates, a 72.0% size reduction). Overall, our pipeline allows for a significant reduction in costs associated with library generation and minimizes natural product redundancy entering into downstream biological screening efforts.

<https://doi.org/10.1021/acs.jnatprod.9b00168>



## A Call to Action: the Need for Standardization in Developing Open-Source Mass Spectrometry-Based Methods for Microbial Subspecies Discrimination

Chase M. Clark,<sup>a</sup> Brian T. Murphy,<sup>a</sup> Laura M. Sanchez<sup>a</sup><sup>a</sup>Department of Pharmaceutical Sciences, University of Illinois at Chicago, Chicago, Illinois, USA

KEYWORDS MALDI-TOF MS, bioinformatics, dereplication, microbial ecology

In the last decade, there has been a renewed push by academic researchers to create rapid and accurate techniques to differentiate, identify, and prioritize culturable microbial isolates. One such technique that continues to gain momentum among microbiologists is matrix-assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF MS). It is an established, inexpensive technique commonly used to rapidly identify microbial taxa and differentiate culturable microbes. This technology has become commonplace in clinical and veterinary laboratories where rigorously validated methods are used in conjunction with commercially available reference databases to identify pathogenic microorganisms. However, the broader community, especially laboratories working with environmental microbes, typically cannot access the expensive software and databases. It is our opinion that this community, which relies on free and open-source software, currently lacks a coherent set of accepted experimental practices, including employment of internal standard strains, statistically driven determination of biological and technical replicates, and deposition of MS data into open-access repositories. Establishing guidelines would enable researchers to better compare microbial typing methods and advance our ability to group and delineate environmental isolates in an effective manner, particularly at the subspecies level.

<https://doi.org/10.1128/mSystems.00813-19>

**jove** Search 10,747 video articles Advanced

BIOPHARMA NEW ABOUT Jove FOR LIBRARIANS PUBLISH VIDEO JOURNAL SCIENCE EDUCATION

ABSTRACT INTRODUCTION PROTOCOL RESULTS DISCUSSION MATERIALS REFERENCES DOWNLOADS

**BIOCHEMISTRY**

Using the Open-Source MALDI TOF-MS IDBac Pipeline for Analysis of Microbial Protein and Specialized Metabolite Data

Chase M. Clark<sup>1</sup>, Maria S. Costa<sup>1,2</sup>, Erin Conley<sup>1</sup>, Emma Li<sup>1</sup>, Laura M. Sanchez<sup>1</sup>, Brian T. Murphy<sup>1</sup>

<sup>1</sup>Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, <sup>2</sup>Faculty of Pharmaceutical Sciences, University of Iceland

This content is OPEN ACCESS.

CHAPTERS

- 0.04 Title
- 1.10 Preparation of MALDI Target Plates and Data Acquisition
- 2.02 Installing the IDBac Software and Starting with Raw Data
- 2.47 Work with Previous Experiments
- 3.22 Setting up Protein Data Analysis and Creating Mirror Plots
- 4.07 Clustering Samples Using Protein Data
- 4.49 Customizing the Protein Dendrogram and Inserting Samples from a Separate Experiment into the Dendrogram
- 5.32 Analyzing Specialized Metabolite Data and Metabolite Association Networks (MANs)

ISSUE 147 DOI: 10.3791/59219 PUBLISHED: 5/15/2019 0 COMMENTS PDF EMBED ADD TO FAVORITES

bioRxiv preprint doi: https://doi.org/10.1101/59219; this version posted May 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

<https://doi.org/10.3791/59219>

# IDBac Publications

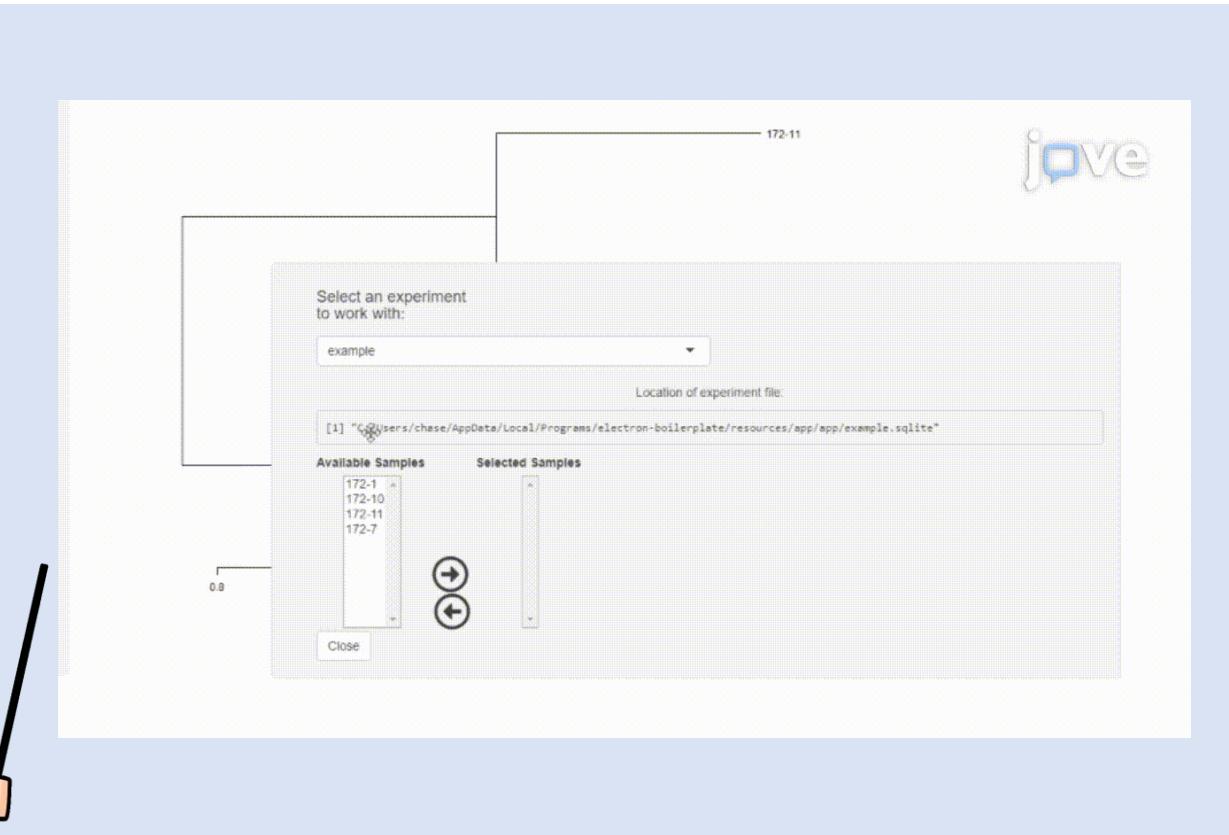
<https://chasemc.github.io/IDBac>

# To Learn More About the Protocol and Software

A screenshot of a Jove video article page. The title is "Using the Open-Source MALDI TOF-MS IDBac Pipeline for Analysis of Microbial Protein and Specialized Metabolite Data". The abstract section includes a biochemistry icon, the authors (Chase M. Clark<sup>1</sup>, Maria S. Costa<sup>1,2</sup>, Erin Conley<sup>1</sup>, Emma Li<sup>1</sup>, Laura M. Sanchez<sup>2</sup>, Brian T. Murphy<sup>1</sup>), and the institutions (<sup>1</sup>Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, <sup>2</sup>Faculty of Pharmaceutical Sciences, University of Iceland). The video player shows a thumbnail of the software interface with a protein dendrogram and metabolite data. The sidebar contains chapters such as "Preparation of MALDI Target Plates and Data Acquisition", "Installing the IDBac Software and Starting with Raw Data", and "Analyzing Specialized Metabolite Data and Metabolite Association Networks (MANs)".

<https://www.jove.com/video/59219/using-open-source-maldi-tof-ms-idbac-pipeline-for-analysis-microbial>

# To Learn More About the Protocol and Software



A screenshot of a video article page on the jove website. The title is "Using the Open-Source MALDI TOF-MS IDBac Pipeline for Analysis of Microbial Protein and Specialized Metabolite Data" by Chase M. Clark, Maria S. Costa, Erin Conley, Emma Li, Laura M. Sanchez, Brian T. Murphy. The page includes sections for ABSTRACT, INTRODUCTION, PROTOCOL, RESULTS, DISCUSSION, MATERIALS, REFERENCES, and DOWNLOADS. The protocol section features a video player showing a dendrogram analysis. The sidebar on the right lists chapters such as "Preparation of MALDI Target Plates and Data Acquisition", "Installing the IDBac Software and Starting with Raw Data", and "Analyzing Specialized Metabolite Data and Metabolite Association Networks (MANs)".

<https://www.jove.com/video/59219/using-open-source-maldi-tof-ms-idbac-pipeline-for-analysis-microbial>

Programming with IDBac x +

127.0.0.1:4984

Programming With IDBac

1 Preamble

- 1.1 Major points:
- 1.2 Download IDBac example file

2 Connect to an IDBac Database

- 2.1 Connect to IDBac database

3 IDBac Databases Explained

- 3.1 Database Tables
- 3.2 How do I use this????

4 Add Data to an IDBac Database

- 4.1 Make a new empty IDBac datab...
- 4.2 Add data from an mzXML file
- 4.3 Adding multiple mzXML files

5 Starting With Bruker Data

6 Simple Analysis

- 6.1 Check database
- 6.2 What samples are in the datab...

7 Working with IDBac from Python

IDBac Website

IDBac PNAS Publication

IDBac Video Protocol

Published with bookdown

# Programming with IDBac

Chase Clark

2020-02-15

## Chapter 1 Preamble

While some familiarity with R is suggested, the examples laid out within this book should be explanatory enough for a novice to comfortable working through.

Suggestions or additions for content are welcome and may contributed at [github.com/chasemc/programmingidbac](https://github.com/chasemc/programmingidbac). Note that this project has a contributor covenant TODO.

### 1.1 Major points:

Things you want to do will revolve around:

- Creating an IDBac database from your raw (or converted) data
- Moving data from one IDBac database to another
- Accessing spectra
- Accessing peak-picked data
- Filtering data by some attribute

### 1.2 Download IDBac example file

The data used in this book uses example data that can be found here:  
<ftp://massive.ucsd.edu/MSV000084291>

```
library(here)
```

Programming with IDBac

127.0.0.1:4984

Programming With IDBac

1 Preamble

- 1.1 Major points:
- 1.2 Download IDBac example file

2 Connect to an IDBac Database

- 2.1 Connect to IDBac database

3 IDBac Databases Explained

- 3.1 Database Tables
- 3.2 How do I use this????

4 Add Data to an IDBac Database

- 4.1 Make a new empty IDBac database
- 4.2 Add data from an mzXML file
- 4.3 Adding multiple mzXML files

5 Starting With Bruker Data

6 Simple Analysis

- 6.1 Check database
- 6.2 What samples are in the database?

7 Working with IDBac from Python

IDBac Website

IDBac PNAS Publication

IDBac Video Protocol

Published with bookdown

# Programming with IDBac

Chase Clark

2020-02-15

## Chapter 1 Preamble

While some familiarity with R is suggested, the examples are enough for a novice to comfortably work through.

Suggestions or additions for content are welcome and can be submitted to [github.com/chasemc/programming\\_idbac](https://github.com/chasemc/programming_idbac). Note that this

### 1.1 Major points:

Things you want to do will revolve around:

- Creating an IDBac database from your raw (or converted) data
- Moving data from one IDBac database to another
- Accessing spectra
- Accessing peak-picked data
- Filtering data by some attribute

### 1.2 Download IDBac example

The data used in this book uses example data that can be found at <ftp://massive.ucsd.edu/MSV000084291>.

```
library(here)
```

Programming With IDBac

1 Preamble

- 1.1 Major points:
- 1.2 Download IDBac example file

2 Connect to an IDBac Database

- 2.1 Connect to IDBac database

3 IDBac Databases Explained

- 3.1 Database Tables
- 3.1.1 The Version table
- 3.1.2 The locale table
- 3.1.3 The metadata table
- 3.1.4 The mass\_index table
- 3.1.5 The xml table
- 3.1.6 The spectra table
- 3.2 How do I use this????

4 Add Data to an IDBac Database

- 4.1 Make a new empty IDBac database
- 4.2 Add data from an mzXML file
- 4.3 Adding multiple mzXML files

5 Starting With Bruker Data

6 Simple Analysis

- 6.1 Check database
- 6.2 What samples are in the database?

7 Working with IDBac from Python

IDBac Website

IDBac PNAS Publication

IDBac Video Protocol

The `xml` table stores the mzML (or mzxml if that was the input type) files and some basic information about the instrument if that information was available (e.g. information won't be present if spectra were converted from txt files). The `xml` field contains the compressed mzML/mzxml file, in full.

### 3.1.6 The spectra table

```
dplyr::tbl(example_pool,  
           "spectra")
```

```
## # Source: table<spectra> [?? x 42]  
## # Database: sqlite 3.29.0  
## # [C:\Users\chase\Documents\GitHub\programming_idbac\data\example_data\idbac_experiment_file..  
## # spectrum_mass_h... spectrum_intens... xml_hash strain_id peak_matrix  
## <chr> <chr> <chr> <chr> <chr>  
## 1 b83311689e011f6f 6028ecf5d313b045 ef1071a... 172-1 {"\\"mass\\"...  
## 2 b83311689e011f6f bac07d6e5744cf4e ef1071a... 172-1 {"\\"mass\\"...  
## 3 a30959c610fd9278 b7238aaee3626ef85 ef1071a... 172-1 {"\\"mass\\"...  
## 4 a30959c610fd9278 7ec32da4207c396a ef1071a... 172-1 {"\\"mass\\"...  
## 5 b83311689e011f6f 60f61ee538773cda cd9f331... 172-10 {"\\"mass\\"...  
## 6 b83311689e011f6f 48f154a0e6d3a584 cd9f331... 172-10 {"\\"mass\\"...  
## 7 a30959c610fd9278 5d8fc715f0478a5a cd9f331... 172-10 {"\\"mass\\"...  
## 8 a30959c610fd9278 95922f4f84fa25e7 cd9f331... 172-10 {"\\"mass\\"...  
## 9 b83311689e011f6f 1f7851d0ae1debcf 03d7c1f... 172-11 {"\\"mass\\"...  
## 10 b83311689e011f6f 502b1ed5bc032cd9 03d7c1f... 172-11 {"\\"mass\\"...  
## # ... with more rows, and 37 more variables: spectrum_intensity <blob>,  
## # max_mass <int>, min_mass <int>, ignore <int>, number <int>,  
## # time_delay <int>, time_delta <dbl>, calibration_constants <chr>,  
## # v1_to_calibration <chr>, data_type <chr>, data_system <chr>,  
## # spectrometer_type <chr>, inlet <chr>, ionization_mode <chr>,  
## # acquisition_method <chr>, acquisition_date <chr>, acquisition_mode <chr>,  
## # tof_mode <chr>, acquisition_operator_mode <chr>, laser_attenuation <int>,  
## # digitizer_type <chr>, flex_control_version <chr>, id <chr>,
```

Programming with IDBac

x +

127.0.0.1:4984

Programming With IDBac

1 Preamble

- 1.1 Major points:
- 1.2 Download IDBac example file

2 Connect to an IDBac Database

- 2.1 Connect to IDBac database

3 IDBac Databases Explained

- 3.1 Database Tables
- 3.2 How do I use this????

4 Add Data to an IDBac Database

- 4.1 Make a new empty IDBac database
- 4.2 Add data from an mzXML file
- 4.3 Adding multiple mzXML files

5 Starting With Bruker Data

6 Simple Analysis

- 6.1 Check database
- 6.2 What samples are in the database?

7 Working with IDBac from Python

IDBac Website

IDBac PNAS Publication

IDBac Video Protocol

Published with bookdown

Chapter 6 Simple Analysis | Prog

x +

127.0.0.1:4984/simple-analysis.html

Programming With IDBac

1 Preamble

- 1.1 Major points:
- 1.2 Download IDBac example file

2 Connect to an IDBac Database

- 2.1 Connect to IDBac database

3 IDBac Databases Explained

- 3.1 Database Tables
- 3.2 How do I use this????

4 Add Data to an IDBac Database

- 4.1 Make a new empty IDBac database
- 4.2 Add data from an mzXML file
- 4.3 Adding multiple mzXML files

5 Starting With Bruker Data

6 Simple Analysis

- 6.1 Check database
- 6.2 What samples are in the database?

7 Working with IDBac from Python

IDBac Website

IDBac PNAS Publication

IDBac Video Protocol

Published with bookdown

# Programming with IDBac

Chase Clark  
2020-02-15

## Chapter 1 Preamble

While some familiarity with R is suggested, the examples are enough for a novice to comfortably work through.

Suggestions or additions for content are welcome and can be submitted to [github.com/chasemc/programmingidbac](https://github.com/chasemc/programmingidbac). Note that this

### 1.1 Major points:

Things you want to do will revolve around:

- Creating an IDBac database from your raw (or converted) data
- Moving data from one IDBac database to another
- Accessing spectra
- Accessing peak-picked data
- Filtering data by some attribute

### 1.2 Download IDBac example

The data used in this book uses example data that can be found at <ftp://massive.ucsd.edu/MSV000084291>.

```
library(here)
```

## Chapter 6 Simple Analysis

Load necessary packages for this tutorial:

If you haven't already, connect to an IDBac database as shown in "01\_connect-to-idbac-database".

Connect to the database

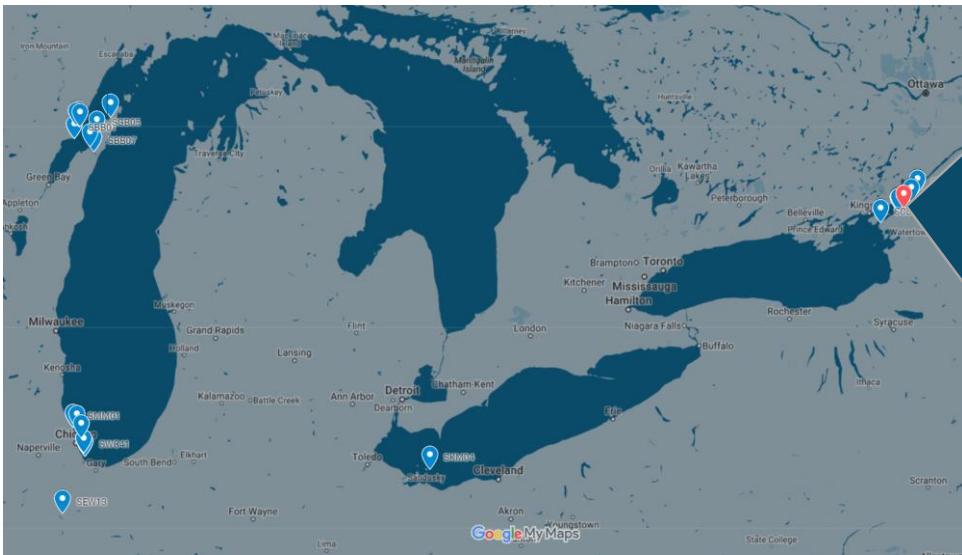
```
example_pool <- IDBacApp::idbac_connect(fileName = "idbac_experiment_file",
                                             filePath = here::here("data",
                                                       "example_data"))
```

```
my_plot <- IDBacApp::assembleMirrorPlots(sampleID1 = "172-7",
                                              sampleID2 = "172-10",
                                              peakPercentPresence = 0.7,
                                              lowerMassCutoff = 3000,
                                              upperMassCutoff = 15000,
                                              minSNR = 4,
                                              tolerance = 0.002,
                                              pool1 = example_pool$idbac_experiment_file,
                                              pool2 = example_pool$idbac_experiment_file)
```

```
IDBacApp::mirrorPlot(my_plot)
```

# **Applying IDBac to the study of freshwater sponge microbiomes**

**5: Years**  
**28: Citizen scientists**  
**93: Freshwater sponges**  
**3: Sponge genera**



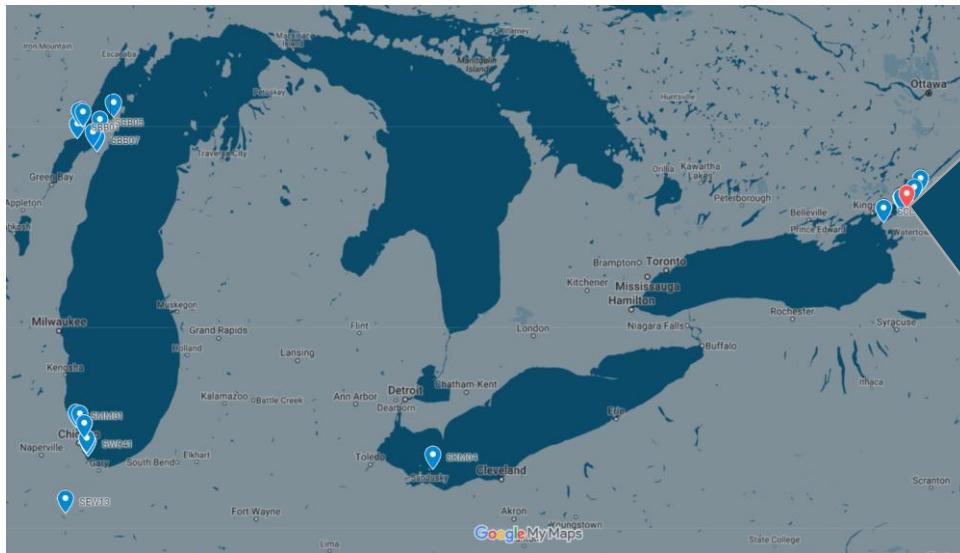
**2:** *Eunapius fragilis* var. *minuta*

**1:** Distance, in miles, of collection sites

**1:** Undergraduate student

**1:** Graduate student

**900:** Bacterial isolates





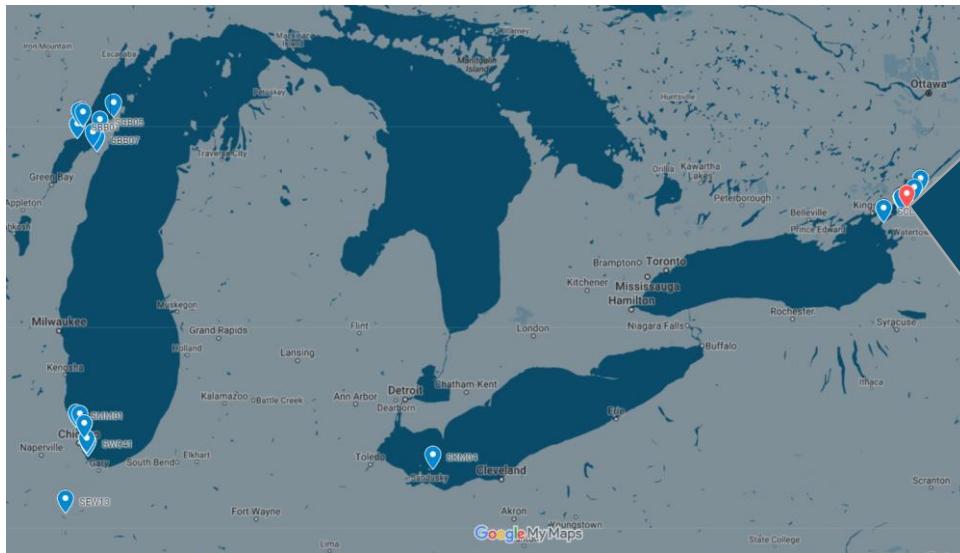
**2:** *Eunapius fragilis* var. *minuta*

**1:** Distance, in miles, of collection sites

**1:** Undergraduate student

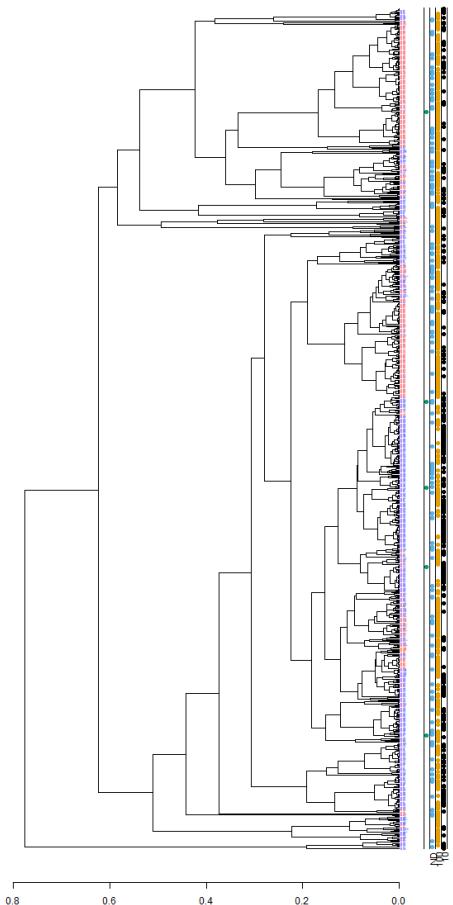
**1:** Graduate student

**900:** Bacterial isolates

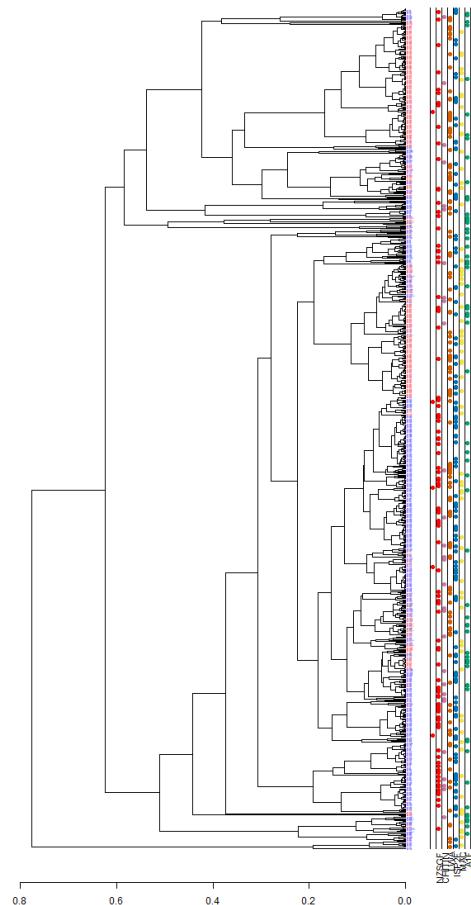


# First Thing to Test: Are These Results Real?

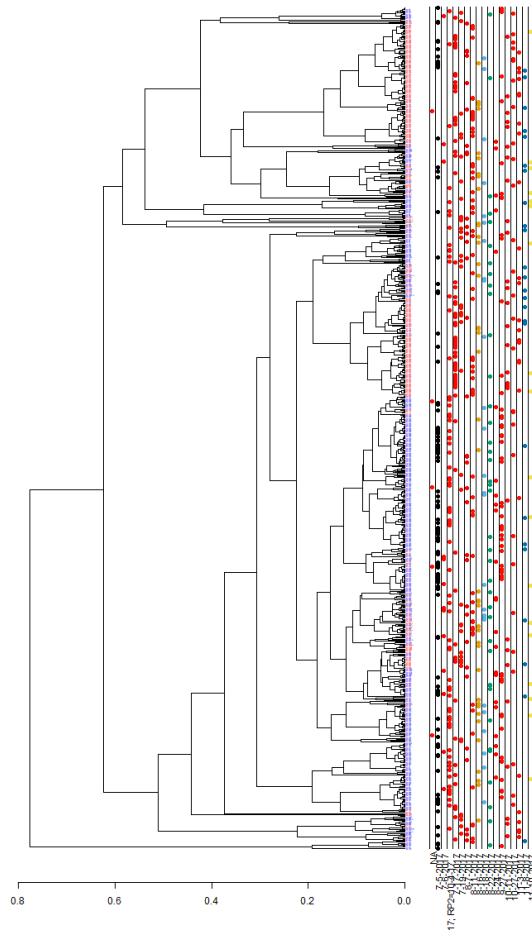
Correlation with sample dilution



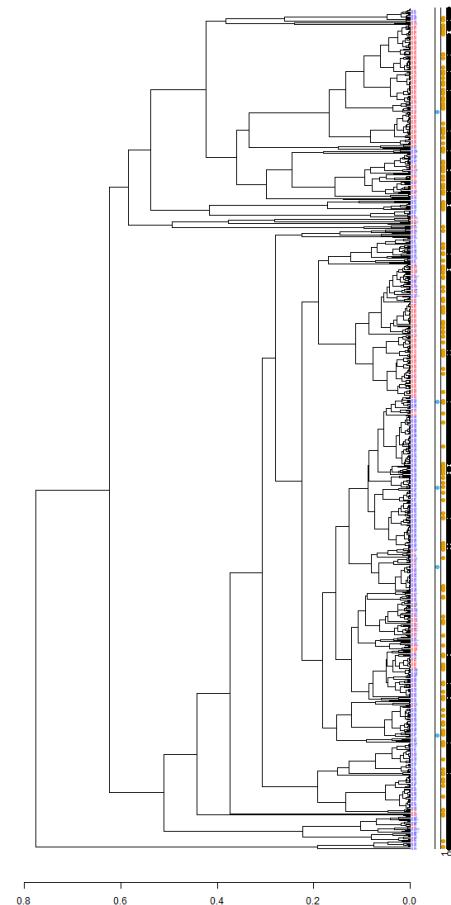
Correlation with isolation media



Correlation with date isolated

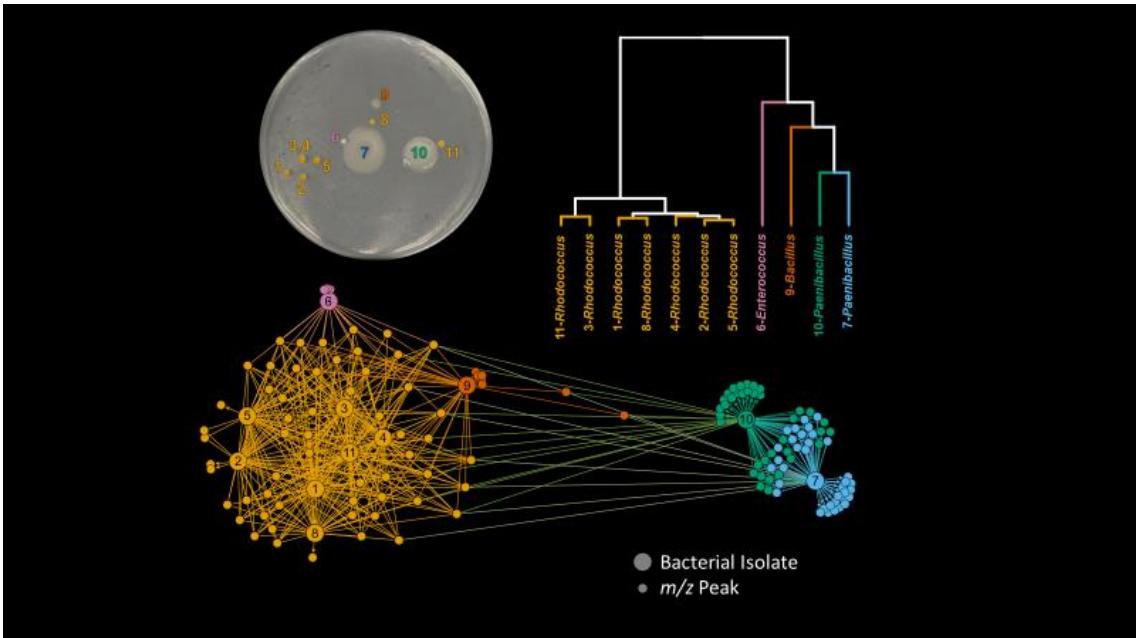


Correlation with heat shock



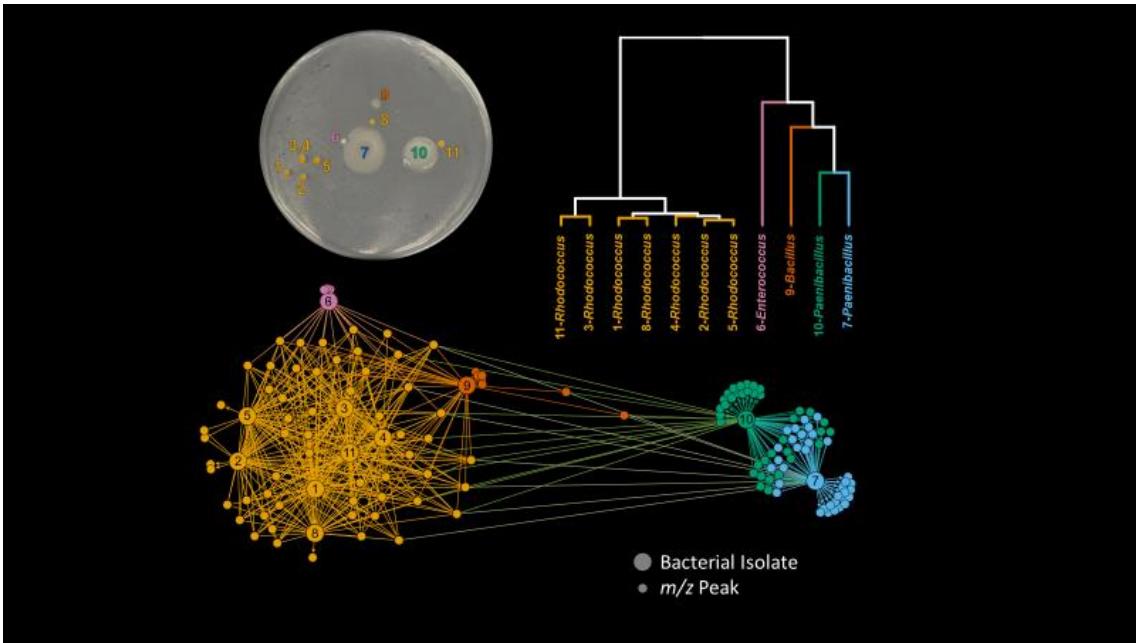
# So, let's choose our strains to study further...

This slide from earlier looked pretty simple.

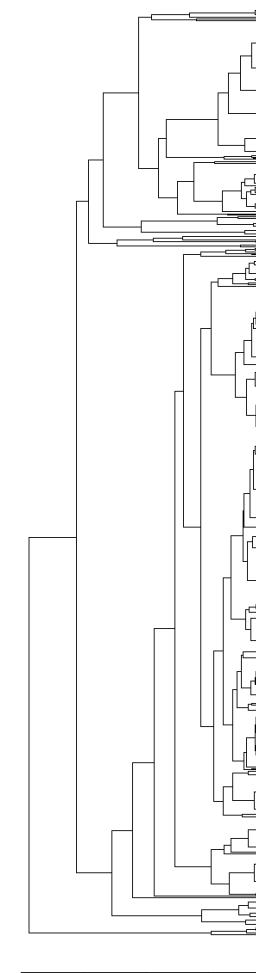


# So, let's choose our strains to study further...

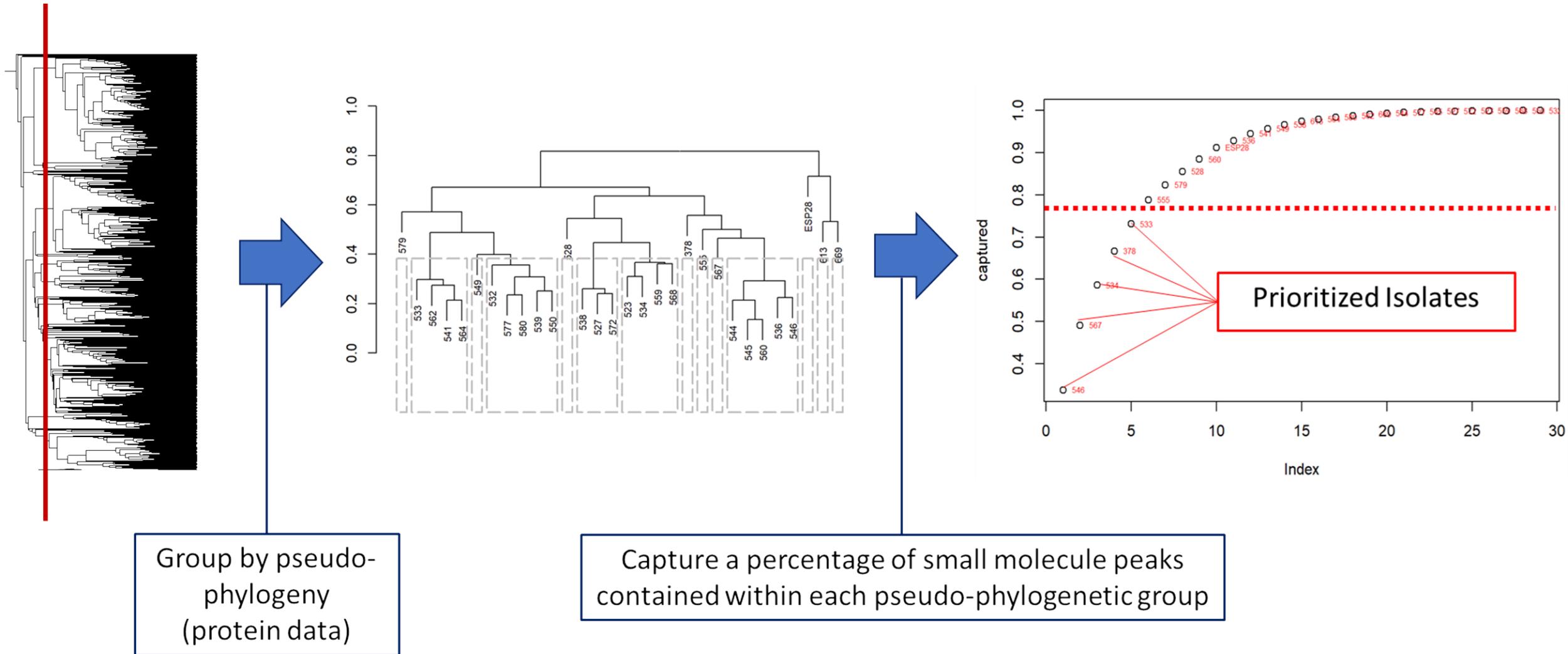
This slide from earlier looked pretty simple.



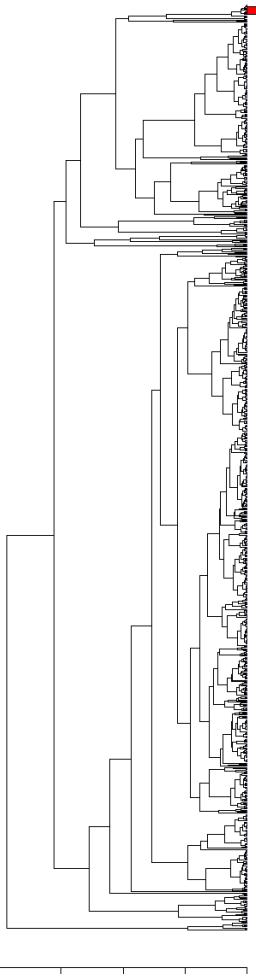
But now we have 90x more isolates!!!!



# Automating the Creation of a Small, Highly-Diverse Libraries

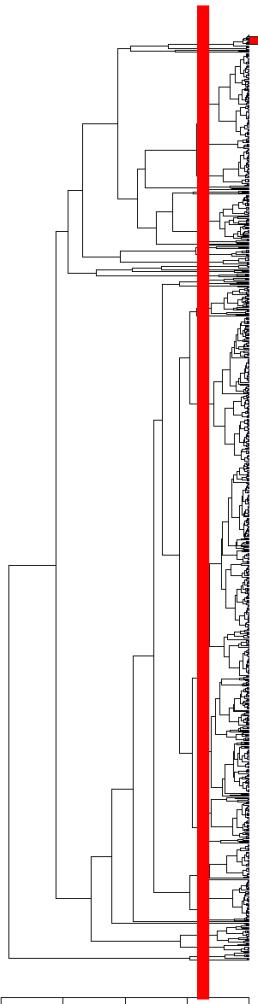


# Automating the Creation of a Small, Highly-Diverse Libraries



900 bacterial isolates grouped by  
their protein fingerprints

# Automating the Creation of a Small, Highly-Diverse Libraries

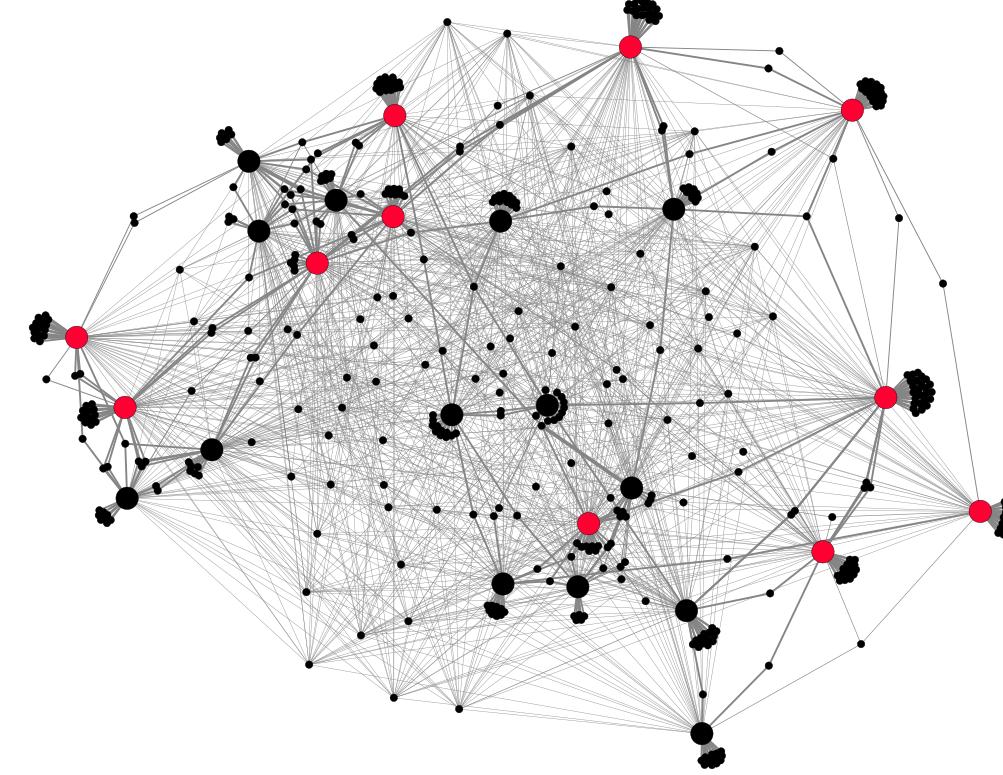


900 bacterial isolates grouped by  
their protein fingerprints

# Automating the Creation of a Small, Highly-Diverse Libraries

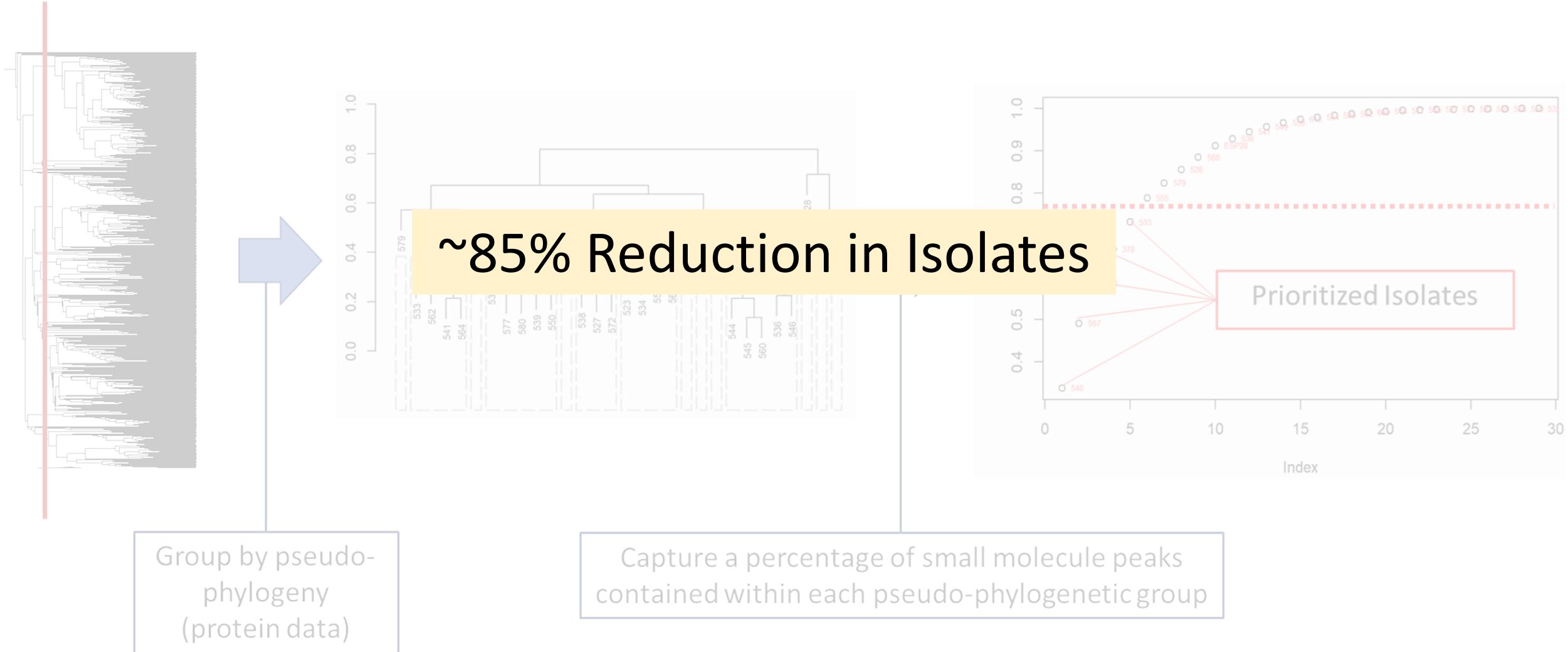


900 bacterial isolates grouped by  
their protein fingerprints



Small molecule data from each group  
is used to choose isolates

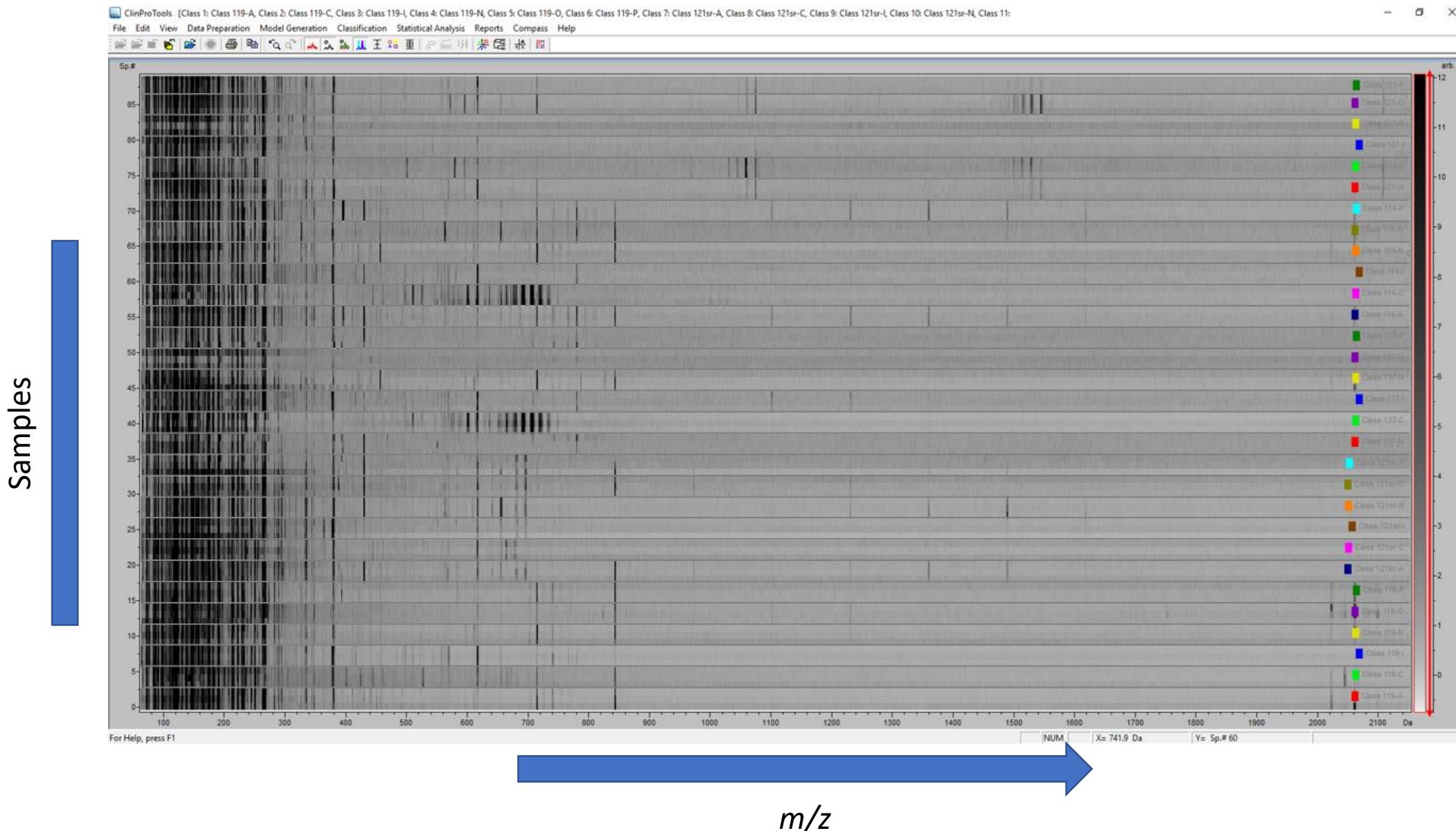
# Automating the Creation of a Small, Highly-Diverse Libraries

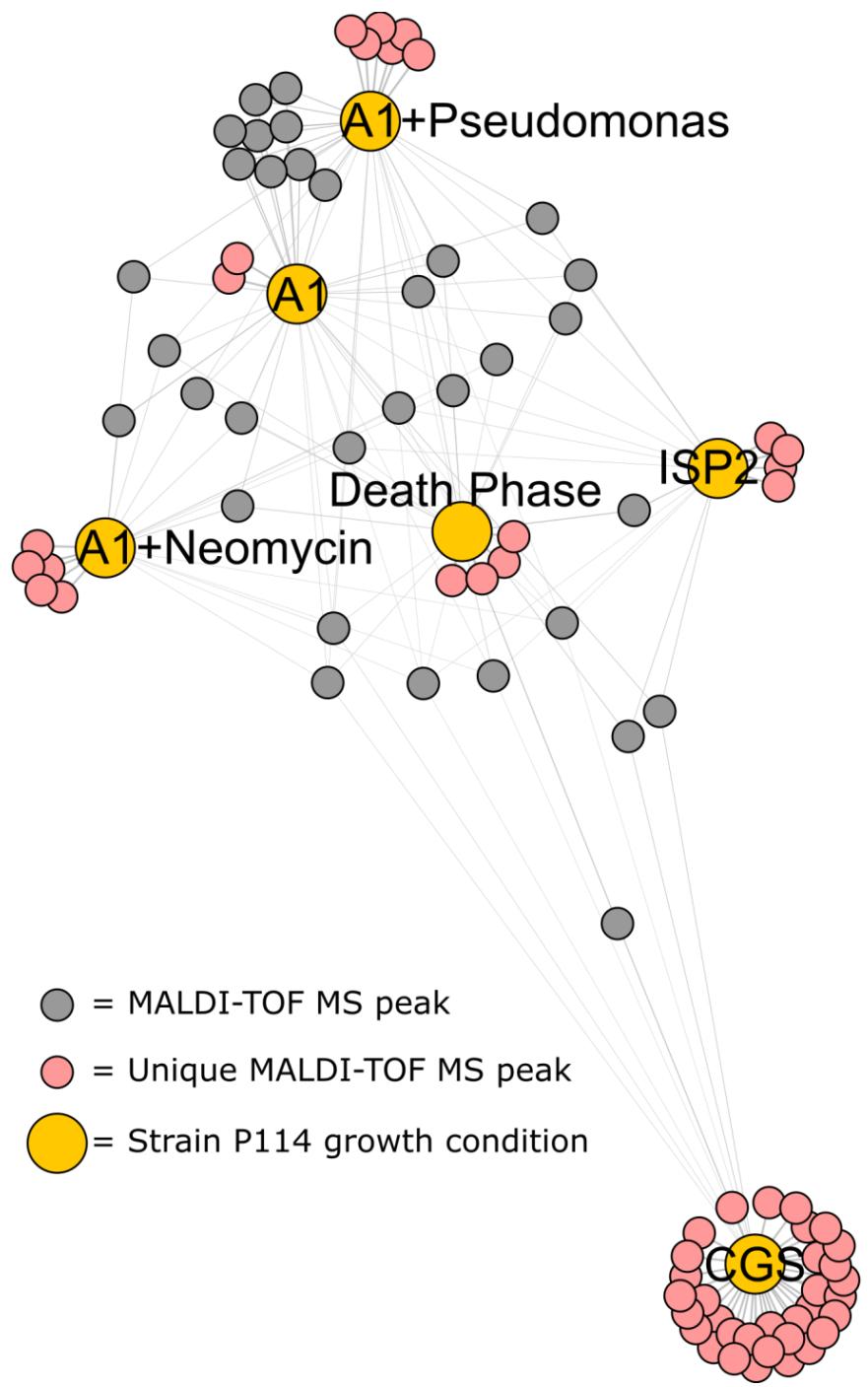


# Using IDBac For More Than Prioritization

# Four Strains Grown in 6 Different Conditions

## MALDI-TOF MS Small Molecule Data



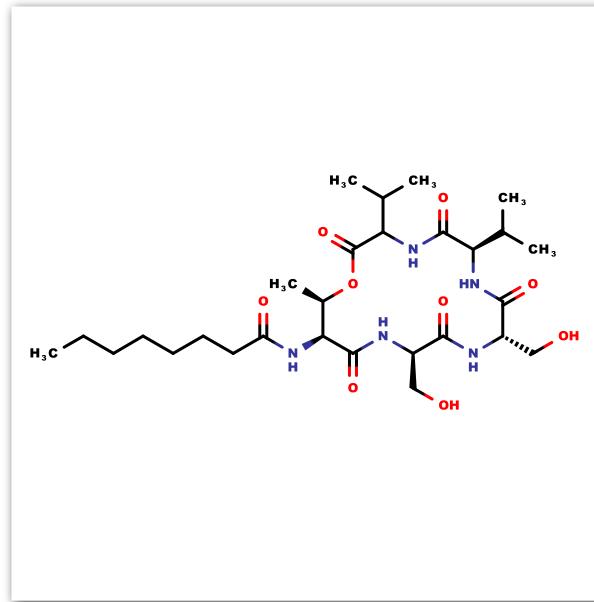
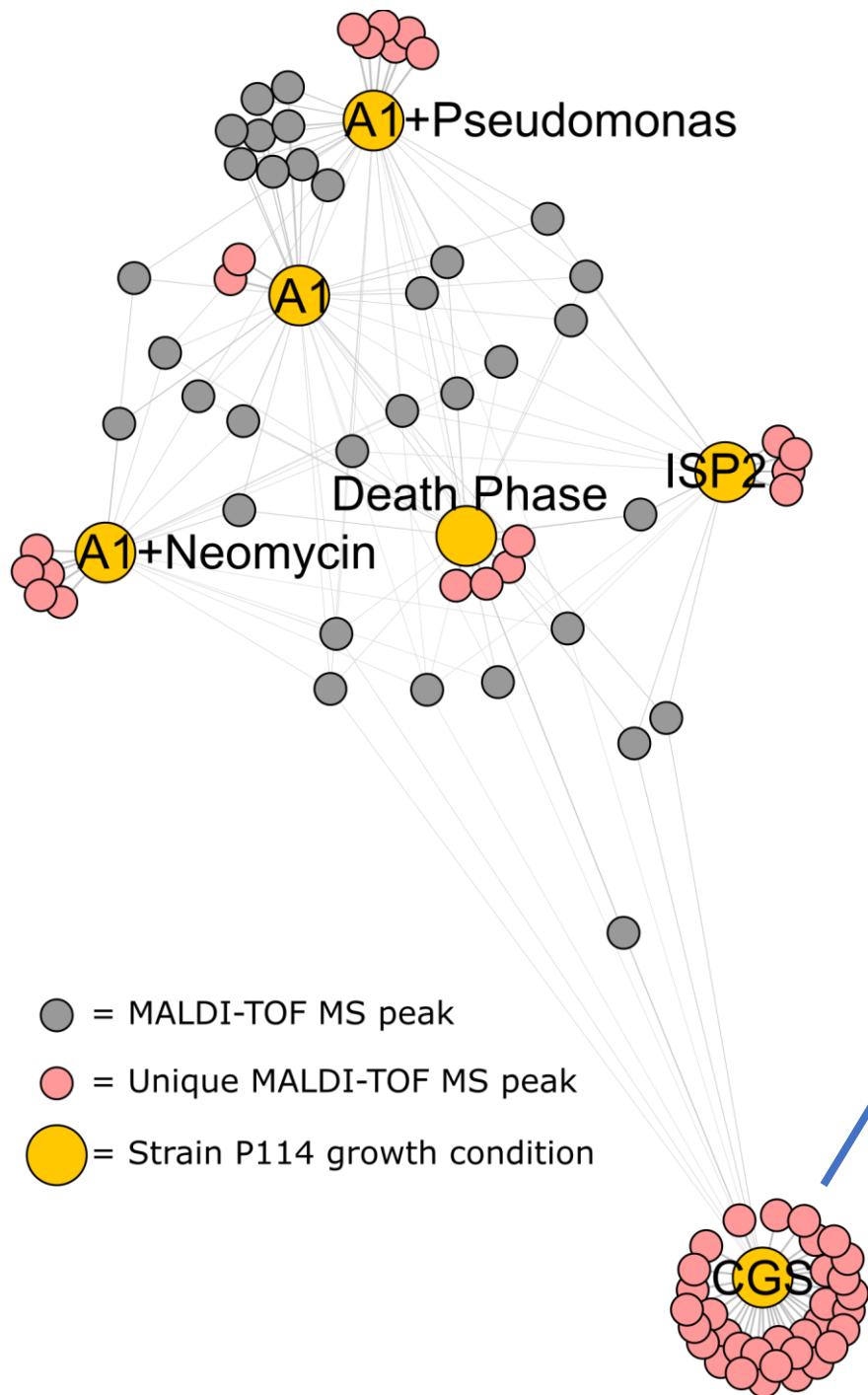


One isolate showed significant differences in MALDI-TOF MS small molecule spectra when grown in CGS + bfe media

CGS + bfe

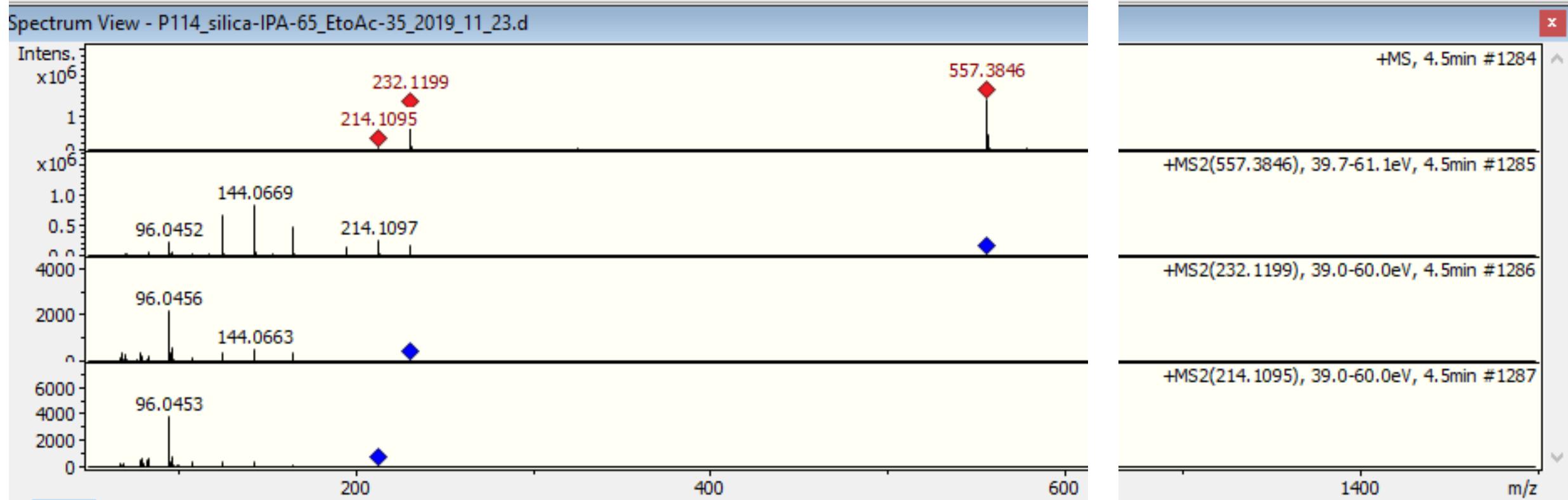
- 2.5 g soytone
- 2 g casamino acids
- 50% glycerol
- 1 L DIH<sub>2</sub>O
- 5 ml 20 mg/mL KBr
- 5 ml 8 mg/mL FeSO<sub>4</sub>



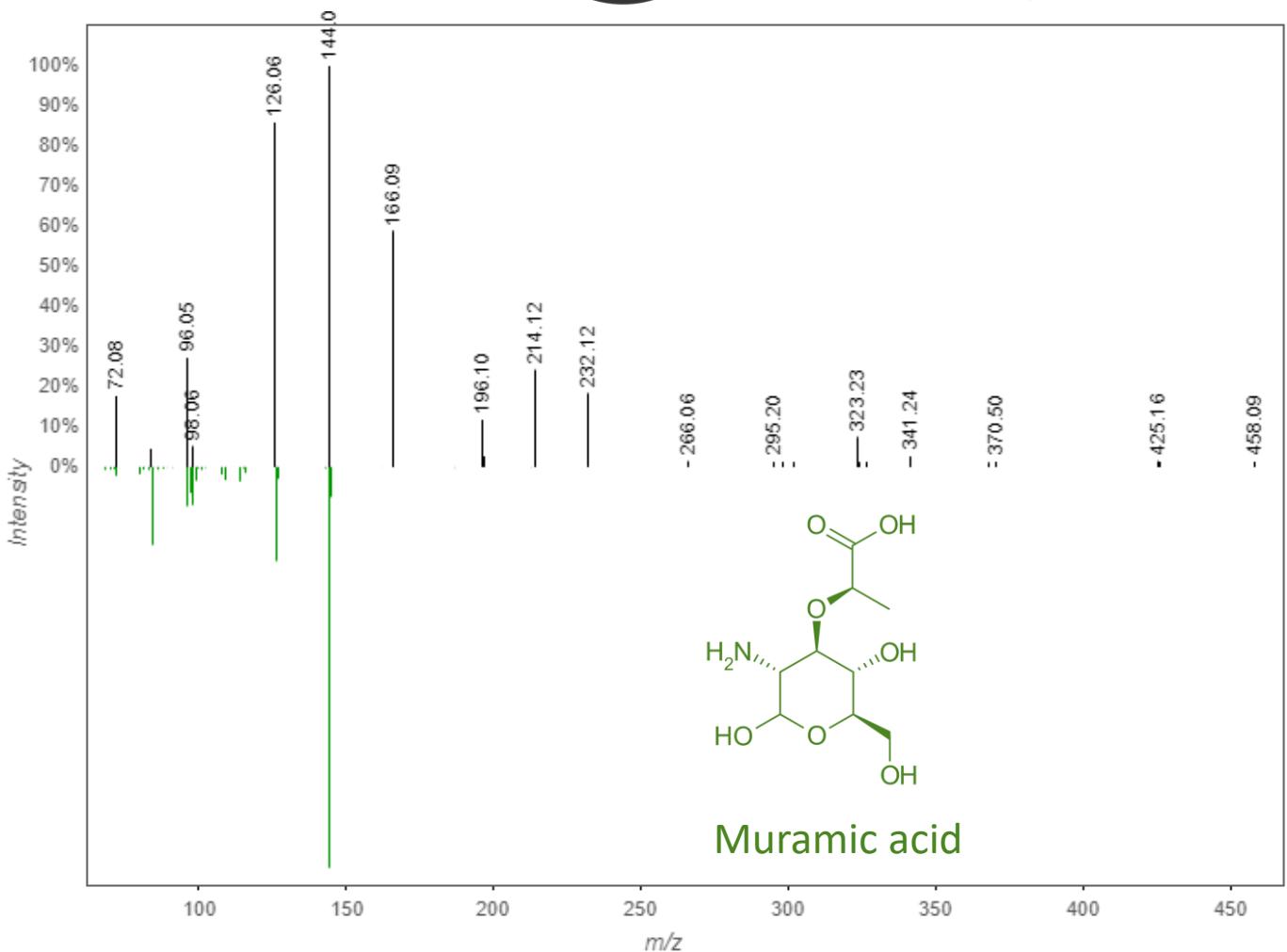
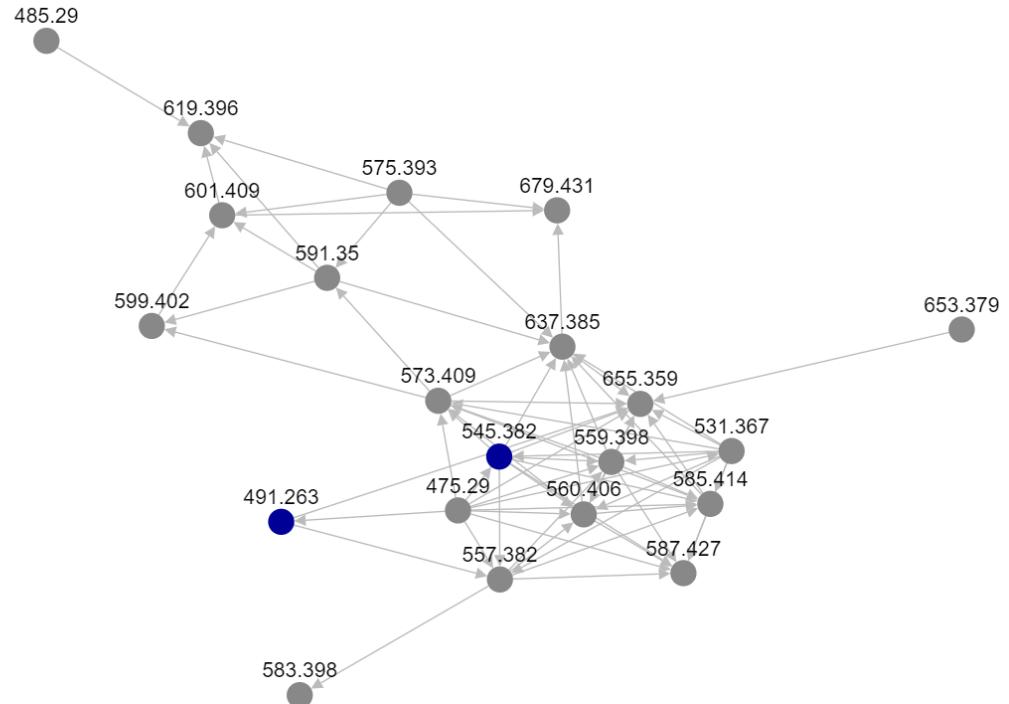


NPA ID	NPA024448
CLUSTER ID	160
NODE ID	62
NAME(S)	Stephensiolide A
FORMULA	C28H49N5O9
MOLECULAR WEIGHT (DA)	599.73
ACCURATE MASS (DA)	599.3530
ORIGIN ORGANISM TYPE	Bacterium
ORIGIN GENUS	Serratia
ORIGIN SPECIES	sp.
INCHIKEY	MSYDJMRXBLIWIS-WCTRMTBBSA-N
EXPORT OPTIONS	<a href="#">TSV</a> <a href="#">MOL</a> <a href="#">PNG</a>
PROJECT MOLECULE TO GLOBAL VIEW	<a href="#">Global View</a>

# A Second Set of Precursors Were Falling Apart in MS1

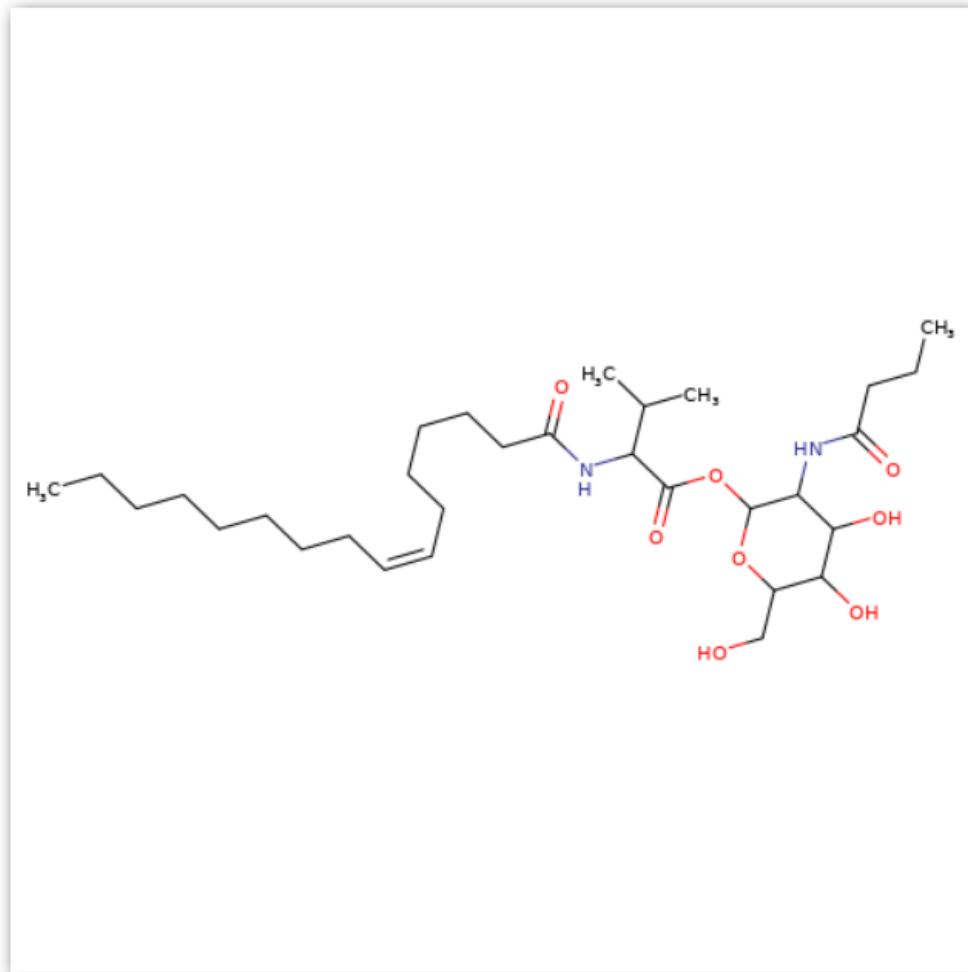


# GNPS Search Partially Matched Muramic Acid



GNPS

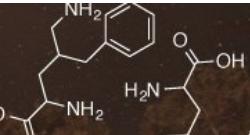
## COMPOUND STRUCTURE



EXPORT OPTIONS	PNG	MOL	TSV
PROJECT MOLECULE TO GLOBAL VIEW	See Global		

## COMPOUND PROPERTIES

NPA ID	NPA002917
CLUSTER ID	1797
NODE ID	1449
NAME(S)	N-Butylglucosamine ester derivative A
FORMULA	C31H56N2O8
MOLECULAR WEIGHT (DA)	584.79
ACCURATE MASS (DA)	584.4037
ORIGIN ORGANISM TYPE	Bacterium
ORIGIN GENUS	Serratia
ORIGIN SPECIES	sp.
INCHIKEY	SMEVICNLKWNVJT-YPKPFQOOSA-N
INCHI	InChI=1S/C31H56N2O8/c1-5-7-8-9-10-11-12-13-14-15-16-17-18-20-25(36)32-26(22(3)4)30(39)41-31-27(33-24(35)19-6-2)29(38)28(37)23(21-34)40-31/h13-14,22-23,26-29,31,34,37-38H,5-12,15-21H2,1-4H3,(H,32,36)(H,33,35)/b14-13-
SMILES	CCCCCCCC/C=C\CCCCC(=O)NC(C(C)C)C(=O)OC1C(C(C(C(O)CO)O)O)NC(=O)CCC



[Back to main page](#)

[Back to status page](#)

[Collapse all](#)

[Download](#)

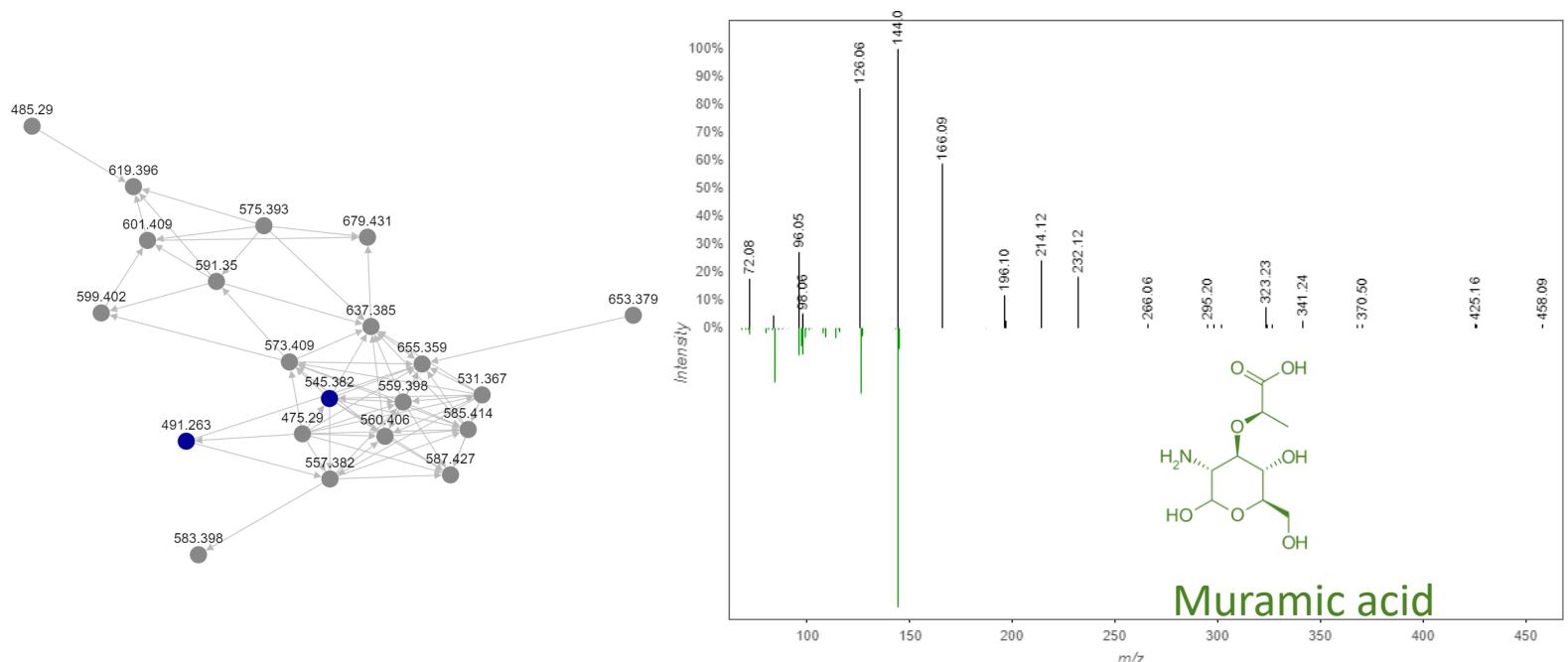
545 Hits 1 ~ 4 out of 4 Go

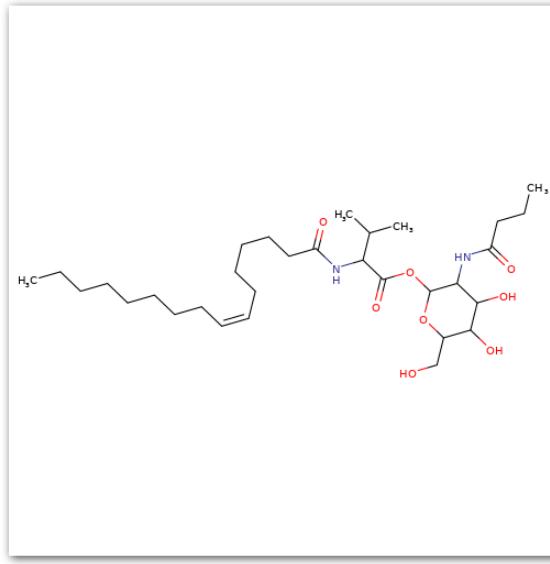
Select columns

Filter	View Dataset	Title	Description	Organisms	Cosine Score	Matched Peaks	MZ Delta	NumFiles
View Mirror Match 1	<a href="#">View MSV000084024</a>	GNPS - IIN - Fungus-growing T. septentrionalis Gardens Extracts LC-MS/MS	<a href="#">Show</a>	Trachymyrmex septentrionalis	0.91	10	0.00	3
View Mirror Match 2	<a href="#">View MSV000082294</a>	GNPS - 20180213_NSF_FGA_metadata_AMCR_MassIVE	<a href="#">Show</a>	Trachymyrmex septentrionalis	0.91	10	0.00	3
View Mirror Match 3	<a href="#">View MSV000082295</a>	GNPS - Fungus-growing T. septentrionalis Gardens Extracts LC-MS/MS	<a href="#">Show</a>	Trachymyrmex septentrionalis	0.91	10	0.00	3
View Mirror Match 4	<a href="#">View None</a>	None	20180213_NSF_FGA_metadata_AMCR	None	None	None	None	None

Only 1 hit to a fungal garden dataset, when searched against all public GNPS datasets

Wang, M., Jarmusch, A.K., Vargas, F. et al. Mass spectrometry searches using MASST. *Nat Biotechnol* **38**, 23–26 (2020).  
<https://doi.org/10.1038/s41587-019-0375-9>





NPA ID	NPA002917
CLUSTER ID	1785
NODE ID	1446
NAME(S)	N-Butylglucosamine ester derivative A
FORMULA	C <sub>31</sub> H <sub>56</sub> N <sub>2</sub> O <sub>8</sub>
MOLECULAR WEIGHT (DA)	584.79
ACCURATE MASS (DA)	584.4037
ORIGIN ORGANISM TYPE	Bacterium
ORIGIN GENUS	Serratia
ORIGIN SPECIES	sp.
INCHIKEY	SMEVICNLKWNVJT-YPKPFQOOSA-N
EXPORT OPTIONS	<a href="#">TSV</a> <a href="#">MOL</a> <a href="#">PNG</a>
PROJECT MOLECULE TO GLOBAL VIEW	<a href="#">Global View</a>

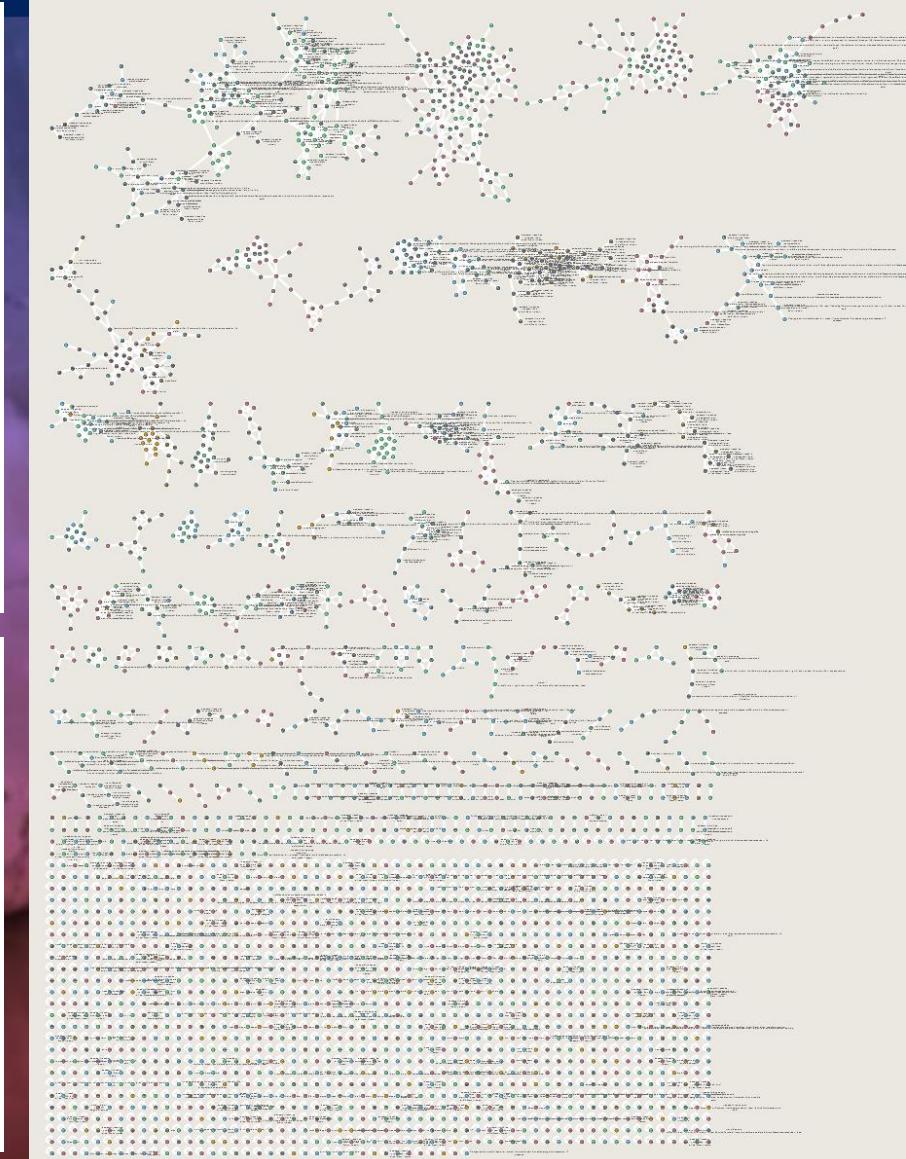
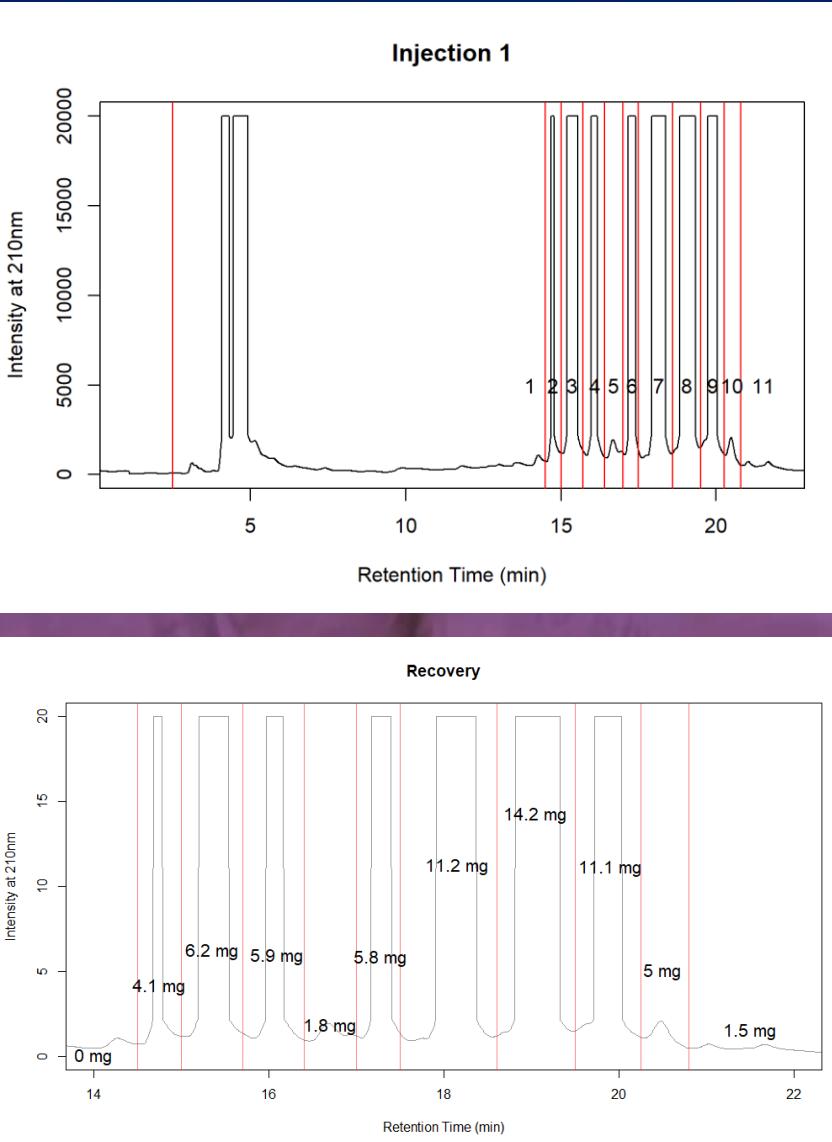
“... possess mild but selective activity against mycobacteria and could be leads for the development of new drugs for use against *M. tuberculosis*.”

Dwivedi D, Jansen R, Molinari G, Nimtz M, Johri BN, Wray V.

Antimycobacterial Serratamolides and Diacyl

Peptoglucosamine Derivatives from *Serratia* sp . *J Nat Prod.*

2008;71(4):637-641. doi:10.1021/np7007126



Institute for Tuberculosis Research

# SCIENCE TODAY

ICELAND -  
THE FUTURE  
OF MEDICINE?

SCIENCE  
MUSEUM

## Funding:

National Center For  
Complementary & Integrative  
Health of the National Institutes  
of Health Award Number:  
F31AT010419



Grant CP-044R-17



## Murphy Lab Current Grad and Post-Doc Members:

Brian Murphy  
Jeongho Lee  
Linh Nguyen  
Tuan Anh Tran

Antonio Hernandez  
Chase Clark  
Maryam Elfeki

Michael Mullowney  
Jhewelle Fitz-Henley



PHARMACEUTICAL  
SCIENCES  
COLLEGE  
OF PHARMACY



Institute for  
Tuberculosis Research



Center for  
Biomolecular Sciences



@ChasingMicrobes

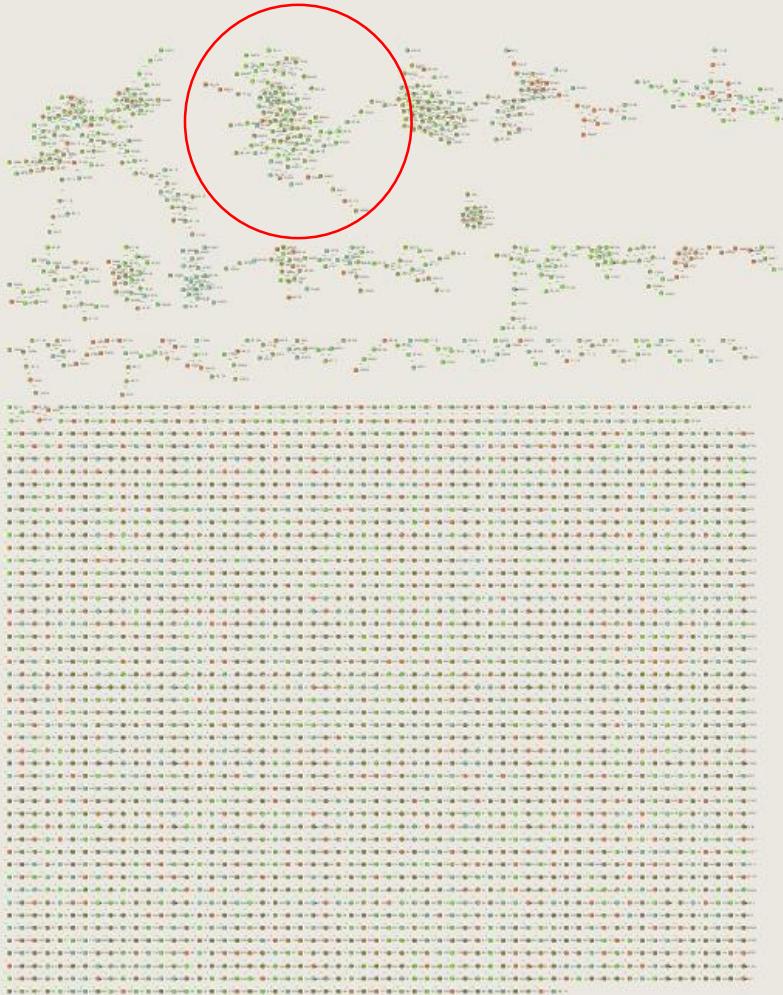
chasemc.github.io

murphylabuic.com

@Murphylabuic

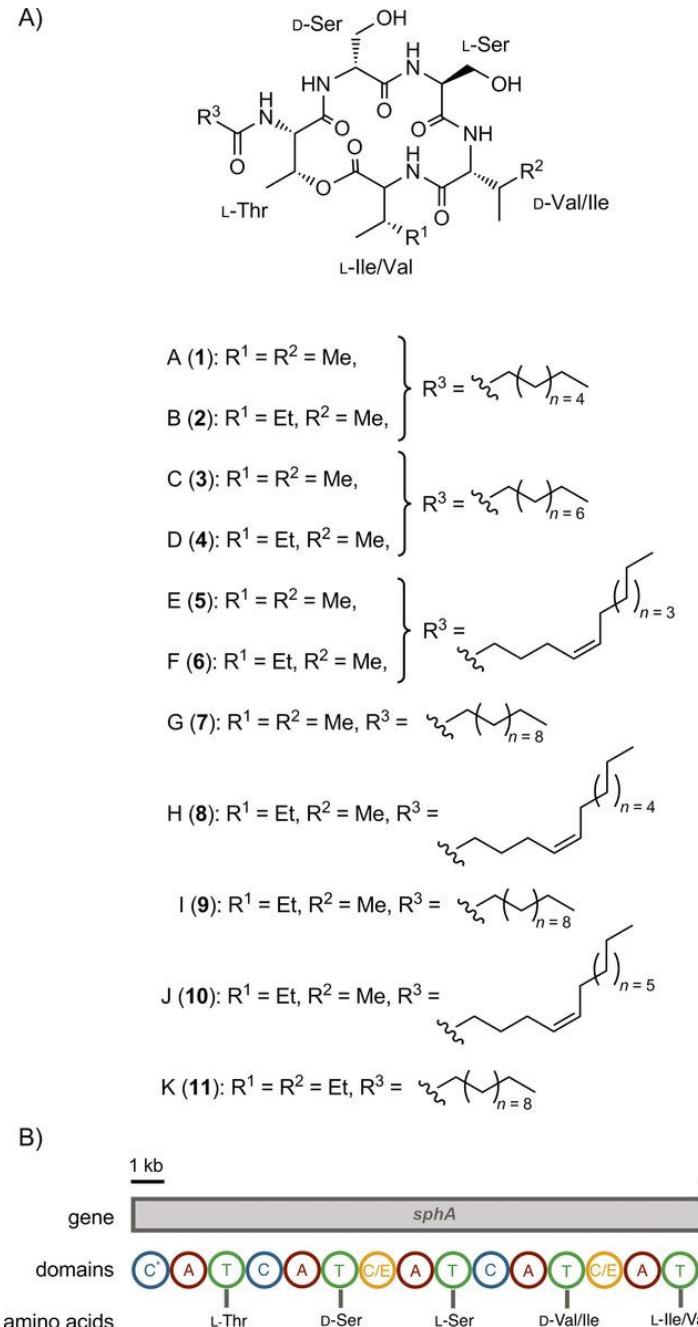
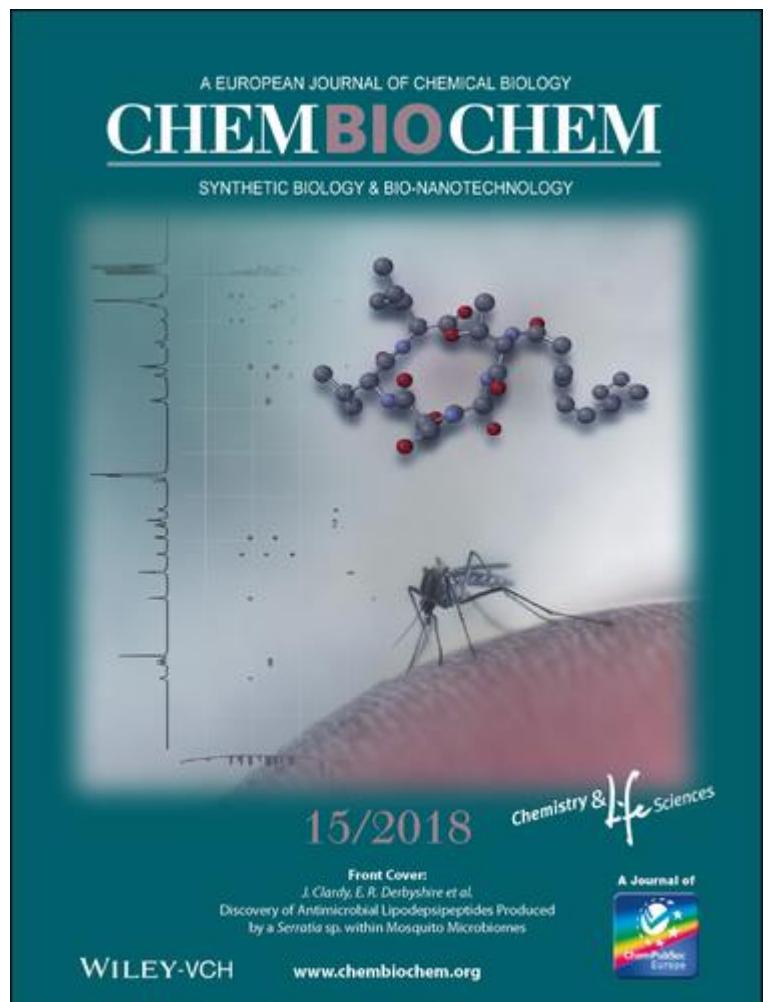


# GNPS Network of 30-L Extract



A molecular network was created using the online workflow at GNPS. The data was 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 +/- 50Da window throughout the spectrum. The data was then clustered with MS-Cluster with a parent mass tolerance of .03 Da and a MS/MS fragment ion tolerance of .03 Da to create consensus spectra . Further, consensus spectra that contained less than 1 spectra were discarded. A network was then created where edges were filtered to have a cosine score above .6 and more than 4 matched peaks. Further edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above .6 and at least 4 matched peaks. Analog search was enabled against the librar with a maximum mass shift of 400 Da.

# Discovery of Antimicrobial Lipodepsipeptides Produced by a *Serratia* sp. within Mosquito Microbiomes



<https://doi.org/10.1002/cbic.201800124>



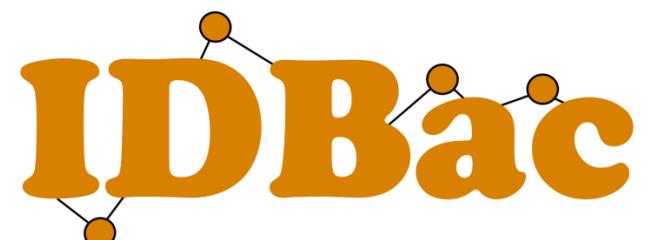
Collect Sponges from the Environment

4

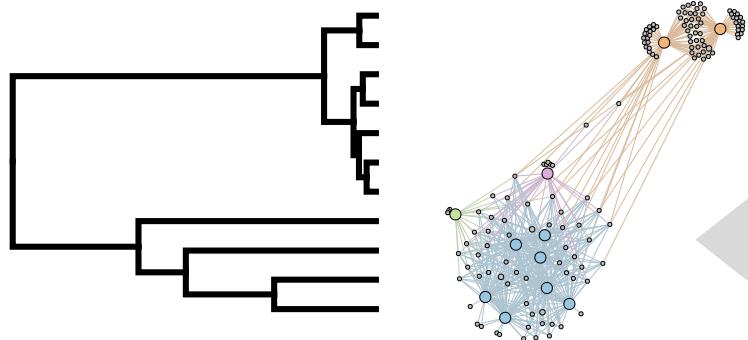


Culture Bacteria

2

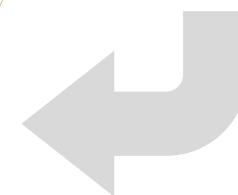


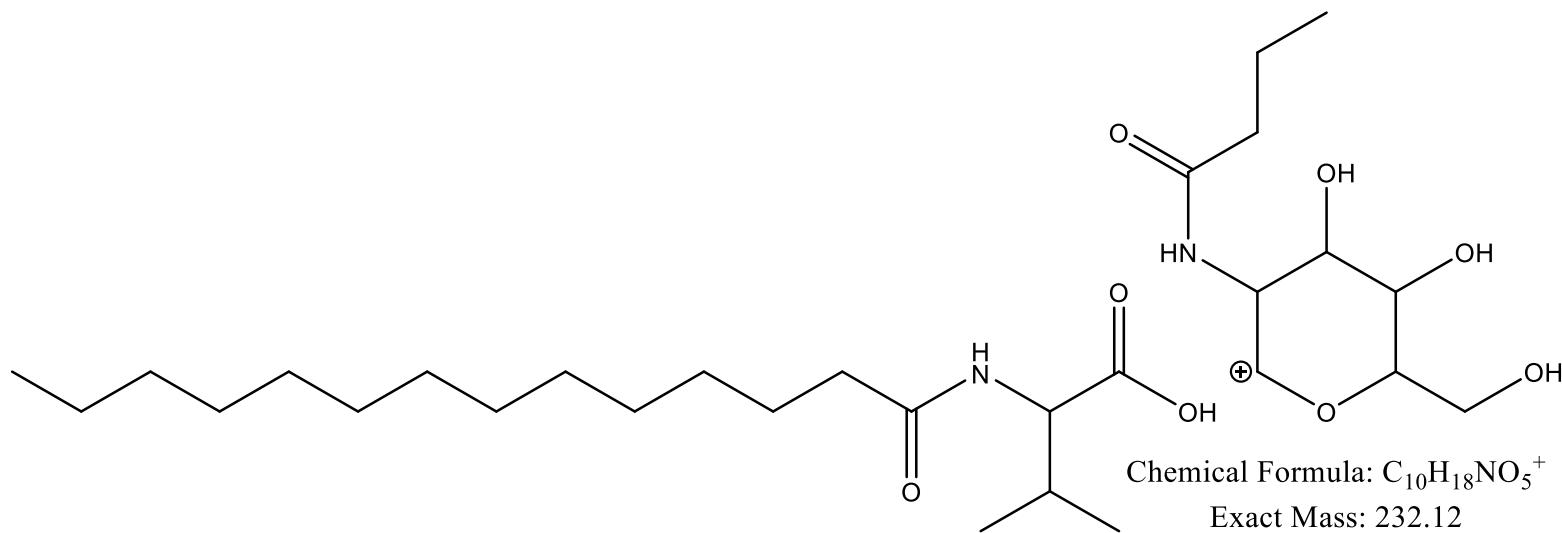
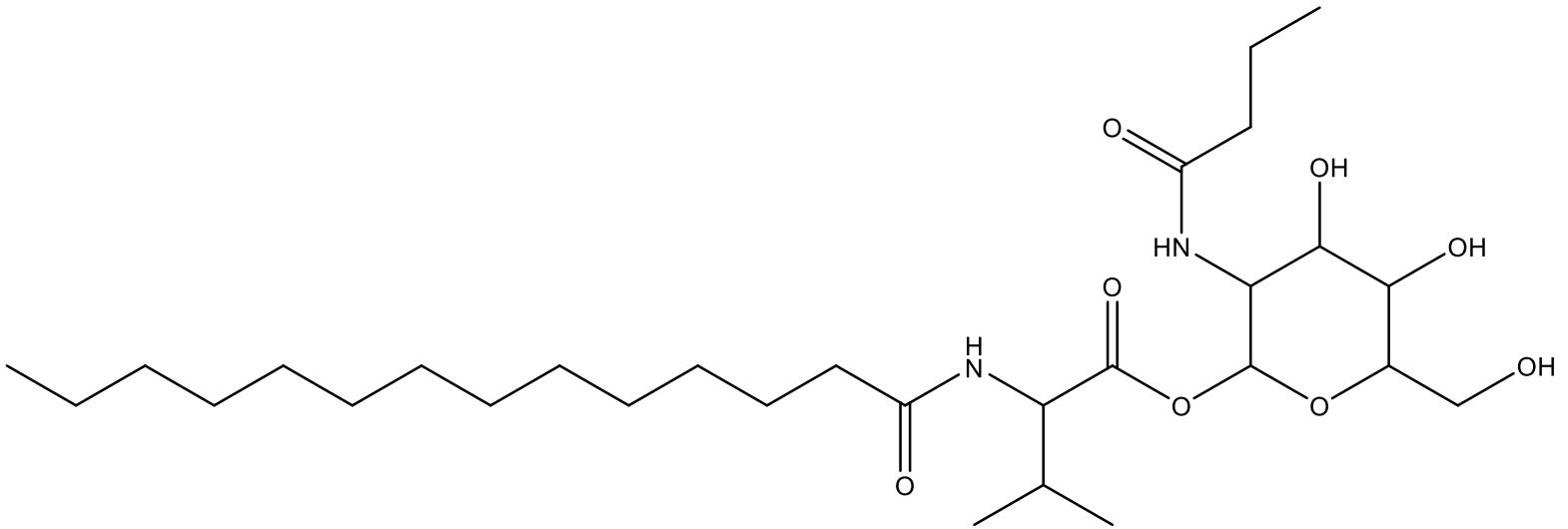
Perform IDBac (MALDI) on Isolates



Protein and Small Molecule Spectra  
Create Diverse Libraries and Inform  
Future Sample Collection

3





Chemical Formula: C<sub>19</sub>H<sub>37</sub>NO<sub>3</sub>  
Exact Mass: 327.28

Chemical Formula: C<sub>10</sub>H<sub>18</sub>NO<sub>5</sub><sup>+</sup>  
Exact Mass: 232.12