

IDBàc

A New Paradigm in
Developing Microbial
Libraries for Drug Discovery



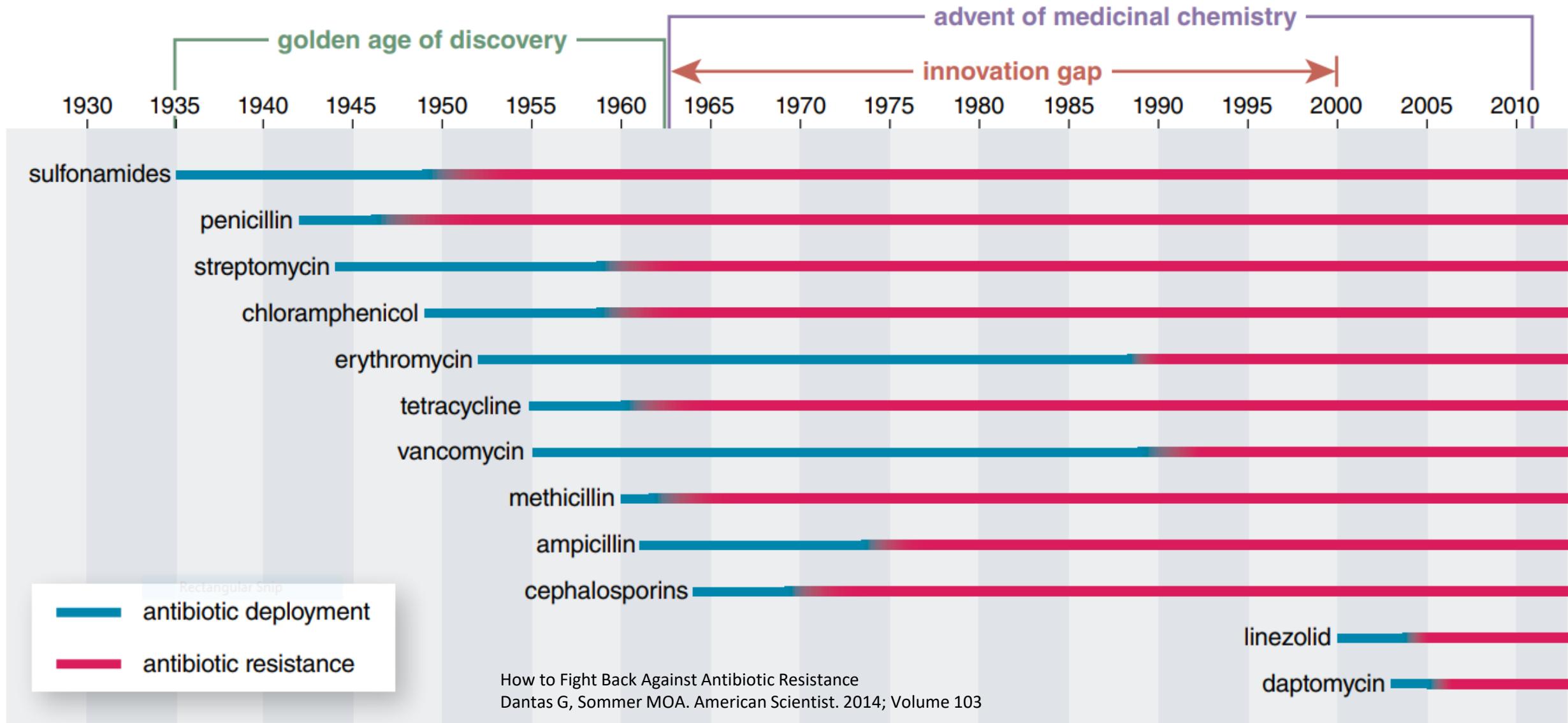
@ChasingMicrobes
chasemc.github.io



Photos of slides OK



Antibiotic Resistance is a Serious Threat



Industry Is Giving Up on Antibiotics



70 HTS campaigns
3 million compounds
19 Hits
5 Leads



65 HTS campaigns
2 million compounds
57 Hits
19 Leads



3 million compounds

Gram-negative
cellular activity

Industry Is Giving Up on Antibiotics



70 HTS campaigns
3 million compounds
19 Hits
5 Leads



65 HTS campaigns
2 million compounds
57 Hits
19 Leads



3 million compounds

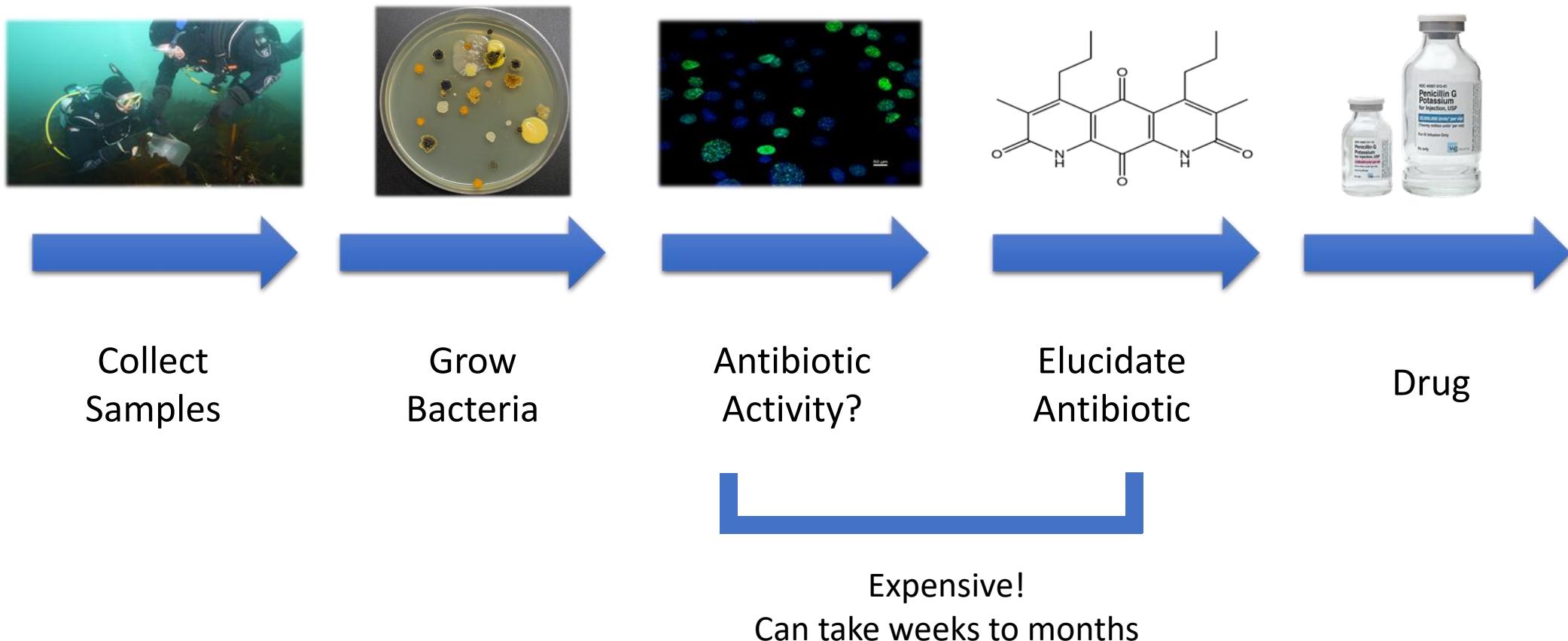
0 Gram-negative
cellular activity



We Scour the Globe for New Bacteria, New Drug Leads



Discovery Pipeline (Simplified)



You need a new antibiotic.

You have resources to
study 5 bacteria.

Which do you choose?

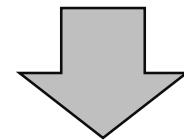


You need a new antibiotic.



You have resources to
study 5 bacteria.

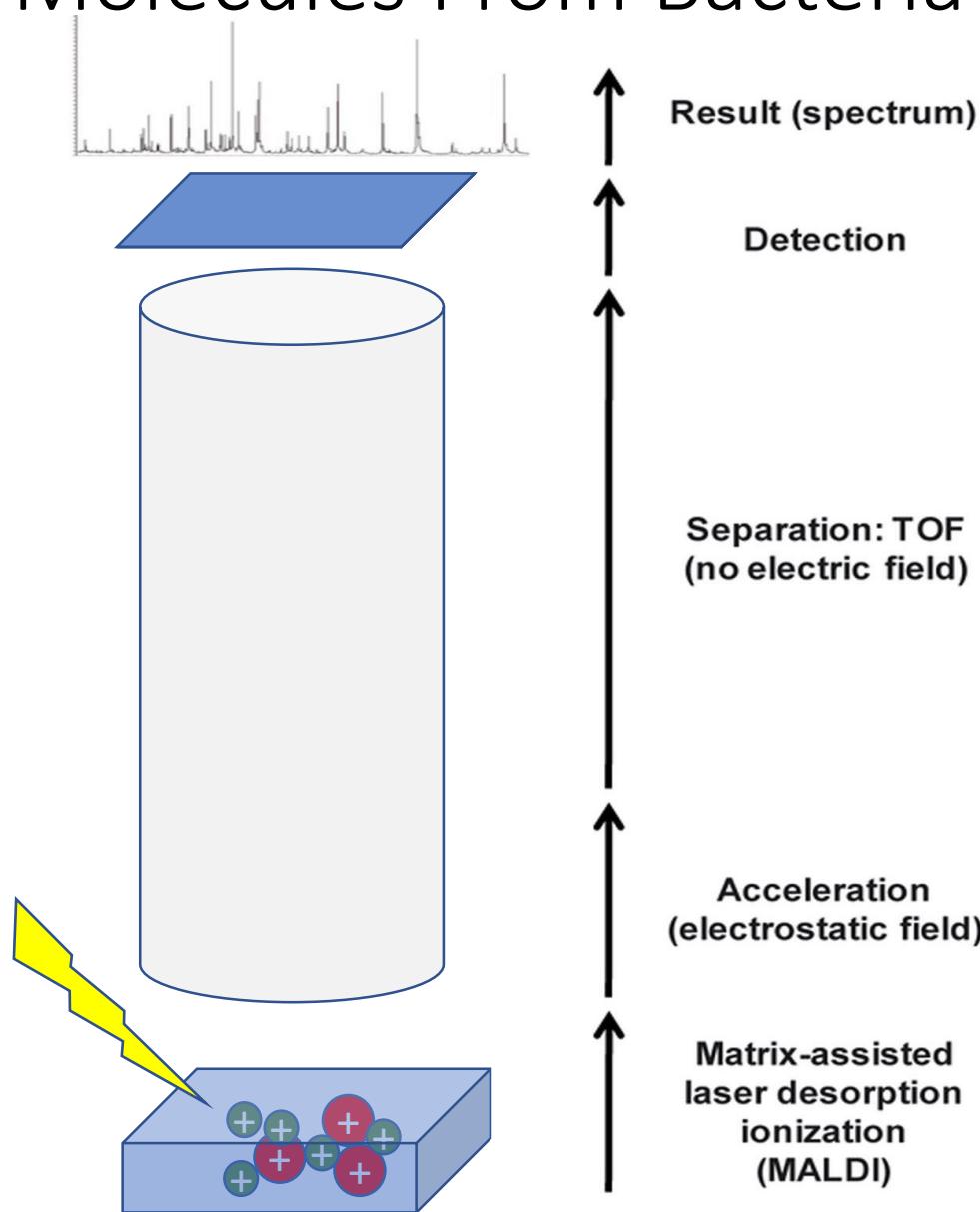
Which do you choose?



MALDI-TOF MS: A Rapid and Simple Way to “Weigh” Molecules From Bacteria



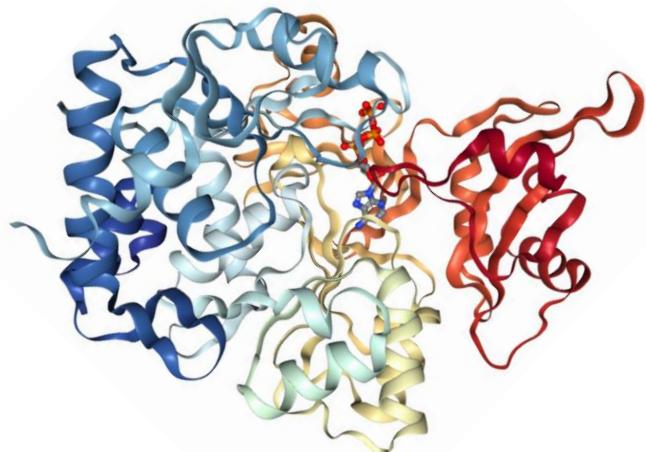
Up to 384 bacteria in one experiment



We Obtain Two Datasets

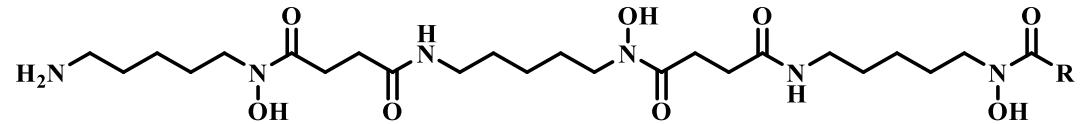
**Protein Spectra
("Phylogenetic Groupings")**

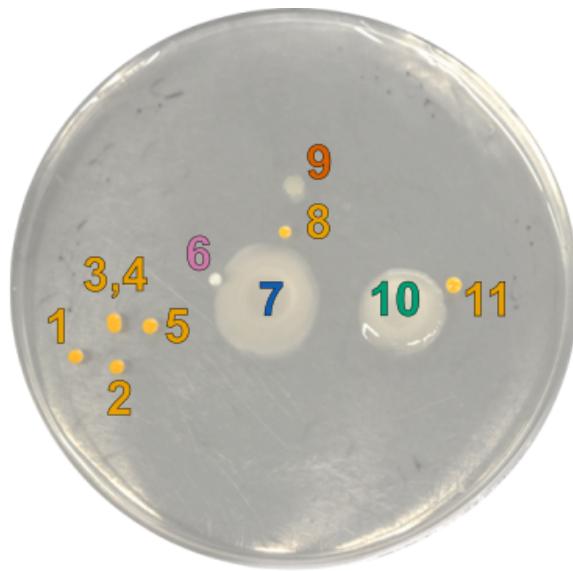
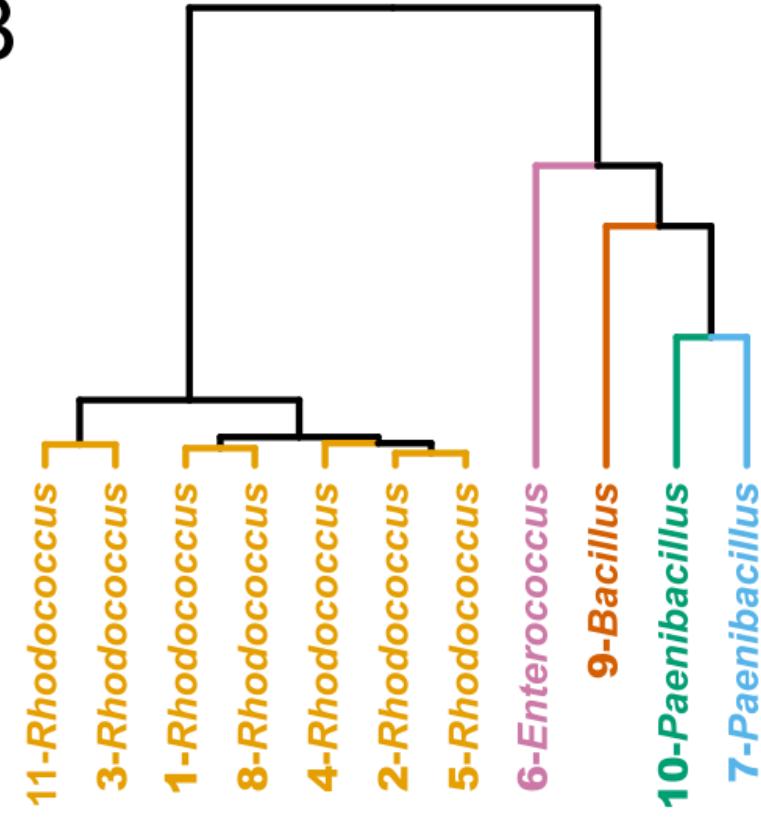
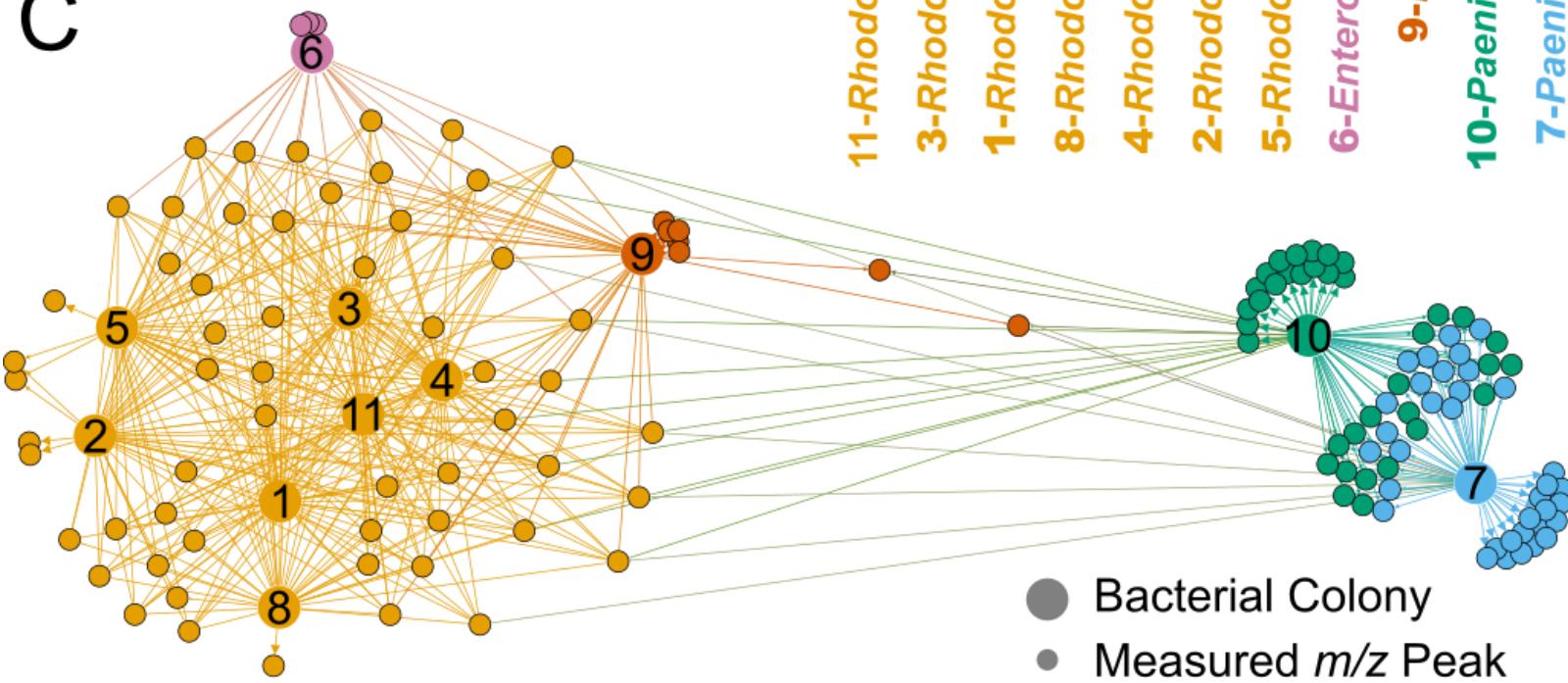
- Large molecules that are evolutionarily stable.



**Small Molecule Spectra
("Potential Drugs")**

- Small molecules that provide a selective advantage.



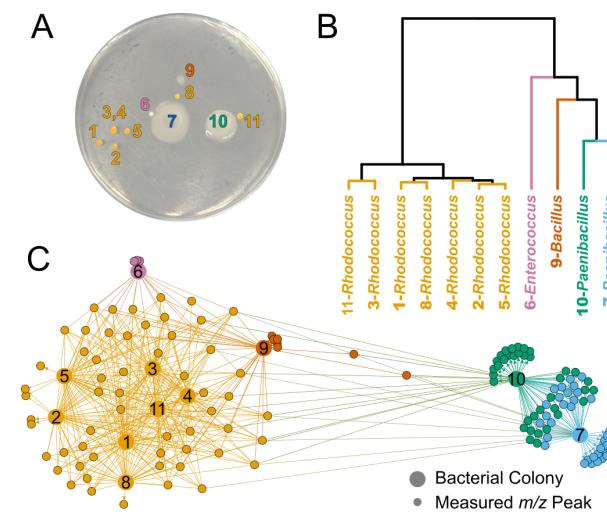
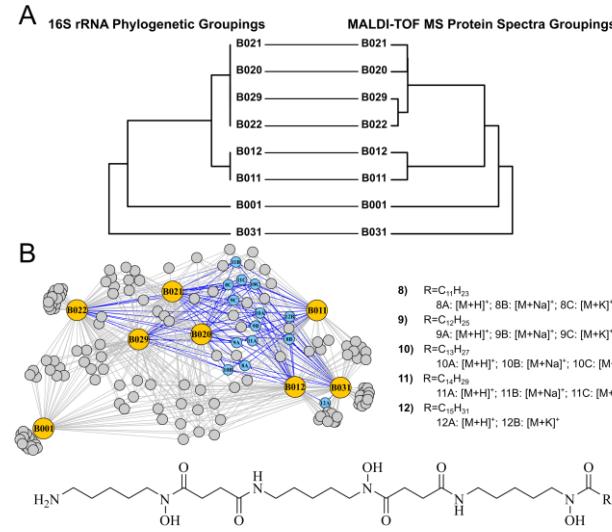
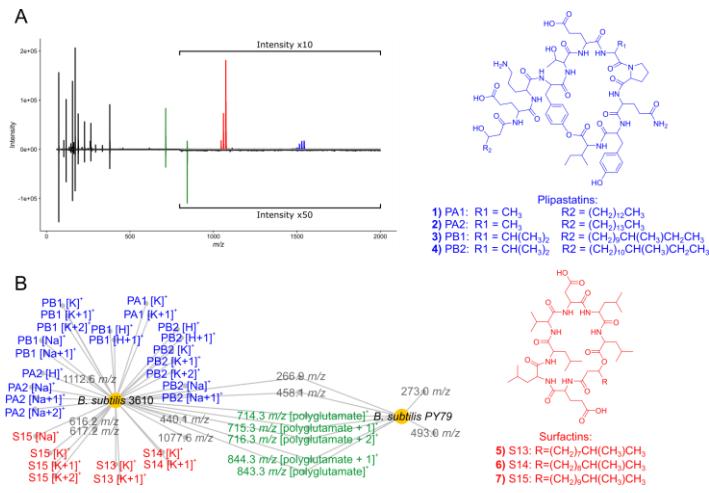
A**B****C**

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

Chase M. Clark^{a,1}, Maria S. Costa^{b,1}, Laura M. Sanchez^{a,2}, and Brian T. Murphy^{a,2,3}

^aDepartment of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL; and ^bFaculty of Pharmaceutical Sciences, University of Iceland, Hagi, IS-107 Reykjavík, Iceland

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

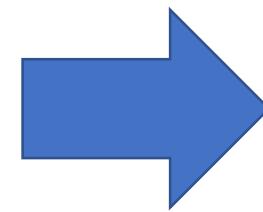
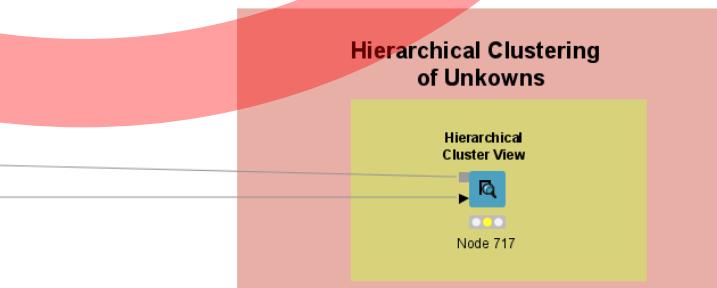
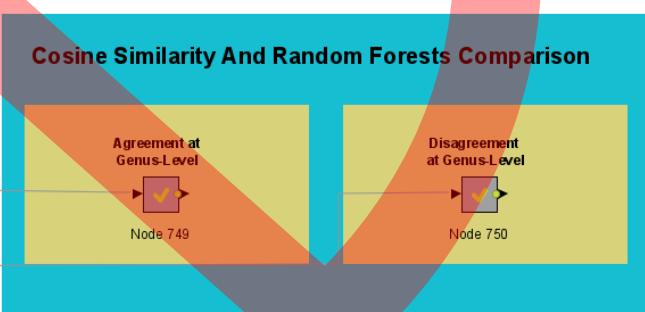
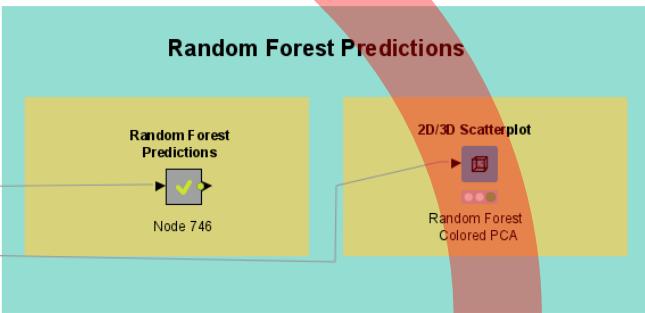
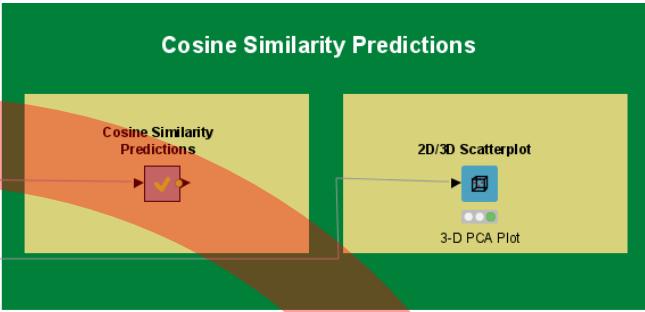
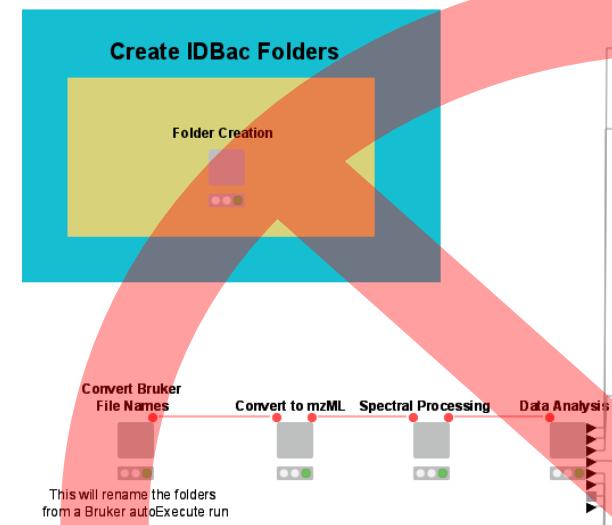


IDBac

A MALDI Protein and Small Molecule Bioinformatics Platform

 GITHUB

 TWITTER



A shiny app as a package:

R packages:
“...the fundamental
unit of shareable code
is the package.”

378 commits	3 branches	17 releases	1 contributor
Branch: AddingDatabase... ▾	New pull request	Create new file	Upload files
This branch is 66 commits ahead, 5 commits behind master.			Pull request Compare
 chasemc fix typo			Latest commit fdaacaa 20 days ago
└ MALDI-Plate_Template	add excel template		10 months ago
└ R	Major rewrite, speed boost		20 days ago
└ ReadMe_Images	add readme images		5 months ago
└ inst	fix typo		20 days ago
└ tests	test for collapsing protein replicates		20 days ago
└ .gitattributes	MajorUpdate		6 months ago
└ .gitignore	gitignore		2 months ago
└ CreateExecutable.R	.		5 months ago
└ DESCRIPTION	add testthat		21 days ago
└ IDBac_App.Rproj	Rename IDBac.Rproj to IDBac_App.Rproj		5 months ago
└ LICENSE	remove pwiz		5 months ago
└ NAMESPACE	MajorUpdate		6 months ago
└ README.md	Add google forum url		3 months ago
└ testing.R	add testthat		21 days ago

A shiny app uses modular code that can be tested

Branch: AddingDatabase... ▾ [IDBacApp / R /](#)

This branch is 66 commits ahead, 5 commits behind master.

 chasemc Major rewrite, speed boost

..

[LibraryCreation.r](#)

[LibrarySearch.r](#)

[LibrarySearchPlots.r](#)

[collapseProteinReplicates.r](#)



Branch: AddingDatabase... ▾ [IDBacApp / tests / testthat /](#)

This branch is 66 commits ahead, 5 commits behind master.

 chasemc test for collapsing protein replicates

..

[test_collapseProteinReplicates.r](#)

[test_spectraProcessingFunction.R](#)

[test_trimProteinSpectra.r](#)

Most people won't "try to get it to work"

Most people won't "try to get it to work" ...a Shiny app is easy to use.



- Install Shiny app like any other Windows program
- Requires no R programming skills
- R is installed if needed
- Logs start up errors
- Does not share data/code with third party or use server software
- Supports continuous installation from GitHub
- Easily uninstalled using the Windows Control Panel

Authors@R: c(
person("Jon", "Hill", email = "jon.mark.hill@gmail.com", role = c("aut", "cre", "cph"))),



Recycle Bin

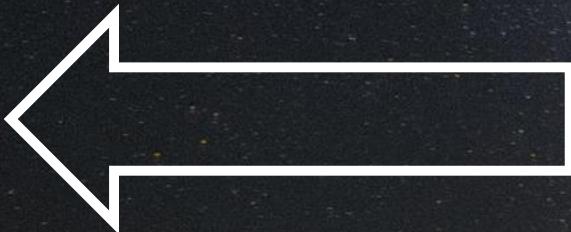
IDBac



IDBac



RinPharma_Chase-C...



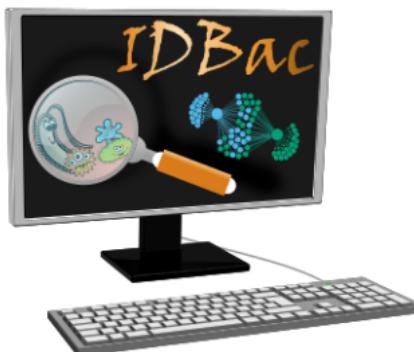
Double-Click to
Start Shiny App



Type here to search



8:54 PM
8/2/2018



Welcome to IDBac

General Information:

- For more information about the method, as well as links to the full code, please visit chasemc.github.io/IDBac
- Bugs and suggestions may be reported on the [IDBac Issues Page on GitHub](#)
- [Check for Updates](#)

How to Cite IDBac:

- Publication:
 - Coupling MALDI-TOF mass spectrometry protein and specialized metabolite analyses to rapidly discriminate bacterial function
Chase M. Clark, Maria S. Costa, Laura M. Sanchez, Brian T. Murphy; bioRxiv 215350; doi: [10.1101/215350](https://doi.org/10.1101/215350)
- Software:
 - For reproducibility, cite this specific version of IDBac as version: [DOI 10.5281/zenodo.1185404](#)

Select "PreProcessing" above to begin

Important to Provide Instruction and Feedback

The screenshot shows the IDBac software interface running in a web browser at 127.0.0.1:3785. A central modal dialog box displays the message "Conversion Complete" with the subtext: "586 files were converted into 100 open data format files. To check what has been converted you can navigate to: C:\Users\chase\Desktop>IDBac-7/Converted_To_mzML". Below the message are two buttons: "Click to continue with Peak Processing" and "Close". To the right of the modal, there are two radio button options: "Select here if you have already converted data and just want to re-analyze it" and "Select here if you have already converted files to mzXML but want to add or subtract files for analyses".

Instructions

1: This directs where on your computer you would like to create an IDBac working directory.
In the folder you select- IDBac will create folders within a main directory named "IDBac":

```
graph TD; Documents[Documents] --> Working_Directory[Working_Directory]; Working_Directory --> IDBac[IDBac]; IDBac --> Converted_To_mzML[Converted_To_mzML]; IDBac --> Peak_Lists[Peak_Lists]; IDBac --> Saved_MANs[Saved_MANs]
```

2: The RAW data will be one folder which itself contains an Excel map and two folders: one containing protein data and another containing small molecule data:

```
graph TD; Documents[Documents] --> MALDI_Plate_1[MALDI_Plate_1]; MALDI_Plate_1 --> ProteinData[ProteinData]
```

Workflow Pane

1: Your Working Directory is where files will be created
Click to select your Working Directory
[1] "C:\\Users\\chase\\Desktop"

2: Your RAW data should be one folder that contains: a folder containing protein data and folder containing small-molecule data
Click to select the location of your RAW data
protein
small_molecule

3: Choose your Sample Map file, the excel sheet which IDBac will use to rename your files.
Browse... Template_data1.xlsx
Upload complete

4: Click "Convert to mzXML" to begin spectra conversion

Spectrum 1 (up)
(Peak matches to bottom spectrum are blue, non-matches are red)

1003

Spectrum 2 (down)

1003

[Download Main Plot](#)

[Download Zoomed Plot](#)

In what percentage of replicates must a peak be present to be kept? (0-100%) (Experiment/Hypothesis dependent)

70

Signal To Noise Cutoff

4

Lower Mass Cutoff

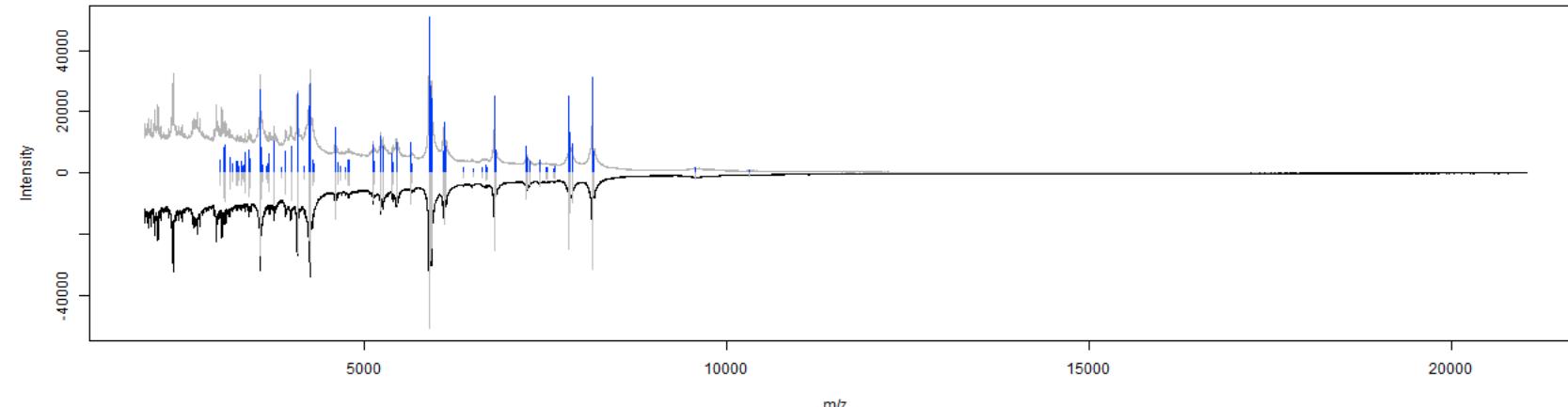
3000

Upper Mass Cutoff

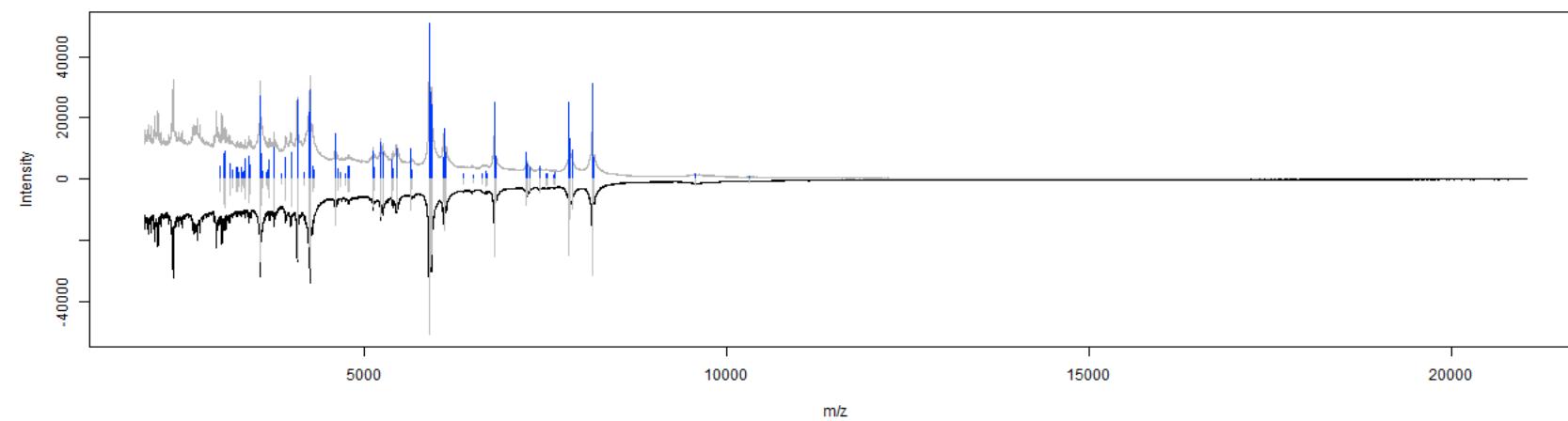
15000

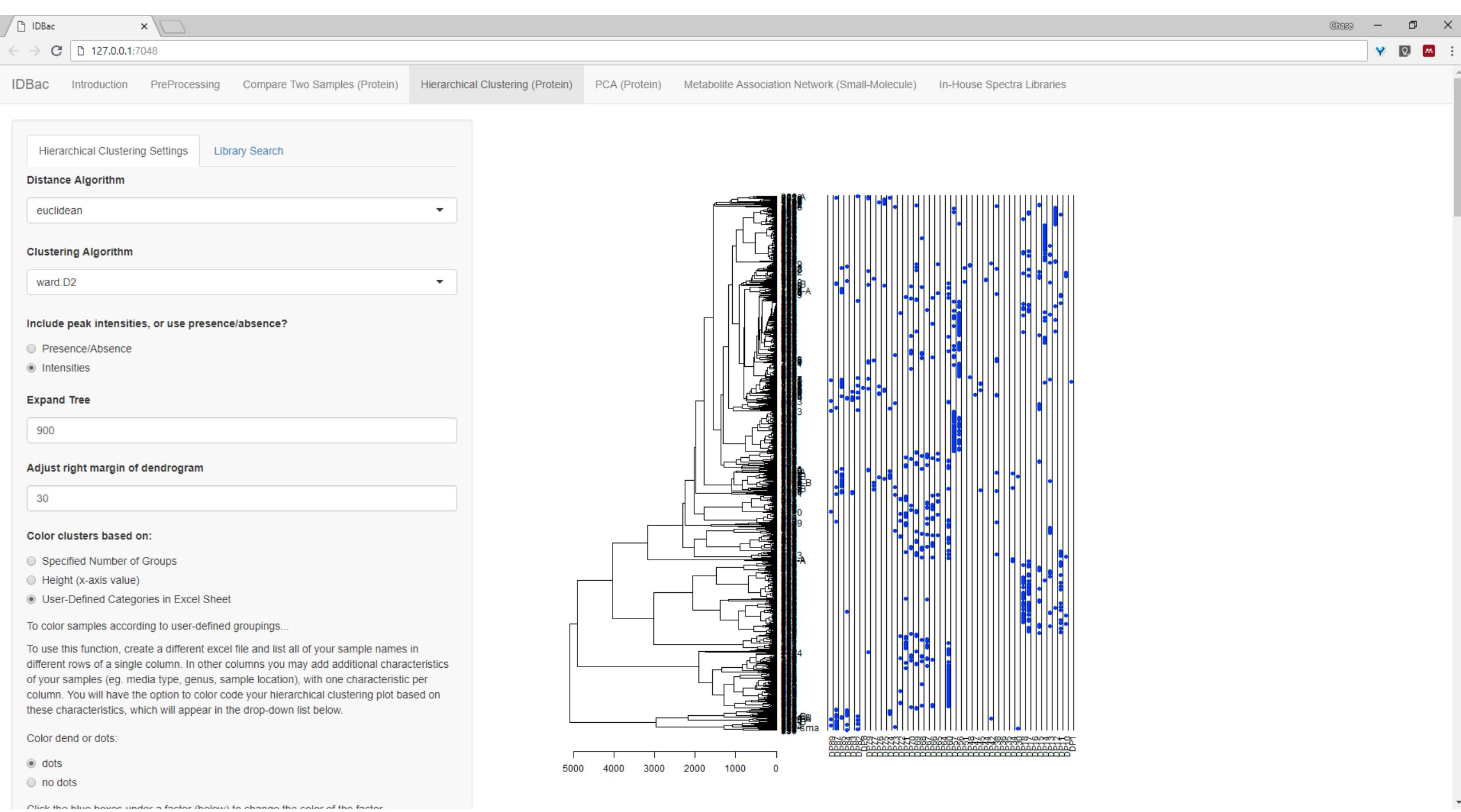
Note: Mass Cutoff and Percent Replicate values selected here will be used in all later analyses.

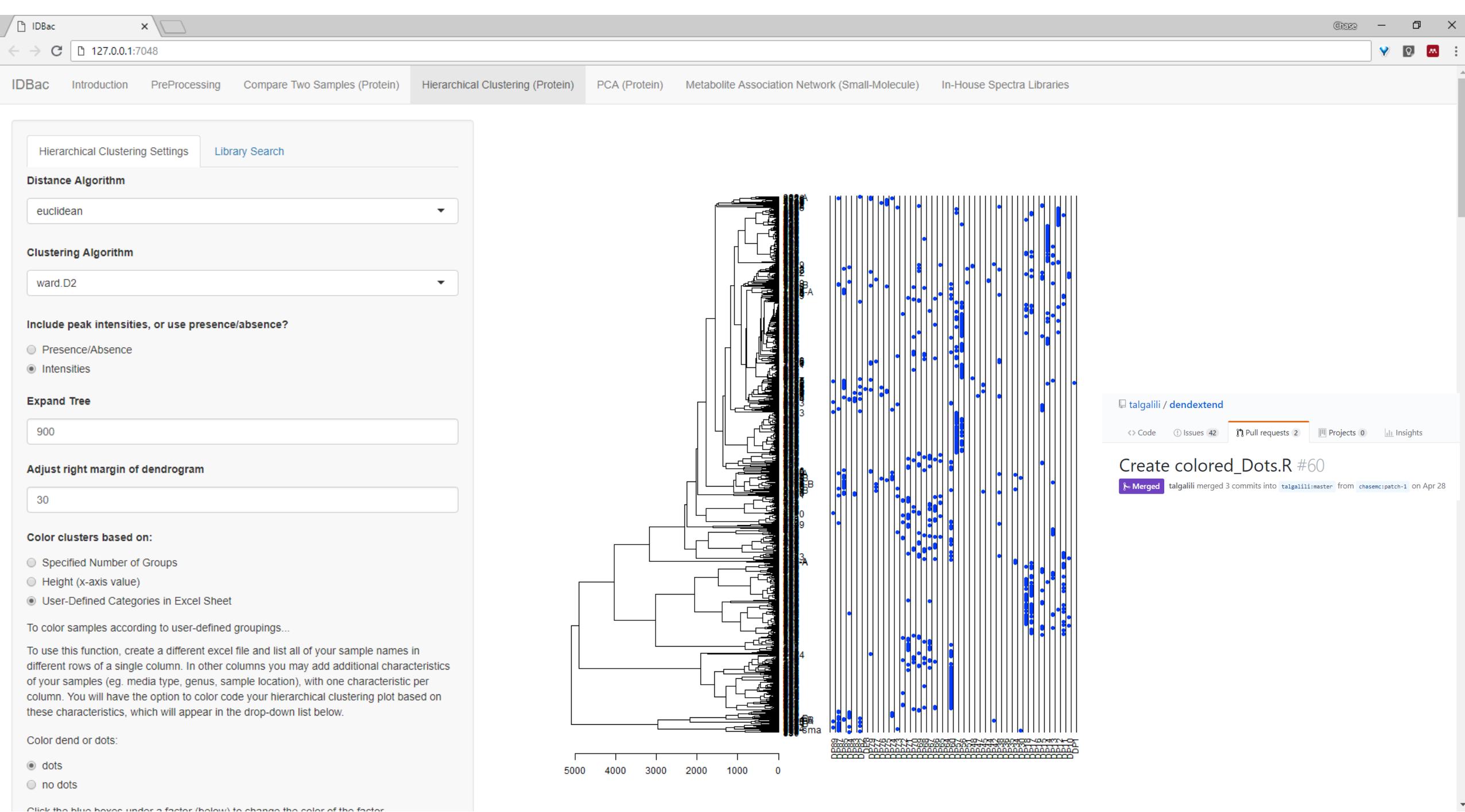
Note 2: Displayed spectra represent the mean spectrum for a sample. Example: if you observe a peak in your mean spectrum but it isn't represented as a red or blue line, then either it doesn't occur often enough across your replicates or its signal to noise ratio is less than what is selected.

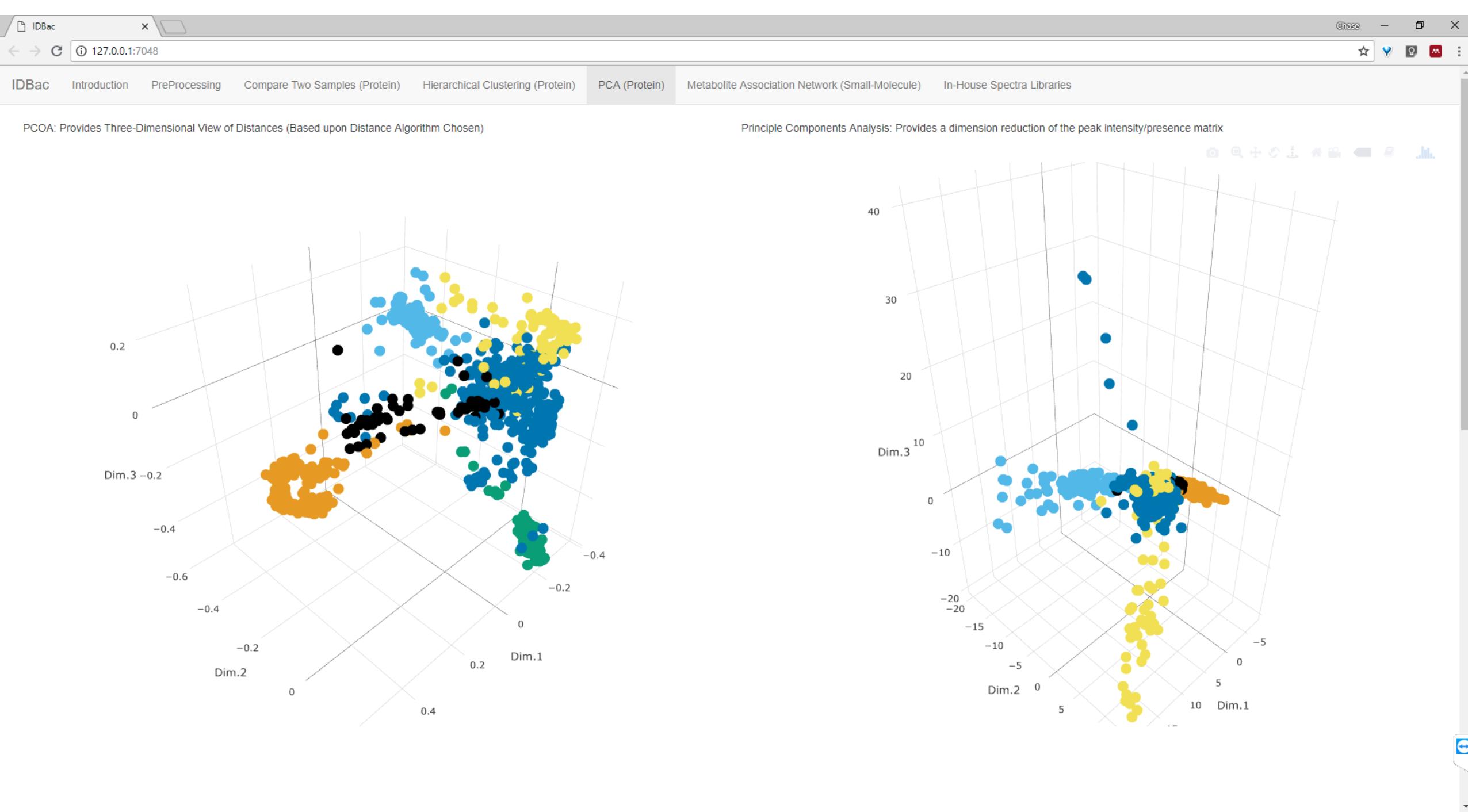


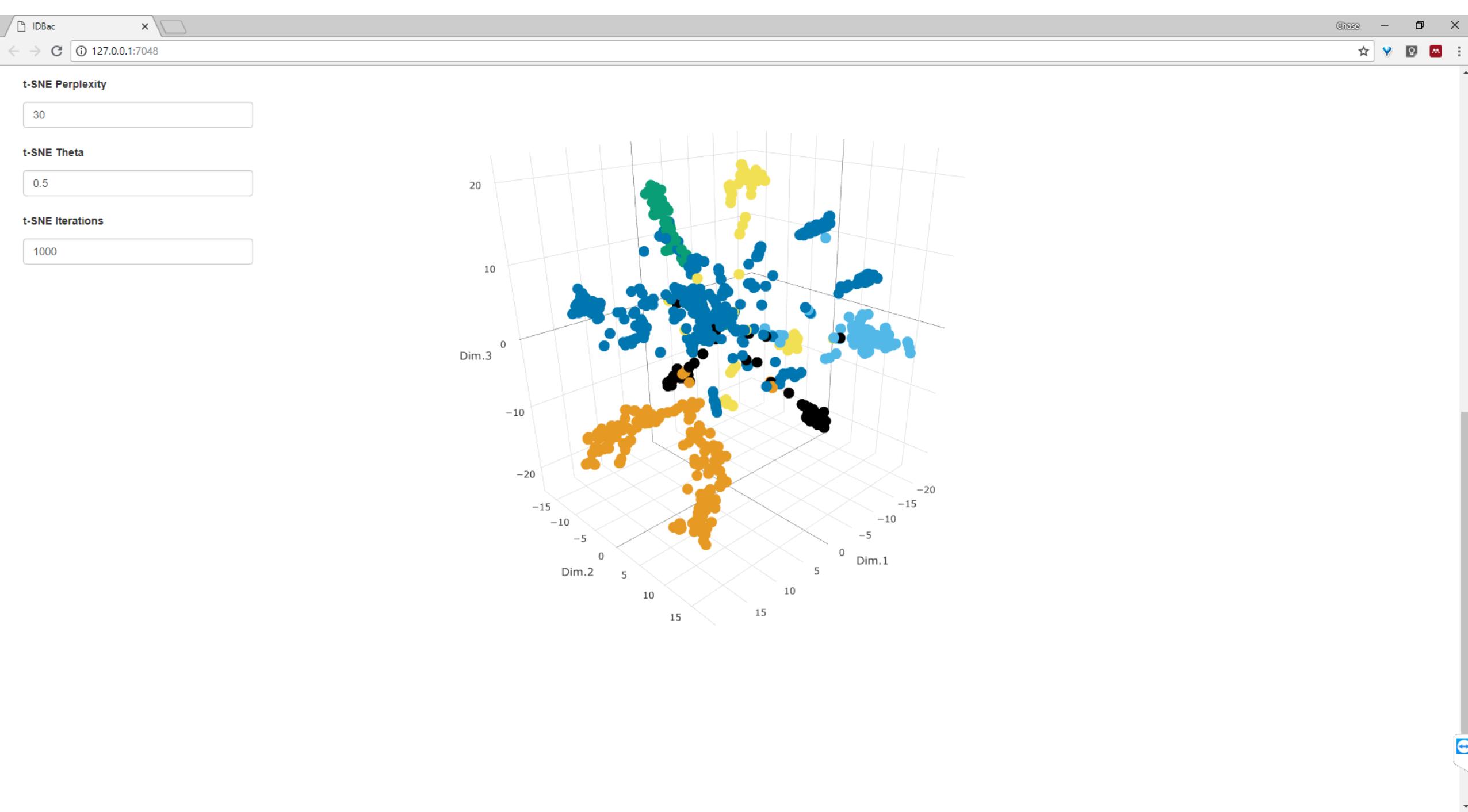
Click and Drag on the plot above to zoom (Will zoom in plot below)







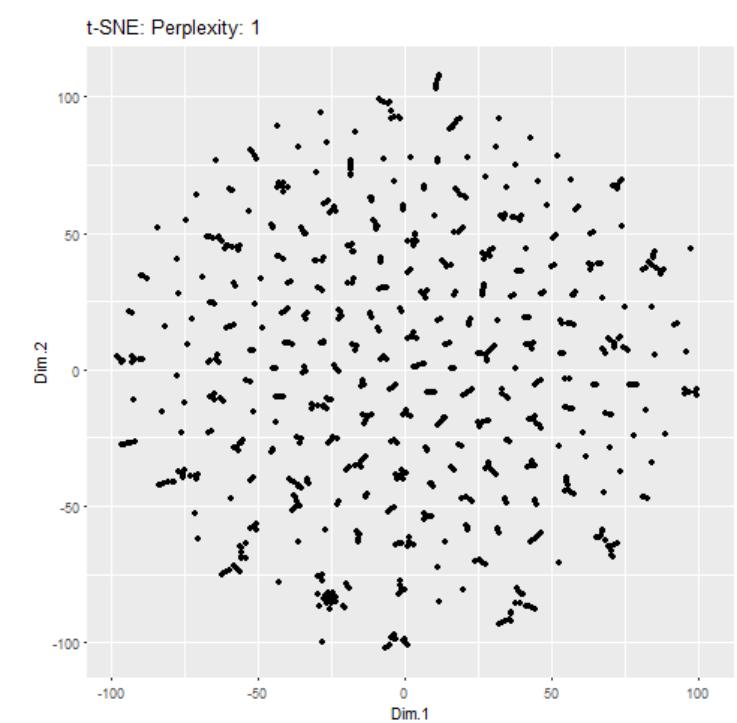
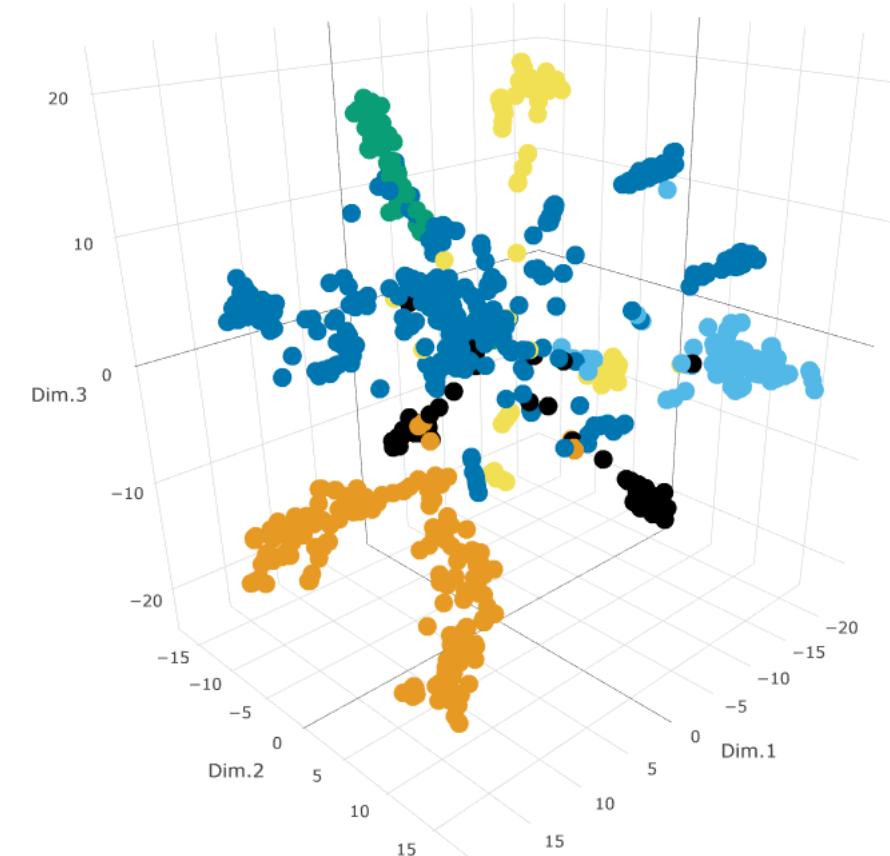




t-SNE Perplexity

t-SNE Theta

t-SNE Iterations



Do you have a matrix blank?

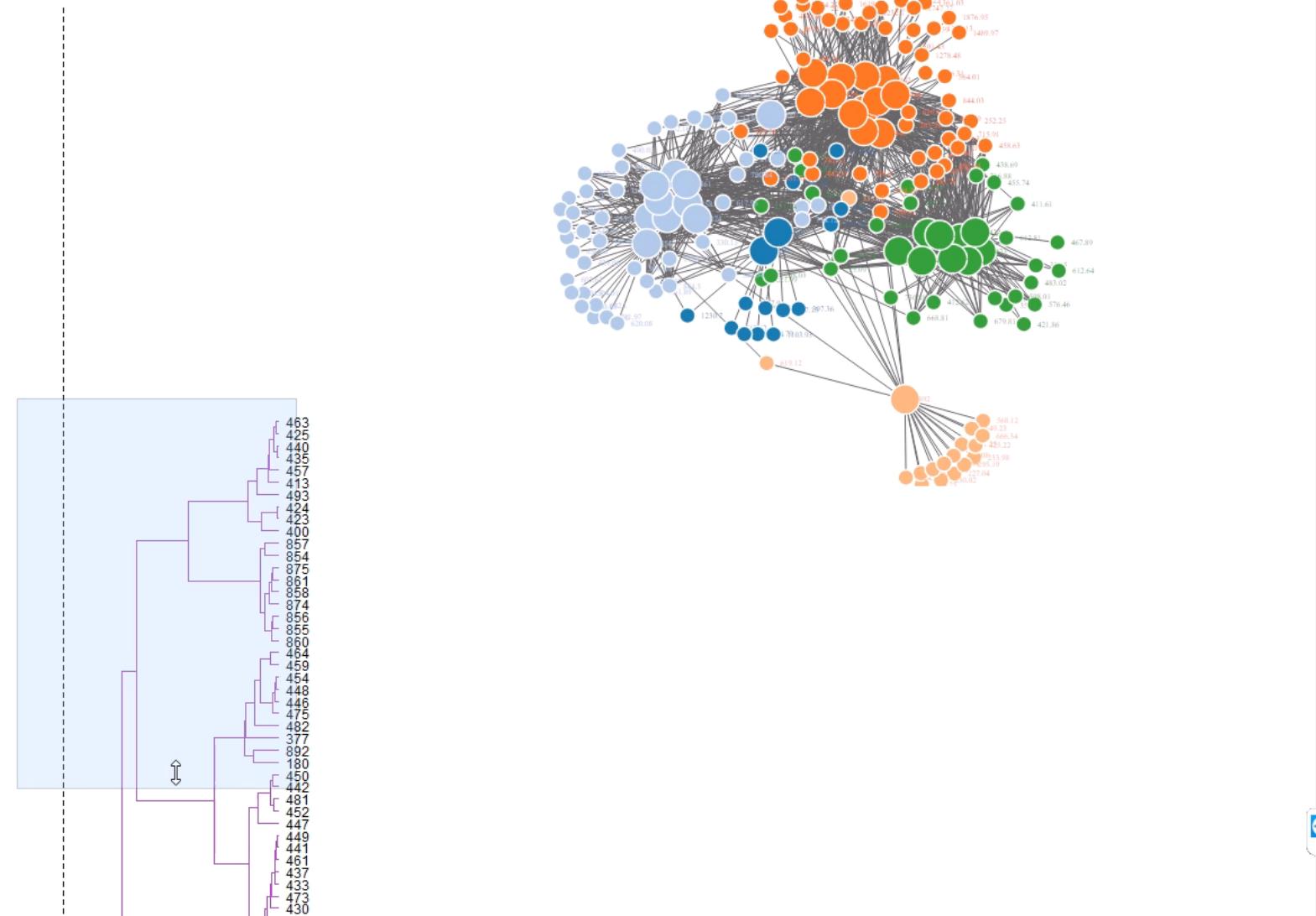
- Yes
 No (Also Turns Off Matrix Subtraction)

In what percentage of replicates must a peak be present to be kept? (0-100%) (Experiment/Hypothesis dependent)

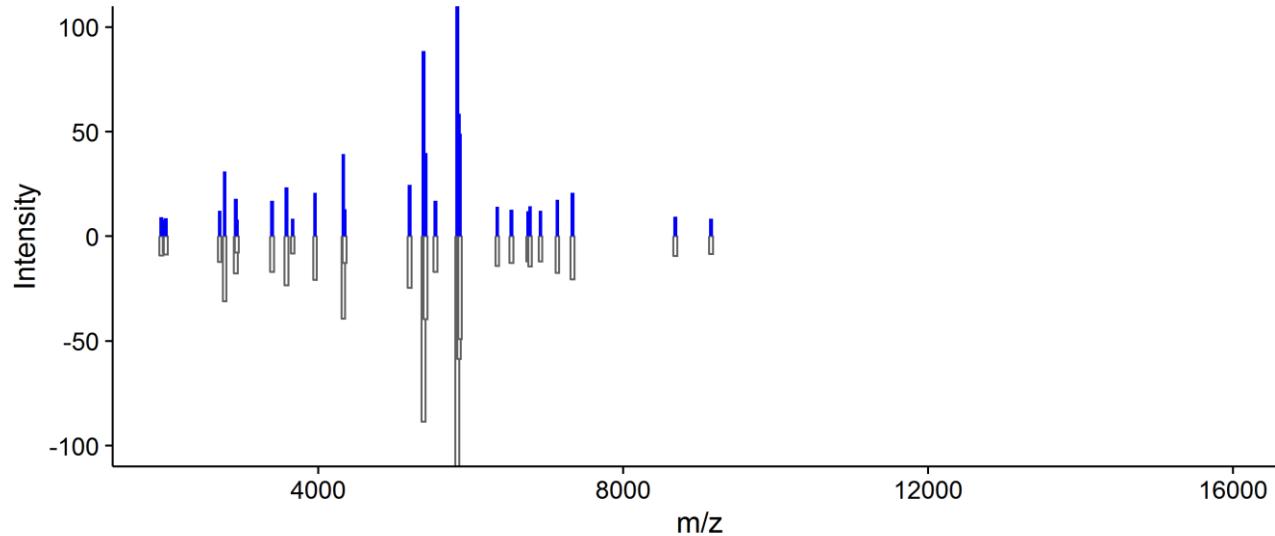
Signal To Noise Cutoff**Upper Mass Cutoff****Lower Mass Cutoff****Expand Tree****Adjust right margin of dendrogram****Download Current Network Data**

Hint 1: Use mouse to select parts of the tree and display the MAN of corresponding samples.

Hint 2: Use mouse to click & drag parts (nodes) of the MAN if it appears congested.



1002 Micrococcaceae vs 1002 Micrococcaceae

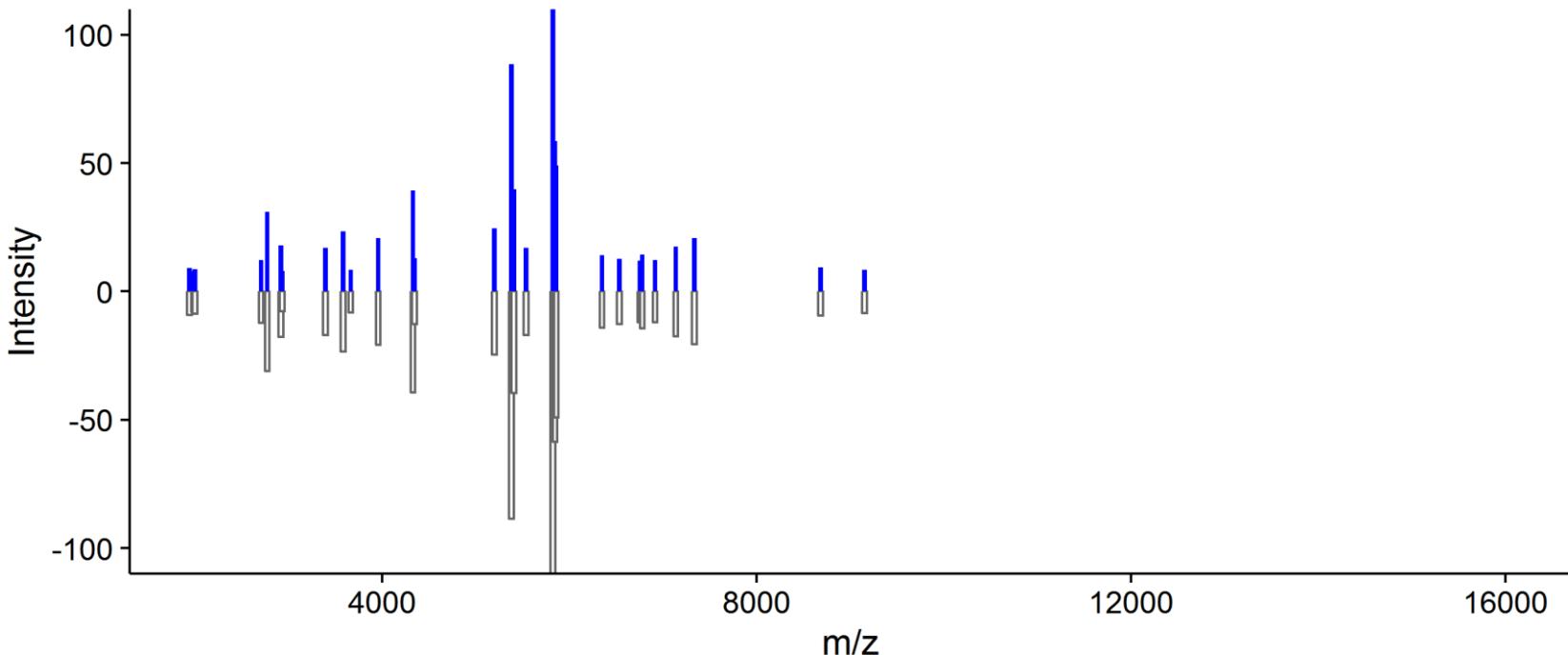


Spectra Library Search:
Cosine Similarity Score

coop::cosine()

$$\cos \theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|}$$

1002 Micrococcaceae vs 1002 Micrococcaceae



Recently Funded to “Digitize” >8,000 Bacteria for Initial Database

IDBac

127.0.0.1:7646

IDBac

Introduction PreProcessing Compare Two Samples (Protein) Hierarchical Clustering (Protein) PCA (Protein)

Metabolite Association Network (Small-Molecule) In-House Spectra Libraries

Create a New Library Add Isolates to an Existing Library Modify an Existing Library

Input Library Name:
Example

Save

	Strain_ID	Genbank_Accession	Kingdom	Phylum	Class	Order	Family	Genus	Species	Strain
1	114A-2									
2	114A-3									
3	114A-4									
4	114B-1									
5	114B-10									
6	114B-2									
7	114B-5_Actinobacteria_Aeromicrobium fastidiosum									
8	114B-6									
9	114B-8									
10	114B-9									
11	114C-1									
12	114C-2									
13	114C-3									
14	114C-4									
15	114C-5_Uncultured bacterium									
16	114C-6									
17	114C-7									
18	114C-8									



You need a new antibiotic.

You have resources to
examine 5 bacteria.

Which do you choose?



Thanks to Rstudio's **Phil Bowsher** for going above and beyond in helping me get to this conference.

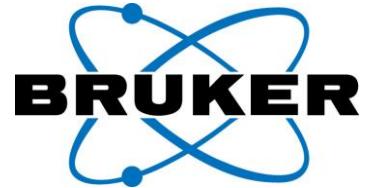
And **B2S Life Sciences** for funding my trip here:



Special Thanks



Dr. Laura Sanchez



**MEDICINAL
CHEMISTRY
AND
PHARMACOGNOSY**
COLLEGE
OF PHARMACY



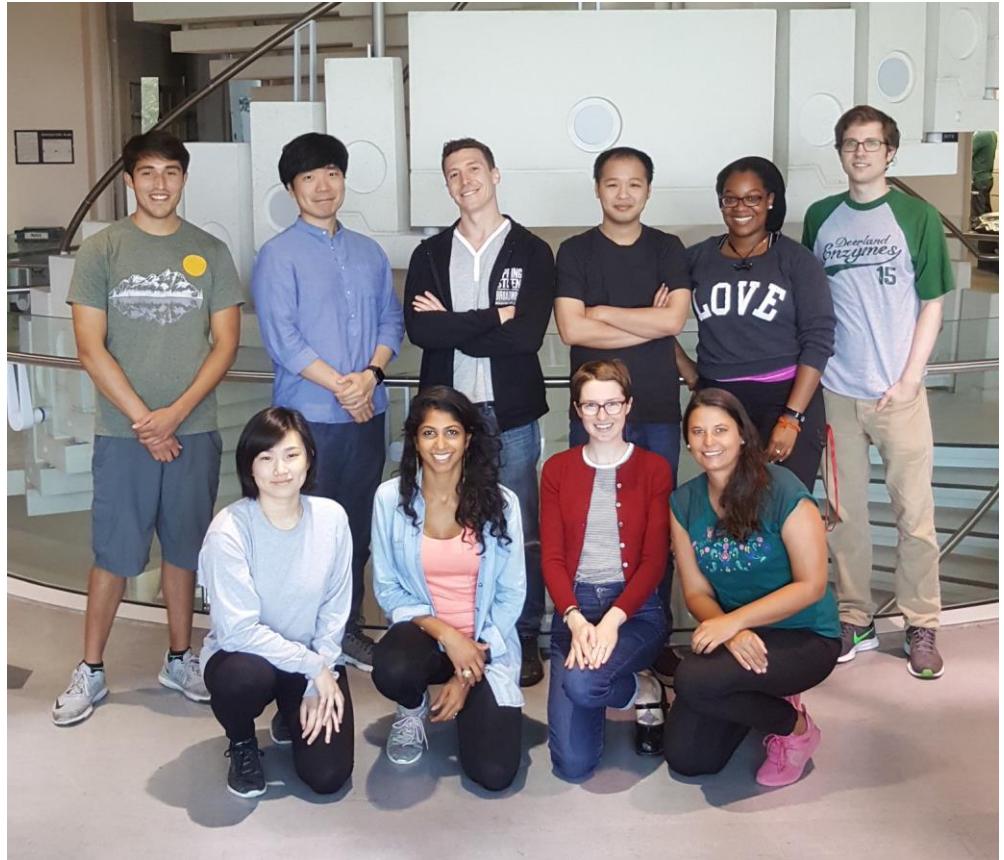
Murphy Lab

Dr. Brian Murphy
Sofia Costa
Antonio Hernandez
Dr. Jeong Ho Lee
Dr. Tuan Anh Tran
Maryam Elfeki
Vanessa Nepomuceno
Emma Ward
Erin Conley

Past Undergraduate Researchers:

Antonio Hernandez
Milan Patel
Jhewelle Fitz-Henley
Adithyan Subramanian
David An

Dr. Roberto Pronzato and Dr. Renata Manconi



@ChasingMicrobes
chasemc.github.io
murphylabuic.com
@Murphylabuic

