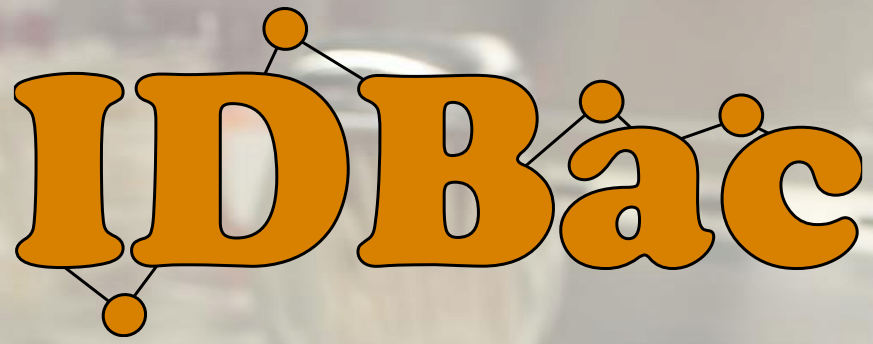
GitHub chasemc.github.io
@ChaseClarkatUIC



A Shiny App to Analyze Bacterial Profiles and Aid the Discovery of New Antibiotics

Chase Clark
PhD Student
Murphy Lab
Department of Medicinal Chemistry and Pharmacognosy
Center for Biomolecular Sciences
University of Illinois at Chicago

```
# Run peakMatching pairwise
  matched <- list(matched,lapply(steppedePairs,peakMatching))


### This version is much less RAM intensive, but we lose the groupings
  peakMatching<-function(pairs)
{as.vector(unlist(lapply(lapply(seq_along(peakLists[pairs[1]][[1]]),function(x){
                    ppmVector <- ppmAcrossVector(peakLists[pairs[1]][[1]],100)
                    rety<-which(peakLists[pairs[2]][[1]] <=
                    peakLists[pairs[1]][[1]][x]+ppmVector[x] &
                    peakLists[pairs[2]][[1]] >=
                    peakLists[pairs[1]][[1]][x]-ppmVector[x])
                    }
                  ),function(z) length((z)))))}
```

add a comment

2 Answers

active oldest votes

1

Consider a nested *apply* family call:

- `mapply()` - outer: pairwise iteration between corresponding columns of *Data_1* and *Data_2*
- `sapply` - inner: vector iteration of each value in the *Data_1* column for element comparison

Below checks if each *Data_1* value is between the two values of each column of *Data_2*:

Data

```
txt = '  data1       data2       data3      data4
-0.710003   -0.714271   -0.709946   -0.713645
-0.710458   -0.715011   -0.710117   -0.714157
-0.71071    -0.714048   -0.710235   -0.713515
-0.710255   -0.713991   -0.709722   -0.713972'

Data_1 <- read.table(text=txt, header=TRUE)

txt = ' dataA       dataB       dataC      dataD
-0.71097    -0.714059   -0.70928    -0.714059
-0.710343   -0.714576   -0.709338   -0.713644'

Data_2 <- read.table(text=txt, header=TRUE)
```

Code

```
check_inbetween <- function(x,y){
  sapply(x, function(i) (i > y[1] & i < y[2]))
}

inbetween_matrix <- mapply(check_inbetween, Data_1, Data_2)

inbetween_matrix
#      data1 data2 data3 data4
# [1,] FALSE FALSE FALSE  TRUE
# [2,]  TRUE FALSE FALSE FALSE
# [3,]  TRUE FALSE FALSE FALSE
# [4,] FALSE FALSE FALSE  TRUE
```

share improve this answer

answered Jul 7 '17 at 1:04

Parfait
34k • 7 • 23 • 55

Thanks @Parfait, I guess even `sweep()` can be used – Chetan Arvind Patil Jul 7 '17 at 1:18

1 Possibly. I never used `sweep()`. And that implementation has an interesting nested sweep. My attempt seems to not recycle the function but only runs on first columns of datasets: `sweep(as.matrix(Data_1), 1, as.matrix(Data_2), FUN = function(x,y) (x > y[1] & x < y[2]), check.margin = FALSE)` – Parfait Jul 7 '17 at 1:58

add a comment

0 Optimize Apply() While() in R

Related

1014 How to sort a dataframe by column(s)?

0 Using mapply() in R over rows, vs. columns

1 efficient way of coding for matching using

ChemicalData returns the correct result for "Sulfuric" but an empty list for "Sulfuric Acid". Why?

No space left, even though I have an untouched 990 GB partition

What is this methodology called?

What are good words to refer to the condition of objects?

If I generate a random symmetric matrix, what's the chance it is positive definite?

Dealing with aggressive student suspected to be cheating

During single-engine taxi, how is the asymetric

Thanks Parfait!

answered Jul 7 '17 at 1:04

Parfait

34k • 7 • 23 • 55

Quick Outline

- Why we are trying to discover new antibiotics
- It’s not easy!!
- A new pipeline to prioritize bacteria for antibiotic discovery
- What is Shiny?
- Brief overview of IDBac
- Some tips about Shiny

On the Precipice of a Pre-Antibiotic Era

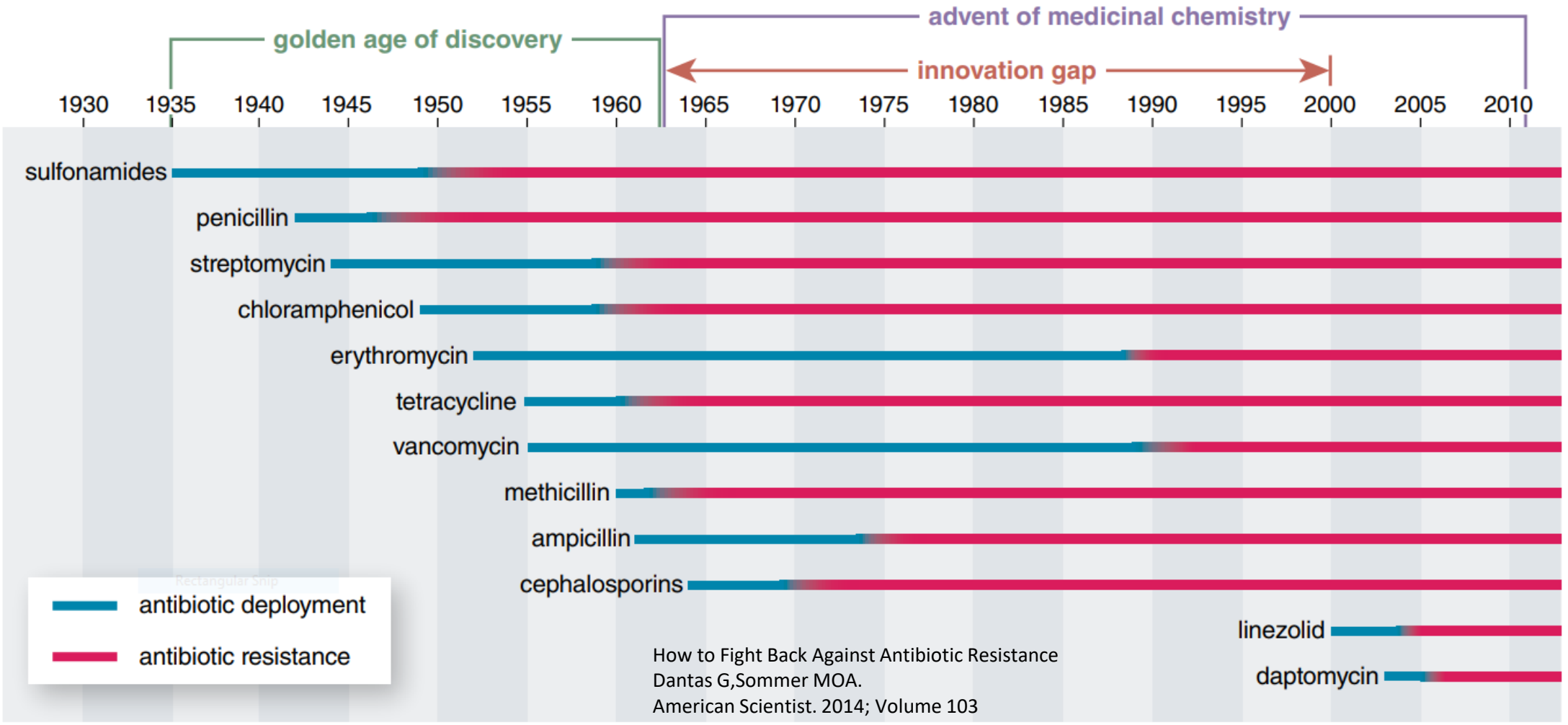
World Health Organization

The world is running out of antibiotics ,
WHO report confirms
News release 20 Septmember 2017 | Geneva

Antibiotics are Like R Packages…

New Packages are Needed to Solve Difficult Problems

But only a couple new packages since the 60's!!!!!!!!



How to Fight Back Against Antibiotic Resistance
Dantas G,Sommer MOA.
American Scientist. 2014; Volume 103

Ironically, the most prolific source of antibiotics has been… bacteria!

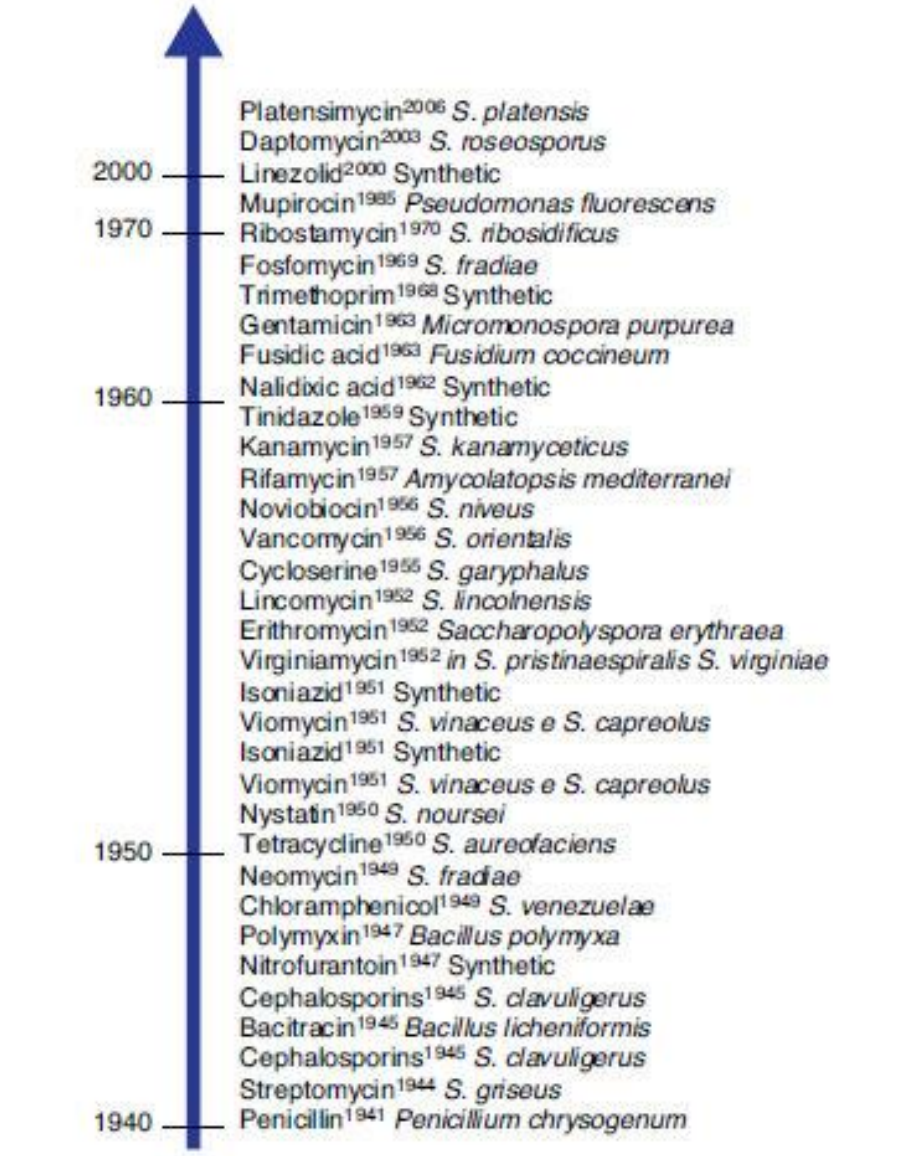
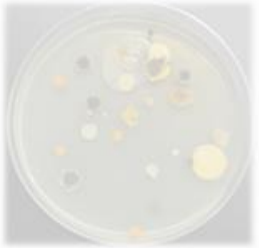


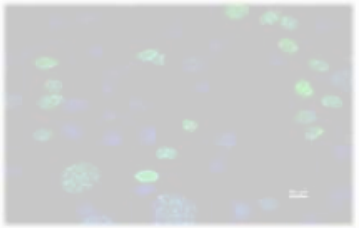
Fig. 1 – Key findings and dates of antibiotics. Highlights of the *Streptomyces*.

http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-86702012000500011


FROM
THE
LAB
SCIENCE
MUSEUM

Current Methods: Are Like Finding a Needle in a Hay Stack… While Blindfolded





Sample collection

Isolate Bacteria

Grow bacteria, crude chemical extract, test for activity

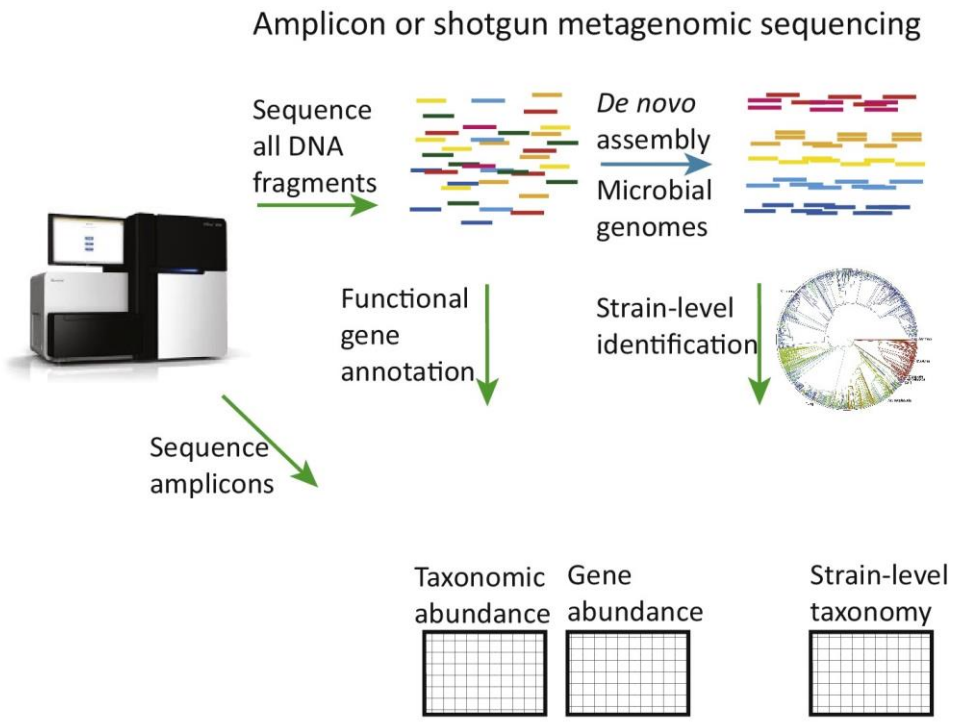
Lots of work

Years of work

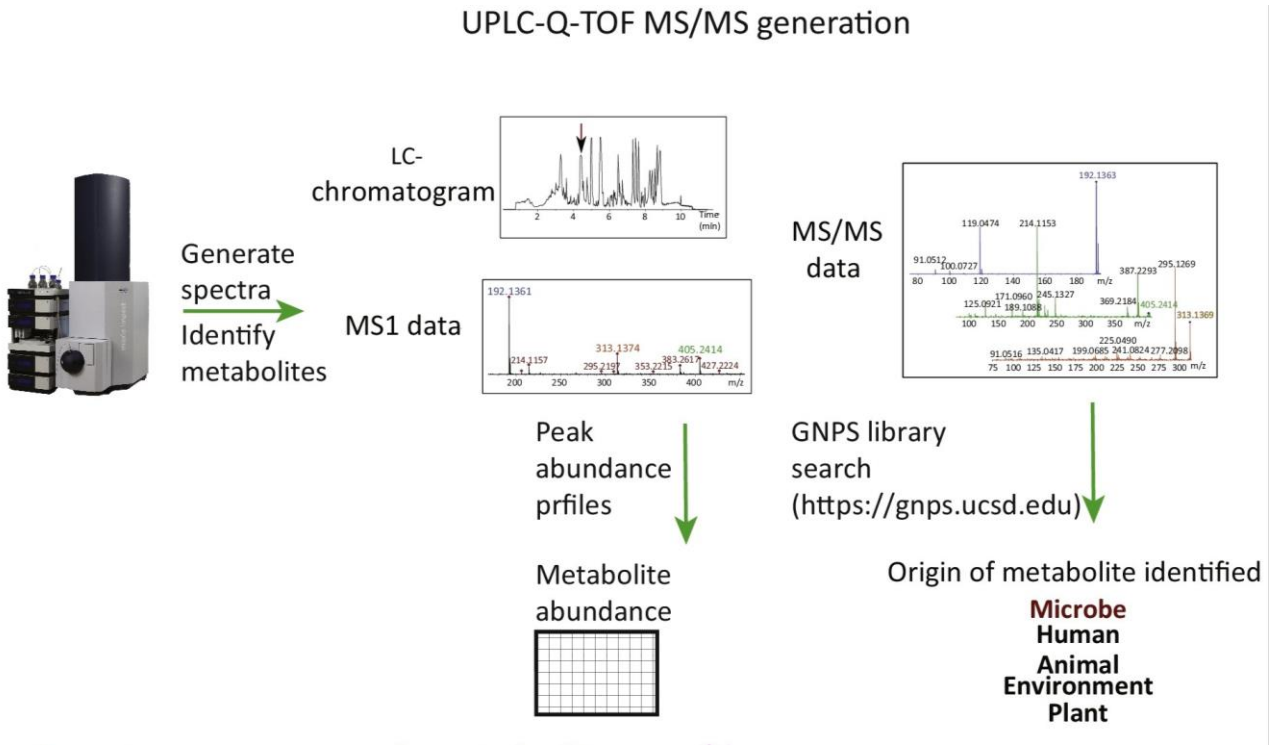
New Drug!

Each Bacterium Studied Requires Significant Time and Monetary Investment

Genomics



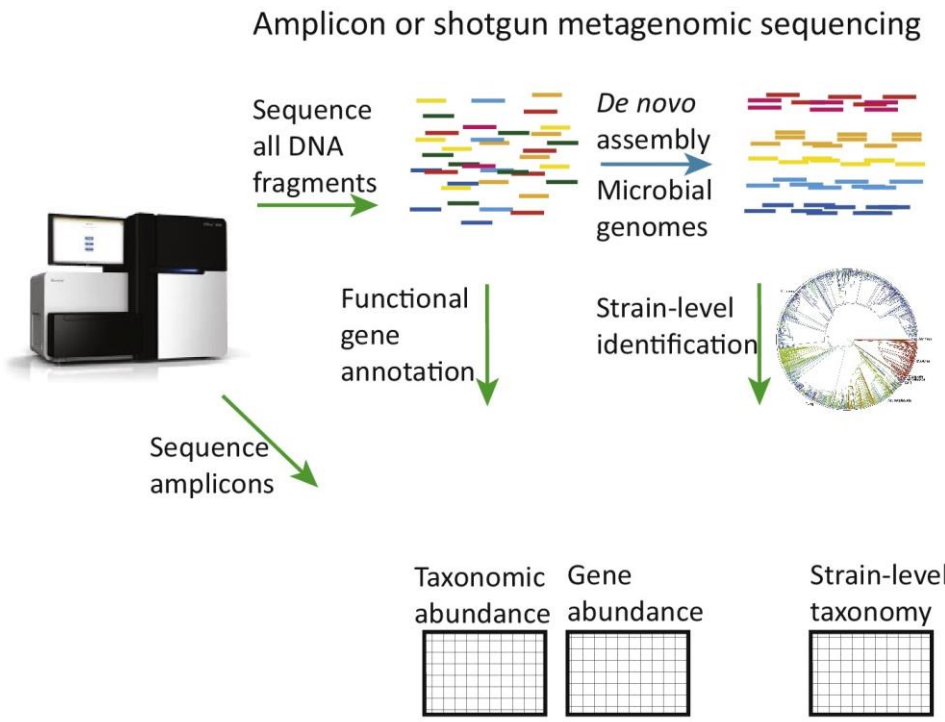
Low-Throughput Metabolomics



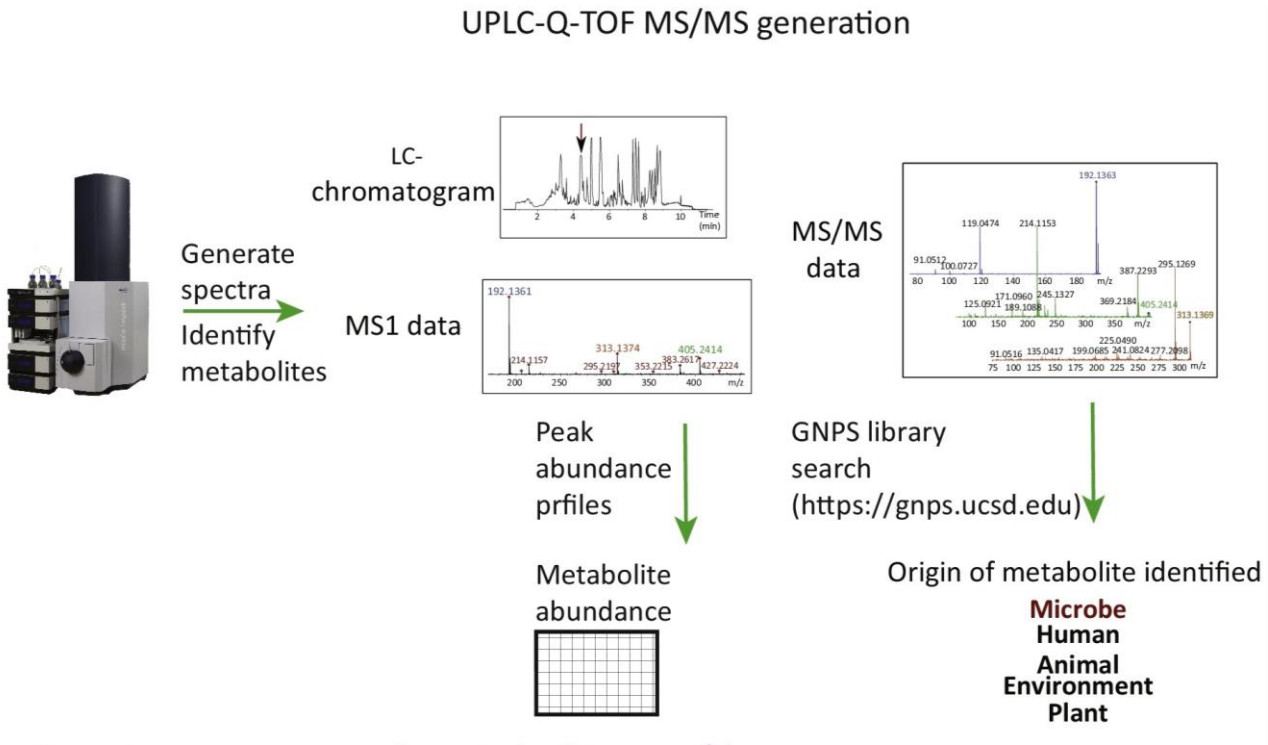
Metcalf, Jessica L. et al.
Trends in Biotechnology , Volume 35 , Issue 9 , 814 – 823
http://dx.doi.org/10.1016/j.tibtech.2017.03.006

Each Bacterium Studied Requires Significant Time and Monetary Investment

Genomics



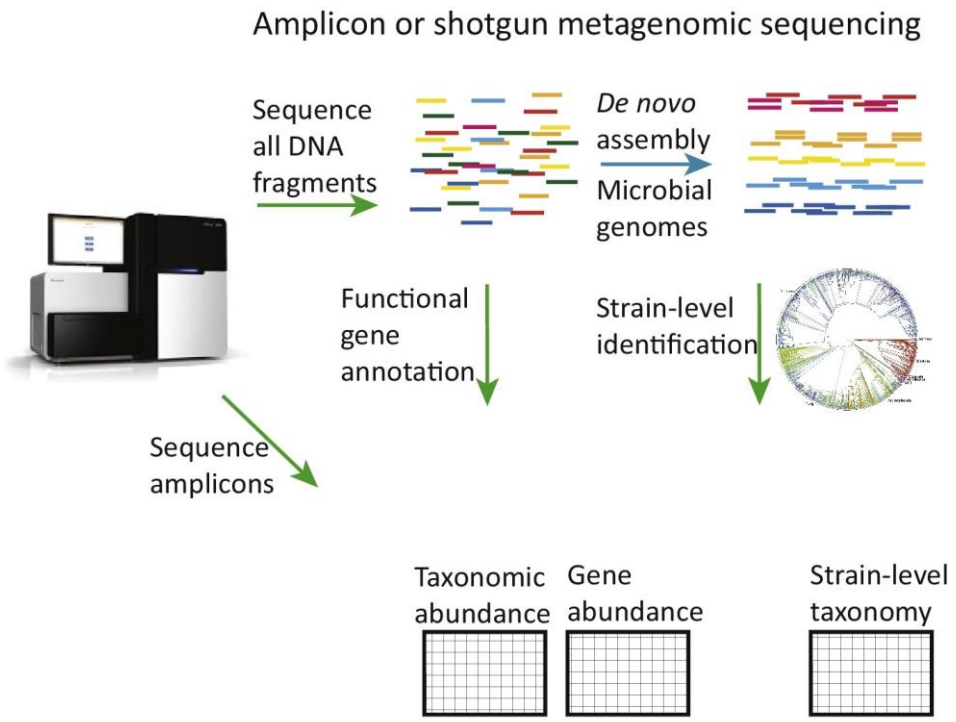
Low-Throughput Metabolomics



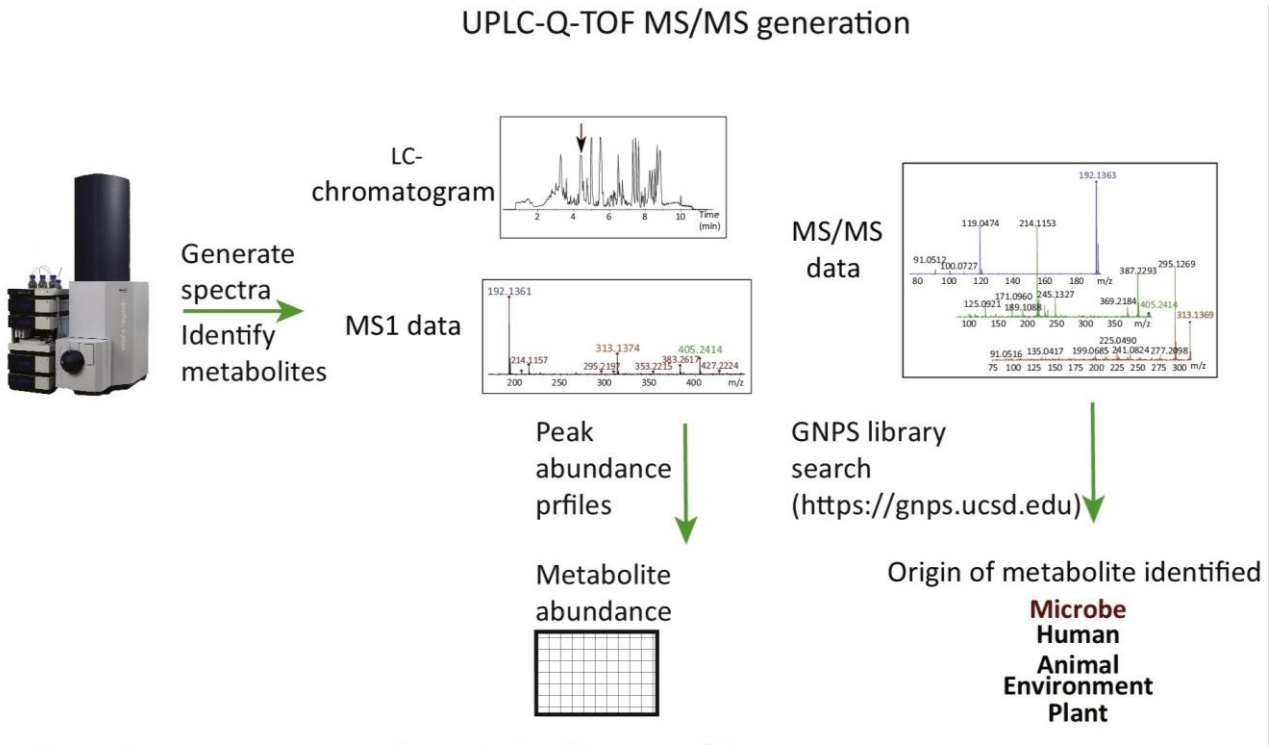
Metcalf, Jessica L. et al.
Trends in Biotechnology , Volume 35 , Issue 9 , 814 – 823
http://dx.doi.org/10.1016/j.tibtech.2017.03.006

Each Bacterium Studied Requires Significant Time and Monetary Investment

Genomics



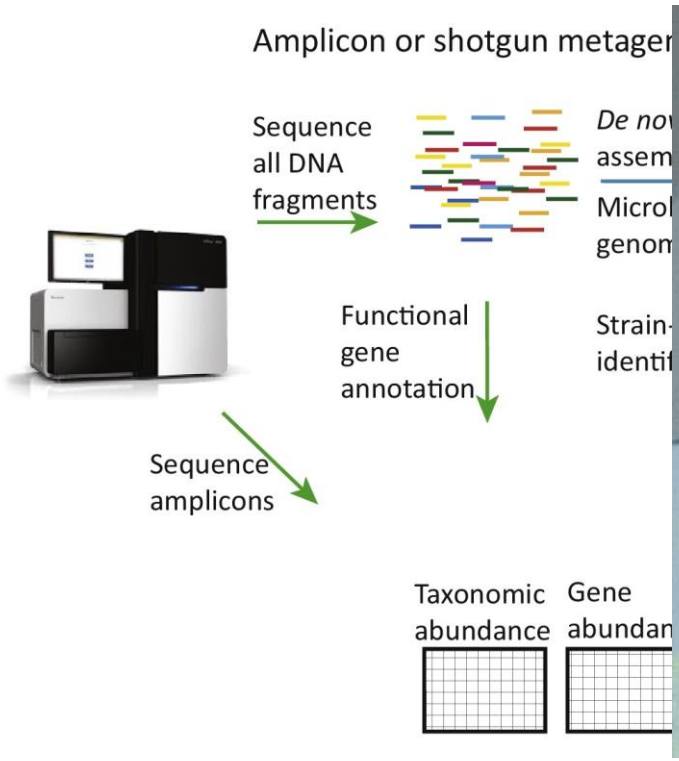
Low-Throughput Metabolomics



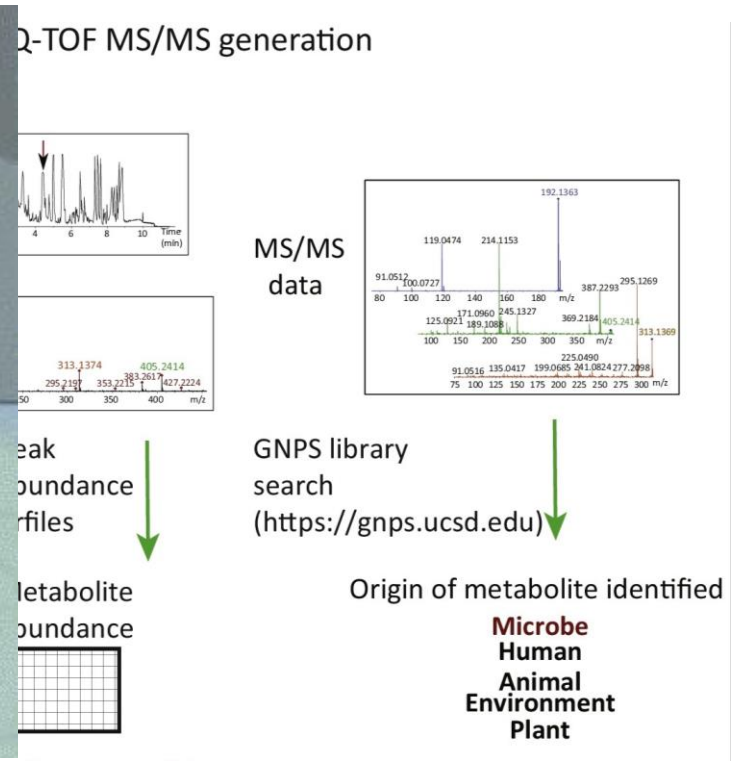
Metcalf, Jessica L. et al.
Trends in Biotechnology , Volume 35 , Issue 9 , 814 – 823
http://dx.doi.org/10.1016/j.tibtech.2017.03.006

Each Bacterium Studied Requires Significant Time and Monetary Investment

Genomics



Low-Throughput Metabolomics

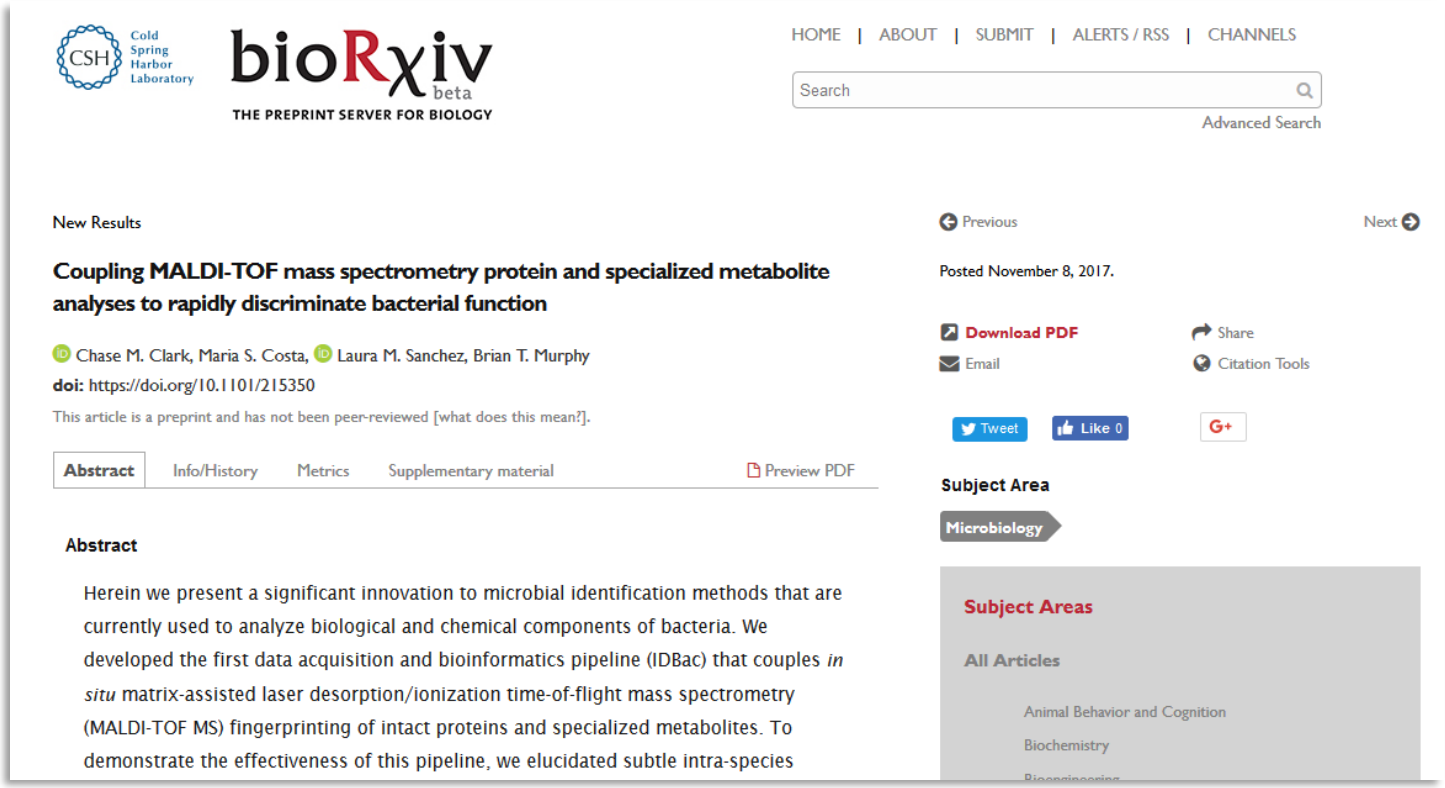


Metcalf, Jessica L. et al.
Trends in Biotechnology , Volume 35 , Issue 9 , 814 – 823
http://dx.doi.org/10.1016/j.tibtech.2017.03.006

Labs have trouble prioritizing which bacteria to study.

- Which orange colonies are actually the same?
  - Visual cues are not enough!
- Do they all produce the same chemistry?
  - Visual cues are not enough!

Our Lab Recently Developed IDBac: A High Throughput Method to Rapidly Distinguish Bacterial Phylogeny and Functional Metabolism



Cold Spring Harbor Laboratory

bioRxiv beta

THE PREPRINT SERVER FOR BIOLOGY

HOME | ABOUT | SUBMIT | ALERTS / RSS | CHANNELS

Search

Advanced Search

New Results

Previous Next

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

Posted November 8, 2017.

Chase M. Clark, Maria S. Costa, Laura M. Sanchez, Brian T. Murphy

doi: https://doi.org/10.1101/215350

This article is a preprint and has not been peer-reviewed [what does this mean?].

Download PDF Share

Email Citation Tools

Tweet Like 0 G+

Abstract Info/History Metrics Supplementary material Preview PDF

Subject Area

Microbiology

Abstract

Herein we present a significant innovation to microbial identification methods that are currently used to analyze biological and chemical components of bacteria. We developed the first data acquisition and bioinformatics pipeline (IDBac) that couples *in situ* matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) fingerprinting of intact proteins and specialized metabolites. To demonstrate the effectiveness of this pipeline, we elucidated subtle intra-species

Subject Areas

All Articles

Animal Behavior and Cognition

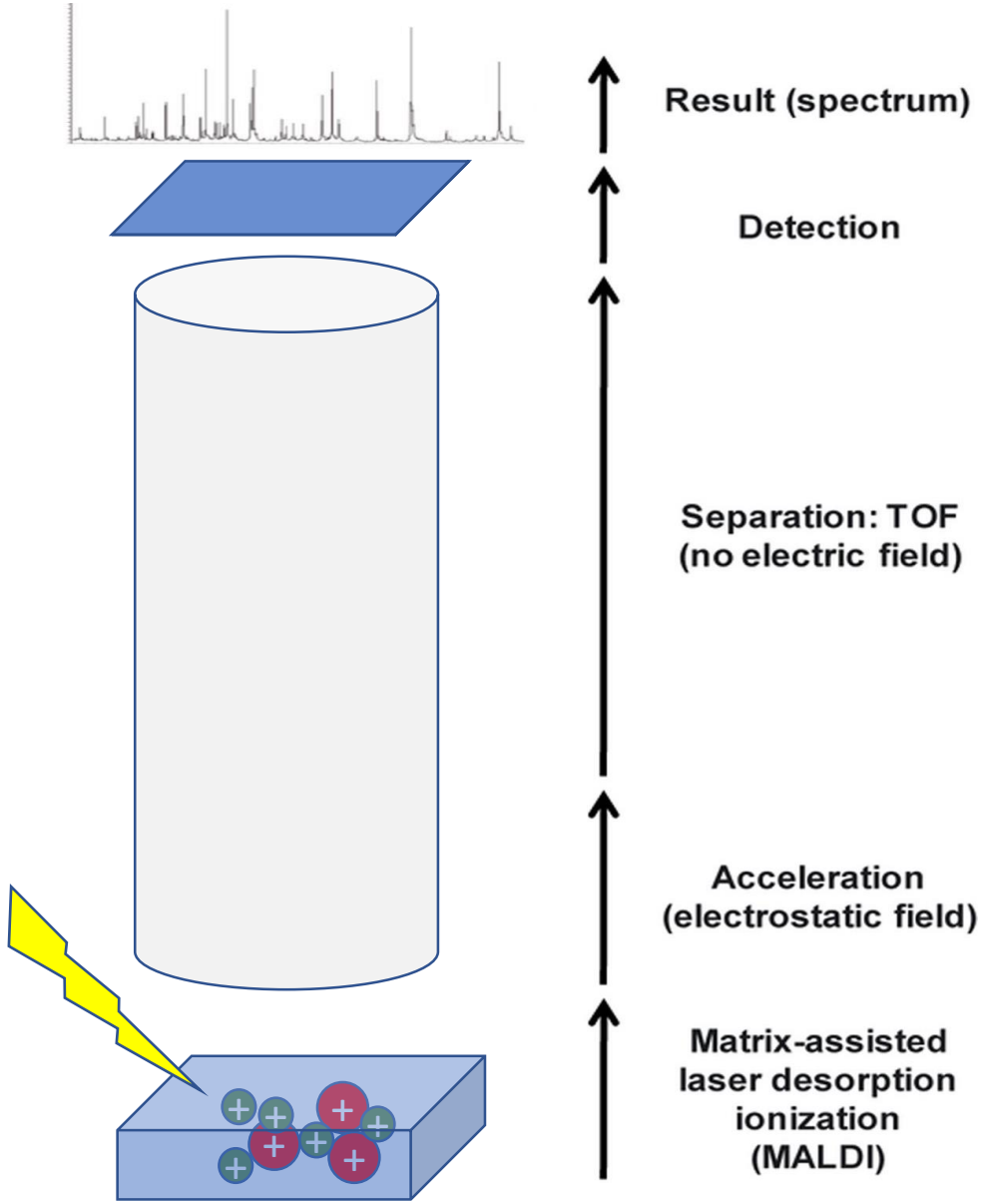
Biochemistry

chasemc.github.io/IDBac

https://github.com/chasemc/IDBac

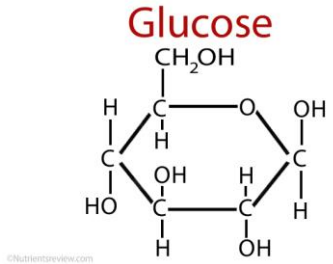
MALDI-TOF: A Rapid and Simple Way to Weigh Molecules From Bacteria

Up to 384 bacteria in one experiment

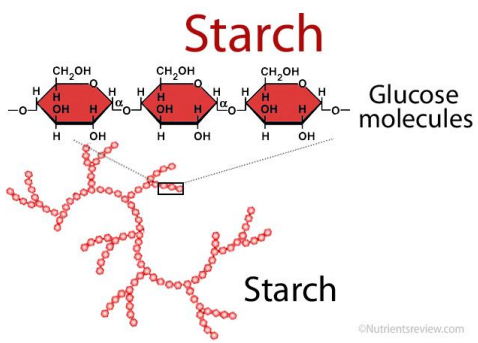


We’re Just Weighing Molecules!

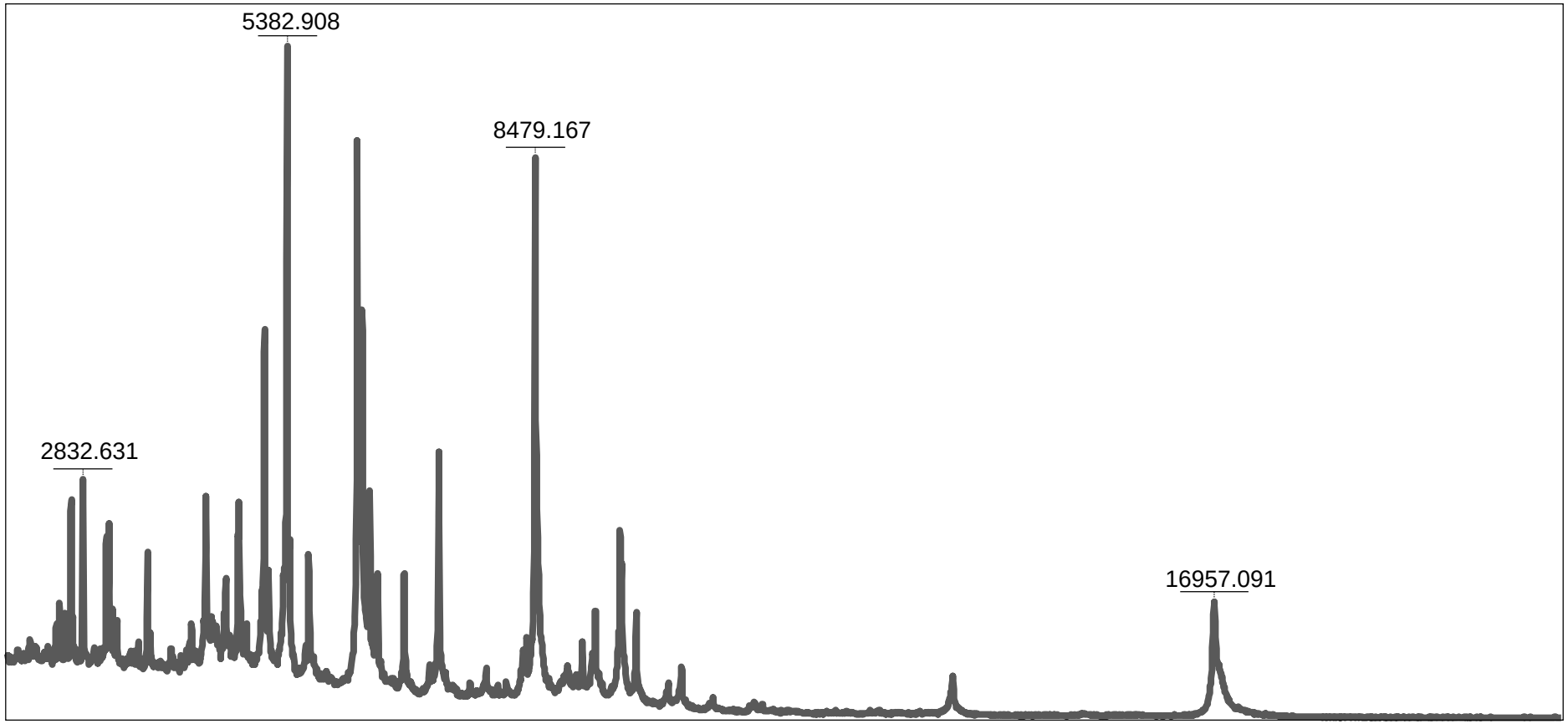
Light Molecules



Heavy Molecules



We're Just Weighing Molecules!



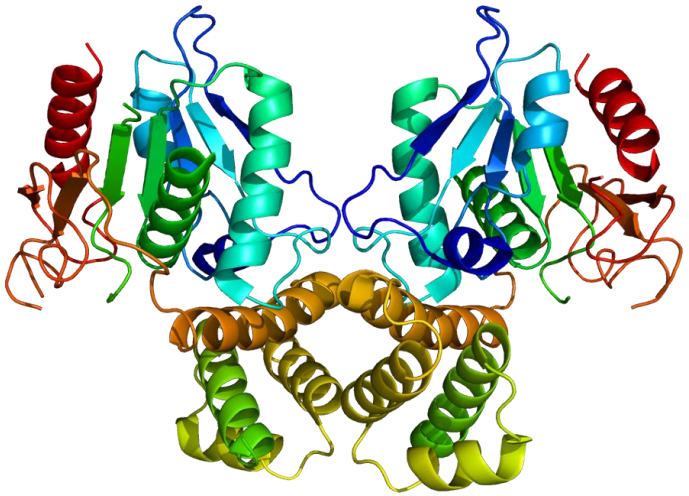
plot(readRDS(list.files(pattern="SummedProteinSpectra.rds")[[1]]),xlim=c(2000,12000))

We Weigh Two Types of Molecules = Two Spectra/Sample

Protein Data

(Identify Bacteria)

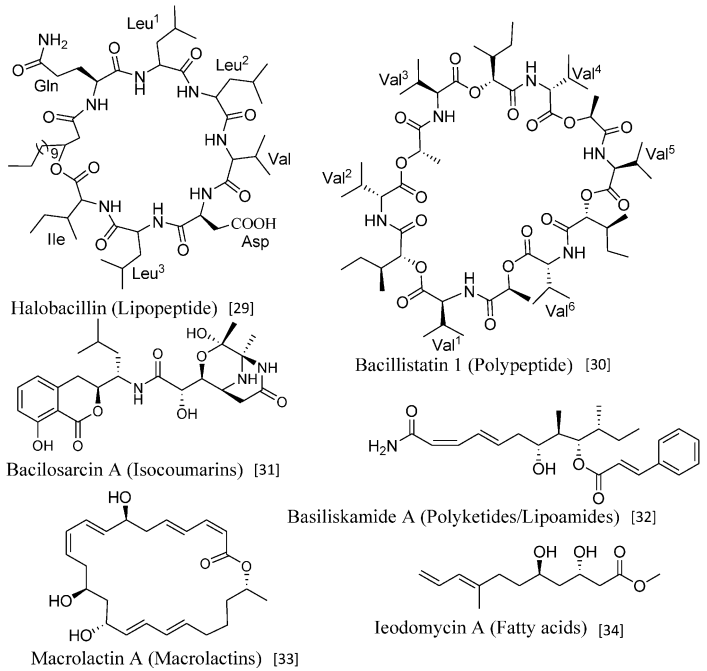
- Big molecules that are evolutionarily stable.



Specialized Metabolite Data

(Potential Drugs)

- Small molecules that provide a selective advantage.

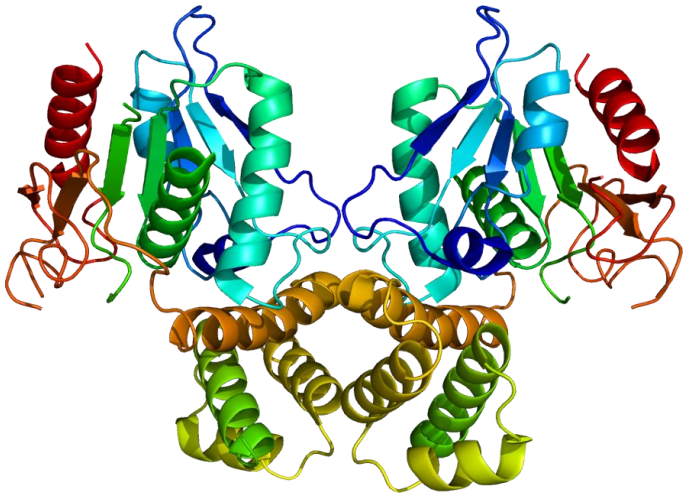


We Weigh Two Types of Molecules = Two Spectra/Sample

Protein Data

(Identify Bacteria)

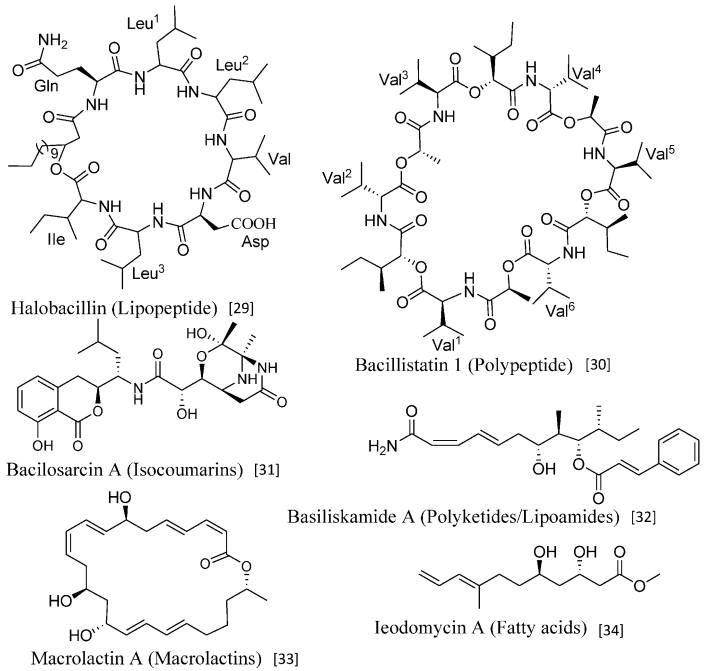
- Big molecules that are evolutionarily stable.



Specialized Metabolite Data

(Potential Drugs)

- Small molecules that provide a selective advantage.

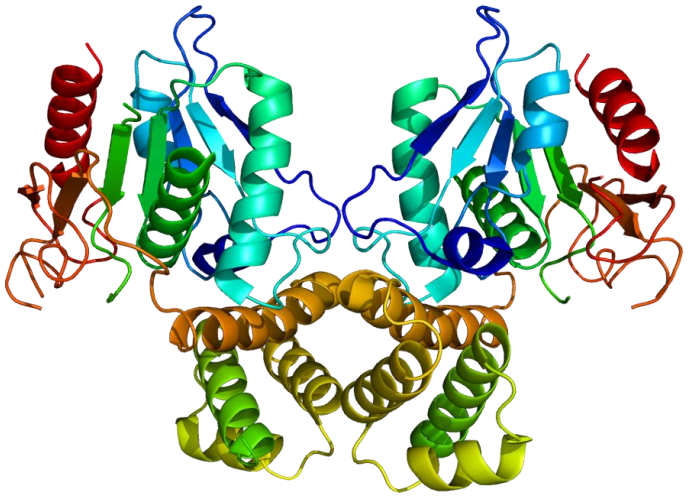


We Weigh Two Types of Molecules = Two Spectra/Sample

Protein Data

(Identify Bacteria)

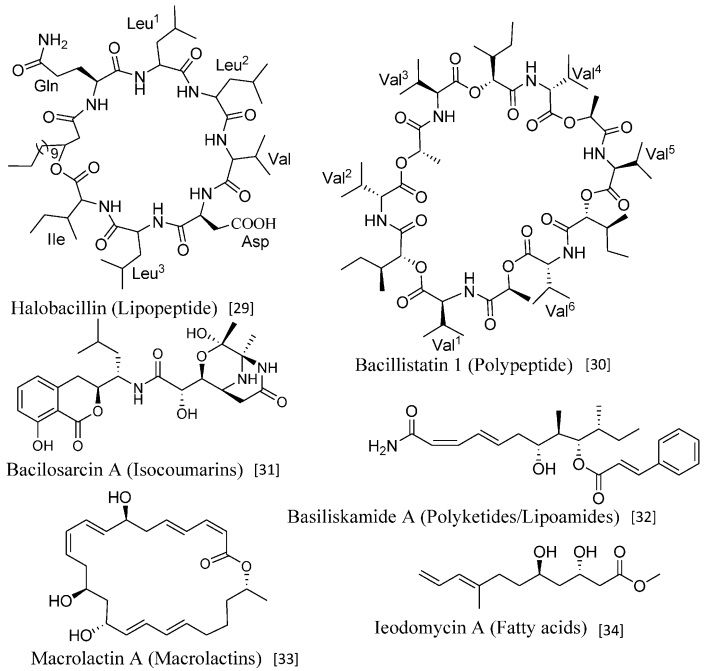
- Big molecules that are evolutionarily stable.



Specialized Metabolite Data

(Potential Drugs)

- Small molecules that provide a selective advantage.



We Take Two Measurements:

- Big molecules
- Small Molecules

Data processing becomes
a major bottleneck

Not “Big Data”…But Not Small Data

$$(384\ spots \times 1\ Protein\ Data\ Set) +$$
$$(384\ spots \times 1\ Small\ Molecule\ Data\ Set\ ) =$$
$$\boldsymbol{768\ Data\ Files\ per\ Run}$$

	Small Mol Total	Protein Total
Instrument Data Format	774 MB	156 MB
mzML format	1.43 GB	251 MB

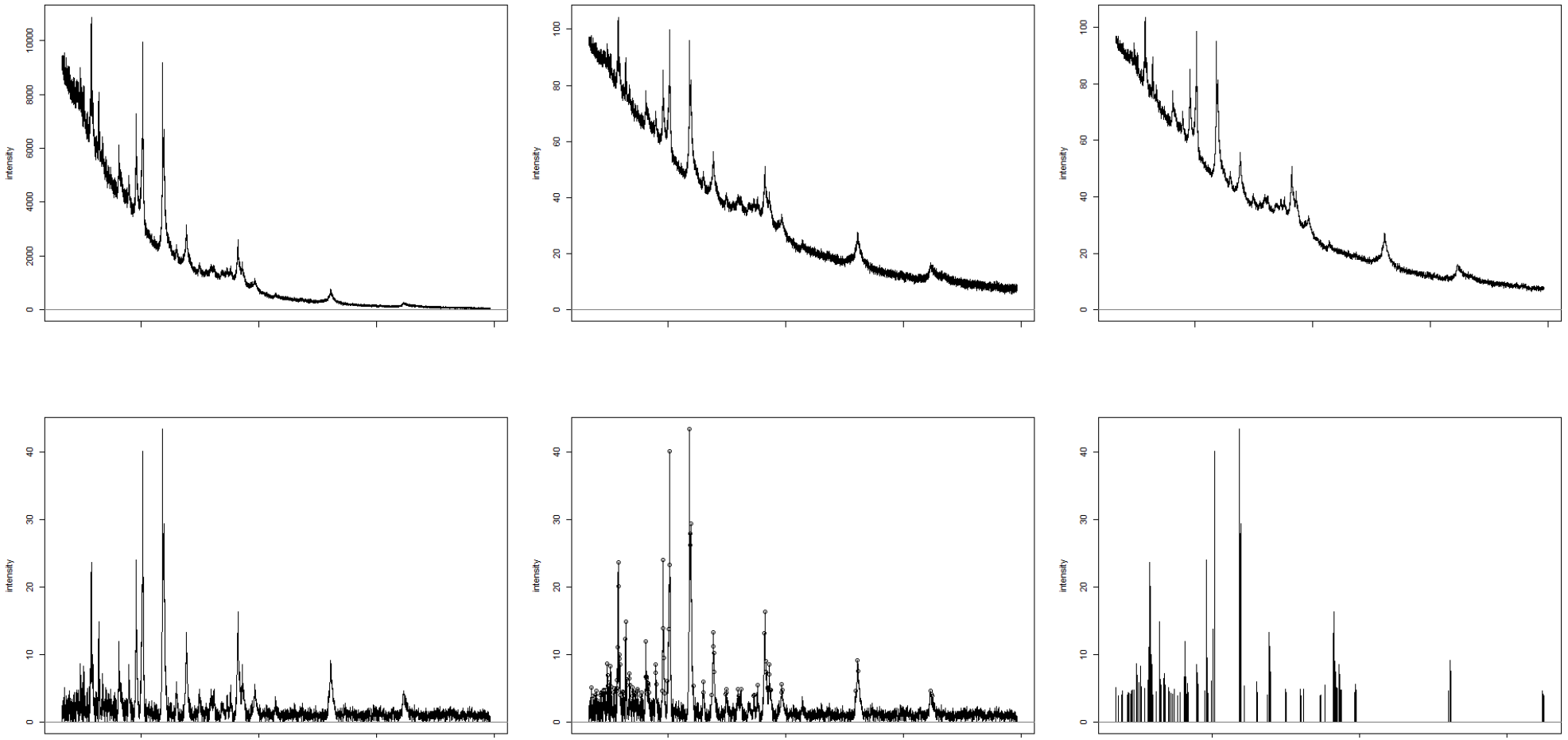
For context, we aren’t dealing with just one diversity plate ->

We are dealing with hundreds of plates. That means thousands of bacteria!

Automated Processing of MALDI Protein Spectra in R

Code here: https://github.com/chasemc/Playground/blob/master/RshinyPres/README.md#spectra-processing

1. Raw Spectrum
2. Intensity Normalization
3. Smoothing
4. Baseline Correction
5. Peak Detection
6. Extract Peak Data

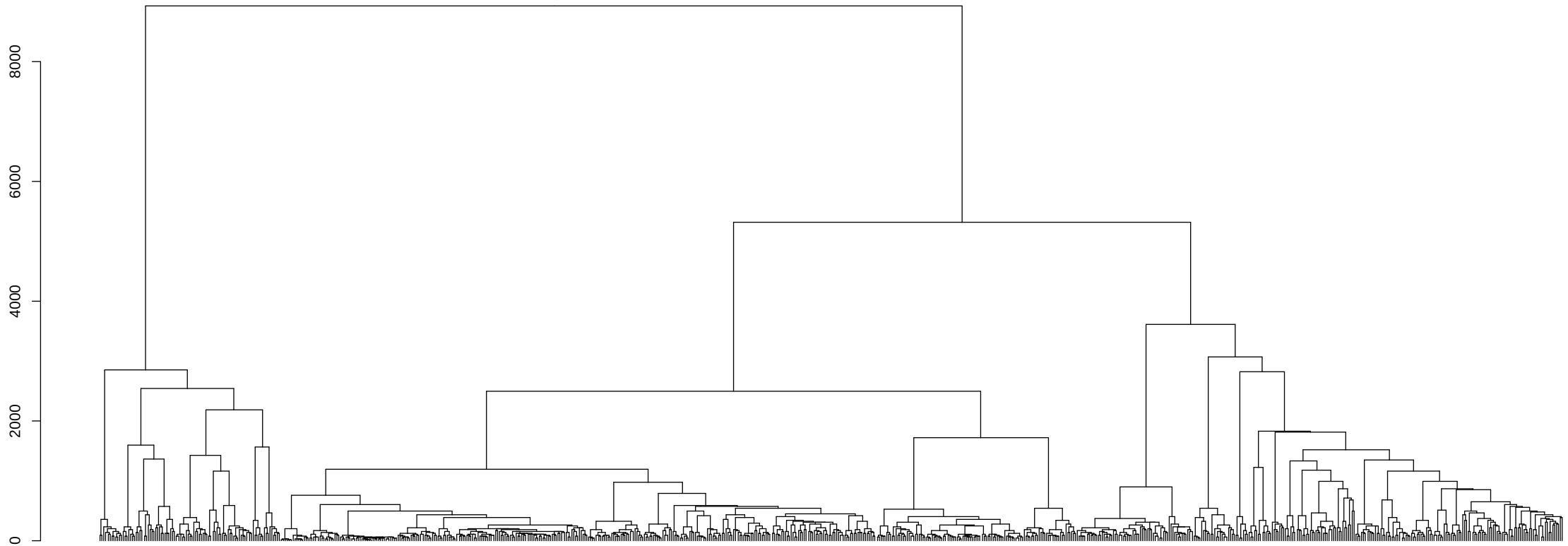


Most of these are from the MALDIquant package by Sebastian Gibb

Some of the Analyses Offered in the IDBac Shiny App… Real Data, Real Results

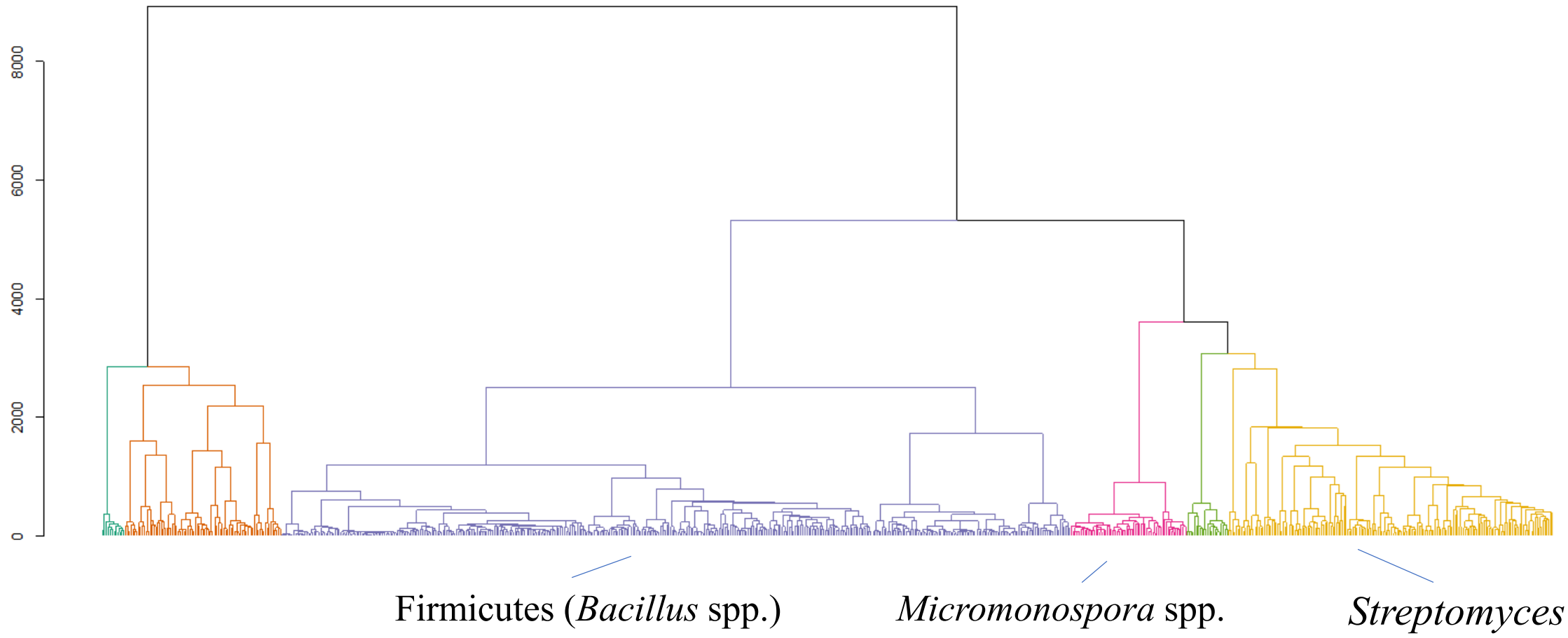
We’re almost to the Shiny Part!

1) Hierarchical Clustering of Protein Data



Sub-sample of ~ 750 isolates

1) Hierarchical Clustering of Protein Data = Type of Bacteria



Sub-sample of ~ 750 isolates

Problem:
Same Species of Bacteria Can Produce Different Chemicals

Problem:
Same Species of Bacteria Can Produce Different Chemicals

That’s why we collect two types of data!

Analyzing Small Molecule Data

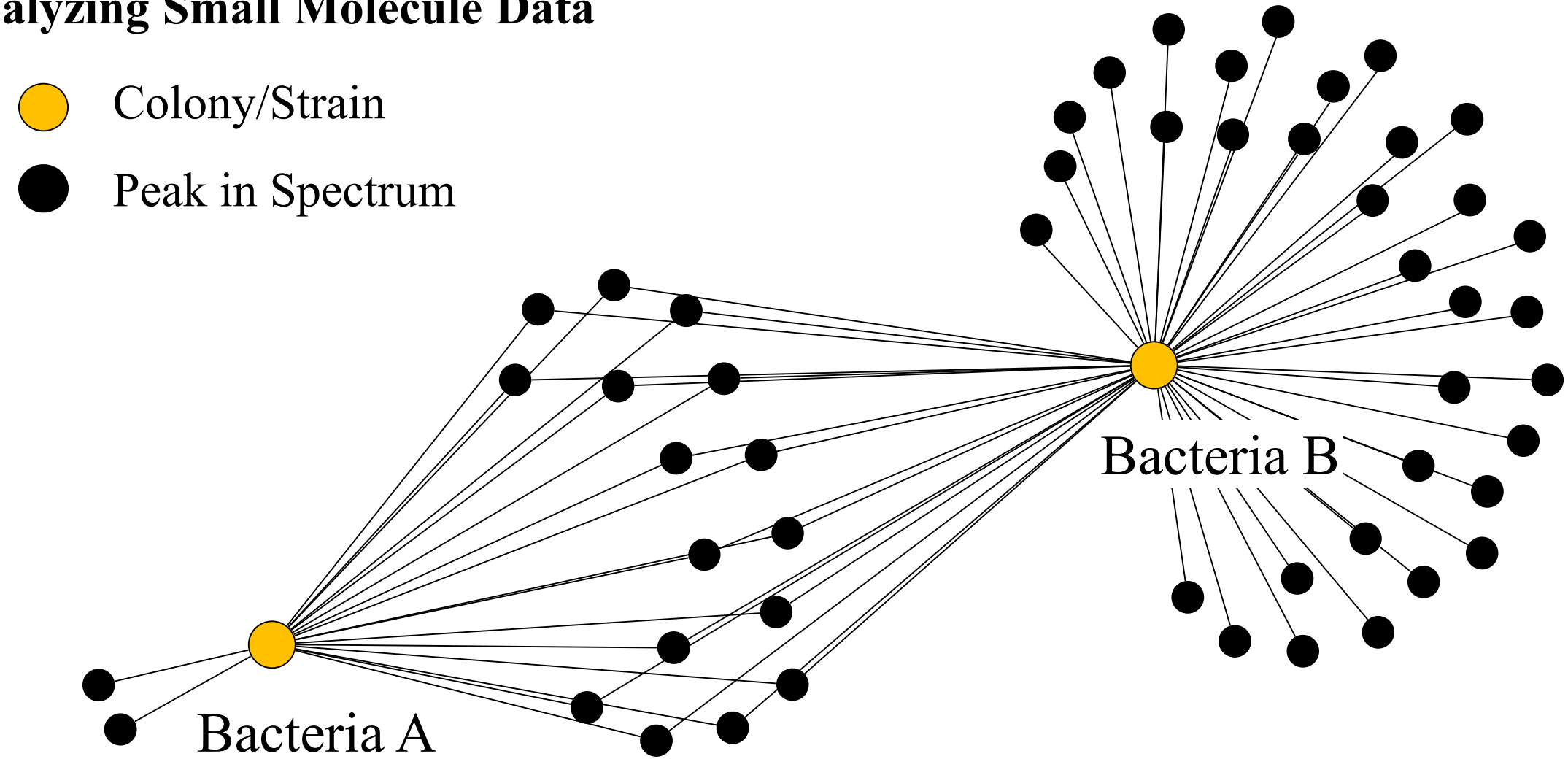
Metabolite Association Network

Analyzing Small Molecule Data

- Colony/Strain
- Peak in Spectrum

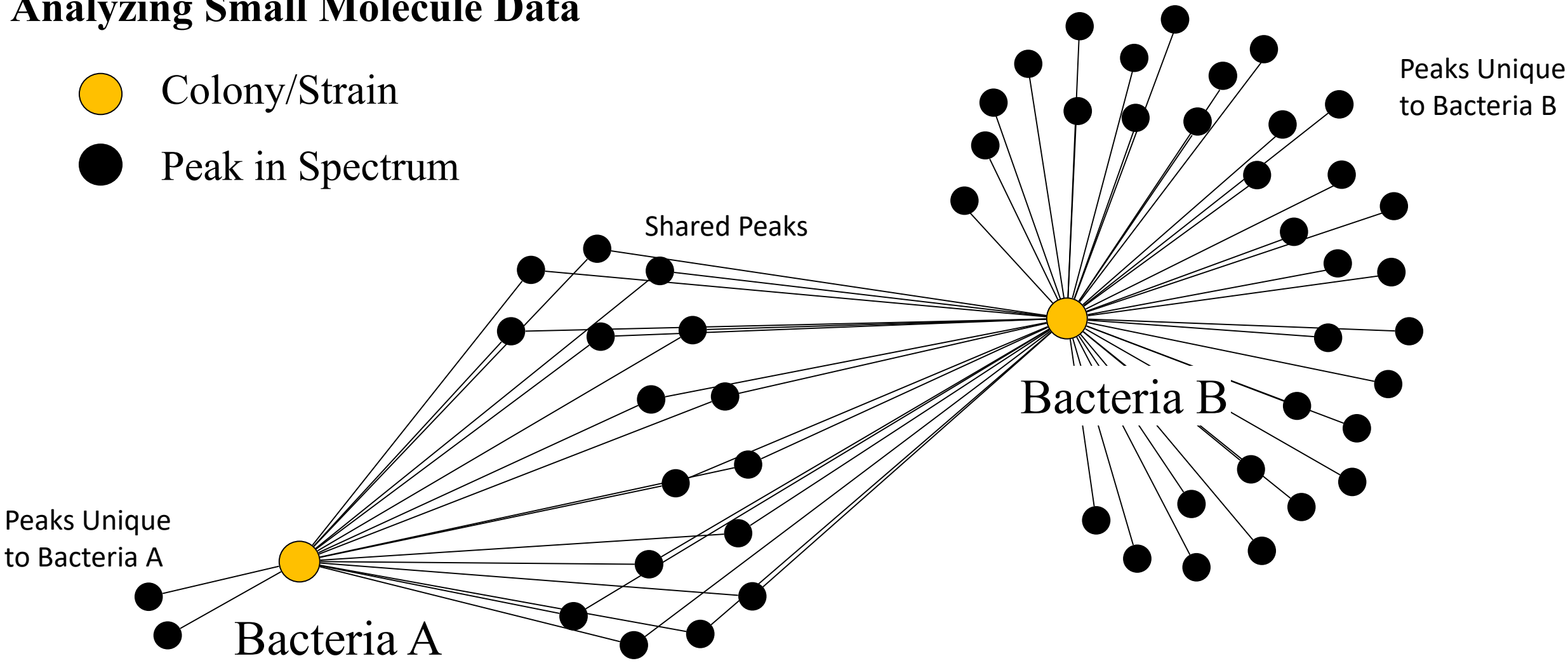
Metabolite Association Network

Analyzing Small Molecule Data



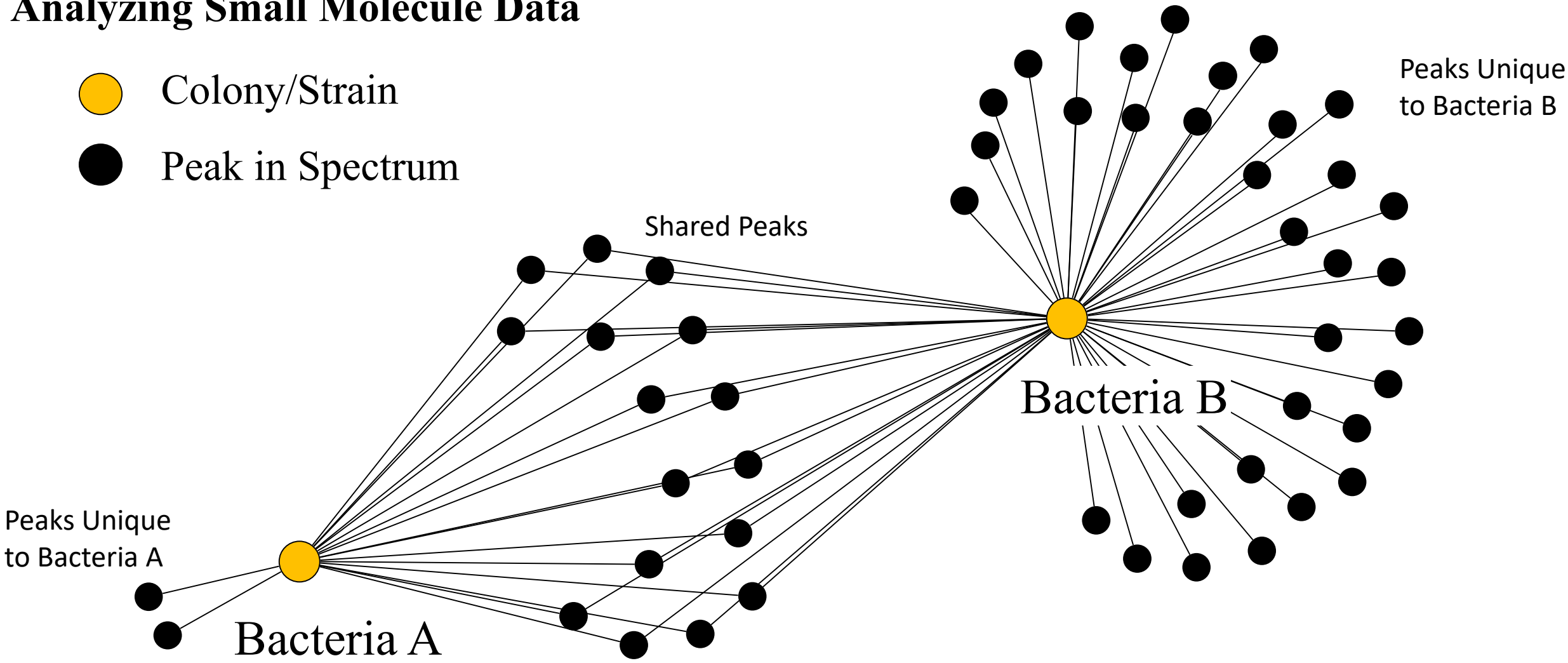
Metabolite Association Network

Analyzing Small Molecule Data

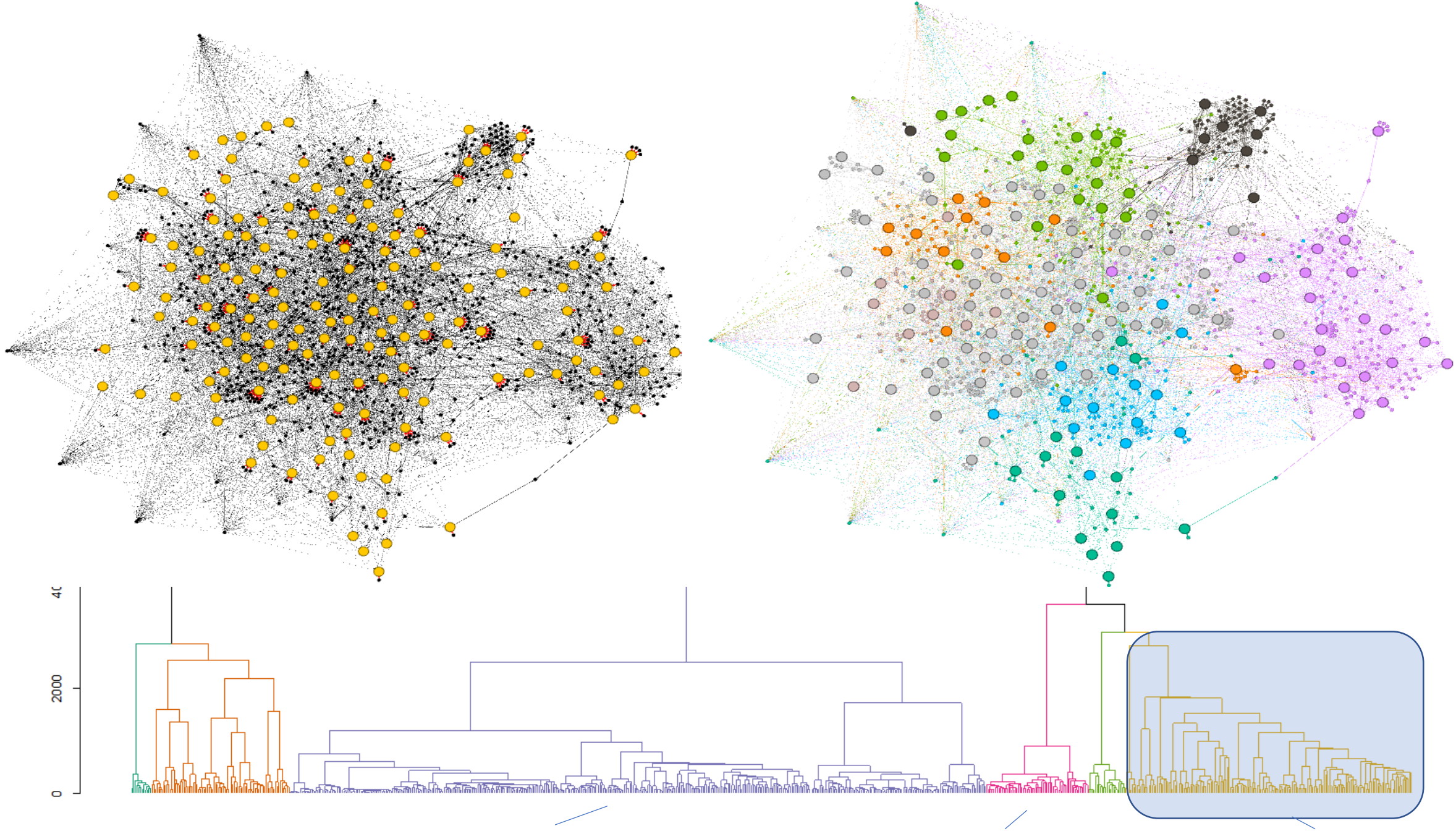


Metabolite Association Network

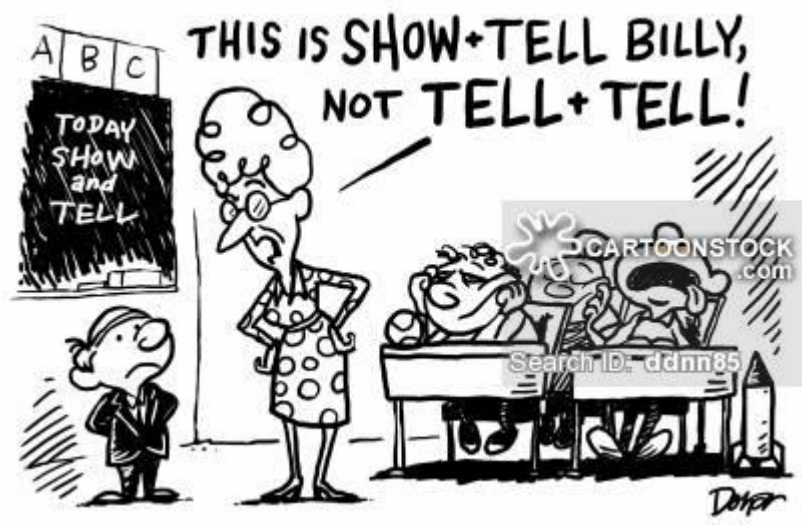
Analyzing Small Molecule Data



Metabolite Association Network



Firmicutes (*Bacillus* spp.) *Micromonospora* spp. *Streptomyces* spp.



“So Chase,
I thought you used Shiny?”

Shiny was created and is currently maintained by Winston Cheng/ RStudio

Maintainer: Winston Chang <winston@rstudio.com>

Winston Chang [aut, cre],
Joe Cheng [aut],
JJ Allaire [aut],
Yihui Xie [aut],
Jonathan McPherson [aut],
RStudio [cph],
jQuery Foundation [cph] (jQuery library and jQuery UI library),
jQuery contributors [ctb, cph] (jQuery library; authors listed in inst/www/shared/jquery-AUTHORS.txt),
jQuery UI contributors [ctb, cph] (jQuery UI library; authors listed in inst/www/shared/jqueryui/AUTHORS.txt),
Mark Otto [ctb] (Bootstrap library),
Jacob Thornton [ctb] (Bootstrap library),
Bootstrap contributors [ctb] (Bootstrap library),
Twitter, Inc [cph] (Bootstrap library),
Alexander Farkas [ctb, cph] (html5shiv library)…

Scott Jehl [ctb, cph] (Respond.js library),
Stefan Petre [ctb, cph] (Bootstrap-datepicker library),
Andrew Rowls [ctb, cph] (Bootstrap-datepicker library),
Dave Gandy [ctb, cph] (Font-Awesome font),
Brian Reavis [ctb, cph] (selectize.js library),
Kristopher Michael Kowal [ctb, cph] (es5-shim library),
es5-shim contributors [ctb, cph] (es5-shim library),
Denis Ineshin [ctb, cph] (ion.rangeSlider library),
Sami Samhuri [ctb, cph] (Javascript strftime library),
SpryMedia Limited [ctb, cph] (DataTables library),
John Fraser [ctb, cph] (showdown.js library),
John Gruber [ctb, cph] (showdown.js library),
Ivan Sagalaev [ctb, cph] (highlight.js library),
R Core Team [ctb, cph] (tar implementation from R)

https://thinkr.fr/wp-content/uploads/meme-shiny-r.jpg

LET'S MAKE YOUR CODE

- Shiny apps are contained in a single script called app.R
- The script app.R lives in a directory and the app can be run with runApp("~/newdir")
- app.R has three components:

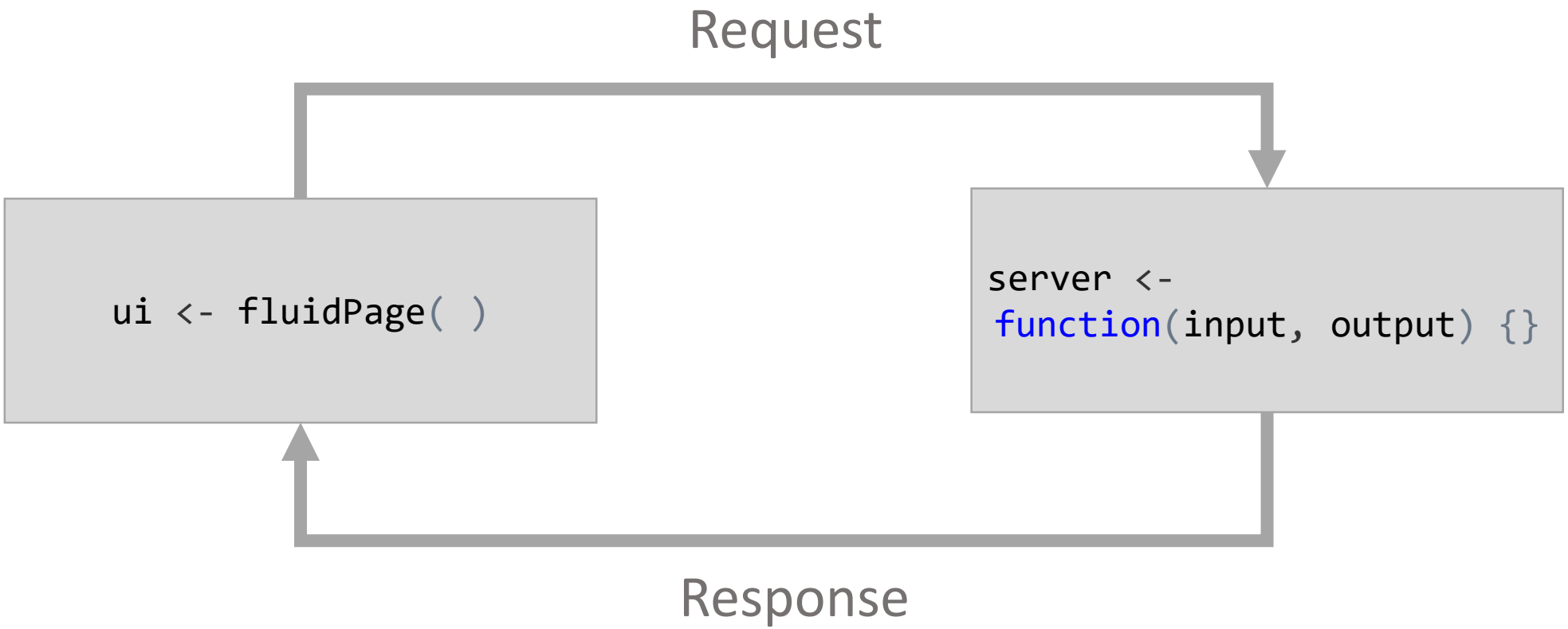
```
# A user interface object---
   ui <- fluidPage( )  #Controls the layout

 # A server logic function ----
   server <- function(input, output) { }

# A call to run the app ----
   shinyApp(ui = ui, server = server)
```

A LITTLE MORE SHINY

https://shiny.rstudio.com/tutorial/

Creating a Shiny App is like making a website.
Requires some understanding of front and backend design

Creating a Shiny App is like making a website.
Requires some understanding of front and backend design



Chase will do a quick walk-through of:

https://shiny.rstudio.com/gallery/kmeans-example.html

Some Tips to Make Coding Shiny Smoother

First Tip = Bad Tip

Unit Tests in Shiny… Not so Easy

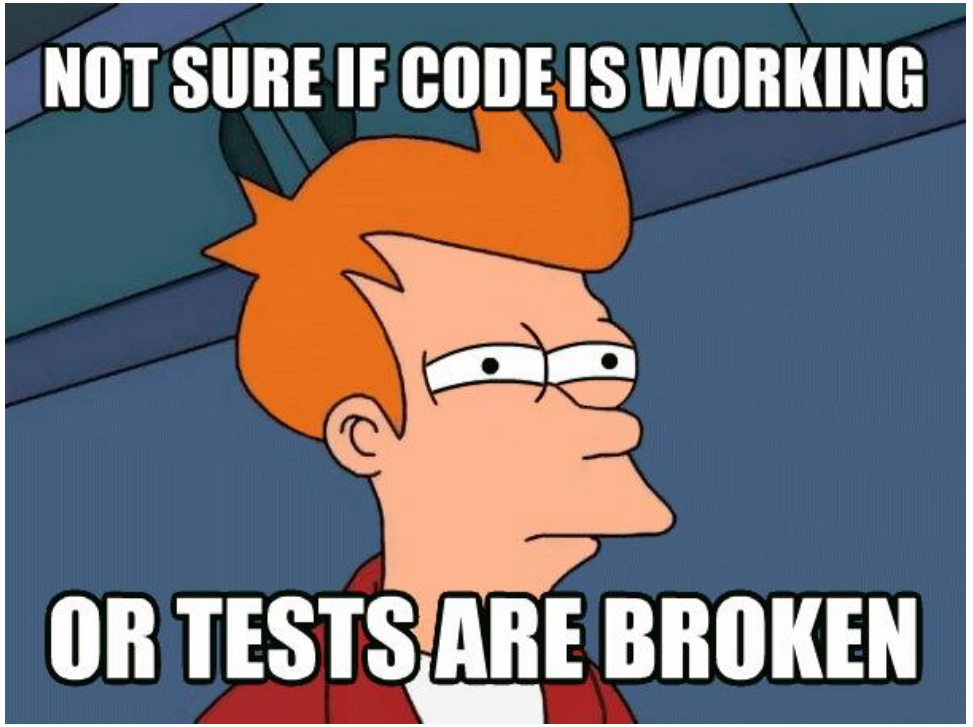
shinytest

> Test Shiny Apps

build failing | build unknown | CRAN not published | downloads 0/month

Installation

```
devtools::install_github("rstudio/shinytest")
```



https://www.rstudio.com/resources/webinars/testing-shiny-applications-with-shinytest/

If you aren't using RStudio, Why?....

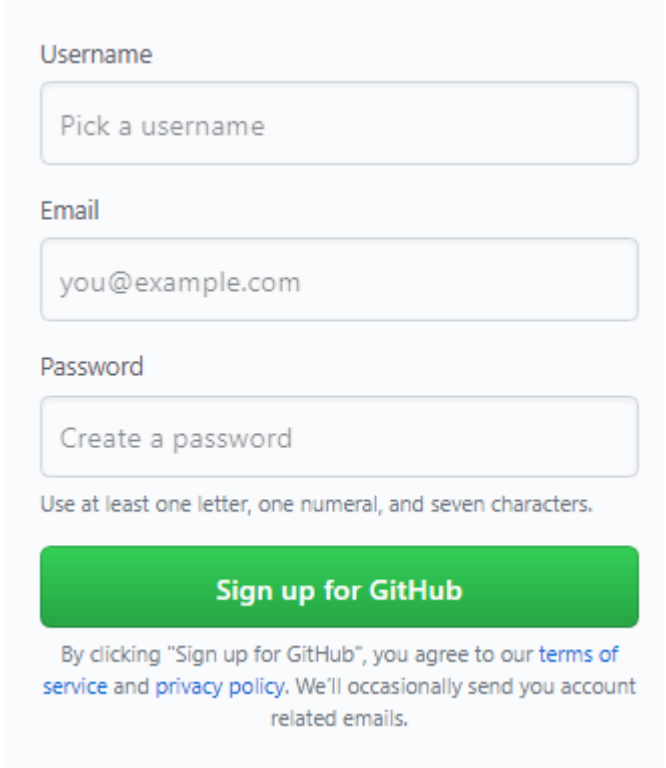
Winston Chang created Shiny…
Winston Chang is a software engineer at RStudio…
Ergo:
RStudio works well with building Shiny apps

RStudio®

Distractions Happen

“Git” GitHub and “Commit” to it

Versioning is essential to reproducibility…and keeping your sanity



= Project Website
= Issue Tracking
= App Distribution

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.

Github +Zenodo =
Automatic DOI Minting for Each Release

- Software:
  - For reproducibility, cite this version of IDBac with the version DOI: "10.5281/zenodo.1115619"
  - Past versions of IDBac can be found here: DOI 10.5281/zenodo.1115620

Tip: Make sure your users have needed packages!

```
# Function to Install and Load R Packages
Install_And_Load <- function(Required_Packages)
{  Remaining_Packages <- Required_Packages[!(Required_Packages %in% installed.packages()[, "Package"])]
 if (length(Remaining_Packages))
 { install.packages(Remaining_Packages)
 }
 for (package_name in Required_Packages)
 {
  library(package_name,
          character.only = TRUE,
          quietly = TRUE)  }}
# Required packages to install and load
Required_Packages = c("parallel","shiny","MALDIquant","MALDIquantForeign","mzR","readxl",
"networkD3","factoextra","ggplot2","ape","FactoMineR","dendextend","networkD3","reshape2","dplyr",
"igraph","rgl")

# Install and Load Packages
Install_And_Load(Required_Packages)
```

https://stackoverflow.com/questions/4090169/elegant-way-to-check-for-missing-packages-and-install-them

Provide Your Users an Option to Install Shiny Apps via an Easy .exe

RInno: Allows Users to Create Shiny Apps as a Shareable .exe Installer

- Not the easiest, but possible.
- Works… but needs work:
  - Continuous installation, devtools::install_github()
  - Package Dependencies (packrat?)

IDBac: Main Page

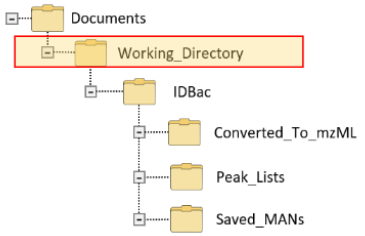
Why make users search for documentation?
- Include directions in-app

Starting with a Single MALDI Plate of Raw Data

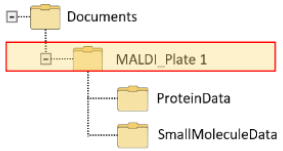
Instructions

1: Working Directory This directs where on your computer you would like to create an IDBac working directory.

In the folder you select, IDBac will create sub-folders within a main directory named "IDBac":



2: Raw Data Your RAW data is a single folder that contains: one subfolder containing protein data and one subfolder containing small-molecule data



*Note: Sometimes the browser window won't pop up, but will still appear in the application bar. See below:

Workflow Pane

1: Working Directory

Click to select your Working Directory

2: Raw Data

Click to select the location of your RAW data

No Folder Selected

3: Choose your Sample Map file, the excel sheet that IDBac will use to rename your files.

Browse... No file selected

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

Convert to mzXML

Note: If you canceled out of the popup after spectra conversion completed, you can process your converted spectra using the button below: (but only after all files have been converted) This step is not necessary otherwise.

Process mzXML spectra

Tips and Tricks

IDBac: Main Page

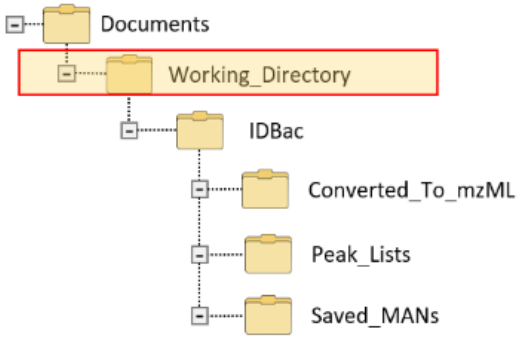
Why make users search for documentation?
- Include directions in-app

Starting with a Single MALDI Plate of Raw Data

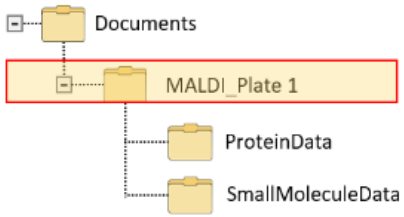
Instructions

1: Working Directory This directs where on your computer you would like to create an IDBac working directory.

In the folder you select, IDBac will create sub-folders within a main directory named "IDBac":



2: Raw Data Your RAW data is a single folder that contains: one subfolder containing protein data and one subfolder containing small-molecule data



*Note: Sometimes the browser window won't pop up, but will still appear in the application bar. See below:

Workflow Pane

1: Working Directory

Click to select your Working Directory

2: Raw Data

Click to select the location of your RAW data

No Folder Selected

3: Choose your Sample Map file, the excel sheet that IDBac will use to rename your files.

Browse... No file selected

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

Convert to mzXML

Note: If you canceled out of the popup after spectra conversion completed, you can process your converted spectra using the button below: (but only after all files have been converted) This step is not necessary otherwise.

Process mzXML spectra

Tips and Tricks

Everyone Dislikes Popups.

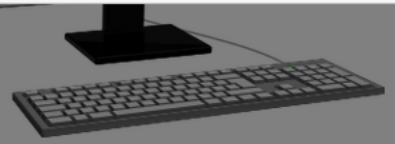
But I recommend their (judicious) use in Shiny

```
popup1<-reactive({
  showModal(modalDialog(
    title = "Important message",
    "When file-conversions are complete this pop-up will be replaced by a summary
     of the conversion.",br(),
    "IDBac uses parallel processing to make these computations faster, unfortunately
    this means we can't show a progress bar.",br(),
    "This also means your computer might be slow during these file conversions.",br(),
    "To check what has been converted you can navigate to:",
    paste0(selectedDirectory(), "/",uniquifiedIDBac(),"/Converted_To_mzML"),
    easyClose = FALSE, size="l",footer=""
  ))
})
```

IDBac

127.0.0.1:3785

Search



Conversion Complete

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

Click to continue with Peak Processing | Close

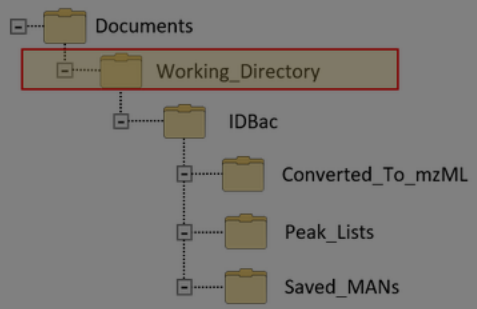
Select here if you have already converted data and just want to re-analyze it

ve already converted
nt to add or subtract

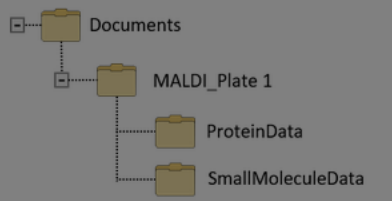
Instructions

1: This directs where on your computer you would like to create an IDBac

In the folder you select- IDBac will create folders within a main directory named "IDBac":



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:



Note: Sometimes the browser window won't pop up, but will still appear in the application bar. See below:

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.

Convert to mzXML

If you canceled out of the popup after spectra conversion completed you can process your converted spectra using the button below:

Process mzXML spectra

Spectra Processing Progress

Scoping Assignment for Debugging

I usually avoid using <<-

But… <<- works great for debugging complicated Shiny apps!

Insert **whatTheHellIsThis <<- shinyReactiveVariable()**

Resources Suggestions for R/Shiny Beginners

New to R

Where to start:

- swirlstats.com
- r4ds.had.co.nz
  - adv-r.had.co.nz

Where to find answers:

- stackexchange.com
- community.rstudio.com
- Web Search! (i.e. google.com)
  - E.g. (search: *convert list to data frame in R*)

New to Shiny

- shiny.rstudio.com/tutorial
- shiny.rstudio.com/gallery
- shiny.rstudio.com/gallery/widget-gallery
- deanattali.com/blog/building-shiny-apps-tutorial/

Active Community on Twitter- #rstats

Suggestions to follow (not inclusive):

@daattali (shiny)
@hadleywickham
@JennyBryan
@drob
@thomasp85
@kierisi
@dataandme

Special Thanks

@ChaseClarkatUIC
murphylabuic.com
@Murphylabuic

Dr. Laura Sanchez

Murphy Lab

Dr. Brian Murphy

Sofia Costa

Maryam Elfeki

Vanessa Nepomuceno

Jeong Ho Lee

Antonio Hernandez

Milan Patel

MEDICINAL CHEMISTRY AND PHARMACOGNOSY
COLLEGE OF PHARMACY

UIC

BRUKER

https://goo.gl/forms/DWnAXBPhHOGtbzHs1


```
library(meme)
  meme("https://imgflip.com/s/meme/X-All-The-Y.jpg", "code all the things!")
```

Contribute

If you find IDBac useful and would like to make it even better, contribute to the code or leave comments/suggestions at:

https://github.com/chasemc/IDBac

Our lab focuses on the most diverse Kingdom of life, bacteria: