Protein MS isn't the only MS…
Programming in R for Metabolomics Mass Spectrometry

Chase Clark
PhD Candidate
Murphy Lab
Dept. of Pharmaceutical Sciences
Center for Biomolecular Sciences
University of Illinois at Chicago

@ChasingMicrobes
chasemc.github.io

May Institute Future Developers Meeting

~~Protein MS isn't the only MS…~~

~~Programming in R for Metabolomics~~

~~Mass Spectrometry~~

Programming in R for Mass-omics

Chase Clark
PhD Candidate
Murphy Lab
Dept. of Pharmaceutical Sciences
Center for Biomolecular Sciences
University of Illinois at Chicago

@ChasingMicrobes

chasemc.github.io

May Institute Future Developers Meeting

RforMassSpectrometry - Home - Mozilla Firefox

RforMassSpectrometry

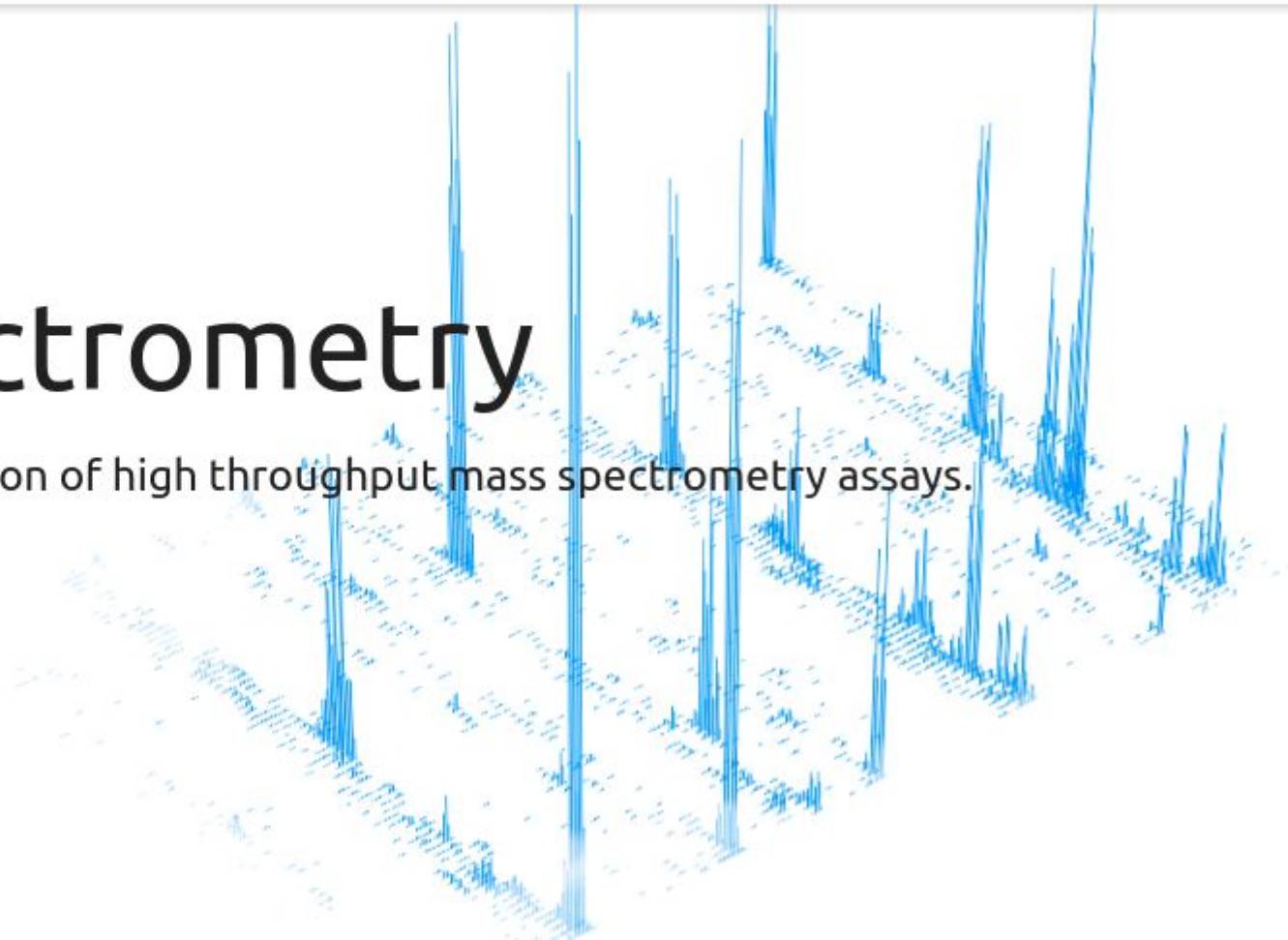
https://www.rformassspectrometry.org

Home Packages Contact

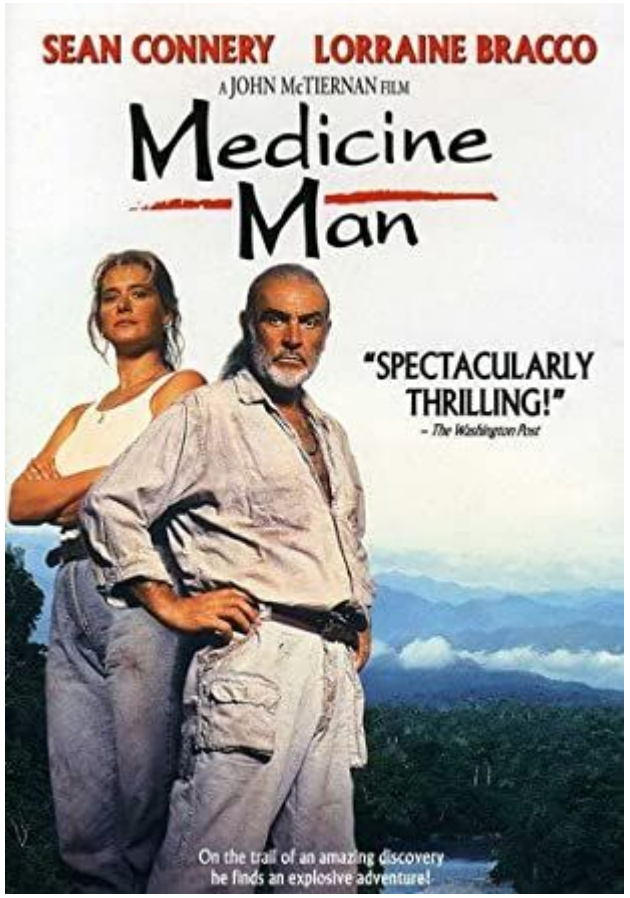
R for Mass Spectrometry

R software for the analysis and interpretation of high throughput mass spectrometry assays.

Contact



Pharmacognosy



Pharmacognosy

The study of the physical, chemical, biochemical and biological properties of drugs, drug substances or potential drugs of natural origin as well as the search for new drugs from natural sources.

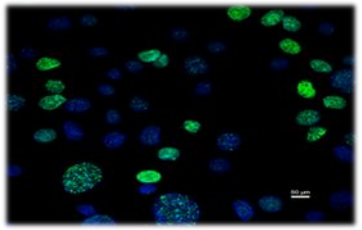
https://blog.sciencemuseum.org.uk/superbugs-at-the-science-museum

https://youtu.be/KoIDErEhaAQ

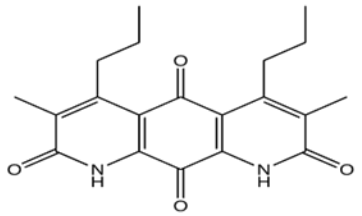
Discovery Pipeline (Simplified)

Collect Samples

Grow Bacteria



Antibiotic Activity?



Elucidate Antibiotic

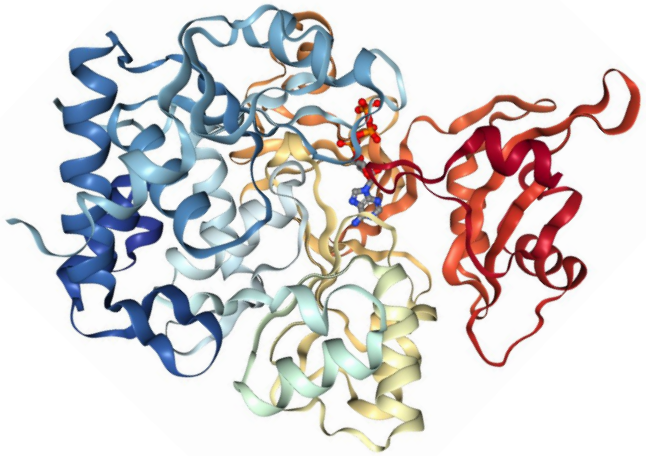
Drug

Which bacteria do we study?

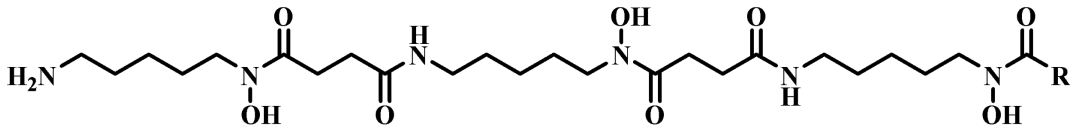
Dr. Laura Sanchez

Collect Two Fingerprints Per Colony

Protein Spectra
2-20 kDa
Pseudo-Phylogenetic Groupings



Small Molecule Spectra
0.2-2 kDa
Measuring Produced Metabolites



Domain Knowledge Existed in R

Species Identification using MALDIquant

Sebastian Gibb* and Korbinian Strimmer †

June 8, 2015

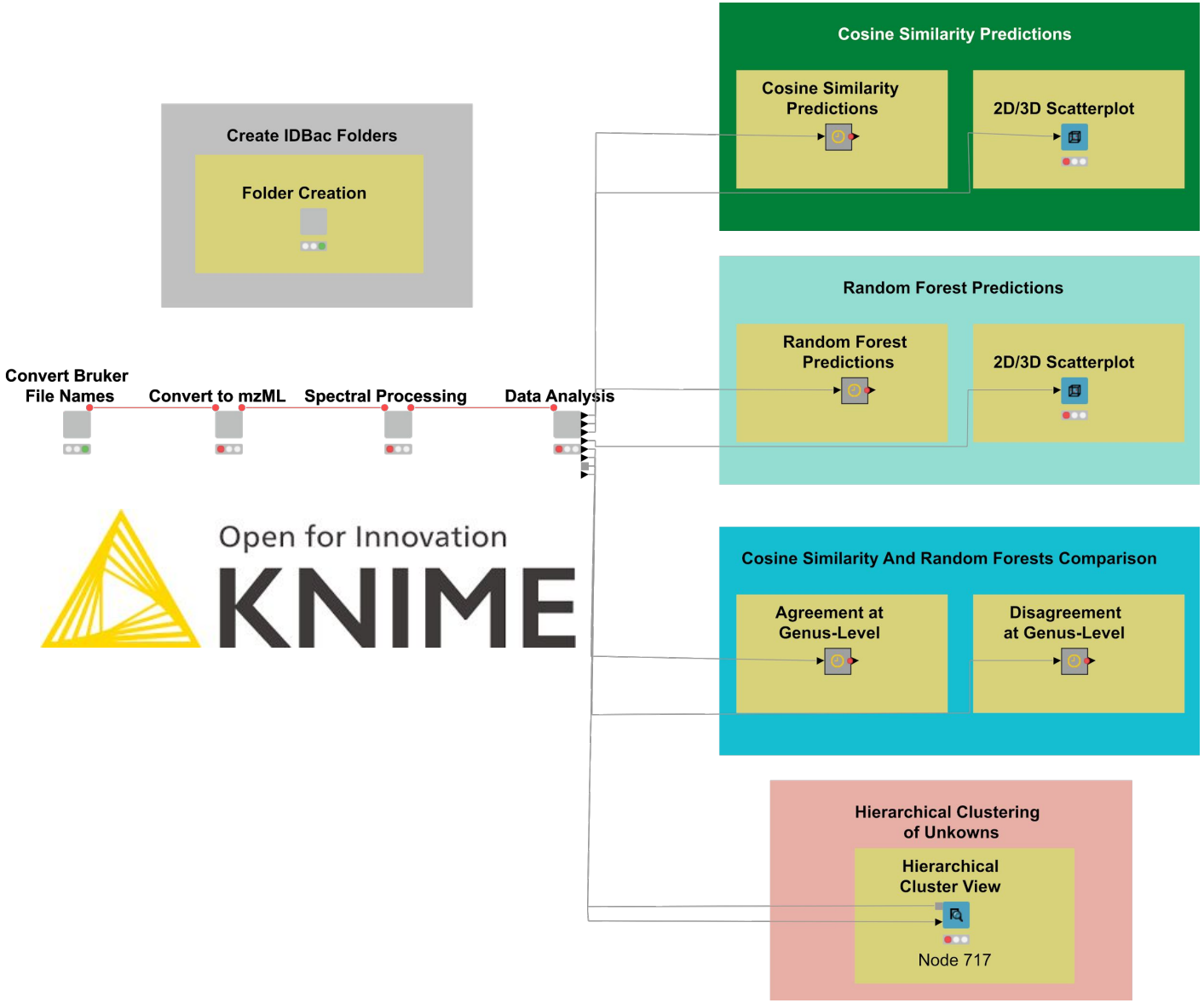
Abstract

This vignette describes how to use MALDIquant for species identification.

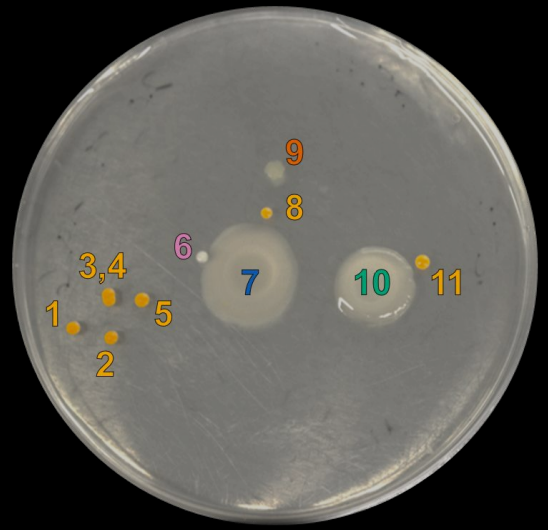
But I didn't know to code!

IDBac

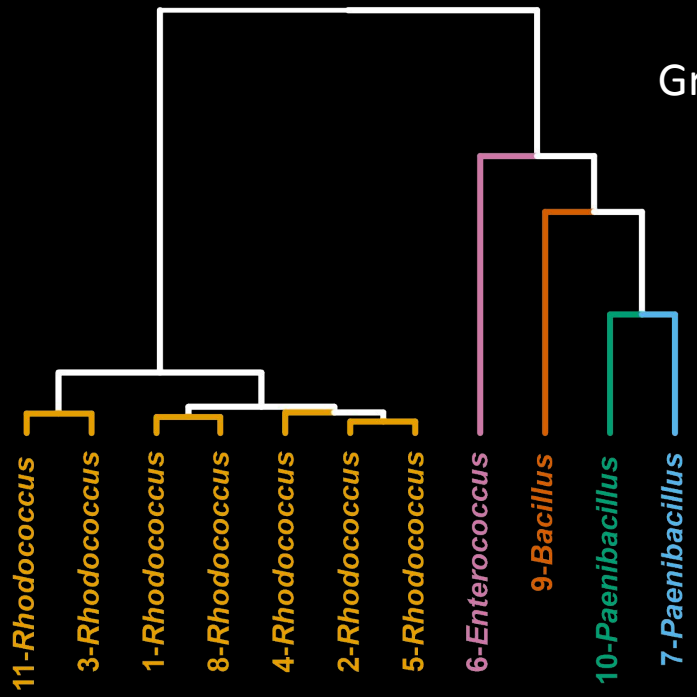
Contents

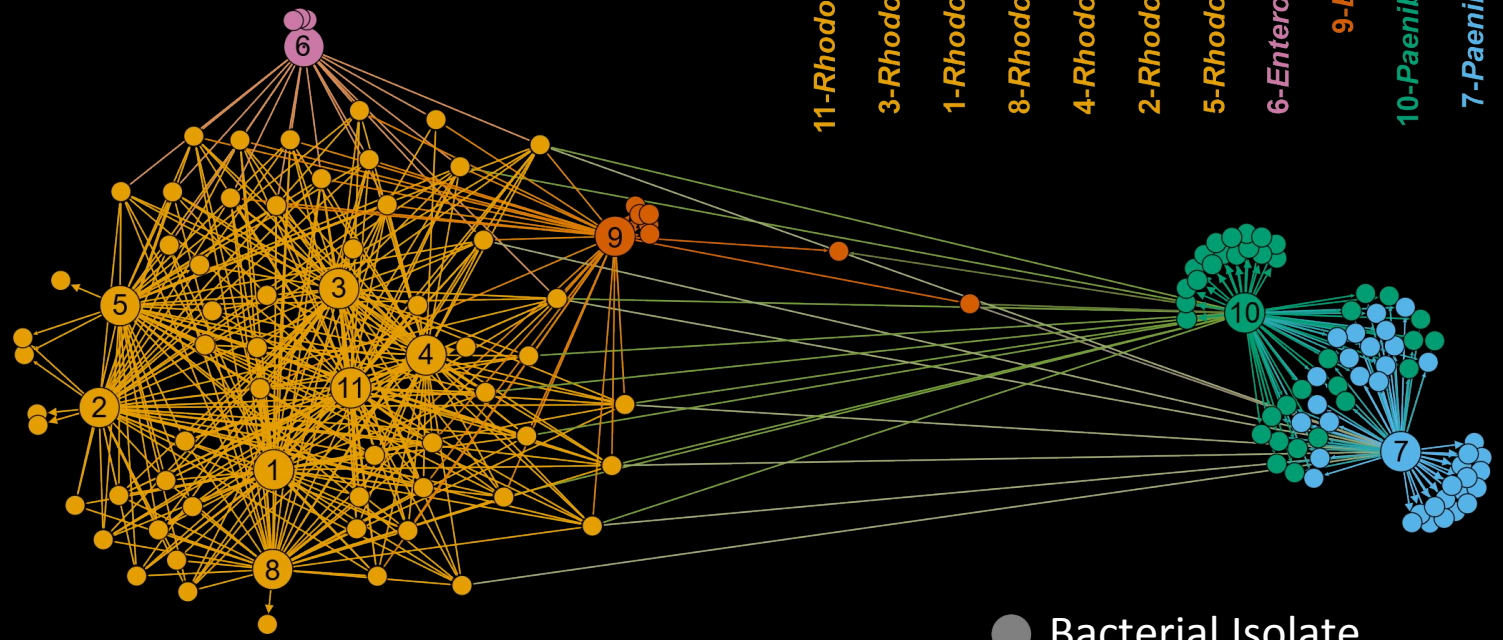


IDBac



Grouped using protein spectra





Visualized small molecule data

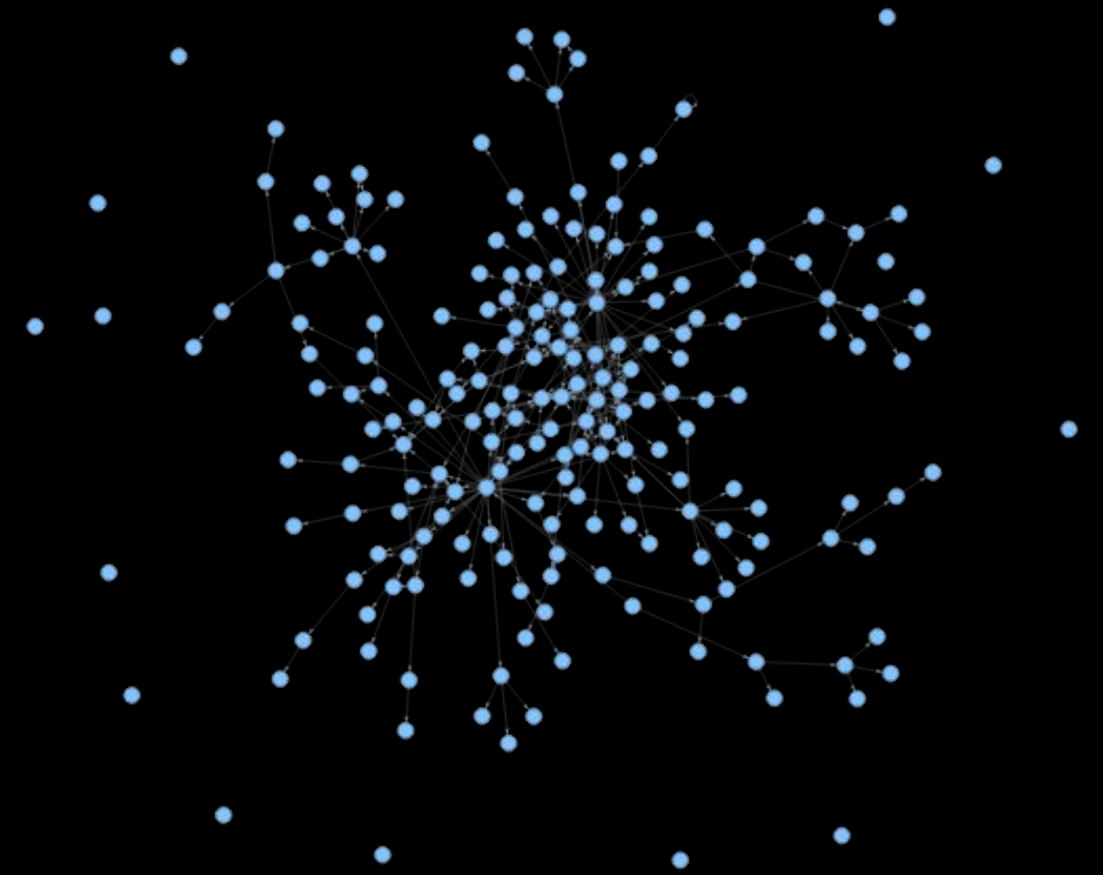
Bacterial Isolate

m/z Peak

```
Package: IDBacApp
Type: Package
Title: MALDI-TOF MS Protien and Small Molecule Analysis
Version: 1.2.0.9000
Author: Chase Clark
Maintainer: Chase Clark [redacted]
Description: This app allows users to take raw data MALDI-TOF MS files
    and create bundled mzXML files for easy sharing/database creation as well as data analysis
    (hierarchical clustering, PCA, Spectra Comparison, Metabolite Association Network).
License: GPL-3
Encoding: UTF-8
LazyData: true
biocViews:
Imports:
    ape (>= 5.3),
    colourpicker (>= 1.0),
    coop (>= 0.6-2),
    data.table (>= 1.12.2),
    DBI (>= 1.0.0),
    digest (>= 0.6.20),
    fst (>= 0.9.0),
    ggplot2 (>= 3.2.0),
    glue (>= 1.3.1),
    graphics,
    grDevices,
    httr (>= 1.4.1),
    igraph (>= 1.2.4.1),
    irlba (>= 2.3.3),
    jsonlite (>= 1.6),
    magrittr (>= 1.5),
    MALDIquant (>= 1.19.3),
    MALDIquantForeign (>= 0.12),
    mzR,
    networkD3 (>= 0.4),
    plotly (>= 4.9.0),
    pool (>= 0.1.4.2),
    remotes,
    reshape2 (>= 1.4.3),
    rhandsontable (>= 0.3.7),
    rmarkdown (>= 1.14),
    RSQLite (>= 2.1.2),
    Rtsne (>= 0.15),
    S4Vectors,
    shinycssloaders (>= 0.2.0),
    sigmajs (>= 0.1.3),
    stats,
    svglite (>= 1.2.2),
    utils,
    Matrix,
    dendextend (>= 1.12.0),
    shiny (>= 1.3.2),
    curl
Suggests:
    knitr (>= 1.23),
    testthat (>= 2.2.1)
RoxygenNote: 7.1.0
VignetteBuilder: knitr
Roxygen: list(markdown = TRUE)
```

```
Loading IDBacApp
Testing IDBacApp
✓ |  OK F W S | Context
✓ |   1       |
✓ |  18       | 00_create_database [7.5 s]
✓ |   6       | 01_create_database [1.2 s]
✓ |   5       | test-createFuzzyVector
✓ |   1       | test-bootstrap [0.8 s]
✓ |   1       | brukerToMzml_popup
✓ |   1       | test-colored_dots [0.3 s]
✓ |   1       | test-colorpalette
✓ |   0     1 | controlBrukerDisplay

test-controlBrukerDisplay.R:18: skip: controlBrukerDisplay works
Reason: On Linux

✓ |   2       | copy_db
✓ |   1       | db_path_from_pool
✓ |   1       | distMatrix
✓ |   2       | findMSconvert
✓ |   6       | getMicrotyperFiles
✓ |   4       | getPeakData
✓ |   1       | test-hashr
✓ |   3       | test-insertlocale
✓ |   1       | test-labelsfrombrusheddendrogram
✓ |   9       | test-map384well
✓ |   3       | mirrorplots
✓ |   1       | test-module_databasetab
✓ |   1       | test-module_tsne
✓ |   4       | test-networkfromdf
✓ |   7       | parseDelimitedMS
✓ |   1       | test-pathsanitize
✓ |   1       | test-plotly_3d_scatter
✓ |   1       | test-runApp
✓ |   1       | test-samplemapviewer
✓ |   3       | test-spectrummatrixtomaldiqaunt
✓ |  12       | sqlTableArchitecture
✓ |   3       | subtree
✓ |   6       | test-utils

== Results =====================================
Duration: 12.1 s

OK:       108
Failed:   0
Warnings: 1
Skipped:  1
```

IDBac



Network of internal/external
IDBac functions

Programming with IDBac

127.0.0.1:4984

Programming with IDBac

Chase Clark

2020-02-15

Chapter 1 Preamble

While some familiarity with R is suggested, the exampl
enough for a novice to comfortable working through.

Suggestions or additions for content are welcome and
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 con
- 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
ftp://massive.ucsd.edu/MSV000084291

```
library(here)
```

Chapter 6 Simple Analysis | Prog

127.0.0.1:4984/simple-analysis.html

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

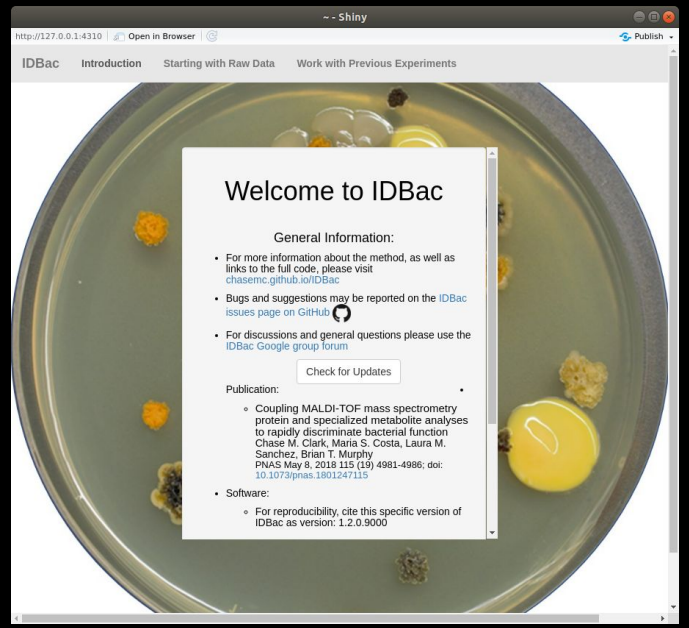
```
example_pool <- IDBacApp::idbac_connect(fileName = "idbac_experiment_file",
                                        filePath = here::here("data",
                                                              "example_data"))
```

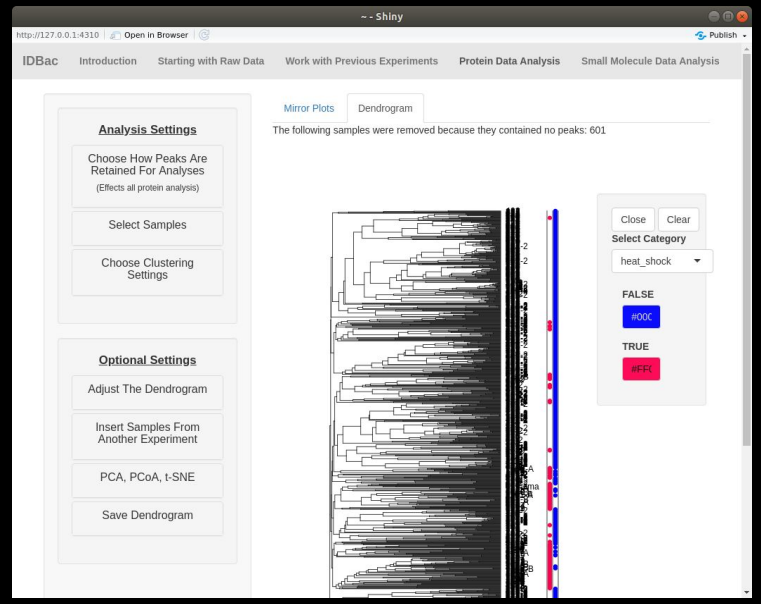
```
my_plot <- IDBacApp:::assembleMirrorPlots(sampleID1 = "172-7",
                                          sampleID2 = "172-10",
                                          peakPercentPresence = 0.7,
                                          lowerMassCutoff = 3000,
                                  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

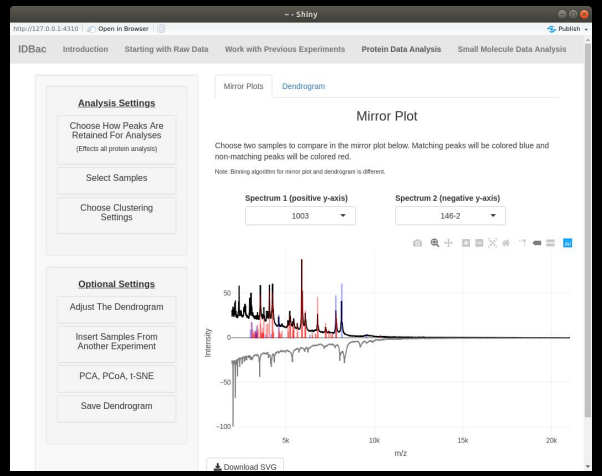
IDBac

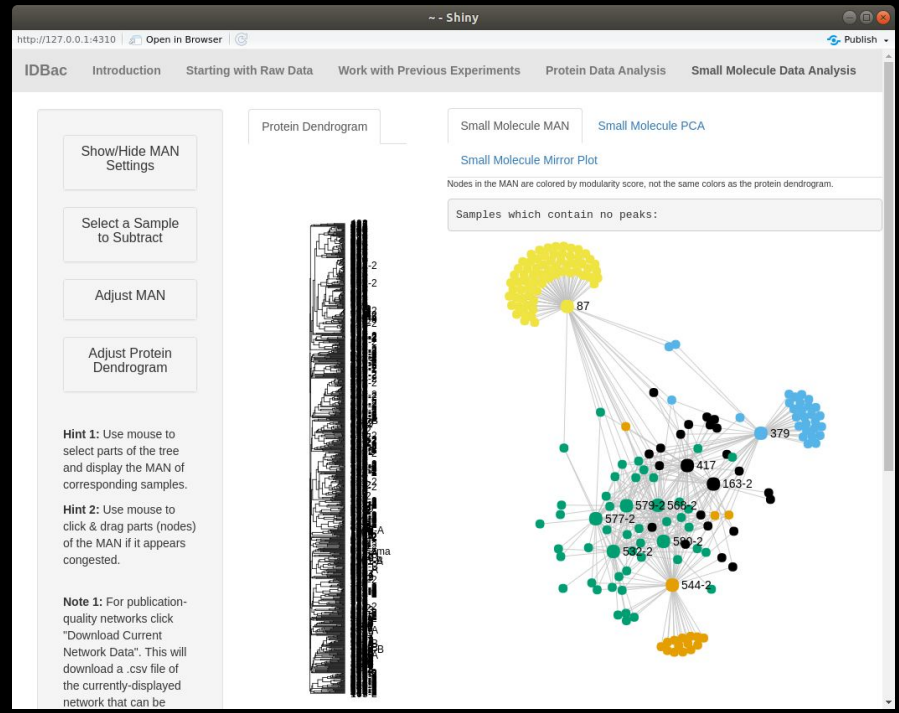
MALDIquant
mzR

SQLite

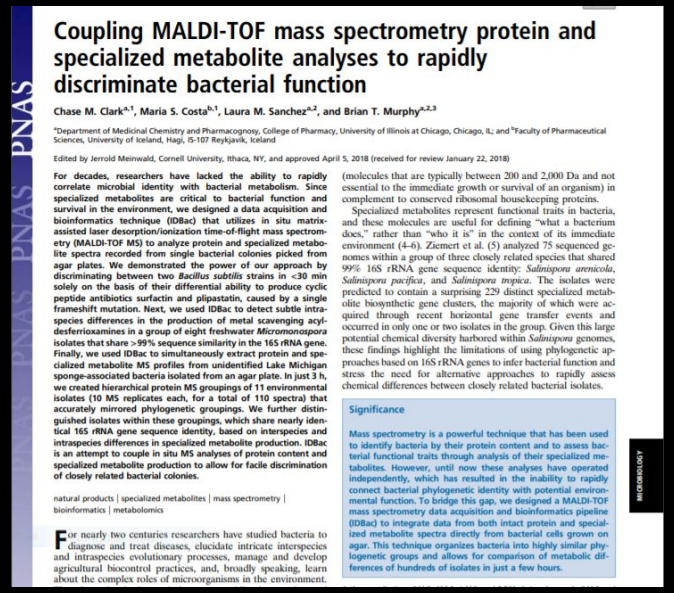








IDBac Publications



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]

[a]Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL; and [b]Faculty of Pharmaceutical Sciences, University of Iceland, Hagi, IS-107 Reykjavik, Iceland

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

For decades, researchers have lacked the ability to rapidly correlate 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 solely on the basis of their differential ability to produce cyclic peptide antibiotics surfactin and plipastatin, caused by a single frameshift mutation. Next, we used IDBac to detect subtle intraspecies differences in the production of metal scavenging acyldesferrioxamines in a group of eight freshwater *Micromonospora* isolates that share >99% sequence similarity in the 16S rRNA gene. Finally, we used IDBac to simultaneously extract protein and specialized metabolite MS profiles from unidentified Lake Michigan sponge-associated bacteria isolated from an agar plate. In just 3 h, we created hierarchical protein MS groupings of 11 environmental isolates (10 MS replicates each, for a total of 110 spectra) that accurately mirrored phylogenetic groupings. We further distinguished isolates within these groupings, which share nearly identical 16S rRNA gene sequence identity, based on interspecies and intraspecies differences in specialized metabolite production. IDBac is an attempt to couple in situ MS analyses of protein content and specialized metabolite production to allow for facile discrimination of closely related bacterial colonies.

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

For nearly two centuries researchers have studied bacteria to diagnose and treat diseases, elucidate intricate interspecies and intraspecies evolutionary processes, manage and develop agricultural biocontrol practices, and, broadly speaking, learn about the complex roles of microorganisms in the environment. (molecules that are typically between 200 and 2,000 Da and not essential to the immediate growth or survival of an organism) in complement to conserved ribosomal housekeeping proteins.

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 (4–6). Ziemert et al. (5) analyzed 75 sequenced genomes within a group of three closely related species that shared 99% 16S rRNA gene sequence identity: *Salinispora arenicola*, *Salinispora pacifica*, and *Salinispora tropica*. The isolates were predicted to contain a surprising 229 distinct specialized metabolite biosynthetic gene clusters, the majority of which were acquired through recent horizontal gene transfer events and occurred in only one or two isolates in the group. Given this large potential chemical diversity harbored within *Salinispora* 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 functional traits through 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 data acquisition and 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.

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

AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

EDITORIAL

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

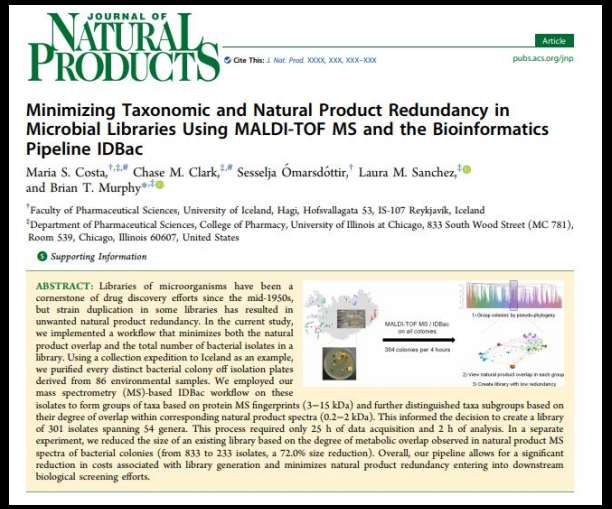
Chase M. Clark,[a] Brian T. Murphy,[a] Laura M. Sanchez[a]

[a]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



JOURNAL OF NATURAL PRODUCTS

Article

Cite This: J. Nat. Prod. XXXX, XXX, XXX−XXX

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,[†,‡,#] Chase M. Clark,[‡,#] Sesselja Ómarsdóttir,[†] Laura M. Sanchez,[‡] and Brian T. Murphy*[,‡]

[†]Faculty of Pharmaceutical Sciences, University of Iceland, Hagi, Hofsvallagata 53, IS-107 Reykjavik, Iceland

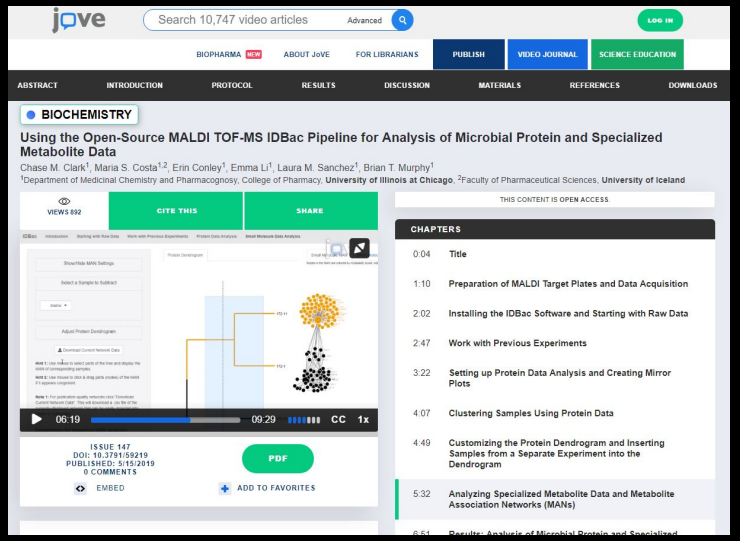
[‡]Department of Pharmaceutical Sciences, College of Pharmacy, University of Illinois at Chicago, 833 South Wood Street (MC 781), Room 539, Chicago, Illinois 60607, United States

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 expedition to Iceland as an example, we purified every distinct bacterial colony off isolation plates 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 only 25 h of data acquisition and 2 h of analysis. In a separate experiment, we reduced the size of an existing library based on the degree of metabolic overlap observed in natural product 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

https://chasemc.github.io/IDBac



jove | Search 10,747 video articles | Advanced | LOG IN

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[1], Maria S. Costa[1,2], Erin Conley[1], Emma Li[1], Laura M. Sanchez[1], Brian T. Murphy[1]

[1]Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago. [2]Faculty of Pharmaceutical Sciences, University of Iceland

VIEWS 892 | CITE THIS | SHARE

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)

06:19 | 09:29 | CC | 1x

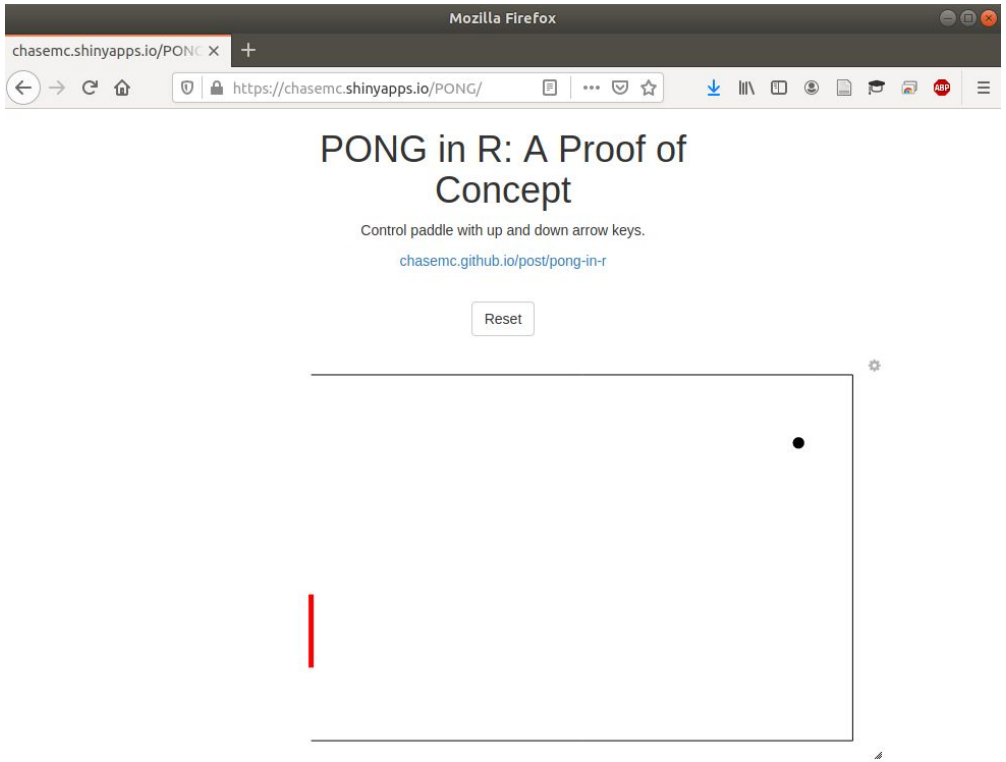
ISSUE 147
DOI: 10.3791/59219
PUBLISHED: 5/15/2019
0 COMMENTS

PDF | EMBED | ADD TO FAVORITES

https://doi.org/ 10.3791/59219

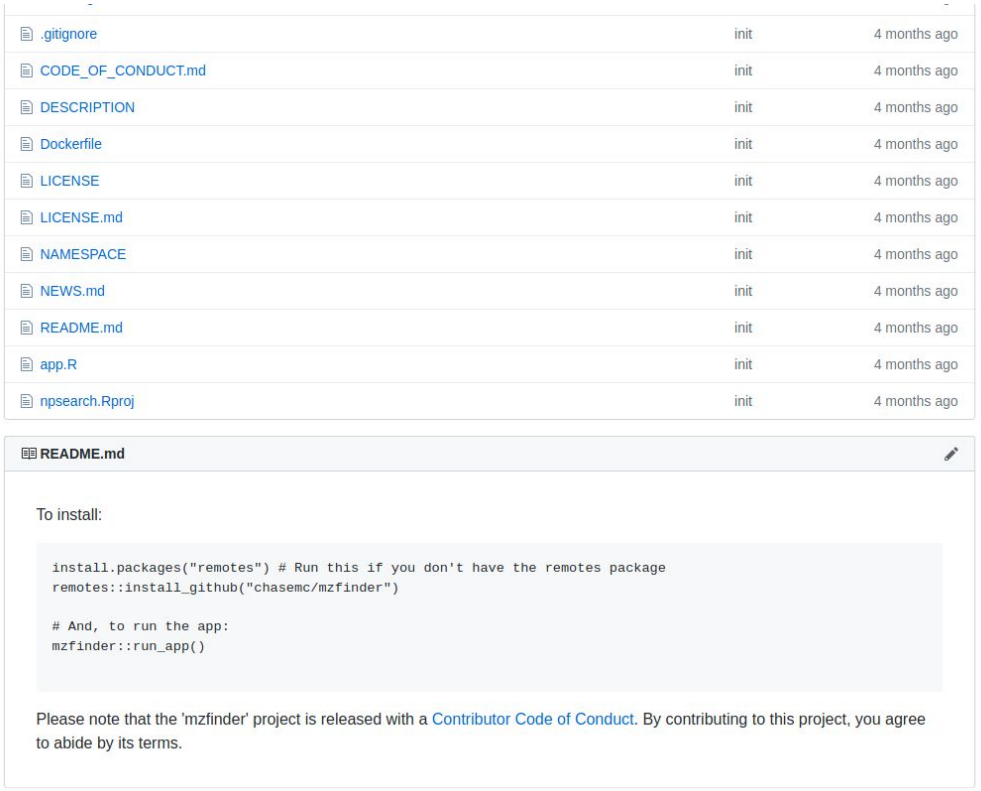
How will users access your software?

Web Deployment (Server)



https://chasemc.shinyapps.io/PONG/

Local GUI



https://github.com/chasemc/mzfinder

How will users access your software?

Web Deployment

large files + long computations

How will users access your software?

Web Deployment

large files + long computations

Local Installation

Large percent of users won't even attempt
Local technical support

{electricShine}
Installable Shiny/Electron Apps

- lifecycle experimental

Windows CI:

- build passing

Mac and Linux CI

- build passing

Is easy! One meta-function:

```
electricShine::buildElectricApp(
app_name = "My_App",
description = "My demo application",
package_name = "demoApp",
semantic_version = "1.0.0",
build_path = buildPath,
mran_date = MRANdate,
function_name = "run_app",
github_repo = "chasemc/demoApp",
local_path  = NULL

)
```

Self-Sufficient

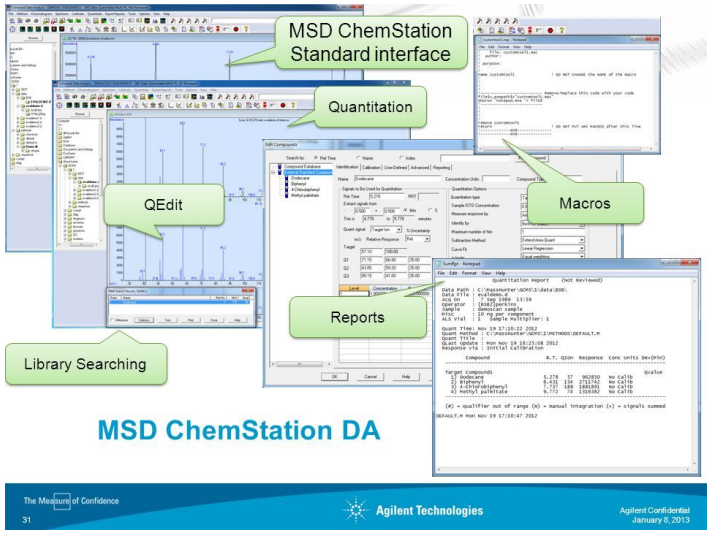
- R- from specified MRAN date
- R packages- from specified MRAN date
- Your Shiny app!

- Currently: Windows/Mac
- Future: Linux
- Compatible with continuous-deployment

https://github.com/chasemc/electricShine

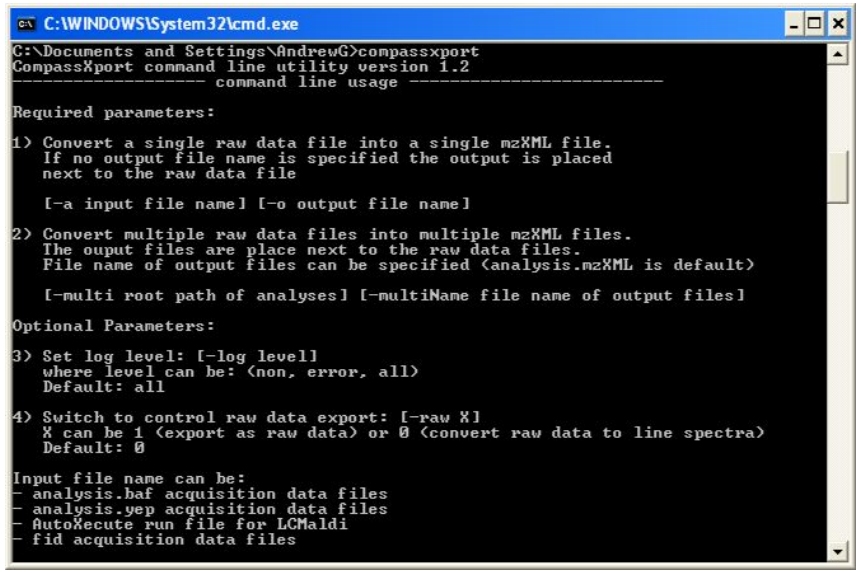
The many faces of mass spectrometry software

GUI



Server

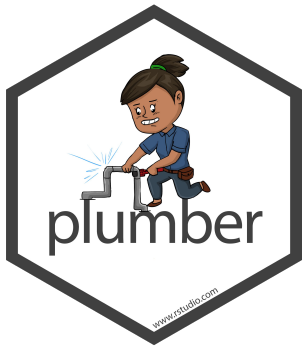
Command line



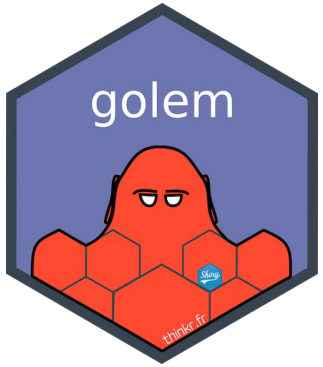
The many faces of mass spectrometry software

GUI ✓

Server ✓



{plumber} converts your existing R code to a web API using a handful of special one-line comments.



{golem} is an opinionated framework for building production-grade shiny applications.

Command line ✓

@ChasingMicrobes
chasemc.github.io

MEDICINAL CHEMISTRY AND PHARMACOGNOSY
COLLEGE OF PHARMACY
UIC

Funding:

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

NATIONAL GEOGRAPHIC
Grant CP-044R-17

Murphy Lab
Current Grad and Post-Doc Members:

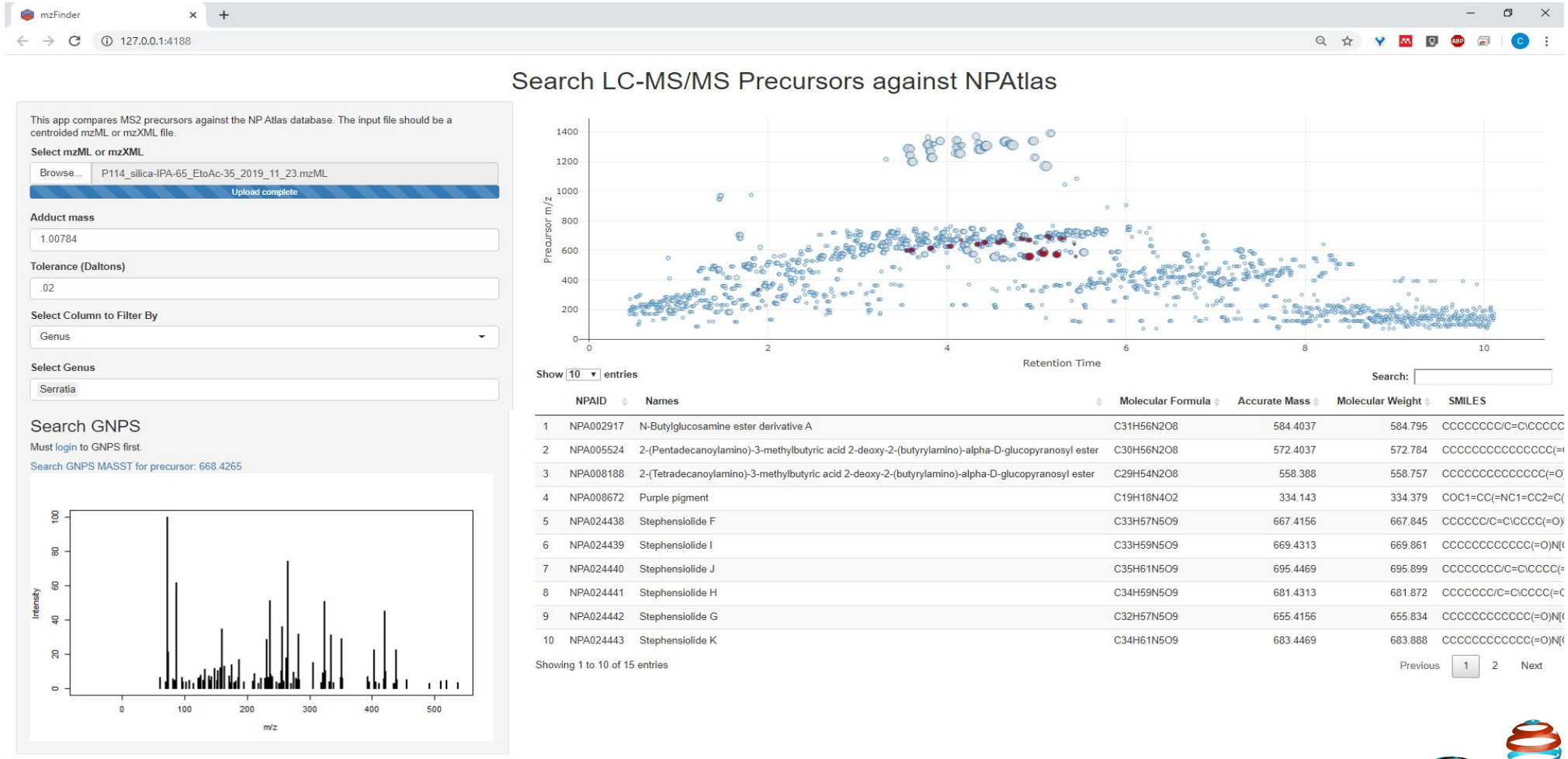
Dr. Brian Murphy
Chase Clark
Antonio Hernandez
Dr. Jeongho Lee
Dr. Tuan Anh Tran
Dr. Linh Nguyen
Maryam Elfeki

@ChasingMicrobes

chasemc.github.io

github.com/chasemc/presentations/may_institute/2020

Is R Fast Enough? Usually Yes



https://github.com/chasemc/mzfinder

golem

data.table

npatlas