Journal of Visualized Experiments

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

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE59219R1
Full Title:	Using the Open-Source MALDI TOF-MS IDBac Pipeline for Analysis of Microbial Protein and Specialized Metabolite Data
Keywords:	mass spectrometry; microbiology; natural products; specialized metabolites; MALDI-TOF; bioinformatics
Corresponding Author:	Brian T Murphy University of Illinois Chicago Chicago, IL UNITED STATES
Corresponding Author's Institution:	University of Illinois Chicago
Corresponding Author E-Mail:	btmurphy@uic.edu
Order of Authors:	Chase M Clark
	Maria Sofia Costa
	Erin Conley
	Emma Li
	Laura M. Sanchez
	Brian T Murphy
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Open Access (US$4,200)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Chicago, IL, USA

TITLE:

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

AUTHORS AND AFFILIATIONS:

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

[1]Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL

[2]Faculty of Pharmaceutical Sciences, University of Iceland, Hagi, IS-107 Reykjavík, Iceland

Corresponding Author:

Brian T. Murphy (btmurphy@uic.edu

Laura M. Sanchez (sanchelm@uic.edu)

Email Addresses of Co-authors:

Chase M. Clark (cclark42@uic.edu)

Maria S. Costa (marysofs@gmail.com)

Erin Conley (econle3@uic.edu)

Emma Li (eli8@uic.edu)

KEYWORDS:

Mass spectrometry, microbiology, natural products, specialized metabolites, MALDI-TOF MS, bioinformatics

SUMMARY:

IDBac is an open-source mass spectrometry-based bioinformatics pipeline that integrates data from both intact protein and specialized metabolite spectra, collected on cell material scraped from bacterial colonies. The pipeline allows researchers to rapidly organize hundreds to thousands of bacterial colonies into putative taxonomic groups, and further differentiate them based on specialized metabolite production.

ABSTRACT:

In order to visualize the relationship between bacterial phylogeny and specialized metabolite production of bacterial colonies growing on nutrient agar, we developed IDBac—a low-cost and high-throughput matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) bioinformatics pipeline. IDBac software is designed for non-experts, is freely available, and capable of analyzing a few to thousands of bacterial colonies. Here, we present procedures for the preparation of bacterial colonies for MALDI-TOF MS analysis, MS instrument operation, and data processing and visualization in IDBac. In particular, we instruct users how to cluster bacteria into dendrograms based on protein MS fingerprints and interactively create Metabolite Association Networks (MANs) from specialized metabolite data.

INTRODUCTION:

A major barrier to researchers who study bacterial function is the ability to quickly and simultaneously assess the taxonomic identity of a microorganism and its capacity to produce specialized metabolites. This has prevented significant advances in understanding the relationship between bacterial phylogeny and specialized metabolite production in the majority of bacteria isolated from the environment. Although MS-based methods that use protein fingerprints to group and identify bacteria are well described[1–4], these studies have generally been performed on small groups of isolates, in a species-specific manner. Importantly, information on specialized metabolite production, a major driver of microbial function in the environment, has remained unincorporated in these studies. Silva et al.[5] recently provided a comprehensive history detailing the underuse of MALDI-TOF MS to analyze specialized metabolites and the shortage of software to relieve current bioinformatics bottlenecks. In order to address these shortcomings, we created IDBac, a bioinformatics pipeline that integrates both linear and reflectron modes of MALDI-TOF MS[6]. This allows users to rapidly visualize and differentiate bacterial isolates based on both protein and specialized metabolite MS fingerprints, respectively.

IDBac is cost-effective, high-throughput, and designed for the lay user. It is freely available (chasemc.github.io/IDBac), and only requires access to a MALDI-TOF mass spectrometer (reflectron mode will be required for specialized metabolite analysis). Sample preparation relies on the simple "extended direct transfer" method[7,8] and data are collected with consecutive linear and reflectron acquisitions on a single MALDI-target spot. With IDBac, it is possible to analyze the putative phylogeny and specialized metabolite production of hundreds of colonies in under four hours, including sample preparation, data acquisition, and data visualization. This presents a significant time and cost advantage over traditional methods of identifying bacteria (such as gene sequencing), and analyzing metabolic output (liquid chromatography–mass spectrometry [LCMS] and similar chromatographic methods).

Using data obtained in linear mode analysis, IDBac employs hierarchical clustering to represent the relatedness of protein spectra. Since the spectra mostly represent ionized ribosomal proteins, they provide a representation of the phylogenetic diversity present in a sample. In addition, IDBac incorporates reflectron mode data to display specialized metabolite fingerprints as Metabolite Association Networks (MANs). MANs are bipartite networks that allow for easy visualization of shared and unique metabolite production between bacterial isolates. The IDBac platform allows researchers to analyze both protein and specialized metabolite data in tandem but also individually if only one data-type is acquired. Importantly, IDBac processes raw data from Bruker and Xiamen instruments, as well as txt, tab, csv, mzXML, and mzML. This eliminates the need for manual conversion and formatting of data sets, and significantly reduces the risk of user error or mishandling of MS data.

PROTOCOL:

1. Preparation of MALDI matrix

1.1. Prepare 10 mg/mL MALDI-grade, and/or recrystallized α-cyano-4-hydroxycinnamic acid (CHCA) in MS-grade solvents: 50% acetonitrile (ACN), 47.5% water (H_2O), 2.5% trifluoroacetic acid (TFA). Example: 100 µL solution = 50 µL ACN + 47.5 µL H_2O + 2.5 µL TFA + 1 mg CHCA

1.1.1. Prepare at least 1 µL of matrix solution per MALDI plate spot and vortex or sonicate until in solution (approximately 5 min sonication or no visible solids).

CAUTION: TFA is a strong acid that should be handled in a chemical fume hood while wearing proper personal protective equipment, as it can damage skin, eyes, and airways with contact or inhalation.

NOTE: CHCA is hygroscopic and light-sensitive and should be stored in amber vials in a desiccator. There are many MALDI matrix options available. CHCA is most common for protein profiling of bacteria, but also works for specialized metabolite analysis. Matrix selection depends on individual user/experiment needs.

2. Preparation of MALDI target plates

NOTE: See Sauer et al.[7], for more details.

2.1. Rinse MALDI plate with methanol (HPLC-grade or higher) and wipe dry with soft paper wipes. Do not use abrasive brushes when cleaning target plates, as this can permanently damage the surface of the target plate.

2.2 Assign protein and specialized metabolite calibrant spots. Organize calibration spots evenly across the sample population, to account for MALDI-plate-irregularities and instrument drift over time. Assign an appropriate number of media/matrix-blank spots for the study; these spots will contain only media and matrix, or only matrix.

2.3. Using a sterile toothpick, transfer a small portion of a bacterial colony to the appropriate spot on the MALDI plate. Spread the bacterial colony evenly over the spot. The spot should appear as flat as possible.

NOTE: It will be easier to flatten bacterial colonies that are more mucoid/amorphous. For more rigid/solid colonies, avoid leaving visible clusters of cell mass on the MALDI spot (**Figure 1**).

[Insert Figure 1 here]

2.4. Prepare a matrix/media control by using a sterile toothpick to transfer a minimal amount of agar/media onto the appropriate spot(s) on the MALDI plate.

2.5. Overlay 1 µL of 70% mass spectrometry grade formic acid onto each sample spot, including the matrix control spots. Allow acid to air dry completely in a chemical fume hood (approximately 5 min).

CAUTION: Formic acid is a caustic chemical and should be handled in chemical fume hoods. It can damage airways if inhaled.

2.6. Add 1 µL of the prepared MALDI matrix solution to each sample spot, as well as to the matrix/media control spots. Allow matrix solution to air dry completely (approximately 5 min).

NOTE: It is possible to store the plate in a desiccator, in the dark, until it can be analyzed on a MALDI-TOF mass spectrometer. Allowable storage times may vary depending on sample stability.

2.7. Add 0.5–1.0 µL calibrant to the assigned calibration spots, followed by 1 µL MALDI matrix solution. Pipette the resulting solution up and down to mix. Allow all spots to air dry completely prior to introduction into the MALDI-TOF mass spectrometer.

NOTE: The protein and specialized metabolite calibrants should be added within 30 min of MALDI analysis, as both are susceptible to degradation.

3. Data acquisition

NOTE: The general parameters for data acquisition are listed in **Table 1**.

[Insert Table 1]

3.1. Following the protocols specific to the instrument being used, set up both protein and specialized metabolite calibrations.

3.2. Test a few separate target spots to determine the optimal laser power and detector gain to use when acquiring spectra (this will vary day-to-day and by instrument).

NOTE: **Figure 2A** and **Figure 3A** show optimal spectra, while **Figure 2D** and **Figure 3D** are examples of poor-quality spectra.

[Insert Figures 2 and 3 here]

3.4. Acquire spectra, saving protein spectra into one folder and specialized metabolite spectra into a second, separate folder.

4. Cleaning the MALDI target plate (adapted from Sauer et al.[7])

4.1. Remove the MALDI target plate from its holder and rinse with acetone.

4.2. Wash with a non-abrasive liquid soap to remove trace proteins and lipids, and soft paper wipes/soft-bristled toothbrush.

4.3. Rinse with de-ionized water for approximately 2 min to completely remove soap.

4.4. Sonicate the target plate in water (HPLC grade or higher) for 5 min.

4.5. Rinse the target plate with water (HPLC grade or higher).

4.6. Rinse the target plate with methanol (HPLC grade or higher).

5. Installing the IDBac Software

5.1. Download the IDBac software.

NOTE: Permanent, versioned backups are also available for download (see the **Table of Materials**).

5.2. Double-click the downloaded "Install_IDBac.exe" to initiate the installer and follow the on-screen instructions.

6. Starting with Raw Data

NOTE: Detailed explanations and instructions of each data processing step are embedded within IDBac, however the main analyses and interactive inputs are described below.

6.1. Double-click the IDBac desktop shortcut to launch IDBac. IDBac will open on the **Introduction** tab by default.

6.2. Use the **Check for Updates** button to ensure that the most current version of IDBac is being used (requires internet access). If a newer version is available, IDBac will automatically download and install the update, after which IDBac will request to be restarted.

6.3. Click on the **Starting with Raw Data** tab and choose from the menu the type of data to be used with IDBac; continue by following the in-app instructions.

6.4. When setting-up the conversion and processing of data files, input a descriptive name for the experiment where prompted (see **Figure 4**). Experiments will later be displayed alphabetically, so a helpful strategy is to start experiment names with a group-attribute (e.g., "bacillus-trials_experiment-1"; "bacillus-trials_experiment-2").

7. Work with previous experiments

7.1. After converting files and processing them with IDBac, or anytime one wishes to reanalyze an experiment, navigate to the **Work with previous experiments** page and **Select an experiment to work with** (**Figure 5**).

7.2. (Optional) Add information about samples using the menu **Click here to modify the selected experiment**. Input information into the auto-populated spreadsheet and press **Save** (**Figure 6**). This option allows the user to color-code data during analyses.

7.3. (Optional) Transfer all, or a subset of samples to a new or another experiment by clicking **Transfer samples from previous experiments to new/other experiments** and following the provided instructions (**Figure 7**).

7.4. When ready to begin analysis, ensure the experiment to work with is selected. Select either **Protein Data Analysis** or **Small Molecule Data Analysis**.

8. **Setting up protein data analysis and creating mirror plots**

8.1. If analyzing protein data, first navigate to the **Protein Data Analysis** page. Choose peak-picking settings and evaluate protein spectra of samples via the displayed mirror plots (**Figure 8**).

NOTE: In the mirror plots, a red peak signifies the presence of that peak only in the top spectrum, while blue peaks represent those occurring in both spectra.

8.2. Adjust the percentage of replicates in which a peak must be present in order for it to be included for analyses (e.g., if the threshold is set to 70% and a peak occurs in at least 7 out of 10 replicates, it will be included).

8.3. Using the mirror plots as visual guidance, adjust the signal to noise cutoff that retains the most "genuine" peaks and the least noise, noting that more replicates and a higher "percentage peak presence" value will allow selection of a lower signal to noise cutoff.

8.4. Specify the lower and upper *m/z* cutoffs, dictating the range of mass values within each spectrum to be used in further analyses by IDBac.

9. **Clustering samples using protein data**

9.1. Within the **Protein Data Analysis** page, select the **Dendrogram** tab. This allows for grouping samples into a dendrogram according to user-selected distance measures and clustering algorithms.

9.2. Click **Select Samples** on the menu and follow the instructions to select samples to include in the analyses. Only samples that contain protein spectra will be displayed within the **Available Samples** box (**Figure 9**).

9.3. Use the default values or, under **Choose Clustering Settings**, select the desired **Distance** and **Clustering** algorithms to be applied to the generation of the dendrogram.

9.4. Select **Presence/Absence** as input. Alternatively, if confident about the peak heights of the samples (e.g., after performing a study to assess variability of peak intensity), select **Intensities** as input.

NOTE: At the time of publication, IDBac provides flexibility in settings for clustering, relying on users to choose the appropriate combinations. If unfamiliar with these options, it is suggested to pair either A: "cosine" distance, and "average (UPGMA)" clustering; or B: "Euclidean" distance, and "Ward.D2" clustering.

9.5. To display bootstrap values on the dendrogram, enter a number between 2 and 1000 under **Bootstraps**.

9.6. When reporting results, copy the text within the **Suggestions for Reporting Protein Analysis** paragraph. This provides the user-defined settings that generated the specific dendrogram.

10. Customizing the protein dendrogram

10.1. To begin customizing the dendrogram, open the **Adjust the dendrogram** menu (**Figure 10**).

10.2. To color the dendrogram's lines and/or labels select the appropriate button: **Click to modify lines** or **Click to modify labels** and select the desired options.

10.3. To plot information from the spreadsheet next to the dendrogram (see step 7.2), select the button **Incorporate info about samples**. This will open a panel where a category (column in the spreadsheet) will self-populate based on the entered values (**Figure 11**).

11. Insert samples from a separate experiment into the dendrogram

11.1. To insert samples from another experiment, select the menu button **Insert Samples from Another Experiment**. Follow the directions in the newly-opened panel (**Figure 12**).

12. **Analyzing specialized metabolite data and metabolite association networks (MANs)**

12.1. Proceed to the **Metabolite Association Network (Small-Molecule)** page. This page allows for data visualization by principle components analysis (PCA) and MANs, which use bipartite networks to display the correlation of small molecule *m/z* values with samples.

12.2. If a protein dendrogram was created (section 9), it will be displayed on this page as well. Click-and-drag on the dendrogram to highlight select samples of interest to be analyzed. If no samples are highlighted or no protein dendrogram was created, a MAN of a either a random subset or all samples will appear, respectfully (**Figure 13**).

12.3. To subtract a matrix/media blank in the MAN, open the menu **Select a Sample to Subtract** and choose the appropriate sample to use as a blank.

12.4. Open the menu **Show/Hide MAN Settings** to select the desired values for percentage of peak presence in replicates, signal to noise, and upper and lower mass cutoffs, as was done for protein spectra in Section 9. Use the small molecule mirror plots to guide the selection of these settings.

12.5. Select "Download Current Network Data" to save the data of the MAN that is currently displayed. These data can be used in network analysis software other than IDBac.

12.6. For reporting results, copy the text within the **Suggestions for Reporting MAN Analysis** paragraph. This provides the user-defined settings used to generate the created MAN.

13. **Sharing data**

13.1. Each IDBac "experiment" is saved as a single SQLite database. It contains the converted mzML raw spectra, detected peaks, and all user-input information about samples. Therefore, to share an IDBac experiment simply share the SQlite file that has the same name as the experiment (the file location is displayed on the **Working with Previous Experiments** page).

REPRESENTATIVE RESULTS:

We analyzed six strains of *Micromonospora chokoriensis* and two strains of *Bacillus subtilis*, which were previously characterized[6], using data available at DOI: 10.5281/zenodo.2574096. Following directions in the **Starting with Raw Data** tab, we selected the option **Click here to convert Bruker files** and followed the IDBac-provided instructions for each dataset (**Figure 14**).

After the automated conversion and preprocessing/peak-peaking steps were completed, we proceeded to create a new combined IDBac experiment by transferring samples from the two experiments into a single experiment containing both *Bacillus* and *Micromonosopora* samples (**Figure 15**). The resulting analysis involved comparing protein spectra using mirror plots, as pictured in **Figure 16**, which was useful for evaluating spectra quality and adjusting peak-picking settings. **Figure 17** displays a screenshot of the protein clustering results with default settings selected. The dendrogram was colored by adjusting the threshold on the plot (appears as a dotted line). Of note is the clear separation between genera, with both *M. chokoriensis* and *B. subtilis* isolates clustering separately.

Figure 18, **Figure 19**, and **Figure 20** highlight the ability to generate MANs of user-selected regions by clicking and dragging across the protein dendrogram. With this we were able to rapidly create MANs to compare only the *B. subtilis* strains (**Figure 18**), only the *M. chokoriensis* strains (**Figure 19**), and all the strains simultaneously (**Figure 20**). The primary function of these networks is to provide researchers with a broad overview of the degree of specialized metabolite overlap between bacteria. With these data in hand, researchers now have the capacity to make informed decisions from only a small amount of material scraped from a bacterial colony.

FIGURE AND TABLE LEGENDS:

Figure 1: MALDI-target plate showing two different isolates before adding formic acid and MALDI matrix (top 3 spots – *Bacillus* sp.; bottom 3 spots – *Streptomyces* sp.). For both, column **3** represents excess sample; column **2** represents the appropriate amount of sample; column **1** represents insufficient sample for MALDI analysis.

Figure 2: Example protein spectra displaying the effect of modifying laser power and detector gain. Spectra quality is best in panel **A**, and decreases until insufficient spectra quality in panels **C** and **D**. While the spectrum in panel **B** may result in useable peaks, panel **A** displays optimal data.

Figure 3: Example specialized metabolite spectra displaying the effect of modifying laser power and detector gain. Spectra quality is best in panel **A** and decreases until insufficient spectra quality in panels **C** and **D**. While the spectrum in panel **B** may result in useable peaks, panel **A** displays optimal data.

Figure 4: IDBac data conversion and preprocessing step. IDBac converts raw spectra into the open mzML format and stores mzML, peak lists, and sample information in a database for each experiment.

Figure 5: "Work with Previous Experiments" page. Use IDBac's "Work with Previous Experiments" page to select an experiment to analyze or modify.

Figure 6: Input sample information. Within the "Work with Previous Experiments" page users can input information about samples such as taxonomic identity, collection location, isolation conditions, etc.

Figure 7: Transfer data. The "Work with Previous Experiments" page contains the option to transfer data between existing experiments and to new experiments.

Figure 8: Choose how peaks are retained for analysis. After selecting an experiment to analyze, visiting the "Protein Data Analysis" page and subsequently opening the "Choose how Peaks are Retained for Analysis" menu allows users to choose settings like signal-to-noise ratio for retaining peaks. The displayed mirror plot (or dendrogram) will automatically update to reflect the chosen settings.

Figure 9: Select samples from the chosen experiment to include within the displayed dendrogram.

Figure 10: Adjust the dendrogram. IDBac provides a few options for modifying how the dendrogram looks, these may be found within the menu "Adjust the Dendrogram". This includes coloring branches and labels by k-means, or by "cutting" the dendrogram at a user-provided height.

Figure 11: Incorporate info about samples. Within the "Adjust the Dendrogram" menu is the option "Incorporate info about samples". Selecting this will allow plotting information about samples next to the dendrogram. Sample information is input within the "Work with Previous Experiments" page.

Figure 12: Insert Samples from Another Experiment menu. Sometimes it is helpful to compare samples from another experiment. Use the "Insert Samples from Another Experiment" menu to choose samples to include within the currently-displayed dendrogram.

Figure 13: Small Molecule Data Analysis" page. If a dendrogram was created from protein spectra, it will be displayed within the "Small Molecule Data Analysis" page. This page will also display Metabolite Associate Networks (MANs) and Principle Components Analysis (PCA) for small molecule data.

Figure 14: Spectra processing. Downloaded Bruker autoFlex spectra were converted and processed using IDBac.

Figure 15: Combined IDBac experiment. Because the *Micromonospora* and *Bacillus* spectra were collected on different MALDI target plates, the two experiments were subsequently combined into a single experiment-"Bacillus_Micromonsopora". This was done within the "Work with Previous Experiments" tab, following directions within the menu "Transfer samples from previous experiments to new/other experiments".

Figure 16: Comparison. *Micromonspora* and *Bacillus* spectra were compared using the mirror plots within the "Protein Data Analysis" page. Ultimately, default peak settings were chosen.

Figure 17: Hierarchical clustering. Hierarchical clustering, using default settings, correctly grouped *Bacillus* and *Micromonospora* isolates. The dendrogram was colored by "cutting" the dendrogram at an arbitrary height (displayed as a dashed-line) and 100 bootstraps used to show confidence in branching.

Figure 18: MAN created by selecting the *Bacillus* sp. strains from the protein dendrogram showed differential production of specialized metabolites.

Figure 19: MAN created by selecting the six *Micromonospora* sp. strains from the protein dendrogram showed differential production of specialized metabolites.

Figure 20: MAN of *Bacillus* sp. and *Micromonospora* sp. strains showing a differential production of specialized metabolites.

DISCUSSION:

The IDBac protocol details bacterial protein and specialized metabolite data acquisition and analysis of up to 384 bacterial isolates in 4 h by a single researcher. With IDBac there is no need to extract DNA from bacterial isolates or generate specialized metabolite extracts from liquid

fermentation broths and analyze them using chromatographic methods. Instead, protein and specialized metabolite data are gathered by simply spreading material from bacterial colonies directly onto a MALDI target plate. This greatly reduces the time and cost associated with alternative techniques such as 16S rRNA gene sequencing and LCMS[9].

It is important to add a matrix blank and calibration spots to the MALDI plate, and we recommend using an appropriate number of replicates to ensure reproducibility and statistical confidence. The numbers of replicates will be experiment-dependent. For example, if a user intends to differentiate thousands of colonies from a collection of environmental diversity plates, fewer replicates may be necessary (our lab collects three technical replicates per colony). Alternatively, if a user wishes to create a custom database of strains from specific bacterial taxa to rapidly determine sub-species classifications of unknown isolates, then more replicates are appropriate (our lab collects eight biological replicates per strain).

IDBac is a tool for rapidly differentiating highly-related bacterial isolates based on putative taxonomic information and specialized metabolite production. It can complement or serve as a precursor to orthogonal methods such as in-depth genetic analyses, studies involving metabolite production and function, or characterization of specialized metabolite structure by Nuclear Magnetic Resonance spectroscopy and/or LC-MS/MS.

Specialized metabolite production (IDBac MANs) is highly susceptible to bacterial growth conditions, especially using different media, which is a potential limitation of the method. However these traits may be exploited by the user, as IDBac can readily generate MANs showing the differences in specialized metabolite production under a variety of growth conditions. It is important to note that while specialized metabolite fingerprints may vary by growth condition, we have shown that protein fingerprints remain relatively stable (see Clark et al.[6]). When dealing with environmental diversity plates, we recommend purifying bacterial isolates prior to analysis in order to reduce possible contributions from neighboring bacterial cross-talk.

Finally, the lack of a searchable public database of protein MS fingerprints is a major shortcoming in the use of this method to classify unknown environmental bacteria. We created IDBac with this in mind, and included automated conversion of data into a community-accepted open-source format (mzML)[10–12] and designed the software to allow searching, sharing, and creation of custom databases. We are in the process of creating a large public database (>10,000 fully characterized strains), which will allow for the classification of some isolates to the species-level, including links to GenBank accession numbers when available.

IDBac is open source and the code is available for anyone to customize their data analysis and visualization needs. We recommend that users consult an extensive body of literature (Sauer et al.[7], Silva et al.[5]) to help support and design their experimental goals. We host a forum for discussion at: https://groups.google.com/forum/#!forum/idbac and a means to report issues with the software at: https://github.com/chasemc/IDBacApp/issues.

ACKNOWLEDGMENTS:

This work was supported by National Institute of General Medical Sciences Grant R01 GM125943, National Geographic Grant CP-044R-17; Icelandic Research Fund Grant 152336-051; and University of Illinois at Chicago startup funds. Also, we thank the following contributors: Dr. Amanda Bulman for assistance with MALDI-TOF MS protein acquisition parameters; Dr. Terry Moore and Dr. Atul Jain for recrystallizing alpha-cyano-4-hydroxycinnamic acid matrix (CHCA).

DISCLOSURES:

The authors have nothing to disclose.

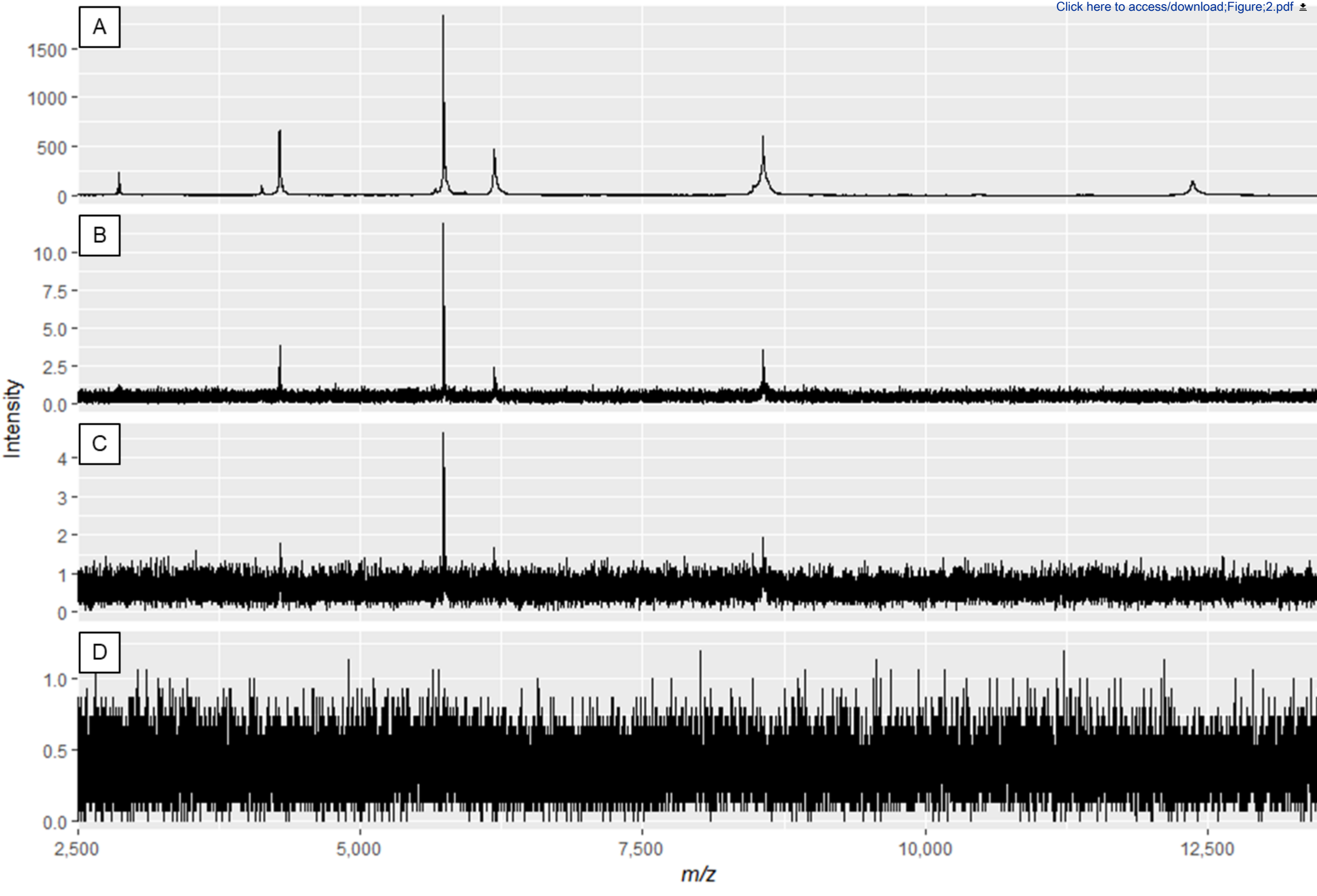
REFERENCES:

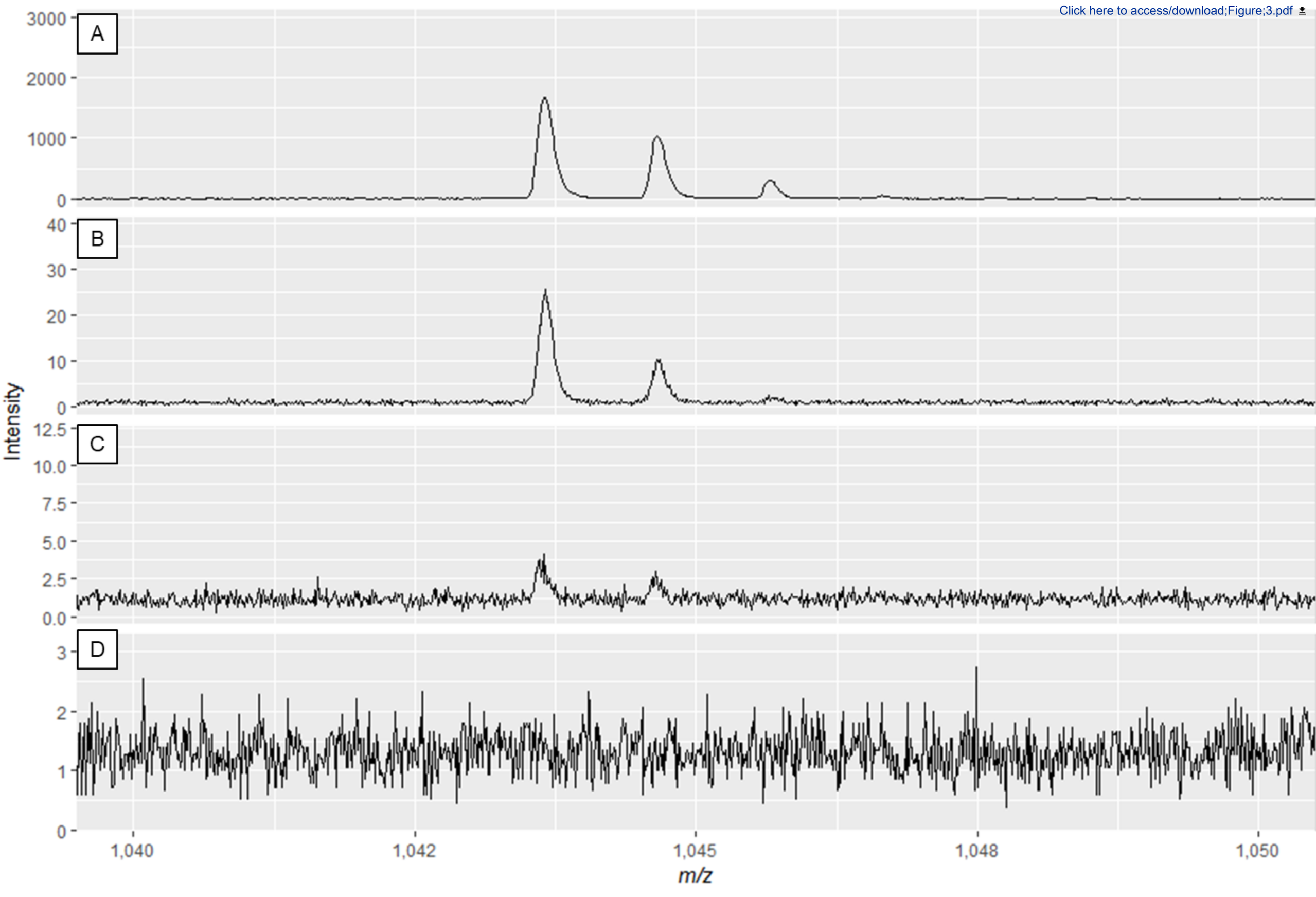
1. Sandrin, T. R., Goldstein, J. E., Schumaker, S. MALDI TOF MS profiling of bacteria at the strain level: A review. *Mass Spectrometry Reviews* **32**, (3) 188–217 (2013).
2. Cain, T. C., Lubman, D. M., Weber, W. J., Vertes, A. Differentiation of bacteria using protein profiles from matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* **8**, (12) 1026–1030 (1994).
3. Holland, R. D., Wilkes, J. G., et al. Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* **10**, (10) 1227–1232 (1996).
4. Rahi, P., Prakash, O., Shouche, Y. S. Matrix-assisted laser desorption/ionization time-of-flight mass-spectrometry (MALDI-TOF MS) based microbial identifications: challenges and scopes for microbial ecologists. *Frontiers in Microbiology* **7**, 1359 (2016).
5. Silva, R., Lopes, N. P., Silva, D. B. Application of MALDI mass spectrometry in natural products analysis. *Planta Medica* **82**, 671–689 (2016).
6. Clark, C. M., Costa, M. S., Sanchez, L. M., Murphy, B. T. Coupling MALDI-TOF mass spectrometry protein and specialized metabolite analyses to rapidly discriminate bacterial function. *Proceedings of the National Academy of Sciences of the United States of America* **115**, (19) 4981–4986 (2018).
7. Freiwald, A., Sauer, S. Phylogenetic classification and identification of bacteria by mass spectrometry. *Nature Protocols* **4**, (5) 732–742 (2009).
8. Schulthess, B., Bloemberg, G. V, Zbinden, R., Böttger, E. C., Hombach, M. Evaluation of the Bruker MALDI Biotyper for identification of Gram-positive rods: development of a diagnostic algorithm for the clinical laboratory. *Journal of Clinical Microbiology* **52**, (4) 1089–97 (2014).
9. Schumann, P., Maier, T. MALDI-TOF mass spectrometry applied to classification and identification of bacteria. *Methods in Microbiology* **41**, 275–306 (2014).
10. Chambers, M. C., Maclean, B., et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology* **30**, (10) 918–920 (2012).
11. Kessner, D., Chambers, M., Burke, R., Agus, D., Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **24**, (21) 2534 (2008).
12. Martens, L., Chambers, M., et al. mzML-a community standard for mass spectrometry data. *Molecular & Cellular Proteomics* **10**, (1) (2011).

1 2 3

B*acillus* sp.

Streptomyces sp.





IDBac

127.0.0.1:5125

IDBac | Introduction | Starting With Raw Data | Work With Previous Experiments

Create an IDBac experiment

- Click here to convert Bruker files
- Click here to convert mzML/mzXML files
- Click here to convert txt files
- Click here to convert Microtyper files

Starting with mzML or mzXML Data:

1: Enter a name for this new experiment

This will become a filename, non-valid characters will be removed.

Hint: Intead of a space, use "_".

Enter Experiment Name Here

Once entered, the filename-friendly version of the entered name will appear here once.

This will be the version of your experiment name that is saved.

2: Click to select the location of your mzML files

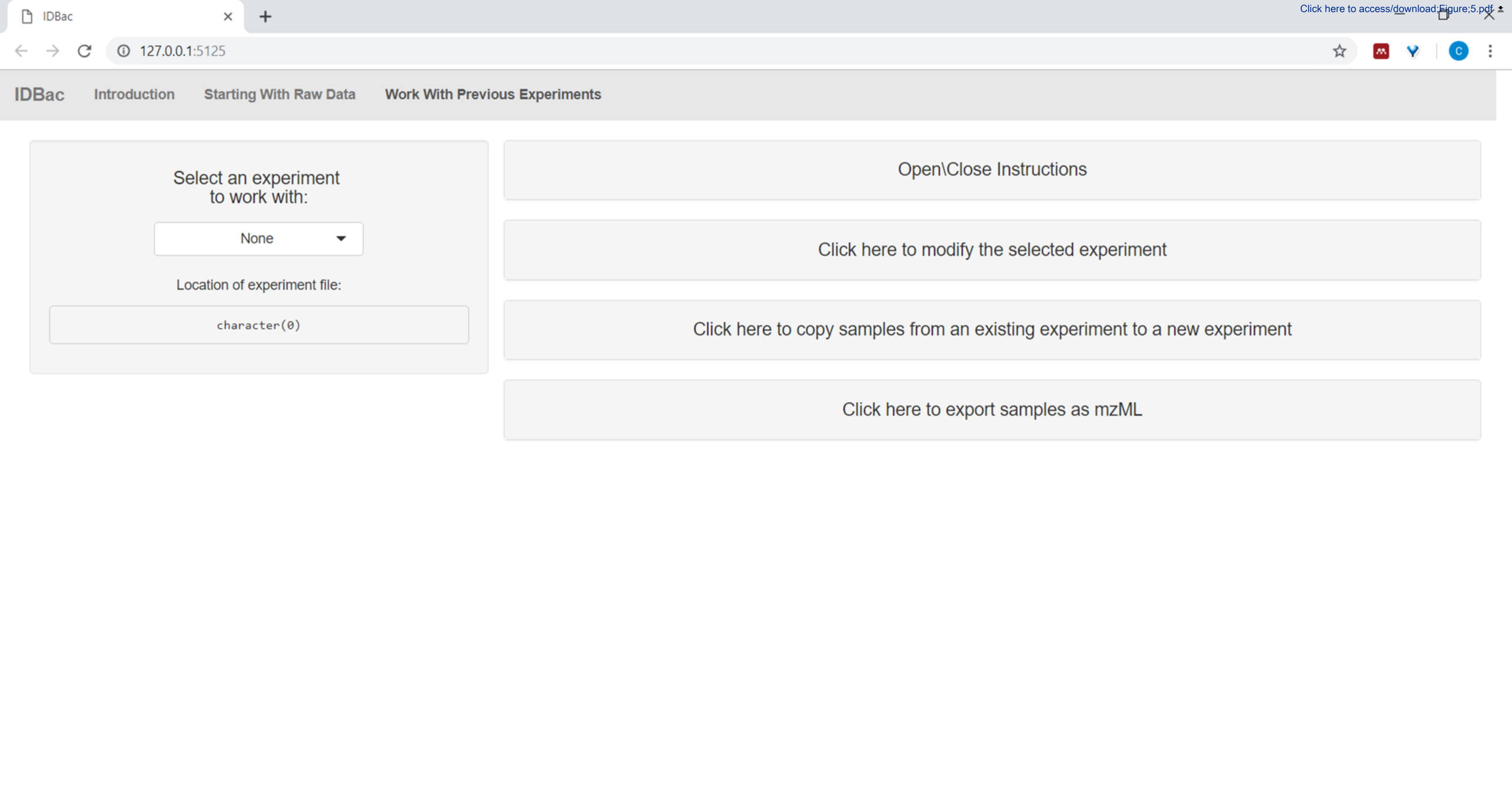
Raw Data Folder

No Folder Selected

Samples will be named according to the file name of the provided files

4: Click "Process Data" to begin spectra conversion.

Process Data



IDBac

127.0.0.1:5125

IDBac | Introduction | Starting With Raw Data | **Work With Previous Experiments** | Protein Data Analysis | Small Molecule Data Analysis

Select an experiment to work with:

example

Location of experiment file:

[1] "C:/Users/CMC/Documents/GitHub/IDBac_App/example.sqli

Open\Close Instructions

Click here to modify the selected experiment

Add/modify information about samples | Experiment Summary

Here is where you can add information about your sample. There are always standard columns like "Genus", but you can add your own columns as well.

After you are finished inputting your information, press "save" to write the information to the database.

New Column Name

Insert Column

save

	Strain_ID	Genbank_Acce	NCBI_TaxID	Kingdom	Phylum	Class	Order	Family	Genus
1	Example_Strai	KY858228	446370	Bacteria	Firmicutes	Bacilli	Bacillales	Paenibacillacea	Paenibacillus
2	172-1								Rhodococcus
3	172-10								Paenibacillus
4	172-11								Paenibacillus
5	172-7								Rhodococcus
6	Matrix								

Click here to copy samples from an existing experiment to a new experiment

IDBac

127.0.0.1:5125

Transfer samples from previous experiments to new/other experiments.

Select an existing experiment to copy samples from:

example

Location of experiment file:

```
[1] "C:/Users/CMC/Documents/GitHub/IDBac_App/example.sqlite"
```

New experiment name:

Enter new experiment name here

Move samples between boxes by clicking the strain's name and then an arrow. Samples in the right box will be used for analysis.

Available Samples

172-1
172-10
172-11
172-7
Matrix

Selected Samples

Add to new Experiment

Analysis Settings

Choose how Peaks are Retained for Analyses

(Effects all protein analysis)

Percent Presence

70

Signal to Noise Cutoff

4

Lower Mass Cutoff

3000

Upper Mass Cutoff

15000

Select Samples

Choose Clustering Settings

Optional Settings

Mirror Plots | Dendrogram

Mirror Plots

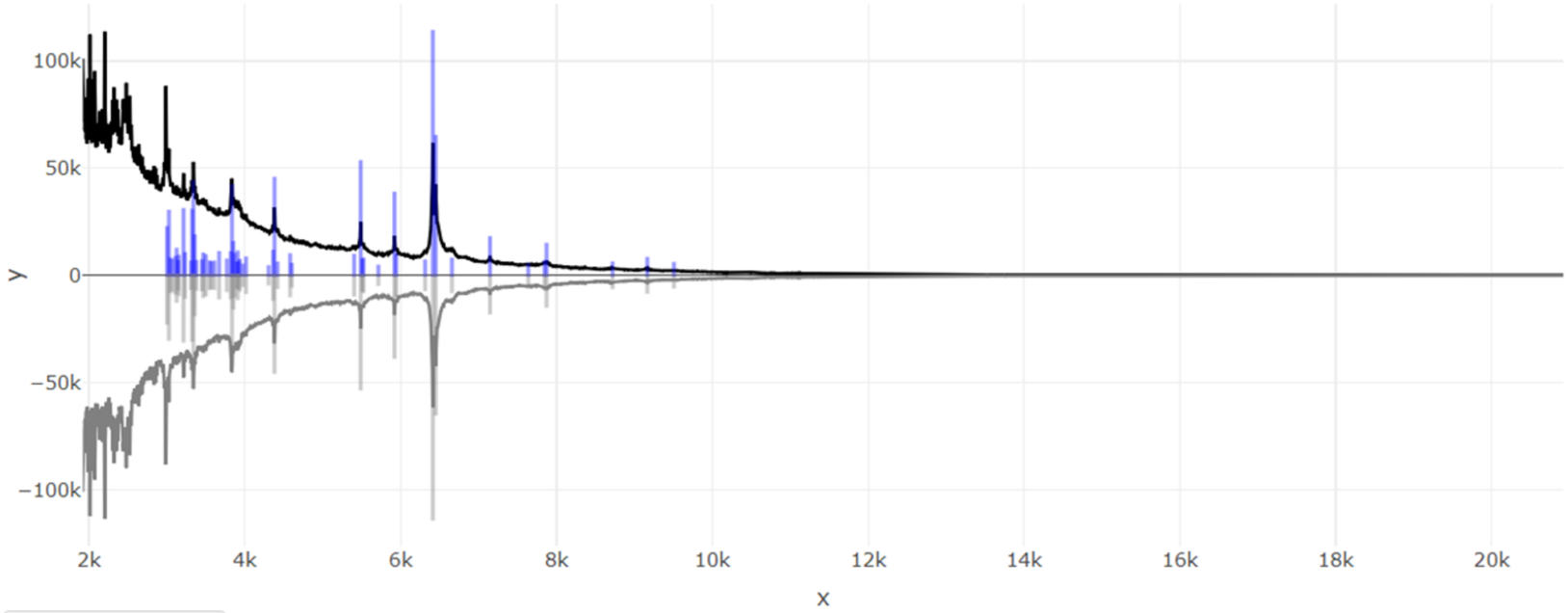
Choose two samples to compare in the mirror plots below. Matching peaks will be colored blue and non-matching peaks will be colored red.

Spectrum 1 (positive y-axis)

172-1

Spectrum 2 (negative y-axis)

172-1



Download SVG

Analysis Settings

Choose how Peaks are Retained for Analyses

(Effects all protein analysis)

Select Samples

Move samples between boxes by clicking the samples's name and then an arrow. Samples in the right box will be used for analysis.

Available Samples

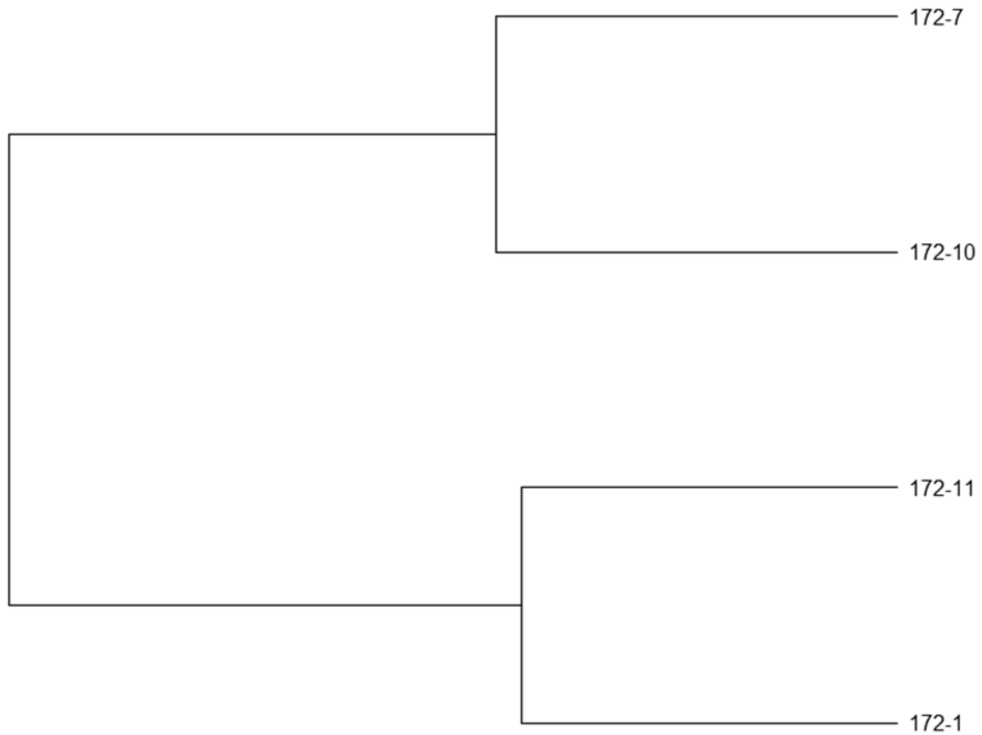
Selected Samples

172-1
172-10
172-11
172-7

Choose Clustering Settings

Optional Settings

Mirror Plots | Dendrogram



IDBac

127.0.0.1:5125

Select Samples

Choose Clustering Settings

Optional Settings

Adjust the Dendrogram

Expand dendrogram

750

Adjust right margin of dendrogram

20

Label Positions:

- Plot all labels at x = 0
- Hang labels

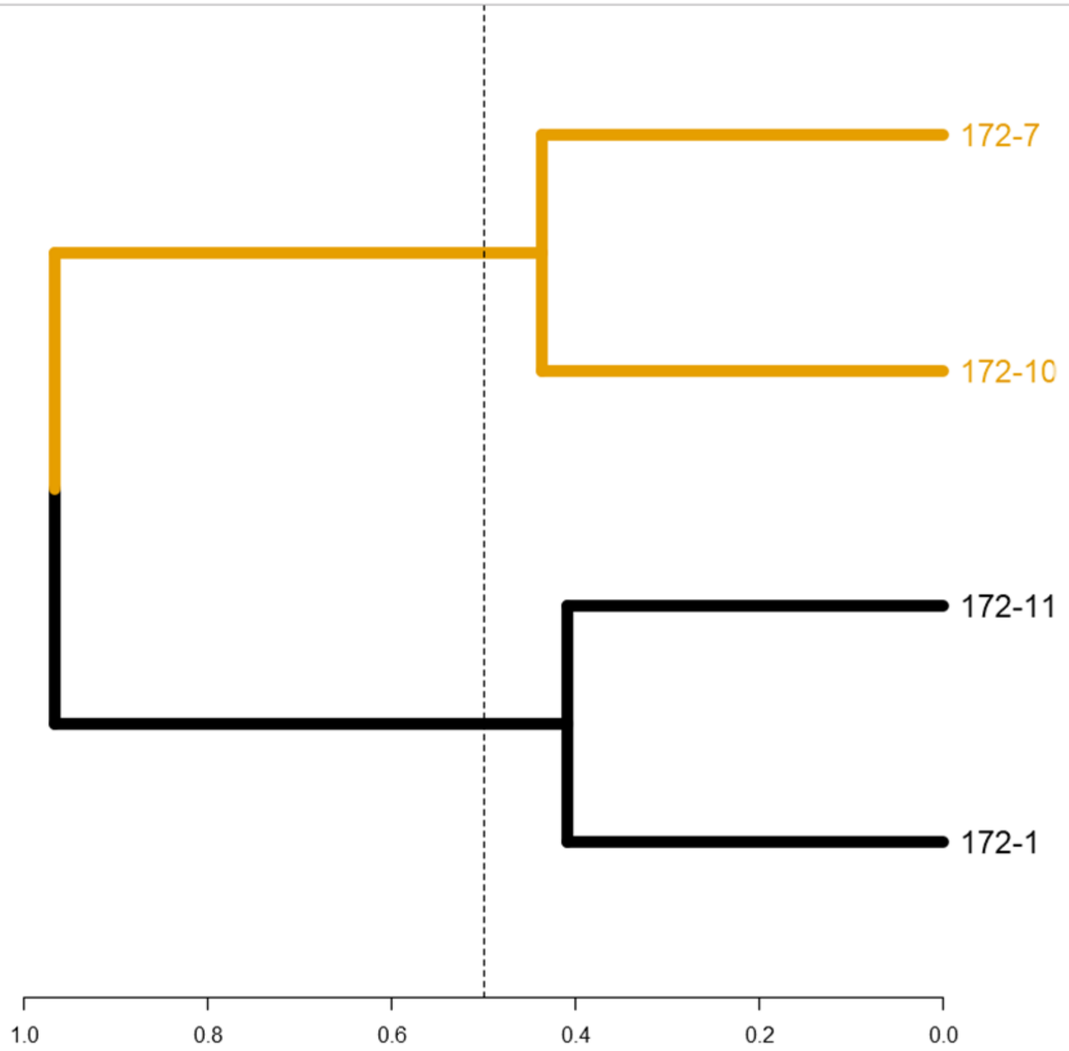
Incorporate info about samples

Click to modify lines

Click to modify labels

Insert Samples from Another Experiment

PCA, PCoA, t-SNE



Adjust Dendrogram Lines

Color By:

Color by cutting at height

Cut Tree at Height

0.5

Line Width

7

Close

Adjust Dendrogram Labels

Color By:

Choose Number of Groups

Choose the number of groups

2

Label Size

1.5

Close

IDBac

127.0.0.1:5125

Mirror Plots | Dendrogram

Analysis Settings

Choose how Peaks are Retained for Analyses

(Effects all protein analysis)

Select Samples

Choose Clustering Settings

Optional Settings

Adjust the Dendrogram

Expand dendrogram

750

Adjust right margin of dendrogram

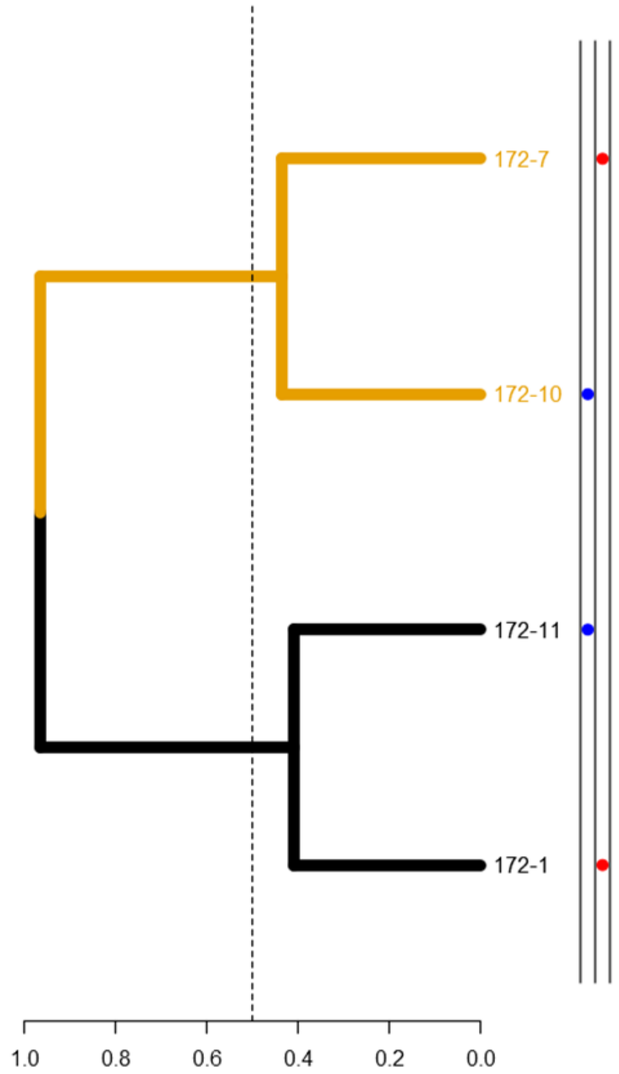
40

Label Positions:

- (•) Plot all labels at x = 0
- ( ) Hang labels

Incorporate info about samples

Click to modify lines | Click to modify labels



Select Category

Genus

Rhodococcus

#FF0000

Paenibacillus

#0000FF

Close

IDBac

127.0.0.1:5125

Analysis Settings

Choose how Peaks are Retained for Analyses

(Effects all protein analysis)

Select Samples

Choose Clustering Settings

Optional Settings

Adjust the Dendrogram

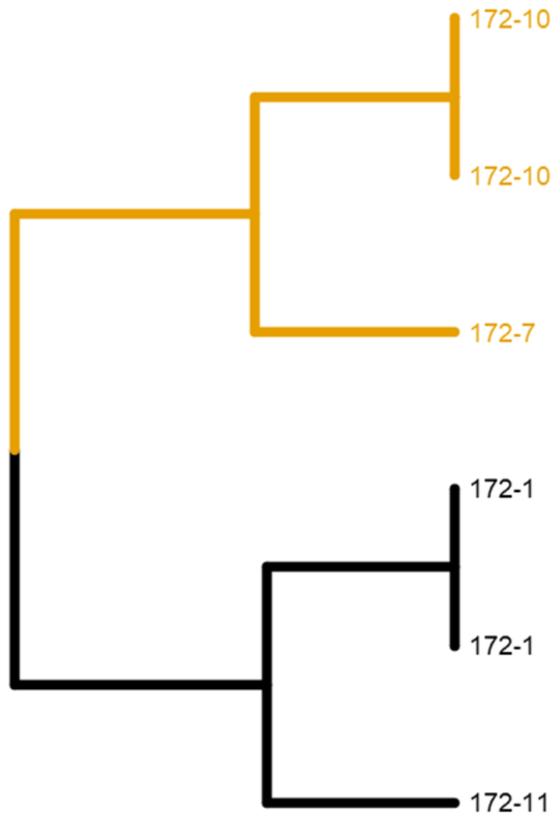
Insert Samples from Another Experiment

Insert samples from another experiment

PCA, PCoA, t-SNE

Save Dendrogram

Mirror Plots | Dendrogram



Select an experiment to work with:

example

Location of experiment file:

[1] "C:/Users/CMC/Documents/GitHub/IDBac_App/example.sqlite"

Available Samples

172-11
172-7

Selected Samples

172-1
172-10

Close

New Tab | IDBac

127.0.0.1:7894

IDBac | Introduction | Starting With Raw Data | Work With Previous Experiments | Protein Data Analysis | **Small Molecule Data Analysis**

Show/Hide MAN Settings

Select a Sample to Subtract

Adjust Protein Dendrogram

Download Current Network Data

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

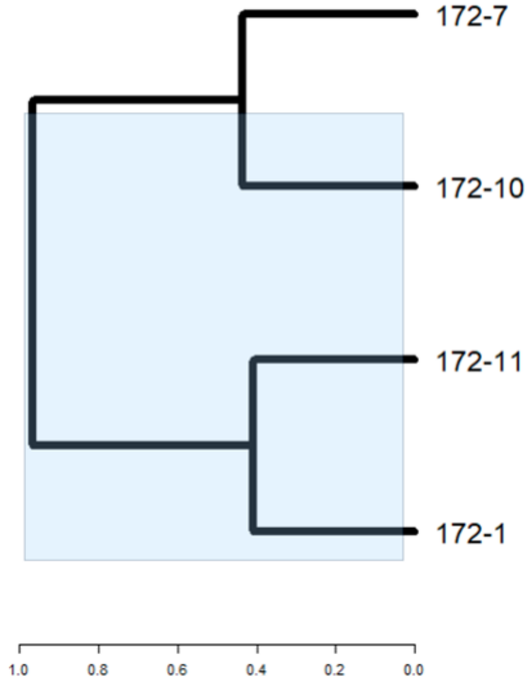
Hint 2: Use mouse to click & drag parts (nodes) of the MAN if it appears congested.

Note 1: For publication-quality networks click "Download Current Network Data". This will download a .csv file of the currently-displayed network that can be easily imported into Gephi or Cytoscape.

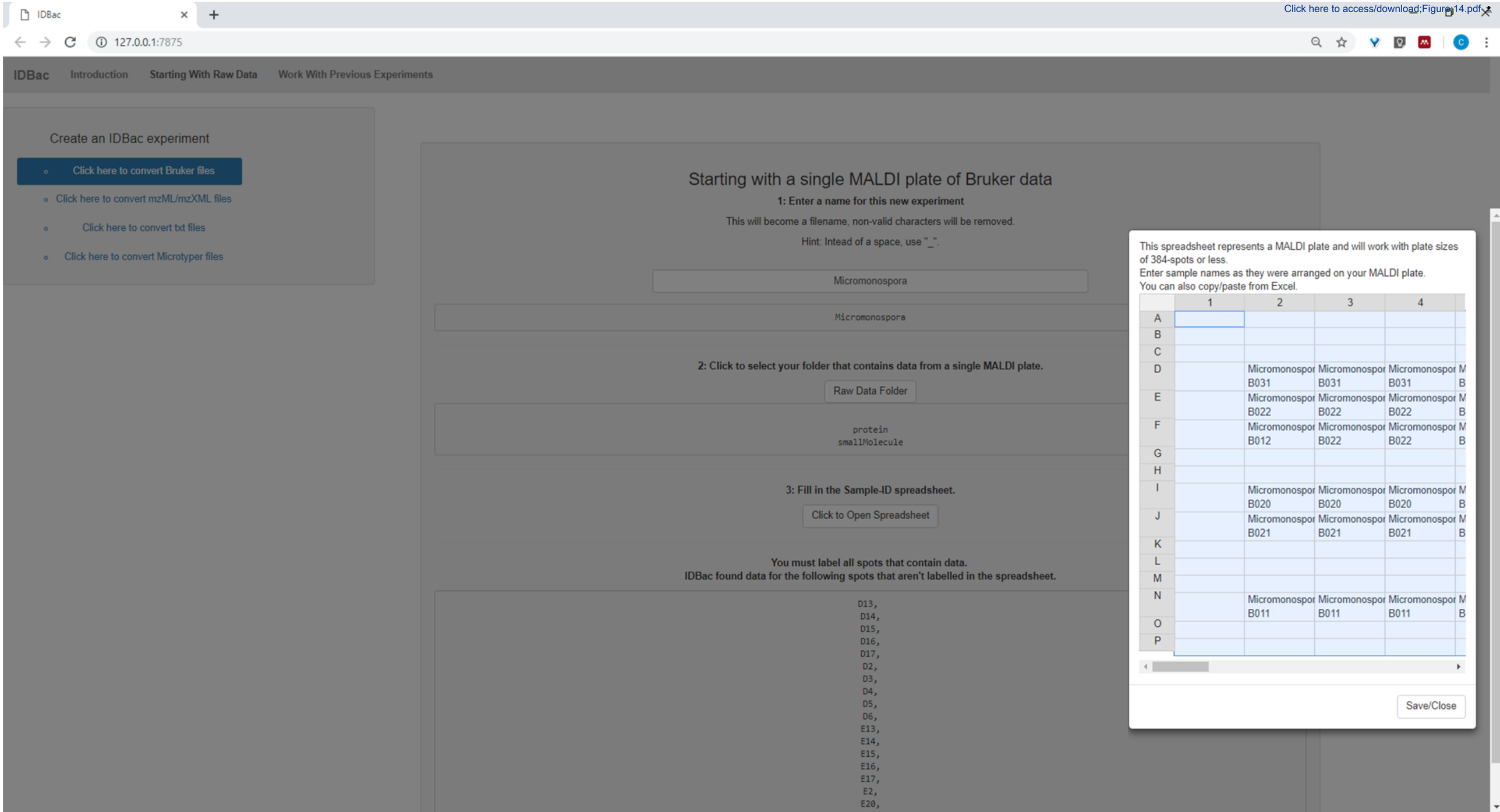
Suggestions for Reporting MAN Analysis:

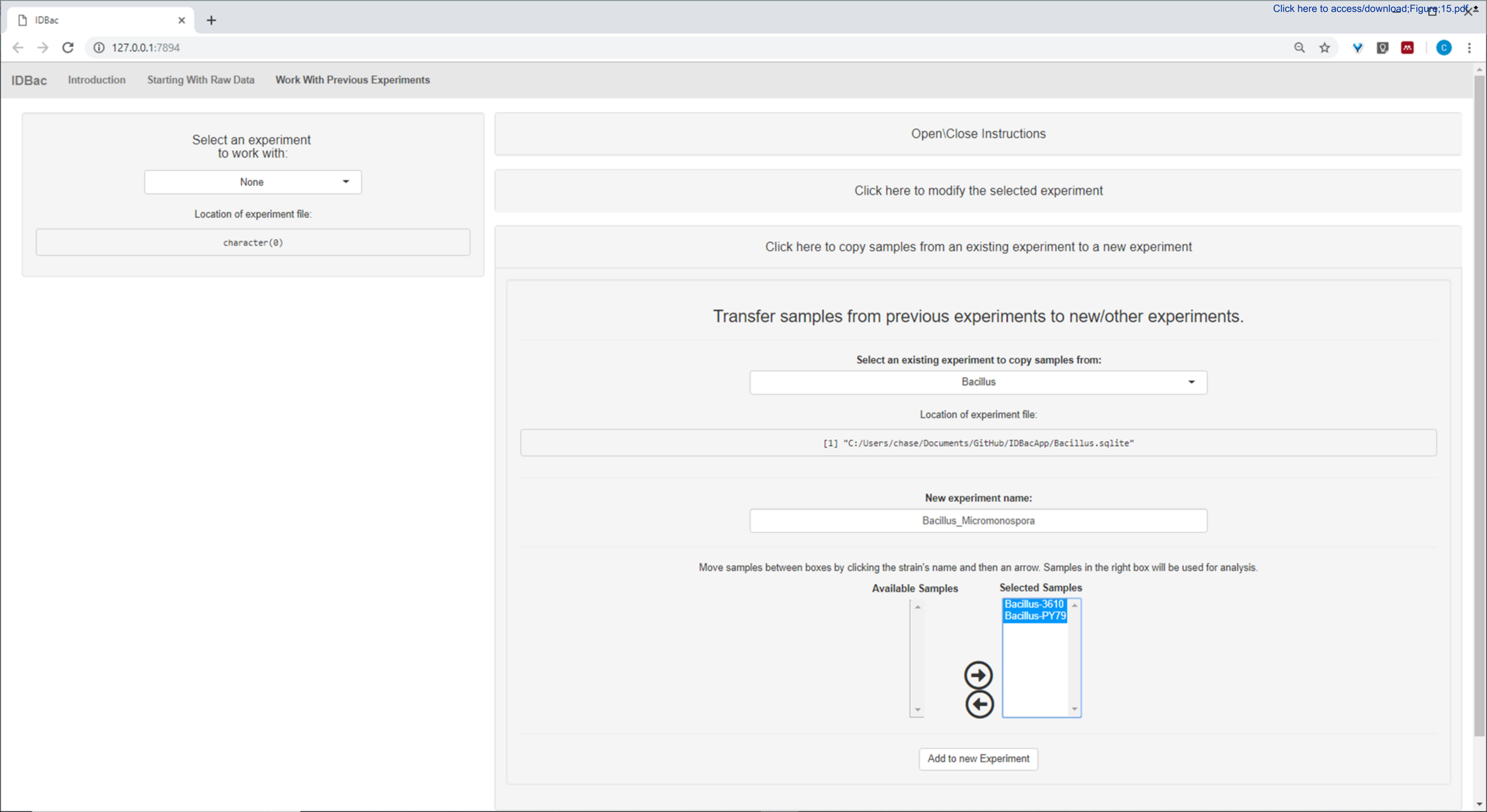
This MAN was created by analyzing `3` samples, retaining peaks with a signal to noise ratio above `4` , and occurring in greater than `70` % of replicate spectra. Peaks occuring below `200` m/z or above `2000` m/z were removed from the analysis.

Protein Dendrogram



Small Molecule MAN | SmallMolecule PCA





IDBac

127.0.0.1:7875

IDBac | Introduction | Starting With Raw Data | Work With Previous Experiments | Protein Data Analysis | Small Molecule Data Analysis

Analysis Settings

Choose how Peaks are Retained for Analyses
(Effects all protein analysis)

Select Samples

Choose Clustering Settings

Optional Settings

Adjust the Dendrogram

Insert Samples from Another Experiment

PCA, PCoA, t-SNE

Save Dendrogram

Suggestions for Reporting Protein Analysis:

This dendrogram was created by analyzing `8` samples, and retaining peaks with a signal to noise ratio above `4` and occurring in greater than `70` % of replicate spectra. Peaks occuring below `3000` m/z or above `15000` m/z were removed from the analyses. For clustering spectra, distance and algorithms were used.

Mirror Plots | Dendrogram

Mirror Plots

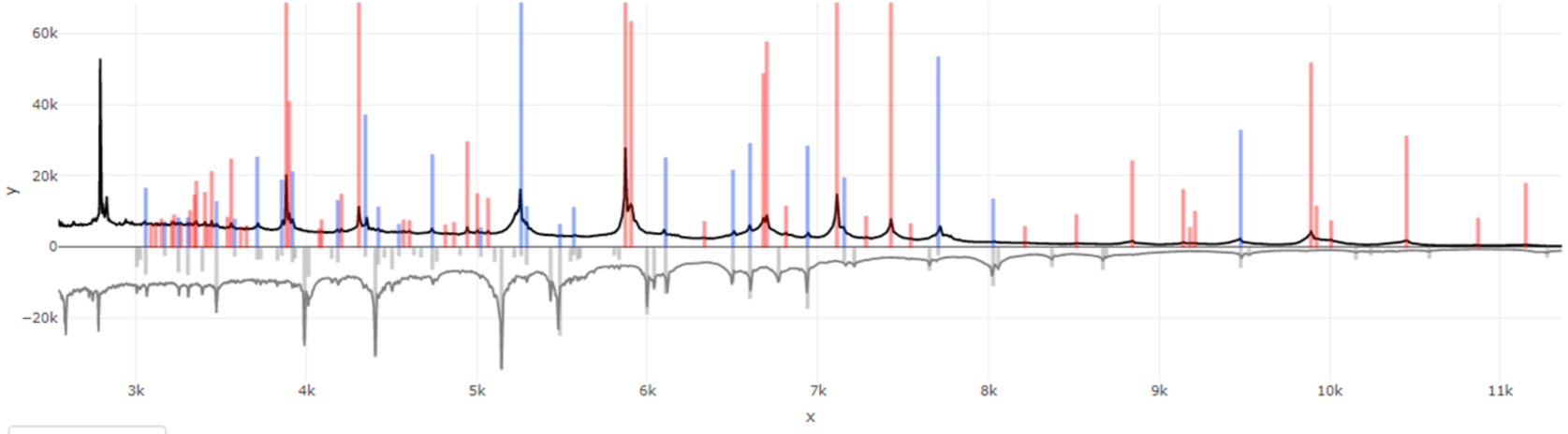
Choose two samples to compare in the mirror plots below. Matching peaks will be colored blue and non-matching peaks will be colored red.

Spectrum 1 (positive y-axis)

Bacillus-3610

Spectrum 2 (negative y-axis)

Micromonospora-B011



Download SVG

Analysis Settings

Choose how Peaks are Retained for Analyses
(Effects all protein analysis)

Select Samples

Choose Clustering Settings

Optional Settings

Adjust the Dendrogram

Expand dendrogram

750

Adjust right margin of dendrogram

20

Label Positions:

- Plot all labels at x = 0
- Hang labels

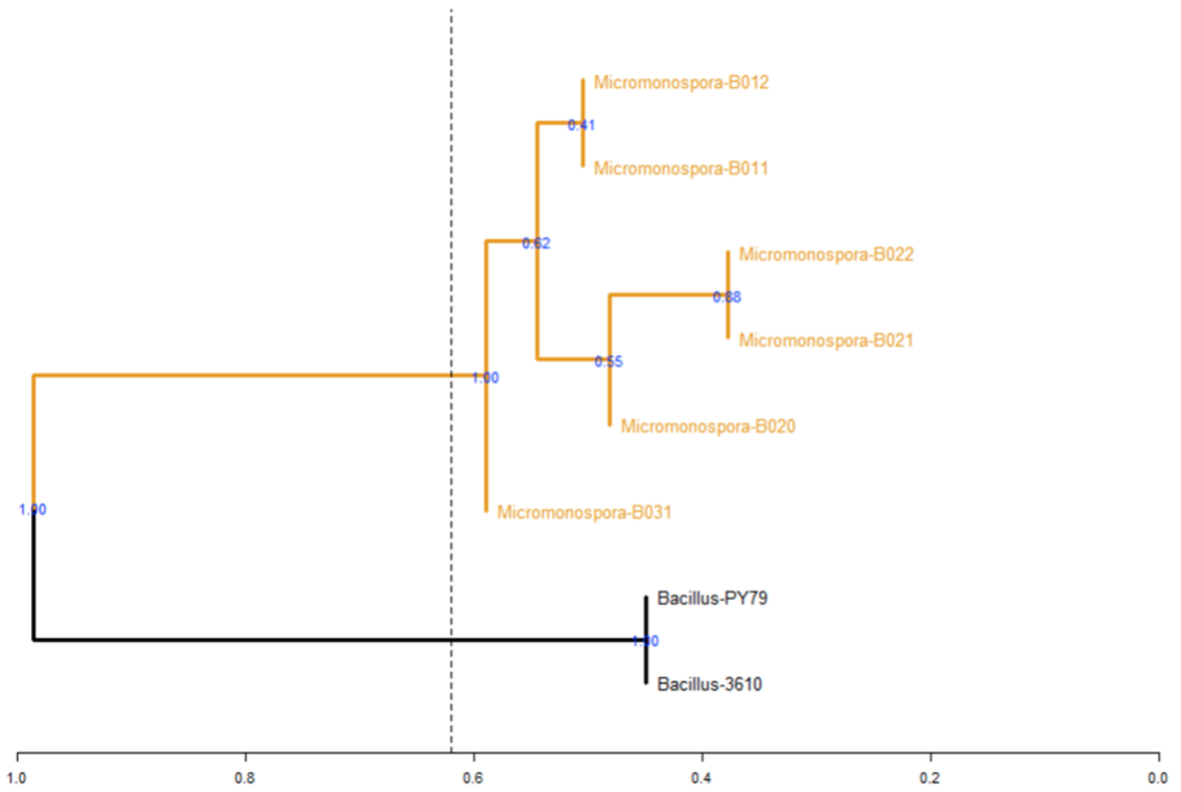
Incorporate info about samples | Click to modify lines | Click to modify labels

Insert Samples from Another Experiment

PCA, PCoA, t-SNE

Save Dendrogram

Mirror Plots | Dendrogram



Suggestions for Reporting Protein Analysis:

IDBac

127.0.0.1:7875

IDBac | Introduction | Starting With Raw Data | Work With Previous Experiments | Protein Data Analysis | **Small Molecule Data Analysis**

Show/Hide MAN Settings

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

70

Signal To Noise Cutoff

4

Upper Mass Cutoff

3000

Lower Mass Cutoff

200

Select a Sample to Subtract

Matrix

Adjust Protein Dendrogram

Download Current Network Data

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

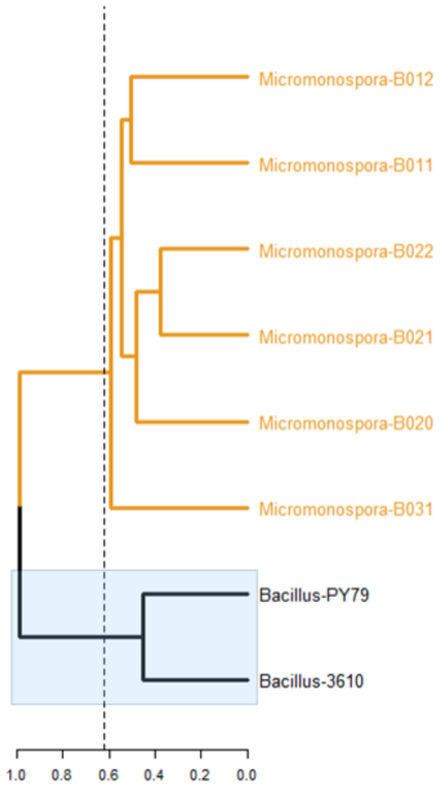
Hint 2: Use mouse to click & drag parts (nodes) of the MAN if it appears congested.

Note 1: For publication-quality networks click "Download Current Network Data". This will download a .csv file of the currently-displayed network that can be easily imported into Gephi or Cytoscape.

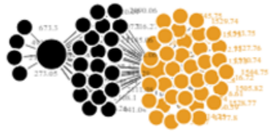
Suggestions for Reporting MAN Analysis:

This MAN was created by analyzing 2 samples, subtracting a matrix blank, retaining peaks with

Protein Dendrogram



Small Molecule MAN | SmallMolecule PCA



IDBac

127.0.0.1:7875

IDBac | Introduction | Starting With Raw Data | Work With Previous Experiments | Protein Data Analysis | Small Molecule Data Analysis

Show/Hide MAN Settings

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

70

Signal To Noise Cutoff

4

Upper Mass Cutoff

3000

Lower Mass Cutoff

200

Select a Sample to Subtract

Matrix

Adjust Protein Dendrogram

Download Current Network Data

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

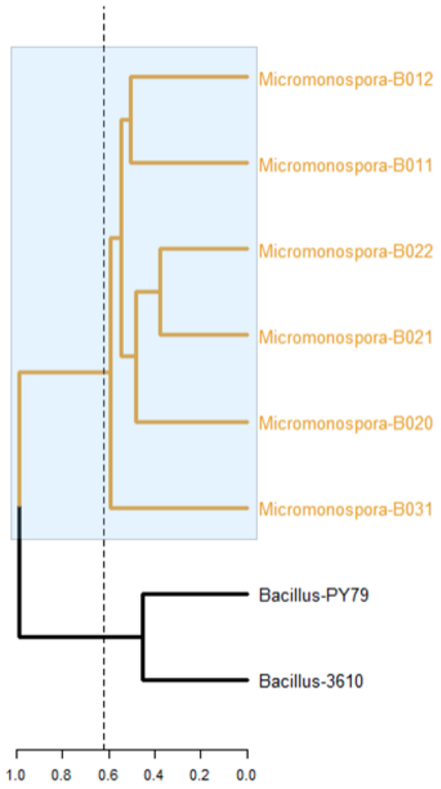
Hint 2: Use mouse to click & drag parts (nodes) of the MAN if it appears congested.

Note 1: For publication-quality networks click "Download Current Network Data". This will download a .csv file of the currently-displayed network that can be easily imported into Gephi or Cytoscape.

Suggestions for Reporting MAN Analysis:

This MAN was created by analyzing 6 samples, subtracting a matrix blank, retaining peaks with

Protein Dendrogram



Small Molecule MAN | SmallMolecule PCA

IDBac

127.0.0.1:7875

IDBac | Introduction | Starting With Raw Data | Work With Previous Experiments | Protein Data Analysis | **Small Molecule Data Analysis**

Show/Hide MAN Settings

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

70

Signal To Noise Cutoff

4

Upper Mass Cutoff

3000

Lower Mass Cutoff

200

Select a Sample to Subtract

Matrix

Adjust Protein Dendrogram

Download Current Network Data

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

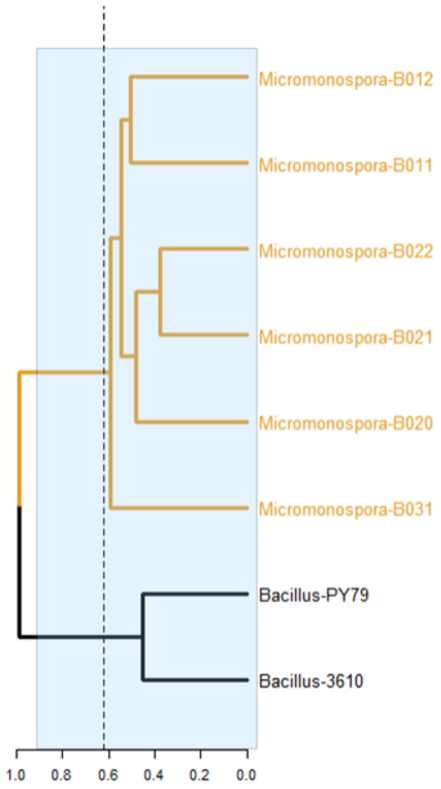
Hint 2: Use mouse to click & drag parts (nodes) of the MAN if it appears congested.

Note 1: For publication-quality networks click "Download Current Network Data". This will download a .csv file of the currently-displayed network that can be easily imported into Gephi or Cytoscape.

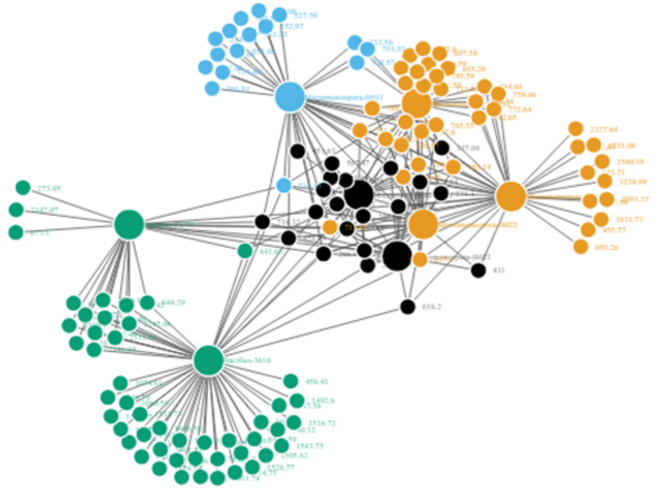
Suggestions for Reporting MAN Analysis:

This MAN was created by analyzing 8 samples, subtracting a matrix blank, retaining peaks with

Protein Dendrogram



Small Molecule MAN | SmallMolecule PCA



Parameter	Protein	Specialized Metabolite
Mass Start	1920	60
Mass End	21000	2700
Mass Deflection	1900	50
Shots	500	1000
Frequency	2000	2000
Laser Size	Large	Medium
MaxStdDev (ppm)	300	30

Table of Materials

Click here to access/download;Table of Materials;JoVE_Material.xlsx

Name of Material/ Equipment	Company	Catalog Number
Acetonitrile	Fisher	60-002-65
Autoflex Speed LEF MALDI-TOF instrument	Bruker Daltonics	
Bruker Daltonics Bacterial test standard	Fisher	NC0884024
Bruker Peptide Calibration standard	Fisher	NC9846988
Formic Acid	Fisher Chemical	A117-50
MALDI-TOF target Plate	Bruker Daltonics	
Methanol	Fisher Chemical	A456-500
Toothpicks		
Trifluoroacetic acid	Fisher	AC293810010
Water	VWR	7732-18-5
α-Cyano-4-hydroxycinnamic acid	Sigma	28166-41-8

Comments/Description
LC-MS Ultra CHROMASOLV
Bruker Daltonics 8604530
Bruker Daltonics 8206195
99.5+%, Optima LC/MS Grade
Optima LC/MS Grade
any is ok
99.5%, for biochemistry, ACROS Organics
LC-MS
(C2020-25G) ≥98% (TLC), powder

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article: IDBac: An open source MALDI MS Platform...

Author(s): See Manuscript.

Item 1: The Author elects to have the Materials be made available (as described at http://www.jove.com/publish) via:

☐ Standard Access ☒ Open Access

Item 2: Please select one of the following items:

☒ The Author is **NOT** a United States government employee.

☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. **Defined Terms.** As used in this Article and Video License Agreement, the following terms shall have the following meanings: "**Agreement**" means this Article and Video License Agreement; "**Article**" means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; "**Author**" means the author who is a signatory to this Agreement; "**Collective Work**" means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; "**CRC License**" means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode; "**Derivative Work**" means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; "**Institution**" means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; "**JoVE**" means MyJove Corporation, a Massachusetts corporation and the publisher of The Journal of Visualized Experiments; "**Materials**" means the Article and / or the Video; "**Parties**" means the Author and JoVE; "**Video**" means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. **Background.** The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. **Grant of Rights in Article.** In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and(c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

4. **Retention of Rights in Article.** Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. **Grant of Rights in Video – Standard Access.** This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. **Grant of Rights in Video – Open Access.** This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this **Section 6** is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. **Government Employees.** If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. **Protection of the Work.** The Author(s) authorize JoVE to take steps in the Author(s) name and on their behalf if JoVE believes some third party could be infringing or might infringe the copyright of either the Author's Article and/or Video.

9. **Likeness, Privacy, Personality.** The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

10. **Author Warranties.** The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

11. **JoVE Discretion.** If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

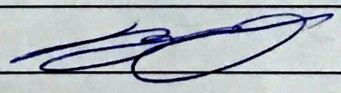
12. **Indemnification.** The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in JoVE or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

13. **Fees.** To cover the cost incurred for publication, JoVE must receive payment before production and publication the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

14. **Transfer, Governing Law.** This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement is required per submission.

CORRESPONDING AUTHOR

Name:	Brian T. Murphy
Department:	Medicinal chemistry + Pharmacognosy
Institution:	U of Illinois Chicago
Title:	Associate Professor
Signature:	
Date:	10/9/18

Please submit a **signed** and **dated** copy of this license by one of the following three methods:

1. Upload an electronic version on the JoVE submission site
2. Fax the document to +1.866.381.2236
3. Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02140

Department of Medicinal Chemistry and Pharmacognosy (MC 781)
College of Pharmacy
833 South Wood Street; Chicago, Illinois 60612-7231

Brian T. Murphy
Associate Professor

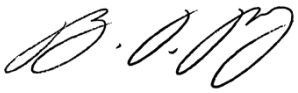
February 28, 2018

JoVE Editorial Staff,

This letter is in reference to our original manuscript submission **JoVE59219 - [EMID:006dbda986a0ecac]** titled *"IDBac: An Open-Source MALDI-MS Platform for Analysis of Microbial Protein and Specialized Metabolite Data."* Note: we revised the title to read: "*Using the Open-Source MALDI TOF-MS IDBac Pipeline for Analysis of Microbial Protein and Specialized Metabolite Data*." Herein we submit a revised manuscript.

Below is a detailed response to each reviewer critique. We only included major issues as comments below; all grammatical or minor structural changes were made per request of the reviewer. Reviewer comments are indented and in blue font, while our responses are left aligned. If you require any further material or wish to discuss this over the phone, we would be happy to speak with you. Thank you for your time and consideration,

Brian T. Murphy, Associate Professor
Department of Medicinal Chemistry and Pharmacognosy
Center for Biomolecular Sciences
University of Illinois at Chicago
P: 312-413-9057; E: btmurphy@uic.edu



Laura M. Sanchez, Assistant Professor
Department of Medicinal Chemistry and Pharmacognosy
Center for Biomolecular Sciences
University of Illinois at Chicago

Response to reviewer comments:

> The authors briefly mention in the discussion, "Performing the full analysis on isolated colonies will provide the best results, as mixed colonies/biofilms may provide spurious peaks that interfere with analyses." While that is the case, some bacteria do not produce all of their specialized metabolites when grown in isolation but may produce them in coculture - couldn't this pipeline then could also be used to analyze the impacts of mixed colonies on the specialized metabolites? I therefore wonder whether these 'spurious peaks' are potentially biologically interesting results, or whether the authors have reason to believe that they are in fact some sort of artifact?

We agree and removed the sentence as we thought it may confuse readers. However, to answer the question- yes, IDBac could be/is being used to study the effect of culture conditions on metabolite production.

> In the procedure, under number 5: Cleaning the target plate the protocol for cleaning the target plates should be included, and the reference just left to the Sauer paper: the article referenced for the cleaning is not an open-access article so not everyone will be able to view it. It would therefore be best to simply include a brief overview of the preferred/recommended ways to clean the plates for the novice user.

We have included a general method for cleaning the plates. We left the reference to the article as we adapted their method.

> Table 1 (which was to include the general parameters for data acquisition) was not included.

We corrected this.

> In addition, it is not clear how challenging it will be to use data collected on instruments other than the Bruker Autoflex - when I tried to input some Microflex files into IDBac, they were not in the correct format for preprocessing (i.e. "The Raw data file will be one folder that contains individual folders for each MALDI plate…" etc.) It is possible I was just using older data that doesn't conform to this format, and newly collected data would (perhaps especially if collected according to the missing Table 1?) but will this be a broader issue for data collected from other MALDI instruments/software? Is there a workaround that involves creating these folders? If so, what are the required parameters to facilitate this?

It can be difficult to maintain compatibility across vendor software versions. To help ameliorate this we have added the ability to use mzXML, mzML, txt, and csv files as input and will incorporate more input-types when users request them.