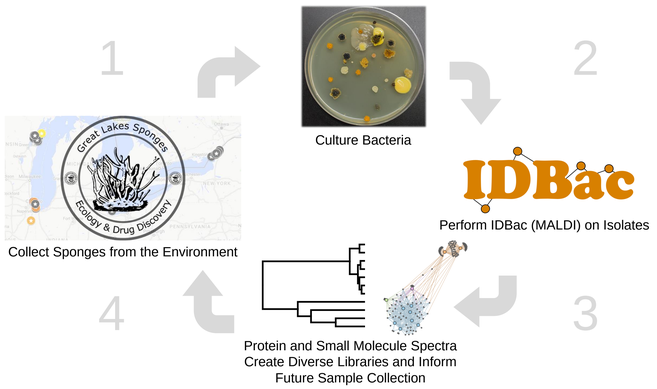
IDBac as a Tool to Evaluate Environmental Bacterial Collections in a Single Week

Chase M. Clark, Michael Mullowney, Antonio Hernandez, Emma Ward, Milan Patel, Laura M. Sanchez and Brian T. Murphy
University of Illinois at Chicago, Chicago, IL, USA

Project Overview



IDBac: A Tool to Understand Geographic Distribution of Bacterial Specialized Metabolites

Our lab recently created IDBac, which uses matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and custom software to allow fingerprinting of up to 384 strains in four hours by a single user.[1] We are currently applying IDBac to a growing collection of in-house strains from the understudied freshwater sponge *Eunapius fragilis* var. *minuta*. IDBac provides the opportunity to compare populations of bacteria between sponges and study the effects of geographic location on microbial and specialized metabolite populations.

The geographic distribution of bacteria and specialized metabolites is still not totally clear.

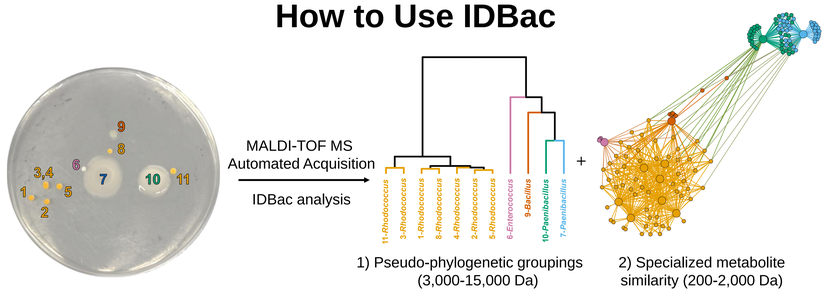
- Zach Charlop-Powers et al. suggested through metagenomic sequencing:
  - *"…that geographic distance and local environment play important roles in the sample-to-sample differences we detected in biosynthetic gene populations."[2]*
- Martina Adamek, et al. recently suggested that at the genus level:
  - *"…it can be concluded that taxonomy is a more important indicator of BGC distribution than geographic origin… In general, these data support the view that geographically distant but ecologically similar habitats share overlapping gene pools."[3]*
- Cimermancic et al. summarized their study of "…1,154 sequenced genomes spanning the prokaryotic tree of life":
  - *"We observe many nodes of high diversity in the tree closer to the leaves, pointing to evolution independent of phylogeny, perhaps indicative of ecologically driven diversification."[4]*

We postulate that everyone above is correct. Some bacterial genera likely have higher vertical transmission of biosynthetic gene clusters than horizontal, and some vice-versa, with horizontal gene gain and loss being largely driven by ecological speciation. **Therefore, we developed IDBac as a tool to rapidly investigate questions related to the interplay between phylogeny, metabolism and ecology through visualizing data driven patterns. Here, we use the cultivatable microbiome of understudied freshwater sponges as a model.**

References

1. Clark CM, Costa MS, Sanchez LM, Murphy BT (2018) Coupling MALDI-TOF mass spectrometry protein and specialized metabolite analyses to rapidly discriminate bacterial function. *Proc Natl Acad Sci USA* 115(19):4981–4986. doi:10.1073/pnas.1801247115
2. Adamek M, et al. (2018) Comparative genomics reveals phylogenetic distribution patterns of secondary metabolites in Amycolatopsis species. *BMC Genomics* 19(1):426.
3. Charlop-Powers Z, et al. (2015) Global biogeographic sampling of bacterial secondary metabolism. eLife 4:e05048.
4. Cimermancic P, et al. (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* 158(2):412–421.

How to Use IDBac



Bacterial Sources Collected for this Study

Citizen scientist divers collected two *Eunapius fragilis* var. *minuta* freshwater sponges from shipwrecks, approximately 1.3 km apart. Each sample was plated onto nine diverse media.

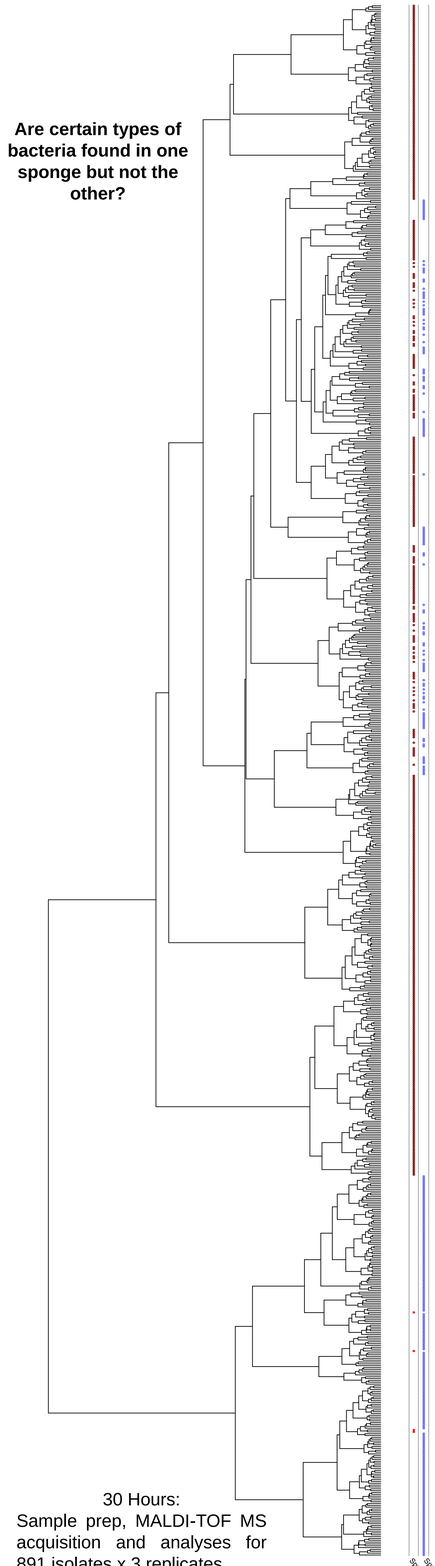
We **isolated all visible colonies over a 6 month period**. At 21 days on A1 agar, using the tip of a toothpick, isolates were smeared onto a MALDI target plate and both protein and specialized metabolite were acquired on a Bruker Autoflex Speed LRF.

Easily Visualize Entire Collection Trips at Once

All visualizations in the middle, and most in the right-panes, were made within the IDBac software

= Isolates From *E. fragilis* Sponge A
= Isolates From *E. fragilis* Sponge B

Are certain types of bacteria found in one sponge but not the other?



30 Hours:
Sample prep, MALDI-TOF MS acquisition and analyses for 891 isolates x 3 replicates.

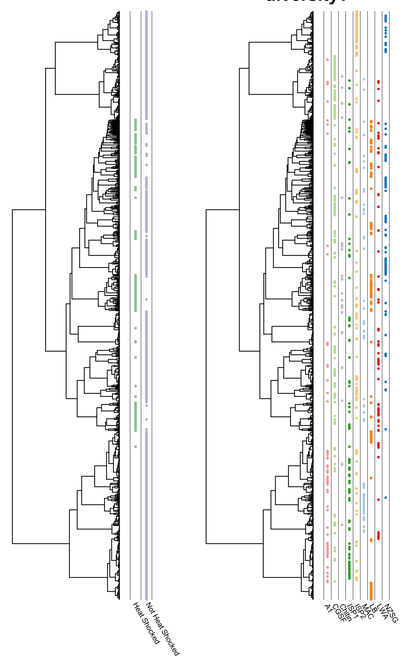
How to Read the Graphs

These dendrograms (trees) show the relatedness of MALDI MS protein fingerprints for each isolate. Each of the 900 tips on the right side of the dendrogram represent a single bacterial isolate.

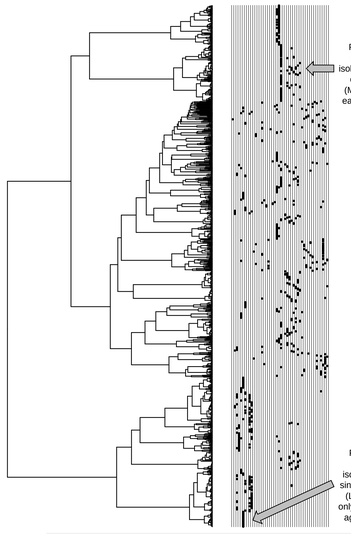
Using IDBac, data for each isolate was mapped next to the dendrogram and represented by a colored circle. The columns of circles represent a user-defined parameter (e.g. media type, source, etc).

How much diversity is lost by pretreating a sample by heat-shock?

Does nutrient media type significantly affect bacterial diversity?



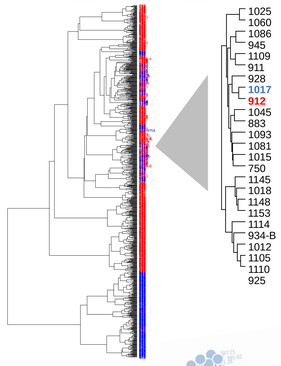
How often are phylogenetically similar isolates observed across all diversity plates?



Suggestions for Reporting Protein Analysis:
This dendrogram was created by analyzing 891 samples, and retaining peaks with a signal to noise ratio above 4 and occurring in greater than 50% of replicate spectra. Peaks occurring below 3,000 *m/z* or above 15,000 *m/z* were removed from the analyses. For clustering spectra, cosine distance and ward.D2 algorithms were used.

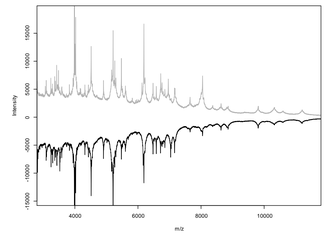
Do Related Bacteria from Different Locations Have the Same Biosynthetic Capacity?

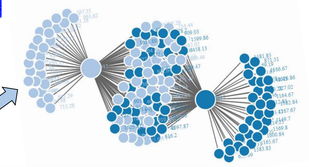
MALDI Protein Data



Here I show two bacteria isolated from the same species of sponge, collected 1 km apart. While both isolates have nearly-identical protein profiles, their specialized metabolite profiles are strikingly different. This highlights the importance of collecting protein and small-molecule spectra when building chemically-diverse strain libraries

MALDI Protein Data





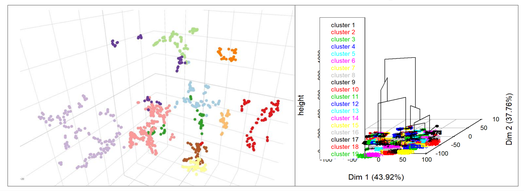
MALDI Metabolite Association Network (Specialized Metabolite Data)

Visualizing Metabolite Comparisons of 900 Isolates

The next version of IDBac will provide the ability to compare large numbers of small molecule fingerprints. While preliminary results suggest this 200-2,000 *m/z* range groups isolates similar to phylogeny, it also preserves the ability to distinguish minor differences in specialized metabolites.

Small molecule t-SNE

Clustering of t-SNE data

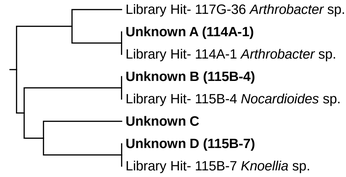


MALDI Spectra Libraries

The next version of IDBac will incorporate the ability to create shareable spectra libraries, perform library searches, and seeding (see figure to right). This will support our recently-funded R01 in which we will "digitize" over 8,000 strains from the ARS bacterial culture collection, among other collections.
This functionality and the extensive spectra database will provide users unprecedented access to understanding the identity of bacteria they are working with.

NRRL Agricultural Research Service Culture Collection

Trees of unknown isolates can be seeded with data from a library of known strains:



It's All Open Source (Free)

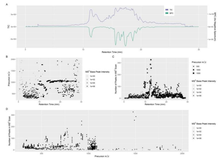
and Designed Keeping Non-Programmers in Mind

IDBac

A MALDI-TOF MS Protein and Small Molecule Bioinformatics Platform
Download at: chasemc.github.io/IDBac

mzEasy
A Simple Interface to MSConvert and Automated Plotting
DOWNLOAD MZEASY

I also recently wrote mzEasy, an easy-to-use interface for converting raw LC-MS/MS data to mzXML for use with GNPS. It automatically produces summarizing plots, to broadly assess quality and give perspective of a run.
Download at: chasemc.github.io/mzEasy



Acknowledgments

Ken Kozen and Joseph Dudiak for collection of the freshwater sponges
Sanchez lab members for MALD-TOF MS instrument maintenance
Jhewelle Fitz-Henley for help with the freshwater sponge project

Funding

National Geographic: CP-044R-17
UIC Startup Funds

@ChasingMicrobes
chasemc.github.io
murphylabuic.com