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 Reykjavík, 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 acyl-desferrioxamines 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. Therefore, developing techniques to rapidly characterize and discriminate between bacteria has been paramount to these efforts. In the past four decades, sequencing of the 16S ribosomal RNA (rRNA) gene has been instrumental to the classification of bacteria due to its widespread presence in the kingdom, degree of conservation, and length (1, 2). This and other genetic-based approaches, such as pulsed field gel electrophoresis, multilocus sequence typing, and DNA–DNA hybridization, have become commonplace (3). However, several limitations to these techniques include cost, turnaround time needed for sequencing and/or analysis, narrow windows of "universal" primers, and, in some cases, low species-level phylogenetic resolution. Most importantly, in the majority of cases, these methods are unable to elucidate how microorganisms interact with one another and function in situ. To address this shortcoming, we developed a pipeline that allows rapid discrimination of bacteria based on mass spectral signatures of specialized metabolite production (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.

Author contributions: C.M.C., M.S.C., L.M.S., and B.T.M. designed research; C.M.C. and M.S.C. performed research; C.M.C., M.S.C., L.M.S., and B.T.M. analyzed data; and C.M.C., M.S.C., L.M.S., and B.T.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Mass spectra data have been deposited in MASSIVE for MALDI (ftp://massive.ucsd.edu/MSV000081619) and LC-MS/MS (ftp://massive.ucsd.edu/MSV000081555). The 16S rRNA sequences of environmental bacteria used in this study were deposited in GenBank (see *SI Appendix*, Table S2) (accession nos. KY858227–KY858239, KY858241–KY858247, MG188670, MG188671, and KP009553).

[1]C.M.C and M.S.C. contributed equally to this work.

[2]L.M.S. and B.T.M. contributed equally to this work.

[3]To whom correspondence should be addressed. Email: btmurphy@uic.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1801247115/-/DCSupplemental.

Published online April 23, 2018.

Shortly after the development of 16S rRNA gene sequencing, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was implemented as a technique to identify large biomolecules (7, 8). Subsequent innovations in instrumentation led to the ability to obtain better-resolved spectra of intact proteins in high throughput, facilitating the rapid and less-costly identification of bacteria based largely on ribosomal MS fingerprints (8–11). Bruker (12) and bioMerieux (13) have successfully applied this technology in the clinical setting, while many others have used it on relatively small strain groupings (from 10 to a few hundred isolates) in a genus- and species-specific manner to environmental and/or clinical microorganisms or to the classification of mammalian cells, as recently summarized by the reviews in refs. 14–17. Despite the aforementioned applications, these methods do not provide information on bacterial specialized metabolite production. Relatively little is known about the relationship between bacterial taxonomy and specialized metabolite production in the majority of bacteria isolated from the environment, yet these characteristics are central to researchers who study bacteria in both academic and industrial settings. Thus, a comprehensive analytical pipeline that allows simultaneous analyses of these factors has been a major obstacle to correlating microbial identity with specialized metabolite production.

In this study, we present a significant innovation to described MS methods that analyze bacteria by utilizing the full capabilities offered by MALDI-TOF mass spectrometers. In addition to linear mode protein analysis, we use reflectron mode to analyze specialized metabolites, as the combination of both information-rich spectral regions has yet to be applied to existing MALDI-TOF MS analysis pipelines. Silva et al. (18) recently provided a comprehensive history detailing the underuse of MALDI MS to analyze specialized metabolites and the shortage of software to relieve current bioinformatics bottlenecks.

To validate our pipeline, we demonstrate the ability to differentiate isolates within closely related species groupings, often on colonies that share indistinguishable morphology, and characterize them based on in situ antibiotic (plipastatin), siderophore (desferrioxamine), and motility factor (surfactin) production. Total acquisition, analysis, and visualization of MALDI-TOF MS data from both intact proteins and specialized metabolites of up to 384 bacterial colonies can be performed in <4 h. Compared with other instrumentation that is commonly used to analyze specialized metabolites, such as quadrupole TOF, Orbitrap, and Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers, MALDI-TOF MS requires minimal expertise to operate and is accompanied by facile sample preparation protocols. Our method provides an alternative to laborious liquid cultivation, metabolite extraction, and chromatographic experiments that are the current standard of practice for studies that focus on specialized metabolite analysis.

Results

To visualize relational patterns between bacterial isolates, we created a bioinformatics pipeline designed to facilitate the multistage analysis of protein and specialized metabolite MS data. Generally speaking, we analyzed MS fingerprints of intact proteins (3,000–15,000 Da) and specialized metabolites (200–2,000 Da) in consecutive MALDI-TOF MS linear and reflectron mode acquisitions, respectively. Principal components analysis and hierarchical clustering of protein spectra placed bacterial isolates into putative genus- and species-level groups. Isolates within each grouping were then further discriminated based on differences in specialized metabolite production through analysis of Metabolite Association Networks (MANs; see *SI Appendix, Text S1* for a description). The IDBac software, written in R, is available for download and installation via a simple Windows installer (*Publication Code and Data Availability*). The software was designed for simplicity and ease of use with documentation provided at each step in the workflow that at the time of publication provides inverted/mirror plots for comparison, hierarchical clustering, principle components analysis, and MAN analysis. However, whereas similar software has required users to convert raw data on their own and format these to software-dependent configurations, IDBac takes raw data as input. This provides a transparent and repeatable data-handling process that is less prone to user error or data-tampering and, as a byproduct, bundles spectra by sample into the widely accepted and readily shareable mzML format. The IDBac software is available under a GNU General Public License, and links to the full code along with data acquisition and analysis tutorials are within *Publication Code and Data Availability*.

Although our platform is intended to visualize interspecies and intraspecies differences in specialized metabolite production between colonies rather than to identify the precise chemical structure of excreted specialized metabolites, it was important to validate that specific nodes (*m*/*z* features) in our MANs represented bacterial chemistry as opposed to matrix peaks, media components, or instrument noise. After preprocessing and peak-picking our MALDI-TOF MS spectra, we accounted for the matrix and media ionizable compounds by subtracting a reference peak list for a matrix and media control during each experimental run. To account for instrument noise, we precluded signals from the peak picking algorithm that fell below a user-defined signal-to-noise ratio (i.e., 4:1) and retained peaks occurring in a minimum of 70% of replicates. We demonstrated the ability of IDBac to use specialized metabolite production to distinguish between the nearly identical *Bacillus subtilis* 3610 and its frame-shift mutant *B. subtilis* PY79 (Fig. 1). The two strains were chosen based on their relation to one another: *B. subtilis* PY79 (19, 20) is deficient in the ability to produce the antibiotics surfactin and plipastatin due to a frameshift mutation in *sfp*, the 4′-phosphopantetheinyl transferase that mediates nonribosomal peptide synthetase apoform activation (21).

B. subtilis 3610 and *B. subtilis* PY79 were grown on separate A1 nutrient agar plates under identical conditions and their specialized metabolite regions (200–2,000 Da) were analyzed using MALDI-TOF MS (ten technical replicates each). The differences in specialized metabolite production between strains 3610 and PY79 were readily observed and visualized in our MAN (Fig. 1*B*), where large nodes represent bacterial colonies and smaller nodes represent *m*/*z* peaks (singly charged molecules) present in their corresponding MALDI-TOF spectra. This MAN allowed for facile visualization of both the differences in specialized metabolite spectra between strains and the peaks that were shared or unique to each strain. Four plipastatin (compounds **1**–**4**) and three surfactin (compounds **5**–**7**) analogs, including their resolved isotopologues and adducts, were detected from strain 3610, and not from PY79. Both strains shared the ability to produce partially characterized polyglutamate polymers (22), which have been implicated in species-specific functions such as virulence factor production, biofilm formation, and sequestration of toxic metal ions (22, 23). Importantly, the majority of *m*/*z* values (33 of 42; 78.6%) in the MAN represented specialized metabolites, and IDBac successfully filtered out ions associated with matrix and media components. MALDI-TOF MS analysis of the specialized metabolite region and data processing using IDBac correctly depicted subspecies antibiotic production differences between genetic variants of *B. subtilis* in <30 min.

Next, we demonstrated that intraspecies strain groupings based on MALDI-TOF MS intact protein profiles can be further discriminated based on in situ specialized metabolite production. From our in-house strain library, we selected eight *Micromonospora chokoriensis* isolates from two sediment samples collected nearly 175 km apart in Lake Michigan, US. These isolates share >99%

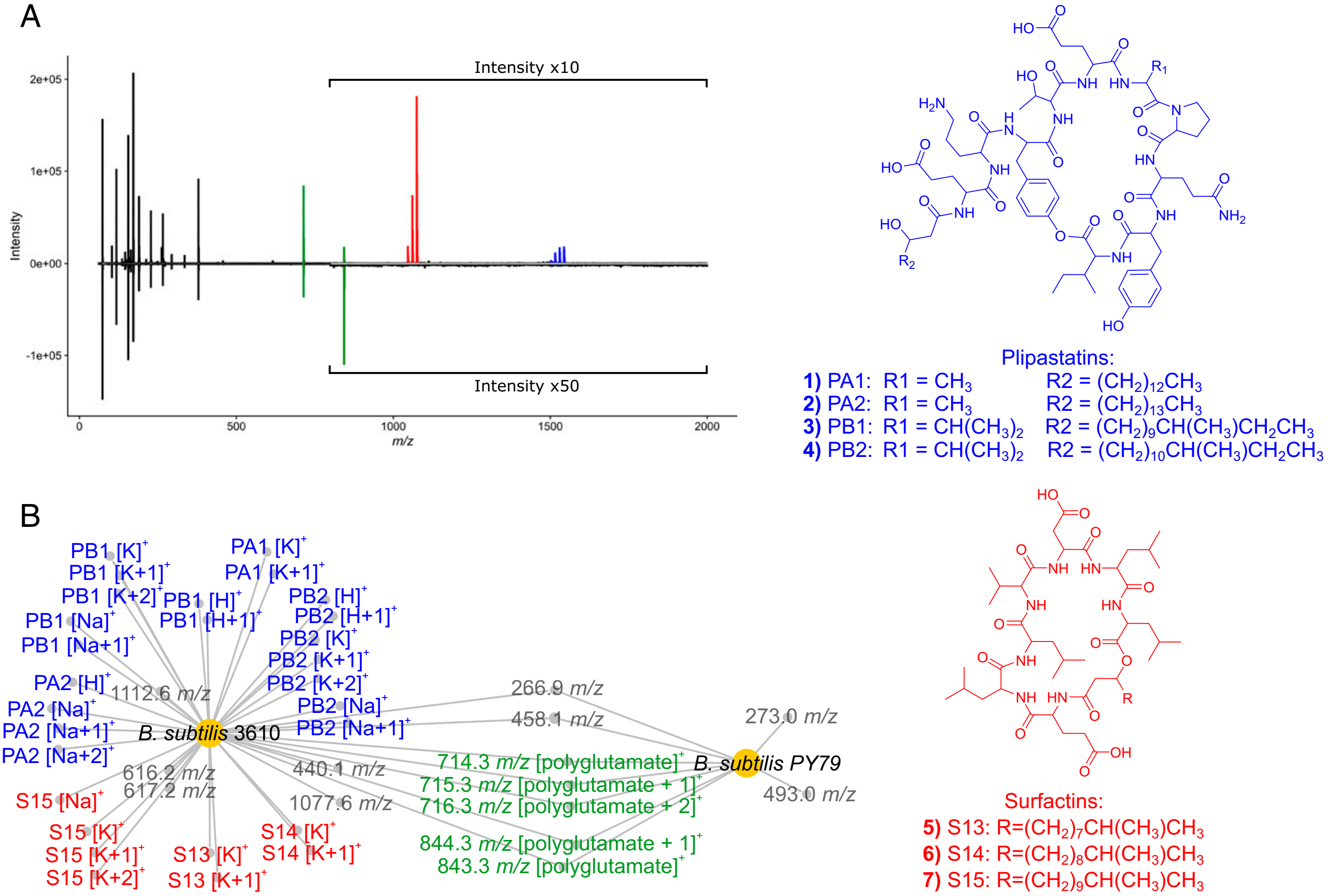

Fig. 1. Analysis of MALDI-TOF MS specialized metabolite data from two *B. subtilis* genetic variants (10 technical replicates each) shows distinct differences in plipastatin and surfactin analog production. (*A*) Inverse spectrum comparison showing representative spectra for *B. subtilis* 3610 (positive spectrum) and *B. subtilis* PY79 (negative spectrum). (*B*) MAN showing the differential production of surfactin and plipastatin antibiotic analogs between the two strains. Large nodes represent individual bacterial colonies, while smaller nodes represent individual *m/z* values in MALDI-TOF spectra that fall within our peak selection criteria. MALDI-TOF spectrum annotations can be found in *SI Appendix*, Fig. S1. Isotopologues are denoted as +1, +2. Based on their precedence in *B. subtilis* 3610 and PY79 in literature, we assigned several *m/z* peaks as polyglutamate-like polymers (22). The remaining peaks were not identified but may represent primary metabolites and/or yet-to-be-characterized specialized metabolites. Importantly, we assigned 33/42 *m/z* values (78.6%), providing evidence that the MAN is primarily composed of specialized metabolites.

16S rRNA gene sequence identity across >1,460 nucleotides (24). We cultivated biological replicates in a random pattern across a 48-well microwell plate, resulting in at least four independent biological replicates of each strain. With a minimum of eight technical replicates, we used MALDI-TOF MS to consecutively record MS spectra of the protein (3,000–15,000 *m/z*) and specialized metabolite regions (200–2,000 *m/z*). This resulted in at least 32 replicate spectra per isolate, with peaks required to be present in 70% or greater of replicates to be included in analyses (a user-defined parameter within IDBac). We observed that hierarchical clustering of MALDI-TOF MS protein data correlated strongly with 16S rRNA similarity (Fig. 2*A*), results consistent with many previous efforts that used MALDI-TOF MS as an alternative to traditional sequence-based taxonomic classification methods, as recently summarized (14, 15).

Importantly, while these MALDI-TOF MS protein fingerprints correlated nearly identically with corresponding phylogenetic groupings (Fig. 2*A*), differences in specialized metabolite production existed and could not be predicted through 16S rRNA phylogenetic analyses or geographic strain distribution patterns. However, analysis of IDBac's MAN rapidly discerned subtle, but significant, variations in bacterial chemistry within these closely related *Micromonospora* isolates (Fig. 2*B*), allowing us to assess the relationship of specialized metabolite production to both strain phylogeny and geographic origin. A major distinguishing pattern exhibited by seven of eight *M. chokoriensis* isolates was a group of features between 650 and 800 *m/z*. This pattern was characterized by successive 14-Da differences beginning at 673.5 *m/z* and extending to 771.6 *m/z*, along with sodium and potassium adducts of each, which we attributed to analogs differing in the addition/subtraction of methylene groups (*SI Appendix*, Fig. S2). An outlier strain, B001, did not share these features, indicating differential specialized metabolite production within this group of highly similar strains.

To further validate the differential specialized metabolite production observed in the MAN, we employed HPLC tandem MS (HPLC-MS/MS) analysis for evaluation with the Global Natural Products Social molecular networking platform (25) and comparative metabolomics XCMS analysis (26). HPLC-MS/MS afforded an orthogonal means of analysis through chromatographic separation, a different mechanism of ionization, and tandem mass data that helped verify relationships between observed metabolites. Using GNPS, XCMS, and simulated spectra for iron isotopologues (*SI Appendix*, Figs. S3–S7), we identified a series of acylated desferrioxamine analogs (compounds **8**–**12**). These belong to a class of siderophores that sequester the essential growth factor ferric iron from the environment. One of

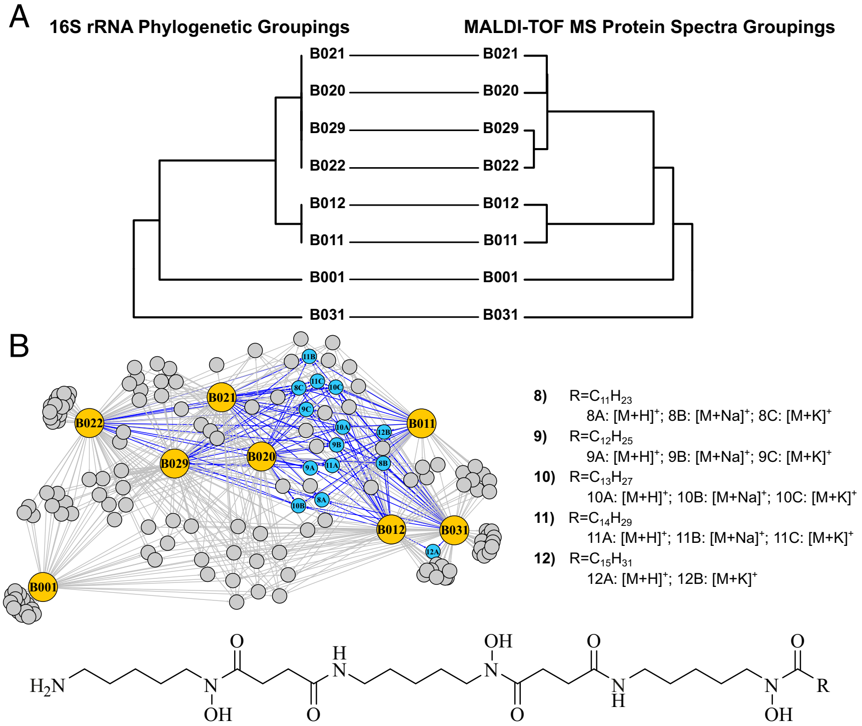
A
16S rRNA Phylogenetic Groupings
MALDI-TOF MS Protein Spectra Groupings
B021
B020
B029
B022
B012
B011
B001
B031
B
B022
B021
B011
B029
B020
B012
B031
B001
8) R=$C_{11}H_{23}$
8A: $[M+H]^+$; 8B: $[M+Na]^+$; 8C: $[M+K]^+$
9) R=$C_{12}H_{25}$
9A: $[M+H]^+$; 9B: $[M+Na]^+$; 9C: $[M+K]^+$
10) R=$C_{13}H_{27}$
10A: $[M+H]^+$; 10B: $[M+Na]^+$; 10C: $[M+K]^+$
11) R=$C_{14}H_{29}$
11A: $[M+H]^+$; 11B: $[M+Na]^+$; 11C: $[M+K]^+$
12) R=$C_{15}H_{31}$
12A: $[M+H]^+$; 12B: $[M+K]^+$
H_2N
N
OH
O
H
N
OH
N
O
O
N
H
O
N
OH
R

Fig. 2. IDBac protein and specialized metabolite analysis of isolates from a single *Micromonospora* species. (*A*) Tanglegram depicts a high degree of similarity between groupings of 16S rRNA gene sequence identity and MALDI-TOF MS protein data of *M. chokoriensis* isolates. (*B*) MAN of MALDI-TOF MS data from *M. chokoriensis* colonies highlights distinct intraspecies differences in specialized metabolite production; this is due to differential production of a specific series of acylated desferrioxamine siderophores (shown as blue nodes), which B001 did not produce.

these, desferrioxamine B, is used clinically to treat metal poisoning. Interestingly, according to our analyses, B001 lacks the capacity to produce these acylated analogs (Fig. 2*B*), and this was readily highlighted upon analysis of the MAN. We also observed several other desferrioxamine analogs produced by these isolates; links to these data are available in *SI Appendix*, Table S1. Of note, while the phylogenetic identity of *Micromonospora* isolate B031 is more closely related to B001, it overlaps the remaining six isolates in its ability to produce acylated desferrioxamines, highlighting the importance of our method to provide a means other than phylogeny to distinguish between in situ function of similar strains.

The HPLC-MS/MS–based observation of strain-specific patterns of siderophore production corroborated our initial MALDI-TOF MS results that showed B001, while able to produce unmodified desferrioxamine B, was deficient in the ability to produce a series of acylated desferrioxamine B analogs (compounds **8**–**12**). This is significant, as B001 was isolated from the same 1-cm^3 sediment sample as six of the seven strains that produced this compound series, highlighting that phylogenetic groupings and geographic location are not sufficient indicators of specialized metabolite production capacity. Our observations of desferrioxamine biosynthetic pathway promiscuity were consistent with previous studies (27–29). However, each of these studies required extensive genome sequencing and/or liquid fermentation experiments, followed by chromatographic analyses, whereas analysis through MALDI-TOF MS/IDBac was able to visualize putative phylogenetic relationships and intraspecies differences in specialized metabolism in a few hours, once each colony appeared on a Petri dish.

In many instances, the taxonomic identity and specialized metabolite production capacity of a group of bacteria are not well characterized or completely unknown, particularly in studies involving bacteria isolated from the environment. Thus, we tested the ability of the IDBac analysis pipeline to rapidly extract protein and specialized metabolite information from unknown environmental bacteria cultivated from a freshwater sponge collected in Lake Michigan. We cultivated sponge-associated bacteria from 1 cm^3 of tissue onto high-nutrient A1 medium (Fig. 3*A*; see *SI Appendix* for further details). Using a sterile toothpick, we selected all colonies that grew on the plate over a 90-d time period. These colonies were then subjected to MALDI-TOF MS analysis with 10 technical replicates, and the data were processed in IDBac. Principal components analysis and unsupervised hierarchical clustering of the bacterial protein range afforded four distinct groupings (Fig. 3*B*). We confirmed that these MS protein groupings aligned with the genera *Enterococcus*, *Bacillus*, *Paenibacillus*, and *Rhodococcus* via 16S rRNA gene sequencing analysis of each of isolate (*SI Appendix*, Figs. S8 and S9).

Next, we generated a MAN (Fig. 3*C*) to further discriminate these groups based on specialized metabolite production. Interestingly, unsupervised subnetworking modularity analysis (30, 31) separated the network into five groups of strains (colored in Fig. 3*C*), which highly correlated to phylogenetic groupings. The seven *Rhodococcus* isolates exhibited a high degree of specialized metabolite associations, while the *Bacillus* and *Enterococcus* isolates were classified as unique via modularity scoring. Importantly, our method quickly highlighted subtle differences in specialized metabolite production between two morphologically identical *Paenibacillus* strains and separated the two strains whose 16S rRNA gene sequences shared a pairwise similarity of 99.87% (1,496 of 1,498 nucleotides) (24). A careful look at their specialized metabolite profiles showed subtle differences in production of a series of specialized metabolite(s) ranging from 900 to 1,250 *m*/*z*. We wanted to confirm that this difference was not due to variations in colony microenvironment, since nearby colonies can affect specialized metabolite production through physical contact and

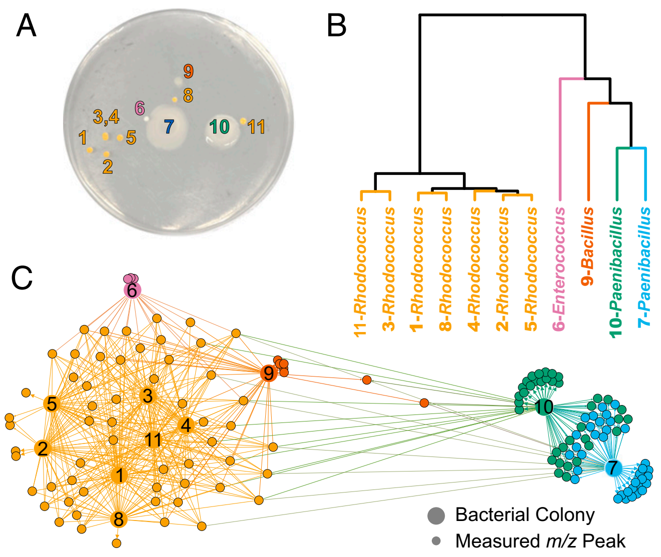

Fig. 3. Rapid protein and specialized metabolite fingerprinting of unknown environmental isolates using MALDI-TOF MS and IDBac. (*A*) Bacterial diversity plate obtained from placing freshwater sponge tissue on high-nutrient A1 agar. (*B*) IDBac allowed for hierarchical clustering of MALDI-TOF MS protein spectra with the option to choose standard distance measures and clustering algorithms. For workflows requiring analysis of hundreds to thousands of strains, protein grouping is essential for data reduction before specialized metabolite analysis is performed. (*C*) MAN, colored via modularity analysis with default thresholds in Gephi (31), allowed for rapid decision-making based on gross in situ specialized metabolite production after matrix and media signals were subtracted automatically from the network in IDBac. Significant outliers were the two *Paenibacillus* strains, which produced several shared and unique high molecular-weight specialized metabolites.

chemical cross-talk (22, 32, 33). To determine this, we grew each *Paenibacillus* isolate by itself over three individual cultivation experiments and MALDI-TOF MS data acquisition events, and after observation of previously observed specialized metabolite patterns in the 900–1,250 *m/z* range, we concluded that both isolates contained partially overlapping, but distinct, metabolic capacities (*SI Appendix*, Fig. S10). These subtle differences in metabolic capacity were readily detected and visualized as a result of the IDBac MAN and could not have been achieved through analysis of protein groupings alone. In just 3 h, our pipeline created statistically robust protein profile groupings of environmental isolates (11 strains, with 10 MS replicates each, for a total of 110 MALDI spots). It further distinguished colonies with nearly identical morphology and 16S rRNA gene sequence identity based on their capacity to produce specialized metabolites.

Discussion

One aspect of the IDBac pipeline that provides significant advantages over existing MS platforms is that it affords information on putative colony phylogeny and specialized metabolite production without the need for extraction and chromatographic analyses. The latter techniques rely on growing pure bacterial isolates (often in liquid culture), generating extracts, and using LC analyses (generally coupled to MS) to separate and detect specialized metabolites. This is often a laborious, costly, and relatively time-consuming process. Our pipeline is an extraction-free process and is ideally suited for researchers who aim to (*i*) compare functional chemistry between closely related isolates (e.g., comparing antibiotic production between two *B. subtilis* genetic variants), (*ii*) probe relationships between taxonomic identity and environmental functionality (e.g., studying intraspecies differences in *M. chokoriensis* siderophore production), or (*iii*) assess the broad relatedness of unknown environmental bacterial isolates (e.g., visualizing phylogenetic and metabolic relatedness within a group of unknown bacteria). The latter points highlight a strength of IDBac as an engine to generate research questions and hypotheses. For example, why have similar *Micromonospora* strains from the same 1 cm^3 of sediment (*SI Appendix*, Table S2) evolved differing capacities to scavenge iron? Is there any correlation between geographic location and desferrioxamine production in an expanded set of Lake Michigan sediment-derived *Micromonospora* isolates? Is there a synergistic role between the specialized metabolites that is unique to two phylogenetically similar *Paenibacillus* strains isolated from the same sponge? MALDI-TOF MS/IDBac analysis allows for easy visualization of global specialized metabolite patterns and, as a result, gives researchers the opportunity to ask these questions based on rapidly generated data visualizations.

IDBac is complementary to other innovative platforms such as GNPS, which aids in the dereplication of previously characterized and identification of potentially new specialized metabolites (25, 34). Toward this point, it is important to note that MALDI-TOF MS and various LC-MS systems use distinct modes of ionization and give rise to different observed features that hinder direct comparison of experimental data. Our work highlights the individual strengths of orthogonal MS-based approaches to extract complementary information from a biological system.

Through the course of our studies, we have documented a few potential limitations of our method. First, although IDBac can create accurate subspecies groupings of bacteria based on protein MS fingerprints, species-level identification is only possible in the presence of a searchable and extensive protein fingerprint database. A greater community effort is required to document MALDI-TOF MS protein fingerprints into a publicly available database to maximize phylogenetic coverage. Such an effort would facilitate the rapid identification of unknown environmental bacteria and elevate this process to be on par with existing commercial platforms used to identify clinical pathogens (14). An effort to make a readily searchable/freely available bacterial protein MS database is ongoing in our laboratory, and the first version of IDBac supports this mission by providing a simple interface for bundling and converting proprietary vendor-format data files to the open-access mzML format for easy, reproducible sharing.

A second limitation is that bacterial specialized metabolite production is sensitive to external factors such as growth medium, temperature, unintended microbial contamination, and the proximity of other microorganisms on a Petri dish (35). For comparison of specialized metabolite profiles to be reliable, it is imperative that strains are cultivated under identical conditions and that several biological/technical replicates are acquired for each isolate (both precautions were taken in this study). Fortunately, the throughput of our pipeline allows for multiple replicates of a sample to be analyzed rapidly. Conversely, MALDI-TOF MS protein fingerprints of bacterial colonies are robust and exhibit minimal fluctuation when the data are acquired on different growth media (*SI Appendix*, Fig. S11).

Finally, our method is limited by the resolution of the MALDI-TOF MS and would be improved through use of instruments with higher resolving power capabilities. High-resolution instruments, such as MALDI-FT-ICR and fragmentation data from a MALDI-TOF/TOF MS, could aid in preliminary compound class dereplication via generation of accurate molecular formulas and structural information, respectively. However, this would increase the length of the experiments and data storage requirements and would require additional technical expertise.

IDBac couples bacterial protein and specialized metabolite MS data to rapidly discriminate between isolates based on both their identity and potential environmental function. Our MS pipeline addresses a need in research communities that study microbial function (e.g., chemical ecology, pathogenesis, taxonomy, drug discovery, agriculture, and food safety). The pipeline is faster than existing MS methods used to analyze nonvolatile metabolite production within microorganisms and requires less technical experience to operate. The experiment requires 10^5 to 10^7 bacterial cells (36)—or approximately the tip of a toothpick—to generate MS profiles and is designed to be performed in high-throughput. Acquisition of MALDI-TOF MS data from both intact proteins and specialized metabolites of single bacterial colonies has been integrated in one single freely available bioinformatics pipeline. This work reports on coupling protein and specialized metabolite MS data in a semiautomated pipeline to make the rapid discrimination of bacteria accessible to the broad research community.

Methods

MALDI-TOF MS Sample Preparation and Data Acquisition. For MALDI-TOF MS analysis, proteins were extracted by using an extended direct transfer method that included a formic acid overlay (36, 37). Measurements were performed in linear and reflectron modes by using an Autoflex Speed LRF mass spectrometer (Bruker Daltonics) equipped with a smartbeam-II laser (355 nm). Detailed instrument settings are available in *SI Appendix*.

MALDI-TOF Data Bioinformatics Pipeline. In brief, MALDI-TOF MS raw data were first converted to the open-source mzML format (38) by using ProteoWizard's command-line MSConvert (39). The mzML files were read into R through mzR (39) and custom code and processed into peak lists by using MALDIquant (40). Lastly, the peak lists informed interactive data analyses and visualizations that were displayed through a default, offline internet browser with RStudio's Shiny (41) package. The IDBac software is being distributed under a GNU General Public License, and links to the full code along with data acquisition and analysis tutorials can be found within *Publication Code and Data Availability*.

Publication Code and Data Availability. IDBac software and updates, method manuals, and information on code contribution are available at chasemc.github.io/IDBac and permanent snapshots at https://zenodo.org/record/1115619. The flexControl, flexAnalysis, and R code used to generate all figures and analyses for this publication along with the publication-version of IDBac is available for download at https://data.mendeley.com/datasets/ysrtr9c5s7/1. Mass spectra were deposited in MASSIVE (accession nos. MSV000081619 and MSV000081555).

ACKNOWLEDGMENTS. We thank the following contributors: Russel Cuhel and crew of R. V. Neeskay (University of Milwaukee Wisconsin) for assistance with sediment collection; Keith Jung for collection of the freshwater sponge; Dr. Amanda Bulman (Bruker) 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); and Dr. Darrell McCaslin (University of Wisconsin Madison) for testing our method on legacy instruments. This work was supported by National Institute of Child Health and Human Development Grant K12HD055892; the National Institutes of Health Office of Research on Women's Health; National Geographic Grant CP-044R-17; Icelandic Research Fund Grant 152336-051; and University of Illinois at Chicago startup funds.

1. Woese CR (1987) Bacterial evolution. *Microbiol Rev* 51:221–271.
2. Yarza P, et al. (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* 12:635–645.
3. Vandamme P, et al. (1996) Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev* 60:407–438.
4. Patin NV, Duncan KR, Dorrestein PC, Jensen PR (2016) Competitive strategies differentiate closely related species of marine actinobacteria. *ISME J* 10:478–490.
5. Ziemert N, et al. (2014) Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *Proc Natl Acad Sci USA* 111:E1130–E1139.
6. Penn K, et al. (2009) Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *ISME J* 3:1193–1203.
7. Karas M, Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 60:2299–2301.
8. Tanaka K, et al. (1988) Protein and polymer analyses up to *m/z* 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 2:151–153.
9. Sandrin TR, Goldstein JE, Schumaker S (2013) MALDI TOF MS profiling of bacteria at the strain level: A review. *Mass Spectrom Rev* 32:188–217.
10. Cain TC, Lubman DM, Weber WJ, Vertes A (1994) Differentiation of bacteria using protein profiles from matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 8:1026–1030.
11. Holland RD, et al. (1996) Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 10:1227–1232.
12. Maier T, Klepel S, Renner U, Kostrzewa M (2006) Fast and reliable MALDI-TOF MS–Based microorganism identification. *Nat Methods* 3:68–71.
13. Dubois D, et al. (2012) Performances of the Vitek MS matrix-assisted laser desorption ionization-time of flight mass spectrometry system for rapid identification of bacteria in routine clinical microbiology. *J Clin Microbiol* 50:2568–2576.
14. Rahi P, Prakash O, Shouche YS (2016) Matrix-assisted laser desorption/ionization time-of-flight mass-spectrometry (MALDI-TOF MS) based microbial Identifications: Challenges and scopes for microbial ecologists. *Front Microbiol* 7:1359.
15. Popović NT, Kazazić SP, Strunjak-Perović I, Čož-Rakovac R (2017) Differentiation of environmental aquatic bacterial isolates by MALDI-TOF MS. *Environ Res* 152:7–16.
16. Munteanu B, Hopf C (2013) Emergence of whole-cell MALDI-MS biotyping for high-throughput bioanalysis of mammalian cells? *Bioanalysis* 5:885–893.
17. Cassagne C, Normand A-C, L'Ollivier C, Ranque S, Piarroux R (2016) Performance of MALDI-TOF MS platforms for fungal identification. *Mycoses* 59:678–690.
18. Silva R, Lopes NP, Silva DB (2016) Application of MALDI mass spectrometry in natural products analysis. *Planta Med* 82:671–689.
19. Zeigler DR, et al. (2008) The origins of 168, W23, and other *Bacillus subtilis* legacy strains. *J Bacteriol* 190:6983–6995.
20. Schroeder JW, Simmons LA (2013) Complete genome sequence of *Bacillus subtilis* strain PY79. *Genome Announc* 1:e01085-13.
21. Stein T (2005) *Bacillus subtilis* antibiotics: Structures, syntheses and specific functions. *Mol Microbiol* 56:845–857.
22. Yang Y-L, et al. (2011) Connecting chemotypes and phenotypes of cultured marine microbial assemblages by imaging mass spectrometry. *Angew Chem Int Ed Engl* 50: 5839–5842.
23. Candela T, Fouet A (2006) Poly-gamma-glutamate in bacteria. *Mol Microbiol* 60: 1091–1098.
24. Yoon S-H, et al. (2017) Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol* 67: 1613–1617.
25. Wang M, et al. (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* 34: 828–837.
26. Gowda H, et al. (2014) Interactive XCMS online: Simplifying advanced metabolomic data processing and subsequent statistical analyses. *Anal Chem* 86:6931–6939.
27. Arias AA, et al. (2015) Growth of desferrioxamine-deficient *Streptomyces* mutants through xenosiderophore piracy of airborne fungal contaminations. *FEMS Microbiol Ecol* 91:fiv080.
28. D'Onofrio A, et al. (2010) Siderophores from neighboring organisms promote the growth of uncultured bacteria. *Chem Biol* 17:254–264.
29. Bruns H, et al. (2018) Function-related replacement of bacterial siderophore pathways. *ISME J* 12:320–329.
30. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech* 2008:P10008.
31. Bastian M, Heymann S, Jacomy M (2009) Gephi: An open source software for exploring and manipulating networks visualization and exploration of large graphs. *International AAAI Conference on Weblogs and Social Media* (Association for the Advancement of Artificial Intelligence, Menlo Park, CA), pp 361–362.
32. Gonzalez DJ, et al. (2012) Observing the invisible through imaging mass spectrometry, a window into the metabolic exchange patterns of microbes. *J Proteomics* 75: 5069–5076.
33. Yang Y-L, Xu Y, Straight P, Dorrestein PC (2009) Translating metabolic exchange with imaging mass spectrometry. *Nat Chem Biol* 5:885–887.
34. Yang JY, et al. (2013) Molecular networking as a dereplication strategy. *J Nat Prod* 76: 1686–1699.
35. Zarins-Tutt JS, et al. (2016) Prospecting for new bacterial metabolites: A glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products. *Nat Prod Rep* 33:54–72.
36. Freiwald A, Sauer S (2009) Phylogenetic classification and identification of bacteria by mass spectrometry. *Nat Protoc* 4:732–742.
37. Schumann P, Maier T (2014) MALDI-TOF mass spectrometry applied to classification and identification of bacteria. *Methods Microbiol* 41:275–306.
38. Martens L, et al. (2011) mzML–A community standard for mass spectrometry data. *Mol Cell Proteomics* 10:000133.
39. Chambers MC, et al. (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* 30:918–920.
40. Gibb S, Strimmer K (2012) MALDIquant: A versatile R package for the analysis of mass spectrometry data. *Bioinformatics* 28:2270–2271.
41. Chang W, Cheng J, Allaire J, Xie Y, McPherson J (2016) Shiny: Web Application Framework for R (RStudio, Boston).