Development of an inexpensive matrix-assisted laser desorption—time of flight mass spectrometry method for the identification of endophytes and rhizobacteria cultured from the microbiome associated with maize

Michael G. LaMontagne[1], Phi L. Tran[1], Alexander Benavidez[2] and Lisa D. Morano[2]

[1] Department of Biology and Biotechnology, University of Houston, Clear Lake, Houston, Texas, United States

[2] Department of Natural Sciences, University of Houston, Downtown, Houston, Texas, United States

ABSTRACT

Many endophytes and rhizobacteria associated with plants support the growth and health of their hosts. The vast majority of these potentially beneficial bacteria have yet to be characterized, in part because of the cost of identifying bacterial isolates. Matrix-assisted laser desorption-time of flight (MALDI-TOF) has enabled culturomic studies of host-associated microbiomes but analysis of mass spectra generated from plant-associated bacteria requires optimization. In this study, we aligned mass spectra generated from endophytes and rhizobacteria isolated from heritage and sweet varieties of *Zea mays*. Multiple iterations of alignment attempts identified a set of parameters that sorted 114 isolates into 60 coherent MALDI-TOF taxonomic units (MTUs). These MTUs corresponded to strains with practically identical (>99%) 16S rRNA gene sequences. Mass spectra were used to train a machine learning algorithm that classified 100% of the isolates into 60 MTUs. These MTUs provided >70% coverage of aerobic, heterotrophic bacteria readily cultured with nutrient rich media from the maize microbiome and allowed prediction of the total diversity recoverable with that particular cultivation method. *Acidovorax* sp., *Pseudomonas* sp. and *Cellulosimicrobium* sp. dominated the library generated from the rhizoplane. Relative to the sweet variety, the heritage variety c ontained a high number of MTUs. The ability to detect these differences in libraries, suggests a rapid and inexpensive method of describing the diversity of bacteria cultured from the endosphere and rhizosphere of maize.

INTRODUCTION

Current agricultural practices will not meet the nutritional needs of a population that will reach nine billion people by the middle of this century (*Anand, Grayston & Chanway, 2013*). In parallel, climate change will increase extreme weather events, including drought (*Dai, 2011*; *Trenberth et al., 2014*), while urbanization will reduce arable land (*Song, Pijanowski & Tayyebi, 2015*). Microbial products can mitigate this food crisis by increasing crop yield (*Tkacz & Poole, 2015*) and helping crops tolerate drought and marginal soils (*Brígido & Glick, 2015*). The plant microbiome abounds with endophytes and plant growth promoting rhizobacteria (PGPR) that can help plants recover more nutrients from the soil and tolerate stressors like drought (*Barnawal et al., 2013*; *Bresson et al., 2014*). PGPR can also control plant pathogens (*Chowdhury et al., 2013*), promote beneficial mycorrhizae colonization (*Labbé et al., 2014*) and produce potentially valuable secondary metabolites (*Brader et al., 2014*; *Kumar, Dubey & Maheshwari, 2012*; *Raaijmakers & Mazzola, 2012*); however, the vast majority of microbes associated with agrosystems, have yet-to-be cultured. This limits the development of microbial products from PGPR and manipulative experiments with defined and representative plant microbiomes that can facilitate hypothesis testing, as demonstrated with maize (*Niu et al., 2017*).

Given its long history of artificial selection and its importance as a global food crop, the maize microbiome is of particular interest. Maize stems (*Kämpfer et al., 2016a*, *2016b*) have yielded novel bacterial species and maize kernels host diverse bacterial endophytes (*Rijavec et al., 2007*). Few studies have compared the diversity of libraries of bacteria isolated from both the rhizosphere and endosphere of maize (*McInroy & Kloepper, 1995*). Further, although Bt and non-Bt modern maize showed little differences in their bacterial communities (*Mashiane et al., 2017*), maize varieties may support different microbiomes. For example, a comparison of libraries generated from a wild maize and a more modern variety yielded three PGPR species from the wild maize that were not recovered from the modern variety (*Mousa et al., 2015*) and ancient maize yielded a bacterial endophyte that inhibits multiple fungal pathogens (*Johnston-Monje & Raizada, 2011*). A deeper understanding of interactions between maize varieties and the microbiomes they host in their endosphere and rhizosphere could inform searches for PGPR but development of deep libraries of isolates for research and microbial product development multiplies the costs associated with microbial identification.

Matrix-assisted laser desorption-time of flight (MALDI-TOF) appears well suited for analysis of libraries isolated from plant microbiomes (*Ghyselinck et al., 2013*). MALDI-TOF systems provide strain-level identification of microbes (*Ahmad, Babalola & Tak, 2012*; *Sauer et al., 2008*; *Singhal et al., 2015*) for pennies an isolate. These systems compare favorably to the widely used method of 16S rRNA sequencing (*Emami et al., 2016*); however, the costs of acquiring and operating these systems and the challenges in interpreting the data they generate have limited the widespread application of MALDI-TOF. These systems use pattern matching between mass and reference spectra; however, these databases have poor representation of environmental microbes. This paucity of representation limits MALDI-TOF effectiveness (*Singhal et al., 2015*).

Compilations of spectra are available (*Böhme et al., 2012*; *Murugaiyan et al., 2018*; *Rau et al., 2016*). These compilations are typically species-specific databases of mass spectra, including databases for *Mycobacterium kansasii* (*Murugaiyan et al., 2018*) and *Vibrio* species (*Erler et al., 2015*) or for particular systems, like spacecraft (*Seuylemezian et al., 2018*). Further, custom databases require proprietary software and do not facilitate investigators sharing mass spectra, which would benefit science. To address this, several teams have developed applications for microbial identification by MALD-TOF by matching of spectra to peaks inferred from genomic (*Tomachewski et al., 2018*) or proteomic databases (*Cheng, Qiao & Horvatovich, 2018*) and cluster analysis allows for rapid dereplication without a reference library (*Clark et al., 2018*; *Dumolin et al., 2019*). Web applications for this analysis are available (*LaMontagne et al., 2017*); however, data analysis of mass spectra generated from bacteria isolated from the plant microbiome needs improvement (*Huschek & Witzel, 2019*). To our knowledge, a consensus designation of how clusters defined by MALDI-TOF corresponds to biological relevant taxonomic units has not emerged.

In this study, we generated a library of readily-culturable bacteria from the endosphere and rhizosphere of two varieties of maize (*Zea mays*): an agronomically important hybrid and a heritage variety. These maize models enabled comparisons of libraries that we predict would differ. Isolates were sorted, using custom scripts, into MALDI-TOF Taxonomic Units (MTUs) based on similarity of mass spectra generated by MALDI-TOF. These clusters, which are analogous to the operational taxonomic units widely reported in metagenomic studies, corresponded to a threshold of cosine similarity coefficient >0.65. These MTUs appeared coherent, as isolates clustered within an MTU showed nearly identical rRNA gene sequences and MTUs were consistent with species identified with a commercial mass spectra database. The throughput of this approach enabled comprehensive sampling of the aerobic, readily-culturable bacteria of the maize microbiomes we selected as a model. Rarefaction analysis indicated that the library we generate provided >70% coverage of the aerobic, readily-culturable bacteria in the maize microbiomes we sampled. This sampling depth facilitates testing hypotheses about what controls the diversity of libraries generated from plant microbiomes. For example, in this study, we observed that the rhizoplane of the heritage variety yielded the most MTUs that were unique to that particular niche compared to libraries generated from either the endosphere or the rhizoplane of the sweet variety.

MATERIALS & METHODS

Sampling maize microbiome

Two maize varieties were chosen for this experiment, a 'Heritage' and a 'Sweet' variety. The Heritage variety was Dent Earth Tones Corn (Botanical Interest Inc., Broomfield, CO, USA). This is a heritage variety selected for propagation based on its colorful kernels. The Sweet variety, Bodacious Hybrid Sweet Corn (American Seed Co, Spring Grove, PA, USA), has been highly bred for both sweetness and disease resistance. This modern variety has the designation "se" which stands for sugar enhancer and "R/M" indicating disease resistance to multiple maize diseases.

Plants were grown inside at the University of Houston-Downtown in fall of 2017. Two seeds per variety were planted in 3.5 inch × 3.5 inch pots containing LadyBug Vortex Potting Soil® (New Earth, Conroe, TX, USA). Plants were grown at room temperature in a two-tier plant growing unit with grow lights. The plants for this experiment were all on the same tier in two batches. Grow lights were set approximately 0.7 m above the plants and set for 12/12 h light/cycle. Pots were watered every 2–3 days. One batch, designated "New" was harvested 60 days after sowing. Another batch, designated "Old" was harvested 120 days after sowing. All plants were brought to the University of Houston—Clear Lake for plant measurements and microbial extraction.

Investigation of both maize varieties began with the cutting of plants into shoots and roots. Soil was knocked from the roots and the fresh weight of shoots and roots was recorded. The height of each shoot was recorded from the soil line to the longest shoot tip. Roots were transferred into a 50 ml conical tube and wetted with 40 ml of 5% DMSO prepared in sterile distilled water. The tubes were then vortexed for 1 min and centrifuged at $4,800 \times g$, in a swing-out rotor, for 10 min at room temperature. The roots were recovered with tweezers and blotted dry to obtain wet weights. The supernatant was discarded and the remaining slurry (~5 ml) was transferred to microtubes with a wide orifice pipet.

Endophytes were recovered from stem sections that were cut into lengths that weighed between 0.20 and 0.25 grams. These pieces were submersed in 70% ethanol for 60 s followed by flaming to remove ethanol. The pieces were then rinsed in sterile distilled water for 60 s. The pieces were then cut into approximately 1 mm fragments and ground to a slurry in 2 ml of sterile water, using a surface sterilized mortar and pestle (*Johnston-Monje et al., 2014*).

Slurries generated from roots and shoots were serially diluted in sterile phosphate buffered saline and inoculated on half strength tryptic soy broth (EMD Millipore, Billerica MA) solidified with 15 g per liter with agar (AmericanBio, Natick, MA) to make TSA plates. The track dilution method was used to spread dilutions (*Jett et al., 1997*). This method entailed spotting 10 µl from four dilutions in a row along one edge of the plate. The plate was then tilted to allow the droplets to run in a track across the plate. The plates were allowed to dry before being sealed with parafilm and incubated at 30 °C for 48 h. Colonies were picked from dilutions (generally the 10^{-4} dilution) that yielded isolated colonies. To maximize diversity, morphology was considered in picking colonies. Representative morphologies were selected from each track. Isolates were re-streaked to fresh plates twice for purity.

Microbial Identification—MALDI-TOF

Isolates were prepared for MALDI-TOF analysis with the ethanol inactivation and formic acid extraction protocols recommended by Bruker Scientific (Billerica, MA, USA). This approach follows the recommendations of Freiwald and Sauer (*Freiwald & Sauer, 2009*), except 50 µl volumes of formic acid and acetonitrile were used. Briefly, bacteria were cultured overnight at 30 °C on TSA plates. Resulting colonies were then suspended in 300 µl of HPLC-grade water and treated by adding 900 µl HPLC-grade ethanol.

The resulting slurry was stored at 4 °C for one week. The ethanol treated cells were then recovered by centrifugation (16,000 × *g*, 2 min) and the resulting pellet was extracted with formic acid/acetonitrile to yield a formic acid extract.

Matrix solution, 10 mg/mL a-cyano-4-hydroxycinnamic acid (HCCA, Sigma, St. Louis, MO, USA) in 50% acetonitrile, 47.5% water, and 2.5% trifluoroacetic acid (ACN, Sigma, St. Louis, MO, USA) was prepared fresh for each use. Bacterial test standard (BTS, Bruker p/n 8255343) was dissolved in 50 µL of 50% aq. ACN, 2.5% TFA following the manufacturer instructions. MALDI target plate (MSP 96 polished steel target, Bruker p/n 8280800) was washed with trifluoroaceitic acid and 70% ethanol as recommended by Bruker. On one or two spots on the target, 1 µl of formic acid extract was spotted on the target followed by 1 µl of matrix solution. Two BTS spots were prepared by applying 1 µl of BTS solution followed by 1 µl of matrix solution to each spot. Two matrix blank spots were included on each target as negative controls. Targets were allowed to dry at room temperature for approximately 15 min and they were shipped overnight, with an ice pack, to the Proteomics and Mass Spectrometry Core Facility at the Huck Institute (The Pennsylvania State University, University Park, PA, USA). Positive-ion mass spectra were acquired on a Bruker Ultraflextreme MALDI TOF/TOF mass spectrometer. Linear detection mode with the following parameters was used: pulsed ion extraction 170 ns; Ion source 1 25 kV; Ion source 2 94% of Ion source 1, and Lens 32% of Ion source 1. Matrix suppression (deflection) was set to 1,500 m/z. The laser repetition rate was 667 Hz; Smartbeam parameter set to "3_medium". Real-time smoothing was Off, baseline offset 0.2%, analog offset 2.1 mV, and the detection was set for low mass range, 1,880–20,000 Da. The target was moved in a random walk, complete sample pattern; 50 shots were fired at 24 raster spots (1,200 total shots) limited to a 2,000-mm diameter.

Mass calibration was performed with BTS as the standard; quadratic calibration curve was constructed based on m/z values of 8 calibrants over 3,637 to 16,952 m/z range. Mass spectra were smoothed using 10 cycles of SavitzkyGolay, with width 2 m/z; and baseline-subtracted using TopHat algorithm. Mass List Find parameters were as follows: peak detection algorithm Centroid, S/N threshold 2, minimum intensity threshold 600, max number of peaks 300, peak width 4 m/z, height 90%. The processed mass spectra were loaded into the MALDI Biotyper Version 3.1 (build 66) software (Bruker) and searched against a Bruker Taxonomy library containing 7,854 entries.

Mass spectra were also analyzed by cluster analysis using an R script that implemented functions in MALDIquant (*Gibb & Strimmer, 2012*), PVclust (*Suzuki & Shimodaira, 2006*) and several other packages. This script included two optimization loops that iteratively sampled random values, within specified ranges, for seven parameters: half-window for smoothing, baseline removal, half-window for alignment, tolerance of alignment, signal to noise ratio (SNR) for alignment, half-window for peak detection and SNR for peak detection. The first loop identified the parameters that optimized the number of peaks shared, as Jaccard coefficients calculated with the philentropy (*Drost, 2018*) between pairs of average mass spectra generated from BTS. The Jaccard coefficients were selected to minimize the variability introduced by peak heights, as described previously (*AlMasoud et al., 2014*). Output from the first optimization was passed to a step then conducted

quality control analysis to identify noisy spectra. On the spectra that passed quality control, the second loop selected the parameters that minimized the overlap in cosine similarity values, calculated following (*Strejcek et al., 2018*), between closely related and distantly related isolates. The cosine similarity coefficient was selected because these similarities, which are analogous to Pearson coefficients, are the most widely used metric for assessing the similarity between mass spectra (*Huber et al., 2021*) and are readily calculated with the R package coop (*Schmidt, 2019*). Cosine similarities were weighted by Jaccard coefficients normalized after with the formula $y = y_0 + x/(x + 0.2)$, where x is Jaccard coefficient, y_0 is the average cosine similarity when x = 0 and y is a predicted cosine value. The script then performed cluster analysis to define MALDI-TOF taxonomic units (MTUs), rarefaction analysis using the iNext package (*Hsieh, Ma & Chao, 2016*) and trained a machine learning algorithm using the RWeka package (*Hornik, Buchta & Zeileis, 2009*). This script is presented in one R markdown file (Supplemental Data). Raw mass spectra and Bruker system identifications are available as dataset MSV000086274 in MassIVE (*Deutsch et al., 2016*).

Curve fitting and ANOVA testing of significance of correlations were completed with SigmaPlot for Windows version 13.0 (Systat Software, San Jose, CA, USA).

Microbial Identification—16S rRNA gene sequencing

Representative isolates of MTUs that were not identified at the species level using the Biotyper database and software (Biotyper score < 2), were identified by 16S rRNA gene sequencing. DNA was extracted from isolates cultured overnight in tryptic soy broth (30 °C, 200 rpm) with the Puregene kit, following the manufacturer's protocol (Qiagen, Carlsbad CA, USA). A near intact fragment of the 16S rRNA gene was amplified with ReadyMade™ Primers 16SrRNA For and 16srRNA Rev supplied by IDT (Coralville, IA, USA) using DreamTaq Hot Start DNA polymerase, following the manufacturer's protocol (ThermoFisher, Waltham, MA, USA). PCR reactions were preheated (94 °C, 2 m) and then cycled 32 times through the following steps: denaturing (94 °C, 30 s), annealing (56 °C, 30 s) and extension (72 °C, 90 s). The final extension step was extended for 5 min and product size was confirmed by electrophoresis with a FlashGel™ DNA Kit (Lonza, Basel, Switzerland). PCR products were then purified with the DNA Clean & Concentrator kit, following the manufacturer's protocol (Zymo, Irvine, CA, USA). Purified fragments were sequenced with the above ReadyMade™ Primers, using Sanger technology, by Lone Star Laboratory (Houston, TX, USA).

Single-pass, 16S rRNA sequences were manually curated in BioEdit (*Hall, 1999*), checked for chimers with Decipher (*Wright, Yilmaz & Noguera, 2012*) and aligned against curated sequences with SINA (*Pruesse, Peplies & Glöckner, 2012*). ModelFinder (*Kalyaanamoorthy et al., 2017*) was used within the IQ-TREE environment (*Trifinopoulos et al., 2016*) to select the appropriate phylogenetic maximum likelihood model. Bootstrap values were calculated with UFBoot2 (*Hoang et al., 2017*) and a tree was generated with Newick Display 1.6 (*Junier & Zdobnov, 2010*). Decipher, ModelFinder, UFBoot2 and Newick Display were all run with default parameters. The 16S sequences were uploaded to GenBank and assigned accession numbers MW092913–MW092939.

Table 1 Number of maize-associated isolates analyzed by MALDI-TOF.

	Maize Variety	
Niche	Heritage	Sweet
Endosphere	22	23
Rhizoplane	51	36

RESULTS

The majority of bacteria isolated from maize did not match reference spectra in the Bruker system. Of mass spectra generated from the 132 representative colony morphologies cultured from the rhizoplane and endosphere of two varieties of maize (Table 1), ten failed the initial quality control of the Bruker system and eight failed a SNR threshold we describe below. Of the remaining 114 isolates, 36 were identified at the species level with probable confidence, as defined by a classification score of > 2 with the Biotyper system. Mass spectra from these 114 isolates were clustered into 60 MTUs with the following pipeline. These MTUs agreed with operational taxonomic units assigned by 16S rRNA gene sequencing.

Iteratively trying different parameters, like tolerance and SNR for aligning mass spectra and detecting peaks respectively, identified a set that optimized the number of peaks shared between spectra generated from BTS references spotted on separate targets. The number of peaks shared between reference spectra reached an asymptote after a dozen iterations (Fig. S1). Extending the number of iterations of the optimization loop to 2,000 identified parameters that corresponded to the highest Jaccard similarity coefficients (0.952), in terms of peaks matched between spectra generated from BTS. Pairwise comparisons of these BTS spectra shared 20/21 peaks and showed cosine similarities that ranged from 0.963–0.970. Jaccard similarity coefficients between these reference spectra showed a modal relationship with the total number of peaks detected in this set of samples (Fig. S2 top). The zenith of this relationship occurred between 389–406 total peaks detected from the 174 spectra that passed initial quality control within the Bruker system. The parameters that yielded the most reproducible peaks were selected for identification of quality spectra, as defined by the number of the SNR for the ten largest peaks and the number of peaks detected. Quality control analysis, using a threshold of a median SNR of 15 for the 10 largest peaks and detection of at least 13 peaks, pruned mass spectra generated from 8 of 122 isolates from subsequent cluster analysis. The second optimization loop found, for pairs of isolates that had practically identical 16S rRNA sequences (>99%), the average Jaccard coefficient for comparisons reached a zenith at 498 peaks detected from 114 isolates (Fig S2 bottom).

Cosine similarities calculated with the parameters that showed the greatest discrimination between comparisons within species than between species, suggested a hyperbolic relationship with the Jaccard similarities (Fig. 1). Most (6472/6670) of the pairwise comparisons showed Jaccard similarity coefficients of less than 0.2. In that range, cosine similarities averaged (±SD) 0.27 ± 0.10. For the 192 pairwise comparisons that
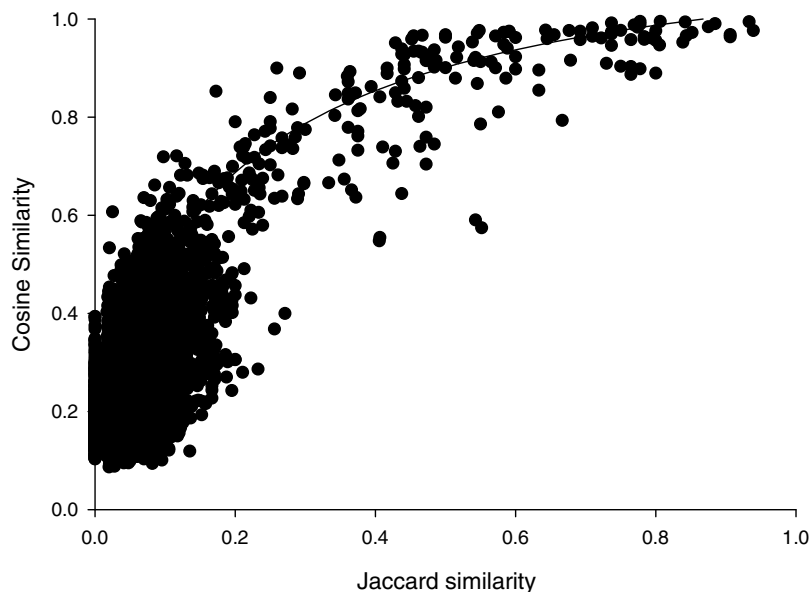
Figure 1 Cosine versus Jaccard similarities for pairwise comparisons. Line illustrates a mathematical model of the hyperbolic relationship described in Methods, not a fit to the data.

showed Jaccard similarities greater than or equal to 0.2, cosine similarities averaged 0.81 ± 0.16. Jaccard similarities of 0.2 corresponded to cosine similarities of about 0.69. This suggests that spectra that share very few peaks may show relatively high cosine similarities, which may reflect the variability in peak heights. Accordingly, we weighted cosine similarities by Jaccard coefficients (see "Methods").

The percent identities of pairwise comparisons 16S rRNA genes fit a sigmoidal model with weighted cosine similarities (Fig. 2). The fit had a modest correlation coefficient (r^2= 0.445) and analysis of variance suggested the trend was highly probable ($P < 0.0001$); however, most (259/325) of the pairwise comparisons showed percent identities of 16S rRNA genes of less than 90%. The midpoint of this sigmoidal model corresponded to a cosine similarity of 0.66. Cosine similarities of >0.66 consistently corresponded to a percent identify of ribosomal sequences ranging from 99–100% (Fig. 2).

Inspection of Fig. 2 suggested that cosine similarity of 0.65 would distinguish species. Using that threshold, clustering of a library of 114 isolates into MALDI-TOF taxonomic units (MTUs) defined 60 MTUs. Of these, 25 clusters contained more than one isolate (replicated MTUs) and 35 were singletons. Extrapolation from rarefaction analysis predicted that a library of 228 isolates would contain 84 MTUs, with a confidence interval of 70–98 (Fig. 3 top). This library of isolates corresponds to 70% coverage, with a confidence interval of 62–77% (Fig. 3 bottom). In other words, as defined by MTUs, the pooled library contained the majority of bacterial species that could be readily isolated from the rhizoplane and endosphere of both maize varieties we sampled. Further extrapolation predicted that a library of 228 isolates would provide 78–95% coverage.

Partial sequencing of 16S rRNA genes yielded sequences that were used to classify isolates that were not matched to the commercial database into 18 phylotypes that
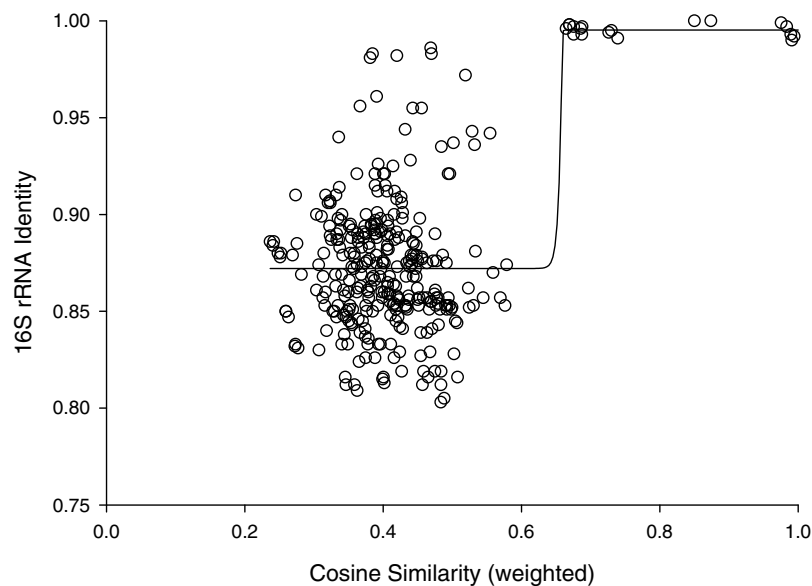
Figure 2 Cosine similarity versus 16S identity. Cosine similarity was weighted with values predicted from Jaccard coefficients using the hyperbolic model in Fig. 1 as described in Methods. Line represents a sigmoidal curve fit. Full-size ⬚ DOI: 10.7717/peerj.11359/fig-2

corresponded to different genera (Fig. 4). Phylotypes defined by rRNA gene sequencing generally appeared robust, as assessed by bootstrap values > 90% (Fig. 4). The exception was separation of the branch containing isolates that classified as *Paenibacillus* and *Bacillus* species, which bootstrapping supported only 76% of iterations. Within this Bacillales order, and in terms of rRNA gene sequences, isolate H1OZ89 showed high similarity to a phosphate solubilizing *Paenibacillus* species previously isolated from rape (*Ding et al., 2019*). Also, within this order, isolate H3OZ122 showed high similarity to a microplastics degrading *Bacillus* strain recently isolated from the Yellow Sea (*Wang et al., 2019*), isolate S4OZ125 showed high similarity to *Bacillus aciditolerans*, a recently described novel species isolated from a rice field (*Ding et al., 2019*) and H3OZ107 showed similarity to a *Staphylococcus hominis* strain isolated from a blood sample (Fig. 4). These isolates (H1OZ89, H3OZ122, S4OZ125 and H3OZ107) clustered into separate MTUs, which suggests the similarity threshold defined MTUs corresponds to species that occupy different niches.

Consistent with MTUs that classified within the Bacillales order, the topology of Fig. 4 generally showed congruence with MTUs. For example, MTUs 33, 34, 46 and 51 each contained two isolates that both clustered together in the same leaf on the tree. For the seven isolates of the *Acidovorax* genera shown in Fig. 4, the threshold for clustering appeared conservative. The rRNA sequences generated from seven isolates showed > 99% identity with species of *Acidovorax*, and each other, but were separated into three MTUs (29, 35 and 54). MTU 35 corresponded to a branch in Fig. 4 supported with 94% bootstrapping value, which suggests that cluster is coherent. On the other hand, MTUs 25 and 54 appear to have isolates of the same species, split into two clusters.
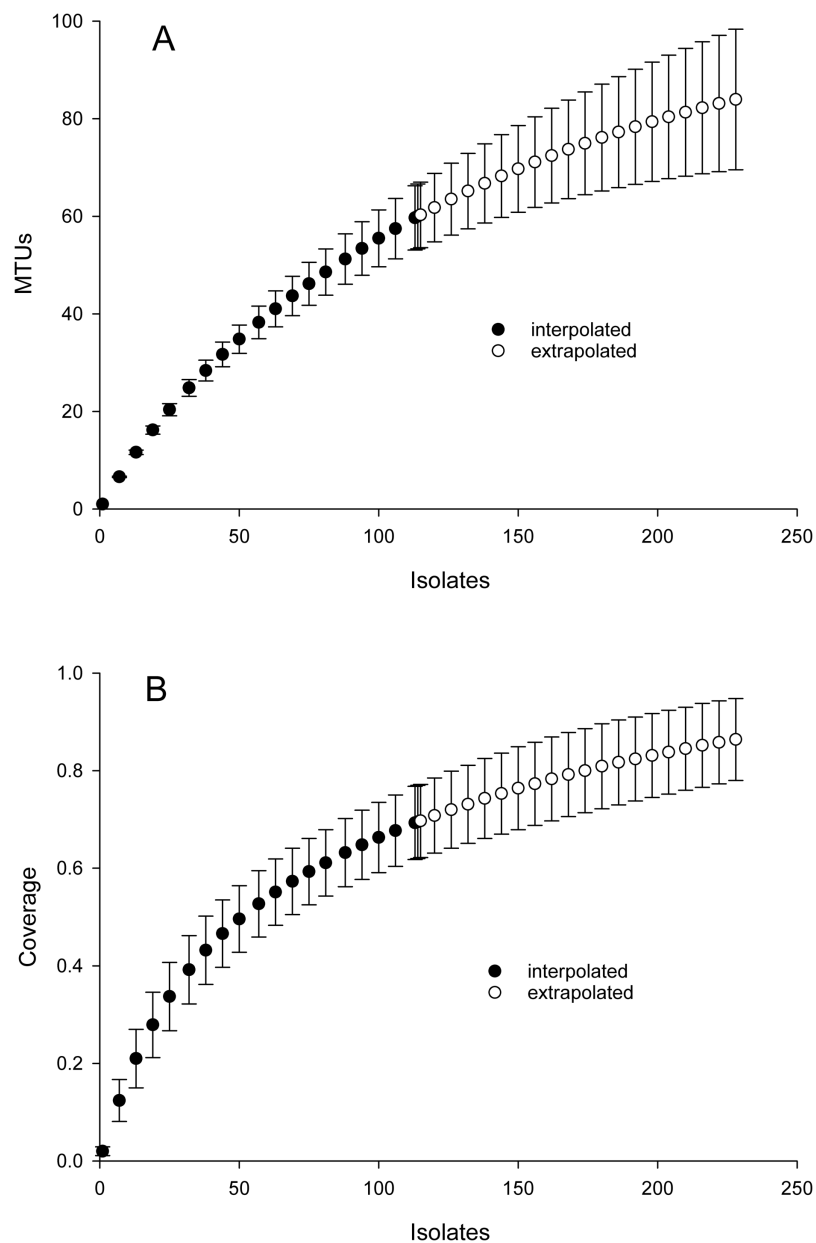
Figure 3 Rarefaction analysis of library generated from the maize microbiome. X-axis presents number of isolates in the library (filled circles) or an extrapolation from the existing library (open circles). Confidence intervals present standard error bars. Figures were generated with the default parameters in iNext. (A) Y-axis presents number of taxonomic units defined by similarity of mass spectra. (B) Y-axis presents fraction of total diversity recovered. Full-size ⬛ DOI: 10.7717/peerj.11359/fig-3

Matching to reference spectra did not appear to depend on the quality of the spectra, as classification scores showed no relationship with the SNR of spectra. The set of identified species contained 35 isolates that classified as bacteria and one isolate that classified as a fungal species (*Trichosporon mucoides*). Of the 35 bacterial isolates 6 were singletons. The 29 replicated isolates were classified into 11 species. These isolates generally corresponded to specific MTUs. For example, the 6 isolates reliably identified as
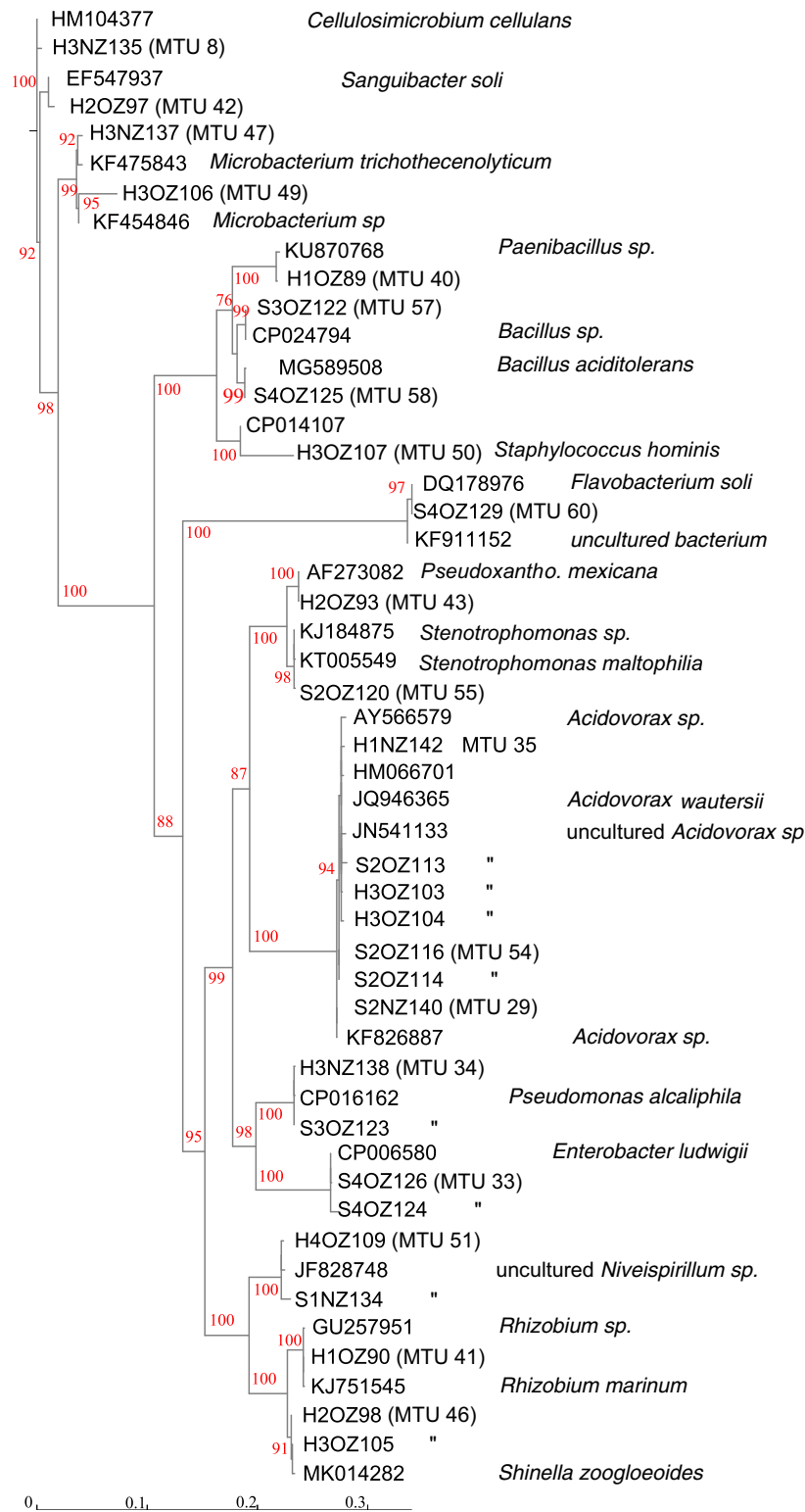
Figure 4 Phenogram generated from 16S rRNA gene sequences. Partial sequences were aligned against curated sequences with SINA. ModelFinder was used select the appropriate phylogenetic maximum likelihood model within IQ-TREE. Bootstrap values were calculated with UFBoot2. Genbank accession numbers and species identifications are provided for reference sequences. Scale indicates substitutions, where a distance of 0.04 corresponds to 96% identity.   Full-size ☒ DOI: 10.7717/peerj.11359/fig-4
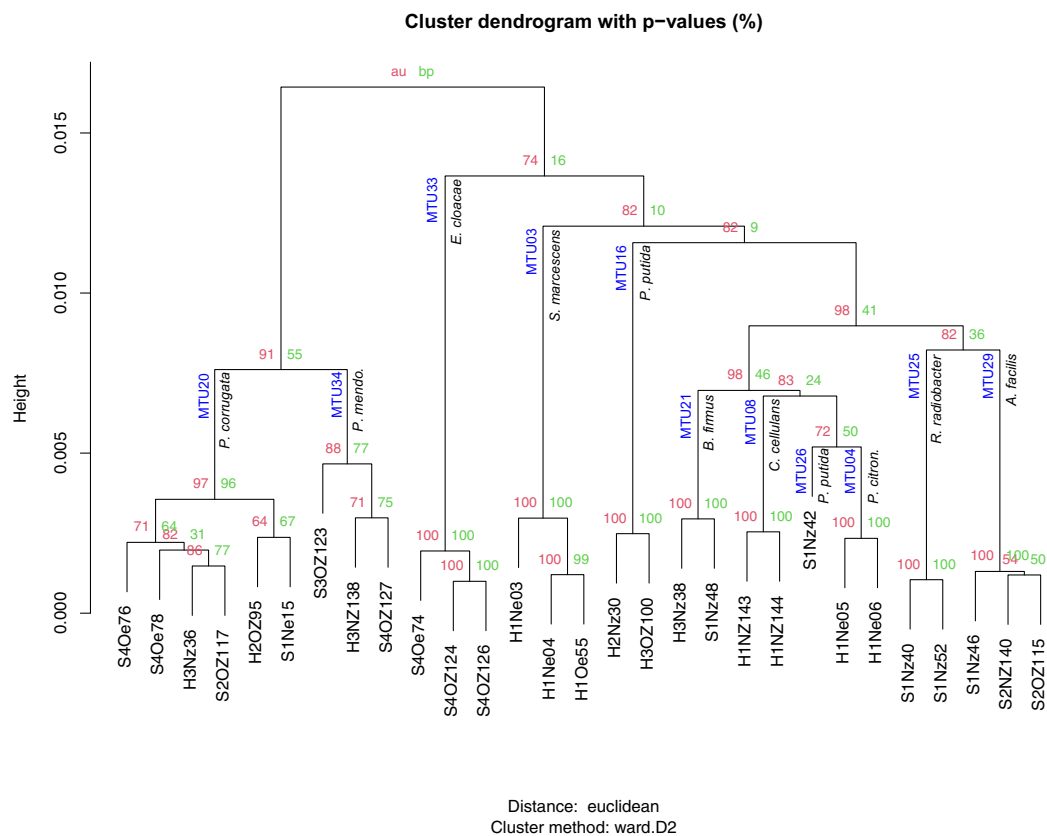
Figure 5 Hierarchical clustering of mass spectra for isolates reliably identified with Bruker MALDI Biotyper system. AU and BP values present approximately unbiased probability values and bootstrap probabilities assigned with pvclust. MTUs and species identification given by the Biotyper system are indicated at each node. Height presents dissimilarity between nodes.

Full-size ⬚ DOI: 10.7717/peerj.11359/fig-5

Pseudomonas corrugata all clustered into MTU 20 (Fig. 5). Similarly, isolates identified by the Bruker system as *Pseudomonas mendocina, Enterobacter cloacae, Serratia marcescens, Bacillus firmus, Cellulosimicrobium cellulans, Pseudomonas citronellolis, Rhizobium radiobacter* and *Acidovorax facilis* clustered within MTUs 34, 33, 3, 21, 8, 4, 25 and 29 respectively (Fig. 5). Bootstrap values, for iterations of cluster analysis, generally supported the congruence of these MTUs with species identifications. Bootstrap values for these clusters exceeded 95%, with the exception of MTU 34 (Fig. 5). However, the dendrogram was not coherent with respect to the phylogeny of the isolates. In particular, *Pseudomonas* species were dispersed throughout the tree and appeared within a branch that contained Gram positive species (Fig. 5) and, for the three isolates the system identified as *P. putida*, the threshold used to define MTUs appeared conservative. These isolates were split into two MTUs (Fig. 5). The lone isolate in MTU 26 showed little similarity, in terms of cosine similarity, to any other isolate.

Machine learning resolved the apparent discrepancy for the isolates that clustered into different MTUs but were identified as one species by the Bruker system. To train the machine learning algorithm, MTUs were identified by the Bruker system and 16S
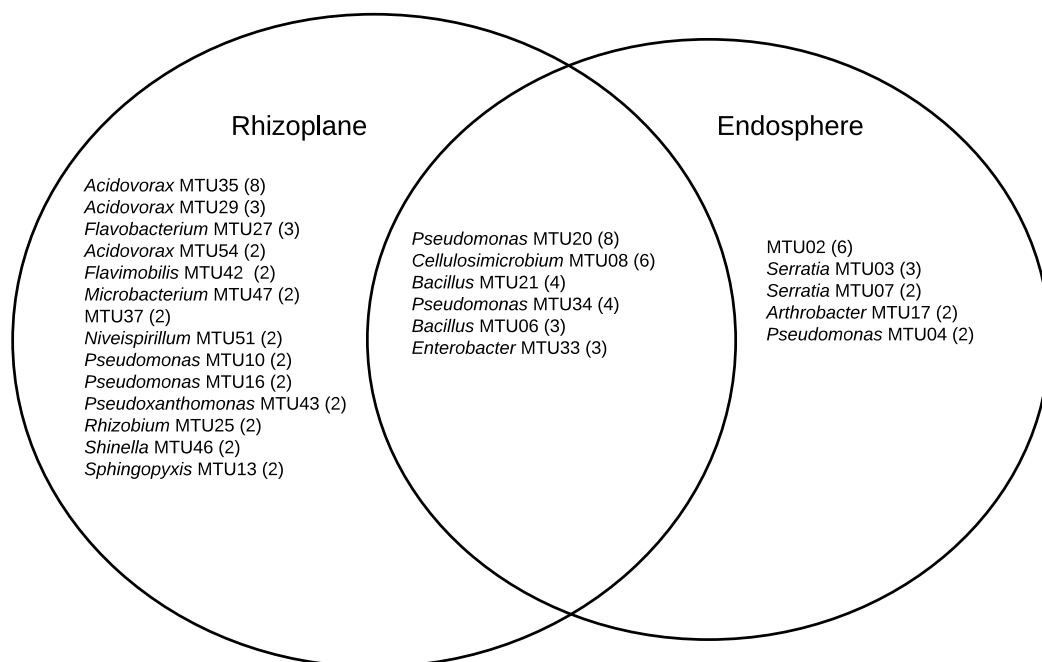
Figure 6 MTUs detected in library generated from the rhizoplane and endosphere of maize. The number of isolates clustered in each MTU is indicated. Singletons are not shown.

Full-size 🖼 DOI: 10.7717/peerj.11359/fig-6

sequencing. When the Bruker and 16S identification differed, the 16S identification was used but this was rare. Identification by the Bruker system and 16S sequencing agreed at the genera level, for Bruker scores greater than or equal to 1.6; however, the two approaches showed some discordance at the species level. For example, MTU 33 was reliably identified (Bruker score = 2.5) as *Enterobacter cloacae* but these isolates clustered with *Enterobacter ludwigii* in Fig. 4. MTU 34 was probably identified (Bruker > 2) as *Pseduomonas mendocina* but these isolates clustered with *P. alcaliphila* in Fig. 4. MTU 35 were identified by the Bruker system as *Acidovorax facilis* with scores > 2.2 but a representative of that cluster showed high similarity, in terms of 16S sequence, to *A. wautersii* (Fig. 4). This MTU was defined as *A. wautersii*. When species were split into two MTUs, they were pooled. For example, the three isolates identified, at the probable level, as *P. putida* by the Bruker system were defined as that species, even though they belonged to two MTUs (Fig. 5). MTUs 54 and 29 were identified as one clade *Acidovorax sp.* based on Fig. 4. After discarding 34 singletons, this classification scheme provided at least genera level identities for all but two of the remaining isolates. The machine learning algorithm classified 100% of the 80 replicated isolates, and the two BTS references, correctly as 22 MTUs. This accuracy of classification is reflected in a Cohen's *Kappa* coefficient of 1, where this coefficient ranges from −1 to 1.

A combination 16S sequencing and MALDI TOF methods identified distinct MTUs that were relatively abundant to particular niches or plant varieties. Libraries generated from the rhizoplane contained the highest number of replicated MTUs overall (Fig. 6). Of those 20 MTUs, 14 did not appear in libraries generated from the endosphere. These

rhizoplane-specific MTUs included numerous representatives of two *Acidovorax* sp. and *Flavobacterium sp. Niveispirillum sp., Pseudoxanthomonas sp., Rhizobium radiobacter and Shinella sp.* (Fig. 6). MTUs specific to the endosphere included a yet-to-be identified cluster (MTU02), *Serratia* sp., *Arthrobacter* sp., and *Pseudomonas* sp. (Fig. 6). MTUs identified as *Pseudomonas* sp., *Cellulosimicrobium* sp*., Bacillus* sp., and *Enterobacter* sp. were found in both the rhizoplane and the endosphere (Fig. 6). Libraries generated from the heritage variety contained more MTUs (11) unique to that variety than libraries generated from the sweet variety (Fig. S3).

DISCUSSION

Optimization of cluster analysis of mass spectra generated from a library of bacteria isolated from maize microbiomes produced coherent MTUs. These MTUs agreed with species defined by 16S rRNA gene sequencing and by the Bruker system and with niches typically occupied by species related to these MTUs. The library of readily culturable bacteria isolated from the rhizoplane of the heritage variety contained the most strains that were specific to a particular niche. That is strains that abound in libraries generated from that sample and were not common in libraries generated from the endosphere of that variety or from rhizoplane and endosphere of the sweet variety. The rhizoplane and endosphere have many species in common, but it is logical that the rhizoplane community is more diverse given its direct contact with soil, which are diverse systems (*Howe et al., 2014*). In fact, endophyte communities may be a subset of the more extensive rhizosphere communities (*Long et al., 2010*).

The relatively high abundance of *Acidovorax* and *Flavobacteria* in libraries generated from the rhizoplane (Fig. 6) is consistent with previous culture-dependent surveys of plant microbiomes. *Acidovorax* is commonly found in both soil and plants (*Long et al., 2010*) and *Flavobacteria* genera includes a terrestrial clade associated with plant roots plant roots (*Kolton et al., 2013*). However, these genera are not consistently associated with maize roots, as assessed by metagenomic analysis (*Niu et al., 2017*). We also isolated *Rhizobium radiobacter (*formerly *Agrobacterium tumefaciens)*. This soil bacteria interacts with plants and strains of this species show potential for bioremediation applications (*Deepika et al., 2016*) and novel species of this genera have been isolated from maize roots (*Gao et al., 2017*). *Microbacterium* has been isolated specifically from maize kernels (*Rijavec et al., 2007*). This endophyte may have been in the seed when it was sown.

The library generated from the maize endosphere included species that colonize plants. The most abundant MTU specific to the endosphere (MTU02) was not identifiable with the Bruker system and was not archived for future analysis. Other MTUs that abounded in libraries generated from the endosphere included *Serratia*, *Arthrobacter* and *Pseudomonas* species. *Serratia marcescens* is a common endophyte in cotton roots (*McInroy & Kloepper, 1995*) and rice (*Tan et al., 2001*). Several *Serratia* species associate with maize roots (*Mosimann et al., 2017*) and *S. marcescens* can cause maize whorl rot (*Wang et al., 2015*). *Arthrobacter spp*. are readily isolated from cotton stems (*McInroy & Kloepper, 1995*) and identified as an endophyte in maize sap (*Ali et al., 2018*).

MTUs found in both the libraries generated from the rhizoplane and endosphere included genera frequently associated with plant hosts. *Pseudomonas* sp. are readily cultured from the rhizoplane of corn on selective media (*Mosimann et al., 2017*). *Cellulosimicrobium sp.* can stimulate plant growth and act as a biocontrol for barley (*Nabti et al., 2014*). Recently the genome of a *Cellulosimicrobium* sp. strain isolated from endosphere of a perennial grass was sequenced because of its growth promoting properties (*Eida et al., 2020*). *Bacillus* sp. are ubiquitous in soils and widely used as microbial treatments of seeds (*Rocha et al., 2019*) and *Enterobacter* sp. can promote the growth of maize (*Naveed et al., 2014*) and appear a core member of that host's microbiome (*Niu et al., 2017*).

Many of the potentially beneficial rhizobacteria and endophytes presented above appear associated with only the heritage and not the sweet variety. For example, *Cellulosimicrobium* and *Serratia* sp. was common to only the heritage variety and *Enterobacter* and *Flavobacteria* sp. were common to only the sweet variety. This can inform microbial discovery efforts, where the goal is to isolate strains of a particular genera; however, for development of microbial products to improve the productivity and sustainability of maize, the variety cultivated by growers should be considered. Plant varieties form close associations with particular beneficial rhizobacterial strains (*Batstone et al., 2020*).

Clustering rhizobacteria and endophytes into MTUs, with the scripts presented herein, suggests a method to identify beneficial strains of plant-associated bacteria. MALDI-TOF has the resolution to differentiate strains of species that differ in functions, like antibiotic production (*Clark et al., 2018*) and resistance (*Flores-Treviño et al., 2019*). This resolution is particularly valuable in distinguishing *Bacillus* species (*Hotta et al., 2011*; *Lasch et al., 2009*). For example, MALDI-TOF can identify pathogenic strains of clinical (*Celandroni et al., 2016*) and environmental isolates of *Bacillus* (*Starostin et al., 2015*). In vitro assays, like 1-aminocyclopropane-1-carboxylate deaminase production (*Glick, 2014*), are widely used to identify potential PGPR. The correspondence between these phenotypes and MTUs requires further investigation.

Identification of maize-associated bacteria involved leveraging species identification with the Bruker system with 16S rRNA gene sequencing. This hybrid approach is widely used in applications of MALDI-TOF to environmental microbiology, where many of the isolates will not match the commercial database. For this study, the percentage of bacteria identified to the probable species level (32%) with the Biotyper system was lower than previous reports (73%) for soil bacteria (*Strejcek et al., 2018*) but in the range reported for a library generated from the rhizosphere of horseradish (*Uhlik et al., 2011*) and bacteria isolated from seawater (*Timperio et al., 2017*). We also observed discrepancies between species identifications for the two approaches, as reported previously and confirmed recently (*Pandey et al., 2019*). This could reflect the paucity of environmental isolates in the reference library we queried. The confidence of identification shows a hyperbolic relationship with the number of reference spectra in the database and approaches an asymptote at about 50 main spectra for each strain (*Erler et al., 2015*). The commercial database we queried only contains a few main spectra for most species.

With cluster analysis, we can begin to understand the diversity of rhizobacteria cultured with any particular technique from a plant's microbiome. Clustering of mass spectra, following optimization of alignment, identified MTUs that corresponded to highly similar (>99%) 16S rRNA gene sequences and species defined by a commercial database. The cosine similarity (0.65) selected as a threshold for defining MTUs was lower than threshold (0.79) reported by Strejcek et al. (*Strejcek et al., 2018*) but the coherence of the clusters, in terms of the consistency with clusters by 16S rRNA sequencing defined by a commercial spectra database and, supports this lower threshold. A more conservative threshold would split clusters that cannot be resolved with 16S rRNA sequencing, which is the current standard for microbial identification (*Edgar, 2018*).

High bootstrap support for a dendogram generated from mass spectra (Fig. 5) suggests the data analysis pipeline used to generated these clusters is robust. This suggests an approach to comprehensively compare libraries of bacteria isolated from different parts of the maize microbiome and test hypotheses about the efficacy of advanced culturing techniques, such as in situ cultivation (*Berdy et al., 2017*), improved media formulations (*Tanaka et al., 2014*) and gnotobiotic systems (*LaMontagne, 2020*). Machine learning also supported the clusters defined herein, and this could lead to an alternative method of identification (*De Bruyne et al., 2011*). Indeed, dozens of machine learning algorithms have recently been applied to this task (*Weis, Jutzeler & Borgwardt, 2020*). Our analysis suggests Rweka shows promise; however, development of such an identification system would require a challenge dataset generated externally, preferably on a different instrument (*Clark, Murphy & Sanchez, 2020*).

The reliance on 16S sequences limits the scope of this study. Phylogenetic analysis of the 16S rRNA gene benefits from a vast public database and user-friendly data analysis packages; however, analysis of this one gene lacks resolution. For example, diverse Bacillus species share identical rRNA gene sequences. Multilocus sequence typing (MLST) provides a higher resolution approach (*Maiden et al., 1998*) but MLST can fail to accurately estimate phylogenetic relationships between bacteria (*Tsang et al., 2017*). Whole genome sequencing provides strain-level identification (*Salipante et al., 2015*) but costs hundreds of dollars per isolate. Continued decreasing costs per bp of next generation sequencing (*Park & Kim, 2016*) and application of low-cost, third generation sequencers (*Petersen et al., 2019*) suggests whole genome sequencing will soon be standard for clinical applications (*Anis et al., 2018*); however, the true cost of whole genome includes library preparation, skilled labor and computational costs (*Sboner et al., 2011*). These costs limit the application of this approach to high-throughput microbial discovery programs.

The pooling of isolates collected at different stages of growth, and isolated on different media, into one library limits the scope of this study and precludes testing of hypothesis about the controls of the diversity and limits our ability to estimate the coverage these libraries provided. Defining the diversity of maize microbiome, within different locations of the host and between different varieties of maize, would require generation of much larger libraries then is feasible with culture-dependent techniques. Metagenomic analysis is

better suited testing hypotheses and provides a more complete census of the microbial communities in nature (*Handelsman, 2004*); however, this dataset revealed trends consistent with metagenomic studies, ecological theory and published observations.

In this study, we used the maize microbiome as model and demonstrate that MALDI-TOF systems provide a relatively low-cost method of identifying bacteria isolated from plant microbiomes. This proteomics approach is available to the general scientific community through core facilities, like the Huck Institute a Pennsylvania State University, which processed the samples herein. The costs of analysis by vendors is about a dollar per isolate for pre-spotted targets. The ethanol treatment/formic acid extraction applied herein is robust. Undergraduates in a teaching laboratory at the University of Houston—Clear Lake generated the protein extracts presented herein and dozens of students in this laboratory routinely generate quality spectra the first time they try this technique. The primary cost associated with this approach is purchasing of the reusable targets (~$500 each) and the scripts presented herein allow users to process the data with packages freely available in R.

The availability of MALDI-TOF analysis, through core facilities and commercial vendors, as well as the development of data analysis pipelines, like the one presented herein, should allow investigators to test the coherence of MTUs in terms of phenotypic traits and genotypes. However, to fully realize the potential of MALDI-TOF for identifying bacteria associated with plants, we need to build public databases of both raw and processed mass spectra and the metadata associated with the isolates. In other words, the community needs a resource analogous to the short archive available for next and third generation sequencing data. MassIVE provides such a resource (*Deutsch et al., 2016*) but only a handful of mass spectra of bacteria isolated from the plant microbiome have been archived in that database to date. We echo the call for the development and use of such a central depository of mass spectra generated from plant-associated microbes (*Ahmad, Babalola & Tak, 2012*) and environmental isolates in general.

CONCLUSIONS

MALDI-TOF facilitated the classification of bacteria isolated from the maize microbiome into 60 MTUs. Our method proved to be a rapid and inexpensive method of describing the types of bacteria cultured from plant microbiomes with freely available spectra alignment and machine learning packages.

ACKNOWLEDGEMENTS

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests

Michael G. LaMontagne is an Academic Editor for PeerJ and the co-Founder of EndoBiome - a company that develops microbial products for agriculture.

Author Contributions

- Michael G. LaMontagne conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Phi L. Tran performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Alexander Benavidez performed the experiments, prepared figures and/or tables, and approved the final draft.
- Lisa D. Morano conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

Patent Disclosures

The following patent dependencies were disclosed by the authors:
LaMontagne MG. 2020. Gnotobiotic rhizobacterial isolation plant systems and methods of use thereof. US Patent 10801079, Oct. 13, 2020.

Data Availability

The following information was supplied regarding data availability:
The 16S sequences are available at GenBank: MW092913 to MW092939.
Raw mass spectra and Bruker system identifications are available in MassIVE: MSV000086274.

The Supplemental Files are comprised of the data analysis scripts in an R markdown file and in html format; the metadata and the 16S identity matrix inputs to this data analysis pipeline; and for convenience, R files are available as input files that allow investigators to input the results of multiple, computationally-intensive, iterations of alignment attempts.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.11359#supplemental-information.

REFERENCES

Ahmad F, Babalola O, Tak H. 2012. Potential of MALDI-TOF mass spectrometry as a rapid detection technique in plant pathology: identification of plant-associated microorganisms. *Analytical and Bioanalytical Chemistry* **404(4)**:1247–1255 DOI 10.1007/s00216-012-6091-7.

Ali S, Isaacson J, Kroner Y, Saldias S, Kandasamy S, Lazarovits G. 2018. Corn sap bacterial endophytes and their potential in plant growth-promotion. *Environmental Sustainability* **1(4)**:341–355 DOI 10.1007/s42398-018-00030-4.

AlMasoud N, Xu Y, Nicolaou N, Goodacre R. 2014. Optimization of matrix assisted desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) for the characterization of *Bacillus* and *Brevibacillus* species. *Analytica Chimica Acta* **840**:49–57 DOI 10.1016/j.aca.2014.06.032.

Anand R, Grayston S, Chanway C. 2013. N_2-fixation and seedling growth promotion of Lodgepole Pine by endophytic *Paenibacillus polymyxa*. *Microbial Ecology* **66(2)**:369–374 DOI 10.1007/s00248-013-0196-1.

Anis E, Hawkins IK, Ilha MRS, Woldemeskel MW, Saliki JT, Wilkes RP. 2018. Evaluation of targeted next-generation sequencing for detection of bovine pathogens in clinical samples. *Journal of Clinical Microbiology* **56(7)**:e00399-00318 DOI 10.1128/JCM.00399-18.

Barnawal D, Maji D, Bharti N, Chanotiya C, Kalra A. 2013. ACC deaminase-containing *Bacillus subtilis* reduces stress ethylene-induced damage and improves mycorrhizal colonization and rhizobial nodulation in *Trigonella foenum*-graecum under drought stress. *Journal of Plant Growth Regulation* **32(4)**:809–822 DOI 10.1007/s00344-013-9347-3.

Batstone RT, O'Brien AM, Harrison TL, Frederickson ME. 2020. Experimental evolution makes microbes more cooperative with their local host genotype. *Science* **370(6515)**:476–478 DOI 10.1126/science.abb7222.

Berdy B, Spoering AL, Ling LL, Epstein SS. 2017. *In situ* cultivation of previously uncultivable microorganisms using the ichip. *Nature Protocols* **12(10)**:2232–2242 DOI 10.1038/nprot.2017.074.

Brader G, Compant S, Mitter B, Trognitz F, Sessitsch A. 2014. Metabolic potential of endophytic bacteria. *Current Opinion in Biotechnology* **27**:30–37 DOI 10.1016/j.copbio.2013.09.012.

Bresson J, Vasseur F, Dauzat M, Labadie M, Varoquaux F, Touraine B, Vile D. 2014. Interact to survive: *Phyllobacterium brassicacearum* improves *Arabidopsis* tolerance to severe water deficit and growth recovery. *PLOS ONE* **9(9)**:e107607 DOI 10.1371/journal.pone.0107607.

Brígido C, Glick B. 2015. Phytoremediation using rhizobia. In: Ansari AA, Gill SS, Gill R, Lanza GR, Newman L, eds. *Phytoremediation*. Berlin: Springer International Publishing, 95–114.

Böhme K, Fernández-No IC, Barros-Velázquez J, Gallardo JM, Cañas B, Calo-Mata P. 2012. SpectraBank: an open access tool for rapid microbial identification by MALDI-TOF MS fingerprinting. *Electrophoresis* **33(14)**:2138–2142 DOI 10.1002/elps.201200074.

Celandroni F, Salvetti S, Gueye SA, Mazzantini D, Lupetti A, Senesi S, Ghelardi E. 2016. Identification and pathogenic potential of clinical *Bacillus* and *Paenibacillus* isolates. *PLOS ONE* **11(3)**:e0152831 DOI 10.1371/journal.pone.0152831.

Cheng D, Qiao L, Horvatovich P. 2018. Toward spectral library-free matrix-assisted laser desorption/ionization time-of-flight mass spectrometry bacterial identification. *Journal of Proteome Research* **17(6)**:2124–2130 DOI 10.1021/acs.jproteome.8b00065.

Chowdhury SP, Dietel K, Rändler M, Schmid M, Junge H, Borriss R, Hartmann A, Grosch R. 2013. Effects of *Bacillus amyloliquefaciens* FZB42 on lettuce growth and health under pathogen pressure and its impact on the rhizosphere bacterial community. *PLOS ONE* **8(7)**:e68818 DOI 10.1371/journal.pone.0068818.

Clark CM, Costa MS, Sanchez LM, Murphy BT. 2018. 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 DOI 10.1073/pnas.1801247115.

Clark CM, Murphy BT, Sanchez LM. 2020. A call to action: the need for standardization in developing open-source mass spectrometry-based methods for microbial subspecies discrimination. *mSystems* **5(1)**:e00813-00819 DOI 10.1128/mSystems.00813-19.

Dai A. 2011. Drought under global warming: a review. *Wiley Interdisciplinary Reviews: Climate Change* **2(1)**:45–65 DOI 10.1002/wcc.81.

De Bruyne K, Slabbinck B, Waegeman W, Vauterin P, De Baets B, Vandamme P. 2011. Bacterial species identification from MALDI-TOF mass spectra through data analysis and machine learning. *Systematic and Applied Microbiology* **34(1)**:20–29 DOI 10.1016/j.syapm.2010.11.003.

Deepika KV, Raghuram M, Kariali E, Bramhachari PV. 2016. Biological responses of symbiotic *Rhizobium radiobacter* strain VBCK1062 to the arsenic contaminated rhizosphere soils of mung bean. *Ecotoxicology and Environmental Safety* **134(6)**:1–10 DOI 10.1016/j.ecoenv.2016.08.008.

Deutsch EW, Csordas A, Sun Z, Jarnuczak A, Perez-Riverol Y, Ternent T, Campbell DS, Bernal-Llinares M, Okuda S, Kawano S, Moritz RL, Carver JJ, Wang M, Ishihama Y, Bandeira N, Hermjakob H, Vizcaíno JA. 2016. The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Research* **45(D1)**:D1100–D1106 DOI 10.1093/nar/gkw936.

Ding MJ, Shang NJ, Xiao ZX, Shao F, Liu L, Huang Y, Zhou LK, Zhou JH, Zhang Y. 2019. *Bacillus aciditolerans* sp. nov., isolated from paddy soil. *International Journal of Systematic and Evolutionary Microbiology* **69(4)**:1155–1161 DOI 10.1099/ijsem.0.003285.

Drost H-G. 2018. Philentropy: information theory and distance quantification with R. *Journal of Open Source Software* **3(26)**:765 DOI 10.21105/joss.00765.

Dumolin C, Aerts M, Verheyde B, Schellaert S, Vandamme T, Van der Jeugt F, De Canck E, Cnockaert M, Wieme AD, Cleenwerck I, Peiren J, Dawyndt P, Vandamme P, Carlier A. 2019. Introducing SPeDE: high-throughput dereplication and accurate determination of microbial diversity from matrix-assisted laser desorption-ionization time of flight mass spectrometry data. *mSystems* **4(5)**:e00437-00419 DOI 10.1128/mSystems.00437-19.

Edgar RC. 2018. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics* **34(14)**:2371–2375 DOI 10.1093/bioinformatics/bty113.

Eida AA, Bougouffa S, Alam I, Saad MM, Hirt H. 2020. Complete genome sequence of the endophytic bacterium *Cellulosimicrobium* sp. JZ28 isolated from the root endosphere of the perennial desert tussock grass *Panicum turgidum*. *Archives of Microbiology* **202(6)**:1563–1569 DOI 10.1007/s00203-020-01859-2.

Emami K, Nelson A, Hack E, Zhang J, Green DH, Caldwell GS, Mesbahi E. 2016. MALDI-TOF mass spectrometry discriminates known species and marine environmental isolates of *Pseudoalteromonas*. *Frontiers in Microbiology* **7(791)**:104 DOI 10.3389/fmicb.2016.00104.

Erler R, Wichels A, Heinemeyer E-A, Hauk G, Hippelein M, Reyes NT, Gerdts G. 2015. VibrioBase: a MALDI-TOF MS database for fast identification of Vibrio spp. that are potentially pathogenic in humans. *Systematic and Applied Microbiology* **38(1)**:16–25 DOI 10.1016/j.syapm.2014.10.009.

Flores-Treviño S, Garza-González E, Mendoza-Olazarán S, Morfín-Otero R, Camacho-Ortiz A, Rodríguez-Noriega E, Martínez-Meléndez A, Bocanegra-Ibarias P. 2019. Screening of biomarkers of drug resistance or virulence in ESCAPE pathogens by MALDI-TOF mass spectrometry. *Scientific Reports* **9(1)**:18945 DOI 10.1038/s41598-019-55430-1.

Freiwald A, Sauer S. 2009. Phylogenetic classification and identification of bacteria by mass spectrometry. *Nature Protocols* **4(5)**:732–742 DOI 10.1038/nprot.2009.37.

Gao J-L, Sun P, Wang X-M, Lv F-Y, Mao X-J, Sun J-G. 2017. *Rhizobium wenxiniae* sp. nov., an endophytic bacterium isolated from maize root. *International Journal of Systematic and Evolutionary Microbiology* **67(8)**:2798–2803 DOI 10.1099/ijsem.0.002025.

Ghyselinck J, Velivelli SLS, Heylen K, O'Herlihy E, Franco J, Rojas M, De Vos P, Prestwich BD. 2013. Bioprospecting in potato fields in the Central Andean Highlands: screening of rhizobacteria for plant growth-promoting properties. *Systematic and Applied Microbiology* **36(2)**:116–127 DOI 10.1016/j.syapm.2012.11.007.

Gibb S, Strimmer K. 2012. MALDIquant: a versatile R package for the analysis of mass spectrometry data. *Bioinformatics* **28(17)**:2270–2271 DOI 10.1093/bioinformatics/bts447.

Glick B. 2014. Bacteria with ACC deaminase can promote plant growth and help to feed the world. *Microbiological Research* **169(1)**:30–39 DOI 10.1016/j.micres.2013.09.009.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**:95–98.

Handelsman J. 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews* **68(4)**:669–685 DOI 10.1128/MMBR.68.4.669-685.2004.

Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. 2017. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35(2)**:518–522 DOI 10.1093/molbev/msx281.

Hornik K, Buchta C, Zeileis A. 2009. Open-source machine learning: R meets Weka. *Computational Statistics* **24(2)**:225–232 DOI 10.1007/s00180-008-0119-7.

Hotta Y, Sato J, Sato H, Hosoda A, Tamura H. 2011. Classification of the genus *Bacillus* based on MALDI-TOF MS analysis of ribosomal proteins coded in S10 and spc operons. *Journal of Agricultural and Food Chemistry* **59(10)**:5222–5230 DOI 10.1021/jf2004095.

Howe AC, Jansson JK, Malfatti SA, Tringe SG, Tiedje JM, Brown CT. 2014. Tackling soil diversity with the assembly of large, complex metagenomes. *Proceedings of the National Academy of Sciences of the United States of America* **111(13)**:4904–4909 DOI 10.1073/pnas.1402564111.

Hsieh TC, Ma KH, Chao A. 2016. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution* **7(12)**:1451–1456 DOI 10.1111/2041-210X.12613.

Huber F, Ridder L, Verhoeven S, Spaaks JH, Diblen F, Rogers S, Van der Hooft JJJ. 2021. Spec2Vec: improved mass spectral similarity scoring through learning of structural relationships. *PLOS Computational Biology* **17(2)**:e1008724 DOI 10.1371/journal.pcbi.1008724.

Huschek D, Witzel K. 2019. Rapid dereplication of microbial isolates using matrix-assisted laser desorption ionization time-of-flight mass spectrometry: a mini-review. *Journal of Advanced Research* **19(3)**:99–104 DOI 10.1016/j.jare.2019.03.007.

Jett BD, Hatter KL, Huycke MM, Gilmore MS. 1997. Simplified agar plate method for quantifying viable bacteria. *BioTechniques* **23(4)**:648–650 DOI 10.2144/97234bm22.

Johnston-Monje D, Mousa WK, Lazarovits G, Raizada MN. 2014. Impact of swapping soils on the endophytic bacterial communities of pre-domesticated, ancient and modern maize. *BMC Plant Biology* **14(1)**:233 DOI 10.1186/s12870-014-0233-3.

Johnston-Monje D, Raizada MN. 2011. Conservation and diversity of seed associated endophytes in *Zea* across boundaries of evolution, ethnography and ecology. *PLOS ONE* **6(6)**:e20396 DOI 10.1371/journal.pone.0020396.

Junier T, Zdobnov EM. 2010. The Newick utilities: high-throughput phylogenetic tree processing in the Unix shell. *Bioinformatics* **26(13)**:1669–1670 DOI 10.1093/bioinformatics/btq243.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14(6)**:587–589 DOI 10.1038/nmeth.4285.

Kolton M, Sela N, Elad Y, Cytryn E. 2013. Comparative genomic analysis indicates that niche adaptation of terrestrial *Flavobacteria* is strongly linked to plant glycan metabolism. *PLOS ONE* **8(9)**:e76704 DOI 10.1371/journal.pone.0076704.

Kumar P, Dubey RC, Maheshwari DK. 2012. *Bacillus* strains isolated from rhizosphere showed plant growth promoting and antagonistic activity against phytopathogens. *Microbiological Research* **167(8)**:493–499 DOI 10.1016/j.micres.2012.05.002.

Kämpfer P, Busse HJ, Kleinhagauer T, McInroy JA, Glaeser SP. 2016a. *Sphingobacterium zeae* sp. nov., an endophyte of maize. *International Journal of Systematic and Evolutionary Microbiology* **66(7)**:2643–2649 DOI 10.1099/ijsem.0.001100.

Kämpfer P, Glaeser SP, McInroy JA, Busse HJ. 2016b. *Nocardioides zeicaulis* sp. nov., an endophyte actinobacterium of maize. *International Journal of Systematic and Evolutionary Microbiology* **66(4)**:1869–1874 DOI 10.1099/ijsem.0.000959.

Labbé JL, Weston DJ, Dunkirk N, Pelletier DA, Tuskan GA. 2014. Newly identified helper bacteria stimulate ectomycorrhizal formation in *Populus*. *Frontiers in Plant Science* **5**:579 DOI 10.3389/fpls.2014.00579.

LaMontagne MG. 2020. Gnotobiotic rhizobacterial isolation plant systems and methods of use thereof. US Patent 10801079, October 13, 2020.

LaMontagne MG, Shetty T, Gajjar T, Kayyuru C, Sriram S, Zhang C, Buddharaju P. 2017. HABase: a web-application for the analysis of protein spectra and identification of microbial species. In: *International Conference on Bioinformatics and Computational Biology, BIOCOMP'17*. 77–78.

Lasch P, Beyer W, Nattermann H, Stämmler M, Siegbrecht E, Grunow R, Naumann D. 2009. Identification of *Bacillus anthracis* by using matrix-assisted laser desorption ionization-time of flight mass spectrometry and artificial neural networks. *Applied and Environmental Microbiology* **75(22)**:7229–7242 DOI 10.1128/AEM.00857-09.

Long HH, Sonntag DG, Schmidt DD, Baldwin IT. 2010. The structure of the culturable root bacterial endophyte community of *Nicotiana attenuata* is organized by soil composition and host plant ethylene production and perception. *New Phytologist* **185(2)**:554–567 DOI 10.1111/j.1469-8137.2009.03079.x.

Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America* **95(6)**:3140–3145 DOI 10.1073/pnas.95.6.3140.

Mashiane RA, Ezeokoli OT, Adeleke RA, Bezuidenhout CC. 2017. Metagenomic analyses of bacterial endophytes associated with the phyllosphere of a Bt maize cultivar and its isogenic parental line from South Africa. *World Journal of Microbiology and Biotechnology* **33(4)**:80 DOI 10.1007/s11274-017-2249-y.

McInroy JA, Kloepper JW. 1995. Survey of indigenous bacterial endophytes from cotton and sweet corn. *Plant and Soil* **173(2)**:337–342 DOI 10.1007/BF00011472.

Mosimann C, Oberhänsli T, Ziegler D, Nassal D, Kandeler E, Boller T, Mäder P, Thonar C. 2017. Tracing of two *Pseudomonas* strains in the root and rhizoplane of maize, as related to their plant growth-gromoting effect in contrasting soils. *Frontiers in Microbiology* **7**:2150 DOI 10.3389/fmicb.2016.02150.

Mousa WK, Shearer CR, Limay-Rios V, Zhou T, Raizada MN. 2015. Bacterial endophytes from wild maize suppress Fusarium graminearum in modern maize and inhibit mycotoxin accumulation. *Frontiers in Plant Science* **6(65)**:781 DOI 10.3389/fpls.2015.00805.

Murugaiyan J, Lewin A, Kamal E, Bakuła Z, van Ingen J, Ulmann V, Unzaga Barañano MJ, Humięcka J, Safianowska A, Roesler UH, Jagielski T. 2018. MALDI spectra database for rapid discrimination and subtyping of *Mycobacterium kansasii*. *Frontiers in Microbiology* **9**:587 DOI 10.3389/fmicb.2018.00587.

Nabti E, Bensidhoum L, Tabli N, Dahel D, Weiss A, Rothballer M, Schmid M, Hartmann A. 2014. Growth stimulation of barley and biocontrol effect on plant pathogenic fungi by a *Cellulosimicrobium* sp. strain isolated from salt-affected rhizosphere soil in northwestern Algeria. *European Journal of Soil Biology* **61**:20–26 DOI 10.1016/j.ejsobi.2013.12.008.

Naveed M, Mitter B, Yousaf S, Pastar M, Afzal M, Sessitsch A. 2014. The endophyte *Enterobacter* sp. FD17: a maize growth enhancer selected based on rigorous testing of plant beneficial traits and colonization characteristics. *Biology and Fertility of Soils* **50(2)**:249–262 DOI 10.1007/s00374-013-0854-y.

Niu B, Paulson JN, Zheng X, Kolter R. 2017. Simplified and representative bacterial community of maize roots. *Proceedings of the National Academy of Sciences of the United States of America* **114(12)**:E2450–E2459 DOI 10.1073/pnas.1616148114.

Pandey A, Jain R, Sharma A, Dhakar K, Kaira GS, Rahi P, Dhyani A, Pandey N, Adhikari P, Shouche YS. 2019. 16S rRNA gene sequencing and MALDI-TOF mass spectrometry based comparative assessment and bioprospection of psychrotolerant bacteria isolated from high altitudes under mountain ecosystem. *SN Applied Sciences* **1(3)**:278 DOI 10.1007/s42452-019-0273-2.

Park ST, Kim J. 2016. Trends in next-generation sequencing and a new era for whole genome sequencing. *International Neurourology Journal* **20(Suppl. 2)**:S76–S83 DOI 10.5213/inj.1632742.371.

Petersen LM, Martin IW, Moschetti WE, Kershaw CM, Tsongalis GJ. 2019. Third-generation sequencing in the clinical laboratory: exploring the advantages and challenges of nanopore sequencing. *Journal of Clinical Microbiology* **58(1)**:e01315-01319 DOI 10.1128/JCM.01315-19.

Pruesse E, Peplies J, Glöckner FO. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28(14)**:1823–1829 DOI 10.1093/bioinformatics/bts252.

Raaijmakers JM, Mazzola M. 2012. Diversity and natural functions of antibiotics produced by beneficial and plant pathogenic bacteria. *Annual Review of Phytopathology* **50(1)**:403–424 DOI 10.1146/annurev-phyto-081211-172908.

Rau J, Eisenberg T, Männig A, Wind C, Lasch P, Sting R. 2016. MALDI-UP—an internet platform for the exchange of MALDI-TOF mass spectra. *eJournal* **1**:2–17.

Rijavec T, Lapanje A, Dermastia M, Rupnik M. 2007. Isolation of bacterial endophytes from germinated maize kernels. *Canadian Journal of Microbiology* **53(6)**:802–808 DOI 10.1139/W07-048.

Rocha I, Ma Y, Souza-Alonso P, Vosátka M, Freitas H, Oliveira RS. 2019. Seed coating: a tool for delivering beneficial microbes to agricultural crops. *Frontiers in Plant Science* **10**:e1357 DOI 10.3389/fpls.2019.01357.

Salipante SJ, SenGupta DJ, Cummings LA, Land TA, Hoogestraat DR, Cookson BT. 2015. Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology. *Journal of Clinical Microbiology* **53(4)**:1072–1079 DOI 10.1128/JCM.03385-14.

Sauer S, Freiwald A, Maier T, Kube M, Reinhardt R, Kostrzewa M, Geider K. 2008. Classification and identification of bacteria by mass spectrometry and computational analysis. *PLOS ONE* **3(7)**:e2843 DOI 10.1371/journal.pone.0002843.

Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB. 2011. The real cost of sequencing: higher than you think!. *Genome Biology* **12(8)**:125 DOI 10.1186/gb-2011-12-8-125.

Schmidt D. 2019. *Co-Operation: fast correlation, covariance, and cosine similarity*. R package version 0.6-2. . *Available at* https://cran.r-project.org/package=coop.

Seuylemezian A, Aronson HS, Tan J, Lin M, Schubert W, Vaishampayan P. 2018. Development of a custom MALDI-TOF MS database for species-level identification of bacterial isolates collected from spacecraft and associated surfaces. *Frontiers in Microbiology* **9**:780 DOI 10.3389/fmicb.2018.00780.

Singhal N, Kumar M, Kanaujia PK, Virdi JS. 2015. MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. *Frontiers in Microbiology* **6(285)**:791 DOI 10.3389/fmicb.2015.00791.

Song W, Pijanowski BC, Tayyebi A. 2015. Urban expansion and its consumption of high-quality farmland in Beijing, China. *Ecological Indicators* **54**:60–70 DOI 10.1016/j.ecolind.2015.02.015.

Starostin KV, Demidov EA, Bryanskaya AV, Efimov VM, Rozanov AS, Peltek SE. 2015. Identification of *Bacillus* strains by MALDI TOF MS using geometric approach. *Scientific Reports* **5(1)**:16989 DOI 10.1038/srep16989.

Strejcek M, Smrhova T, Junkova P, Uhlik O. 2018. Whole-cell MALDI-TOF MS versus 16S rRNA gene analysis for identification and dereplication of recurrent bacterial isolates. *Frontiers in Microbiology* **9**:1294 DOI 10.3389/fmicb.2018.01294.

Suzuki R, Shimodaira H. 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22(12)**:1540–1542 DOI 10.1093/bioinformatics/btl117.

Tan Z, Hurek T, Gyaneshwar P, Ladha JK, Reinhold-Hurek B. 2001. Novel endophytes of rice form a taxonomically distinct subgroup of *Serratia marcescens*. *Systematic and Applied Microbiology* **24(2)**:245–251 DOI 10.1078/0723-2020-00002.

Tanaka T, Kawasaki K, Daimon S, Kitagawa W, Yamamoto K, Tamaki H, Tanaka M, Nakatsu CH, Kamagata Y. 2014. A hidden pitfall in agar media preparation undermines cultivability of microorganisms. *Applied and Environmental Microbiology* **80(24)**:7659–7666 DOI 10.1128/AEM.02741-14.

Timperio AM, Gorrasi S, Zolla L, Fenice M. 2017. Evaluation of MALDI-TOF mass spectrometry and MALDI BioTyper in comparison to 16S rDNA sequencing for the identification of bacteria isolated from Arctic sea water. *PLOS ONE* **12(7)**:e0181860 DOI 10.1371/journal.pone.0181860.

Tkacz A, Poole P. 2015. Role of root microbiota in plant productivity. *Journal of Experimental Botany* **66(8)**:2167–2175 DOI 10.1093/jxb/erv157.

Tomachewski D, Galvão CW, De Campos Júnior A, Guimarães AM, Da Rocha JCF, Etto RM. 2018. Ribopeaks: a web tool for bacterial classification through m/z data from ribosomal proteins. *Bioinformatics* **34(17)**:3058–3060 DOI 10.1093/bioinformatics/bty215.

Trenberth KE, Dai A, Van der Schrier G, Jones PD, Barichivich J, Briffa KR, Sheffield J. 2014. Global warming and changes in drought. *Nature Climate Change* **4(1)**:17–22 DOI 10.1038/nclimate2067.

Trifinopoulos J, Nguyen L-T, Von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research* **44(W1)**:W232–W235 DOI 10.1093/nar/gkw256.

Tsang AKL, Lee HH, Yiu S-M, Lau SKP, Woo PCY. 2017. Failure of phylogeny inferred from multilocus sequence typing to represent bacterial phylogeny. *Scientific Reports* **7(1)**:4536 DOI 10.1038/s41598-017-04707-4.

Uhlik O, Strejcek M, Junkova P, Sanda M, Hroudova M, Vlcek C, Mackova M, Macek T. 2011. Matrix-assisted laser desorption ionization (MALDI)-time of flight mass spectrometry- and MALDI biotyper-based identification of cultured biphenyl-metabolizing bacteria from contaminated horseradish rhizosphere soil. *Applied and Environmental Microbiology* **77(19)**:6858–6866 DOI 10.1128/AEM.05465-11.

Wang X-Q, Bi T, Li X-D, Zhang L-Q, Lu S-E. 2015. First report of corn whorl rot caused by *Serratia marcescens* in China. *Journal of Phytopathology* **163(11–12)**:1059–1063 DOI 10.1111/jph.12366.

Wang X, Qu C, Wang W, Zheng Z, Liu F, An M, Miao J. 2019. Complete genome sequence of marine *Bacillus* sp. Y-01, isolated from the plastics contamination in the Yellow Sea. *Marine Genomics* **43(1)**:72–74 DOI 10.1016/j.margen.2018.05.002.

Weis CV, Jutzeler CR, Borgwardt K. 2020. Machine learning for microbial identification and antimicrobial susceptibility testing on MALDI-TOF mass spectra: a systematic review. *Clinical Microbiology and Infection* **26(10)**:1310–1317 DOI 10.1016/j.cmi.2020.03.014.

Wright ES, Yilmaz LS, Noguera DR. 2012. DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. *Applied and Environmental Microbiology* **78(3)**:717–725 DOI 10.1128/AEM.06516-11.