APPLICATION OF WHOLE GENOME SEQUENCING AND MALDI-TOF TO IDENTIFICATION OF BACILLUS SPECIES ISOLATED FROM CLEANROOMS AT NASA JOHNSON SPACE CENTER

by

Farnaz Mazhari, B.S.

THESIS

Presented to the Faculty of

The University of Houston-Clear Lake

In Partial Fulfillment

Of the Requirements

For the Degree

MASTER OF SCIENCE

in Biotechnology

THE UNIVERSITY OF HOUSTON-CLEAR LAKE

DECEMBER, 2021

APPLICATION OF WHOLE GENOME SEQUENCING AND MALDI-TOF TO IDENTIFICATION OF BACILLUS SPECIES ISOLATED FROM CLEANROOMS AT NASA JOHNSON SPACE CENTER

by

Farnaz Mazhari

APPROVED BY

_____
Michael G. LaMontagne, Ph.D., Chair

_____
Richard Davis, Ph.D., Committee Member

_____
Aaron Regberg, Ph.D., Committee Member

_____
Lory Z. Santiago-Vázquez, Ph.D., Committee Member

RECEIVED/APPROVED BY THE COLLEGE OF SCIENCE AND ENGINEERING:

_____
David Garrison, Ph.D., Interim Associate Dean

_____
Miguel A. Gonzalez, Ph.D., Dean

Acknowledgements

To begin, I want to express my gratitude to the University of Houston-Clear Lake and NASA Johnson Space Center's ARES division for allowing me to perform collaborative research. This experience has broadened my horizons and helped me become a more well-rounded scientist. Thank you to my parents who overcame a lot to build a life in the U.S that allowed me opportunities I otherwise would not have had.

Thank you, Dr. LaMontagne, for believing in me as an undergraduate student with a strong curiosity about research. The checkerboard assay project means so much to me. It symbolizes the beginning of me finding my niche and passion as a researcher. It was that feeling of really believing in something that drove me and still drives me to this day. Working in your lab is why I chose a career in microbiology. This project would not have been possible without your expertise and serving as my thesis chair. Thank you, Dr. Santiago, for serving on my committee, offering your guidance and helping me write my proposal. Thank you to my UHCL friends who have always had my back.

Thank you, Dr. Aaron Regberg and Dr. Richard Davis, for serving on my committee and allowing me the extraordinary opportunity to be a part of your team at NASA Johnson Space Center. I am fortunate to be able to contribute to the future of space exploration. Your guidance and expertise have helped me develop into a far better scientist than I was when I first started. I truly love my work and have a drive to conduct research that will advance the goals of our geo-microbiology team.

This would not have been possible without the indispensable help and support of the microbiology lab at NASA Johnson Space Center. Thank you all for making me feel welcome and always going out of your way to help me accomplish my goals. A great deal of my knowledge about microbiology, sequencing technologies and genomics is because of the scientists in that lab volunteering their time and expertise

ABSTRACT

APPLICATION OF WHOLE GENOME SEQUENCING AND MALDI-TOF TO IDENTIFICATION OF BACILLUS SPECIES ISOLATED FROM CLEANROOMS AT NASA JOHNSON SPACE CENTER

Farnaz Mazhari
University of Houston-Clear Lake, 2021

Thesis Chair: Michael LaMontagne, Ph.D.

Astromaterial cleanrooms at NASA Johnson Space Center are built environments that hold samples, such as lunar rocks, from different space exploration missions. *Bacillus sp.* are frequently detected in routine microbial monitoring of these facilities. Since this, and related genera, can form endospores that can withstand harsh conditions, they could contaminate astromaterials. This could confound searches for extraterrestrial life. Whole genome sequencing (WGS) is widely used for identifying bacterial strains and tracking their source; however, WGS is expensive and time consuming. Matrix-assisted laser desorption ionization– time of flight mass spectrometry (MALDI-TOF) shows promise as a low-cost, rapid method of identifying strains of bacteria, but few studies have compared this proteomics method to WGS. To evaluate a high throughput method of tracking the source of contamination of this built environment, WGS and MALDI-TOF was conducted on 18 bacterial strains isolated from surfaces in astromaterials cleanrooms. WGS identified 14 *Bacillus*, 2 *Paenibacillus*, 1 *Solibacillus* and 1 *Alcaligenes* strains.

These isolates showed similarity to strains commonly observed in spacecraft assembly cleanrooms at other facilities. Cluster analysis of mass spectra generated by MALDI-TOF grouped strains together that were greater than 94% similar to each other in terms of amino acid sequences of single copy core genes, as assessed by WGS. This suggests that MALDI-TOF and WGS results are consistent with each other and MALDI-TOF can rapidly identify strains of *Bacillus sp.* isolated from cleanroom environments with a resolution comparable to WGS. Based on phylogenomic analysis, these results also suggest the presence of a cosmopolitan class of *Bacillus sp*. that are more likely to be found in cleanrooms and similar built environments than in natural systems.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I:

INTRODUCTION

Advanced Curation of Astromaterials Research and Exploration Science (ARES) Curation laboratories

Advanced curation is a cross-disciplinary field developed to improve curation techniques for astromaterial collections. Continued research and development of sample collection, handling, characterization and analysis has improved curation and acquisition practices (McCubbin et al., 2021). Beginning with the lunar samples collected during the moon landing in 1969, the NASA Astromaterials Acquisition and Curation Office has maintained clean rooms for storing extraterrestrial samples from the moon, meteorites, cosmic dust, asteroids, comets, solar wind particles, and space exposed hardware (Figure 1). Low nutrition levels (oligotrophic) and low humidity levels in certified cleanrooms inhibit proliferation and abundance of microbial organisms (Carosso, n.d.) Regular cleaning, air filtration, with high-efficiency particulate air (HEPA) filters, and continual humidity and temperature control, render these facilities inhospitable to microbial life. Despite these fastidious controls, these cleanrooms contain bacteria and fungi (Regberg et al., 2018).

In the context of microbial ecology, these environments are extreme built environments and are analogous to cleanrooms maintained in medical centers and semiconductor manufacturing (Favero et al., 1968; Venkateswaran et al., 2001). To protect people and materials in the environments from harmful microorganisms, it is important to identify these contaminants and track their source.

Figure 1. *Photo of the Lunar Laboratory at Nasa Johnson Space Center* (“*CollectSPACE”, n.d.)*

Microbiological Contamination of Astromaterials

Limiting microbial contamination and growth is important to preserve the integrity of astromaterials (Regberg et al., 2018). Microorganisms can degrade and change the composition of minerals. This could create organic molecules that can be misinterpreted as biosignatures of extraterrestrial life (Rummel, 2001). Organic acid generation and direct enzymatic oxidation/reduction of transition metals by microbes can alter the mineral composition of rocks. These microbial weathering processes may include enzymatic oxidation of iron and manganese. Production of organic acids can dissolve and chelate mineral matrices, chemically reduce mineral and dissolve rock surfaces (Lian et al., 2008). Diverse microbes can also create secondary metabolites that have unique amino acid analogs and non-ribosomal peptides with both L- and D-chirality (Gokulan et al., 2014). These organic signatures are used to detect life (Pohorille and Sokolowska, 2020).

Microbes in built environments, such as spacecraft cleanroom facilities, must resist extreme physical and chemical conditions. Cleanroom practices have been consistently and strictly implemented, under planetary protection requirements, for decades; these protocols appear to have selected for bacteria that are multi-resistant to various contamination control procedures such as peroxide and UV-radiation (Tirumalai et al, 2018; Link et al, 2004).

Contamination of astromaterials and their facilities could confound identification of extraterrestrial life because it would be difficult to differentiate between terrestrial and extraterrestrial organisms. Plans for Mars exploration and the development of a commercial space flight industry have elevated these concerns because of the prospect of microbes being able to survive interplanetary transfer (Tirumalai et al, 2018; Schuerger et al, 2003). Microbes can consume amino acids, which are of particular interest as biosignatures (Pohorille and Sokolowska, 2020). The search for extraterrestrial life involves ultrasensitive technologies that can detect cells and biomarkers. If these technologies incorrectly detected Earth-derived cells or biomarkers, it would jeopardize the search for extraterrestrial life (Rummel, 2001).

Bacillus Species

Bacillus species comprise a large percentage of the microbes in the library of isolates generated from cleanrooms at NASA Johnson Space Center (Figure 2). Members of this genus are rod-shaped, endospore-forming, aerobic or facultatively anaerobic, Gram-positive bacteria (Burns, 1991). Many of *Bacillus sp*. isolated from cleanrooms show distinct phenotypes, such as various growth patterns and colors.

This represents a significant knowledge gap because this group can form endospores that can survive for years under extreme conditions (Venkateswaran et al., 2004). Spores in their dormant stage have no observable metabolism and resist

inactivation by wet and dry heat, UV and gamma radiation, intense desiccation, and oxidizing chemicals like peroxide (Nicholson et al., 2000). *Bacillus* strains isolated from spacecraft assembly facilities at the NASA Jet Propulsion Laboratory (JPL) show unusually high resistance to environmental stressors (Tirumalai et al., 2013; Tirumalai et al, 2018; Schuerger et al, 2003). For example, *B. pumilus* SAFR-032 tolerates simulated Mars environmental conditions (Sella, et al., 2015).

Sequencing the 16s rRNA gene is routinely used to identify bacterial isolates. This approach depends on a gene of 1,500 or fewer base-pairs. This method is time consuming and has limited resolving capability for closely related *Bacillus* species. For example, 16S rRNA gene sequences of three *Bacillus* species are 99 % identical (Yamada et al., 1999). This lack of variation *Bacillus* species limits the utility of rRNA gene sequencing in monitoring these genera (Tirumalai et al., 2018; Espariz et al., 2016). This represents a significant knowledge gap because of the high degree of functional diversity of species. To address this, multiple locus sequence typing and, increasingly, whole genome sequencing are commonly used to identify *Bacillus* species.

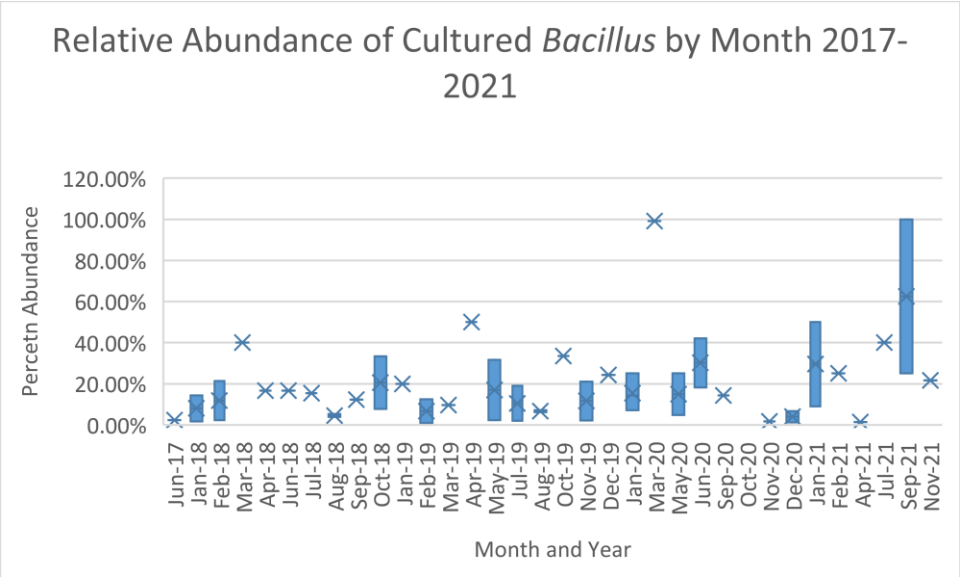

Figure 2. *Relative abundance of Bacillus isolates found in astromaterials clean rooms at NASA Johnson Space Center by month between 2017 to 2021 based on routine microbial sampling results. The median relative abundance is 45%.*

MALDI – TOF MS

Matrix-assisted laser desorption – time of flight mass spectrometry (MALDI-TOF MS) systems provide strain-level identification of microbes at a low cost. These systems use pattern matching between mass and reference spectra; (Ahmad et al., 2012; Singhal et al., 2015). Bacterial identification utilizing MALDI-TOF MS is superior to conventional biochemical systems for species identification (Seng et al., 2009). However, mass spectral databases used by MALDI-TOF MS systems contain a poor representation of environmental bacteria. This hampers the application of MALDI-TOF's efficiency to routine monitoring of built and natural environments. To address this compilation of spectra are available for particular species (Böhme et al., 2012) or for systems, such as spacecraft assembly facilities (Seuylemezian et al., 2018).

MALDI-TOF MS collects unique molecular signatures that are representative of a larger range of proteins and can clearly differentiate between two closely related species. This technology shows promise to replace routine 16S rRNA sequencing. MALDI-TOF's

capacity to distinguish taxonomic groups can detect novel species. Identification of isolates quickly and accurately is necessary for culture-dependent monitoring programs to effectively track contamination (Seuylemezian et al., 2018).

Sample are prepared for MALDI TOF MS analysis by coating with an energy-absorbent organic matrix substance, such as a-cyano-4-hydroxycinnamic acid($C_{10}H_7NO_3$). After the matrix crystallizes, the sample encased within it is ionized with a laser to generate ions that are accelerated to a detector. The time it takes the ions to reach the detector corresponds to their mass-to-charge ratio (m/z) (Yates, 1998). A range of dimension less m/z values, from 2,000 to 12,000 are used herein for microbial identification (Figure 3).

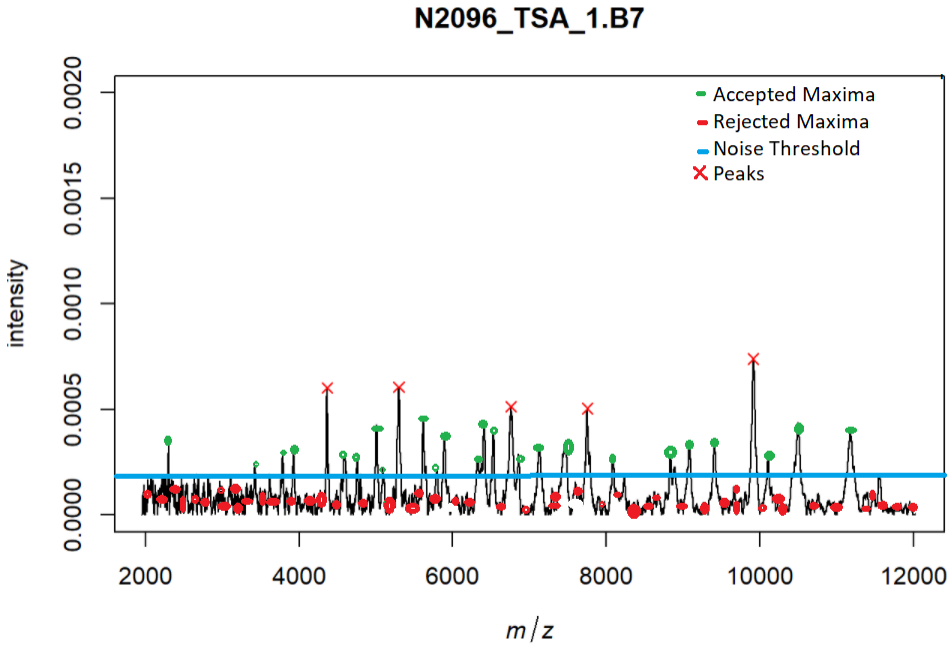

Figure 3. *Detailed view of spectra for sample 2096 TSA-1 showing plot of signal to noise ratio and peak detection. Accepted maxima indicate acceptable range for peak detection*

Hybrid Whole Genome Sequencing

The emergence of next-generation sequencing technology (NGS) has resulted in the widespread ability to generate a large amount of microbial sequencing data in a short amount of time (Chen et al, 2020). As a result, whole genome sequencing (WGS) has

grown in popularity as a method for microbial identification (Didelot et al., 2012). The most popular short read sequencing platforms are made by Illumina (De Maio et al., 2019). These NGS devices can generate millions of paired-end reads at 50-300 bp with a low (0.1 percent) error rate. However, short reads fail to resolve repeated, ambiguous regions of the genome, resulting in fragmented assemblies. Long-read technology, such as the Oxford Nanopore Technologies (ONT) device, can generate read in the range of millions of base pairs, which can cross repeated regions. These long reads improve quality of assemblies (Chen et al., 2020). However, ONT reads have high error rates, ranging from 5 and 15 percent (De Maio et al., 2019; Rang et al., 2018). Hybrid genome assembly combines the strengths of these two sequencing platforms and gives rise to an accurate and near-complete genome assembly (Chen et al, 2020; Didelot et al., 2012).

Hypothesis

Overall, the goal of this project is to see if MALDI-TOF MS can discriminate *Bacillus* strains isolated from cleanrooms with resolution comparable to whole genome sequencing.

CHAPTER II:

MATERIALS AND METHODOLOGY

Sample Collection and Processing

As part of routine microbial monitoring of clean room laboratories, samples were collected from the Meteorite, Cosmic Dust, Star Dust, Lunar, Genesis Hayabusa, and Cold Curation laboratories (Figures 4 - 8). Puritan Brand Sterile, DNA Free, Foam Tipped Applicator (Part Number: 25-1805 1PF RND FDNA) and Puritan Brand Sterile Polyester Tipped Applicators (Part Number: 25-1000 1PD) were used to sample an area of 300 cm^2 modified according to the guidelines of the NASA standard assay ("*Handbook for the microbial examination of space hardware", 2010)*. Samples were then transported to a lab and the swabs were resuspended in 15 mL of phosphate buffered saline (PBS) and vortexed for 10 seconds. A total of 4 Tryptic soy agar (TSA), 2 Blood agar (BA), 2 Reasoner's 2A agar (R2A), 2 Sabouraud Dextrose agar (SDA), 1 Sabouraud Dextrose Chloramphenicol agar (SDA+C), and 1 Potato Dextrose agar (PDA) plate were inoculated with the PBS suspension. TSA plates were analyzed following a 48 H incubation at 35°C. BA plates were analyzed following a 48 H incubation at 37°C. R2A plates were analyzed following a 7 days incubation at 26°C. PDA plates were analyzed following a 7-day incubation at 30°C.

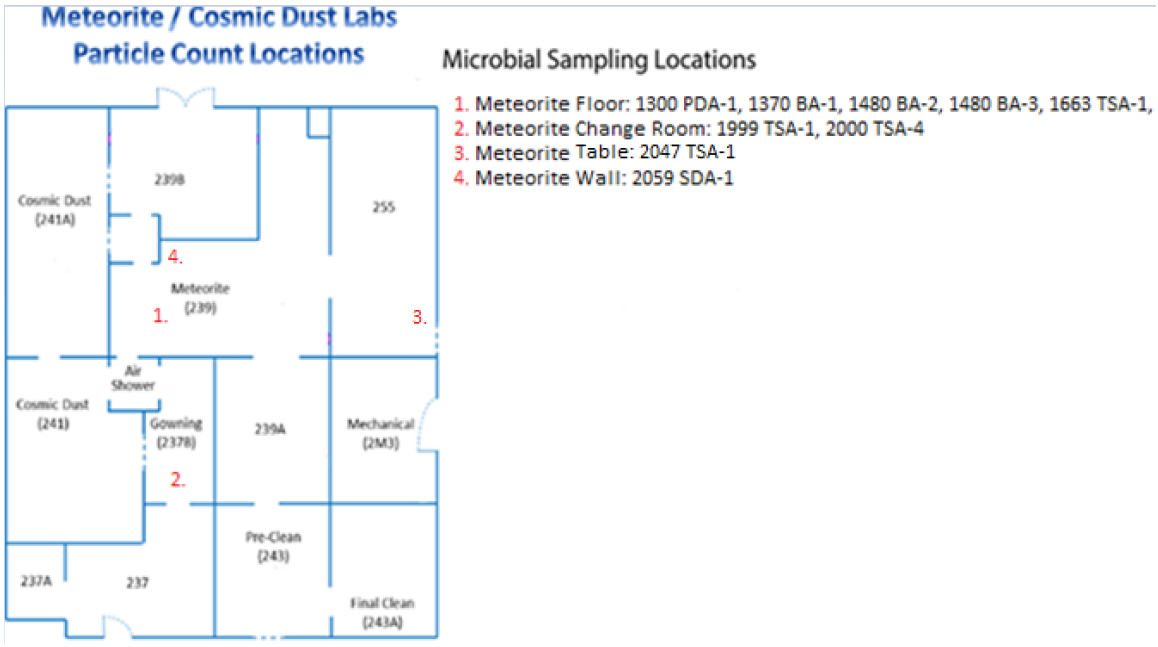

Figure 4. *Meteorite and Cosmic Dust Lab floor plan and sampling locations.*

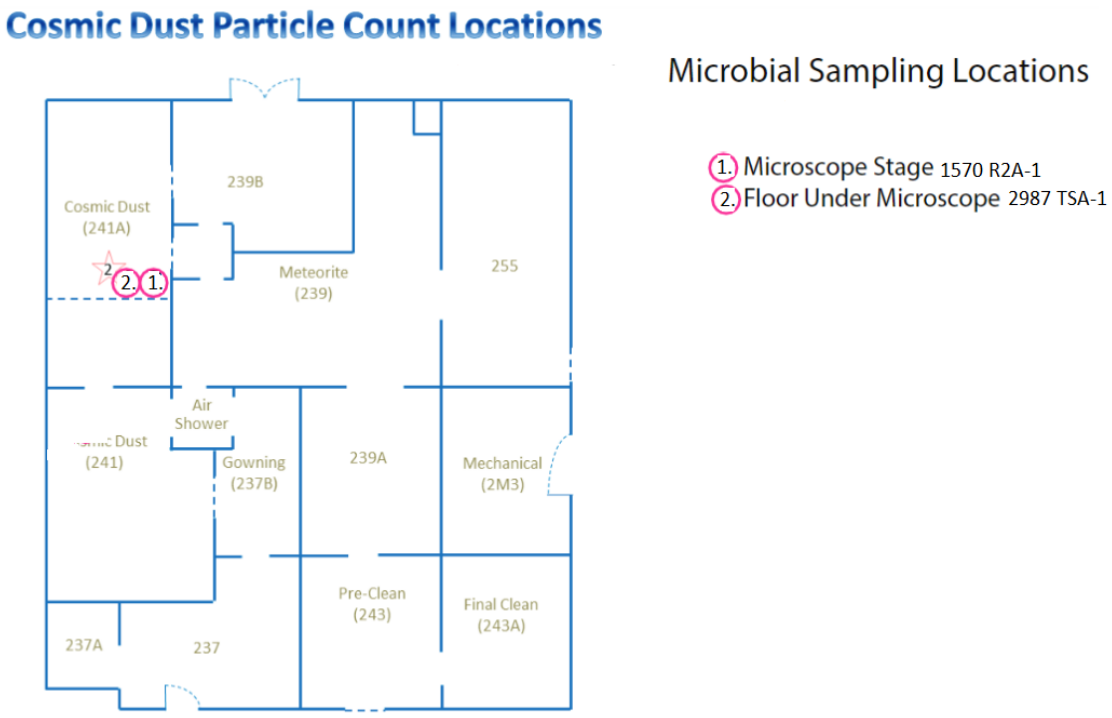

Figure 5. *Cosmic Dust Lab floor plan and sampling locations*

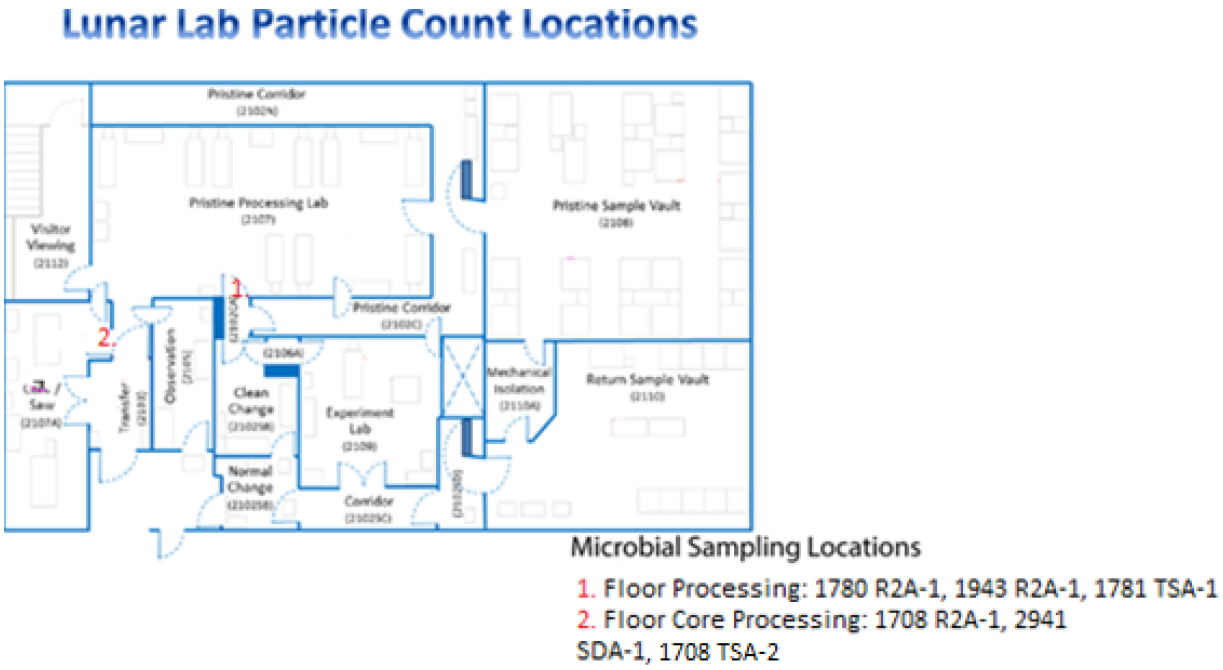

Figure 6. *Lunar lab floor plan and sampling locations.*

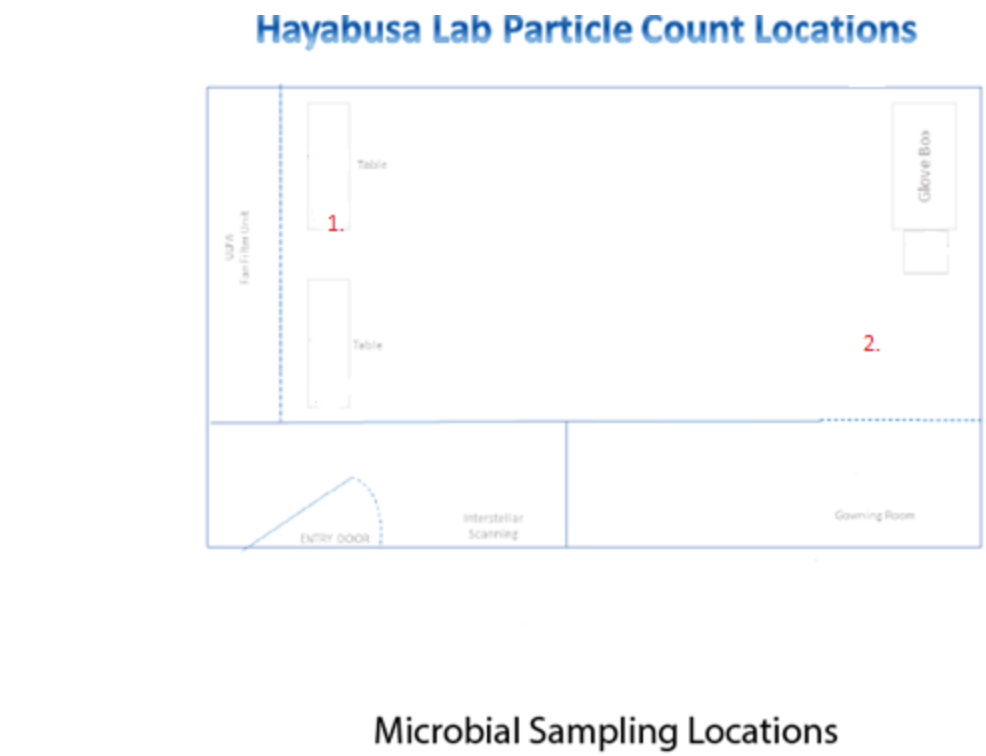

Figure 7. *Hayabusa lab floor plan and sampling locations.*

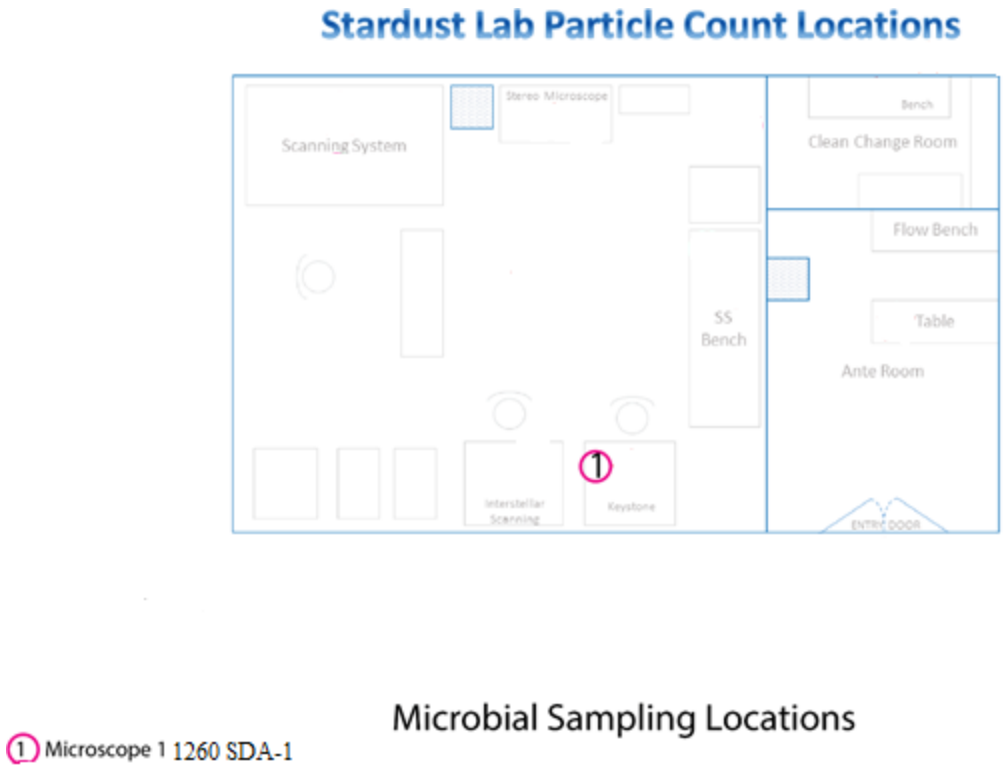

Figure 8. *Stardust lab floor plan and sampling location.*

Sample Identification and Representative Data Set

A total of 23 isolates were sub-cultured onto TSA plates and identified at the genus level using the Vitek 2 (bioMérieux USA, St. Louis, MO) or Applied Biosystems Microseq 3500 16S rDNA Bacterial Identification System (Applied Biosystems, Waltham, MA) prior to the start of this study. These isolates were selected at random and consisted of 20 *Bacillus, 2 Actinobacter* and *1 Paenibacillus*. Isolates were sub-cultured onto TSA plates and incubated for 24 hours at 35°C. A 10 ul loopful of bacteria was inoculated into a micro bank tube (Pro-Lab Diagnostics, Round Rock, TX) vortexed for 30 seconds and placed into the -80° C freezer.

MALDI TOF MS

A total of 18 Isolates were cultured for 24 hours at 35º C on TSA plates. 300 µl of Ultra-Pure Water, High Performance Liquid Chromatography Mass Spectrometer (HPLC/MS) Grade 11 was added to each 1.5 ml microfuge tube (Eppendorf, Hamburg, Germany). A loopful of a single colony was inoculated from the plates into separate microfuge tubes for each isolate and vortexed thoroughly before treating with900 ul of Ethanol, 100% HPLC/MS Grade 12 and then vortexed for 30 seconds. Tubes were centrifuged at maximum (13,000 g) speed for 2 minutes. Ethanol was decanted and pellet was centrifuged for another minute. Excess ethanol was removed with a pipette and left to air dry. After about five minutes of drying, 50 µl of 70% HPLC/MS grade formic acid was added and tubes were vortexed for 30 seconds and incubated for 5 minutes at room temperature. An equal volume (50 µl) of 100% HPLC/MS Grade acetonitrile was added under a fume hood and centrifuged at maximum speed (13,000 g) for 2 minutes. Taking care to avoid the pellet, 70 µL of supernatant was transferred to a fresh 1.5 ml microtube and these extracts were stored at -20° C.

For analysis, one µL of supernatant was pipetted onto a MALDI steel target in appropriate position and left to air dry. Two size standards were prepared by applying one µl of bacterial test standard (BTS, Bruker p/n 8255343) solution and two spots were left blank. After the spots dried, each spot received one µl of matrix solution and the target was allowed to thoroughly air dry. Targets 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 16802). Positive-ion mass spectra were acquired on a Bruker Ultraflextreme MALDI TOF/TOF mass spectrometer as described previously (LaMontagne et al., 2021). A Bruker Ultraflextreme MALDI TOF/TOF mass spectrometer was used to collect positive-ion mass spectra. The

following parameters were employed in linear detection mode: 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. Suppression (deflection) of the matrix was adjusted to 1,500 m/z. The repetition rate of the laser was 667 Hz, and the Smartbeam parameter was set to "3 medium." The detection was set for a low mass range of 1,880–20,000 Da, with real-time smoothing was turned off, baseline offset 0.2 percent, analog offset 2.1 mV, and real-time smoothing turned off. The target was moved in a complete sample pattern, with 50 shots fired at 24 raster spots (1,200 total shots) and a 2,000-mm diameter limit.

DNA Isolation

For WGS, 24 Isolates were removed from -80° C freezer and sub-cultured onto TSA plates and incubated at 35ºC for 24 hours. Isolates were once again sub-cultured from incubated plates after 24 hours on tryptic soy agar plates at 35ºC. Individual colonies were inoculated onto Hardy dx (cat no. Q85) liquid TSA tubes. Tubes were incubated at 35ºC on shaker at 200 rpm with loosened lids. After 24 hours, liquid cultures were centrifuged at 10,000 x g for 10 minutes. The cell pellet was resuspended in 1 ml sterile PBS was centrifuged at 13,000 x g for 2 minutes to pellet the cells and the pellets were then were resuspended in 480 ul Ethylenediaminetetraacetic acid (EDTA) and 120 ul of 10 mg/ml, lysozyme, egg white, ultra-pure grade (Amresco, Solon, OH) was added to the resuspended cell pellet for cell lysis. Isolates were incubated at 37ºC for 30 - 60 minutes and centrifuged at 13,000 x g for 2 minutes. The supernatant was removed. Promega Wizard™ Genomic DNA Purification Kit was used for DNA extraction and the protocol was followed by the given manufacturer (Promega, Madison, Wisconsin) DNA size and integrity was assessed by Agilent 4200 TapeStation System.

Invitrogen™ Qubit™

The Qubit 1X dsDNA High Sensitivity assay kit was used to quantify DNA concentration from extracted DNA of samples. Reagents were left at room temperature for 30 minutes. For samples with concentrations above the kit's detection limit of 100 ng a 10-fold dilution was made by combining 9 µl water and 1 µl of DNA extraction sample. Low and high standards were used by combining 190 µl buffer and 10 µl standard for each. Samples were prepared by combining 198 µl buffer and 2 µl sample in duplicates for verification. Samples were vortexed and left at room temperature for 2 minutes. Qubit was run using the 1X dsDNA parameter. Standards were read first from low to high followed by the samples.

Agilent 4200 TapeStation System

The TapeStation gel was used to measure the size and integrity of the DNA (DIN). The DNA concentration was not used as a primary deciding factor because Oxford Nanopore Technology recommends the Qubit as the DNA quantification method. Samples with low molecular weight and DIN were re-extracted.

MinION Mk1C Sequencing

DNA isolated samples were sequenced with the MinION Mk1C from Oxford Nanopore Technologies using the Rapid Barcoding Sequencing (SQK-RBK004) RBK_9054_v2_revQ_14Aug2019 kit and protocol. Run parameters are described in Table 1.

Table 1. *MinION Mk1C run parameters*

Flow Cell Type	FLO-MIN106
Kit	SQK-RBK004
Initial Bias Voltage	-180 mV
FAST5 Output	Enabled
FASTQ Output	Enabled
Active Channel Selection	Enabled
Basecalling	on
Specified Run Length	72 hours
FAST5 Reads per File	4000
FASTQ Reads per File	4000
Mux Scan Period	1 hour 30 minutes
Reserved Pores	0%
Basecall Model	High-accuracy basecalling
Barcoding	trim_barcodes="off",require_barcodes_both_ends="off",
	detect_mid_strand_barcodes="off",min_score=40
Read Filtering	min_qscore=7
MinKNOW Core	Version 4.2.4
Bream	Version 6.1.4
Guppy	Version 4.3.4

Illumina MiSeq Sequencing

Isolates were sequenced on the Illumina MiSeq using the MiSeq Reagent Kit v3 with paired-end reads. The Illumina DNA prep reference guide Document # 1000000025416 v09 was followed.

Bioinformatics

(A) MALDI-TOF MS Data Analysis

MALDI-TOF MS data analysis was conducted according to the methods described previously (LaMontagne et al., 2021). Mass spectra were analyzed by cluster

analysis using an R script containing a compilation of packages. MALDIquantFOREIGN was used to import data and MALDIquant for smoothing and transformation, baseline correction, intensity calibration, peak detection, peak alignment, peak binning and feature matrix (Gibb & Strimmer, 2012). Pvclust was used to provide boot strap probability values for clusters in the dendrogram (Suzuki & Shimodaira, 2006). Philentropy was used to calculate distance and similarity measures (Drost, 2018). iNEXT was used for rarefaction analysis (Hsieh et al, 2016). ggplot2 was used to for biplots (Wilkinson, 2011). RWeka was to test the coherence of MALDI-TOF taxonomic units (MTUs) (Hornik et al., 2009). This script had two optimization loops that iteratively sampled random *values* for seven parameters: half-window for smoothing, baseline *removal, half-*window for alignment, alignment tolerance, signal to noise ratio (SNR) for alignment, half-window for peak detection, and peak detection SNR. The parameters that improved the number of peaks shared between pairs of average mass spectra generated from BTS were discovered in the first loop as Jaccard coefficients calculated with the the R package philentropy (Drost, 2018) between pairs of average mass spectra created from BTS. The first optimization's output was transferred to a step which then performed a quality control analysis to identify noisy spectra. On the spectra that passed quality check, the second loop chose the parameters that minimized the overlap in cosine similarity values used to distinguish between species of closely related and distantly related isolates. These cosine similarities were normalized with the method $y = y0 + x/ (x + 0.2)$, where x is the Jaccard coefficient, y0 is the average cosine similarity when $x = 0$, and y is a prediction cosine value. The script then used the RWeka package to train a machine learning algorithm and did cluster analysis to construct MALDI-TOF taxonomic units (MTUs). (Deutsch et al., 2017). A pairwise similarity of amino acids derived from single core

genes was used to compare to clustering of MTUs to estimate identification accuracy (Figure 12).

(B) Quality Control of MinION Mk1C Reads

FastQC ("*Babraham Bioinformatics,"* n.d.) and MultiQC (Ewels et al., 2016) were used for quality control on the *MinION Mk1C* reads.

(C) EDGE Bioinformatics

The Bioinformatics platform EDGE Bioinformatics (Li et al., 2017) was used to analyze and assemble the raw sequence data using the hybrid whole genome pipeline with a set of tools described below.

a. Count Fastq

Count for total raw reads, bases and mean read length was conducted.

b. Quality Trim and Filter

Reads were trimmed at a quality level of 30. The minimum sequence length was set to 50. The "N" base cut off was set to 10. Low complexity filter ratio, Maximum fraction of mono-/di-nucleotide sequence was set to 0.85. Reads with adapters or contamination sequences were trimmed using Porechop ("Porechop", n.d.) and greater than 15 poly A was trimmed off.

c. Assembly

De novo assembly was done using Unicycler (Wick et al., 2017) at a minimum contig length of 200 bp and a minimum of 2000 reads. Miniasm (Li, 2018) was used to find consensus sequences at a minimum of 3.

d. Reads Mapping to Contigs

Read Mapping was done using bowtie (Langmead, 2015) at a max clip (number of clipped read characters) of 50 and a min mapq (mapping quality score) of 42.

e. Reads Taxonomy Classification

Taxonomy classification was done using the tools speDB-b, gottcha-speDB-v, gottcha2-speDB-b, pangia, metaphlan2, kraken2, centrifuge (Altschul et al., 1990)

f. Contigs Taxonomy Classification

Contig taxonomy classification was done using BLAST ("National Center for Biotechnology Information" n.d.)

g. Contigs Annotation

Contig annotation was done using Prokka (Seemann et al., 2014) at a cut size of 700 bp.

h. Gene Family Analysis

Rapsearch2 (Zhao et al., 2012) was used for gene family analysis and detected antibiotic resistance and virulence genes.

(D) Bandage

Bandage was used to visualize the assembly graphs with connections (Wick, et al 2015).

(E) Phylogeny

a. 16S rRNA Phylogenetic Tree

16s genes were extracted from the whole genome FASTA files and used to make a phylogenetic tree on the phylogeny.r platform (Eddy, 2011). MUSCLE (Edgar, 2004) (v3.8.31) was used to align the sequences, with the maximum accuracy setting (MUSCLE with default settings).

Gblocks (Castresana, 2000) (v0.91b) was used to remove ambiguous sections (i.e., regions with gaps and/or poorly aligned) after alignment using the following parameters:

-after gap clearing, a block's minimum length is 10

-In the final alignment, no gaps were allowed.

-all segments with more than 8 consecutive nonconserved locations were rejected

-For a flank position, the minimum number of sequences is 85 percent.

The maximum likelihood technique employed in the PhyML program (Guindon and Olivier, 2003) (v3.1/3.0 aLRT) was used to reconstruct the phylogenetic tree. To account for rate heterogeneity between sites, the HKY85 substitution model (Hasegawa et al., 1985) was chosen, with an estimated fraction of invariant sites (of 0.591) and four gamma-distributed rate categories. The gamma shape parameter (gamma=0.712) was calculated directly from the data. The aLRT test was used to examine internal branch reliability (SH-Like) (Anisimova & Olivier, 2006).

TreeDyn (Chevenet et al., 2006) was used to create a graphical representation and alter the phylogenetic tree (v198.3).

b. Phylogenomic Tree

A phylogenomic tree using 119 single copy genes (SCGs) from the phylum Firmicutes was created using GToTree. Genes sets were included already in the program based on genes that occur 90 percent of the time in the phylum. Accession assemblies from genomes of similar taxa were used as references. The amino acid sequences for each of the reference accessions was downloaded. Open-reading frames on all the input fasta files (using prodigal) were called. Target genes within them (using HMMER3) were identified. Completion/redundancy based on the target genes were estimated. Gene hits based on length and genomes based on how many hits they have of the target genes were filtered out. Required gap sequences for genomes missing any genes were added in. Each individual gene-set was aligned with muscle. Automated trimming of alignments was performed with Trimal (Capella-Gutierrez et al., 2009). All alignments were concatenated together. Trees were made with FastTree by default (Price et al., 2019).

CHAPTER III:

RESULTS

Representative Data Set

Samples 1461 R2A-1, 1708 TSA-2, 2090 TSA-1, 1480 BA-2, 1735 TSA-3, 2069 TSA-4, 2059 SDA-1, 2069 TSA-3 were identified using MALDI-TOF only. 1663 TSA-1, 1570 R2A-1, 1813 SDA-1, 1480 BA-3, 1708 R2A-1, 1943 R2A-1, 2047 TSA-1, 1370 BA-1, 1735 SDA-2, 1781 TSA-1 and 1735 TSA-3 were identified using both MALDI-TOF and whole genome sequencing. 1708 R2A-1, 2069 SDA-1, 2096 TSA-1, 2933 TSA-1, 2941 SDA-1, 2987 TSA-1, 3103 SDA-1 and 2090 TSA-1 were identified only using whole genome sequencing (Table 1 - 5).

Table 2. *Bacillus representative data set identified at genus level using Vitek2 or Applied Biosystems Microseq 3500 16S rDNA Bacterial Identification System*

Sample ID	Date Sampled	Lab recovered from	Sample Number	Media
Bacillus species	8/27/2019	Stardust	1260	SDA-1
Bacillus species	9/3/2019	Meteorite	1300	PDA-1
Bacillus species	10/1/2019	Meteorite	1370	BA-1
Bacillus species	10/31/2029	Hamburg	1461	R2A-1
Bacillus species	11/5/2019	Meteorite	1480	BA-2
Bacillus species	11/5/2019	Meteorite	1480	BA-3
Bacillus species	12/3/2019	Cosmic	1570	R2A-1
Bacillus species	1/21/2020	Meteorite	1663	TSA-1
Bacillus species	1/29/2020	Lunar	1708	TSA-2
Bacillus species	2/12/2020	Hayabusa	1735	SDA-2
Bacillus species	2/12/2020	Hayabusa	1735	TSA-3
Bacillus species	2/25/2020	Lunar	1780	R2A-1
Paenibacillus amylolyticus	2/25/2020	Lunar	1781	TSA-1
Bacillus species	3/2/2020	Cold	1813	SDA-1
Bacillus species	3/30/2020	Lunar	1943	R2A-1
Actinobacter radiosistans	4/21/2020	Meteorite	1999	TSA-1
Acinetobacter radioresistens	4/21/2020	Meteorite	2000	TSA -4
Bacillus fortis	5/19/2020	Meteorite	2047	TSA-1
Bacillus simplex	5/19/2020	Meteorite	2059	SDA-1
Bacillus species	5/19/2020	Meteorite	2069	TSA -4
Bacillus species	5/19/2020	Meteorite	2069	TSA-3
Bacillus species	5/19/2020	Meteorite	2069	SDA-1
Bacillus species	5/20/2020	Hayabusa	2090	TSA -1
Bacillus licheniformis	5/20/2020	Hayabusa	2096	TSA -1
Bacillus species	3/23/2021	Lunar	2933	TSA-1
Paenibacillus species	3/23/2021	Lunar	2941	SDA-1
Bacillus species	3/30/2021	Cosmic	2987	TSA-1
Bacillus species	5/18/2021	Meteorite	3103	SDA-1

Whole Genome Sequencing

(A) Invitrogen™ Qubit™ 4 Fluorometer

Yield from DNA extractions ranged from 5 – 181 ng µl (Table 3).

Table 3. *Qubit DNA concentrations averages from successful DNA extractions*

Isolate	Genus	Concentration (mean) ng/ul
2090 TSA-1	*Bacillus*	9
1570 R2A-1	*Bacillus*	132
1461 R2A-1	*Bacillus*	71
1663 TSA-1	*Bacillus*	17
1813 SDA-1	*Bacillus*	7
1735 SDA-2	*Bacillus*	54
2096 TSA-1	*Bacillus*	48
1708 SDA-2	*Brevibacterium*	60
2047 TSA-1	*Solibacillus*	50
2069 SDA-1	*Bacillus*	29
1708 R2A-1	*Bacillus*	21
1735 TSA-3	*Alcaligenes*	48
3103 SDA-1	*Bacillus*	181
1781 TSA-1	*Paenibacillus*	90
2933 TSA-1	*Bacillus*	37
2987 TSA-1	*Bacillus*	5
2941 SDA-1	*Paenibacillus*	43
2059 SDA-1	*Bacillus*	10
1370 BA-1	*Bacillus*	14
3089 SDA-1	*Bacillus*	19
1480 BA-3	*Bacillus*	28
1480 BA-2	*Bacillus*	21
1943 R2A-1	*Bacillus*	8
2069 TSA-4	*Bacillus*	5

The Rapid Barcoding Sequencing (SQK-RBK004) protocol requires 400 ng high molecular weight DNA at a concentration of 53ng/µl. This threshold concentration of DNA was not met for all isolates even after multiple rounds of DNA extraction. The decision was made to sequence all isolates. The Illumina MiSeq protocol called for an optimal 100-400 ul ng in 30 ul and every isolate met this threshold.

(B) Agilent 4200 TapeStation System

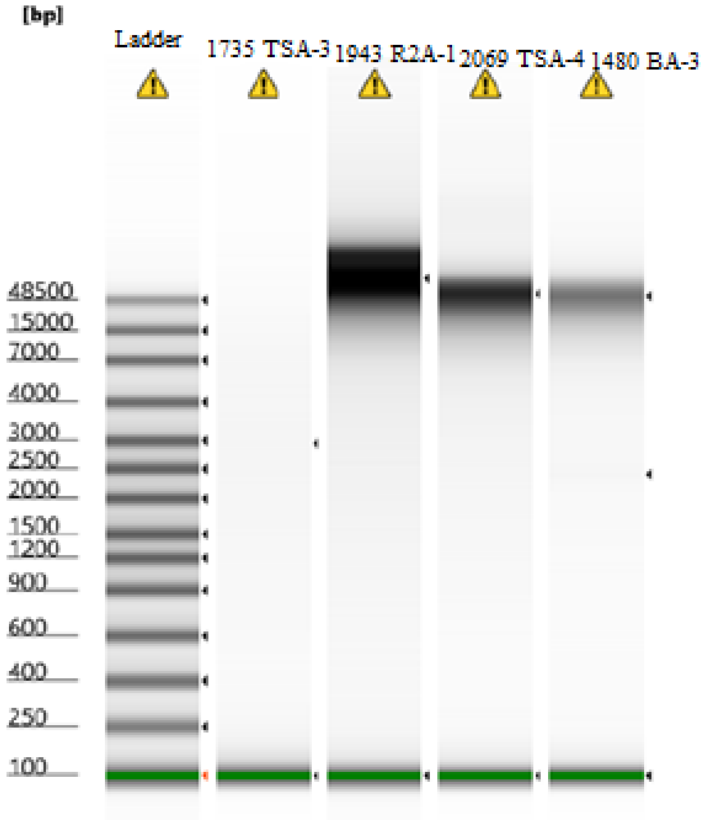

Figure 9. *Tape station gel for first DNA extraction round bands at 48500bp indicating high molecular weight DNA. 1735 TSA-3 was barely visible and had a low DNA integrity number (DIN). Green line at 100 bp represents the lower limit*

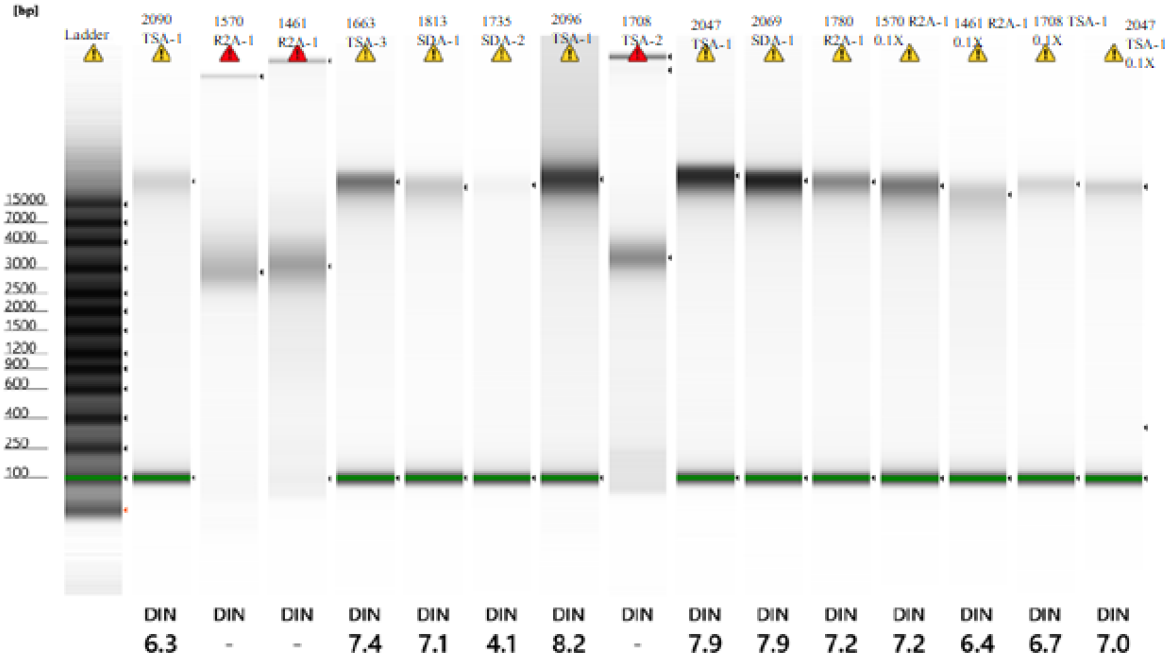

Figure 10. *Tape Station gel for second DNA extraction round. Green line at 100bp represents lower limit*

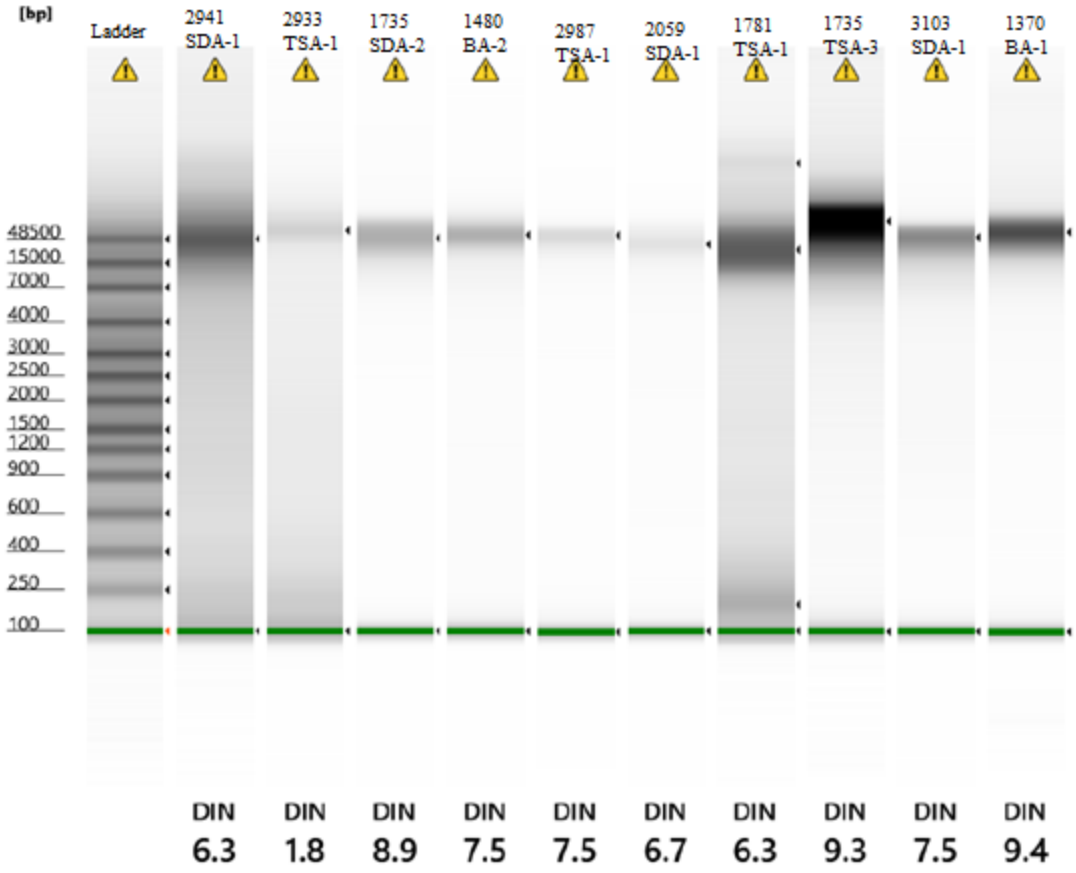

Figure 11. *Tape Station gel for third DNA extraction round. Green line at 100bp indicates lower limit*

MALDI TOF MS

All except two of the isolates were reliably identified by the Bruker Biotyper system (Table 5), on the scale provided by the company (Table 4). The identifications were consistent, at the genera level, for all isolates, but there were two notable discrepancies, for identifications from the same isolate spotted multiple times.

Table 4. *Meaning of score values for Bruker Daltonik MALDI Biotyper Classification Results*

Range	Description	Color
2.300-3.000	highly probable species identification	Green
2.000-2.299	secure genus identification, probable species identification	Green
1.700-1.999	probable genus identification	Yellow
0.000-1.699	not reliable identification	Red

Table 5. *Bruker Daltonik MALDI Biotyper Classification Results*

Analyte Name	Organism (Best Match)	Score Value
1461 R2A-1	*Bacillus licheniformis*	2.168
1708 TSA-2	*Bacillus thuringiensis*	2.177
1663 TSA-1	*Bacillus cereus*	2.303
1570 R2A-1	*Bacillus cereus*	2.251
1813 SDA-1	*Bacillus cereus*	2.328
2090 TSA-1	*Bacillus cereus*	2.435
1480 BA-3	*Bacillus flexus*	2.259
1708 R2A-1	*Bacillus flexus*	2.255
1943 R2A-1	not reliable identification	1.682
2047 TSA-1	not reliable identification	1.623
1370 BA-1	*Bacillus pumilus*	1.935
1480 BA-2	*Bacillus megaterium*	2.386
1735 SDA-2	*Bacillus pumilus*	1.967
1781 TSA-1	*Paenibacillus amylolyticus*	2.243
1735 TSA-3	*Alcaligenes faecalis*	2.336
2069 TSA-4	*Bacillus marisflavi*	2.199
2059 SDA-1	*Bacillus simplex*	2.152
2069 TSA-3	not reliable identification	1.472

Bioinformatics

(A) Comparison of MALDI TOF MS to Whole Genome Identification

64% of species-level and 82% of genus-level identifications by the Biotyper system coincided with those identified by WGS sequencing analysis (Table 7).

Table 6. *Comparison of Bruker Daltonik MALDI Biotyper Classification Results to Whole Genome Identification*

Isolate	MALDI TOF MS ID	Whole Genome ID
1663 TSA-1	*Bacillus cereus*	*Bacillus cereus*
1570 R2A-1	*Bacillus cereus*	*Bacillus subtilis*
1813 SDA-1	*Bacillus cereus*	*Bacillus cereus*
1480 BA-3	*Bacillus flexus*	*Bacillus flexus*
1708 R2A-1	*Bacillus flexus*	*Bacillus flexus*
1943 R2A-1	not reliable identification	*Bacillus altitudinis*
2047 TSA-1	not reliable identification	*Solibacillus silvestris*
1370 BA-1	*Bacillus pumilus*	*Bacillus safensis*
1735 SDA-2	*Bacillus pumilus*	*Bacillus pumilus*
1781 TSA-1	*Paenibacillus amylolyticus*	*Paenibacillus xylanexedens*
1735 TSA-3	*Alcaligenes faecalis*	*Alcaligenes faecalis*

(B) Cluster Analysis of Mass Spectra

Cluster analysis of spectra generated by MALDI-TOF MS performed with MALDIquant showed three clades strongly supported by bootstrap values > 80% (Fig. 13). Isolates identified as *B. flexus* by WGS clustered together. Isolates identified as *B. cereus* and *B. subtilis* also formed a distinct clade.

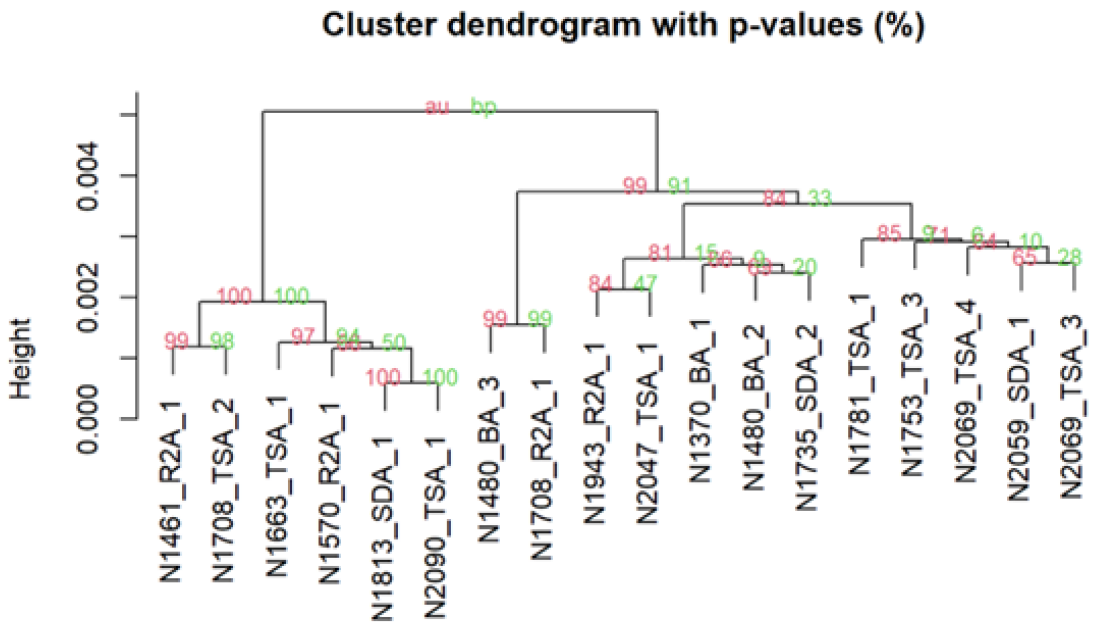

Figure 12. *MALDI TOF MS Cluster Dendrogram. P-values indicate bootstrap values and distance is Euclidean. AU (red values) and BP (green values) indicate unbiased probability values and bootstrap probabilities assigned using pvclust*

(C) Pairwise Similarity of Amino Acids with Mass Spectra

Pairwise similarity scores of mass spectra generated by MALDI-TOF MS and amino acid sequences of single core genes predicted from draft genomes showed good agreement. (Figure. 14). At a 94% (species-level) pairwise amino acid similarity mass spectra were consistently a 0.65 cosine similarity. This level of similarity corresponds to clusters defined as MALDI-TOF taxonomic units (MTUs) previously

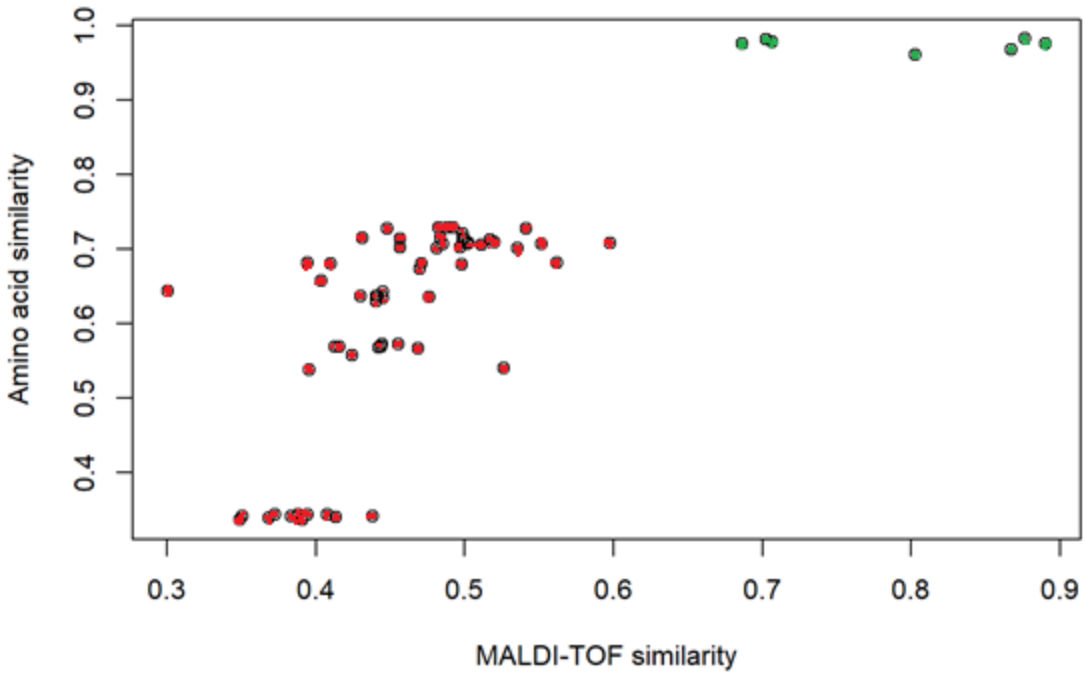

Figure 13. *Pairwise similarity of amino acids from single core genes with MALDI-TOF clusters. Green indicates a species-level pairwise similarity and red indicates interspecies similarity*

(D) Species Diversity of MTUs

Rarefaction analysis, performed with iNEXT on the number of MTUs, calculated assuming a similarity threshold of > 0.65 (Figure 14.), suggested that the diversity of the library was limited. That is a collector's curve, which serves to identify the expected number of distinct species, suggests there are only about 15 - 20 species of readily culturable species in the system (Figure 15).

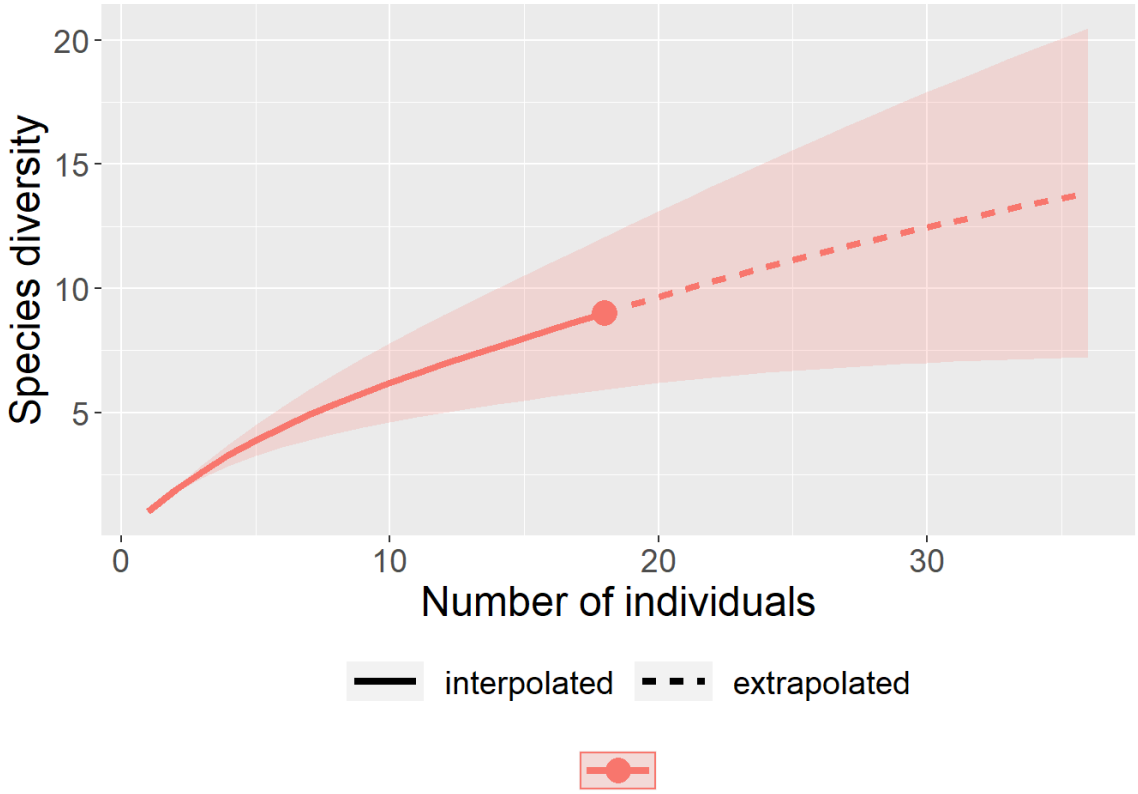

Figure 14. *Species diversity using MTUs calculated by rarefaction analysis*

(E) Summary of Whole Genomes

The Edge Bioinformatics pipeline assembled 18 draft genomes; two failed (Table 8). The percent mapped reads for all assembled genomes was greater than 99%, which indicates highly accurate sequencing. Every isolate had less than 10 contigs except 1663 TSA-1, which was assembled into 48 contigs. High N50 size also indicated good assembly (Table 8). The mean GC contents were consistent with the averages within their specific genus of 43 percent for Bacillus (Akashi and Hirofumi, 2013) and 56 percent for Alcaligenes (Basharat Z and He T, 2018) (Figure 29).

Table 7. Summary table of whole genomes

Isolate	Contigs	Max Contig Size (bp)	N50 (bp)	% total mapped reads	Species Taxa (% reads)	Percent Composition (%)	Redundancy (%)
1370 BA-1	1	3,658,909	1,976,068	99.91%	*Bacillus safensis (99.6%)*	99.16%	0.84
1461 R2A-1	assemblyFail						
1480 BA-3	6	3,935,672	3,935,672	99.95%	*Bacillus flexus (90.1%)*	98.32%	0
1570 R2A-1	6	5,353,115	5,353,115	99.98%	*Bacillus subtilis (94.4%)*	N/A	N/A
1663 TSA-1	48	1,045,468	614,346	99.23%	*Bacillus cereus (95.6%)*	100%	0.84
1708 R2A-1	1	4,184,416	4,184,416	99.93%	*Bacillus flexus (77.2%)*	99.16%	0.84
1735 SDA-2	4	4,035,936	4,035,936	99.93%	*Bacillus pumilus (93.8%)*	100%	0.84
1735 TSA-3	1	4,258,302	4,258,302	99.98%	*Alcaligenes faecalis (92.7%)*	91.60%	3.36
1781 TSA-1	2	7,075,035	7,075,035	99.93%	*Paenibacillus xylanexedens(85.5%)*	99.16%	2.52
1813 SDA-1	2	5,204,374	5,204,374	99.95%	*Bacillus cereus (98.4%)*	100%	0
1943 R2A-1	1	3,749,496	3,749,496	99.94%	*Bacillus altitudinis (99.3%)*	98.32%	0.84
2047 TSA-1	3	3,692,152	3,692,152	99.62%	*Solibacillus silvestris (34.5%)*	99.16%	0.84
2059 SDA-1	assemblyFail						
2069 SDA-1	3	5,505,279	5,505,279	99.93%	*Brevibacterium frigoritolerans(91.1%)*	98.32%	1.68
2096 TSA-1	1	4,180,554	4,180,554	99.93%	*Bacillus licheniformis (99.1%)*	99.16%	1.68
2933 TSA-1	3	5,043,971	5,043,971	99.76%	*Bacillus infantis (79.1%)*	98.32%	4.2
2941 SDA-1	8	6,868,407	6,868,407	99.91%	*Paenibacillus xylanexedens(32.1%)*	99.16%	5.04
2987 TSA-1	1	4,088,622	4,088,622	99.80%	*Bacillus velenzensis (99.7%)*	99.16%	0.84
3103 SDA-1	2	4,752,604	4,752,604	99.97%	*Bacillus pseudomycoides (58.4%)*	100	0
2090 TSA-1	7	5,311,671	5,311,671	99.85%	*Bacillus Cereus (97.8%)*	100.00%	0

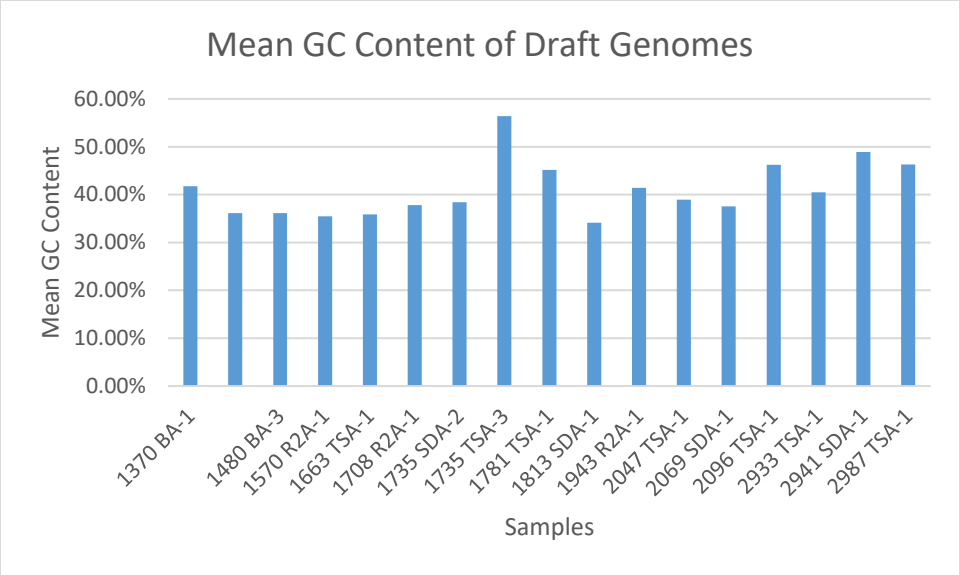

Figure 15. *Mean GC content of the 18 draft genomes from the sample data set. Mean is consistent within genera norms. 1735 TSA-3 is the outlier because it is an unrelated genus*

(F) Phylogeny

a. 16S rRNA Phylogenetic Tree

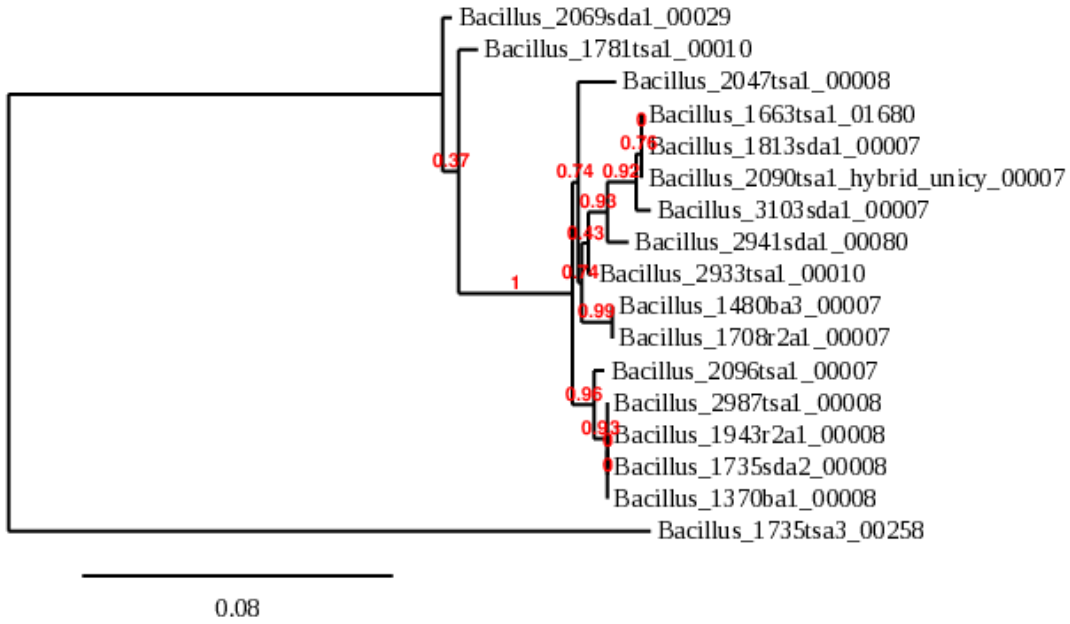

Figure 16. *16s rRNA phylogenetic tree done with Phylogeny.r (Dereeper et al., 2008) Branches having support value less than 50% were collapsed. Boot strap values of 1*

b. Phylogenomic Tree

References assembly accessions pulled from the NCBI assembly database include GCF_019037645.1, GCF_019037245.1, GCF_019037745.1, and GCF_019037445.1 which were originally isolated from a spacecraft assembly cleanroom at the JPL. GCF_019749155.1 and GCF_019749315.1 were isolated from rodents on the International Space Station by the JPL. GCF_019749315.1, GCF_013345915.1 and GCF_014267515.1 were isolated from various areas of the ISS during Microbial tracking investigation by the JPL. GCF_014876505.1 is a radiation-resistant extremophile isolated from the Xinjiang Uygur Autonomous Region of China. GCF_017303255.1 is a polyhydroxyalkanoates-producing bacteria from the coastal area of Shenzhen, China.

GCF_017874615.1 DOE Joint Genome Institute, GCF_001677985.1 California State University Fresno isolated from poultry probiotics.

Phlyogenomic analysis of single copy core genes showed a greater increase in species diversity compared to the 16s rRNA tree. There are more distinct branches and nodes indicating species divergence. The tree was rooted with 1735 TSA-. Genomes showed close relatability to reference genomes of which came from spacecraft assembly cleanrooms and five from the international space station and one radiation-resistant extremophile.

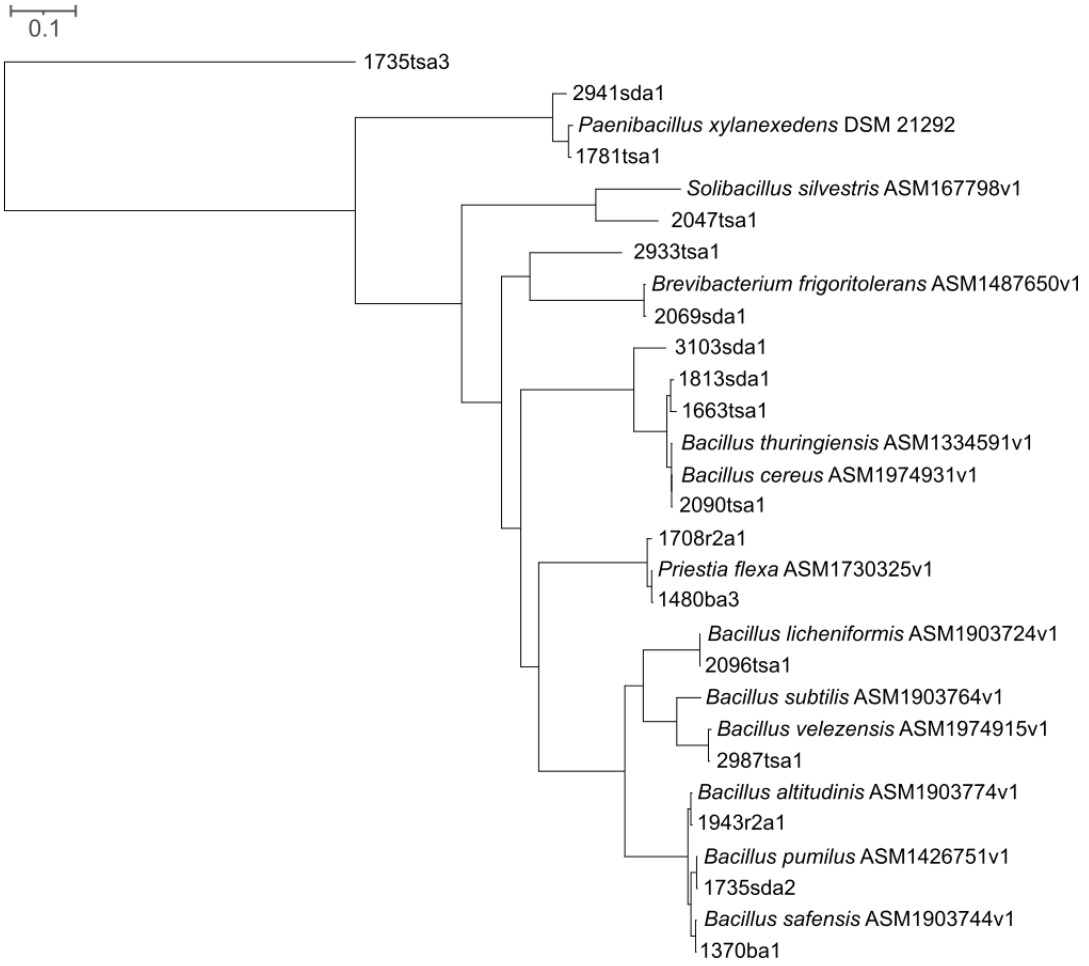

Figure 17. *Phylogenomic tree done with GToTree (Lee, 2019) using single core genes from the phylum Firmicutes using reference assembly accessions from NCBI. Visualized in iTOL interactive tree of life (Letunic & Peer, 2016).*

CHAPTER IV:

DISCUSSION

This main goal of this project was to create reference genomes of *Bacillus* isolates from NASA Johnson Space Center astromaterials clean rooms and compare the resolution of MALDI-TOF to WGS. Draft genomes were assembled from representative *Bacillus* species selected from a library and compared to the accuracy of MALDI TOF MS. Out of a library of 20 isolates, 18 draft genomes were assembled. These genomes were phylogenomically compared using the single copy core genes from the phylum Firmicutes. Cluster analysis results suggest that MALDI-TOF results agreed with WGS. Strains with a single core copy gene amino acid similarity of more than 94% based on pairwise similarity were clustered together.

16S rRNA Sequencing

The 16S rRNA tree shows less species resolution than the cluster dendrogram and phylogenomic tree. Some of the bootstrap values were low (< 0.43) indicating low reliability. There are more unresolved nodes compared to the phylogenomic tree and the MALDI-TOF cluster dendrogram indicating a lack of species diversity. When depending entirely on 16S rRNA gene sequencing for *Bacillus* identification, there is a knowledge gap. 16S sequencing has limited resolving power when compared to WGS and MALDI-TOF (LaMontagne et al., 2021 & Seuylemezian et al., 2018) and is unable to distinguish between *Bacillus* species. We would be missing out on the potential to identify certain novel species and conduct accurate microbial tracking if we relied solely on the 16S gene for routine microbe monitoring (Seuylemezian et al., 2018). The limitations of 16S rRNA sequencing limit the scope of microbial diversity.

Draft Genomes

Genome completeness for all strains was above 98% (Table 8) and a genome redundancy of less than 1%, indicating that repeated and ambiguous sections of the genome were resolved. The quality of these draft genomes may reflect the employment of hybrid assembly, which combines the strengths of NGS and Nanopore sequencing. Nanopore Minion reads have lengthy reads with no theoretical limit (Chen and Meng, 2020), allowing them to span the genome's longest repeating element at roughly 7,000 bp and generate a complete genome with few contigs. However, compared to Illumina reads, they have high error rates. Illumina readings are short (max 300 bp) yet provide good accuracy ("Melbourne Bioinformatics", n.d.). Together reads generated from these platforms consistently produced high quality, near complete genomes.

Visualization of gemomic structure in Bandage revealed numerous short fragment contigs that matched the properties of plasmids, which are more abundant in *Bacillus sp.* The importance of possible contamination, regardless of the amount, should not be underestimated. Small amounts of contamination can make a large impact on identifying genetic variants (Going et al., 2020) and many genomic sequences in public databases are contaminated (Lu, 2018; Merchant et al, 2014). This can introduce biases in variant analysis regardless of what type of mapping and variant calling cutoffs are used. It would be best to analyze and remove possible contaminated reads before studying genetic variants to get the most accurate results (Goig, et al, 2020). Gap closure was not performed to complete genomes, but this was not a necessary step for the scope of this study. However, these gaps could contain information that is functionally important. It would be a necessary step for gene function analysis.

Phylogenomic Tree

The draft genomes were successfully used to build a phylogenomic tree with high gene hit rates using single copy core genes. Presumably these genes share similar evolutionary pressure and therefore improve inference of evolutionary relationships (Creevey et al., 2011). Target genes were predicted based on genes that exist in 90% of reference genomes for the Firmicutes. This tree indicated distinct evolutionary relationships that were not apparent using the 16s rRNA tree.

Strain similarity to genomes of bacteria isolated from similar built environments, such as spacecraft assembly cleanrooms, the international space station and a radiation-resistant extremophile. This suggests that these environments harbor similar microorganisms. These extremophiles could impact astromaterials.

MALDI TOF Comparison to Whole Genome Sequencing

A total of 8 bacterial samples were used to compare WGS with MALDI TOF identification methods. Optimization of cluster analysis of mass spectra generated from a library of bacteria isolated from astromaterials cleanrooms produced coherent MTUs that corresponded to species defined by WGS. The samples do not appear to be clustering in any pattern according to the lab or sample location they were collected from. 1943 R2A-1 was the only isolate that did not cluster similarly but was given a not reliable identification according to the MALDI Biotyper. The amino acid matrix produced as the GToTree (Lee, 2019) phylogenomic tree output file was used to do pairwise similarity between spectras and amino acids. The results showed that a 94% (species-level) pairwise amino acid similarity would produce clusters. MALDI-TOF shows to have a comparable resolution to that of WGS. However, the MALDI Biotyper only identified 64% of samples consistently with their species-level WGS identification and 82% with their genus-level WGS identification. The lack of representation of environmental isolates in this database is evident here.

MALDI-TOF MS Theoretical Implications

MALDI-TOF MS was confirmed to be a quick, accurate, and cost-effective method of microbial identification. The paucity of available environmental microorganism representation in spectral databases can be overcome by building mass spectral profile databases, which are referenced with 16s rRNA and WGS sequences. This serves to mitigate the bias toward clinical isolates and foster better environmental microorganism representation (Deutsch et al., 2017).

It is critical to not only identify microbes in the astromaterials cleanroom, but also to assess species diversity and novel strains. If we can resolve microbes down to the

strain level, we will have better knowledge of their metabolic processes and what kind of harm or modifications they can make to astromaterials.

Endospore forming bacteria are a common source of contamination of built environments and constitute the vast majority of bacteria cultivated from cleanrooms at NASA Johnson Space Center (Figure 1). *Bacillus* strains isolated from clean rooms at the JPL and have extreme resistances to UV radiation and peroxide and can withstand extreme temperatures, high-alkalinity, space vacuum and simulated Mars environments (Tirumalai et al., 2013).

Limitations

It is recommended by Bruker Daltonics that 12 spots are used for reference spectra generation by MALDI TOF MS. In this study only samples were only spotted twice. This narrowed the breadth and number of isolates compared. Not all the isolates from the sample set had MALDI TOF identification conducted.

CHAPTER V:

CONCLUSION AND FUTURE DIRECTIONS

MALDI TOF clustered Bacillus isolates in this sample set at a resolution comparable to that of WGS. This suggests MALDI TOF provides a quick, cost-effective, and accurate method of tracking microbial contamination of astromaterial curation facilities. This proteomics approach could replace 16s rRNA sequencing for routine monitoring. Only a handful of isolates were analyzed by both MALDI TOF and WGS; however, rarefaction suggested this sample was representative.

The availability of draft genomes generated in this work creates the opportunity for further research into the nature of this set of *Bacillus* strains that appear associated with clean rooms. It would improve this analysis if contaminating reads were removed from these draft genomes gaps were closed within scaffolds. Functional genomics could then be used to identify secondary pathways, metabolites and natural products that could harm astromaterials. The availability of a well characterized library of isolates will also facilitate manipulative experiments, such as simulated Mars and zero gravity environments, to investigate how these isolates adapt to extreme environments.

REFERENCES

Ahmad, Faheem, et al. “Potential of MALDI-TOF Mass Spectrometry as a Rapid Detection Technique in Plant Pathology: Identification of Plant-Associated Microorganisms.” *Analytical and Bioanalytical Chemistry*, vol. 404, no. 4, Sept. 2012, pp. 1247–55. *DOI.org (Crossref)*, https://doi.org/10.1007/s00216-012-6091-7.

Akashi, Motohiro, and Hirofumi Yoshikawa. “Relevance of GC Content to the Conservation of DNA Polymerase III/Mismatch Repair System in Gram-Positive Bacteria.” *Frontiers in Microbiology*, vol. 4, 2013. *DOI.org (Crossref)*, https://doi.org/10.3389/fmicb.2013.00266.

Altschul, Stephen F., et al. “Basic Local Alignment Search Tool.” *Journal of Molecular Biology*, vol. 215, no. 3, Oct. 1990, pp. 403–10. *DOI.org (Crossref)*, https://doi.org/10.1016/S0022-2836(05)80360-2.

Anisimova, Maria, and Olivier Gascuel. “Approximate Likelihood-Ratio Test for Branches: A Fast, Accurate, and Powerful Alternative.” *Systematic Biology*, edited by Jack Sullivan, vol. 55, no. 4, Aug. 2006, pp. 539–52. *DOI.org (Crossref)*, https://doi.org/10.1080/10635150600755453.

Babraham Bioinformatics - FastQC A Quality Control Tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 10 Dec. 2021.

Basharat, Zarrin, et al. “Genome Sequencing and Analysis of *Alcaligenes faecalis* Subsp. Phenolicus MB207.” *Scientific Reports*, vol. 8, no. 1, Dec. 2018, p. 3616. *DOI.org (Crossref)*, https://doi.org/10.1038/s41598-018-21919-4.

Böhme, Karola, et al. “SpectraBank: An Open Access Tool for Rapid Microbial Identification by MALDI-TOF MS Fingerprinting: Proteomics and 2DE.” *ELECTROPHORESIS*, vol. 33, no. 14, July 2012, pp. 2138–42. *DOI.org (Crossref)*, https://doi.org/10.1002/elps.201200074.

Burns, Chester R. “The University of Texas Medical Branch at Galveston: Origins and Beginnings.” *JAMA*, vol. 266, no. 10, Sept. 1991, p. 1400. *DOI.org (Crossref)*, https://doi.org/10.1001/jama.1991.03470100092039.

Capella-Gutierrez, S., et al. “TrimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses.” *Bioinformatics*, vol. 25, no. 15, Aug. 2009, pp. 1972–73. *DOI.org (Crossref)*, https://doi.org/10.1093/bioinformatics/btp348.

Castresana, J. “Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis.” *Molecular Biology and Evolution*, vol. 17, no. 4, Apr. 2000, pp. 540–52. *DOI.org (Crossref)*, https://doi.org/10.1093/oxfordjournals.molbev.a026334

Chen, Zhao, et al. “Benchmarking Hybrid Assembly Approaches for Genomic Analyses of Bacterial Pathogens Using Illumina and Oxford Nanopore Sequencing.” *BMC Genomics*, vol. 21, no. 1, Sept. 2020, p. 631. *BioMed Central*, https://doi.org/10.1186/s12864-020-07041-8.

Chevenet, François, et al. “TreeDyn: Towards Dynamic Graphics and Annotations for Analyses of Trees.” *BMC Bioinformatics*, vol. 7, no. 1, Dec. 2006, p. 439. *DOI.org (Crossref)*, https://doi.org/10.1186/1471-2105-7-439.

Contamination Engineering Design Guidelines. https://epact2.gsfc.nasa.gov/tycho/STEREOContamControl.htm. Accessed 10 Dec. 2021.

Creevey, Christopher J., et al. “Universally Distributed Single-Copy Genes Indicate a Constant Rate of Horizontal Transfer.” *PLoS ONE* vol. 6, no. 8, Aug. 2011, p. e22099. *DOI.org (Crossref)*, https://doi.org/10.1371/journal.pone.0022099.

De Maio, Nicola, et al. “Comparison of Long-Read Sequencing Technologies in the Hybrid Assembly of Complex Bacterial Genomes.” *Microbial Genomics*, vol. 5, no. 9, Sept. 2019. *DOI.org (Crossref)*, https://doi.org/10.1099/mgen.0.000294.

Dereeper A.*, Guignon V.*, Blanc G., Audic S., Buffet S., Chevenet F., Dufayard J.F., Guindon S., Lefort V., Lescot M., Claverie J.M., Gascuel O. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 2008 Jul 1;36(Web Server issue):W465-9. Epub 2008 Apr 19. (PubMed) *: joint first authors

Deutsch, Eric W., et al. “The ProteomeXchange Consortium in 2017: Supporting the Cultural Change in Proteomics Public Data Deposition.” *Nucleic Acids Research*, vol. 45, no. D1, Jan. 2017, pp. D1100–06. *DOI.org (Crossref)*, https://doi.org/10.1093/nar/gkw936.

Didelot, Xavier, et al. “Transforming Clinical Microbiology with Bacterial Genome Sequencing.” *Nature Reviews Genetics*, vol. 13, no. 9, Sept. 2012, pp. 601–12. *DOI.org (Crossref)*, https://doi.org/10.1038/nrg3226.

Drost, Hajk-Georg. “Philentropy: Information Theory and Distance Quantification with R.” *Journal of Open Source Software*, vol. 3, no. 26, June 2018, p. 765. *DOI.org (Crossref)*, https://doi.org/10.21105/joss.00765.

Eddy, Sean R. “Accelerated Profile HMM Searches.” *PLoS Computational Biology*, vol. 7, no. 10, Oct. 2011, p. e1002195. *DOI.org (Crossref)*, https://doi.org/10.1371/journal.pcbi.1002195.

Edgar, Robert C. “MUSCLE: A Multiple Sequence Alignment Method with Reduced Time and Space Complexity.” *BMC Bioinformatics*, vol. 5, no. 1, Aug. 2004, p. 113. *BioMed Central*, https://doi.org/10.1186/1471-2105-5-113.

Espariz, Martín, et al. “Taxonomic Identity Resolution of Highly Phylogenetically Related Strains and Selection of Phylogenetic Markers by Using Genome-Scale Methods: The Bacillus Pumilus Group Case.” *PLOS ONE*, vol. 11, no. 9, Sept. 2016, p. e0163098. *DOI.org (Crossref)*, https://doi.org/10.1371/journal.pone.0163098.

Ewels, Philip, et al. “MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report.” *Bioinformatics*, vol. 32, no. 19, Oct. 2016, pp. 3047–48. *DOI.org (Crossref)*, https://doi.org/10.1093/bioinformatics/btw354.

Favero, M. S., et al. “Microbiological Sampling of Surfaces.” *Journal of Applied Bacteriology*, vol. 31, no. 3, Sept. 1968, pp. 336–43. *DOI.org (Crossref)*, https://doi.org/10.1111/j.1365-2672.1968.tb00375.x.

Gibb, S., and K. Strimmer. “MALDIquant: A Versatile R Package for the Analysis of Mass Spectrometry Data.” *Bioinformatics*, vol. 28, no. 17, Sept. 2012, pp. 2270–71. *DOI.org (Crossref)*, https://doi.org/10.1093/bioinformatics/bts447.

Goig, Galo A., et al. “Contaminant DNA in Bacterial Sequencing Experiments Is a Major Source of False Genetic Variability.” *BMC Biology*, vol. 18, no. 1, Dec. 2020, p. 24. *DOI.org (Crossref)*, https://doi.org/10.1186/s12915-020-0748-z.

Gokulan, K., et al. “METABOLIC PATHWAYS | Production of Secondary Metabolites of Bacteria.” *Encyclopedia of Food Microbiology*, Elsevier, 2014, pp. 561–69. *DOI.org (Crossref)*, https://doi.org/10.1016/B978-0-12-384730-0.00203-2.

Guindon, Stéphane, and Olivier Gascuel. “A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood.” *Systematic Biology*, edited by Bruce Rannala, vol. 52, no. 5, Oct. 2003, pp. 696–704. *DOI.org (Crossref)*, https://doi.org/10.1080/10635150390235520.

Hasegawa, Masami, et al. “Dating of the Human-Ape Splitting by a Molecular Clock of Mitochondrial DNA.” *Journal of Molecular Evolution*, vol. 22, no. 2, Oct. 1985, pp. 160–74. *DOI.org (Crossref)*, https://doi.org/10.1007/BF02101694

Hornik, Kurt, et al. “Open-Source Machine Learning: R Meets Weka.” *Computational Statistics*, vol. 24, no. 2, May 2009, pp. 225–32. *DOI.org (Crossref)*, https://doi.org/10.1007/s00180-008-0119-7.

Hsieh, T. C., et al. “INEXT: An R Package for Rarefaction and Extrapolation of Species Diversity (H Ill Numbers).” *Methods in Ecology and Evolution*, edited by Greg McInerny, vol. 7, no. 12, Dec. 2016, pp. 1451–56. *DOI.org (Crossref)*, https://doi.org/10.1111/2041-210X.12613.

Hybrid Genome Assembly - Nanopore and Illumina - Bioinformatics Documentation. https://www.melbournebioinformatics.org.au/tutorials/tutorials/hybrid_assembly/nanopore_assembly/. Accessed 10 Dec. 2021.

Konstantinidis, K. T., and J. M. Tiedje. "Genomic Insights That Advance the Species Definition for Prokaryotes." *Proceedings of the National Academy of Sciences*, vol. 102, no. 7, Feb. 2005, pp. 2567–72. *DOI.org (Crossref)*, https://doi.org/10.1073/pnas.0409727102

LaMontagne, Michael G., et al. "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." *PeerJ*, vol. 9, May 2021, p. e11359. *DOI.org (Crossref)*, https://doi.org/10.7717/peerj.11359.

Langmead, Ben, et al. "Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome." *Genome Biology*, vol. 10, no. 3, 2009, p. R25. *DOI.org (Crossref)*, https://doi.org/10.1186/gb-2009-10-3-r25.40. Tracey Allen K. Freitas, Po-E Li, Matthew B. Scholz and Patrick S. G. Chain (2015) Accurate read-based metagenome characterization using a hierarchical suite of unique signatures, Nucleic Acids Research

Lee, Michael D. "GToTree: A User-Friendly Workflow for Phylogenomics." *Bioinformatics (Oxford, England)*, vol. 35, no. 20, Oct. 2019, pp. 4162–64. *PubMed*, https://doi.org/10.1093/bioinformatics/btz188.

Lee, Michael. "Happy Belly Bioinformatics: An Open-Source Resource Dedicated to Helping Biologists Utilize Bioinformatics." *Journal of Open Source Education*, vol. 2, no. 19, Sept. 2019, p. 53. *DOI.org (Crossref)*, https://doi.org/10.21105/jose.00053.

Letunic, Ivica, and Peer Bork. "Interactive Tree of Life (ITOL) v3: An Online Tool for the Display and Annotation of Phylogenetic and Other Trees." *Nucleic Acids Research*, vol. 44, no. W1, July 2016, pp. W242–45. *DOI.org (Crossref)*, https://doi.org/10.1093/nar/gkw290.

Li, Heng. "Minimap and Miniasm: Fast Mapping and de Novo Assembly for Noisy Long Sequences." *Bioinformatics*, vol. 32, no. 14, July 2016, pp. 2103–10. *DOI.org (Crossref)*, https://doi.org/10.1093/bioinformatics/btw152.

Li, Po-E., et al. "Enabling the Democratization of the Genomics Revolution with a Fully Integrated Web-Based Bioinformatics Platform." *Nucleic Acids Research*, vol. 45, no. 1, Jan. 2017, pp. 67–80. *DOI.org (Crossref)*, https://doi.org/10.1093/nar/gkw1027.

Lian, Bin, et al. “Effect of Microbial Weathering on Carbonate Rocks.” *Earth Science Frontiers*, vol. 15, no. 6, Nov. 2008, pp. 90–99. *DOI.org (Crossref)*, https://doi.org/10.1016/S1872-5791(09)60009-9.

Link, L., et al. “Extreme Spore UV Resistance of *Bacillus pumilus* Isolates Obtained from an Ultraclean Spacecraft Assembly Facility.” *Microbial Ecology*, vol. 47, no. 2, Feb. 2004, pp. 159–63. *DOI.org (Crossref)*, https://doi.org/10.1007/s00248-003-1029-4.

Lu, Jennifer, and Steven L. Salzberg. “Removing Contaminants from Databases of Draft Genomes.” *PLOS Computational Biology*, edited by Fengzhu Sun, vol. 14, no. 6, June 2018, p. e1006277. *DOI.org (Crossref)*, https://doi.org/10.1371/journal.pcbi.1006277.

McCubbin, Francis, et al. “Advanced Curation of Astromaterials for Planetary Science Over the Next Decade.” *Bulletin of the AAS*, vol. 53, no. 4, Mar. 2021. *DOI.org (Crossref)*, https://doi.org/10.3847/25c2cfeb.1c20e2ca.

Merchant, Samier, et al. “Unexpected Cross-Species Contamination in Genome Sequencing Projects.” *PeerJ*, vol. 2, Nov. 2014, p. e675. *DOI.org (Crossref)*, https://doi.org/10.7717/peerj.675.

Moissl, Christine, et al. “Molecular Bacterial Community Analysis of Clean Rooms Where Spacecraft Are Assembled: Bacteria in Spacecraft-Associated Clean Rooms.” *FEMS Microbiology Ecology*, vol. 61, no. 3, Sept. 2007, pp. 509–21. *DOI.org (Crossref)*, https://doi.org/10.1111/j.1574-6941.2007.00360.x.

“Moon Rock Laboratory at NASA Turns 25: Collectspace.” *CollectSPACE.com*, http://www.collectspace.com/news/news-072204a.html.

National Aeronautics and Space Administration. 2010. *Handbook for the microbial examination of space hardware*. NASA-HDBK-6022. National Aeronautics and Space Administration, Washington, DC.

“Home - Assembly - NCBI.” *National Center for Biotechnology Information*, U.S. National Library of Medicine.

Nicholas Stoler, Anton Nekrutenko, Sequencing error profiles of Illumina sequencing instruments, *NAR Genomics and Bioinformatics*, Volume 3, Issue 1, March 2021, lqab019, https://doi.org/10.1093/nargab/lqab019

Nicholson, Wayne L., et al. “Resistance of *Bacillus* Endospores to Extreme Terrestrial and Extraterrestrial Environments.” *Microbiology and Molecular Biology Reviews*, vol. 64, no. 3, Sept. 2000, pp. 548–72. *DOI.org (Crossref)*, https://doi.org/10.1128/MMBR.64.3.548-572.2000.

Pohorille, Andrew, and Joanna Sokolowska. “Evaluating Biosignatures for Life Detection.” *Astrobiology*, vol. 20, no. 10, Oct. 2020, pp. 1236–50. *DOI.org (Crossref)*, https://doi.org/10.1089/ast.2019.2151.

Porechop (RRID:SCR_016967)

Rahi, Praveen, and Parag Vaishampayan. “Editorial: MALDI-TOF MS Application in Microbial Ecology Studies.” *Frontiers in Microbiology*, vol. 10, Jan. 2020, p. 2954. *DOI.org (Crossref)*, https://doi.org/10.3389/fmicb.2019.02954.

Rang, Franka J., et al. “From Squiggle to Basepair: Computational Approaches for Improving Nanopore Sequencing Read Accuracy.” *Genome Biology*, vol. 19, no. 1, Dec. 2018, p. 90. *DOI.org (Crossref)*, https://doi.org/10.1186/s13059-018-1462-9.

Regberg, Aaron, et al. *Microbial Ecology of NASA Curation Clean Rooms*. July 2018, p. PPP.3-18-18. *NASA ADS*, https://ui.adsabs.harvard.edu/abs/2018cosp...42E2821R.

Rummel, J. D. “Planetary Exploration in the Time of Astrobiology: Protecting against Biological Contamination.” *Proceedings of the National Academy of Sciences*, vol. 98, no. 5, Feb. 2001, pp. 2128–31. *DOI.org (Crossref)*, https://doi.org/10.1073/pnas.061021398.

Schuerger, Andrew C., et al. “Survival of Endospores of Bacillus Subtilis on Spacecraft Surfaces under Simulated Martian Environments:” *Icarus*, vol. 165, no. 2, Oct. 2003, pp. 253–76. *DOI.org (Crossref)*, https://doi.org/10.1016/S0019-1035(03)00200-8.

Seemann, T. “Prokka: Rapid Prokaryotic Genome Annotation.” *Bioinformatics*, vol. 30, no. 14, July 2014, pp. 2068–69. *DOI.org (Crossref)*, https://doi.org/10.1093/bioinformatics/btu153.

Sella, Sandra R. B. R., et al. “*Bacillus atrophaeus:* Main Characteristics and Biotechnological Applications – a Review.” *Critical Reviews in Biotechnology*, vol. 35, no. 4, Oct. 2015, pp. 533–45. *DOI.org (Crossref)*, https://doi.org/10.3109/07388551.2014.922915.

Seng, Piseth, et al. “Ongoing Revolution in Bacteriology: Routine Identification of Bacteria by Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry.” *Clinical Infectious Diseases*, vol. 49, no. 4, Aug. 2009, pp. 543–51. *DOI.org (Crossref)*, https://doi.org/10.1086/600885.

Seuylemezian, Arman, et al. “Development of a Custom MALDI-TOF MS Database for Species-Level Identification of Bacterial Isolates Collected From Spacecraft and Associated Surfaces.” *Frontiers in Microbiology*, vol. 9, May 2018, p. 780. *DOI.org (Crossref)*, https://doi.org/10.3389/fmicb.2018.00780.

Singhal, Neelja, et al. “MALDI-TOF Mass Spectrometry: An Emerging Technology for Microbial Identification and Diagnosis.” *Frontiers in Microbiology*, vol. 6, Aug. 2015. *DOI.org (Crossref)*, https://doi.org/10.3389/fmicb.2015.00791.

Suzuki, R., and H. Shimodaira. “Pvclust: An R Package for Assessing the Uncertainty in Hierarchical Clustering.” *Bioinformatics*, vol. 22, no. 12, June 2006, pp. 1540–42. *DOI.org (Crossref)*, https://doi.org/10.1093/bioinformatics/btl117.

Tirumalai, Madhan R., et al. “*Bacillus safensis* FO-36b and *Bacillus pumilus* SAFR-032: A Whole Genome Comparison of Two Spacecraft Assembly Facility Isolates.” *BMC Microbiology*, vol. 18, no. 1, Dec. 2018, p. 57. *DOI.org (Crossref)*, https://doi.org/10.1186/s12866-018-1191-y.

Tirumalai, Madhan R., et al. “Candidate Genes That May Be Responsible for the Unusual Resistances Exhibited by *Bacillus pumilus* SAFR-032 Spores.” *PLoS ONE*, , vol. 8, no. 6, June 2013, p. e66012. *DOI.org (Crossref)*, https://doi.org/10.1371/journal.pone.0066012.

Venkateswaran, Kasthuri, et al. “Evaluation of Various Cleaning Methods to Remove Bacillus Spores from Spacecraft Hardware Materials.” *Astrobiology*, vol. 4, no. 3, Sept. 2004, pp. 377–90. *DOI.org (Crossref)*, https://doi.org/10.1089/ast.2004.4.377.

Venkateswaran, Kasthuri, et al. “Molecular Microbial Diversity of a Spacecraft Assembly Facility.” *Systematic and Applied Microbiology*, vol. 24, no. 2, Jan. 2001, pp. 311–20. *DOI.org (Crossref)*, https://doi.org/10.1078/0723-2020-00018.

Wick, Ryan R., et al. “Bandage: Interactive Visualization of *de Novo* Genome Assemblies: Fig. 1.” *Bioinformatics*, vol. 31, no. 20, Oct. 2015, pp. 3350–52. *DOI.org (Crossref)*, https://doi.org/10.1093/bioinformatics/btv383.

Wick, Ryan R., et al. “Unicycler: Resolving Bacterial Genome Assemblies from Short and Long Sequencing Reads.” *PLOS Computational Biology*, vol. 13, no. 6, June 2017, p. e1005595. *DOI.org (Crossref)*, https://doi.org/10.1371/journal.pcbi.1005595.

Wilkinson, Leland. “Ggplot2: Elegant Graphics for Data Analysis by WICKHAM, H.” *Biometrics*, vol. 67, no. 2, June 2011, pp. 678–79. *DOI.org (Crossref)*, https://doi.org/10.1111/j.1541-0420.2011.01616.x.

Yamada, Shoichi, et al. “Cloning and Nucleotide Sequence Analysis of *GyrB* of *Bacillus Cereus* , *B. Thuringiensis* , *B. Mycoides* , and *B. Anthracis* and Their Application to the Detection of *B. Cereus* in Rice.” *Applied and Environmental Microbiology*, vol. 65, no. 4, Apr. 1999, pp. 1483–90. *DOI.org (Crossref)*, https://doi.org/10.1128/AEM.65.4.1483-1490.1999.

Yates, John R. “Mass Spectrometry and the Age of the Proteome.” *Journal of Mass Spectrometry*, vol. 33, no. 1, Jan. 1998, pp. 1–19. *DOI.org (Crossref)*, https://doi.org/10.1002/(SICI)1096-9888(199801)33:1<1::AID-JMS624>3.0.CO;2-9.

Zhao, Y., et al. “RAPSearch2: A Fast and Memory-Efficient Protein Similarity Search Tool for next-Generation Sequencing Data.” *Bioinformatics*, vol. 28, no. 1, Jan. 2012, pp. 125–26. *DOI.org (Crossref)*, https://doi.org/10.1093/bioinformatics/btr595.

APPENDIX A:

AGILENT 4200 TAPESTATION SYSTEM TABLES

Table 8. *Table describing corresponding samples with wells and DIN and DNA concentration from tape station results*

Sample Info

Well	DIN	Conc. [ng/µl]	Sample Description	Alert	Observations
A1	-	71.2	Ladder	⚠	Caution! Expired ScreenTape device; Ladder
B1	4.1	6.57	1735 tsa3	⚠	Caution! Expired ScreenTape device; Sample concentration outside recommended range
C1	9.3	82.9	1943 r2a1	⚠	Caution! Expired ScreenTape device
D1	8.4	40.6	2069 tsa4	⚠	Caution! Expired ScreenTape device
E1	8.0	22.4	1480 ba3	⚠	Caution! Expired ScreenTape device

Table 9. *Table describing corresponding samples with wells and DIN and DNA concentration from second DNA extraction tape station results*

Well	DIN	Conc. [ng/µl]	Sample Description	Alert	Observations
A1	-	158	Ladder	⚠	Caution! Expired ScreenTape device; Peak out of Sizing Range; Ladder
B1	6.3	10.4	2090 tsa-1	⚠	Caution! Expired ScreenTape device
C1	-		1570 r2a-1	▲	Marker(s) not detected; Caution! Expired ScreenTape device
D1	-		1461 r2a-1	▲	Marker(s) not detected; Caution! Expired ScreenTape device
E1	7.4	18.0	1663 tsa-1	⚠	Caution! Expired ScreenTape device
F1	7.1	8.42	1813 sda-1	⚠	Caution! Expired ScreenTape device; Sample concentration outside recommended range
G1	4.1	4.89	1735 sda2	⚠	Caution! Expired ScreenTape device; Sample concentration outside functional range for DIN and the assay
H1	8.2	58.6	2096 tsa-1	⚠	Caution! Expired ScreenTape device
A2	-		1708 tsa-2	▲	Marker(s) not detected; Caution! Expired ScreenTape device
B2	7.9	32.2	2047 tsa-1	⚠	Caution! Expired ScreenTape device
C2	7.9	32.7	2069 sda-1	⚠	Caution! Expired ScreenTape device
D2	7.2	16.2	1780 r2a-1	⚠	Caution! Expired ScreenTape device
E2	7.2	18.6	1570 r2a-1 0.1x	⚠	Caution! Expired ScreenTape device
F2	6.4	12.2	1461 r2a-1 0.1x	⚠	Caution! Expired ScreenTape device

Table 9. *Continued*

G2	6.7	6.70	1708 tsa-2 0.1x	⚠	Caution! Expired ScreenTape device; Sample concentration outside recommended range
H2	7.0	7.16	2047 tsa-1 0.1x	⚠	Caution! Expired ScreenTape device; Sample concentration outside recommended range