ASSESSMENT OF OXFORD NANOPORE TECHNOLOGIES AS A SEQUENCING PLATFORM FOR INCREASED TAXONOMIC RESOLUTION OF MICROBIAL COMMUNITIES

INTERNSHIP PRACTICUM REPORT

Presented to the Graduate Council of the

University of North Texas

Health Science Center

at Fort Worth

in Partial Fulfillment of the Requirements

for the Degree of

MASTER OF SCIENCE

By Kiana Valenti, B.S.

April 2019

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION AND BACKGROUND

Microbial Forensics

DNA analysis for the purpose of forensic human identification was introduced in the 1980's and has been finely tuned over the past several decades [1]. In recent years, a novel concept has emerged regarding forensic identification that unites principles of trace evidence analysis and microbiology through analysis of the microbiome of a biological or environmental specimen. Microorganisms exist in every possible location including in and on the human body in proportions equal to or greater than our own cells [2]. The microorganisms that inhabit the surface of person's skin or live within the gut make up the microbiota; the microbiota is the product of the environment a person is subjected to [3]. Unlike DNA analysis, human microbial analysis includes not only the genome of humans, but also the billions- and up to trillions of microbial cohabitants that humans' host within and on their bodies [4]. Approximately 30 million bacterial cells per hour are shed from the surface of our skin and are left in the vicinity we occupied [5]. Recent research has determined that humans can leave behind unique bacterial signatures indicative of particular body regions and types of bodily contact [6]. For forensic microbiota profiling to be effective, the microbial signature of a perpetrator would need to be detected at a crime scene [5]. By collecting these residual microbial cells left at a crime scene by

a perpetrator, it may be possible to utilize microbiota profiling to complement traditional DNA profiling for forensic investigations [5].

Humans are about 99.9% identical to one another based on their genomes, but can be 80-90% different based on their respective microbiota [3]. This principle can be used to an advantage for forensic identification purposes, as individuals may have different types of microbes present at varying abundances that can be useful for discriminating between individuals. Historically, two strategies for microbial human identification have been explored. The first strategy involves use of phylogenetic distance among microbes within an individual, relying on the precept that microbes within the host individual will be more closely related than microbes across individuals [4]. The second strategy relies on identifying the microbial taxa present in a sample and their relative abundances [4].

Preceding Research

A previous study conducted by Foley (2018) utilized a mixed microbial standard from the American Type Culture Collection, ATCC® MSA-2002™ (ATCC, Manassas, Virginia), an even mixture of 20 bacterial species in the form of whole cells containing fully sequenced, characterized, and authenticated cultures [7]. The purpose of Foley's study was to evaluate the performance of three DNA extraction protocols used in conjunction with three polymerases by sequencing the V4 hypervariable region of the 16S rRNA gene using Illumina MiSeq (Illumina Inc., San Diego, California) [7]. Results of Foley's study (2018) indicated identification of taxa from the microbial standard was not always possible at the species level, and an uneven distribution of bacterial species was observed. In this project an expanded region that includes the entire 16S rRNA gene will be amplified and sequenced using Oxford Nanopore

Technologies, a novel platform capable sequencing the 16S gene to determine whether improved species resolution is possible.

Use of the ATCC® MSA-2002™ standard is beneficial as it is representative of bacterial species commonly found in the environment, including species of Gram positive and Gram negative bacteria. The standard was not manufactured specifically for forensic microbiology research; however, it contains bacteria that are commonly found in a variety of forensically relevant locations such as the human gut, surface of the skin, and soil **(Figure 1)**.

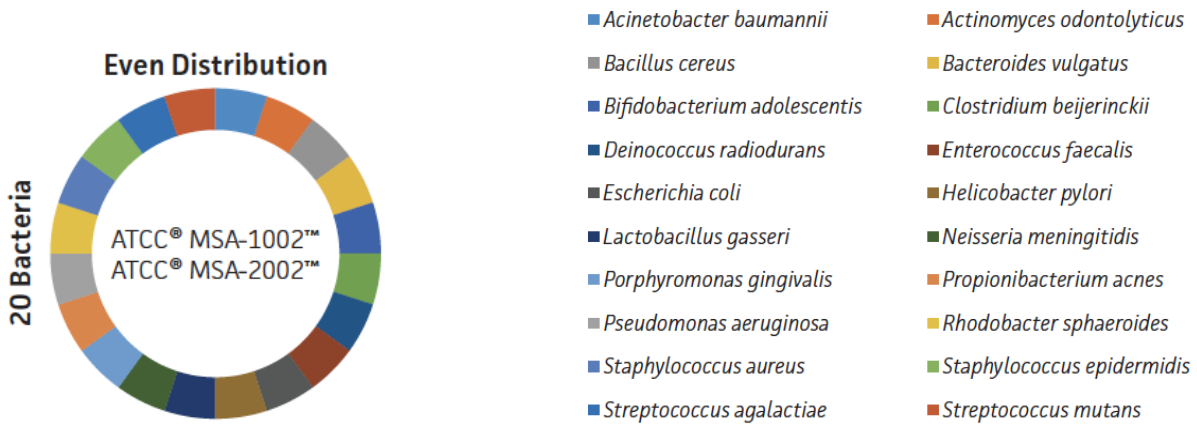

Figure 1. The ATCC® MSA-2002™ standard. The standard consists of an even distribution of 20 bacterial species [obtained from ATCC Data Sheet, 9].

V4 Region vs. 16S Gene

The 16S rRNA gene is comprised of 9 hypervariable regions interspersed across the gene [8]. The preceding project conducted by Foley employed sequencing of the V4 region of the 16S gene, which is a semi-conserved hypervariable region between bacterial species [7]. Each hypervariable region exhibits a different degree of sequence diversity, and it is not possible to distinguish all bacterial species by sequencing any single hypervariable region [9]. The V4 region is comprised of 154 bp while the entire 16S gene is greater than 1500 bp in length. Yang *et al.* evaluated the sensitivity of 7 different 16S rRNA sub-regions as biomarkers and found that

a combination of V4-V6 sub-regions were optimal for bacterial phylogenetic classification [8]. By sequencing only the V4 region in the prior project, valuable information is not being utilized from other hypervariable regions of the 16S gene.

Until recently, most studies of microbiota have relied upon second-generation sequencing platforms that examine 1-2 of the hypervariable regions of the 16S gene. Second-generation sequencing produces high quality short read length (<300bp) sequences; however, it can be difficult to obtain taxonomic assignment of the sequences down to the species level [10]. The introduction of third-generation sequencing technologies with ultra-long read capabilities have made it possible to garner species level and even sub-species level taxonomic identification. Nanopore technology is capable of sequencing ultra-long reads, allowing us to sequence the 16S gene in its entirety [11].

Oxford Nanopore Technologies

A main goal of this project is to generate consensus among application and analysis of a mixed microbial community with a known composition to determine a streamlined, effective means of analyzing microbial casework samples. By using the MinION™ device from Oxford Nanopore Technologies as a sequencing platform in this project, we will examine the entire 16S gene as opposed to only a single variable region within the gene, to determine whether improved taxonomic resolution is possible.

Oxford Nanopore Technologies (ONT) is a revolutionary platform capable of sequencing long reads of DNA [10–14].The science of how the Nanopore functions is what allows for deeper sequencing. A nanopore is a nanoscopic pore composed of α-hemolysin, a heptameric protein that has an inner diameter of less than one nanometer [15] **(Figure 2)**. The pore is only large enough to allow a single strand of DNA through it at a time and the nanopore will sequence

a fragment of DNA that is presented to it, regardless of length [16]. Sequencing adapters that facilitate strand capture within the flow cell are ligated to both ends of genomic DNA fragments before sequencing [12]. The DNA strand is directed to a nanopore by a processive enzyme capable of controlling DNA movement. The strand of DNA is electrophoretically translocated through the pore, facilitated by the processive enzyme which ensures unidirectional flow of the strand [12]. As the bases of the DNA strand pass through the pore, each base registers a characteristic change in ionic current, duration, mean amplitude, and variance unique to each nucleotide that is measured and recorded [12,17]. The signal can then be used to determine the order of bases on that DNA strand. The resulting DNA strand can be base-called in real-time as the MinION™ device sequences the rest of the sample.

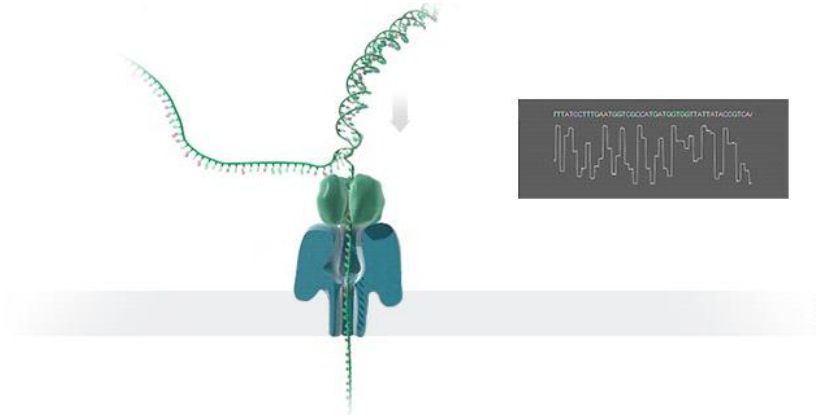

Figure 2. Nanopore sequencing. A single strand of DNA is sequenced at a time. Sequencing adapters are attached to either end of the fragment of DNA. A processive enzyme guides the strand of DNA to an active pore and facilitates the unidirectional translocation of the strand through the pore. As the individual bases of the strand pass through the pore, a characteristic change in current is measured and recorded. The signal can be base-called in real-time or locally [16].

Future Implications

In criminal investigations where human DNA is absent or depleted, profiling the microbiota can serve as a practical identification tool. In the forensic community, microbial composition data has been used for body fluid recognition in place of presumptive testing [18], and in microbial forensic identification research, stable signatures of the microbiome are being identified and targeted [4]. Moving forward in forensic identification, use of Nanopore as a platform could become more prevalent as the means of microbiota taxonomic characterization to make identifications becomes more reliable. Oxford Nanopore Technologies will have the capabilities to decrease sample processing time and allow for sequencing in suboptimal laboratory conditions.

In sum, this study will use nanopore technology as a platform with the aims of determining whether the MinION™ devices from ONT can effectively characterize species present in a mixed microbial standard; determine whether there is consensus of results between two analytical pipelines from Oxford Nanopore Technologies; identify possible sources of error in the protocol; and compare results to the prior study from Foley (2018).

CHAPTER II

MATERIALS AND METHODS

DNA Extraction and Amplification

Samples from two extraction protocols from Foley's study were utilized in this study; FastDNA™ Spin Kit for Soil (MP Biomedicals, LLC, Santa Ana, California), and PowerSoil DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, California). Three samples and one reagent blank from each extraction protocol were sequenced for a total of 8 libraries. Each sample was assigned a barcode **(Table 1).**

Sample	Barcode ID
FastDNA 1	1
FastDNA 2	2
FastDNA 3	3
FastDNA Reagent Blank	4
PowerSoil 1	5
PowerSoil 2	6
PowerSoil 3	7
PowerSoil Reagent Blank	8

Table 1. Barcode assignment to sample.

The SQK-RAB-204 16S Barcoding Kit from Oxford Nanopore Technologies was used for this study. The DNA quantity and quality of the samples were assessed using a Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, California) and a NanoDrop™ 2000 Spectrophotometer (Thermo Fisher Scientific Corporation, Carlsbad, California). Next the samples were prepared

for amplification by adjusting input DNA amount to optimal parameters set forth by the SQK-RAB204 protocol which recommends an input of 10 ng of DNA per sample [19]. All samples from the PowerSoil extraction were concentrated using Microcon® Cetrifugal Filter Devices (Merck Millipore Ltd., Cork, Ireland). The SQK-RAB204 kit contains 12 barcoded primers. These primers enrich the 16S gene through PCR amplification; specifically 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1492R (5'-TACCTTGTTACGACTT-3') [19]. The barcodes contain 5' tags that facilitate the ligase-free attachment of rapid 1D sequencing adapters to the region of interest **(Figure 3)**, [19]. The samples were prepared for amplification by adding 14 µL nuclease-free water, 10 µL (10 ng) input DNA from each sample, 1 µL of the corresponding 16S barcode, and 25 µL LongAmp® Taq 2X Master Mix (New England Biolabs, Inc., Ipswich, Massachusetts) for a total of 50µL per PCR reaction tube. The samples were amplified on an Eppendorf® Mastercycler® pro S thermal cycler (Eppendorf, Hamburg, Germany) under the conditions set forth by the protocol.

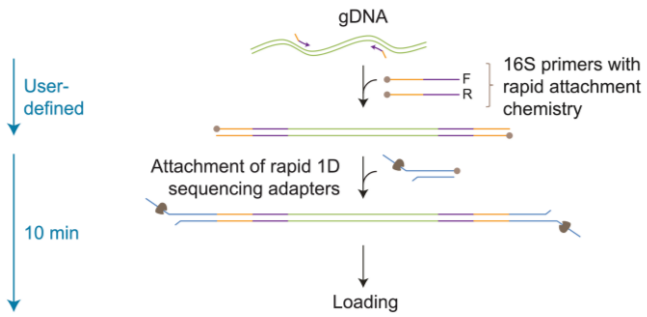

Figure 3. Workflow for the SQK RAB-204 kit. The kit contains 12 barcoded primers that target the 16S gene during PCR amplification using specific 16S primers (27F and 1492R). Within the barcodes are 5' tags that facilitate the ligase-free attachment of rapid 1D sequencing adapters to the region of interest [19].

Library Preparation and Sequencing

Following amplification, each sample was concentrated using the Agencourt AMPure XP purification system (Beckman Coulter, Brea, California) and re-quantified using a Qubit® 2.0 Fluorometer. Following quantification, barcoded libraries were normalized and pooled based on DNA quantitation results to maintain an even distribution of input DNA per sample. An additional Agencourt AMPure XP bead clean-up was performed to concentrate the library, resulting in 10µL of library in Tris-HCl NaCl solution. QC runs for the MinION™ device (FLO-MIN106) were performed to determine the number of active pores within the device **(Figure 4).** The flow cell was primed before the DNA library was deposited onto the MinION™ device's sample port and closed securely. The lid of the MinION™ device was closed and the device was plugged into the computer's USB port. Sequencing was set to run for 48 hours at -180mV. Base-calling using MinKNOW (Oxford Nanopore Technologies) software was performed in real-time during the initial run and after sequencing for the replicate experimental run.

Figure 4. The MinION™ device plugs directly into a PC or laptop. Weighing less than 90 g, the device is highly portable and can be used in non-laboratory environments [12].

Gene Sequence Analysis

Once the run completed, the barcoded sequencing data was exported as fastq files. The fastq files were deconvoluted using EPI2ME, which demultiplexes the barcoded sequences and sorts the reads by barcode [19]. What's in my Pot? (WIMP) v2.3.7 and 16S Taxonomic

Classification v2.2.13 are analysis pipelines on the Metrichor platform capable of classifying and identifying microbial species in real time were used in this study.

Per the ONT website, WIMP identifies microbial species by comparing taxonomic sequences to bacterial reference databases such as NCBI and SILVA [10]. WIMP incorporates the software package Centrifuge, which can accurately identify reads, even when dealing with multiple, highly similar reference genomes, making it ideal for differentiating between strains of bacterial species [20]. Centrifuge is a novel classification engine capable of identifying and aligning unique segments from these reference genomes. It works by building an FM-index (based on the Burrows-Wheeler transform) that has been optimized for differentiation of taxa in a mixed microbial community [21]. WIMP takes results from Centrifuge to assign taxonomic placement on the phylogenetic tree [20]. Taxa can be sorted by classification, barcode number, and minimum relative abundance cutoff.

16S Classification software from ONT was also used to classify the data. 16S is able to make genus level classifications of mixed samples based on BLAST results [22]. 16S is also capable of making species level identification using event data, reference alignment, or consensus based on multiple reads [22]. Like WIMP, 16S displays taxa in a phylogenetic tree and can be sorted based on minimum relative abundance cutoff, as well as taxonomic level. The main difference between WIMP and 16S software is that 16S software uses a compiled reference database, NCBI Bacterial 16S, which contains more organisms than the reference database used for the WIMP workflow [22]. The 16S software also displays a percent average accuracy of read classification.

Percent relative abundance at the species and genus level was calculated by excluding any classified reads for unused barcodes, then by obtaining the number of cumulative reads per

species/genus and dividing by the total number of classified reads for each sequencing run. The average percentage of expected species-level or genus-level identification was performed for WIMP and 16S software. Percent relative abundance at the species level was compared to the expected abundances of the ATCC® MSA-2002™ mock community.

Species and genera identified as present via each analytical pipeline but not included in the mock microbial standard were also identified. Alignment of the 16S rRNA gene of *E. coli* against that of *E. itctaluri* was performed via BLAST. Copy number variation was corrected for by dividing classified reads per species by the corresponding number of copies of the 16S rRNA gene. The expected percentage of classified reads was calculated for each species to normalize the expected distribution of sequenced reads.

CHAPTER III

RESULTS AND DISCUSSION

Overview of What's in my Pot? Taxonomic Classifications

Fastq data was deconvoluted using EPI2ME software. The data was processed via WIMP software v2.3.7 (Instance ID: 187149) with a total of 2,328,830 reads analyzed and 2,302,237 of those reads classified to the lowest taxonomic level, upon exclusion of any reads classified other than barcodes 1-8. In the second run (Instance ID: 192704), WIMP software analyzed a total of 6,907,614 reads and classified 6,864,008 of these at the lowest taxonomic level, upon exclusion of any reads classified other than barcodes 1-8. The reads classified to the reagent blanks accounted for 0.001% or less of the total reads. **Tables 2 and 3** detail the number of reads and the corresponding relative abundance of species and genera respectively from the sequencing runs. The relative abundances listed in the charts have not been normalized for copy number variation and represent raw relative abundances obtained from classified reads. **Figures 5 and 6** display the corresponding phylogenetic trees at the species and genus level respectively from the WIMP pipeline. A minimum relative abundance cut off of 0.5% at the species level and 1.0% at the genus level was applied to exclude taxa present at low abundances and not well characterized by WIMP software and include taxa that were well characterized.

	Run 1		Run 2	
Species	Number of reads	%	Number of reads	%
Acinetobacter baumannii	16,945	0.74%	42,512	0.59%
Bacillus cereus	298,928	12.98%	800,310	11.17%
Bifidobacterium adolescentis	36	0.00%	106	0.00%
Deinococcus radiodurans	31,750	1.38%	57,808	0.81%
Escherichia coli	152,391	6.62%	433,228	6.04%
Lactobacillus gasseri	25,379	1.10%	93,115	1.30%
Porphyromonas gingivalis	1486	0.06%	2,777	0.04%
Pseudomonas aeruginosa	7,807	0.34%	13,521	0.19%
Staphylococcus aureus	72,919	3.17%	249,745	3.48%
Streptococcus agalactiae	108,329	4.71%	515,092	7.19%
Actinomyces odontolyticus	0	0.00%	0	0.00%
Bacteroides vulgatus	2315	0.10%	5,061	0.07%
Clostridium beijerinckii	21,220	0.92%	64,295	0.90%
Enterococcus faecalis	196,890	8.55%	852,748	11.90%
Helicobacter pylori	18,448	0.80%	51,855	0.72%
Neisseria meningitidis	16,594	0.72%	43,085	0.60%
Propionibacterium/Cutibacterium acnes	1,092	0.05%	2,903	0.04%
Rhodobacter sphaeroides	22,902	0.99%	53,741	0.75%
Staphylococcus epidermidis	30,816	1.34%	120,774	1.69%
Streptococcus mutans	198,204	8.61%	631,292	8.81%
Percent expected identifications:		53.18%		56.28%

Table 2. Overview of species level identification using WIMP. An average of 53.23% of reads were assigned to a species expected from the microbial standard. A table providing number of reads and their corresponding percentages grouped by run and extraction method is provided in the appendix.

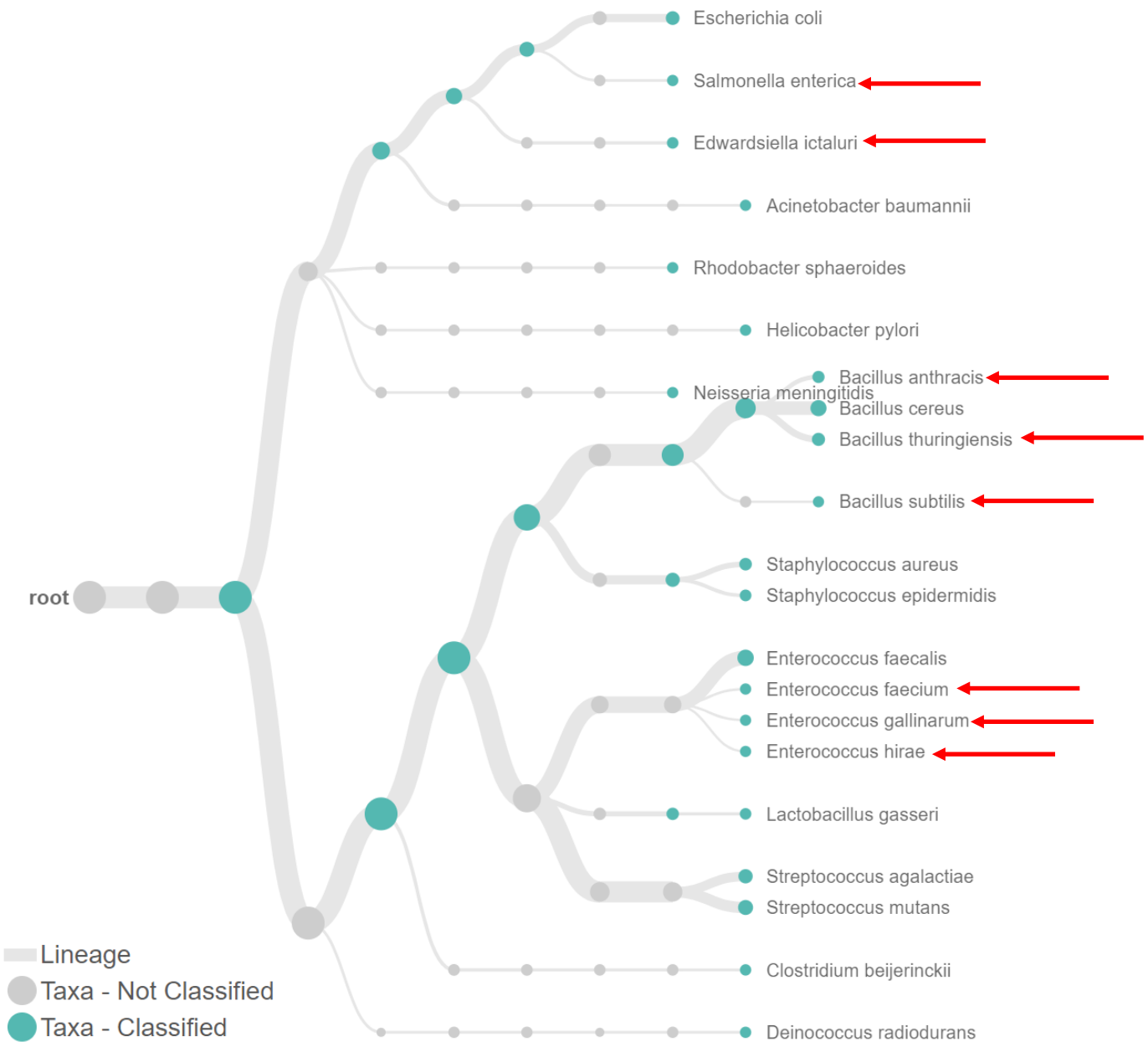

Figure 5. WIMP Classification species level phylogenetic tree with a 0.5% minimum abundance cutoff from the second experimental run. Species identified by WIMP that were not included in the mock microbial standard are indicated with a red arrow.

	Run 1		Run 2	
Genus	Number of reads	%	Number of reads	%
Acinetobacter	22,524	0.98%	55,538	0.81%
Bacillus	619,471	26.91%	1,723,362	25.11%
Bifidobacterium	481	0.02%	811	0.01%
Deinococcus	32,587	1.42%	59,481	0.87%
Escherichia	156,533	6.80%	442,218	6.44%
Lactobacillus	71,336	3.10%	220,503	3.21%
Porphyromonas	1493	0.06%	2,784	0.04%
Pseudomonas	25,882	1.12%	48,166	0.70%
Staphylococcus	140,215	6.09%	498,905	7.27%
Streptococcus	370,103	16.08%	1,311,286	19.10%
Actinomyces	432	0.02%	1,382	0.02%
Bacteroides	2407	0.10%	5,152	0.08%
Clostridium	30,812	1.34%	85,437	1.24%
Enterococcus	245,966	10.68%	1,000,149	14.57%
Helicobacter	19,998	0.87%	55,491	0.81%
Neisseria	19,641	0.85%	50,266	0.73%
Propionibacterium/Cutibacterium	1,102	0.05%	2,909	0.04%
Rhodobacter	27,239	1.18%	57,732	0.84%
Percent expected identifications:		77.67%		81.90%

Table 3. Overview of genus level identification using WIMP software. An average of 79.79% of reads were assigned to a genus expected from the microbial standard.

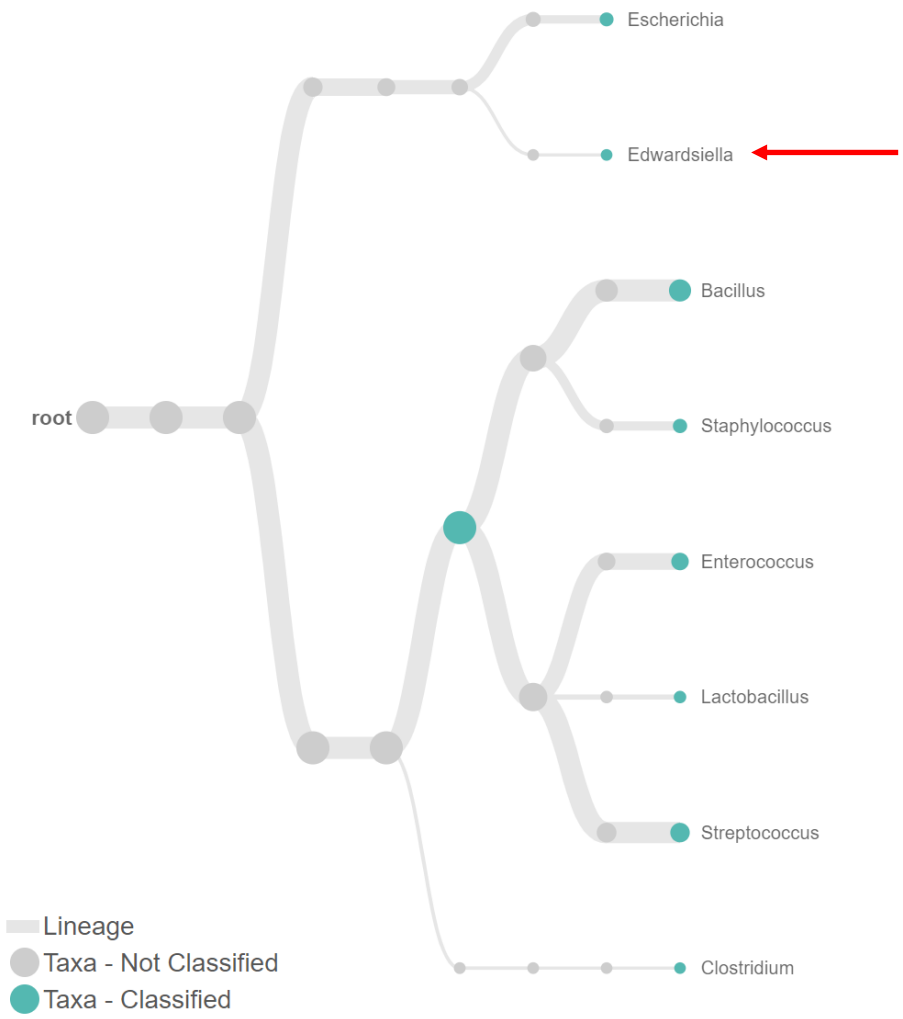

Figure 6. WIMP genus level phylogenetic tree with a 1% minimum abundance cutoff from the second experimental run. The thickness of the branches displayed is proportional to the relative read counts for each genus. *Edwardsiella* is from the same order as *Escherichia*.

Use of WIMP software in conjunction with the MinION™ device resulted in detection of all but one species expected to be present in the ATCC® MSA-2002™ mock microbial community for both sequencing runs. The uncharacterized species, *Actinomyces odontolyticus* was not detected in either sequencing run using WIMP software. WIMP was able to make the classification only at the genus level. Other species were identified that were not expected, though many of these unexpected species belonged to the same genus as expected species in the

mock community. The unexpected species identified at high relative abundance include *Bacillus thuringiensis* and *Bacillus anthrasis.* 5.04% of classified reads were attributed to *B. thuringiensis* in first run and 4.84% of reads in second run. 2.41% of classified reads were attributed to *B. anthrasis* in the first run and 2.71% of reads in second the run. Both species belong to the *Bacillus* genus which are Gram positive bacteria with a high number of copies of the 16S gene in their chromosomes. WIMP was still able to make the classification to a genus from the microbial standard. An unexpected genus was detected using WIMP software at high relative abundance; *Edwardsiella* which comprised 2.40% of classified reads in the first run and 1.33% of reads in the second run.

Edwardsiella is derived from the same order as *Escherichia coli*, Enterobacterales. *Edwardsiella ictaluri* is a Gram negative bacterium that causes acute septicemia or chronic encephalitis in catfish as well as other fish and reptiles. To determine whether misidentification of the species was due to similarities in sequence structure, a preliminary BLAST alignment of the 16S rRNA gene for both species resulted in nucleotide matches at 1401 of 1469 total sites for an overall 95% identity. While the 16S gene is highly conserved across species, this is still a relatively high percent similarity. Identities greater than 90% usually result in genus-level identification, and identities greater than 97% result in species level identification [23]. Further investigation into likely misidentification of *Edwardsiella* species led to a study from Reichley *et al.*, which states that the role of 16s rRNA sequence for differentiation of *Edwardsiella* species has recently been called into question [24]. A limitation cited by Reichley *et al.* that may lead to inaccurate microbial identifications is the reliance on partial 16S rRNA sequences from NCBI; only partial sequences of all three *Edwardsiella* species are available on NCBI. Issues with genus and/or species level resolution within the Enterobacteriacea family using the 16S rRNA gene

have been documented [23]. The data was run through the 16S Taxonomic Classification pipeline and a search for *Edwardsiella* showed very few reads classified by 16S software; 32 of 2,293,970 (0.001%) reads from the first run and 56 of 6,927,881 (0.0008%) of the reads from the second run. Detection of *Edwardsiella* in one pipeline but not the other when the composition of the standard is known provides further support that the databases WIMP uses to align may be the cause for misidentification, and not PCR induced errors. Though it may be more uncommon to encounter *Edwardsiella ictaluri*, a fish-disease inducing bacterium, opposed to the human gut bacterium *Escherichia coli* in a microbial casework sample, it is still necessary to make accurate species identifications when possible, and identifying sources of potential error. A suggestion from the study by Reichley *et al.* suggests utilization of the *gyrB* gene in place of the 16S rRNA gene specifically for identifying members of the Enterobacteriaceae family [24].

A high relative abundance was attributed to species and genera that were not included in the mock microbial standard. 79.79% of the reads were assigned to a genus provided in the mock standard, and 53.23% of the reads were assigned to species from the standard using WIMP software.

Overview of 16S Taxonomic Classifications

Fastq data were deconvoluted using EPI2ME software. The data were processed via 16S Classification software v2.2.13 (Instance ID: 188812) with the first run resulting in a total of 2,293,970 reads classified to the lowest taxonomic level and an average accuracy of 88%, excluding any reads classified other than barcodes 1-8. In the second run (Instance ID: 195070), Fastq WIMP software classified a total of 6,927,881 reads at the lowest taxonomic level and an average accuracy of 90%, excluding any reads classified other than barcodes 1-8. Reads classified to the reagent blanks accounted for 0.001% or less of the total reads in both sequencing

runs. **Tables 4 and 5** detail the number of reads corresponding relative abundance of species and genera respectively from the sequencing runs. The relative abundances listed in the charts have not been normalized for copy number variation and represent raw relative abundances obtained from classified reads. **Figures 7 and 8** display the corresponding phylogenetic trees at the species and genus level respectively from the 16S pipeline. A minimum relative abundance of 1.0% was applied at the species and genus level to exclude taxa present at low abundances and not well characterized by WIMP software and include taxa that were well characterized.

	Run 1		Run 2	
Species	**Number of reads**	**%**	**Number of reads**	**%**
Acinetobacter baumannii	54,305	2.37%	120,809	1.74%
Bacillus cereus	251,631	10.97%	826,027	11.92%
Bifidobacterium adolescentis	0	0.00%	0	0.00%
Deinococcus radiodurans	31,173	1.36%	57,559	0.83%
Escherichia coli	2,558	0.11%	3,189	0.05%
Lactobacillus gasseri	27,946	1.22%	94,754	1.37%
Porphyromonas gingivalis	1,463	0.06%	2,786	0.04%
Pseudomonas aeruginosa	35,749	1.56%	73,770	1.06%
Staphylococcus aureus	105,004	4.58%	387,611	5.59%
Streptococcus agalactiae	270,865	11.81%	960,499	13.86%
Actinomyces odontolyticus	223	0.01%	625	0.01%
Bacteroides vulgatus	2,352	0.10%	5,151	0.07%
Clostridium beijerinckii	7,711	0.34%	20,400	0.29%
Enterococcus faecalis	247,939	10.81%	918,430	13.26%
Helicobacter pylori	19,566	0.85%	53,728	0.78%
Neisseria meningitidis	7,652	0.33%	16,835	0.24%
Propionibacterium/Cutibacterium acnes	819	0.04%	2,105	0.03%
Rhodobacter sphaeroides	17,677	0.77%	40,716	0.59%
Staphylococcus epidermidis	29,688	1.29%	94,735	1.37%
Streptococcus mutans	209,342	9.13%	660,324	9.53%
Percent expected identifications:		57.70%		62.65%

Table 4. Overview of species level identification using 16S software. An average of 60.18% of reads were assigned to a species expected from the microbial standard. A table providing number of reads and their corresponding percentages grouped by run and extraction method is provided in the appendix.

Figure 7. 16S Classification species-level phylogenetic tree with a 1% minimum abundance cutoff from the second experimental run. The thickness of the branches displayed is proportional to the relative read counts for each genus.

	Run 1		Run 2	
Genus	**Number of reads**	**%**	**Number of reads**	**%**
Acinetobacter	56,297	2.45%	124,425	1.80%
Bacillus	306,387	13.36%	952,481	13.75%
Bifidobacterium	0	0.00%	0	0.00%
Deinococcus	31,556	1.38%	58,336	0.84%
Escherichia	13,132	0.57%		0.00%
Lactobacillus	37,104	1.62%	121,928	1.76%
Porphyromonas	1,465	0.06%	2,791	0.04%
Pseudomonas	38,830	1.69%	79,447	1.15%
Staphylococcus	179,216	7.81%	608,327	8.78%
Streptococcus	569,003	24.80%	1,870,002	26.99%
Actinomyces	226	0.01%	626	0.01%
Bacteroides	2,360	0.10%	5,178	0.07%
Clostridium	28,074	1.22%	77,286	1.12%
Enterococcus	272,504	11.88%	1,002,293	14.47%
Helicobacter	19,969	0.87%	54,665	0.79%
Neisseria	24,957	1.09%	61,050	0.88%
Propionibacterium/Cutibacterium	820	0.04%	2,106	0.03%
Rhodobacter	26,231	1.14%	55,726	0.80%
Percent expected identifications:		70.10%		73.28%

Table 5. Overview of genus level identification using 16S software. An average of 71.69% of reads were assigned to a genus expected from the microbial standard.

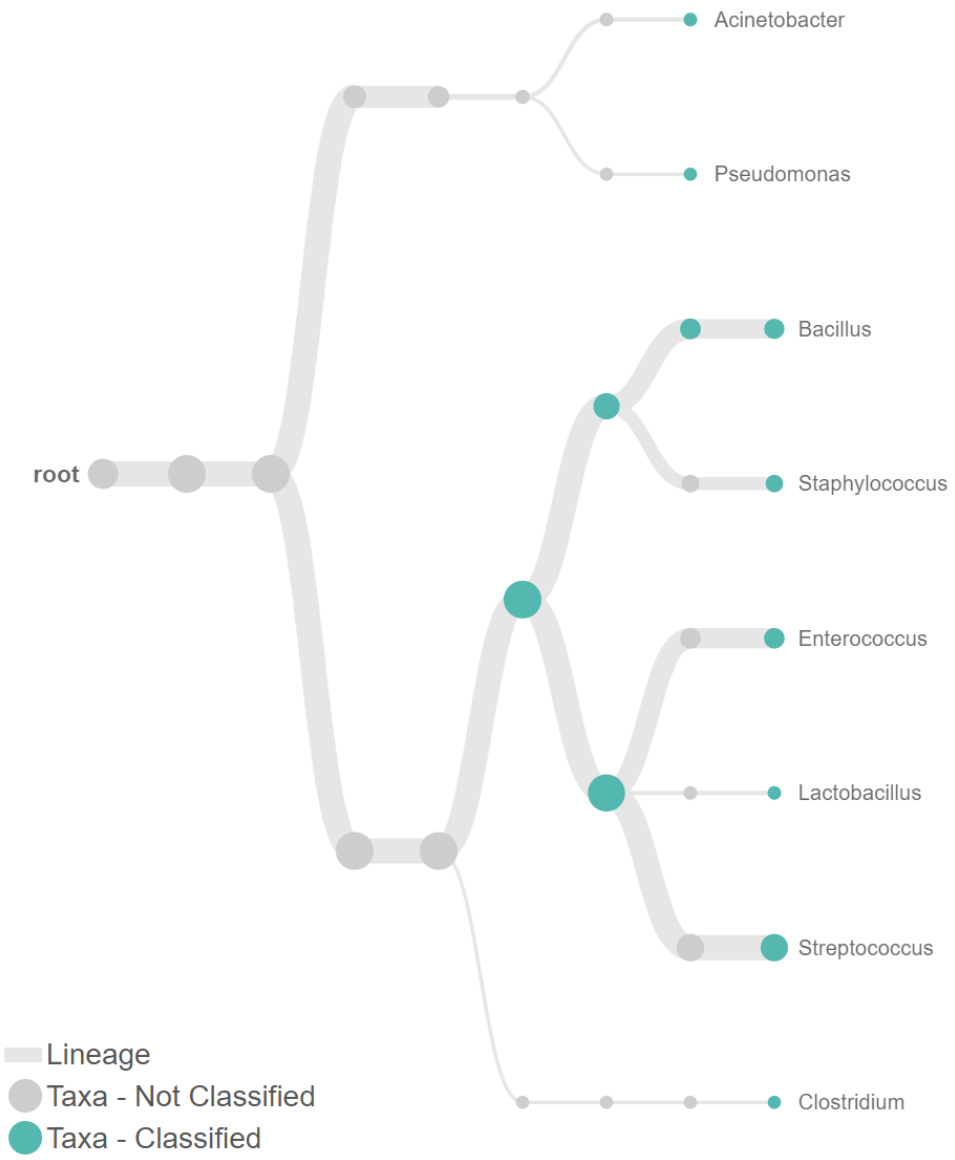

Figure 8. 16S Classification genus level phylogenetic tree with a 1% minimum abundance cutoff from the second experimental run. The thickness of the branches displayed is proportional to the relative read counts for each genus.

Use of 16S software in conjunction with the MinION™ device resulted in detection of all but one species expected to be present in the ATCC® MSA-2002™ mock microbial community for both sequencing runs. The undetected species, *Bifidobacterium adolescentis* was not detected in either sequencing run using 16S software. This is a different undetected species than the

species that was not identified by WIMP software, *Actinomyces odontolyticus*. 16S was only able to assign one read from to the correct genus in the second run. There was higher overall resolution at the genus level. Unexpected species that were identified belonged to the same genus as other species that were expected in the mock community. The species present at high relative abundance included *Streptoccoccus dysgalactiae* accounting for 2.40% of the classified reads in the first run and 2.67% of reads in the second run and *Staphylococcus capitis* which comprised 0.61% of the reads in the first run and 0.65% of the reads in the second run.

Comparison of WIMP and 16S Software Pipelines

16S software assigned a higher percentage of reads (60.18%) to an expected species than WIMP software (53.23%); however, WIMP software assigned a higher percentage of reads (79.79%) to an expected genus than 16S software (71.69%). To determine whether there was consensus between WIMP and 16S pipelines, the average percent species/genus read assignment from both experimental runs was calculated for each pipeline. In **Figures 6 and 7,** side by side comparison of average percentage read assignment for each genus and species is displayed. In general, there was consensus between the relative abundance of classified reads identified at both the genus and species level. There was a difference in relative abundance of *Escherichia*, with WIMP assigning more reads to the genus than 16S software. WIMP also classified a higher percentage of reads to *Bacillus* species than 16S, and 16S classified a higher percentage of reads to *Streptococcus* species than WIMP as detailed in **Figure 6.** At the species level, there was a more even assignment of reads, with the most striking difference between the two platforms being identification of *Escherichia coli* in **Figure 7** where 16S identified very low levels of the species compared to WIMP. It is also apparent from **Figure 7** that identification at the species level differed between the two pipelines for *Streptococcus agalactiae*.

One of the major aims of this study was to determine whether there was agreement between the What's In My Pot? and 16S Taxonomic Classification pipelines. Though both pipelines were run through EPI2ME on the Metrichor platform, WIMP aligns reads from a mixed sample to the NCBI database of bacteria, viruses, and fungi using an algorithm that incorporates Centrifuge, while the 16S pipeline BLASTs the base-called sequences against the specifically curated NCBI 16S bacterial database that contains 16S rRNA sequences of 18,927 species. This database is nested within the NCBI database but offers designated sequences for alignment in the region of interest. As mentioned previously, the method by which WIMP aligns sequenced reads is reliant on the databases it uses. This can be a pitfall in the system if sequences within the database are inaccurate, incomplete, or misclassified [24].

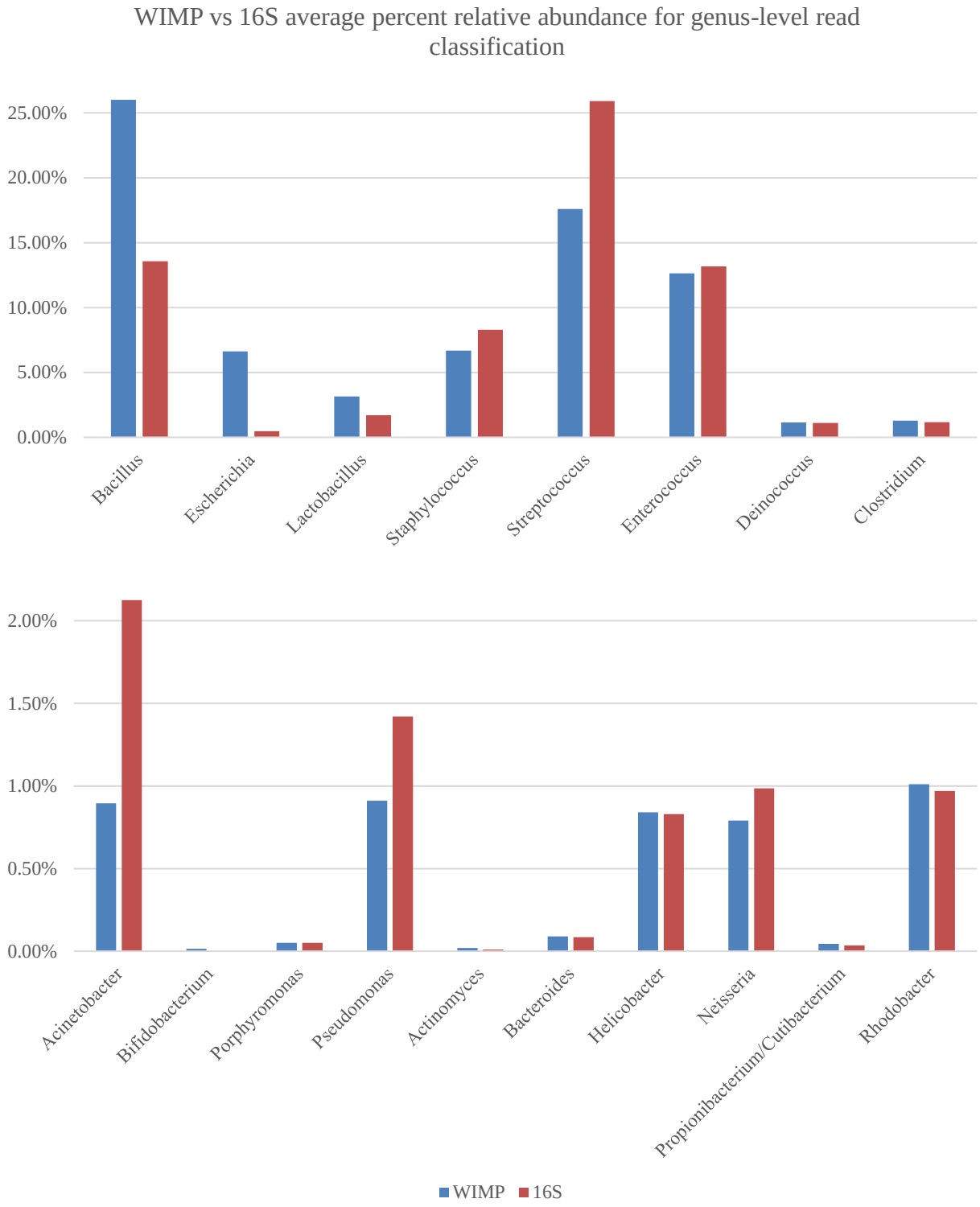

Figure 9. Pipeline comparison of average percent of reads assigned to genera.

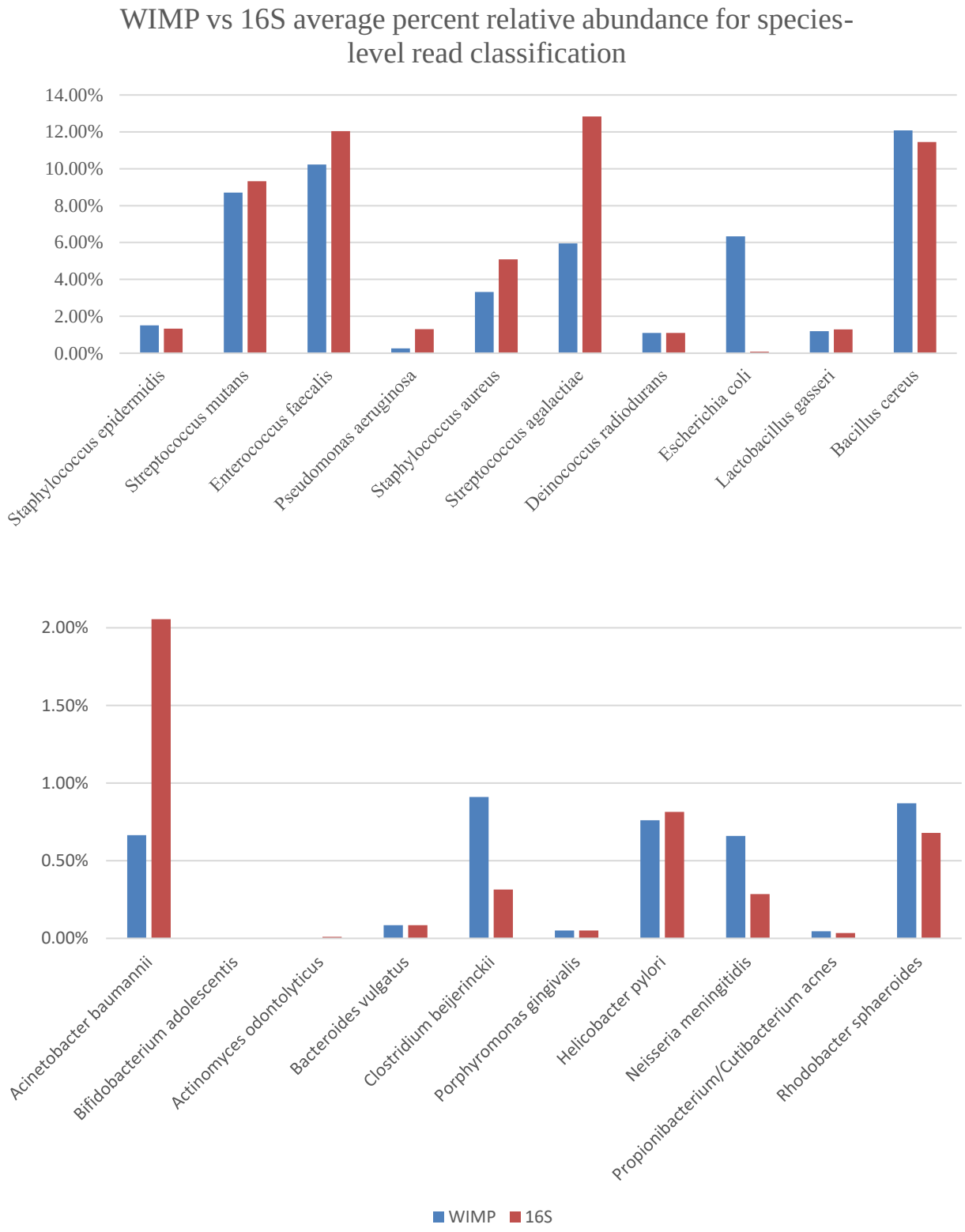

Figure 10. Pipeline comparison of average percent of reads assigned to species.

16S rRNA Copy Number Variation

The number of copies of the 16S gene in the genome differs from species to species, which may lead to biases when determining relative abundances in a mixed microbial sample. The mixed standard used for this study is supposedly prepared as an even mixture of whole cells from 20 species of bacteria. The potential to distort interpretation may present if a classified sequence represents a high copy number taxon of lesser abundance, or a low copy number taxon of higher abundance [26]. Since methods of identifying sample to host individual rely on determination of taxa present and their corresponding abundances, there is a great deal of importance placed on assignment of the correct relative abundances to casework samples. To normalize reads classified by WIMP and 16S, number of reads for each species was divided by the species' corresponding number of copies of the 16S rRNA gene. Reads were normalized by dividing the copy number corrected classified species reads by the corrected total number of classified reads, assuming an equal distribution of species in the mock community. Results were displayed in stack plot format in **Figures 11 and 12** for both analytical pipelines to demonstrate differences in observed and expected percent classified reads per species. The distribution of percentage reads across species is expected to be more balanced; with species that had a high percent of observed reads and a high number of copies of the 16S gene to have a lower expected percent after correcting for copy number. However, this was not always the case, as in *Enterococcus faecalis* which has a copy number of 4 and increased from 11.90% observed reads to 30.33% expected reads.

Species	ATCC Copy Number	Run 1 Observed Percent Reads	Run 1 Copy Number Corrected Percent	Run 2 Observed Percent Reads	Run 2 Copy Number Corrected Percent
Acinetobacter baumannii	6	0.74%	1.37%	0.59%	1.01%
Bacillus cereus	12	12.98%	12.07%	11.17%	9.49%
Bifidobacterium adolescentis	5	0.00%	0.00%	0.00%	0.00%
Deinococcus radiodurans	7	1.38%	2.20%	0.81%	1.17%
Escherichia coli	7	6.62%	10.55%	6.04%	8.80%
Lactobacillus gasseri	6	1.10%	2.05%	1.30%	2.21%
Porphyromonas gingivalis	4	0.06%	0.18%	0.04%	0.10%
Pseudomonas aeruginosa	4	0.34%	0.95%	0.19%	0.48%
Staphylococcus aureus	6	3.17%	5.89%	3.48%	5.92%
Streptococcus agalactiae	7	4.71%	7.50%	7.19%	10.47%
Actinomyces odontolyticus	2	0.00%	0.00%	0.00%	0.00%
Bacteroides vulgatus	7	0.10%	0.16%	0.07%	0.10%
Clostridium beijerinckii	14	0.92%	0.73%	0.90%	0.65%
Enterococcus faecalis	4	8.55%	23.85%	11.90%	30.33%
Helicobacter pylori	2	0.80%	4.47%	0.72%	3.69%
Neisseria meningitidis	4	0.72%	2.01%	0.60%	1.53%
Propionibacterium/Cutibacterium acnes	4	0.05%	0.13%	0.04%	0.10%
Rhodobacter sphaeroides	3	0.99%	3.70%	0.75%	2.55%
Staphylococcus epidermidis	5	1.34%	2.99%	1.69%	3.44%
Streptococcus mutans	5	8.61%	19.21%	8.81%	17.96%

Table 6. WIMP sequence reads per species corrected for 16S rRNA copy number.

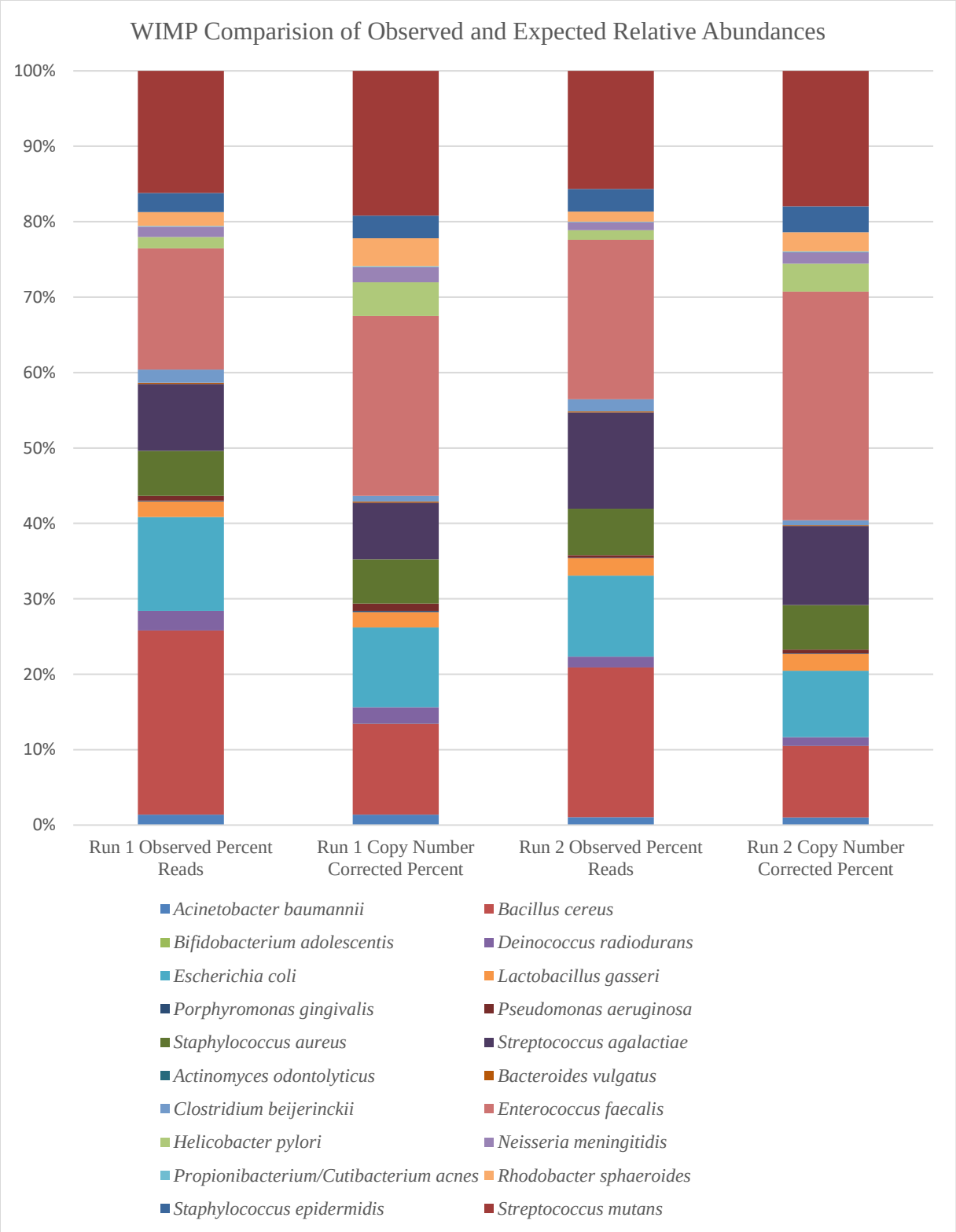

Figure 11. Stack plot of differences in observed and expected percent reads following correction for copy number from WIMP workflow.

Species	ATCC Copy Number	Run 1 Observed Percent Reads	Run 1 Copy Number Corrected Percent	Run 2 Observed Percent Reads	Run 2 Copy Number Corrected Percent
Acinetobacter baumannii	6	2.37%	3.87%	1.74%	2.64%
Bacillus cereus	12	10.97%	8.98%	11.92%	9.03%
Bifidobacterium adolescentis	5	0.00%	0.00%	0.00%	0.00%
Deinococcus radiodurans	7	1.36%	1.91%	0.83%	1.08%
Escherichia coli	7	0.11%	0.16%	0.05%	0.06%
Lactobacillus gasseri	6	1.22%	1.99%	1.37%	2.07%
Porphyromonas gingivalis	4	0.06%	0.16%	0.04%	0.09%
Pseudomonas aeruginosa	4	1.56%	3.83%	1.06%	2.42%
Staphylococcus aureus	6	4.58%	7.49%	5.59%	8.47%
Streptococcus agalactiae	7	11.81%	16.57%	13.86%	17.99%
Actinomyces odontolyticus	2	0.01%	0.05%	0.01%	0.04%
Bacteroides vulgatus	7	0.10%	0.14%	0.07%	0.10%
Clostridium beijerinckii	14	0.34%	0.24%	0.29%	0.19%
Enterococcus faecalis	4	10.81%	26.54%	13.26%	30.10%
Helicobacter pylori	2	0.85%	4.19%	0.78%	3.52%
Neisseria meningitidis	4	0.33%	0.82%	0.24%	0.55%
Propionibacterium/Cutibacterium acnes	4	0.04%	0.09%	0.03%	0.07%
Rhodobacter sphaeroides	3	0.77%	2.52%	0.59%	1.78%
Staphylococcus epidermidis	5	1.29%	2.54%	1.37%	2.48%
Streptococcus mutans	5	9.13%	17.92%	9.53%	17.32%

Table 7. 16S sequence reads per species corrected for 16S rRNA copy number.

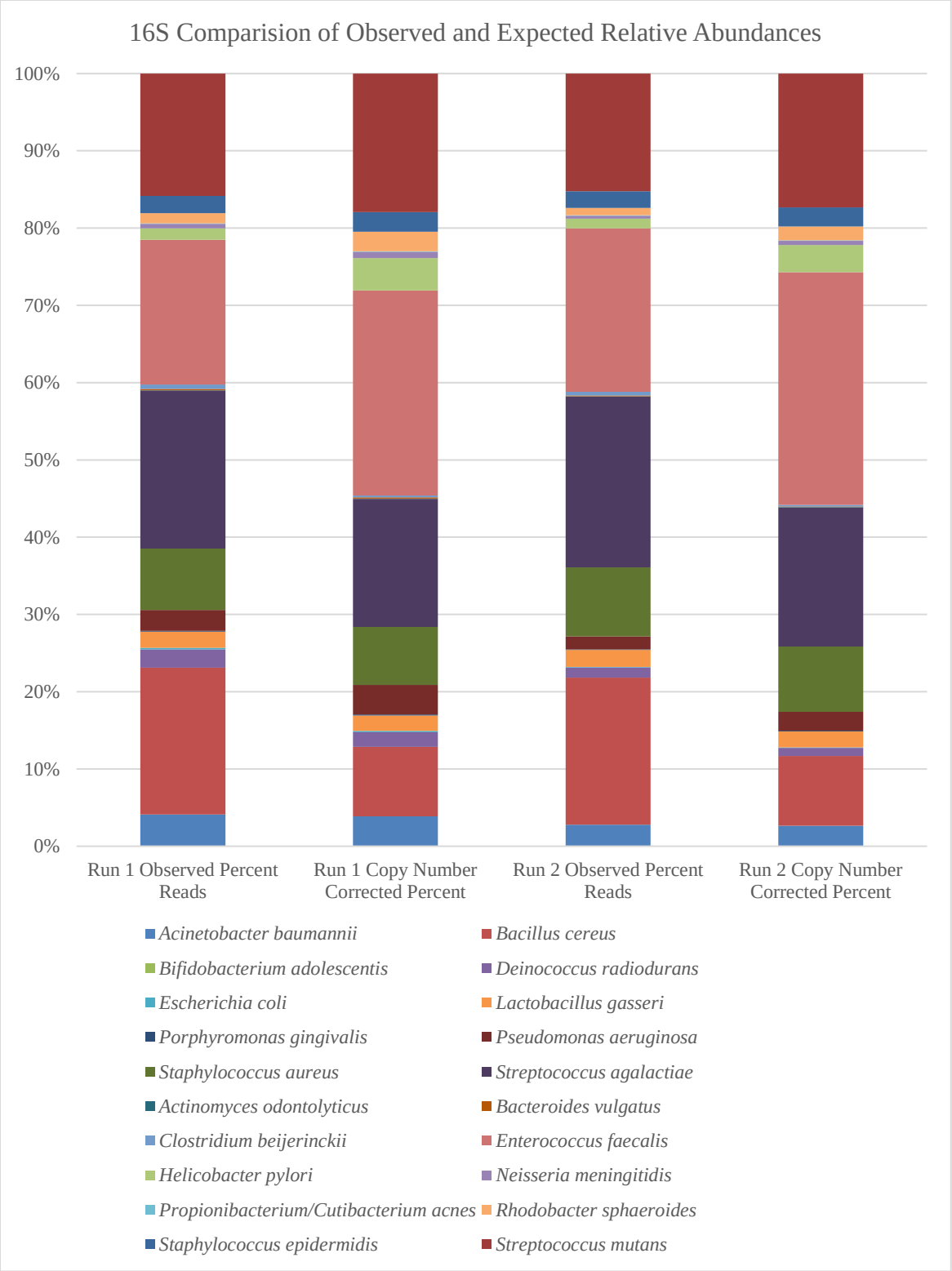

Figure 12. Stack plot of differences in observed and expected percent reads following correction for copy number from 16S workflow.

Comparison to Previous Study

A limitation cited by Foley (2018) of sequencing a region within the 16S rRNA gene was the lack of species-level taxonomic resolution [7]. Foley (2018) found that identification of *P. acnes* and *P. aeruginosa* was not possible past the family level, however both species were identified using WIMP software at low relative abundances [7]. In this study, identification of *A. odontolyticus* was not possible at the species level (genus level only) using WIMP software, and *B. adolescentis* identification was not possible at the species level using 16S software. Regarding classified reads from each extraction protocol, there was a higher number of cumulative classified reads from both sequencing runs obtained from extracts of FastDNA compared to PowerSoil as in Foley's study **(Figure 13).** A breakdown of read assignment to species by workflow and extraction method is provided in the appendix (WIMP page 40, 16S page 41).

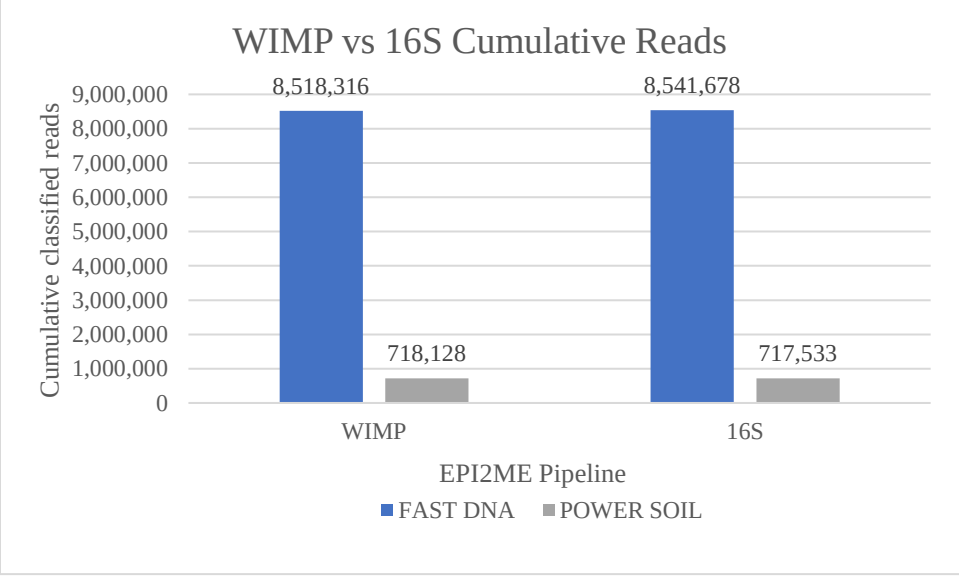

Figure 13. Comparison of cumulative reads from both sequencing runs for FastDNA and PowerSoil extraction kits.

Sources of Error

There was variation in the relative abundance of the species present in the mock microbial community for both analytical packages. The mock community contains an even mixture of 20 bacterial species each with a relative abundance of 5% according to the manufacturer. However, upon sequencing it is apparent that certain species were present in much higher or lower proportion than expected. Foley (2018) detailed a possible source of error as inefficient lysis of Gram positive bacterial species during extraction, or inefficient lysis in general [7]. Shearing of DNA during extraction is also a concern, which may result in poor primer binding. An alternative cause for the discrepancy in relative abundance is variability of 16S rRNA copy number and genome size across different species [27]. Even after correcting for copy number variation, the percent relative abundance by species did not reflect the expected percent relative abundance (5%) from the standard. It is also possible that the reason some species were undetected or present at low relative abundances was due to unsuitable primer design for certain species. Since the primers included in the SQK-RAB204 were designed to flank the entire region of the 16S gene, species that do not align correctly to the forward or reverse primer, or if the strand of DNA has been broken in anyway, the strand will fail to be amplified and thus not detected downstream.

Most classified species that were not included in the microbial standard were classified at a genus level that was included in the microbial standard. The reason for spurious identification of species may be attributed to the species having highly similar sequences that are not always discernible by WIMP or 16S. Species-level identifications are made by reference databases when sequences are 97% similar or greater, and at the genus level when a sequence has 90% identity or

greater, though there is no universally defined threshold [23]. Highly similar sequences may have arisen due to the presence of chimeric sequences, which are formed when two or more biological sequences are joined together. Chimeric sequences may arise during PCR amplification when incomplete extension results in a partially extended strand of DNA binding to another similar but different strand of DNA. The new chimeric sequence is duplicated during the cycles of PCR and results in erroneous identification at the genus and species level downstream. Nanopore's EPI2ME platform currently does not support the removal of chimeric sequences [26].

CHAPTER IV

CONCLUSIONS

In this study the entire length of the 16S rRNA gene was sequenced using ONT's MinION™ device with the hopes of obtaining increased species resolution of a mock microbial community compared to the previous study from Foley (2018) which employed traditional methods. In this study taxa were identified beyond the 20 species that were expected to be included in the mock microbial community. The presence of unexpected species could be due to issues with bioinformatic pipelines, unsuitable primer design, chimeric sequences, or less than optimal extraction or amplification parameters.

For any novel protocol to be used in a court of law, certain standards must be met. Determination of the most effective means of producing reliable, reproducible and accurate results should be evaluated. While this study demonstrates the abilities of the MinION™ device from Oxford Nanopore Technologies to quickly sequence microbial samples, certain parameters of the experimental design have room for improvement.

A challenge of this study is presented by the samples used for this study. The samples used were extracted as part of a previous study. Since the samples were extracted by someone else, it is not possible to evaluate this step in the analytical process. Whether contamination was introduced at some point in the extraction phase is indeterminate.

An aim of this study was to determine whether there is consensus between EPI2ME's 16S and WIMP software packages. The two pipelines made similar genus and species level

identifications with corresponding relative abundances; however, each pipeline identified different species and genera that were not expected in the mock community. While both pipelines are run through EPI2ME, each makes lowest taxonomic identifications in a different manner. Further study into which pipeline employs a more effective analysis should be considered in future studies. Future directions of this project should include examination of chimeric sequences and at what point in the process they are introduced. Amplification bias is a well-documented phenomenon so studies of adjusted number of PCR cycles or an altered protocol to bypass PCR entirely should be explored.

Results from this study are comparable to the results detailed in Foley's study (2018); yet the MinION™ offers a more time efficient procedure with longer read lengths. This study has demonstrated the ability of Nanopore's MinION™ device to characterize and quantify taxa present in a microbial standard containing species that might be found in a casework sample. The MinION™ device's portability and ease of application makes it suitable for use in the field. For now, the applicability of using the MinION™ would be effective for preliminary assessment of samples. It should be noted that two species, *A. odontolyticus*, and *B. adolescentis*, present in the defined community were weakly or not detected with either Oxford Nanopore pipeline, but were identified by Foley (2018) [7]. These results suggest possible problems with amplification of these species using the SQK-RAB204 workflow and the 27F/1492R primer set or problems in bioinformatic taxonomic classification. There were very few reads classified at the higher classifications of these species, consistent with primer binding issues, bioinformatic classification error, or amplification inefficiencies. Though it was possible to garner species level resolution, there is need for validation of this protocol to develop the most effective means of generating reliable and accurate species-level identification and relative abundance results.

APPENDIX

Phylum	Order	Family	Genus	Species	Gram +/-	Identified with WIMP?	Identified with 16S?
Proteobacteria	Pseudomonadales	Moraxellaceae	*Acinetobacter*	*baumanni*	NEG	YES	YES
Actinobacteria	Actinomycetales	Actinomycetaceae	*Actinomyces*	*odontolyticus*	POS	NO	YES
Firmicutes	Bacillales	Bacillaceae	*Bacillus*	*cereus*	POS	YES	YES
Bacteroidetes	Bacteroidales	Bacteroidaceae	*Bacteroides*	*vulgatus*	NEG	YES	YES
Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	*Bifidobacterium*	*adolescentis*	POS	YES	NO
Firmicutes	Clostridiales	Clostridiaceae	*Clostridium*	*beijerinckii*	POS	YES	YES
Deinococcus-Thermus	Deinococcales	Deinococcaceae	*Deinococcus*	*radiodurans*	POS	YES	YES
Firmicutes	Lactobacillales	Enterococcaceae	*Enterococcus*	*faecalis*	POS	YES	YES
Proteobacteria	Enterobacteriales	Enterobacteriaceae	*Escherichia*	*coli*	NEG	YES	YES
Proteobacteria	Campylobacterales	Helicobacteraceae	*Helicobacter*	*pylori*	NEG	YES	YES
Firmicutes	Lactobacillales	Lactobacillaceae	*Lactobacillus*	*gasseri*	POS	YES	YES
Proteobacteria	Neisseriales	Neisseriaceae	*Neisseria*	*meningitidis*	NEG	YES	YES
Bacteroidetes	Bacteroidales	Porphyromonadaceae	*Porphyromonas*	*gingivalis*	NEG	YES	YES
Actinobacteria	Propionibacteriales	Propionibacteriaceae	*Propionibacterium/ Cutibacterium*	*acnes*	POS	YES	YES
Proteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas*	*aeruginosa*	NEG	YES	YES
Proteobacteria	Rhodobacterales	Rhodobacteraceae	*Rhodobacter*	*sphaeroides*	NEG	YES	YES
Firmicutes	Bacillales	Staphylococcaceae	*Staphylococcus*	*aureus*	POS	YES	YES
Firmicutes	Bacillales	Staphylococcaceae	*Staphylococcus*	*epidermidis*	POS	YES	YES
Firmicutes	Lactobacillales	Streptococcaceae	*Streptococcus*	*agalactiae*	POS	YES	YES
Firmicutes	Lactobacillales	Streptococcaceae	*Streptococcus*	*mutans*	POS	YES	YES

All species belong to the Kingdom Bacteria.

	WIMP							
	Run 1				Run 2			
Species	FastDNA	PowerSoil	% FastDNA	% PowerSoil	FastDNA	PowerSoil	% FastDNA	% PowerSoil
Acinetobacter baumannii	10,092	6,853	0.53%	1.77%	36,488	6,024	0.56%	1.86%
Bacillus cereus	262,174	36,754	13.69%	9.49%	771,258	29,052	11.79%	8.95%
Bifidobacterium adolescentis	34	2	0.00%	0.00%	99	7	0.00%	0.00%
Deinococcus radiodurans	21,454	10,296	1.12%	2.66%	44,498	13,310	0.68%	4.10%
Escherichia coli	81,238	71,153	4.24%	18.36%	369,939	63,289	5.66%	19.49%
Lactobacillus gasseri	22,367	3,012	1.17%	0.78%	90,425	2,690	1.38%	0.83%
Porphyromonas gingivalis	1,104	382	0.06%	0.10%	2,399	378	0.04%	0.12%
Pseudomonas aeruginosa	48	3,008	0.00%	0.78%	11,271	2,250	0.17%	0.69%
Staphylococcus aureus	69,760	3,159	3.64%	0.82%	247,203	25	3.78%	0.01%
Streptococcus agalactiae	102,969	5,360	5.38%	1.38%	508,805	6,287	7.78%	1.94%
Actinomyces odontolyticus	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Bacteroides vulgatus	1,617	698	0.08%	0.18%	4,331	7	0.07%	0.00%
Clostridium beijerinckii	150	6,176	0.01%	1.59%	59,014	53	0.90%	0.02%
Enterococcus faecalis	189,305	7,585	9.89%	1.96%	844,812	7,936	12.92%	2.44%
Helicobacter pylori	11,993	6,455	0.63%	1.67%	46,811	50	0.72%	0.02%
Neisseria meningitidis	110	5,580	0.01%	1.44%	38,170	4,915	0.58%	1.51%
Propionibacterium/Cutibacterium acnes	11	17	0.00%	0.00%	2,876	27	0.04%	0.01%
Rhodobacter sphaeroides	148	81	0.01%	0.02%	42,446	11,295	0.65%	3.48%
Staphylococcus epidermidis	29,178	8,081	1.52%	2.09%	119,137	1,637	1.82%	0.50%
Streptococcus mutans	181,322	16,882	9.47%	4.36%	616,786	14,506	9.43%	4.47%

Number of reads classified per species separated by extraction protocol and sequencing run from the What's In My Pot? (WIMP) workflow.

Results can be accessed at the following websites.

[Instance ID: 187149] Run 1

https://epi2me.nanoporetech.com/workflow_instance/187149?token=48100702-0248-11E9-B933-A89B9F7EF595

[Instance ID: 192704] Run 2

https://epi2me.nanoporetech.com/workflow_instance/192704?token=DFA9DA10-3A9D-11E9-9631-0D40B41D277E

	16S							
	Run 1				Run 2			
Species	FastDNA	PowerSoil	% FastDNA	% PowerSoil	FastDNA	PowerSoil	% FastDNA	% PowerSoil
Acinetobacter baumannii	32,526	21,779	1.70%	5.65%	103,634	17,175	1.57%	5.23%
Bacillus cereus	215,929	35,702	11.31%	9.26%	793,345	32,682	12.02%	9.96%
Bifidobacterium adolescentis	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Deinococcus radiodurans	20,881	10,292	1.09%	2.67%	44255	13,304	0.67%	4.05%
Escherichia coli	1614	944	0.08%	0.24%	2751	438	0.04%	0.13%
Lactobacillus gasseri	24,730	3,216	1.30%	0.83%	92,066	2688	1.39%	0.82%
Porphyromonas gingivalis	1082	381	0.06%	0.10%	2397	389	0.04%	0.12%
Pseudomonas aeruginosa	22,068	13,681	1.16%	3.55%	61,309	12,461	0.93%	3.80%
Staphylococcus aureus	101,197	3,807	5.30%	0.99%	384,143	3,468	5.82%	1.06%
Streptococcus agalactiae	257,677	13,188	13.50%	3.42%	948,642	11,857	14.37%	3.61%
Actinomyces odontolyticus	209	14	0.01%	0.00%	594	31	0.01%	0.01%
Bacteroides vulgatus	1642	710	0.09%	0.18%	4410	741	0.07%	0.23%
Clostridium beijerinckii	5504	2207	0.29%	0.57%	18696	1704	0.28%	0.52%
Enterococcus faecalis	238,548	9,391	12.50%	2.44%	909,881	8,549	13.79%	2.61%
Helicobacter pylori	12675	6,891	0.66%	1.79%	48,488	5,240	0.73%	1.60%
Neisseria meningitidis	5216	2436	0.27%	0.63%	14926	5240	0.23%	1.60%
Propionibacterium/Cutibacterium acnes	804	15	0.04%	0.00%	2084	21	0.03%	0.01%
Rhodobacter sphaeroides	11339	6338	0.59%	1.64%	32085	8631	0.49%	2.63%
Staphylococcus epidermidis	28,176	1512	1.48%	0.39%	93,466	1269	1.42%	0.39%
Streptococcus mutans	190,990	18,352	10.01%	4.76%	645,042	15,282	9.77%	4.66%

Number of reads classified per species separated by extraction protocol and sequencing run from the 16S Taxonomic Classification workflow.

Results can be accessed at the following websites.

[Instance ID: 188812] Run 1

https://epi2me.nanoporetech.com/workflow_instance/188812?token=36952D56-1510-11E9-B83A-EF655D9A848

[Instance ID: 195070] Run 2

https://epi2me.nanoporetech.com/workflow_instance/195070?token=DF4C70C2-4F1B-11E9-9442-E15273559BE7

REFERENCES

[1] A.J. Jeffreys, V. Wilson, S.L. Thein, Individual-specific ‘fingerprints’ of human DNA, Nature. 316 (1985) 76–79. doi:10.1038/316076a0.

[2] R. Sender, S. Fuchs, R. Milo, Revised Estimates for the Number of Human and Bacteria Cells in the Body., PLoS Biol. 14 (2016) e1002533. doi:10.1371/journal.pbio.1002533.

[3] N. Peng, H. Poon, C. Quirk, K. Toutanova, W. Yih, Cross-Sentence N-ary Relation Extraction with Graph LSTMs, 70 (2017) 1–12. doi:10.1111/j.1753-4887.2012.00493.x.Defining.

[4] A.E. Woerner, N.M.M. Novroski, F.R. Wendt, A. Ambers, R. Wiley, S.E. Schmedes, B. Budowle, Forensic human identification with targeted microbiome markers using nearest neighbor classification, Forensic Sci. Int. Genet. 38 (2019) 130–139. doi:10.1016/j.fsigen.2018.10.003.

[5] J.T. Hampton-Marcell, J. V. Lopez, J.A. Gilbert, The human microbiome: an emerging tool in forensics, Microb. Biotechnol. 10 (2017) 228–230. doi:10.1111/1751-7915.12699.

[6] J. Stenson, G. Brown, A.C. Bateman, J.L. Green, B.J.M. Bohannan, J.F. Meadow, A.E. Altrichter, Humans differ in their personal microbial cloud, PeerJ. 3 (2015) e1258. doi:10.7717/peerj.1258.

[7] B. Foley, Use of ATCC MSA-2002 for validation of extraction and amplification

techniques in 16S microbial community profiling, (2018). The University of North Texas Health Science Center.

[8] B. Yang, Y. Wang, P.Y. Qian, Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis, BMC Bioinformatics. 17 (2016) 1–8. doi:10.1186/s12859-016-0992-y.

[9] R. Sinha, M. Nitin, M.P. Sinha, S. Chakravorty, D. Helb, M. Burday, N. Connell, 16s rDNA Based Identification of Bacteria in the Organophosphates Treated Agricultural Soil, 69 (2008) 330–339. doi:10.1016/j.mimet.2007.02.005.A.

[10] A. Benítez-Páez, K.J. Portune, Y. Sanz, Species-level resolution of 16S rRNA gene amplicons sequenced through the MinION™ portable nanopore sequencer, Gigascience. 5 (2016) 1–9. doi:10.1186/s13742-016-0111-z.

[11] S.T. Calus, U.Z. Ijaz, A.J. Pinto, NanoAmpli-Seq: a workflow for amplicon sequencing for mixed microbial communities on the nanopore sequencing platform, Gigascience. 7 (2018) 1–16. doi:10.1093/gigascience/giy140.

[12] M. Jain, H.E. Olsen, B. Paten, M. Akeson, The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community, Genome Biol. 17 (2016) 1–11. doi:10.1186/s13059-016-1103-0l.

[13] A.D. Tyler, L. Mataseje, C.J. Urfano, L. Schmidt, K.S. Antonation, M.R. Mulvey, C.R. Corbett, Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications, Sci. Rep. 8 (2018) 1–12. doi:10.1038/s41598-018-29334-5.

[14] S. Goldstein, L. Beka, J. Graf, J.L. Klassen, Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing, BMC Genomics. 20

(2019) 1–18. doi:10.1186/s12864-018-5381-7.

[15] Oxford Nanopore Technologies. Types of Nanopores, (Accessed March 2019). https://nanoporetech.com/how-it-works/types-of-nanopores.

[16] Oxford Nanopore Technologies. DNA: nanopore sequencing, (Accessed February 2019). https://nanoporetech.com/applications/dna-nanopore-sequencing.

[17] Oxford Nanopore Technologies How it Works, (Accessed December 2018). https://nanoporetech.com/how-it-works.

[18] E.N. Hanssen, K.H. Liland, P. Gill, L. Snipen, Optimizing body fluid recognition from microbial taxonomic profiles, Forensic Sci. Int. Genet. 37 (2018) 13–20. doi:10.1016/j.fsigen.2018.07.012.

[19] SQK RAB-204 Protocol, (2018). version: RAB_9053_v1_revE_.

[20] Oxford Nanopore Technologies. Community: WIMP workflow, (2018). https://community.nanoporetech.com/technical_documents/epi2me-tech-doc/v/metd_5000_v1_revam_29feb2016/what-s-in-my-pot-wimp.

[21] Center for Computational Biology, (2016). https://ccb.jhu.edu/software/centrifuge/manual.shtml.

[22] Oxford Nanopore Community. 16S workflow, (2018). https://community.nanoporetech.com/technical_documents/epi2me-tech-doc/v/metd_5000_v1_revam_29feb2016/16s-taxonomic-classificati.

[23] J.M. Janda, S.L. Abbott, 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls, J. Clin. Microbiol. 45 (2007) 2761–2764. doi:10.1128/JCM.01228-07.

[24] S.R. Reichley, C. Ware, J. Steadman, P.S. Gaunt, J.C. García, B.R. LaFrentz, A. Thachil,

G.C. Waldbieser, C.B. Stine, N. Buján, C.R. Arias, T. Loch, T.J. Welch, R.C. Cipriano, T.E. Greenway, L.H. Khoo, D.J. Wise, M.L. Lawrence, M.J. Griffin, Comparative Phenotypic and Genotypic Analysis of Edwardsiella Isolates from Different Hosts and Geographic Origins, with Emphasis on Isolates Formerly Classified as E. tarda, and Evaluation of Diagnostic Methods, J. Clin. Microbiol. 55 (2017) 3466–3491. doi:10.1128/jcm.00970-17.

[25] D. Kim, L. Song, F.P. Breitwieser, S.L. Salzberg, Centrifuge : rapid and sensitive classification of metagenomic sequences, (2016) 1721–1729. doi:10.1101/gr.210641.116.Freely.

[26] S.F. Stoddard, B.J. Smith, R. Hein, B.R.K. Roller, T.M. Schmidt, rrnDB: Improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development, Nucleic Acids Res. 43 (2015) D593–D598. doi:10.1093/nar/gku1201.

[27] P. Baldrian, The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses, 8 (2013) 1–10. doi:10.1371/journal.pone.0057923.