*Microbacterium*, *Acinetobacter*, and Nonpathogenic Bacteria Found in Both San Francisco Public Transportations – BART and Muni

Jacqueline Imperial

November 25, 2019

# Introduction

Bacterial microbes present in urban locations have been a major public health issue especially during the peak of the flu season. These major concerns arise from the transfer of bacteria from person to person in public locations such as parks, schools, hospitals, and public transportation, all containing abundant bacteria (Reynolds *et al.*, 2005). These bacteria may pose a threat to the health of individuals due to a high probability of the bacteria being pathogenic (Onat *et al.*, 2017). According to numerous studies, some pathogens found in public transit include the *Staphylococcus* genus (Stepanović *et al.*, 2008), commonly found on human skin, and *S. aureus*, a pathogen responsible for giving rise to Methicillin-resistant *Staphylococcus aureus*, or MRSA (Otter and French, 2009). MRSA can cause mild infections on the skin, but some exposure may be life-threatening (Carleton *et al.*, 2004). Therefore, due to this public health concern, it is crucial to analyze the possibly dangerous strains of microbial communities present and growing in such locations.

The aim of my study is to shed light on the lack of cleanliness and maintenance done on public transportation in major cities such as San Francisco’s Bay Area Rapid Transit (BART) and Municipal Railway (Muni). San Francisco State University’s Biology Department has recently extracted microbial communities in micro-locations of public transportation harboring abundant bacteria. These locations included hand-touch surfaces, such as handrails and doors, seats, and the floor. The findings of this particular study found that the majority of bacteria was located on the floor (Yeh *et al.*, 2011). Numerous other similar research has been done in in different cities such as London and Seoul in order to investigate the culturable airborne bacterial and fungal communities in public transportation locations (Hwang and Park, 2014).

Similar to the study previously mentioned, the purpose of my study was to examine and determine the types of bacterial communities harbored on the floor of the BART and Muni. Particularly, my primary focal question for this study was “what kinds of bacteria contaminated these environments?” Additionally, I wanted to know how drastically the diversity and abundance of the bacteria differed between the two types of public transportation in the city area. Based on these two leading questions, I hypothesized that *Staphylococci* and other less harmful pathogens would be present in both locations. This hypothesis is derived from both the lack of cleanliness from my personal experiences using BART and Muni, and previous findings from similar studies (Redman *et al.*, 2013). I also believe that the diversity will be larger in the Muni due to the lack of maintenance and very high volumes of passengers on a day-to-day basis.

In order to test these hypotheses, I obtained 3 individual swab samples from each the Muni and BART floors. I performed both a culture and culture free procedure for the two different treatments of each of the 3 samples. For the culture samples, the 6 samples in total were diluted 1, 1:10, and 1:100 respectively and cultured on an agar dish. I extracted and amplified the DNA from the 6 most isolated colonies through PCR. Afterwards, I utilized gel electrophoresis in order to separate these DNA fragments according to the size and charge. Lastly, I used unidirectional Sanger sequencing, which allowed for the further analysis of the specific DNA composition. I compared the successful sequences with BLAST, which helped me determine what types of microbial strains were present within the colony samples (Griffen *et al.*, 2011).

In addition, to test the culture-free samples, I utilized PCR to amplify the sequences and add a barcode later for Illumina sequencing. I also quanitified the data through Qubit, separated the products with gel electrophoresis, and performed Illumina sequencing (Meyer and Kircher, 2010). With the Illumina data, I used sequence analysis methods such as fastqc, trimmometric, bioawk, and BLAST to assess the quality, trim the data, and BLAST the results to determine the different taxa within the abundant microbiome (Monger *et al.*, 2017).

The primary findings of this study from the culture samples revealed that the Muni contained more colonies after culturing. These abundant colonies from the Muni were also more diverse in morphotypes compared to the cultured BART samples. According to the molecular data, both BART 1 diluted 1:10 and 1:100 possessed *Microbacterium* strains and BART 2 1:100 contained *Acinetobacter schindleri* strain. Muni 2 1:10 also contained a strain of *Acinetobacter* and Muni 3 1:100 contained *Marinobacterium* (Takeuchi and Hatano, 1998).

# Methods

## Sampling

Materials required for the extraction of samples were 12 sterile 1.5mL tubes and swabs, gloves, zip block bags, and sterile phosphate-buffered saline (PBS). I collected all the samples on the same day. For the BART, I obtained the samples from the Richmond train from three different carts. I extracted the Muni samples from three difference buses on the 5R route. I specifically obtained all samples from the entrances of all trains and buses, where foot-traffic is the greatest. I took two swab samples from each specific location. For sample collection, I first dipped the sterile swab and moistened it into the sterile PBS before gathering the sample from the floor. I gently wiped the floor surface for approximately 10 seconds with one end of the swab. I placed the swab into its designated tube, and broke off the long end of the swab in order to properly fit it in the tube. Each BART and Muni samples were labeled appropriately.

## Culture Samples

### Culturing and DNA Extraction

24 hours after extraction, I diluted and cultured the samples. This culturing method required PBS, or buffer solution, 18 100mm TSA plates, and rattler beads. Before plating, all 6 samples were diluted 1:1, 1:10, and 1:100. I created the dilutions with the samples and designated amounts of PBS. In order to culture all 18 samples and and the dilutions, I pipetted 100μl of each sample onto the plate. I used 18 of the 100mm TSA plates for plating samples from each of my 2 treatments (n=3 per group, 6 total). In order to spread the solution evenly throughout the surface of the plate, I used rattler beads. All 18 samples were then incubated at 37ºC for approximately 4 days before the analysis of colony abundance and morphotypes. Afterwards, I then recorded my observations based on the culture abundance counts and number of distinguishable morphotypes.

I extracted DNA from the cultures with the most isolated colonies. For this study, these 6 target colonies derived from B1 1:10, B1 1:100, B2 1:100, M2 1:10, M3 1:10, and M3 1:100. For DNA extraction, I followed the manufacturer’s protocol from the Sigma REDExtract-N-Amp kit (Weber and Douglas). Afterwards, I used Qubit to quantify the target colony DNA extractions (Haines and Linacre, 2019). The Qubit machine quantified each sample and I recorded the DNA concentrations in ng/μL.

### PCR Amplification and Gel Electrophoresis

During the following week, a PCR procedure was utilized in order to amplify the DNA extractions (Kuno, 1998). Before beginning the protocol, I calculated the necessary volumes of reagents needed for the master mix using (n+1) + 10%, where n=7. The volumes for one reaction were 10μl Amp, 0.8μl 27f primer, 0.8μl 1492r primer, 1μl BSA, and 6.4μl H2O. I aliquoted 19μl of the master mix into all sample tubes and the negative control tube. Additionally, I pipetted 1μl of DNA extraction sample into its designated tube and 1μl H2O into the negative control. The PCR ran at 95º C for 5 minutes in order to denature and separate the long DNA strands. The temperature was then lowered to 94ºC for 30 seconds. Temperature was further lowered to 65ºC for 30 seconds, which is critical for annealing and for the primers to sit on the DNA. The temperature was increased again to 72ºC for 1 minute for extension and elongation by DNA polymerase. The temperature was raised to 95ºC for 30 seconds, 55ºC for 30 seconds, and 72ºC for 1 minute for a total of 25 cycles. Afterwards, the temperature was held at 72ºC for 10 minutes and held 4ºC. The touchdown PCR contained 2% agarose gel, SYBR safe dye, and TAE buffer. After the PCR image was taken, I recorded all observations based on the brightness or faintness of the band. Following the PCR, I ran the products on a gel electrophoresis gel at 140 volts for 30 minutes in order to separate the DNA fragments according to their size and charge (Aaij and Borst, 1972).

### Sanger Sequencing

After checking for successful amplification from PCR utilizing gel electrophoresis, I cleaned the PCR products using Invitrogen Exo-SAP (Exonuclease – Shrimp Alkaline Phosphatase). They were then sent for unidirectional Sanger sequencing at MCLAB located in South San Francisco, CA (Blazej *et al.*, 2006). The Sanger sequencing results were finalized and sent back the following day for later analysis such as trimming and cleaning.

## Culture-Free Samples

### PCR and Gel Electrophoresis

For the culture-free samples, I began with PCR prep to amplify the DNA later for Illumina sequencing and analysis. For the PCR master mix, similar to the culture sample procedure, I calculated the necessary volumes of reagents needed using (n+1) + 10%, where n=7. The volumes for one reaction were 10μl Amp, 0.8μl iseq 16SF, 0.8μl iseq 16R, 1μl BSA, 6.4μl H2O, and, 1μl of the DNA extraction. The negative control sample contained 1μl of H2O instead. The automated PCR cycles ran at 95ºC for 3 minutes, 95ºC for 30 more seconds, lowered to 55ºC for 30 seconds, 72ºC for an additional 30 seconds, and 72ºC for 5 minutes. Lastly, the PCR products were finally held at a temperature of 4ºC. I repeated this automated process for a total of 25 cycles, allowing for amplification and the addition of a barcode for later Illumina sequencing. The touchdown PCR contained 2% agarose gel, SYBR safe dye, and TAE buffer. The image of the PCR bands were taken and I recorded observations based on the brightness and faintness of each band. Afterward, I utilized Qubit to quantify the data and recorded each of the concentrations.

### Illumina Sequencing

Following electrophoresis, PCR products were purified using AmpureXP magnetic beads (Beckman-Coulter) and quantitated using a PicoGreen fluorescent assay (Invitrogen) on a Tecan Infinite M Plex plate reader. Purified PCR products were used as the template for a second round of PCR, which served to attach unique pairs of forward and reverse Illumina barcodes (Nextera XT Index 2 kit). All other components of the PCR mixtures were as for the first-round PCR. These reactions were cycled at 95 degrees C for 3 minutes, then 8 cycles of: 95 degrees C for 30 s, 55 degrees C for 30s, and 72 degrees C for 30s, followed by a 5 minute elongation cycle at 72 degrees C. Following this amplification, PCR products were purified and normalized with a SequelPrep normalization plate (Invitrogen), pooled, and then quantified once again with a Qubit 4 fluorometer (Invitrogen) and a TapeStation 4200 (Agilent) to verify library size and concentration. This library was then diluted to the loading concentration (50 pM) and combined with an Illumina PhiX spike-in library (5% spike in) and sequenced on an Illumina iSeq using a 2 x 150 bp consumable cartridge.

## Sequence Analysis

In order to clean and trim the culture sequences, I used Geneious software for data analysis. The alignment program used is MAFFT, which is a fast and accurate multiple sequence alignment program in Geneious (Katoh *et al.*, 2009). I first viewed all samples to determine which ones were successful and which can be cleaned and analyzed through BLAST. I then began the process of trimming the successful samples by taking out the beginning and ends of the sequence. I also cleaned the sequence data by using the IUPAC ambiguity codes to fixed those that contained mixed peaks. After cleaning and trimming the sequences, I used NCBI’s Nucleotide BLAST to determine what strains the colony contained (Johnson *et al.*, 2008). The alignment program used is MAFFT, which is a fast and accurate multiple sequence alignment program in Geneious (Katoh *et al.*, 2009). All observations of the strains of each successful sample were recorded. Additionally, I created a Maximum Likelihood Phylogeny and Bayesian Phylogeny Tree. The two phylogeny programs used were MrBayes and PHYML. The relevant parameters for MrBayes included the substitution model set to GTR and the outgroup was a *Thermus aquaticus* strain. The MCMC settings were set as well. The chain length was 1,100,000, heated chains were 4, heated chain temperature was 0.2, subsampleing freq was 200, burn-in length was 100,000, and random seed was 11,383. I kept the priors paremters in the defaut settings. The parameters for PHYML included the substitution model as GTR, branch support as Bootstrap, the number of bootstraps as 100, and the optimize was set as topology/length/rate. Additionally, I set the proportion of invariable sites as estimated as well as the gamma distribution parameter.

To analyze the sequences for the culture-free data, I ran a quality assessment of the fastqc files. I then used trimmometric to remove the unusable sequences with a secure file transfer protocol. Bioawk was then utilized to convert the fastq files to fasta. Afterwards, I carried out BLAST to determine the top match in NCBI for each sample.

# Results

Overall, to gather my culure results I cultured the bacterial colonies, extracted and quantified the DNA concentration, ran PCR and gel electrophoresis, and sent the following products for Sanger Sequencing. From the resulting Sanger products, I cleaned and trimmed the sequences and aligned the sequences using MAFFT. Additionally, I created both a Bayesian Tree, Maximum Likelihood Phylogeny, and ran BLAST on the successful samples in order to determine the strains within each sample.

After quantifying DNA PCR products with Qubit, DNA concentrations from BART significantly varied for the culture data, ranging approximately 5.92-34.1 ng/μl. However, concentrations from Muni were consistent around 8.00-9.00 ng/μl (Table 1).The Qubit DNA concentrations also varied among culture-free samples, ranging from ~4.0-10.0 ng/μl. The lowest concentration was 4.44 ng/μl and the highest was 10.0 ng/μl (Table 2).

DNA product is present within most samples except for the negative control wells. Two separate images of the PCR bands from the culture and culture-free products show the presence of PCR product. According to the image of culture PCRs derived from the DNA extractions, the majority of the samples contain faint, but distinguishable, bands and ~200 to 300bp in reference to the ladder located in the far right lane (Figure 1). M3 1:10 is the only faint band and the rest of the bands are relatively normal-bright (Figure 1). For the culture-free samples, B1, B3, and M3 are faint and B2, M1, and M2 are bright (Figure 2). All PCR products were sufficient for later analysis.

The mean values of the 10x dilution colonies were not significantly different between the two sites (p=0.82, Figure 3). There was also no significant difference between the number of morphortypes (Wilcox p= 0.16, Figure 4).

After Sanger sequencing, the samples’ sequences were cleaned and trimmed and afterwards aligned through Geneious’s MAFFT alignment program. All samples’ loci were 16S. After analysis, B1 1:10, B1 1:100. B2 1:100, and M3 1:100 were determined usable and M2 1:10 and the negative control were not usable for the next step of BLAST analysis. The lengths of the sequences after trimming and cleaning, and the manual corrections made are also changed after alignment (Table 3).

The Maximum Likelihood Phylogeny shows the phylogeny of the successful samples. M2 1:10 and B2 1:100 are more closely related than B1 1:100. The bootstrap value is 100 and the posterior probability value is 0.06 (Figure 5). A Bayesian Phylogeny is also indicated between these same samples. The Bayesian Tree indicates similar results as the phylogeny. The bootstrap value is 1 and the Bayesian posterior probability is 0.07 (Figure 6).

Lastly, the results of Nucleotide BLAST revealed which strains predominated within each sample colonies (Table 4). From this, I discovered that B1 1:10 contains *Macrobacterium phyllosphaerae* strain. B1 1:100 contains *Microbacterium hydrocarbonoxydans* strain, B2 1:100 has *Acinetobacter schindleri* strain, M2 1:100 has *Acinetobacter sp.* strain, and M3 1:100 contains a *Marinobacterium sp* MB3 gene. M3 1:10 and the negative control were failed sequences.

After running FastQC Reports on all the 6 culture-free data, I was able the sequence quality for all samples. 0 sequences were flagged as poor quality in samples B1, B2, B3, M1, M2, and M3. Sample B1 contained 5014 sequences and 147-151 sequence length, B2 contained 7668 sequences and the same sequence length, B3 had 4990 sequences and the same sequence length, M1 had 7573 sequences and 51-151 sequence length, M2 had 10693 sequences and 43-151 sequence length, and lastly M3 had 14803 sequences and 43-151 sequence length.

# Discussion

Many forms of bacteria grow on the surface of the floor in public transportation locations, especially San Francisco’s BART and Muni. The main questions going into this study were to determine what kinds of bacterial strains contaminated the two environments and whether or not the diversity and abundance greatly differed between the two treatments. From these leading questions and in reference to previous similar studies, I hypothesized that *Staphylococci* and other less harmful pathogenic bacteria would be found in both BART and Muni. Furthermore, hypothesized that the diversity would be significantly greater in the Muni samples compared to Bart samples due to the lack of maintenance and cleanliness of the city buses as well as the high volume of people going through that specific form of transportation.

According to my results, the abundances and morphotypes did not vary significantly between treatments (Figure 3 and Figure 4). This is attributed to high median value of colony abundance in the Muni samples and the mean values not varying significantly. I fail to reject the null hypothesis which states that both treatment abundances are similar (p=0.83, Figure 3). Moreover, I fail to reject the null hypothesis, which states that the morphotypes between BART and Muni are the same. Although both the mean and median vary greatly according to the boxplot, the p-value equals 0.16. Therefore overall, the results reveal that the bacterial communities growing on the BART and Muni floor surfaces do not greatly differ in abundance and diversity.

Factors that may influence these findings may be that the same groups of people may be using both types of public transportation. Or the bacteria on the bottom of individuals’ shoes do not make a major difference in the bacteria found on the public transportation floor because much of the bacteria picked up from the ground within the Bay Area are transferred equally to both locations. Additionally, the maintenance of both the BART and Muni may be more similar that I thought before the study. Despite the apparent lack of cleanliness in the Muni, it is nearly equal in the BART as well. These factor may have caused the results found in the boxplots and may have led to the conclusion that the diversity and abundance of bacterial colonies in the BART and Muni do not differ greatly. These particular results align with much of the results found by other similar studies.

My second question for this study addressed the contents of the bacteria on the floor and what strains existed in these locations. After cleaning and trimming the Sanger Sequencing data and matching it with BLAST results, I discovered the taxa in such locations. According to Table 4 of my results, I found that the majority of taxa were *Microbacterium* and *Acinetobacter* strains. Both strains are nonpathogenic and mostly not harmful for human health. These taxa findings greatly differed from the findings of studies in London, where research has found *Staphylococcus aureus* on hand-touch surfaces in public transportation (Otter and French, 2009). These findings also did not align with my initial hypothesis. I’m confident in the sample BLAST results that contained a percent identity of more than 90%. Therefore, I am confident in the BLAST results for B1 1:10, B1 1:100, B2 1:100, M2 1:10, and M3 1:10.

Factors that may influence that deviation from my taxa findings to previous studies such as the one done in London may be the difference in culture and overall location. London contains higher volumes of people of different cultures, both locals and tourists, compared to San Francisco, where most people opt to take a car to travel, rather than public transportation. Also the streets of London are less clean and maintained than the Bay Area. Lastly, the climate in the different places may also play a role in the growth of certain bacterial strains.

Additionally, the sequence for M3 1:10 was unusable for both phylogeny and BLAST. This may be due to a mistake in pipetting or extraction of colony DNA. It may also be because the colony was mixed with other neighboring colonies. Due to these possible factors, the sequence failed.

Overall, this study revealed that nonpathogenic bacterial communities grow on the floor public transportation in the San Francisco community. This indicates the improvement of maintenance within the city’s public locations. It also may reveal the future improvement of public health overall.

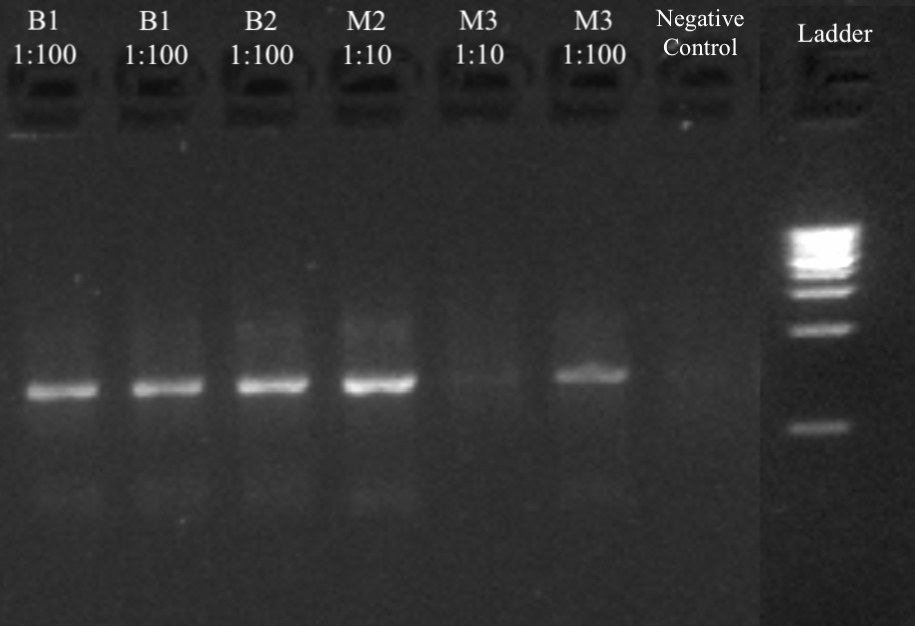
# Figures and Tables

|  |  |  |  |
| --- | --- | --- | --- |
| Sample I.D. | Sample Group | Sample Dilution I.D. | DNA Concentration (ng/μl) |
| B1 | BART | 10 | 34.1 |
| B1 | BART | 100 | 11.5 |
| B2 | BART | 100 | 5.92 |
| M2 | Municipal Railway | 10 | 8.27 |
| M3 | Municipal Railway | 10 | 8.33 |
| M3 | Municipal Railway | 100 | 8.37 |

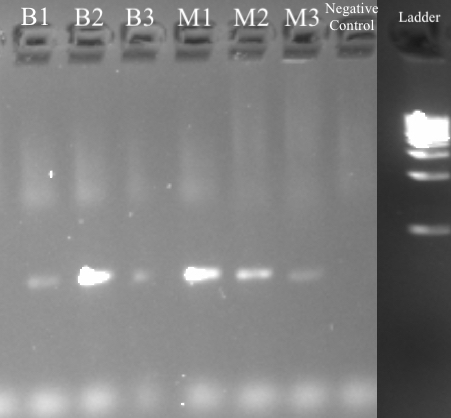
**Table 1:** Table of Qubit culture DNA concentration results. BART DNA concentrations signficantly vary, however dsDNA concentrations were sufficient for PCR.

|  |  |  |
| --- | --- | --- |
| Sample I.D. | Sample Group | DNA Concentration (ng/μl) |
| B1 | BART | 4.90 |
| B2 | BART | 8.48 |
| B3 | BART | 5.28 |
| M1 | Municipal Railway | 10.0 |
| M2 | Municipal Railway | 6.98 |
| M3 | Municipal Railway | 5.94 |
| Negative Control |  | 4.44 |

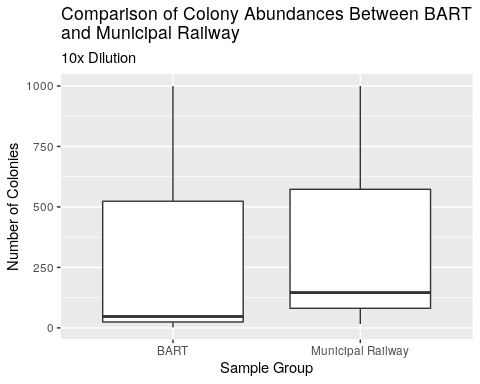
**Table 2:** Table of Qubit culture-free dsDNA concentration results measured in (ng/μl).



**Figure 1:** Image of culture PCRs from DNA extractions of most abundant and isolated colonies. The ladder indicates that the sequences are approximately 300bp.

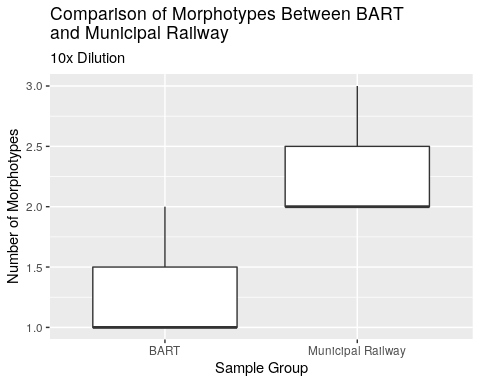


**Figure 2:** Image of culture-free PCRs from all 6 original DNA extractions. The ladder is located on the right, indicated that sequences are approximately 200bp.



**Figure 3:** Boxplot of colony abundances at different sites, 10x dilution. Despite a higher median number of colonies from the Municipal Railway samples, the mean values of colonies were not significantly different between the two sites (Wilcox p = 0.82)

|  |  |  |  |
| --- | --- | --- | --- |
| statistic | p.value | method | alternative |
| 3.5 | 0.8247781 | Wilcoxon rank sum test with continuity correction | two.sided |

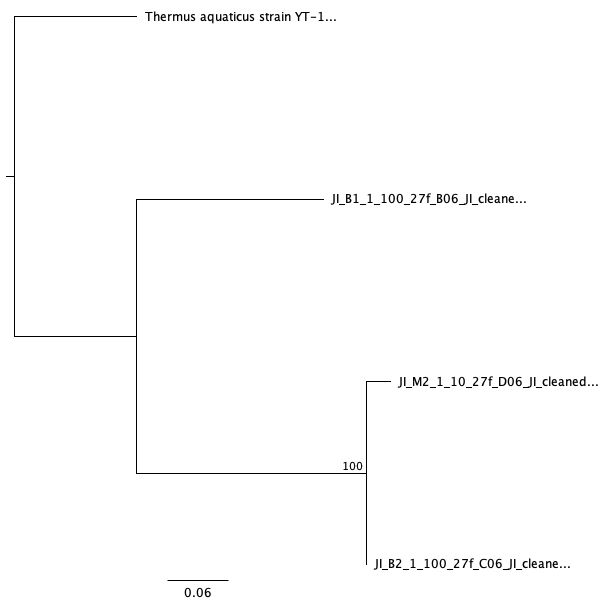


**Figure 4:** Boxplot showing the number of morphotypes from the two different sites. There was a difference in the mean number of morphotypes, W = 1 and p = 0.16.

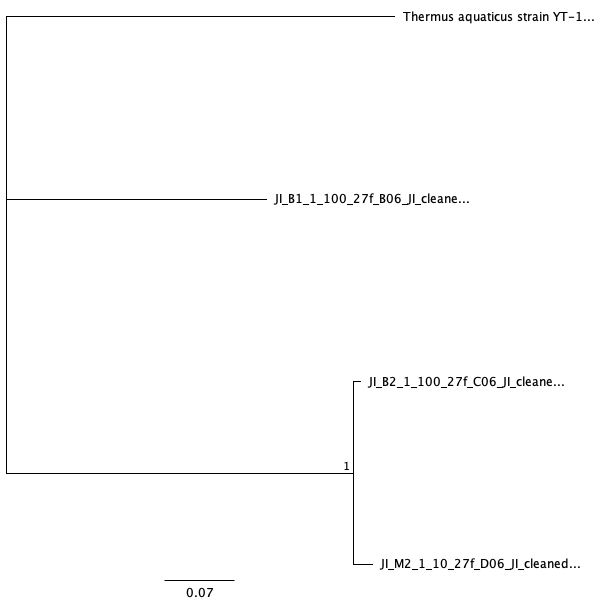
|  |  |  |  |
| --- | --- | --- | --- |
| statistic | p.value | method | alternative |
| 1 | 0.1572992 | Wilcoxon rank sum test with continuity correction | two.sided |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sequence Locus | Original Filename | Filename after Correction | Usability | Length After Trimming | Number of Manual Corrections |
| 16S | JI\_B1\_1\_10\_27f\_A06.ab1 | JI\_B1\_1\_10\_27f\_A06\_JI\_cleaned\_mixed.ab1 | Usable | 280 | 1 |
| 16S | JI\_B1\_1\_100\_27f\_B06.ab1 | JI\_B1\_1\_100\_27f\_B06\_JI\_cleaned.ab1 | Usable | 771 | 1 |
| 16S | JI\_B2\_1\_100\_27f\_C06.ab1 | JI\_B2\_1\_100\_27f\_C06\_JI\_cleaned.ab1 | Usable | 609 | 4 |
| 16S | JI\_M2\_1\_10\_27f\_D06.ab1 | JI\_M2\_1\_10\_27f\_D06\_JI\_cleaned.ab1 | Usable | 633 | 1 |
| 16S | JI\_M3\_1\_10\_27f\_E06.ab1 | JI\_M3\_1\_10\_27f\_E06\_JI\_failed.ab1 | Unusable | 5 | 0 |
| 16S | JI\_M3\_1\_100\_27f\_F06.ab1 | JI\_M3\_1\_100\_27f\_F06\_JI\_cleaned\_failed\_blast.ab1 | Usable | 824 | 3 |
| 16S | JI\_neg\_27f\_G06.ab1 | JI\_neg\_27f\_G06\_JI\_failed.ab1 | Unusable | 5 | 0 |

**Table 3:** Sequence analysis after cleaning and trimming using MAFFT in Geneious.



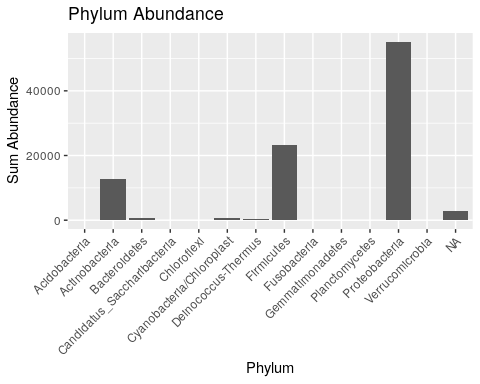
**Figure 5:** Maximum Likelihood Phylogeny of the three most successful samples with the root being *Thermus aquaticus*. M2 1:10 and B2 1:100 are more closely related than B1 1:100, with a bootstrap of 100.

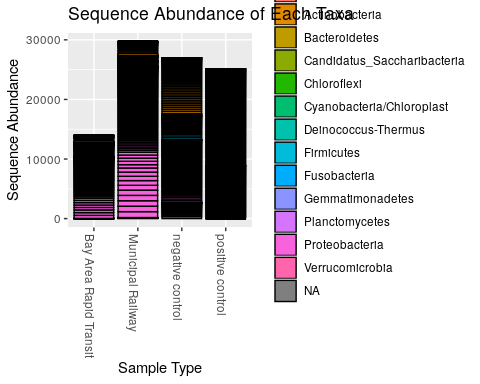


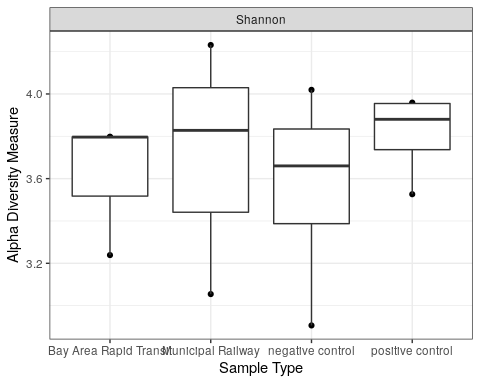
**Figure 6:** A Bayesian Phylogeny indicating that sample B2 1:100 and M2 1:10 are closely related with bootstrap 1 and Bayesian posterior probability of 0.07.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sample I.D. | Description | Percent Identity | Query Cover | E-Value | Accession |
| B1 1:10 | *Microbacterium phyllosphaerae* strain Y137 16S ribosomal RNA gene, partial sequence | 100% | 100% | 6e-143 | MN477985.1 |
| B1 1:100 | *Microbacterium hydrocarbonoxydans* strain T11.3 16S ribosomal RNA gene, partial sequence | 99.87% | 99% | 0.0 | MN198007.1 |
| B2 1:100 | *Acinetobacter schindleri* strain H3 chromosome, complete genome | 99.34% | 100% | 0.0 | CP030754.1 |
| M2 1:10 | *Acinetobacter sp.* strain YX3 16S ribosomal RNA gene, partial sequence | 99.84% | 99% | 0.0 | MK138620.1 |
| M3 1:10 | Sequence Failed |  |  |  |  |
| M3 1:100 | *Marinobacterium sp.* MB3 gene for 16S rRNA, partial sequence | 76.74% | 36% | 2e-38 | AB490785.1 |
| Negative Control | Sequence Failed |  |  |  |  |

**Table 4:** A table of BLAST results from Sanger Sequence samples. The table indicates which strains are present in the bacterial colonies.

 **Figure 7:** This figure is a summary chart of the culture-free BLAST sequence results showing each phylum and its sequence sum abundance. *Actinobacteria*, *Firmicutes*, and *Proteobacteria* are the most abundant phylum strains present in the samples. *Actinobacteria* is ~1300, *Firmicutes* is ~ 2300, and *Proteobacteria* ~5500.

 **Figure 8:** The bar chart in this figure depicts sequence abundance of each taxa found for each of the sample data groups: BART, Muni, negative control, and positive control. BART samples contained mostly *Proteobacteria* and *Verrucomicrobia*. Muni contained *Verrucomicrobia*. Negative control contained *Firmicutes* and *Actinobacteria* and the positive control does not contain any of the taxa, all indicated by the color on the bar chart.



## List of 1  
## $ axis.text.x:List of 11  
## ..$ family : NULL  
## ..$ face : NULL  
## ..$ colour : NULL  
## ..$ size : NULL  
## ..$ hjust : num 1  
## ..$ vjust : num 1  
## ..$ angle : num 45  
## ..$ lineheight : NULL  
## ..$ margin : NULL  
## ..$ debug : NULL  
## ..$ inherit.blank: logi FALSE  
## ..- attr(\*, "class")= chr [1:2] "element\_text" "element"  
## - attr(\*, "class")= chr [1:2] "theme" "gg"  
## - attr(\*, "complete")= logi FALSE  
## - attr(\*, "validate")= logi TRUE

**Figure 9:** The boxplot shows a Shannon diversity index of each sample type and the alpha diversity measure.

|  |  |  |
| --- | --- | --- |
| type | Phylum | sum\_abundance |
| Bay Area Rapid Transit | Proteobacteria | 10397 |
| Municipal Railway | Proteobacteria | 25424 |
| negative control | Proteobacteria | 10519 |
| positive control | Firmicutes | 16232 |

# Sources Cited

Aaij,C. and Borst,P. (1972) The gel electrophoresis of dna. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis*, **269**, 192–200.

Blazej,R.G. *et al.* (2006) Microfabricated bioprocessor for integrated nanoliter-scale sanger dna sequencing. *Proceedings of the National Academy of Sciences*, **103**, 7240–7245.

Carleton,H.A. *et al.* (2004) Community-adapted methicillin-resistant staphylococcus aureus (mrsa): Population dynamics of an expanding community reservoir of mrsa. *The Journal of infectious diseases*, **190**, 1730–1738.

Griffen,A.L. *et al.* (2011) CORE: A phylogenetically-curated 16S rDNA database of the core oral microbiome. *PloS one*, **6**, e19051.

Haines,A.M. and Linacre,A. (2019) Detection of latent dna on tape-lifts using fluorescent in situ detection. *Australian Journal of Forensic Sciences*, **51**, 455–465.

Hwang,S.H. and Park,J.B. (2014) Comparison of culturable airborne bacteria and related environmental factors at underground subway stations between 2006 and 2013. *Atmospheric environment*, **84**, 289–293.

Johnson,M. *et al.* (2008) NCBI blast: A better web interface. *Nucleic acids research*, **36**, W5–W9.

Katoh,K. *et al.* (2009) Multiple alignment of dna sequences with mafft. In, *Bioinformatics for dna sequence analysis*. Springer, pp. 39–64.

Kuno,G. (1998) Universal diagnostic rt-pcr protocol for arboviruses. *Journal of virological methods*, **72**, 27–41.

Meyer,M. and Kircher,M. (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, **2010**, pdb–prot5448.

Monger,C. *et al.* (2017) A bioinformatics pipeline for the identification of cho cell differential gene expression from rna-seq data. In, *Heterologous protein production in cho cells*. Springer, pp. 169–186.

Onat,B. *et al.* (2017) The relationship between particle and culturable airborne bacteria concentrations in public transportation. *Indoor and Built Environment*, **26**, 1420–1428.

Otter,J. and French,G. (2009) Bacterial contamination on touch surfaces in the public transport system and in public areas of a hospital in london. *Letters in applied microbiology*, **49**, 803–805.

Redman,L. *et al.* (2013) Quality attributes of public transport that attract car users: A research review. *Transport policy*, **25**, 119–127.

Reynolds,K.A. *et al.* (2005) Occurrence of bacteria and biochemical markers on public surfaces. *International journal of environmental health research*, **15**, 225–234.

Stepanović,S. *et al.* (2008) Public transport as a reservoir of methicillin-resistant staphylococci. *Letters in applied microbiology*, **47**, 339–341.

Takeuchi,M. and Hatano,K. (1998) Union of the genera microbacterium orla-jensen and aureobacterium collins et al. In a redefined genus microbacterium. *International Journal of Systematic and Evolutionary Microbiology*, **48**, 739–747.

Weber,S. and Douglas,D. Extract-n-amp™ tissue pcr kit: Rapid genomic dna extraction from tissue coupled with pcr.

Yeh,P.J. *et al.* (2011) A diversity of antibiotic-resistant staphylococcus spp. In a public transportation system. *Osong public health and research perspectives*, **2**, 202–209.