

POSTMORTEM MICROBIOME COMPUTATIONAL METHODS AND APPLICATIONS

By

Sierra Frances Kaszubinski

A THESIS

Submitted to
Michigan State University
in partial fulfilment of the requirements
for the degree of

Integrative Biology—Master of Science

2020

ABSTRACT

POSTMORTEM MICROBIOME COMPUTATIONAL METHODS AND APPLICATIONS

By

Sierra Frances Kaszubinski

Microbial communities have potential evidential utility for forensic applications. However, bioinformatic analysis of high-throughput sequencing data varies widely among laboratories and can potentially affect downstream forensic analyses and data interpretations. To illustrate the importance of standardizing methodology, we compared analyses of postmortem microbiome samples using several bioinformatic pipelines, while varying minimum library size or the minimum number of sequences per sample, and sample size. Using the same input sequence data, we found that pipeline significantly affected the microbial communities. Increasing minimum library size and sample size increased the number of low abundant and infrequent taxa detected. Our results show that bioinformatic pipeline and parameter choice significantly affect the resulting microbial communities, which is important for forensic applications. One such forensic application is the potential postmortem reflection of manner of death (MOD) and cause of death (COD). Microbial community metrics have linked the postmortem microbiome with antemortem health status. To further explore this association, we demonstrated that postmortem microbiomes could differentiate beta-dispersion among M/COD, especially for cardiovascular disease and drug-related deaths. Beta-dispersion associated with M/COD has potential forensic utility to aid certifiers of death by providing additional evidence for death determination. Additional supplemental files including tables of raw data and additional statistical tests are available in supplemental files online, denoted in the text as table 'S'.

This thesis is dedicated to my family- my biggest supporters and inspirations. Thank you for believing in me, and not letting me drop out to be a police officer. It was for the best.

ACKNOWLEDGMENTS

I first would like to thank my committee members, Drs. Mariah H. Meek, M. Eric Benbow, Jennifer L. Pechal, and Sarah E. Evans for their contribution to my scientific growth. Dr. Meek, I am forever grateful for the day I became your ‘bonus’ student, thank you for your endless support. Dr. Benbow, I’m glad I finally realized all my favorite forensic studies were coming from one floor below me, thank you for all the adventures. Dr. Pechal, thank you for all your help, you truly shaped me into a scientist I am proud to be. Dr. Evans, I appreciate your comments that made instrumental improvements to my studies, thank you for your support.

I would also like to thank Drs. Carl J. Schmidt and Heather R. Jordan for their contributions to my research. I truly enjoyed our forensic conversations. Thank you to Drs. Ruth Smith and Mary Finn, you are a large part of the reason this thesis is completed. I also want to thank the Meek lab: Charlene, Miranda, Nadya, Sara, and Shannon. Thank you for all the coffee runs and conversations. To the Benbow lab (past and present): Breanna, Courtney L., Courtney W., Emily, Joe, Katelyn, and Nick. I enjoyed every part of fieldwork, most of lab work, and some of coding with you. Thanks for making my time in grad school so much more enjoyable.

Thank you to the SMART scholarship program, and the Defense Forensic Science Center. Because of you, I have my dream job. To Thomas Meyer and Henry Maynard, thank you for your mentorship and great discussions. Thank you to the Department of Integrative Biology, despite my unusual admittance, I felt at home in the program. Finally, thank you to my friends and family. Thank you for believing in me, even when I didn’t believe in myself. Thank you for listening to my endless complaints about Michigan weather, and being there for me through every defeat and every triumph. I literally could not have done this without you, your support means everything to me. This one’s for you.

TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	vii
CHAPTER I: EVALUATING BIOINFORMATIC PIPELINE PERFORMANCE FOR FORENSIC MICROBIOME ANALYSIS	1
Introduction.....	2
Material and Methods	8
Sample Collection, DNA Extraction, and Sequencing	9
Pipeline Comparison.....	10
Data Analysis and Bioinformatics	13
<i>A Priori</i> Power Analysis	16
Data Availability.....	16
Results.....	17
Pipeline Comparison.....	18
Minimum Library Size.....	23
Sample Size.....	26
Discussion	29
Pipeline Comparison.....	30
Minimum Library Size.....	31
Sample Size.....	33
Conclusion	35
CHAPTER II: DYSBIOSIS IN THE DEAD: HUMAN POSTMORTEM MICROBIOME BETA-DISPERSION AS ONE INDICATOR OF MANNER AND CAUSE OF DEATH DURING AUTOPSY	37
Introduction.....	38
Materials and Methods.....	43
Sample Collection, DNA Extraction, and Sequencing	44
Data Analysis and Bioinformatics	44
Method Selection	45
Model Selection	46
Case Studies	48
Data Availability.....	49
Results.....	50
Method Selection	51
Model Selection	52
Case Studies	59
Discussion	62
Conclusion	68
APPENDIX.....	70
REFERENCES	75

LIST OF TABLES

Table 1: Summary of parameter differences among pipelines compared.....	13
Table 2: Summary of sample number and sequence read differences among pipelines.....	20
Table 3: Random forest classification error among pipelines.	20
Table 4: Total number of sequences that remained after filtering, and unclassified sequences (phylum and family taxonomic level) among sample area, sample size, and minimum library size.	25
Table 5: Random forest classification error among sample sizes and minimum library sizes.	28
Table 6: Case metadata stratified by manner/cause of Death (M/COD).	54
Table 7: Summary of logistic regression models classifying natural vs. accident, cardiovascular vs. drug-use, and disease vs. non-diseased state.....	58

LIST OF FIGURES

Figure 1: Workflow for bioinformatic analysis.	12
Figure 2: <i>In silico</i> data error among pipelines.	19
Figure 3: Microbial community composition among pipelines.	22
Figure 4: Microbial community composition among sample sizes and minimum library sizes.	24
Figure 5: Core microbiome among sample sizes and minimum library sizes.	27
Figure 6: Beta-dispersion values among body sites and manners/causes of death (M/CODs).	53
Figure 7: Multinomial logistic regression model comparison among full community and random forest indicator community beta-dispersion for manners/causes of death (M/CODs).	57
Figure 8: Logistic regression models of the best performing pair-wise comparisons.	59
Figure 9: Proposed workflow with suicide matched-design case study.	61
Supplemental Figure 1: Summary of Kruskal-Wallis results for normalization strategies determining beta-dispersion.	71
Supplemental Figure 2: Summary of Fligner-Killeen results for normalization strategies determining beta-dispersion.	72
Supplemental Figure 3: Alpha-diversity metrics across normalization strategies and levels.	73
Supplemental Figure 4: Shannon diversity across sequencing depth among microbial samples.	74

CHAPTER I:
EVALUATING BIOINFORMATIC PIPELINE PERFORMANCE FOR FORENSIC
MICROBIOME ANALYSIS

Introduction

Before the widespread use of next-generation sequencing (NGS), forensic microbiology was limited to identifying pathogens of bio-crimes with culture-based methods, such as the 2001 anthrax letter attacks (Budowle et al. 2003). Now with the advent of NGS technology, amplicon sequencing can describe entire microbial communities from evidence rather than just targeting a single microbe of interest. Microbes are an important forensic resource as they are ubiquitous organisms, with community compositions specific to different environments or hosts (i.e., the location of a body or body part) and that vary over time, such as during decomposition (Pechal et al. 2018; Metcalf et al. 2016).

NGS has expanded microbial forensics to many potential applications including: body fluid identification, human identification, and postmortem interval estimation (Schmedes, Sajantila, and Budowle 2016). Targeted amplicon sequencing of 16S rRNA identified potential microbial biomarkers for sensitive body fluid identification (Dobay et al. 2019), and a clade based, single nucleotide polymorphism approach identified human individuals using their skin microbiome (Schmedes, Woerner, and Budowle 2017). Postmortem microbiome studies have included a diverse set of investigative circumstances to better understand how microbial communities after death can inform forensic investigation. Studies have used human-surrogates, such as swine (Benbow et al. 2015), humans (Pechal et al. 2018), and grave soil (Metcalf et al. 2016), as well as experiments in anthropologic facilities (Pechal et al. 2014) and during routine autopsy for death investigation (Pechal et al. 2018). Researchers have developed models for forensic applications (i.e., postmortem interval estimation), and described microbial community function and succession during decomposition (Benbow et al. 2015; Metcalf et al. 2016; Pechal et al. 2018). While recent research suggests excellent potential for microbial community use in

forensics (Metcalf et al. 2013, 2017; Carter et al. 2016; Johnson et al. 2016), additional foundational work is needed before forensic microbiology using postmortem microbial community assemblages can be applied in the criminal justice system. Needed foundational work includes standardizing parameters for transforming raw microbial sequences into a usable data format for downstream analyses.

For bioinformatic analysis of microbial data, raw sequence data files generated from NGS undergo a series of transformations using executable command line software known as pipelines (Leipzig 2017). Forensic microbiological data created by high-throughput sequencing platforms are processed using complex analyses that require users to make processing decisions along these pipelines (e.g., Should samples be normalized? Which taxonomic database should be used?). It is hypothesized that different decisions could have downstream effects on results and their interpretation (Sivarajah et al. 2017; Golob et al. 2017). For any downstream application of forensic microbiology in the criminal justice system, there is a need for streamlined standard operating procedures (SOPs) (Carter et al. 2016).

There are self-contained pipelines for processing sequence data for characterizing microbial communities. Three of the most often used self-contained pipelines are: QIIME2 (Bolyen et al. 2019), mothur (Schloss et al. 2009), and MG-RAST (Keegan, Glass, and Meyer 2016). QIIME2 and mothur require some command line experience, whereas MG-RAST has a web-based graphical user interface (GUI) (Plummer and Twin 2015). Several studies determined that the microbial communities generated from different pipelines were comparable; however, these studies used simulated data (Golob et al. 2017; Siegwald et al. 2017; Mysara et al. 2017; Nilakanta et al. 2014; McMurdie and Holmes 2014; Weiss et al. 2017), small sample sizes ($n < 40$)

(Plummer and Twin 2015; McMurdie and Holmes 2014; Weiss et al. 2017; D'Argenio et al. 2014) or were composed of the same sample type (i.e., human gut microbial data) (Plummer and Twin 2015; D'Argenio et al. 2014), which does not readily apply to many forensic datasets that can include highly variable sampling locations or contextual circumstances (Benbow et al. 2015; Metcalf et al. 2016; Pechal et al. 2018).

In addition to choosing the pipeline, microbial sequence data analyses can be confounded by different library sizes, or the number of sequence reads per sample (McMurdie and Holmes 2014). The common library size normalization procedure for microbial analysis is referred to as rarefying (McMurdie and Holmes 2014). To rarefy data, a minimum library size is chosen, samples with too few reads are discarded, and the remaining sample reads are subsampled without replacement to the minimum library size (McMurdie and Holmes 2014). Minimum library size is often chosen by selecting the smallest library size of a non-defective sample, which is a subjective assessment that can add uncertainty to microbial community analysis (McMurdie and Holmes 2014). Minimum library size can also be chosen based on rarefaction curves: taxon-based re-sampling curves that indicate species richness and coverage analysis to justify a certain library size. Potentially useful data are omitted during rarefying, which can decrease the power and specificity of analyses as samples are discarded, and the samples that remain may not be distinguishable using a fraction of the original reads (McMurdie and Holmes 2014). While rarefying can decrease power, sample size is an additional factor to consider for downstream analysis of microbial data.

Most postmortem microbiome studies have small sample sizes (Pechal et al. 2018). Often, this is due to space, time or resource limitations of using human donated bodies or surrogate carcasses, such as swine. Prior to 2018, when Pechal et al. (2018) characterized and

modeled postmortem microbiomes using 188 autopsies, sample sizes for postmortem microbiome studies range from 2 to 48 (Metcalf et al. 2016; Hyde et al. 2013). Small sample sizes can decrease statistical power and inflate effect sizes, which can lead to unreliable conclusions that may not be generalizable to other data sets (Ioannidis 2005), which can have important consequences in forensics (Clarke et al. 2017). Given this, it is important to evaluate data analysis parameters with large, spatially and temporally heterogeneous datasets with multiple sample areas (i.e., body locations) in order to improve the reliability of postmortem microbiome data.

Downstream analytical methods for forensic microbiology studies are still being evaluated. Recently, machine learning algorithms and parameterization were comprehensively tested for potential direct forensic applications to postmortem interval estimation, manner of death determination, and location of death (i.e. inside, outside, hospital death) for this particular dataset (Zhang et al. 2019). Therefore, we focus on preceding steps of microbial analysis before modeling applications for comparing forensic predictions. The overall goal was to provide an initial assessment of how pipeline choice and data processing parameter differences affect data outcomes that are used as inputs for downstream modeling and prediction, with the intention that forensic researchers and examiners could make informed decisions about study design, data analysis methods, and applications relevant to their forensics needs. A better understanding of performance among bioinformatic pipelines and parameters is needed to reveal potentially significant differences in downstream analysis and data interpretation, especially for future use in the forensic sciences. To determine how different microbiome data processing parameters affect downstream analytical outcomes, we determined how pipeline software, library size normalization by rarefying, and sample size affect common microbiome community metrics and

machine learning model results using a large postmortem microbiome dataset (n=188). We have emphasized using standardized, recognizable, and user- friendly methods encountered in open-access microbiome analyses tools for forensic applications.

Materials and Methods

Sample Collection, DNA Extraction, and Sequencing

Postmortem microbiome data were obtained from our previously published study that characterized the microbial communities from multiple body sites in 188 routine autopsy cases (Pechal et al. 2018). The postmortem microbiome data for this study represented the microbial communities from the mouth and rectum, with cases that spanned all four seasons (Spring, Summer, Fall, and Winter), manners of death (accident, homicide, suicide, natural), postmortem intervals (< 24h - > 73+ h), and ages (18-55+ years) (Table S1). Detailed methods for sample collection, DNA extraction, and sequencing are available in Pechal et al. (2018). In short, trained personnel at Wayne County Medical Examiner's Office in Detroit, Michigan took swab samples during routine autopsy using DNA-Free sterile cotton-tipped applicators (25–806 1WC FDNA, Puritan®, Guilford, MA, USA). Each swab was rotated for 3-5 seconds on the body location to sample the microbial community. Samples were placed in sterile microfuge tubes and 200 µl of 100% molecular grade ethanol (BP2818-4, Fisher Scientific, Waltham, MA, USA) and stored at -20°C. Metadata were collected for each case including: sampling date (season), anatomic region sampled, sex, ethnicity, estimated age (years), postmortem interval (PMI), body mass index (BMI; kg/m²), event location (indoors, outdoors, hospital, vehicle), and manner of death (Table S1). Manner of death and PMI estimates were determined by a board-certified forensic pathologist at the time of autopsy.

To determine the microbial communities, total DNA was extracted in a biological safety cabinet with aseptic conditions using PureLink® Genomic DNA Mini Kit (Thermo Fisher Scientific, Waltham, MA, USA) following manufacture instructions except an additional 5 ng/µl of lysozyme was added during the lysis step for each sample reaction (Pechal et al. 2017). The samples were quantified by Qubit 2.0 and the Quant-iT dsDNA HS Assay kit (Thermo Fisher

Scientific, Waltham, MA, USA). Microbial DNA was sequenced at the MSU Genomics Core Facility (East Lansing, MI, USA) using an Illumina MiSeq. The MSU Genomics Core Facility prepared the 16S rRNA gene amplicon library and sequenced the samples using a modified version of the protocol adapted for the Illumina HiSeq2000 and MiSeq. The V4 region 16S gene amplicon 2 x 250 bp paired-end reads were generated with region specific primers [515 f (5' GTGCCAGCMGCCGCGGTAA) and 806 r (5' GGACTACHVGGGTWTCTAAT)] that included Illumina flowcell adapters (Caporaso et al. 2010). PCR products were normalized and pooled with Invitrogen SequalPrep DNA Normalization Plates. A combination of Qubit dsDNA HS, Caliper LabChipGX HS DNA, and Kapa Illumina Library Quantification qPCR assays were used to quantify the pooled library. Amplicons were sequenced with custom primers complementing amplification primers to avoid primer sequencing after cluster formation as described by Kozich et al. (Kozich et al. 2013). Pooled sequences were loaded on an Illumina MiSeq standard flow cell (v2) and sequenced using a 500 cycle reagent cartridge. Filtering parameters were optimized for detecting low abundance phylogenetic diversity (Caporaso, Knight, and Kelley 2011; Caporaso et al. 2012). Bases were called by the Illumina Real Time Analysis (RTA) v1.18.54, and the output was demultiplexed and converted to FastQ format by Illumina Bcl2fastq v1.8.4.

Pipeline Comparison

Sequence reads from postmortem microbiome samples were analyzed with mothur v1.39.5 (Schloss et al. 2009), QIIME2 v2018.11 (Bolyen et al. 2019), and MG-RAST v4.0.3 (Keegan, Glass, and Meyer 2016) to determine how the microbial community composition and

diversity metrics (alpha- and beta-diversity) varied among pipelines. The SOP for mothur (Kozich et al. 2013), the QIIME 2 moving pictures tutorial (Caporaso et al. 2011), and the MG-RAST manual (Glass et al. 2010) were used for reference to analyze the samples. For mothur and QIIME2, the SILVA small subunit database v132 (Quast et al. 2013) was used at a 99% similarity cutoff for taxonomic classification. The database version of SILVA used by MG-RAST is unknown, as it is not reported by the authors or on the website, even after multiple inquiries. An overview of each pipeline workflow, including commands used to run the mothur and QIIME2, are represented in Figure 1. Some steps were conserved among pipelines (i.e., quality control and classify sequences/assign taxonomy) but the number and order of steps occurred in different succession among pipelines depending on how the pipeline developers created the software (Figure 1). For QIIME2, DADA2 v1.8.0 (Callahan et al. 2016) corrected Illumina amplicon sequencing data, including removing phiX reads. Sequence were aligned using MAFFT v7.397 (Kato and Standley 2013) and FastTree v2.1 (Price, Dehal, and Arkin 2010) created the phylogenetic tree. Biom tables were created using biom-format package v2.1 (McDonald et al. 2012) and exported. For mothur, OTUs (Operational Taxonomic Units) and taxonomy tables were exported as column separated values (csv) files. For MG-RAST, identified sequences were clustered using UCLUST v6 (Edgar 2010). OTUs and taxonomy tables were exported as tab separated values (tsv) files. As many parameters were standardized as possible, but analysis methods among pipelines differed (Table 1). For example, all pipelines used the SILVA database for taxonomic classification, and VSEARCH v2.8.0 (Rognes et al. 2016) for chimera detection and removal, while the taxonomic classification algorithms and alignment methods differed (Table 1).

To calculate error rate for each pipeline, *in silico* sequences were taken from “mockrobiota” (Bokulich et al. 2016), an online repository of sequences used to assess pipeline error rate. The mock-3 16S rRNA dataset contained sequence data, corresponding taxonomy, and relative abundance of the OTUs. Using pipeline workflows mentioned above, four samples were run through each of the pipelines: two samples representative of an “even” (all taxa have the same relative abundance) community composition and two “staggered” (taxa have varied relative abundances) community composition samples.

Figure 1: **Workflow for bioinformatic analysis.** Several of the steps were shared among pipelines including quality control and aligning sequences. Commands from the code are in square brackets, while parameters are in parentheses. OTU: Operational taxonomic unit.

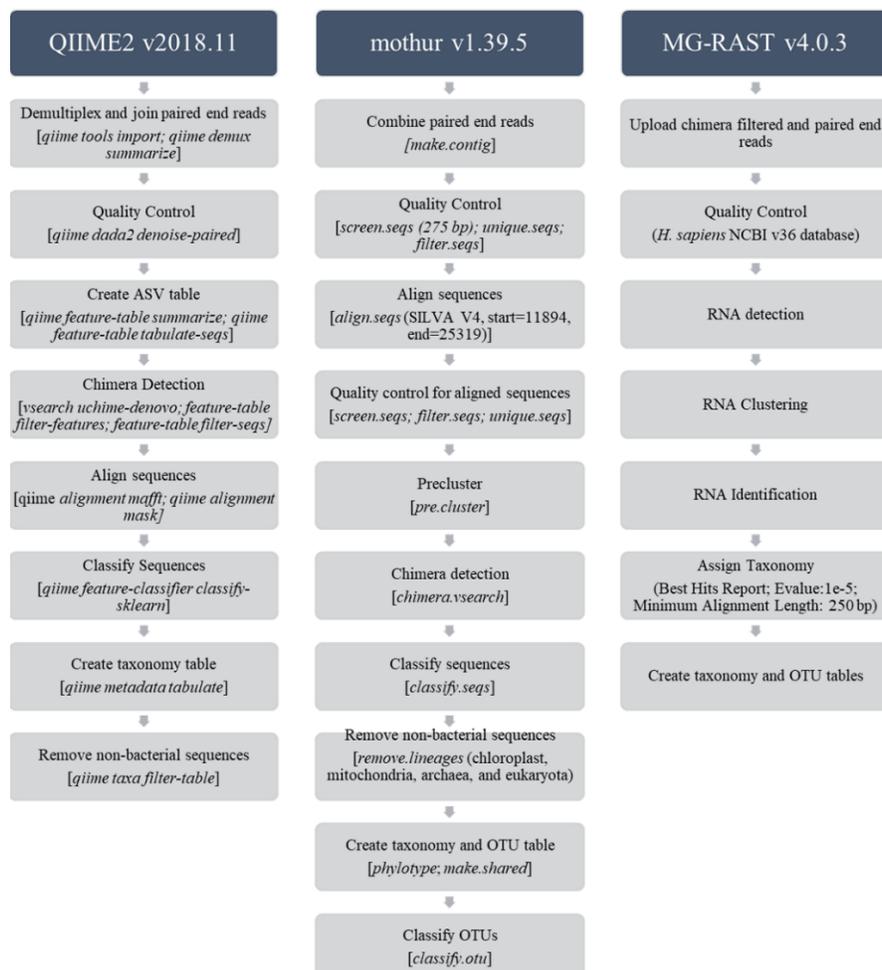


Table 1: **Summary of parameter differences among pipelines compared.** Adapted from Plummer et al. (2015). Bolded text were parameters chosen for running the pipelines.

	QIIME2	mothur	MG-RAST
License	Open-Source	Open-Source	Open-Source
Language	Python	C++	Perl
Current Version	2019.1	1.39.5	4.0.3
Cited (Google Scholar)	18,800	14,300	4,620
Web Based Interface	GUI, API, CLI	API, CLI	GUI , API
Quality Control	YES	YES	YES
16S rRNA Database	SILVA , Greengenes, UNITE	SILVA , Greengenes, RDP	SILVA , Greengenes, RDP, ITS
Alignment Method	mafft	Needleman-Wunsch	BLAT
Taxonomic Assignment	Naive Bayes classifier	Wang	BLAT
Chimera Detection	VSEARCH	VSEARCH	VSEARCH
User Support	Forum, tutorials, FAQs	Forum, SOPs, FAQs, user manual	Video tutorials, FAQs, user manual, 'How to' section on website

Data Analysis and Bioinformatics

Pipeline outputs were quantitatively compared using Biom files from QIIME2, and OTUs and taxonomy files from mothur and MG-RAST were combined with metadata as phyloseq v1.24.0 (McMurdie and Holmes 2013) objects in R v3.5.1 (R Core Team) and rarefied to 1,000 sequences to account for the variability of library size among pipelines. Taxonomic names were corrected so that outputs could be properly merged with comparable taxa names (Table S2). Two methods were used to compare pipelines: statistical analyses of pipeline outputs from postmortem autopsy case microbiome data and pipeline error analysis using *in silico* data. These

comparison methods were chosen based on previous pipeline comparison research, which either compared pipeline output or *in silico* data (Plummer and Twin 2015; Sivarajah et al. 2017). Relative abundance was determined by combining all samples analyzed within each pipeline and sample area, removing taxa that were less than 1% abundance, and determining the proportional contribution of each taxa to the total community. Differentially abundant taxa among pipelines were determined from relative abundance at the phylum and family level using ANCOM (Mandal et al. 2015). Alpha-diversity metrics (observed richness, Chao1, Shannon diversity (1-D), Inverse Simpson diversity (1/D)) were calculated using phyloseq. Alpha-diversity metrics were compared using Kruskal-Wallis and Nemenyi *post hoc* tests with the R stats and PMCMR packages (Pohlert 2018). Beta-diversity metrics, evaluated using Principal Coordinates Analysis (PCoA) of Jaccard distances, were plotted using phyloseq. Jaccard was chosen as a presence/absence method to buffer against relative abundance differences found among parameters. PERMANOVA (permutational multivariate analysis of variance) from the vegan v2.5-2 (Oksanen et al. 2019) package confirmed beta-diversity and dispersion differences among pipelines. In addition to diversity metrics, a measure of divergence was assessed (Kullback-Leibler Divergence) (la Rosa et al. 2012) and found to be generally consistent with trends of diversity metrics (Table S3). Classifications were made of sample area and manner of death using random forest (randomForest package v4.6-14) (Breiman et al. 2018) among pipelines. Out-of-bag (OOB) error rates were reported. However, test-set validation (70% training sets, 30% test sets) was also tested and the error rates were within 2% of the OOB error rates. For a more extensive comparison of random forest methods on the larger dataset of these postmortem microbiome data, please see Zhang et al. (2019).

We also used “mockrobiota” (Bokulich et al. 2016) reference samples to compare the accuracy of the pipeline outputs. For the *in silico* data analysis, three metrics were assessed: correct taxa, false positives, and false negatives. Taxa that were present in both mockrobiota taxonomic reference dataset and the pipeline output was labeled ‘correct taxa’. Taxa present in only the mockrobiota taxonomic reference dataset were considered false negatives, while taxa present only in the pipeline outputs were considered false positives. Concordance with the mockrobiota dataset was assessed based on pipeline outputs of abundance (sequence number) regressed with the expected abundances of mockrobiota as the R^2 value. False negative taxa were indicated by negative abundance values, while false positive taxa fell along the X- axis with no expected abundance value.

Based on the results of comparisons (see Results: Pipeline Comparison), the QIIME2 pipeline was chosen for conducting sample size and library size comparisons. Subsamples of the original 188 cases were chosen at random [R; sample()] without replacement for 60 and 120 cases (see *A Priori* Power Analysis below). Rarefaction levels used for comparison were 1,000 sequences, 7,000 sequences, and no rarefaction. Rarefying to 7,000 sequences was based on the alpha-rarefaction curve generated in QIIME2, while rarefying to 1,000 sequences was based on a subjectively chosen minimum library size. Outputs among subsamples and rarefaction levels using statistical analyses were compared using methods described above, including relative abundance, ANCOM, alpha-diversity, and beta-diversity. Changes in core microbiome characterization and random forest accuracy were also evaluated. In this case, the core microbiome was defined as shared OTUs among defined groups (i.e., sample areas, minimum

library sizes, and sample sizes) (Shade and Handelsman 2012) and determined using log abundance vs. occupancy plots. Classifications were made of sample area and manner of death using random forest (randomForest package v4.6-14) (Breiman et al. 2018), and the model error rate was compared among subsamples and normalization methods.

A Priori Power Analysis

To determine statistically significant sample size, an *a priori* power analysis was completed using G*Power 3 v3.0.5 (Faul et al. 2007). Body locations (mouth and rectum) were compared using the mean and standard deviation of Faith's phylogenetic diversity from previous bioinformatic analysis (Pechal et al. 2018) calculated using R. An independent mean two tail t-test was used to determine sample size needed for significant power. For each body location (mouth and rectum), 27 samples were required for significant power ($\alpha = 0.05$; $1-\beta = 0.80$). Three random subsample sizes were chosen based on the number of cases to include that all have significant power: 60, 120, 188. Random subsamples were generated using R, sampling without replacement.

Data Availability

Postmortem microbiome samples collected, extracted, and targeted amplicon 16S sequenced from 188 cases in Pechal et al. (2018) were used as a large dataset for comparing pipelines, sample size, and minimum library size for downstream analyses. Sequence data are archived through the European Bioinformatics Institute European Nucleotide Archive (www.ebi.ac.uk/ena) under accession number: PRJEB22642. The microbial community analysis is available as R code on GitHub (<https://github.com/sierrakasz/postmortem-analysis>).

Results

Pipeline Comparison

Despite having an easy to implement web-based GUI, MG-RAST was not an appropriate tool for this forensic dataset. Although the concordance of MG-RAST ($R^2 = 0.24$) was comparable to QIIME2 ($R^2 = 0.27$) for the *in silico* dataset (Figure 2), the *in silico* dataset represented very low diversity (< 20 taxa present) and simplistic community structure (even and staggered), with a low sample size ($n = 4$) and was not an accurate approximation of forensic data. For postmortem data, MG-RAST had a much smaller effect size than mothur and QIIME2 due to the twofold reduction in samples (Table S4). MG-RAST had an average library size ~ 20-times smaller than QIIME2 and mothur, tenfold reduction in sequences after filtering, and the highest numbers of unclassified sequences at the family and phylum level (Table 2, Table S5), representing a loss of valuable and forensically relevant data. MG-RAST also had the highest error rates for the random forest modeling in all cases (Table 3; Table S6).

Compared to mothur and QIIME2, MG-RAST had more differentially abundant taxa and significantly reduced alpha- and beta-diversity metrics indicative of a very different microbial community structure, despite the input of the same data. Although the top five most abundant phyla were consistent among pipelines, MG-RAST's mean relative abundances at the phylum and family taxonomic level were reduced for both mouth and rectum sample areas (Figure 3A, Figure 3B, Table S7-S8). MG-RAST had the most differentially abundant taxa (taxa for which relative abundances significantly differ from each other based on statistical tests) at both the phylum (9) and family level (39), which were represented by about 20-times more sequences than mothur and QIIME (ANCOM; Table S9-S11). When considering all samples, MG-RAST produced around 50% lower alpha-diversity for all three metrics compared to mothur and QIIME2 (Kruskal-Wallis $p < 0.05$; *post hoc* Nemenyi $p < 0.05$; Table S12B; Table S13A-

Figure 2: *In silico* data error among pipelines. *In silico* raw sequence data were obtained from mockrobiota and processed through each pipeline using the same methods as the postmortem microbiome dataset. Four samples were processed: two ‘even’ and two ‘staggered’ samples representing different microbial community compositions. (A) Taxa output from each pipeline were compared to the taxonomic reference dataset available on mockrobiota. Taxa that were present in both the mockrobiota taxonomic reference dataset and the pipeline output was labeled ‘correct taxa.’ Taxa present in only the mockrobiota taxonomic reference dataset were considered false negatives, while taxa present only in the pipeline outputs were considered false positives. Samples within each pipeline are ordered along the x-axis as: even, even, staggered, staggered. (B-D) Abundance of taxa (based on sequence number) from pipeline output versus expected abundance of taxa based on the mockrobiota abundance reference dataset for each sample. Negative abundances were assigned to samples that were considered false negatives. The R^2 line represented the actual concordance. (B) MG-RAST (C) mothur (D) QIIME2.

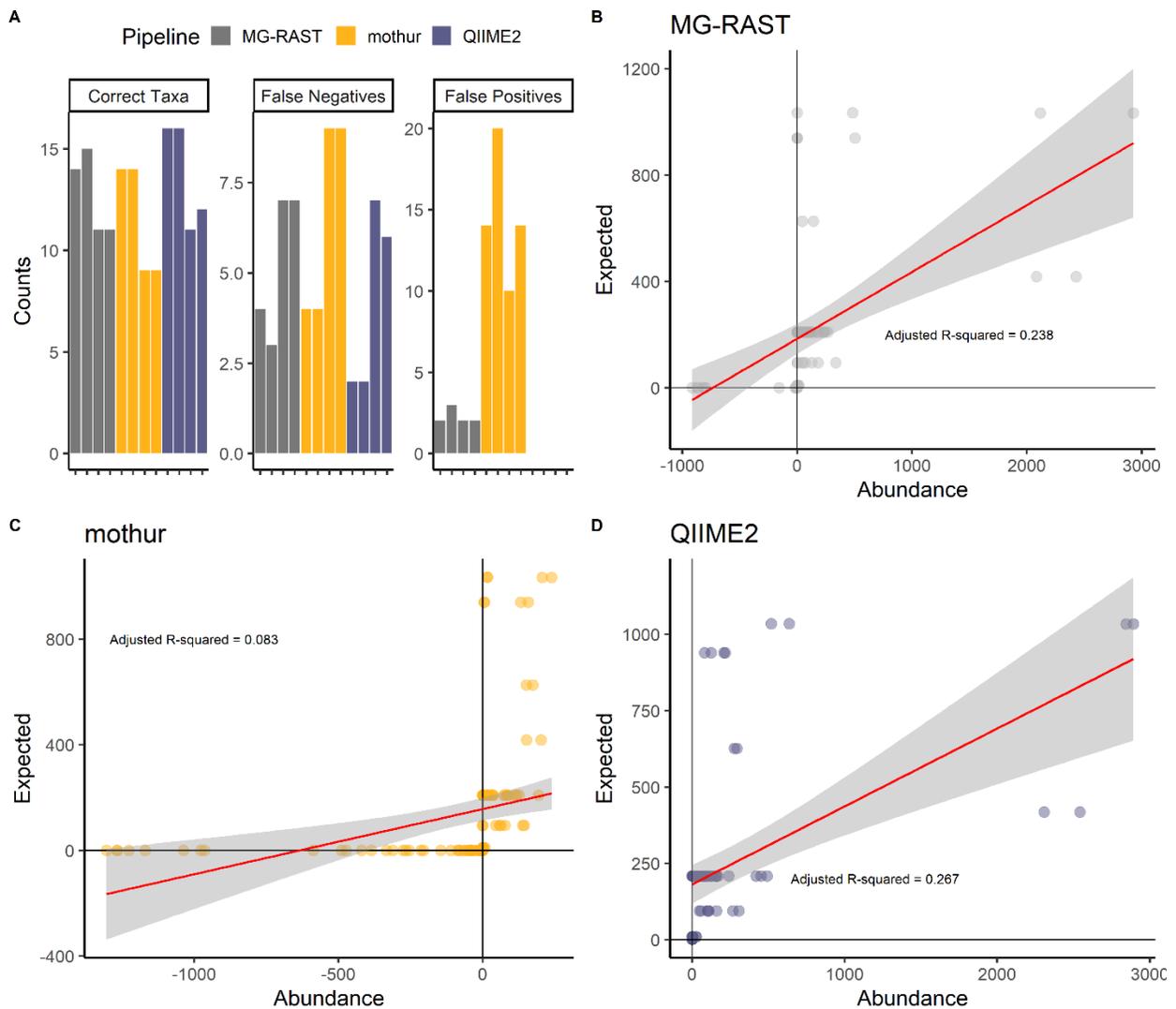


Table 2: **Summary of sample number and sequence read differences among pipelines.** The total number of samples among pipeline differed, as some samples were removed for not having the minimum required sequences (1,000), or in the case of MG-RAST, for not having the minimum 1,000,000 base pairs before filtering.

Pipeline	Total Number of Samples	Number of Mouth Samples	Number of Rectum Samples	Total Number of Reads after Filtering	Number of Unclassified Phylum Reads	Number of Unclassified Family Reads
QIIME2	324	169	155	11,375,659	2,399	2,832
mothur	324	169	155	12,861,356	594	2,419
MG-RAST	150	97	53	2,167,164	40,530	43,576

Table 3: **Random forest classification error among pipelines.** Random forest classifications were made with 1,000 trees, and out-of-bag (OOB) error was reported. Classifications of sample area (mouth and rectum) and manner of death (suicide, homicide, accident, natural) were made for phylum and family taxonomic level. Percentages and number of misclassifications from the total number of samples were included.

Taxonomic Level	Pipeline	Sample Area Error Rate (Misclassifications/ Total)	Manner of Death Error Rate (Misclassifications/ Total)
Phylum	QIIME2	16.05% (52/324)	58.64% (190/324)
Phylum	mothur	12.65% (41/324)	56.79% (184/324)
Phylum	MG-RAST	16.67% (25/150)	72% (108/150)
Family	QIIME2	4.94% (16/324)	57.72% (187/324)
Family	mothur	4.01% (13/324)	57.41% (186/324)
Family	MG-RAST	5.33% (8/150)	60% (90/150)

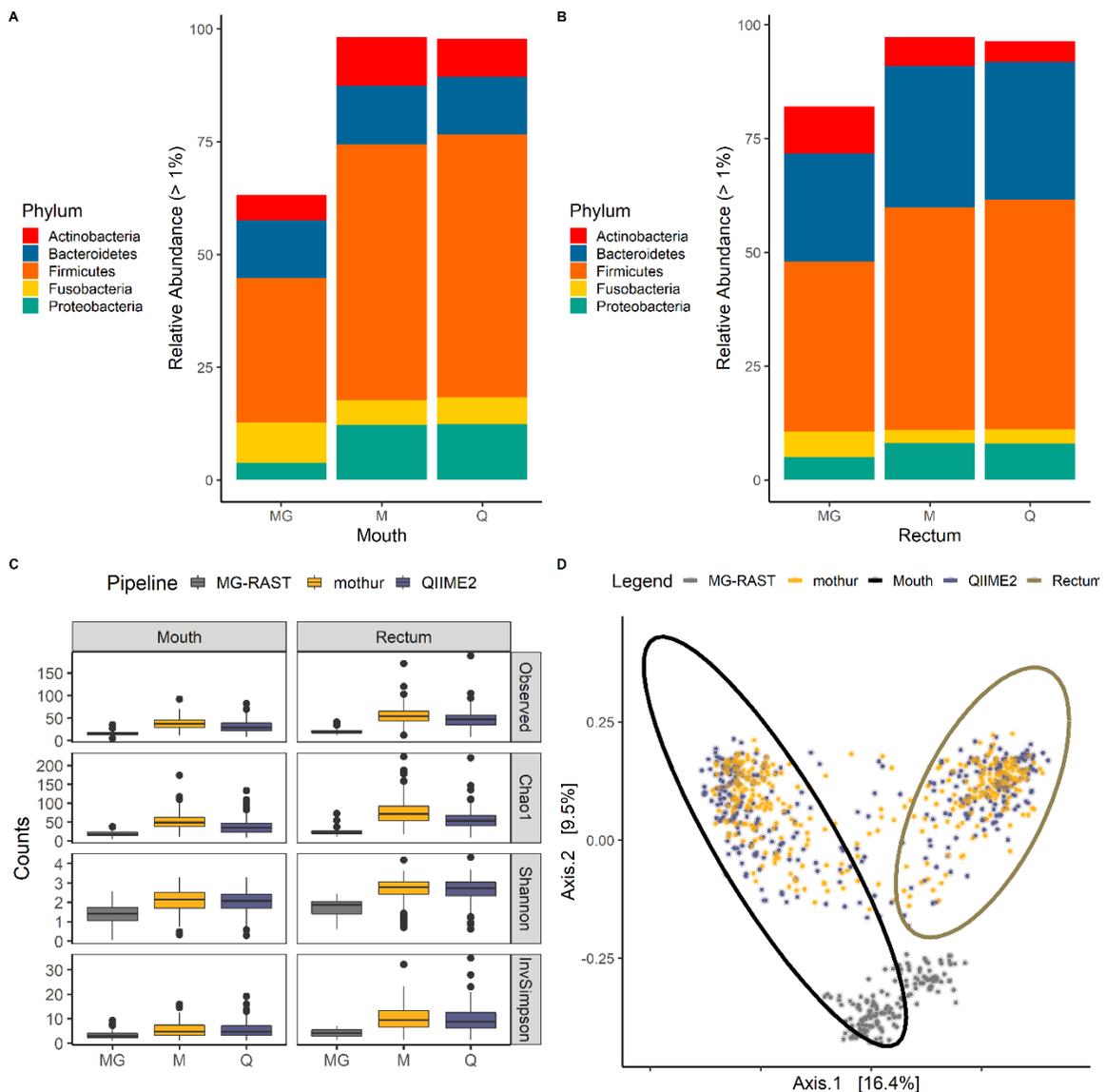
13B). MG-RAST samples clustered closely together (Figure 3D), reflective of MG-RAST's significantly lower dissimilarity (0.850 ± 0.139) than QIIME2 and mothur (PERMANOVA, $p < 0.05$, Table S14). The loss of sequences and samples during the filtering process was also

reflected in the reduced microbial community metrics downstream for MG-RAST.

QIIME2 and mothur had more comparable results but key differences among the pipeline outputs made QIIME2 the most appropriate pipeline to use for downstream analysis. Richness and diversity (evenness and richness) metrics from mothur and QIIME2 quantified similar levels of diversity (Figure 3C, Table S12A-12B). However, QIIME2 had around 20% lower observed richness than mothur (Table S12B, Table S13B) in contrast to previous studies (Plummer and Twin 2015). While QIIME2 had slightly higher levels of unclassified sequences (Table 2), mothur had a higher number of unclassified taxa at the phylum (3) and family level (4) (ANCOM, Table S9-S10), which artificially inflated richness measured by mothur. For random forest classification, mothur had slightly lower error rates than QIIME2 (Table 3), indicator taxa differed among mothur and QIIME2 possibly due to the inflated richness from mothur (Table S15). Despite mothur and QIIME2 samples' admixture (Figure 3D) and comparable dissimilarity (QIIME2: 0.874 ± 0.146 ; mothur: 0.872 ± 0.144), microbial community structure was distinguishable (PERMANOVA, $p < 0.05$, Table S14). However, the effect size was small ($R^2 < 0.1$, Table S14).

Despite the similar microbial communities, mothur and QIIME2 diverged when testing the *in silico* dataset. QIIME2 had no false positives, while mothur had the most false positives (range: 10-20) of all pipelines (Figure 2A, Table S16). Many of mothur's false positives were unclassified taxa at multiple taxonomic levels, which disrupted the relative abundances of the whole microbial community reflected in concordance. QIIME2 had the highest concordance to the mockrobiota taxonomy reference dataset ($R^2 = 0.27$), while mothur had the lowest ($R^2 = 0.08$) (Figure 2C-2D).

Figure 3: **Microbial community composition among pipelines.** (A) Relative abundances of the five most predominate phyla for mouth samples among pipelines (MG: MG-RAST, M: mothur, Q: QIIME2). (B) Relative abundances of the five most predominate phyla for rectum samples among pipelines (MG: MG-RAST, M: mothur, Q: QIIME2). (C) Alpha-diversity metrics for each sample area (mouth and rectum) and pipeline including: observed richness, Chao1, Shannon diversity, and Inverse Simpson diversity (InvSimpson). Kruskal-Wallis and *post hoc* Nemenyi tests for alpha-diversity metrics detected significant differences ($p < 0.05$) among pipelines and sample areas. (D) The principal coordinate analysis (PCoA) of Jaccard distances among sample area and pipeline. Ellipses indicated sample area, for a 95% confidence interval. Permutational multivariate analysis of variance (PERMANOVA) detected significant differences ($p < 0.05$) among pipelines for each sample area, and all pairwise differences were statistically significant ($p < 0.05$).



Based on the pipeline comparison results above, QIIME2 was chosen as the best pipeline to examine the effect of rarefying and sample sizes on modeling results due to the lowest error rate and reduced unclassified taxa compared to mothur, and therefore more accurate microbial community structure.

Minimum Library Size

Key differences in microbial community metrics were among minimum library sizes, indicating the importance of considering minimum library size during forensic microbiome analysis. The number of unclassified sequences at phylum and family level decreased proportionally as the minimum library size decreased from no rarefaction, 7,000 sequences, to 1,000 sequences for each sample area (Table 4). Relative abundances of the top five phyla changed among minimum library size, as smaller minimum library sizes (1,000 vs. 7,000) showed low abundance taxa (not the top five) phyla as a larger component of the relative abundances due to the random subsampling to a specified library size (Figure 4A, Table S17). The differences in microbial community structure among minimum library sizes were evident in the very distinct clustering of minimum library sizes (Figure 4F). Minimum library size had large effects beta-diversity (PERMANOVA, $p < 0.05$, $R^2 = 0.297-0.426$, Table S18). Due to the effects of minimum library size on microbial community structure, including beta-diversity and relative abundance, minimum library size should be considered for downstream analyses.

However, a large majority of taxa were captured within each minimum library size, as indicated by alpha-diversity and core microbiome analysis. Only the minimum library size of 1,000 had around 25% reduced richness compared to 7,000 and no rarefaction library sizes.

Figure 4: **Microbial community composition among sample sizes and minimum library sizes.** (A) Relative abundances of the five most predominate phyla among subsamples with no rarefaction. (B) Relative abundances of the five most predominate phyla among minimum library sizes of subsample 188. (C) Alpha-diversity metrics for each sample size with no rarefaction including: observed richness, Chao1, Shannon diversity, and Inverse Simpson diversity (InvSimpson). Kruskal-Wallis and *post hoc* Nemenyi tests for alpha-diversity metrics did not detect significant differences ($p < 0.05$) among subsamples. (D) Alpha-diversity metrics for each minimum library size of subsample 188 including: observed richness, Chao1, Shannon diversity, and Inverse Simpson diversity. Kruskal-Wallis and *post hoc* Nemenyi tests for alpha-diversity metrics detected significant differences ($p < 0.05$) among pairwise comparisons including the 1,000 sequence minimum library size for observed richness. (E) The principal coordinate analysis (PCoA) of Jaccard distances among subsamples with no rarefaction. Permutational multivariate analysis of variance (PERMANOVA) did not detect significant differences ($p < 0.05$) among sample sizes. (F) The principal coordinate analysis (PCoA) of Jaccard distances among minimum library sizes of subsample 188. PERMANOVA detected significant differences ($p < 0.05$) among rarefaction levels and all pairwise differences were statistically significant ($p < 0.05$).

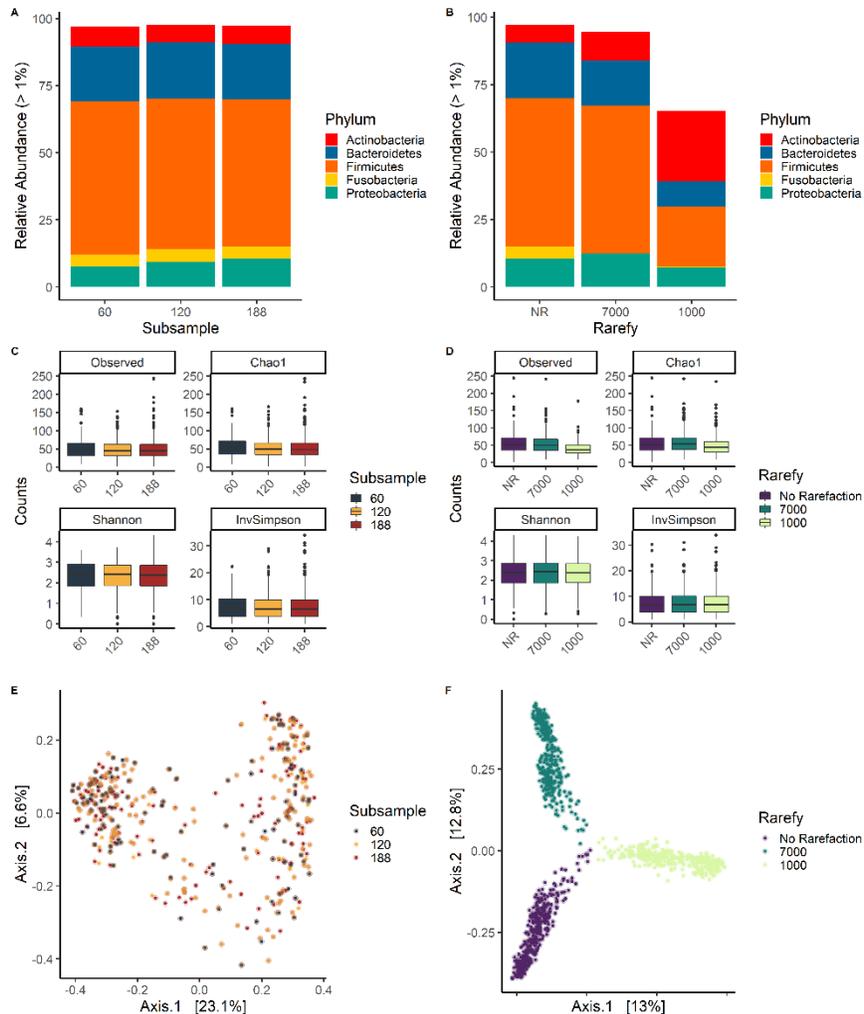


Table 4: Total number of sequences that remained after filtering, and unclassified sequences (phylum and family taxonomic level) among sample area, sample size, and minimum library size.

Pipeline	Sample Area	Sample Size	Minimum Library Size	Total Number of Sequences	Unclassified Phylum	Unclassified Family
QIIME2	Mouth	60	No Rarefaction	2,115,130	4,988	7,660
QIIME2	Mouth	60	7,000	405,998	1,059	1,481
QIIME2	Mouth	60	1,000	60,000	157	207
QIIME2	Rectum	60	No Rarefaction	2,349,815	27,026	30,707
QIIME2	Rectum	60	7,000	405,981	5,200	5,438
QIIME2	Rectum	60	1,000	59,996	733	772
QIIME2	Mouth	120	No Rarefaction	4,182,126	7,721	12,999
QIIME2	Mouth	120	7,000	811,998	1,935	2,703
QIIME2	Mouth	120	1,000	119,000	265	381
QIIME2	Rectum	120	No Rarefaction	3,834,632	35,019	36,583
QIIME2	Rectum	120	7,000	734,993	6,154	6,762
QIIME2	Rectum	120	1,000	106,999	862	958
QIIME2	Mouth	188	No Rarefaction	5,874,406	14,099	22,323
QIIME2	Mouth	188	7,000	1,154,999	4,044	5,424
QIIME2	Mouth	188	1,000	168,999	612	819
QIIME2	Rectum	188	No Rarefaction	5,463,539	63,667	70,089
QIIME2	Rectum	188	7,000	1,056,993	12,498	13,963
QIIME2	Rectum	188	1,000	155,000	1,798	2,017

(Kruskal-Wallis, *post hoc* Nemenyi; $p < 0.05$, Figure 4C, Table S19-S20) When comparing core microbiome taxa among rarefaction levels, most OTUs (695) were shared among all minimum library sizes and were represented in higher log abundance (greater than the mean log abundance) compared to non-core taxa (Figure 5B, Table S21). Some taxa only occurred in non-rarefied samples (62) and amounted to about 10% of the core taxa among all minimum library sizes (Figure 5B, Table S21). Despite changing minimum library sizes, a majority of taxa

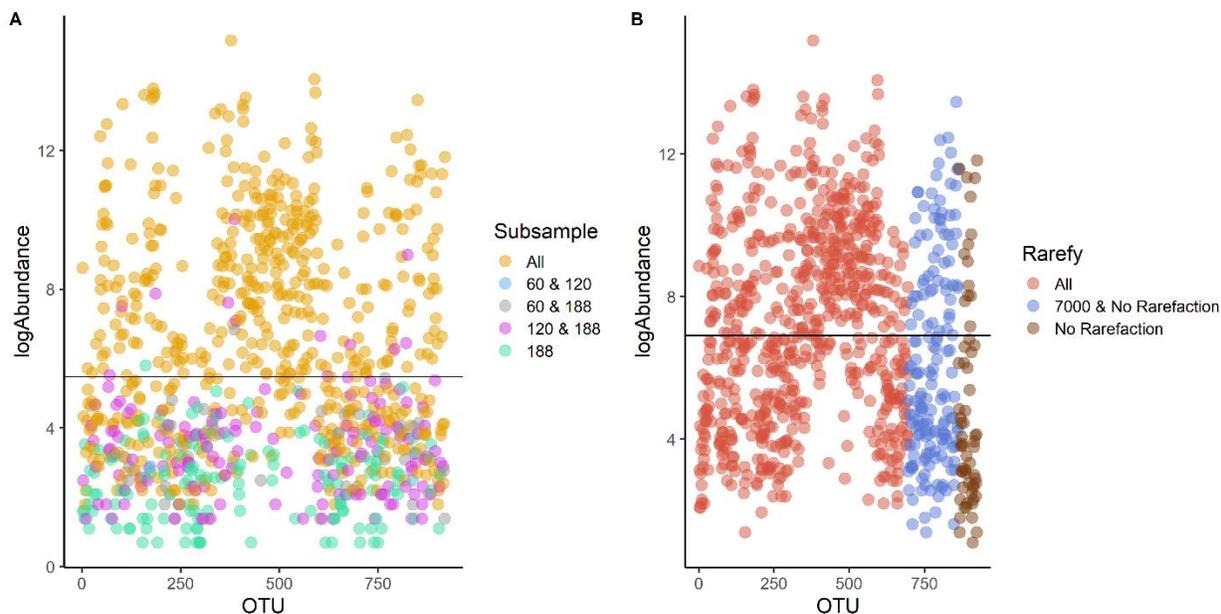
were still represented. However, the low abundant and infrequent taxa still have downstream effects for forensic analysis.

Random forest modeling was affected by low abundant and infrequent taxa. While sample area classification error (14% for phylum and 4% for family) was lower than manner of death classification error (phylum and family error rates were nearly 56%), among all minimum library sizes, sample area classification error and manner of death classification error were relatively similar (Table 5, Table S22). Although the error rates among library sizes were overall similar, some of the important indicator taxa for classifications were unique to certain groups. For example, Bifidobacteriaceae was an important (one of the highest mean Gini decrease) indicator of manner of death for subsamples 60 and 120 and at rarefaction levels 0 and 1,000, while Spirochaetales was an indicator for sample area using subsamples 120 and 188 (Table S23). For downstream applications in forensic microbiology, changing indicator taxa among minimum library sizes indicated that lack of standardization among studies can lead to differentiation among results interpretation.

Sample Size

Overall, sample size did not affect overall microbial community metrics in a significant way. Sample size remained relatively constant among minimum library sizes (standard deviations- no rarefaction: 47.4; 7,000: 45.1; 1,000: 46.0) (Table S24). The total number of sequences, as well as the number of unclassified sequences at phylum and family level, increased 1.5 to 2-times with increasing sample number from 60, 120, to 188 (Table S24), but there were no differentially abundant taxa among subsamples (Figure 4B). Alpha-diversity metrics did not significantly differ among sample sizes (Kruskal-Wallis; $p > 0.05$; Figure 4D, Table S19-

Figure 5: **Core microbiome among sample sizes and minimum library sizes.** In this case, core microbiome is considered by OTU (operational taxonomic unit) membership among defined groups (sample size or minimum library size). (A) Occupancy vs. log abundance of shared OTUs among sample sizes at no rarefaction. The horizontal line indicated mean log abundance. (B) Occupancy vs. log abundance of shared OTUs among minimum library sizes for 188 subsample. The horizontal line indicated mean log abundance.



S20), but observed richness did increase with increasing sample size, as larger sample sizes captured more infrequent taxa. For subsamples within each sample area and minimum library size, there were no significant differences (PERMANOVA; $p > 0.05$) among beta-diversity and beta-dispersion and no clear clustering by subsample (Figure 4E; Table S18).

However, increasing sample size changed the number of low abundant and infrequent taxa.

Comparing the core microbiome, most OTUs (569) were shared among all sample sizes, and were represented in higher log abundance (greater than the mean log abundance) compared to non-core taxa (Figure 5A, Table S25). Although, there were a few OTUs shared among subsamples and unique to the subsample 188 (Figure 5A, Table S25). Much like minimum library size, error rates for predicting sample area among sample sizes were around 14% for

Table 5: Random forest classification error among sample sizes and minimum library sizes. Random forest classifications were made with 1,000 trees, and out-of-bag (OOB) error was reported. Classifications of sample area (mouth and rectum) and manner of death (suicide, homicide, accident, natural) were made for each taxonomic level, rarefaction level, and subsample. Percentages and number of misclassifications from the total number of samples were included.

Taxonomic Level	Minimum Library Size	Sample Size	Sample Area Error Rate (Misclassifications/ Total)	Manner of Death Error Rate (Misclassifications/ Total)
Phylum	1,000	60	14.17% (17/120)	55.83% (67/120)
Phylum	1,000	120	15.49% (35/226)	58.41% (132/226)
Phylum	1,000	188	17.59% (57/324)	57.41% (186/324)
Phylum	7,000	60	14.66% (17/116)	57.76% (67/116)
Phylum	7,000	120	14.03% (31/221)	57.47% (127/221)
Phylum	7,000	188	14.24% (45/316)	56.33% (178/316)
Phylum	No Rarefaction	60	12.50% (15/120)	54.17% (65/120)
Phylum	No Rarefaction	120	14.29% (33/231)	58.44% (135/231)
Phylum	No Rarefaction	188	14.50% (48/331)	58.61% (194/331)
Family	1,000	60	5.83% (7/120)	54.17% (65/120)
Family	1,000	120	3.10% (7/226)	56.64% (128/226)
Family	1,000	188	4.63% (15/324)	56.17% (182/324)
Family	7,000	60	5.17% (6/116)	55.17% (64/116)
Family	7,000	120	4.07% (9/221)	59.28% (131/221)
Family	7,000	188	5.06% (16/316)	58.54% (185/316)
Family	No Rarefaction	60	4.17% (5/120)	54.17% (65/120)
Family	No Rarefaction	120	4.76% (11/231)	57.58% (133/231)
Family	No Rarefaction	188	3.93% (13/331)	58.31% (193/331)

phylum and 4% for family, and around 56% for manner of death classifications (Table 5). Again, there were indicator taxa unique to certain subsamples (Table S23). The increase of infrequent and low abundant taxa among the sample sizes are indicative that sample size should be taken into consideration for downstream application for forensic microbiology studies, and comparisons across studies.

Discussion

Our bioinformatic parameter comparison using a large postmortem microbiome dataset showed that parameters can affect downstream analyses, including microbial community structure results. We also show that sample size and minimum library size affect the resulting number of low and infrequent taxa and potential indicator taxa for model building. While the results here are specific to this postmortem dataset of targeted amplicon (16S rRNA) sequencing on the Illumina platform, similar considerations should be made for other pipelines and platforms that may be used in downstream applications of forensic microbiology in the criminal justice system.

Pipeline Comparison

Pipelines differed in many ways, including the development, parameters, and usability (Plummer and Twin 2015). To accurately compare pipelines, we standardized as many steps as possible; i.e. using SILVA as a reference database. The steps that could not be standardized among pipelines, including different taxonomic assignment and alignment algorithms, were likely responsible for the differences in microbial community characterizations (Siegwald et al. 2017).

Overall, we showed that MG-RAST finds a different microbial community structure compared to mothur and QIIME2, due to reduced diversity metrics and increased unclassified and differentially abundant taxa. These results are similar to previous work comparing MG-RAST to other pipelines (D'Argenio et al. 2014). However, there are few studies that did not find the same reduction in diversity metrics and increased unclassified reads (Plummer and Twin 2015; Golob et al. 2017). Although, both studies focused on a smaller sample size of similar

microbial communities (35 infant gut microbiome samples) or *in silico* generated data that does not approximate forensic data. Due to the results from this postmortem dataset, the reduction in microbial information (samples, classified sequences), the varied community structure (differentially abundant taxa, reduced alpha- and beta-diversity), and the higher random forest classification error, MG-RAST is not the appropriate tool for this type of forensic dataset moving forward.

The overall microbial community structure had minimal differences among QIIME2 and mothur, which was a similar result to previous studies (Plummer and Twin 2015; Golob et al. 2017). It is important to note that we rarefied samples to 1,000 sequences, to account for the sequence reduction in MG-RAST, but this could be limiting the differences among mothur and QIIME2. This study was comparison of bioinformatic pipelines using machine learning outcomes. Rather than an exhaustive search of machine learning algorithms (Zhang et al. 2019), we used a standardized user-friendly methodology of random forest classification (out-of-bag error) after comparison to the test-set validations were within 2% error rate as OOB error rate. QIIME2 had a higher overall classification error than mothur but resulted in less than 1% difference from mothur at the family taxonomic level. Overall, the increased unclassified taxa and *in silico* data error for mothur made QIIME2 the more appropriate choice for downstream analyses in this study.

Minimum Library Size

Rarefying, as a method of normalizing varying library sizes, should be of interest for forensic applications in the future as standardizing methodology among labs will be important for actual casework. Many of the current postmortem microbiome studies have a variety of

minimum library sizes chosen based on alpha-rarefaction curves (Metcalf et al. 2016; Pechal et al. 2018) or based on minimum library size of the samples (DeBruyn and Hauther 2017). We chose three levels of rarefying for the postmortem data, based on these different approaches: no rarefaction, 7,000 (based on alpha-rarefaction curves), and 1,000 (minimum library size).

Previous research determined that in large enough datasets of very different microbial communities (postmortem data across body site) (Pechal et al. 2018), rarefying should not negatively impact downstream analysis (Weiss et al. 2017). However, we found differences among the normalization strategies, indicating that rarefying should be taken into consideration for forensic microbiome analysis. While we showed that more low abundance and infrequent taxa among cases are captured without rarefying, downstream statistical analyses commonly used in forensic microbiology studies (i.e. ANCOM, PERMANOVA) assume similar library sizes (Weiss et al. 2017). Demonstrated by the PCoA plot (Figure 4F) a large majority of the postmortem samples in this study clustered by minimum library size, a result that was found in previous normalization research (Weiss et al. 2017). This clustering, and indication of distinct microbial communities, could possibly limit the ability for laboratories to compare data across studies if normalization was not standardized.

For standardization in forensic analyses, normalization of library size should be taken into consideration due to the effect on downstream statistical analyses. Random forest classification has potential forensic applications for many analyses including manner of death determination. Error rates of classifications (both of sample area and manner of death) were generally stable among minimum library size but decreased with larger library sizes, perhaps as more of the infrequent taxa become important for classification. While classification of sample area is generally consistent with other studies (Pechal et al. 2018), the classifications of manner

of death in this study are generally lower than previously reported studies with the same dataset (Zhang et al. 2019). However, a previous study provided a comprehensive evaluation of machine learning algorithms, which was not the major focus of this study. Rather, by using a standardized, user-friendly open-access methodology of random forest classification, we illustrated the importance to future researchers to consider parameters, such as library size, even if they are not a primary focus of the study. Indicator taxa that are important for classifications changed among minimum library sizes, which has implications for forensic cases, as indicator taxa among classifications will be potentially very important for downstream applications in casework.

Sample Size

There is a tradeoff for researchers to consider when including more samples in analyses. Sample sizes for studies are mostly limited by resource availability, including funding. It is not realistic for all studies to have very large sample sizes. However, to improve forensic microbiology for future use in the criminal justice system, more robust tools that capture low abundant and infrequent taxa encountered in real cases are needed.

Our postmortem microbiome dataset is the largest analyzed to date for postmortem microbiome studies (Pechal et al. 2018). For downstream forensic applications, including more samples in predictive models makes those models more robust to the variability present in real forensic cases. We found patterns of relative abundance did not change as more samples were included in the analysis, but observed richness did increase with increasing sample size, as larger sample sizes captured more infrequent taxa. As expected, increasing subsample size increased

error rate for the random forest classifications (sample area and manner of death). However, those random forest models are arguably more robust for downstream forensic applications (Zhang et al. 2019), therefore including as many samples as possible in forensic microbiology studies is best practice moving forward.

Conclusion

This work was the first to compare pipeline and parameters for a forensically relevant, large, and heterogenous dataset. Based on the results of this study, we make the following recommendations for future forensic microbiology studies: 1. The QIIME2 pipeline is the most suitable pipeline for this type of postmortem microbiome dataset; 2. Rarefying data is the best normalization practice for downstream statistical analyses (Weiss et al. 2017). However, an appropriate minimum library size should be chosen based on richness captured (alpha-rarefaction plots), instead of the smallest library size among samples; and 3. Sample size should be maximized to captures lower abundant and infrequent taxa among the data for more robust model building. However, sample size must be weighed with other practical considerations, such as financial constraints.

Considering the potential application of forensic microbiology to the criminal justice system, continued research to optimize computational methodology will be important for downstream applications. While model building was not the focus of this study, the preliminary results show how parameter choice can potentially affect downstream applications, which is important for future research and casework. Applying bioinformatic workflows necessary for microbial data in forensic casework will be challenging as command line software and microbial data analysis is not already part of the examiner workflow. The constant influx of pipelines available, changing parameter options, and updates will be a barrier to creating an SOP for forensic casework. However, this study is an important part of laying the groundwork for standardizing computational methodology for forensic microbiology research, and will help set the precedent for forensic casework in the future.

CHAPTER II:
DYSBIOSIS IN THE DEAD: HUMAN POSTMORTEM MICROBIOME BETA-
DISPERSION AS ONE INDICATOR OF MANNER AND CAUSE OF DEATH DURING
AUTOPSY

Introduction

The organisms represented in microbiomes have important functional roles for host life, influencing health, development, and disease susceptibility, among many others (Zaneveld, McMinds, and Thurber 2017; Turnbaugh et al. 2007). Microbes also play an important functional role in decomposition (Pechal et al. 2014), as the communities change with dispersal, competition, and other interactions after host death (Pechal et al. 2014, 2018; Metcalf et al. 2016). These dynamic, yet predictable, microbial community profile changes after death make the postmortem microbiome a potential forensic resource for postmortem interval (PMI) estimation. PMI estimation is indeed the most studied forensic application of the postmortem microbiome (Pechal et al. 2014; Metcalf et al. 2013); but, this community has additional potential for forensic applications as well, like indicating antemortem health conditions (e.g., cardiovascular disease or violent death) (Pechal et al. 2018) and the living environment (e.g., neighborhood blight) (Pearson et al. 2019).

The postmortem microbiome is thought to be structured by a decedents' antemortem health condition and the suite of stressors that impact the human host including drug/alcohol abuse or high stress lifestyle conditions such as neighborhood blight, that are associated with certain manners of death (i.e., homicide) (Pechal et al. 2018; Pearson et al. 2019; Zhang et al. 2019). Importantly, the postmortem microbiome does not significantly change from the antemortem microbiome for approximately 48 hours after death (Pechal et al. 2018). Due to the stability of the postmortem microbiome within 48 hours after death, and potential connection to lifestyle condition, microbial community metrics have been shown to indicate certain manners of deaths (MOD) or causes of deaths (COD). However, there have been few studies that have tested associations of postmortem microbial communities with MOD or COD (Pechal et al. 2018; Pearson et al. 2019; Zhang et al. 2019).

In past work, microbial diversity and indicator taxa were shown to reflect antemortem health conditions and MOD (Pechal et al. 2018; Pearson et al. 2019; Zhang et al. 2019). In some cases lower microbial alpha-diversity was associated with cardiovascular disease, non-violent deaths, and neighborhood blight (e.g., abandoned building, inactivity, and dumping) (Pechal et al. 2018; Pearson et al. 2019); however, it is difficult to capture the variability in a large sample set using diversity alone, as it does not account for taxon relative composition that makes up the community. In one of two studies, Zhang et al. combined microbial indicator taxa and case metadata found in autopsy reports (e.g., decedents age, sex, race, etc.) to test machine learning models for classifying M/COD (Zhang et al. 2019). While indicator taxa are a useful reflection of antemortem conditions, microbial indicator taxa may not be present in all cases (e.g., *Haemophilus influenzae*), or they may be ubiquitous (e.g., *Staphylococcus*), and therefore less useful (Pechal et al. 2018). For this study, we tested a new metric that captures microbial variability and does not specifically rely on indicator taxa: beta-dispersion. We hypothesized that postmortem microbiome beta-dispersion could be an additional tool for determining M/CODs during death investigation.

Increased beta-dispersion among living individuals has been associated with obesity, infection, and smoking (Zaneveld, McMinds, and Thurber 2017; Barbian et al. 2015). Following the conceptual context of the ‘Anna Karenina Principle’ (AKP), after prolonged exposure to any array of stressors, the microbiomes of unhealthy individuals becomes more variable compared to the microbial communities of healthy individuals (Zaneveld, McMinds, and Thurber 2017). In other words, increased variation in the microbial communities reflects dysbiosis, and this community variability can be quantified through calculations of beta-dispersion (Zaneveld, McMinds, and Thurber 2017; Barbian et al. 2015). Beta-dispersion is calculated within the

context of a dataset, based on the multivariate distance from the centroid for each case sample, as defined by a grouping factor (Oksanen et al. 2019). Based on the link between increased beta-dispersion and health status, we considered M/COD as grouping factors to quantify microbial signatures associated with M/COD determinations. Such association of beta-dispersion with M/COD could conceivably be additional evidence during death investigation.

Microbial community metrics could potentially aid medical examiners and other certifiers of death, as determining M/COD can be error prone. While COD is the injury/disease a person died from which spans a variety of causes, MOD encompasses only five major categories: natural, accident, suicide, homicide, and undetermined (Randy, Hunsaker III, and Davis 2002). Medical examiners and other certifiers of death qualify their MOD determination with incremental degrees of certainty considering multiple pieces of evidence (Randy, Hunsaker III, and Davis 2002). Given the possibility for mismatch between the MOD determination and the actual MOD, the postmortem microbiome could provide another piece of evidence to justify M/COD determination.

To evaluate how postmortem microbiome variability associated with M/CODs, we modeled postmortem microbiome beta-dispersion from five body sites of 188 routine autopsy with known M/COD (as determined by a board-certified medical examiner). We predicted that certain M/CODs, such as natural deaths and cardiovascular disease, would have higher beta-dispersion than other M/CODs due to the previous antemortem health condition links found in previous studies (Pechal et al. 2018). However, the effect of life environment was predicted to increase beta-dispersion as well, in a way that could potentially factor into homicides or blunt force trauma/gunshot wounds (Pearson et al. 2019). Quantifying beta-dispersion by using M/COD as a

grouping factor could provide reliable and usable tool in death investigation for M/COD determination.

Materials and Methods

Sample Collection, DNA Extraction, and Sequencing

The postmortem microbiome data used in this study were acquired from Pechal et al. (2018) but re-analyzed to test M/COD determination from beta-dispersion. This dataset contains postmortem microbiome samples obtained from 188 Wayne County Medical Examiner's Office autopsy cases (Detroit, MI 2014-2016), representing multiple MODs (accident, homicide, suicide, natural) and CODs (asphyxiation, blunt force trauma, cardiovascular disease, drug-related deaths, gunshot wounds, etc.) (Table S1) (Pechal et al. 2018). The cases also represent a cross-section of the Detroit area populace and were nearly evenly divided among females and males (83:105) and black and white (90:98) (Table S1). Cases in the dataset used were from people 18–88 years with a body mass index (BMI) from 8.5–67.5 kg/m² (Table S1). The dataset is the largest postmortem microbiome available to test beta-dispersions potential to aid M/COD determination.

Detailed methods for sample collection, DNA extraction, and sequencing can be found in (Pechal et al. 2018). In summary, trained personnel at Wayne County Medical Examiner's Office took microbial community swab samples from five body sites (nose, mouth, rectum, ears, and eyes) during routine autopsy. Microbial DNA was extracted and sequenced to characterize the microbial communities. The Michigan State University (MSU) Genomics Core Facility (East Lansing, MI, USA) sequenced the 16S rRNA V4 region using an Illumina MiSeq standard flow cell (v2) using a 500-cycle reagent cartridge.

Data Analysis and Bioinformatics

Sequence reads from postmortem microbiome samples were analyzed with QIIME2 (v2018.11) (Bolyen et al. 2019) following the methodology outlined in (Kaszubinski et al. 2019). Taxonomy and amplicon sequencing variant (ASV) tables were exported as csv files to be used

as input data for all downstream analysis. ASV and taxonomy files were combined with metadata obtained from autopsy reports (age, sex, race, BMI etc.) as phyloseq (v1.28.0) objects in R (v3.6.1) (McMurdie and Holmes 2013; R Core Team). ASVs less than 0.01% of the mean library size were trimmed, removing 22,214 ASVs for a total of 8,692 ASVs. Phyloseq objects were split among the body sites (nose, mouth, rectum, ears, and eyes) and analyzed separately.

Method Selection

We needed to determine the optimal methodology for calculating beta-dispersion before moving forward with classifying M/COD. To do this, we compared standardization approaches of the microbial communities, distance matrices, the number of significant M/COD comparisons, and alpha-diversity to select the optimal method for beta-dispersion calculation. For standardization, rarefying (randomly subsampling ASVs to a specified minimum library size) and normalizing (removing ASVs not present in a specified percentage of samples) were compared for each body site using three minimum library sizes (3,000, 5,000, and 7,000 sequences) sample percentage cut off (1%, 3%, 10%), respectively. While rarefaction has been debated (Weiss et al. 2017; McMurdie and Holmes 2014), we sought to eliminate bias associated with different library sizes that could inflate differences in beta-dispersion among M/CODs (Kaszubinski et al. 2019).

We also compared unweighted and weighted unifracs distance matrices for calculating beta-dispersion. Unifrac is commonly used in forensic studies (Javan et al. 2017; Pechal et al. 2018). We wanted to determine whether considering abundances would affect the beta-dispersion calculation and should be considered for downstream modeling. Beta-dispersion was calculated among MODs and CODs using the vegan package (v2.5-5) in R (v3.6.1) for each minimum library size and sample percentage cutoff (Oksanen et al. 2019). The betadisper

function from *vegan* reports the distance from the centroid for each sample, as defined by a grouping factor (in this case, M/COD). Every postmortem microbiome sample had two corresponding beta-dispersion values, with either MOD or COD as a grouping factor. Kruskal-Wallis and Fligner-Killeen tests among beta-dispersion values determined differences among M/CODs and were reported with a Bonferroni correction (Pohlert 2018). Additionally, alpha-diversity metrics (Chao1 and Shannon diversity) for each minimum library size and sample percentage cutoff were calculated using *phyloseq* (v1.28.0). We selected a methodology (standardization approach, distance matrix) for calculating beta-dispersion based on the number of significant differences in beta-dispersion identified by Kruskal-Wallis as well as the highest alpha-diversity (see Results: Method Selection). We wanted to select a method that had potential for distinguishing M/COD but did not lose microbial diversity. For subsequent analyses, microbial communities were rarefied to 5,000 sequences and beta-dispersion was calculated using unweighted unifracs distance.

Model Selection

We chose logistic regression modeling to showcase how beta-dispersion and case metadata could be a reliable and usable tool in death investigation for M/COD determination. We built multinomial logistic regression models to classify M/COD from beta-dispersion values and case metadata (Böhning 1992) using the *lme4* (v1.1-21) and *mlogit* package (v1.0-1) (Bates et al. 2019; Croissant 2019). Multinomial logistic regression models included all categories of interest for classifying M/COD (e.g., homicide, suicide, natural, and accidental death for MOD; cardiovascular disease, drug related deaths, blunt force trauma, asphyxiation, gunshot wounds, and other deaths for COD). Metadata of interest [age, BMI, sex, race, PMI (≤ 48 h; > 49 h), season, and event location (outdoors, indoors, hospital, vehicular)] were summarized and tested

with Kruskal-Wallis among M/CODs with a Bonferroni correction to determine the covariates for model building. (Pohlert 2018). We wanted to identify if metadata were indicative of certain M/CODs (e.g., age significantly higher in natural deaths or for cardiovascular disease) and were useful to supplement beta-dispersion in downstream modeling.

We built and tested multiple logistic regression models for each body site, classifying M/COD. Only models with significant ($p < 0.1$) beta-dispersion contribution were selected for further modeling with case metadata. We chose a slightly less conservative cut off for the p-value so that potentially important metadata were not prematurely removed. The best performing models were considered based on Akaike information criterion (AIC) (Bozdogan 1987), McFadden R^2 , and classification success (correct classifications / total number of samples). We used multinomial logistic regression model results were used to determine which body sites beta-dispersion classified M/COD well. Based on initial model building results, models could be improved (see Results: Model Selection).

To improve logistic regression models, we considered three microbial community types for beta-dispersion calculation: full communities, random forest indicator communities, and ‘non-core’ communities. While the grouping factor (M/COD) remained the same, the microbial communities used to calculate beta-dispersion differed. “Full community” beta-dispersion was calculated from the entire filtered and rarefied community within each body site. “Random forest indicators” were determined by Boruta (v6.0.0) using confirmed and tentative taxa of importance ($p < 0.05$) (Kursa and Rudnicki 2018), and beta-dispersion was calculated from identified significant indicator taxa. Random forest classification error was determined using the randomForest package (v4.6-14) (Breiman et al. 2018). While the “core microbiome” definition is widely debated (Shade and Handelsman 2012), we determined ‘non-core’ taxa in this case as

taxa present in only one M/COD of interest (e.g., accidents, blunt force trauma, etc.). ‘Non-core’ taxa were removed, and beta-dispersion was calculated from the remaining taxa.

Lastly, we also test binary logistic regression (classifying between two categories) compared to multinomial logistic regression (multiple categories). Binary logistic regression models were also built using the lme4 (v1.1-21) and mlogit package (v1.0-1) (Bates et al. 2019; Croissant 2019), and the best performing models were considered based on Akaike information criterion (AIC) (Bozdogan 1987), McFadden R^2 , and classification success (correct classifications / total number of samples) much like the multinomial logistic regression models.

Case Studies

Using the methodology outlined above, we tested data from two case studies for classifying MOD from nose communities, to showcase the forensic potential beta-dispersion has as a tool for medical examiners. For the first case study, we chose a matched design of cases examining suicide vs other manners of death, including similar ages, races, and sexes to limit the effect of metadata for a total of 43 cases (22 suicides; 21 accidents/homicide/natural) (Table S26). For the second case study, we examined MOD within COD; examining gunshot wound homicides (n = 25) vs. suicides (n = 4) (Table S26). As previous studies have identified indicator taxa (Pechal et al. 2018; Pearson et al. 2019; Zhang et al. 2019), we determined potential indicator taxa using Boruta. We also evaluated the potential beta-dispersion differences using Permutational multivariate analysis of variance (PERMANOVA), and classified MOD using binary logistic regression (Anderson 2017). We calculated *a priori* power using G*Power 3 v3.0.5 (Faul et al. 2007). Case study beta-dispersion was compared using the mean and standard deviation using an independent mean two-tailed t-test ($\alpha = 0.05$; $1-\beta = 0.80$).

Data Availability

All postmortem microbiome case data can be found as supplemental material or associated archived data of (Pechal et al. 2018). Sequence data were archived through the European Bioinformatics Institute European Nucleotide Archive (www.ebi.ac.uk/ena) under accession number: PRJEB22642. The microbial community analysis is available as R code on GitHub (<https://github.com/sierrakasz/AKP-betadisp-paper>).

Results

Method Selection

Beta-dispersion among M/CODs was calculated using either unweighted or weighted unifrac distance matrices from microbial communities that were either rarefied or normalized (Figure S1-2, Table S27). Microbial communities were rarefied to three minimum library sizes (3,000, 5,000, and 7,000 sequences), or normalized to three sample percentage cut offs (1%, 3%, 10%). Each body site was considered separately. Therefore, we had a total of 120 comparisons to determine which beta-dispersion calculation method was most appropriate for downstream model building.

We determined that unweighted compared to weighted unifrac distances were the optimum distance matrix for this study. Unweighted unifrac distances had three-times as many significant comparisons than weighted unifrac (Kruskal-Wallis $p < 0.05$; Figure S1, Table S27), while unweighted unifrac distances had four-times less significant deviations of variance compared to weighted unifrac (Fligner-Killeen $p < 0.05$; Figure S2, Table S27). We showed that unweighted unifrac distances were more robust against rarefying and normalizing, as significant comparisons occurred with both standardization (Kruskal-Wallis and *post hoc* Nemenyi $p < 0.05$; Figure S1; Table S27). For normalizing combined with weighted unifrac, we found a bias of library size among the significant comparisons (Table S28). Most of the significant comparisons (7 out of 10) had differential library sizes, reflecting that the significance did not originate from M/COD differences in beta-dispersion (Kruskal-Wallis and *post hoc* Nemenyi $p < 0.05$; Table S28). Due to the number of significant comparisons across standardization strategies, less deviations of variance, and lack of library size bias, we chose unweighted unifrac as the appropriate metric for additional analyses.

We determined that rarefaction was the more appropriate standardization strategy than

normalization for this study. While normalizing had more than double (19 vs. 9) the significant comparisons than rarefying, normalizing microbial communities led to a significant decrease in richness (Chao1) and alpha-diversity (Shannon diversity) compared to rarefying (Kruskal-Wallis and *post hoc* Nemenyi $p < 0.05$; Figure S1, S3, Table S27, S29). Due to loss of alpha-diversity and library size bias of normalizing, we chose rarefying as the appropriate standardization strategy.

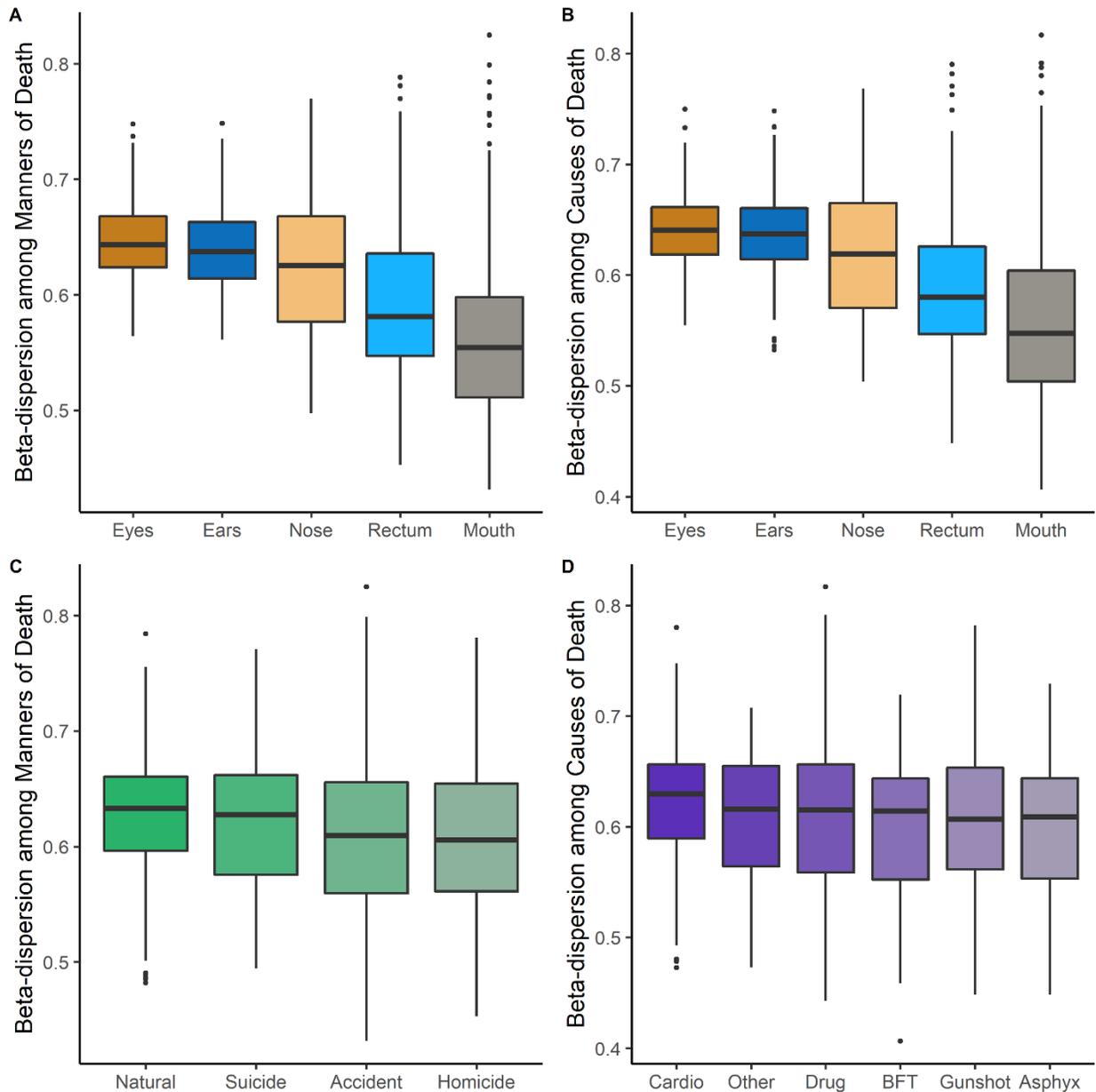
In conclusion, we chose a minimum library size of 5,000 sequences as more body sites yielded significant comparisons (Kruskal-Wallis $p < 0.05$) and was the appropriate minimum library size based on alpha-rarefaction curves of sequencing depth (Figure S4). For downstream model building, samples were rarefied to 5,000 sequences and beta-dispersion was calculated using the unweighted unifrac dissimilarity matrix.

Model Selection

Beta-dispersion significantly differed among body sites and M/CODs (Kruskal-Wallis $p < 0.05$, Figure 6, Table S30). Every postmortem microbiome sample had two corresponding beta-dispersion values (distances from centroid), with MOD or COD as the grouping factor. On average, eye microbiomes had highest beta-dispersion [MOD: 0.646 (SD = 0.0346); COD: 0.642 (SD = 0.035); Table S30] while mouths had the lowest beta-dispersion [MOD: 0.567 (SD = 0.0779); COD: 0.563 (SD = 0.0800); Table S30]. Beta-dispersion for all body site communities was significantly different for M/COD except the ears and eyes, but we considered all body sites for downstream modeling with logistic regression (Kruskal-Wallis $p < 0.05$; Table S30). Natural death postmortem microbiomes had the highest average beta-dispersion [0.628 (SD = 0.056); Table S30] compared to homicides [0.606 (SD = 0.0694)] and accidents [0.608 (SD = 0.0683); Kruskal-Wallis $p < 0.05$; Table S30]. Microbiomes of cases with cardiovascular disease had

Figure 6: **Beta-dispersion values among body sites and manners/causes of death (M/CODs).**

A) Beta-dispersion values for MODs among body sites. B) Beta-dispersion values for CODs among body sites. C) Beta-dispersion values among MODs. D) Beta-dispersion values among CODs (Cardio = cardiovascular disease; Drug = drug-related deaths; BFT = blunt force trauma; Asphyx = Asphyxiation).



significantly higher beta-dispersion among all body sites [0.625 (SD= 0.0565); Table S30]

compared to asphyxiation, which had the lowest [0.598 (SD= 0.0678); Table S30]. Beta-

dispersion for cardiovascular disease was significantly higher than gunshot wounds [0.605 (SD

Table 6: **Case metadata stratified by manner/cause of death (M/COD)**. Age and BMI (kg/m²) were reported with means and standard deviations, while sex, race, PMI, event location, and season were reported as counts within M/COD.

Manner of Death	Age		BMI		Sex		Race		PMI		Event Location				Season			
	mean	sd	mean	sd	Female	Male	Black	White	<48 hr	> 49 hr	Hospital	Indoors	Outdoors	Vehicular	Autumn	Spring	Summer	Winter
Accident (n=71)	39.9	13.8	30	8.48	35	36	26	45	62	9	10	51	7	3	4	41	17	9
Homicide (n=39)	36.2	12.3	26.9	6.42	9	30	34	5	37	2	9	14	12	4	2	20	6	11
Natural (n=57)	53.7	10.7	29.8	10.1	25	32	29	28	45	12	6	49	1	1	2	33	14	8
Suicide (n=22)	44.9	15.6	27.2	6.57	11	12	3	20	21	2	1	18	3	1	2	10	6	5

Cause of Death	Age		BMI		Sex		Race		PMI		Event Location				Season			
	mean	sd	mean	sd	Female	Male	Black	White	<48 hr	> 49 hr	Hospital	Indoors	Outdoors	Vehicular	Autumn	Spring	Summer	Winter
Asphyxiation (n=11)	46.5	20.1	25.4	4.35	4	7	2	9	10	1	1	7	2	1	1	3	3	4
Blunt force trauma (n=21)	42.6	14.7	28.9	10.3	12	9	14	7	21	0	4	7	8	2	4	9	3	5
Cardiovascular disease (n=42)	54.5	11.6	29.9	9.56	20	22	23	19	34	8	6	35	1	0	2	26	7	7
Drug-related deaths (n=70)	41	12.7	29.7	7.51	34	36	22	48	58	12	5	60	3	2	2	43	17	8
Gunshot wounds (n=30)	33.8	11.8	27.5	6.34	5	25	24	6	28	2	7	10	9	4	1	16	6	7
Other (n=16)	47.3	12	28.8	12.1	5	11	7	9	14	2	3	13	0	0	0	7	7	2

=0.0708)], blunt force trauma [0.601 (SD = 0.0624)] and drug-related deaths [0.611(SD = 0.0684); Kruskal-Wallis $p < 0.1$; Table S30]. While beta-dispersion means differed significantly among MODs and CODs, there was an overlap among beta-dispersion values, indicating that other variables contribute to microbiome beta-dispersion (Figure 6). Therefore, we considered additional case metadata for downstream modeling. Metadata were differentially represented in MODs and CODs (Table 6; Table S31). Age, sex, race, and event location were significantly different among MODs/CODs (Kruskal-Wallis $p < 0.05$; Table S31); however, so as to not prematurely remove potentially important metadata we included all metadata of interest in downstream modeling [age, BMI, sex, race, PMI (≤ 48 h; > 49 h), season, and event location (outdoors, indoors, hospital, vehicular)].

We found that initial multinomial logistic regression models were useful for determining which body site beta-dispersion had the best classification potential for M/COD. Classifying among all MODs, nose and mouth community was a significant covariate, while nose, mouth, and ear beta-dispersion were a significant covariate for classifying all CODs ($p < 0.1$; Table S32). Nose community beta-dispersion on average successfully classified MODs at 61.05%, while ears and nose beta-dispersion successfully classified CODs on average 62.87% and 62.79%, respectively (Table S32).). Based on these results, we used only body sites with statistically significant ($p < 0.05$) beta-dispersion as a covariate in the models for further model building: nose, mouth, and ears.

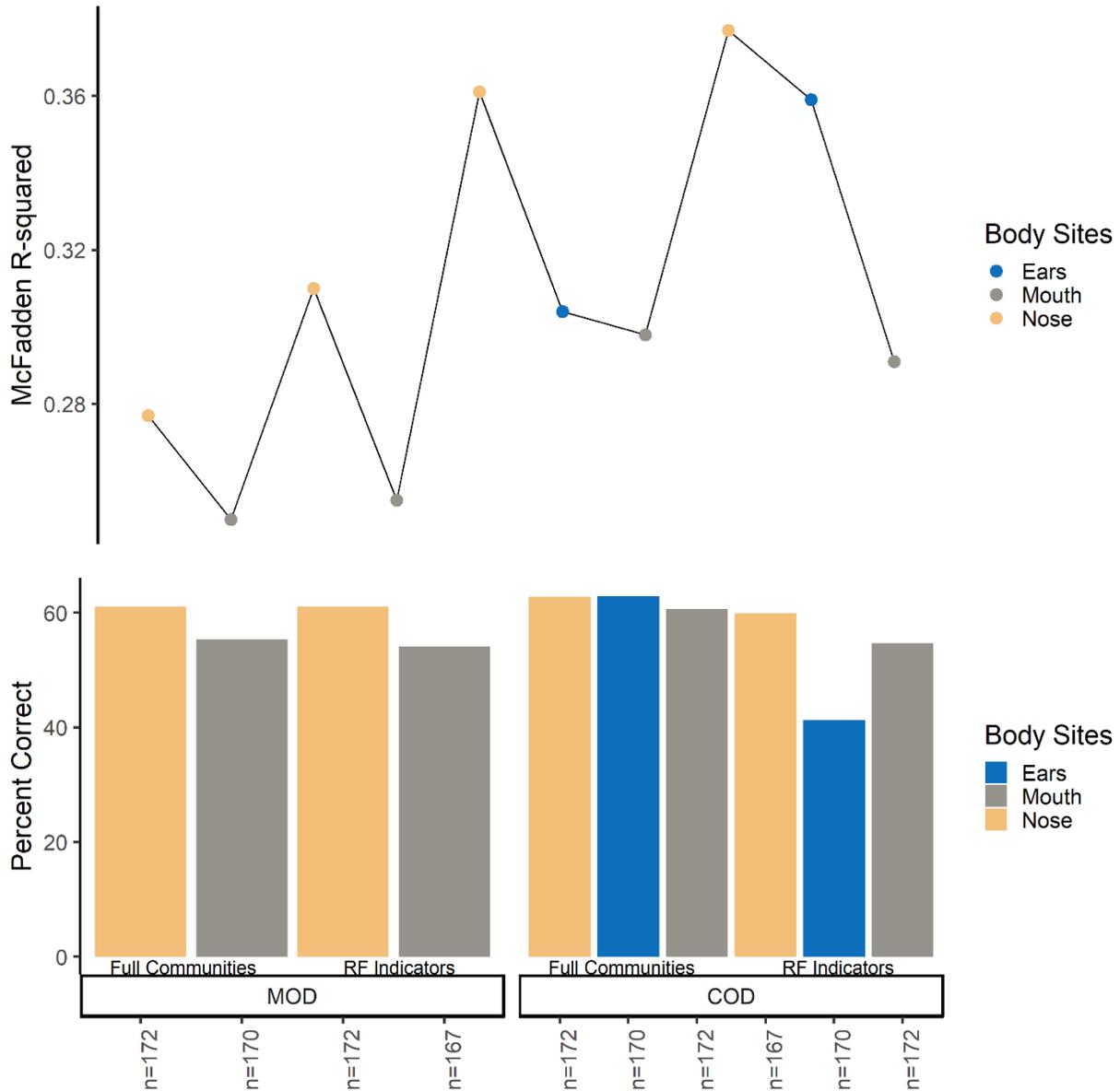
While initial multinomial logistic regression models were able to classify among all M/CODs at higher success rate than random (overall average 50.23%; random chance: MOD: 25%; COD: 16.67%), models could be improved (Table S32). Models classifying M/COD with only beta-dispersion (no case metadata) were significant ($p < 0.05$) but had low classification

success and fit (~40% average classification success; McFadden R^2 of ~ 0.0244; Table S32). Adding case metadata to the models led to better fit (~60% average classification success; McFadden R^2 : ~0.298; Table S32). However, we attempted to improve our models by testing different microbial community beta-dispersion and binary logistic regression.

Overall, we determined microbial community type (i.e., full communities, random forest indicator communities, and ‘non-core’ communities) did not improve logistic regression models (Figure 7; Table S33; Table S34). Models of non-core taxa community beta-dispersion were not significant ($p > 0.05$) and so were removed from further consideration (Table S33). Even though random forest indicator communities were specific to M/COD, associated beta-dispersion models were not more successful than full communities in classifying M/COD, and were less successful in some cases [full: 0.590(SD=0.0264); RF: 0.578(SD=0.0306); Figure 7; Table S34]. For the random forest indicator communities and full communities, models were within 7% of each other [full: 0.298(SD=0.0367); RF: 0.318 (SD=0.0445); Figure 7; Table S34]. As changing the microbial communities did not significantly improve regression models, we considered binary logistic regression.

For some M/COD comparisons (natural vs. accidental death; cardiovascular disease vs. drug-related death; disease vs. non-diseased state), binary logistic regression models performed best with an average classification success of 83.04% (Table S35). For nose communities, pairwise comparisons beta-dispersion was a significant ($p < 0.1$) covariate (full community and random forest indicator communities) included natural vs accidental death, cardiovascular disease vs. drug-use, and disease (natural deaths) vs. non-diseased (accidental, homicide, suicide) deaths (Table 7; Table S35). While random forest indicator communities compared to full communities had marginally higher successful classification (full: 78.9%; RF: 83.6%) and higher

Figure 7: **Multinomial logistic regression model comparison among full community and random forest indicator community beta-dispersion for manners/causes of death (M/CODs).** For the bottom panel, the y axis indicated percent correct, or the number of correct classifications/ total number of samples. Each bar represented a multinomial logistic model. For the top panel, the y axis indicated McFadden R².

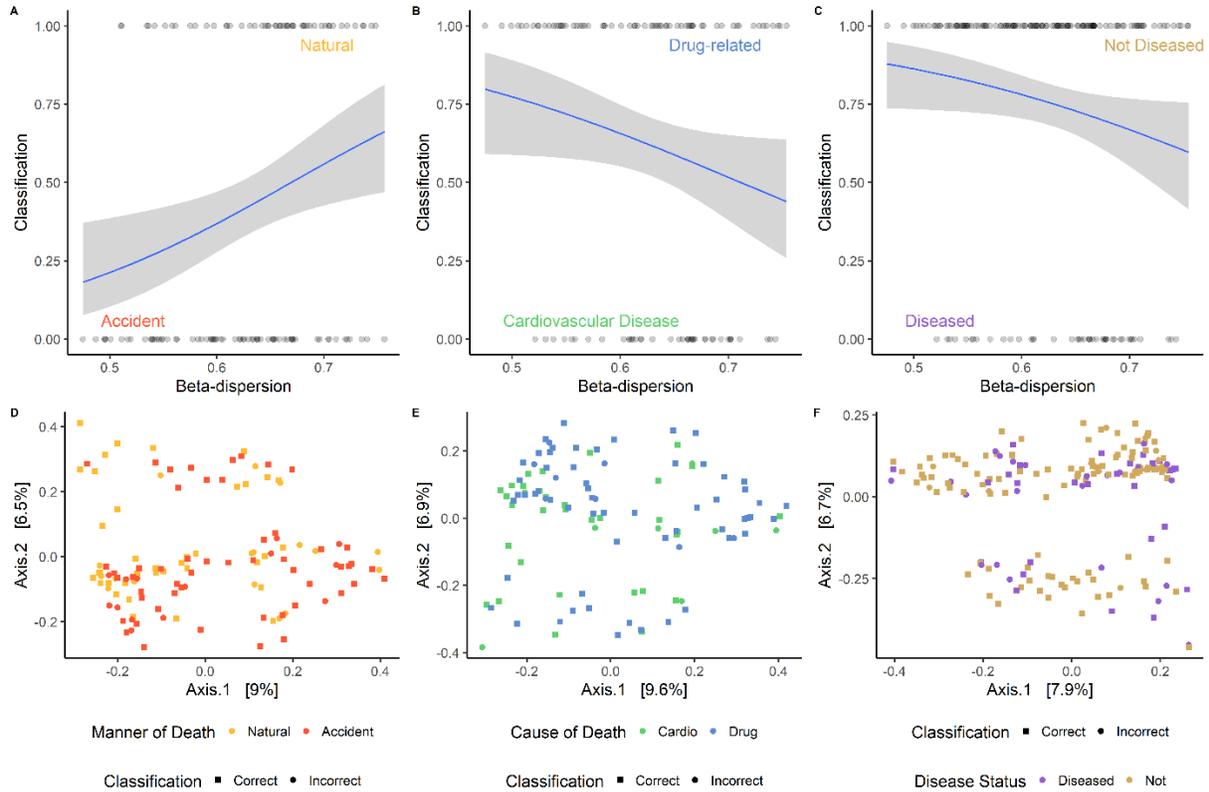


McFadden R² (full: 0.347; RF: 0.369), in all cases the sample size was smaller (Table 7; Table S35). Therefore, we considered full community beta-dispersion as the most appropriate metric

Table 7: Summary of logistic regression models classifying natural vs. accident, cardiovascular vs. drug-use, and disease vs. non-diseased state. Significance of the model was determined by the p-value. For model comparisons, the McFadden R² and percent correct (correct classifications/total number of samples) were considered within each body site.

Comparison	Beta-dispersion Community Profile	Significant Metadata	McFadden R ²	Degrees of Freedom	Chi Sq	p	Percent correct
Natural - Accident	Full community (n=120)	Race + Event Location + Age	0.314	6	51.9	0	0.783
	Random Forest Indicators (n=117)	Event Location + Age	0.388	5	62.5	0	0.829
Cardio - Drug	Full community (n=107)	BMI + Event Location + PMI + Age	0.399	5	56.9	0	0.804
	Random Forest Indicators (n=100)	BMI + Age	0.356	3	47.6	0	0.820
Disease - Non	Full community (n=172)	BMI + Race + Event Location + Age	0.328	5	70.3	0	0.779
	Random Forest Indicators (n=163)	BMI + Race + Event Location + Age	0.364	7	73.1	0	0.859

Figure 8: **Logistic regression models of the best performing pair-wise comparisons.** Nose body site was selected, as well as beta-dispersion from full communities. A-C) Logistic regression models with a 95% confidence interval of beta-dispersion on the x-axis and the binary classification on the y-axis. A) Accident: 0, Natural: 1. B) Drug-related:0, Cardiovascular disease:1. C) Disease: 0, Non-diseased: 1. D-F) Principle Coordinate Analysis (PCoA) plots of microbial samples included in the logistic regression model. Colors corresponded with the M/COD, while shape indicated if the sample was correctly classified by the model. D) Natural vs. Accidental deaths. E) Drug-related vs. cardiovascular disease deaths. F) Diseased vs. non-diseased deaths.



(Figure 8). There were no distinct clustering of samples suggesting that misclassification was randomly distributed among samples (Figure 8).

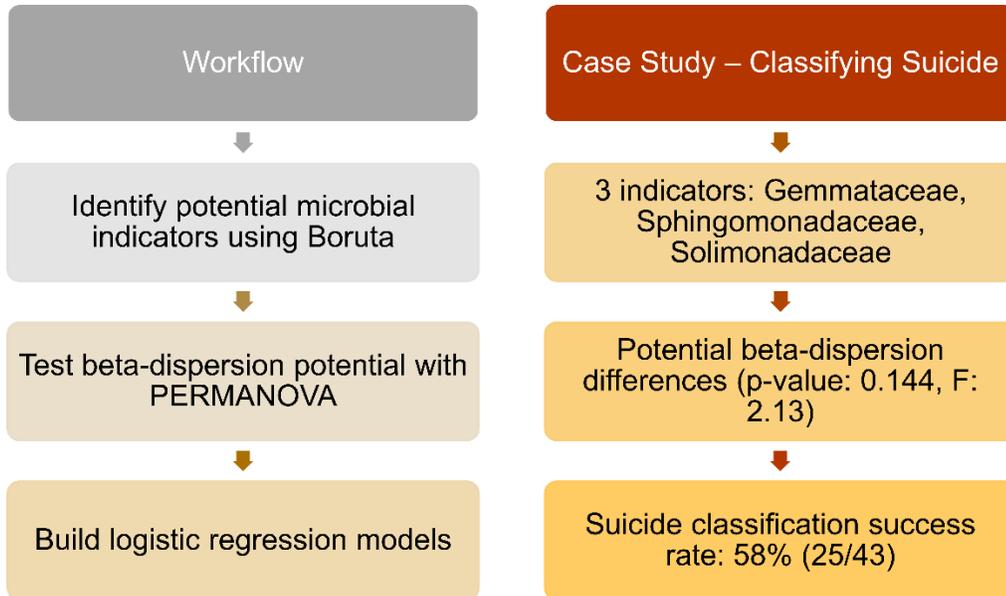
Case Studies

Of the 188 cases, we matched nose communities of 22 cases by age, sex, and race with other deaths (natural, homicide, and accidental) for a total of 43 cases (Table S26). We identified

three significant indicator taxa of suicide (Table S36). Suicide communities had higher beta-dispersion [0.659(SD=0.0433)] than non-suicides [0.654(SD=0.0427); Table S36]. Using PERMANOVA there were no significant differences in beta-dispersion among suicides ($p = 0.144$; Table S36). Using logistic regression beta-dispersion without metadata classified suicide cases with a 58% success rate (Table S36), likely associated with low power (*A priori* power sample size needed: 916). For future studies, we proposed a potential workflow using this matched-design case study for other researchers to use as a reference (Figure 9; Table S36).

Despite the low sample size, we identified ten potential indicator taxa for gunshot wound homicide vs. suicide (Boruta; Table S37). Beta-dispersion among gunshot wound homicides was significantly higher [0.626(SD=0.049)] than gunshot wound suicides [0.543(SD=0.0959); PERMANOVA permuted $p < 0.05$; Table S37]. Significant logistic regression models of gunshot wound homicides vs. suicides accurately classified MOD 93.1% of the time (Table S37), despite uneven and low sample sizes ($n=25$ homicides; $n=4$ suicides; Table S37). For significant power, an *a priori* power sample size needed for each category was 14.

Figure 9: **Proposed workflow with suicide matched-design case study.** Left column indicates potential steps researchers and practitioners can follow for future studies. Right column indicates the results from the matched-design case study, following the workflow. Twenty-three suicide cases were matched by age, sex, and race with other deaths (natural, homicide, and accidental), and nose samples were included in the analyses. Indicator taxa were identified by Boruta, while beta-dispersion was calculated using unweighted unifrac distances and tested with PERMANOVA. A logistic regression model of beta-dispersion was constructed to classify suicide vs. non-suicide deaths.



Discussion

In previous research the AKP concept was tested with distinct treatment vs. control groups based on living health conditions and in living hosts (Zaneveld, McMinds, and Thurber 2017). However, we hypothesized that using the postmortem microbiome within 48 h or death, M/CODs would associate with differential beta-dispersion, which could potentially be used as additional evidence in future forensic investigation. We hypothesized that cardiovascular disease and/or natural death would have the highest beta-dispersion, as AKP predicts with disease state in the antemortem life condition (Zaneveld, McMinds, and Thurber 2017; Barbian et al. 2015); and diversity metrics in previous studies also associated with manner and cause of death (Pechal et al. 2018). High microbiome beta-dispersion was also predicted to be related to a stressful life environment often associated with homicides, gunshot wound, and blunt force trauma deaths (Pearson et al. 2019). Using published case data from 188 postmortem microbiome autopsies our best performing binary logistic regression models were able to confirm medical examiner M/COD assessment ~79% of the time. Our best performing multinomial logistic regression models were able to confirm medical examiner M/COD assessment ~62% of the time. While better than random chance, including all M/CODs for classification with uneven sample sizes was likely resulted reduced classification success in multinomial logistic regression models.

Our dataset represents a diverse cross-section of death cases from a large metropolitan area, with multiple body sites, using targeted sequencing of the 16S rRNA gene. Cases included were predominately natural cardiovascular disease deaths and accidental drug related deaths. Therefore, direct comparison of the results of this study would be most applicable for cities with similar demographics (“U.S. Census Bureau QuickFacts: Detroit City, Michigan; Michigan”), such as Chicago (“U.S. Census Bureau QuickFacts: Chicago City, Illinois”) or Cincinnati (“U.S. Census Bureau QuickFacts: Cincinnati City, Ohio”). While the metadata lends classification

ability in multinomial logistic regression for cases in Detroit, areas with differing demographics will also need to make their own baselines. It is also important to note that beta-dispersion is calculated in reference to the samples included in the dataset (Anderson, Ellingsen, and McArdle 2006). Therefore, future work should include data and cases from other geographic areas with a range of socio-economic diversity and overall living conditions.

While we included five body sites (ears, nose, mouth, eyes, and rectum) that had differential success in classifying M/COD with beta-dispersion, there are more body sites of interest for the forensic community. As body site drives the microbial community composition more than any other factor (Pechal et al. 2018), comparisons to other body sites may be limited. For example, body sites sampled for the “thanatomicrobiome” (i.e., internal organs and blood) (Javan et al. 2016), or skin microbiome (Kodama et al. 2019) could harbor different microbial communities than the ones included in this study. Beta-dispersion among all body sites was significantly different, but mouth, nose, and ears from this data set showed the most potential for downstream forensic applications.

This work revealed that beta-dispersion has potential to inform the M/COD decision making process during death determination. Accidental deaths, which were predominately drug related deaths, had overall lower beta-dispersion than natural deaths, mirroring the dysbiosis found in non-forensic studies (Meckel and Kiraly 2019). Accidental deaths and homicides were not distinguishable by beta-dispersion. While we hypothesized that high-stress lifestyle associated with homicidal deaths would increase beta-dispersion (Pearson et al. 2019), homicides had the lowest beta-dispersion among MODs. The antemortem link of high-stress lifestyle was not as strong as antemortem disease status in this study, compared to previous results that indicated higher microbial diversity associated with neighborhood blight and vacancy (Pearson

et al. 2019). This may be because those decedents who were victims of homicide lived relatively healthy lifestyles compared to decedents with a disease status. However, we do not have access to that information, and are constrained to information in the autopsy reports.

Suicide postmortem microbiomes, while representing the lowest sample size, had similar beta-dispersion to natural deaths, similar to other, antemortem studies (Liang et al. 2018; Naseribafrouei et al. 2014). Microbiomes of suicidal people have higher diversity than healthy controls, specifically increased taxa associated with inflammation (Naseribafrouei et al. 2014). Therefore, there is a potential link between high microbial beta-dispersion and mental health that would be a promising area of future research. A previous study has documented postmortem microbiome diversity and other metrics being associated with heart disease (Pechal et al. 2018). In the current research, cardiovascular disease had significantly higher beta-dispersion than any other type of death. Dysbiosis in the microbiomes of people with cardiovascular disease has been documented, specifically as there may be a microbiome link to disease pathogenesis (Wilson Tang and Hazen 2017). Based on our results, some deaths may benefit from microbial evidence more than others. Specifically, drug-related deaths, cardiovascular disease, and suicides prompt further investigation with the postmortem microbiome.

We chose multinomial logistic regression (MLR) as a simple, easy to apply model that is often used in clinical settings (de Jong et al. 2019). However, MLR has limitations, and biases towards classifying categories with larger sample sizes (de Jong et al. 2019). We achieved marginal improvement in these models with random forest indicator taxa compared to models using the full microbial community data set. This result was not entirely unexpected, as random forest model error rates ranged from: 53.1% - 64.4% (Table S38). Random forest indicators derived from error prone random forest models did not improve multinomial regression models

classifying M/COD. This illustrates that beta-dispersion can be calculated in a variety of ways, which has downstream effects on distinguishing categories of interest. Therefore, an objective approach to selecting beta-dispersion calculation should be used as outlined in this study.

Instead, binary logistic regression models were most effective at improving model success. The categories with the highest classification success also had largest sample size (natural deaths/ accidents; cardiovascular disease/ drug-related deaths), and were highly correlated, as most natural deaths were cardiovascular disease (42/57) and most accidents were drug-related deaths (59/71). There was some overlap in pathology among cardiovascular disease deaths and drug-related deaths (Molina et al. 2020), showcasing how our best performing logistic regression models have potential applications in forensic death determination. While our case studies would benefit from further exploration with larger datasets, we provided strong evidence that other comparisons differentiating MOD, such suicide vs. non-suicides could also prove useful for forensic death determination.

This was not the first study to classify M/COD from microbial communities; however, compared to previous studies using random forest classification, logistic regression performance of beta-dispersion was similar for the same dataset (Kaszubinski et al. 2019; Zhang et al. 2019; Pechal et al. 2018). Using random forest, MOD classification success with microbial communities alone (Kaszubinski et al. 2019) was comparable to multinomial logistic regression models built with beta-dispersion alone (~40%). Adding case metadata, random forest classification accuracy was consistent with multinomial logistic regression models in two body site communities (~60%) (Zhang et al. 2019). MLR does not depend on specific indicator postmortem microbial taxa, which can change across studies (Kaszubinski et al. 2019; Pechal et al. 2018). Furthermore, we suggest that microbial community information, either taxon

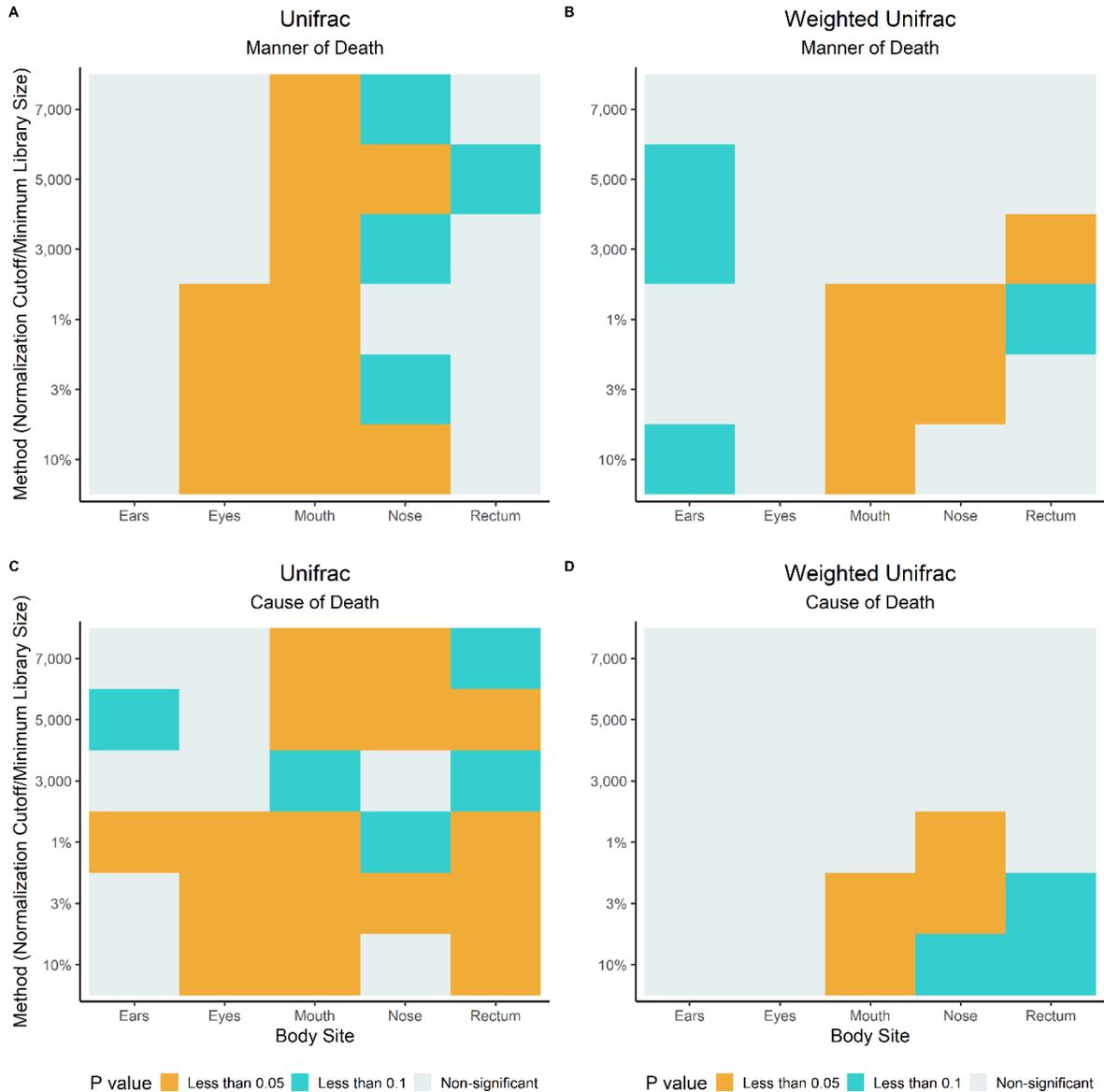
dependent (e.g., indicator taxa) or not (e.g., beta-dispersion) could be an additional piece of evidence in M/COD determination. By including metadata into our models, successful classification was improved rather than using microbial data alone, something to consider in future death investigation. Indeed, to our knowledge case metadata have never been modeled or used alone for M/COD determination.

Conclusion

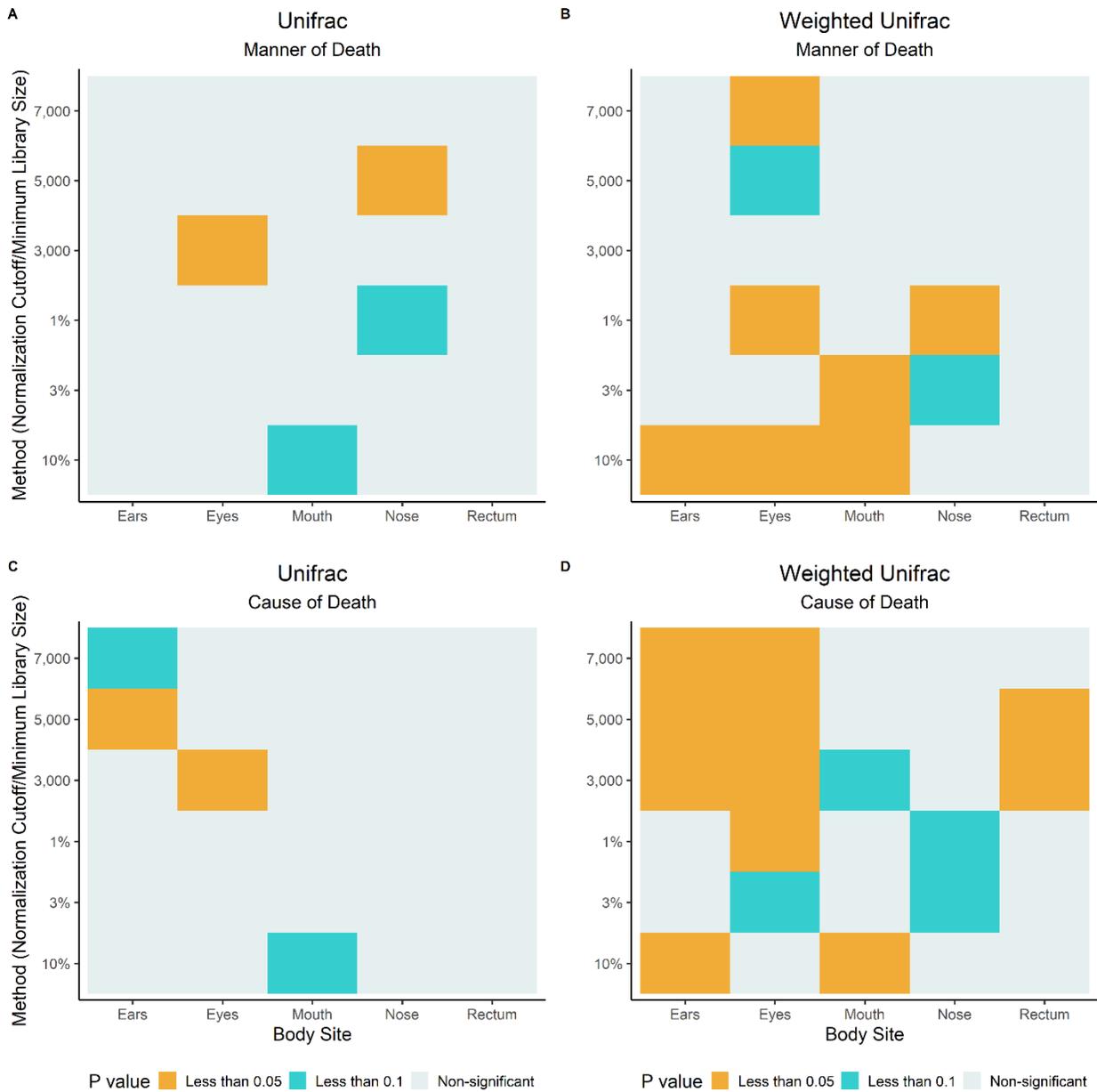
Microbial community metrics, such as beta-dispersion, have potential forensic use in contributing to classification of M/COD during death investigation. This reflection is due to the antemortem link to the postmortem microbiome. We showed beta-dispersion increased based on disease status (cardiovascular disease) according to AKP, and beta-dispersion reflected M/COD, especially for cardiovascular disease and drug related deaths. While random forest is a useful tool for these types of datasets, MLR produced comparable results without reliance on specific indicator taxa. Furthermore, using two case studies we demonstrated circumstances where beta-dispersion could be used to distinguish MOD; however, low and uneven sample size was an issue for all case studies. Despite the reduced power of these case studies, this workflow may be useful for other forensic practitioners to test within their own sample set, that encompass new locations and metadata, to strengthen the antemortem link to the postmortem microbiome. The methods outlined in this study serve as a guide to developing non-taxonomic microbiome tools for other researchers and medical examiners in other geographic locations and investigation contexts. Ultimately, modeling beta-dispersion with case metadata is a tool that could be useful for medical examiners during death investigation, to combine with other methods of M/COD determination.

APPENDIX

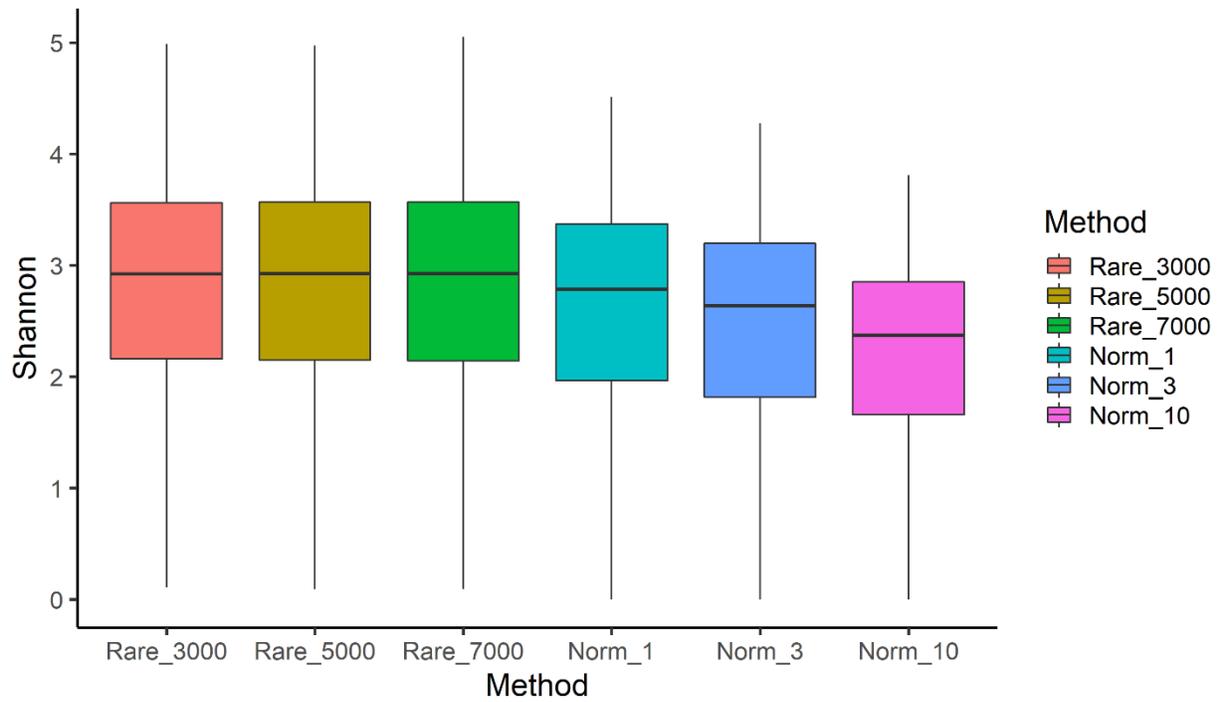
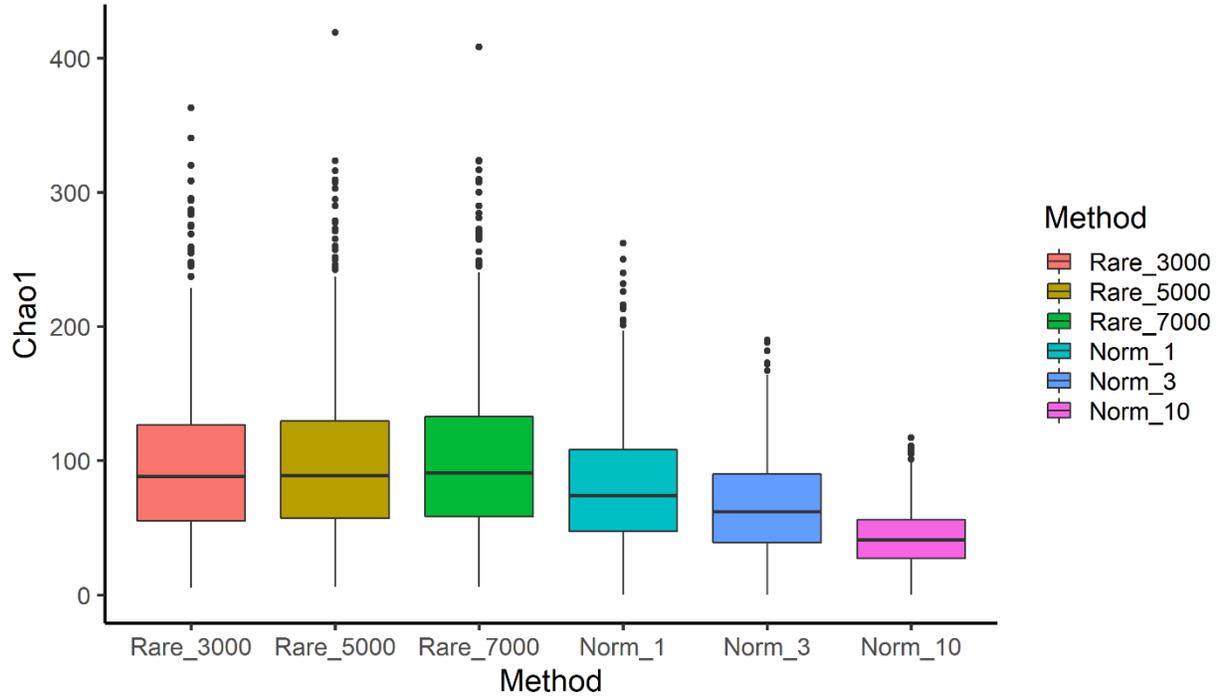
Supplemental Figure 1: **Summary of Kruskal-Wallis results for normalization strategies determining beta-dispersion.** Each body site is represented across the x-axes. Each level for the corresponding normalization strategy (percent cutoff and minimum library sizes) are on the y-axes. Each box is indicative of the results of a Kruskal-Wallis test for either manner of death (A and B) or cause of death (C and D). Left-hand side of the figure includes unweighted unifrac distances (A and C), while the right-hand side includes weighted unifrac distances (B and D). Significant ($p < 0.05$) Kruskal-Wallis tests are indicated by gold, while nearly significant ($p < 0.1$) are indicated by blue.



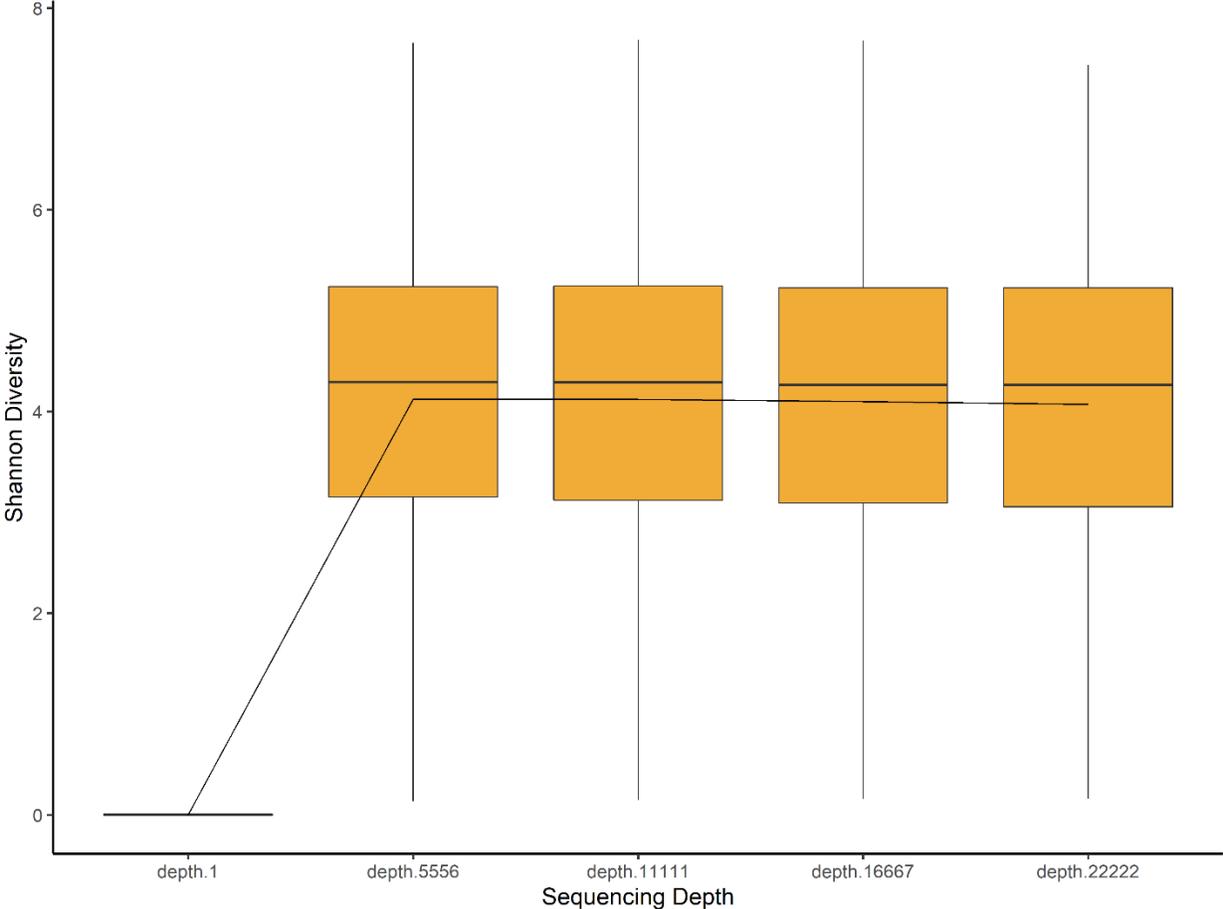
Supplemental Figure 2: **Summary of Fligner-Killeen results for normalization strategies determining beta-dispersion.** Each body site is represented across the x-axes. Each level for the corresponding normalization strategy (percent cutoff and minimum library sizes) are on the y-axes. Each box is indicative of the results of a Fligner-Killeen test for either manner of death (A and B) or cause of death (C and D). Left-hand side of the figure includes unweighted unifracs (A and C), while the right-hand side includes weighted unifracs (B and D). Significant ($p < 0.05$) Fligner-Killeen tests are indicated by gold, while nearly significant ($p < 0.1$) are indicated by blue.



Supplemental Figure 3: **Alpha-diversity metrics across normalization strategies and levels.** Chao1 (richness) and Shannon diversity (richness and evenness) are reported. A significant decrease in alpha-diversity was found among percent cutoffs, and compared to minimum library sizes (Kruskal-Wallis and *post hoc* Nemenyi $p < 0.05$).



Supplemental Figure 4: **Shannon diversity across sequencing depth among microbial samples.**



REFERENCES

REFERENCES

- Anderson, Marti J. 2017. “Permutational Multivariate Analysis of Variance (PERMANOVA).” In *Wiley StatsRef: Statistics Reference Online*, 1–15. Chichester, UK: John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118445112.stat07841>.
- Anderson, Marti J., Kari E. Ellingsen, and Brian H. McArdle. 2006. “Multivariate Dispersion as a Measure of Beta Diversity.” *Ecology Letters* 9 (6): 683–93. <https://doi.org/10.1111/j.1461-0248.2006.00926.x>.
- Barbian, Hannah J., Yingying Li, Miguel Ramirez, Zachary Klase, Iddi Lipende, Deus Mjungu, AH. Moeller, et al. 2015. “Destabilization of the Gut Microbiome Marks the End-Stage of Simian Immunodeficiency Virus Infection in Wild Chimpanzees.” *American Journal of Primatology* 80 (1): 1–29. <https://doi.org/10.1002/ajp.22515>.
- Bates, Douglas, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai, et al. 2019. “Package ‘lme4.’”
- Benbow, Mark Eric, Jennifer L. Pechal, Jennifer M. Lang, Racheal Erb, and John R. Wallace. 2015. “The Potential of High-Throughput Metagenomic Sequencing of Aquatic Bacterial Communities to Estimate the Postmortem Submersion Interval.” *Journal of Forensic Sciences* 60 (6): 1500–1510. <https://doi.org/10.1111/1556-4029.12859>.
- Böhning, Dankmar. 1992. “Multinomial Logistic Regression Algorithm.” *Annals of the Institute of Statistical Mathematics* 44 (1): 197–200. <https://doi.org/10.1007/BF00048682>.
- Bokulich, Nicholas A., Jai Ram Rideout, William G. Mercurio, Arron Shiffer, Benjamin Wolfe, Corinne F. Maurice, Rachel J. Dutton, Peter J. Turnbaugh, Rob Knight, and J. Gregory Caporaso. 2016. “Mockrobiota: A Public Resource for Microbiome Bioinformatics Benchmarking.” *MSystems* 1 (5). <https://doi.org/10.1128/msystems.00062-16>.
- Bolyen, Evan, Jai Ram Rideout, Matthew R. Dillon, Nicholas A. Bokulich, Christian C. Abnet, Gabriel A. Al-Ghalith, Harriet Alexander, et al. 2019. “Reproducible, Interactive, Scalable and Extensible Microbiome Data Science Using QIIME 2.” *Nature Biotechnology* 37 (8): 852–57. <https://doi.org/10.1038/s41587-019-0209-9>.
- Bozdogan, Hamparsum. 1987. “Model Selection and Akaike’s Information Criterion (AIC): The General Theory and Its Analytical Extensions.” *Psychometrika* 52 (3): 345–70. <https://doi.org/10.1007/BF02294361>.
- Breiman, Leo, Adele Cutler, Andy Liaw, and Matthew Wiener. 2018. “Package ‘RandomForest.’” 1–29. <https://doi.org/10.1023/A:1010933404324>.
- Budowle, Bruce, Steven E. Schutzer, Anja Einseln, Lynda C. Kelley, Anne C. Walsh, Jenifer A.L. Smith, Babetta L. Marrone, James Robertson, and Joseph Campos. 2003. “Building

- Microbial Forensics as a Response to Bioterrorism.” *Science* 301 (5641): 1852–53.
<https://doi.org/10.1126/science.1090083>.
- Callahan, Benjamin J., Paul J. McMurdie, Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, and Susan P. Holmes. 2016. “DADA2: High-Resolution Sample Inference from Illumina Amplicon Data.” *Nature Methods* 13 (7): 581–83.
<https://doi.org/10.1038/nmeth.3869>.
- Caporaso, J. Gregory, Rob Knight, and Scott T. Kelley. 2011. “Host-Associated and Free-Living Phage Communities Differ Profoundly in Phylogenetic Composition.” *PLoS ONE* 6 (2).
<https://doi.org/10.1371/journal.pone.0016900>.
- Caporaso, J. Gregory, Justin Kuczynski, Jesse Stombaugh, Kyle Bittinger, Frederic D. Bushman, Elizabeth K. Costello, Noah Fierer, et al. 2010. “QIIME Allows Analysis of High-Throughput Community Sequencing Data.” *Nature Methods*.
<https://doi.org/10.1038/nmeth.f.303>.
- Caporaso, J. Gregory, Christian L. Lauber, Elizabeth K. Costello, Donna Berg-Lyons, Antonio Gonzalez, Jesse Stombaugh, Dan Knights, et al. 2011. “Moving Pictures of the Human Microbiome.” *Genome Biology* 12 (5): R50. <https://doi.org/10.1186/gb-2011-12-5-r50>.
- Caporaso, J. Gregory, Christian L. Lauber, William A. Walters, Donna Berg-Lyons, James Huntley, Noah Fierer, Sarah M. Owens, et al. 2012. “Ultra-High-Throughput Microbial Community Analysis on the Illumina HiSeq and MiSeq Platforms.” *ISME Journal* 6 (8): 1621–24. <https://doi.org/10.1038/ismej.2012.8>.
- Carter, David O., Jeffery K. Tomberlin, M. Eric. Benbow, and Jessica L. Metcalf. 2016. *Forensic Microbiology*. John Wiley & Sons, Incorporated.
- Clarke, Thomas H., Andres Gomez, Harinder Singh, Karen E. Nelson, and Lauren M. Brinkac. 2017. “Integrating the Microbiome as a Resource in the Forensics Toolkit.” *Forensic Science International: Genetics* 30 (2017): 141–47.
<https://doi.org/10.1016/j.fsigen.2017.06.008>.
- Croissant, Yves. 2019. “Package ‘Mlogit,’” 1–40. <https://doi.org/10.1017/CBO9780511805271>.
- D’Argenio, Valeria, Giorgio Casaburi, Vincenza Precone, and Francesco Salvatore. 2014. “Comparative Metagenomic Analysis of Human Gut Microbiome Composition Using Two Different Bioinformatic Pipelines.” *Biomed Res Int* 2014: 325340.
<https://doi.org/10.1155/2014/325340>.
- DeBruyn, Jennifer M., and Kathleen A. Hauther. 2017. “Postmortem Succession of Gut Microbial Communities in Deceased Human Subjects.” *PeerJ*.
<https://doi.org/10.7717/peerj.3437>.
- Dobay, Akos, Cordula Haas, Geoffrey Fucile, Nora Downey, Hilary G. Morrison, Adelgunde

- Kratzer, and Natasha Arora. 2019. “Microbiome-Based Body Fluid Identification of Samples Exposed to Indoor Conditions.” *Forensic Science International: Genetics* 40 (May): 105–13. <https://doi.org/10.1016/j.fsigen.2019.02.010>.
- Edgar, Robert C. 2010. “Search and Clustering Orders of Magnitude Faster than BLAST.” *Bioinformatics* 26 (19): 2460–61. <https://doi.org/10.1093/bioinformatics/btq461>.
- Faul, Franz, Edgar Erdfelder, Albert Georg Lang, and Axel Buchner. 2007. “G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences.” In *Behavior Research Methods*, 39:175–91. Psychonomic Society Inc. <https://doi.org/10.3758/BF03193146>.
- Glass, Elizabeth M., Jared Wilkening, Andreas Wilke, Dionysios Antonopoulos, and Folker Meyer. 2010. “Using the Metagenomics RAST Server (MG-RAST) for Analyzing Shotgun Metagenomes.” *Cold Spring Harbor Protocols* 5 (1): pdb.prot5368. <https://doi.org/10.1101/pdb.prot5368>.
- Golob, Jonathan L., Elisa Margolis, Noah G. Hoffman, and David N. Fredricks. 2017. “Evaluating the Accuracy of Amplicon-Based Microbiome Computational Pipelines on Simulated Human Gut Microbial Communities.” *BMC Bioinformatics* 18 (1): 1–12. <https://doi.org/10.1186/s12859-017-1690-0>.
- Hyde, Embriette R., Daniel P. Haarmann, Aaron M. Lynne, Sibyl R. Bucheli, and Joseph F. Petrosino. 2013. “The Living Dead: Bacterial Community Structure of a Cadaver at the Onset and End of the Bloat Stage of Decomposition.” Edited by Corrie S. Moreau. *PLoS ONE* 8 (10): e77733. <https://doi.org/10.1371/journal.pone.0077733>.
- Ioannidis, John P. A. 2005. “Why Most Published Research Findings Are False.” *PLoS Medicine* 2 (8): e124. <https://doi.org/10.1371/journal.pmed.0020124>.
- Javan, Gulnaz T., Sheree J. Finley, Ismail Can, Jeremy E. Wilkinson, J. Delton Hanson, and Aaron M. Tarone. 2016. “Human Thanatobiome Succession and Time since Death.” *Scientific Reports*. <https://doi.org/10.1038/srep29598>.
- Javan, Gulnaz T., Sheree J. Finley, Tasia Smith, Joselyn Miller, and Jeremy E. Wilkinson. 2017. “Cadaver Thanatobiome Signatures: The Ubiquitous Nature of Clostridium Species in Human Decomposition.” *Frontiers in Microbiology*. <https://doi.org/10.3389/fmicb.2017.02096>.
- Johnson, Hunter R., Donovan D. Trinidad, Stephania Guzman, Zenab Khan, James V. Parziale, Jennifer M. DeBruyn, and Nathan H. Lents. 2016. “A Machine Learning Approach for Using the Postmortem Skin Microbiome to Estimate the Postmortem Interval.” *PLoS ONE* 11 (12): 1–23. <https://doi.org/10.1371/journal.pone.0167370>.
- Jong, Valentijn M.T. de, Marinus J.C. Eijkemans, Ben van Calster, Dirk Timmerman, Karel G.M. Moons, Ewout W. Steyerberg, and Maarten van Smeden. 2019. “Sample Size

- Considerations and Predictive Performance of Multinomial Logistic Prediction Models.” *Statistics in Medicine* 38 (9): 1601–19. <https://doi.org/10.1002/sim.8063>.
- Kaszubinski, Sierra F, Jennifer L Pechal, Carl J Schmidt, Jordan R Heather, M Eric Benbow, and Mariah H Meek. 2019. “Evaluating Bioinformatic Pipeline Performance for Forensic Microbiome Analysis.” *Journal of Forensic Sciences*, 1–13. <https://doi.org/10.1111/1556-4029.14213>.
- Katoh, Kazutaka, and Daron M. Standley. 2013. “MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability.” *Molecular Biology and Evolution* 30 (4): 772–80. <https://doi.org/10.1093/molbev/mst010>.
- Keegan, Kevin P., Elizabeth M. Glass, and Folker Meyer. 2016. “MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function.” In *Methods in Molecular Biology*, 1399:207–33. Humana Press Inc. https://doi.org/10.1007/978-1-4939-3369-3_13.
- Kodama, Whitney A., Zhenjiang Xu, Jessica L. Metcalf, Se Jin Song, Nicholas Harrison, Rob Knight, David O. Carter, and Christopher B. Happy. 2019. “Trace Evidence Potential in Postmortem Skin Microbiomes: From Death Scene to Morgue.” *Journal of Forensic Sciences* 64 (3): 791–98. <https://doi.org/10.1111/1556-4029.13949>.
- Kozich, James J., Sarah L. Westcott, Nielson T. Baxter, Sarah K. Highlander, and Patrick D. Schloss. 2013. “Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform.” *Applied and Environmental Microbiology* 79 (17): 5112–20. <https://doi.org/10.1128/AEM.01043-13>.
- Kursa, Miron Bartosz, and Witold Remigiusz Rudnicki. 2018. “Package ‘Boruta.’” <https://notabug.org/mbq/Boruta/>.
- Leipzig, Jeremy. 2017. “A Review of Bioinformatic Pipeline Frameworks.” *Briefings in Bioinformatics* 18 (3): 530–36. <https://doi.org/10.1093/bib/bbw020>.
- Liang, Shan, Xiaoli Wu, Xu Hu, Tao Wang, and Feng Jin. 2018. “Recognizing Depression from the Microbiota–Gut–Brain Axis.” *International Journal of Molecular Sciences*. MDPI AG. <https://doi.org/10.3390/ijms19061592>.
- Mandal, Siddhartha, Will Van Treuren, Richard A. White, Merete Eggesbø, Rob Knight, and Shyamal D. Peddada. 2015. “Analysis of Composition of Microbiomes: A Novel Method for Studying Microbial Composition.” *Microbial Ecology in Health & Disease* 26 (0). <https://doi.org/10.3402/mehd.v26.27663>.
- McDonald, Daniel, Jose C. Clemente, Justin Kuczynski, Jai R. Rideout, Jesse Stombaugh, Doug Wendel, Andreas Wilke, et al. 2012. “The Biological Observation Matrix (BIOM) Format or: How I Learned to Stop Worrying and Love the Ome-Ome.” *GigaScience*. Oxford

- University Press. <https://doi.org/10.1186/2047-217X-1-7>.
- McMurdie, Paul J., and Susan Holmes. 2013. "Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data." *PLoS ONE* 8 (4). <https://doi.org/10.1371/journal.pone.0061217>.
- McMurdie, Paul J., and Susan Holmes. 2014. "Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible." *PLoS Computational Biology* 10 (4). <https://doi.org/10.1371/journal.pcbi.1003531>.
- Meckel, Katherine R., and Drew D. Kiraly. 2019. "A Potential Role for the Gut Microbiome in Substance Use Disorders." *Psychopharmacology*. Springer Verlag. <https://doi.org/10.1007/s00213-019-05232-0>.
- Metcalf, Jessica L., Laura Wegener Parfrey, Antonio Gonzalez, Christian L. Lauber, Dan Knights, Gail Ackermann, Gregory C. Humphrey, et al. 2013. "A Microbial Clock Provides an Accurate Estimate of the Postmortem Interval in a Mouse Model System." *ELife* 2013 (2): 1–19. <https://doi.org/10.7554/eLife.01104.001>.
- Metcalf, Jessica L., Zhenjiang Z. Xu, Amina Bouslimani, Pieter Dorrestein, David O. Carter, and Rob Knight. 2017. "Microbiome Tools for Forensic Science." *Trends in Biotechnology* 35 (9): 814–23. <https://doi.org/10.1016/j.tibtech.2017.03.006>.
- Metcalf, Jessica L., Zhenjiang Zech Xu, Sophie Weiss, Simon Lax, Will Van Treuren, Embriette R Hyde, Se Jin Song, et al. 2016. "Microbial Community Assembly and Metabolic Function during Mammalian Corpse Decomposition." *Science* 351 (6269): 158–62. <https://doi.org/10.1126/science.aad2646>.
- Molina, D. Kimberly, Kathryn Vance, Maci L. Coleman, Veronica M. Hangrove. 2020. "Testing an Age-old Adage: Can Autopsy Findings be of Assistance in Differentiating Opioid Versus Cardiac Deaths?" *Journal of Forensic Sciences* 65 (1):112–6. <https://onlinelibrary.wiley.com/doi/abs/10.1111/1556-4029.14174>
- Mysara, Mohamed, Mercy Njima, Natalie Leys, Jeroen Raes, and Pieter Monsieurs. 2017. "From Reads to Operational Taxonomic Units: An Ensemble Processing Pipeline for MiSeq Amplicon Sequencing Data." *GigaScience* 6 (2): 1–10. <https://doi.org/10.1093/gigascience/giw017>.
- Naseribafrouei, A., K. Hestad, E. Avershina, M. Sekelja, A. Linløkken, R. Wilson, and K. Rudi. 2014. "Correlation between the Human Fecal Microbiota and Depression." *Neurogastroenterology & Motility* 26 (8): 1155–62. <https://doi.org/10.1111/nmo.12378>.
- Nilakanta, Haema, Kimberly L. Drews, Suzanne Firrell, Mary A. Foulkes, and Kathleen A. Jablonski. 2014. "A Review of Software for Analyzing Molecular Sequences." *BMC Research Notes* 7 (1). <https://doi.org/10.1186/1756-0500-7-830>.

- Oksanen, Jari, F Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan Mcglinn, Peter R Minchin, et al. 2019. "Vegan: Community Ecology Package."
- Pearson, Amber L., Amanda Rzotkiewicz, Jennifer L. Pechal, Carl J. Schmidt, Heather R. Jordan, Adam Zwickle, and M. Eric Benbow. 2019. "Initial Evidence of the Relationships between the Human Postmortem Microbiome and Neighborhood Blight and Greening Efforts." *Annals of the American Association of Geographers* 109 (3): 958–78. <https://doi.org/10.1080/24694452.2018.1519407>.
- Pechal, Jennifer L., Tawni L. Crippen, M. Eric Benbow, Aaron M. Tarone, Scot Dowd, and Jeffery K. Tomberlin. 2014. "The Potential Use of Bacterial Community Succession in Forensics as Described by High Throughput Metagenomic Sequencing." *International Journal of Legal Medicine* 128 (1): 193–205. <https://doi.org/10.1007/s00414-013-0872-1>.
- Pechal, Jennifer L., Carl J. Schmidt, Heather R. Jordan, and M. Eric Benbow. 2017. "Frozen: Thawing and Its Effect on the Postmortem Microbiome in Two Pediatric Cases." *Journal of Forensic Sciences* 62 (5): 1399–1405. <https://doi.org/10.1111/1556-4029.13419>.
- Pechal, Jennifer L., Carl J. Schmidt, Heather R. Jordan, and M. Eric Benbow. 2018. "A Large-Scale Survey of the Postmortem Human Microbiome, and Its Potential to Provide Insight into the Living Health Condition." *Scientific Reports* 8 (1): 1–15. <https://doi.org/10.1038/s41598-018-23989-w>.
- Plummer, Erica, and Jimmy Twin. 2015. "A Comparison of Three Bioinformatics Pipelines for the Analysis of Preterm Gut Microbiota Using 16S rRNA Gene Sequencing Data." *Journal of Proteomics & Bioinformatics* 8 (12): 283–91. <https://doi.org/10.4172/jpb.1000381>.
- Pohlert, Thorsten. 2018. "PMCMR: Calculate Pairwise Multiple Comparisons of Mean Rank Sums."
- Price, Morgan N, Paramvir S Dehal, and Adam P Arkin. 2010. "FastTree 2--Approximately Maximum-Likelihood Trees for Large Alignments." *PloS One* 5 (3). <https://doi.org/10.1371/journal.pone.0009490>.
- Quast, Christian, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, and Frank Oliver Glöckner. 2013. "The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools." *Nucleic Acids Research* 41 (January). <https://doi.org/10.1093/nar/gks1219>.
- Randy, Hanzlick, John C. Hunsaker III, and Gregory J. Davis. 2002. "A Guide for Manner of Death Classification."
- Rognes, Torbjørn, Tomáš Flouri, Ben Nichols, Christopher Quince, and Frédéric Mahé. 2016. "VSEARCH: A Versatile Open Source Tool for Metagenomics." *PeerJ*, 1–22. <https://doi.org/10.7717/peerj.2584>.

- Rosa, Patricio S. Ia, J. Paul Brooks, Elena Deych, Edward L. Boone, David J. Edwards, Qin Wang, Erica Sodergren, George Weinstock, and William D. Shannon. 2012. "Hypothesis Testing and Power Calculations for Taxonomic-Based Human Microbiome Data." *PLoS ONE* 7 (12): e52078. <https://doi.org/10.1371/journal.pone.0052078>.
- Schloss, Patrick D., Sarah L. Westcott, Thomas Ryabin, Justine R. Hall, Martin Hartmann, Emily B. Hollister, Ryan A. Lesniewski, et al. 2009. "Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities." *Applied and Environmental Microbiology* 75 (23): 7537–41. <https://doi.org/10.1128/AEM.01541-09>.
- Schmedes, Sarah E., Antti Sajantila, and Bruce Budowle. 2016. "Expansion of Microbial Forensics." *Journal of Clinical Microbiology* 54 (8): 1964–74. <https://doi.org/10.1128/JCM.00046-16>.
- Schmedes, Sarah E., August E. Woerner, and Bruce Budowle. 2017. "Forensic Human Identification Using Skin Microbiomes." *Applied and Environmental Microbiology* 83 (22). <https://doi.org/10.1128/AEM.01672-17>.
- Shade, Ashley, and Jo Handelsman. 2012. "Beyond the Venn Diagram: The Hunt for a Core Microbiome." *Environmental Microbiology* 14 (1): 4–12. <https://doi.org/10.1111/j.1462-2920.2011.02585.x>.
- Siegwald, Léa, Hélène Touzet, Yves Lemoine, David Hot, Christophe Audebert, and Ségolène Caboche. 2017. "Assessment of Common and Emerging Bioinformatics Pipelines for Targeted Metagenomics." *PLoS ONE* 12 (1): 1–26. <https://doi.org/10.1371/journal.pone.0169563>.
- Sivarajah, Uthayasankar, Muhammad Mustafa Kamal, Zahir Irani, and Vishanth Weerakkody. 2017. "Critical Analysis of Big Data Challenges and Analytical Methods." *Journal of Business Research* 70 (January): 263–86. <https://doi.org/10.1016/j.jbusres.2016.08.001>.
- Team, R Core. "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing. Accessed October 23, 2019. <https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing>.
- Turnbaugh, Peter J., Ruth E. Ley, Micah Hamady, Claire M. Fraser-Liggett, Rob Knight, and Jeffrey I. Gordon. 2007. "The Human Microbiome Project." *Nature*. Nature Publishing Group. <https://doi.org/10.1038/nature06244>.
- "U.S. Census Bureau QuickFacts: Chicago City, Illinois." <https://www.census.gov/quickfacts/chicagocityillinois>.
- "U.S. Census Bureau QuickFacts: Cincinnati City, Ohio." <https://www.census.gov/quickfacts/cincinnati-city-ohio>.

“U.S. Census Bureau QuickFacts: Detroit City, Michigan; Michigan”

<https://www.census.gov/quickfacts/fact/table/detroitcitymichigan,MI/PST045219>.

Weiss, Sophie, Zhenjiang Zech Xu, Shyamal Peddada, Amnon Amir, Kyle Bittinger, Antonio Gonzalez, Catherine Lozupone, et al. 2017. “Normalization and Microbial Differential Abundance Strategies Depend upon Data Characteristics.” *Microbiome* 5 (1): 1–18. <https://doi.org/10.1186/s40168-017-0237-y>.

Wilson Tang, W. H., and Stanley L. Hazen. 2017. “The Gut Microbiome and Its Role in Cardiovascular Diseases.” *Circulation* 135 (11): 1008–10. <https://doi.org/10.1161/CIRCULATIONAHA.116.024251>.

Zaneveld, Jesse R., Ryan McMinds, and Rebecca Vega Thurber. 2017. “Stress and Stability: Applying the Anna Karenina Principle to Animal Microbiomes.” *Nature Microbiology* 2 (August): 1–8. <https://doi.org/10.1038/nmicrobiol.2017.121>.

Zhang, Yu, Jennifer L. Pechal, Carl J. Schmidt, Heather R. Jordan, Wesley W. Wang, M. Eric Benbow, Sing Hoi Sze, and Aaron M. Tarone. 2019. “Machine Learning Performance in a Microbial Molecular Autopsy Context: A Cross-Sectional Postmortem Human Population Study.” *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0213829>.