

PAPER**PATHOLOGY/BIOLOGY**

Sierra F. Kaszubinski ¹ B.Sc.; Jennifer L. Pechal ² Ph.D.; Carl J. Schmidt,^{3,4} M.D.; Heather R. Jordan ⁵ Ph.D.; Mark E. Benbow ^{2,6,7} Ph.D.; and Mariah H. Meek ^{1,7,8} Ph.D.

Evaluating Bioinformatic Pipeline Performance for Forensic Microbiome Analysis^{*,†,‡}

ABSTRACT: Microbial communities have potential evidential utility for forensic applications. However, bioinformatic analysis of high-throughput sequencing data varies widely among laboratories. These differences can potentially affect microbial community composition and downstream analyses. To illustrate the importance of standardizing methodology, we compared analyses of postmortem microbiome samples using several bioinformatic pipelines, varying minimum library size or minimum number of sequences per sample, and sample size. Using the same input sequence data, we found that three open-source bioinformatic pipelines, MG-RAST, mothur, and QIIME2, had significant differences in relative abundance, alpha-diversity, and beta-diversity, despite the same input data. 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 affect results in important ways. Given the growing potential application of forensic microbiology to the criminal justice system, continued research on standardizing computational methodology will be important for downstream applications.

KEYWORDS: forensic science, bioinformatic pipelines, forensic microbiology, postmortem microbiome, microbial communities, next-generation sequencing

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 (1). 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) that vary over time, such as during decomposition (2,3).

Next-generation sequencing has expanded microbial forensics to many potential applications including body fluid identification, human identification, and postmortem interval estimation (4). Targeted amplicon sequencing of 16S rRNA identified potential microbial biomarkers for sensitive body fluid identification (5), and a clade-based, single nucleotide polymorphism approach identified human individuals using their skin microbiome (6). 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 (7), humans (2), and grave soil (3), as well as experiments in anthropologic facilities (8) and during routine autopsy for death investigation (2). Researchers developed models for forensic applications (i.e., postmortem interval estimation) and described microbial community function and succession during decomposition (2,3,7). While recent research suggests excellent potential for microbial community use in forensics (9–12), 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 (13). Forensic microbiological data created by high-

¹Department of Integrative Biology, Michigan State University, East Lansing, MI, 48824.

²Department of Entomology, Michigan State University, East Lansing, MI, 48824.

³Wayne County Medical Examiner's Office, Detroit, MI, 48207.

⁴Department of Pathology, University of Michigan, Ann Arbor, MI, 48109.

⁵Department of Biological Sciences, Mississippi State University, Mississippi State, MS, 39762.

⁶Ecology, Evolutionary Biology and Behavior Program, Michigan State University, East Lansing, MI, 48824.

⁷Michigan Department of Osteopathic Medical Specialties, Michigan State University, East Lansing, MI, 48824.

⁸AgBio Research, Michigan State University, East Lansing, MI 48824.

Corresponding author: Mark E. Benbow, Ph.D. E-mail: benbow@msu.edu

*Presented in part at the 71st Annual Scientific Meeting of the American Academy of Forensic Sciences, February 18–23, 2019, in Baltimore, MD.

†This research was funded by a grant from the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice, awarded (2014-DN-BX-K008) to J.L.P., C.J.S., H.R.J., and M.E.B. S.F.K. is funded by the Department of Defense SMART Scholarship program.

‡Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice or Department of Defense.

Received 19 June 2019; and in revised form 1 Sept. 2019; accepted 19 Sept. 2019.

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 (14,15). For any downstream application of forensic microbiology in the criminal justice system, there is a need for streamlined standard operating procedures (SOPs) (11).

There are self-contained pipelines for processing sequence data and characterizing microbial communities. Three of the most often used self-contained pipelines are QIIME2 (16), mothur (17), and MG-RAST (18). QIIME2 and mothur require some command line experience, whereas MG-RAST has a Web-based graphical user interface (GUI) (19). Several studies determined that the microbial communities generated from different pipelines were comparable; however, these studies used simulated data (15,20–24) or small sample sizes ($n < 40$) (19,23–25) or were composed of the same sample type (i.e., human gut microbial data) (19,25), which does not readily apply to many forensic datasets that can include highly variable sampling locations or contextual circumstances (2,3,7).

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 (23). The common library size normalization procedure for microbial analysis is referred to as rarefying (23). 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 (23). Minimum library size is often chosen by selecting the smallest library size of a nondefective sample, which is a subjective assessment that can add uncertainty to microbial community analysis (23). Minimum library size can also be chosen based on rarefaction curves: taxon-based resampling 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 (23). 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 (2). 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. (2) characterized and modeled postmortem microbiomes using 188 autopsies, sample sizes for postmortem microbiome studies ranged from 2 to 48 (3,26). Small sample sizes decrease statistical power and inflate effect sizes, which can lead to unreliable conclusions that may not be generalizable to other datasets (27), and have consequences for forensics (28). 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 (29). 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 analysis 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 (2). 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 (<24 h→73+ h), and ages (18–55+ years) (Table S1). Detailed methods for sample collection, DNA extraction, and sequencing are available in Pechal et al. (2). 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). Each swab was rotated for 3–5 sec on the body location to sample the microbial community. Samples were placed in sterile microfuge tubes and 200 μ L of 100% molecular-grade ethanol (BP2818-4; Fisher Scientific, Waltham, MA) 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/m^2), 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) following the manufacturer's instructions except an additional 5 $\text{ng}/\mu\text{L}$ of lysozyme was added during the lysis step for each sample reaction (30). The samples were quantified by Qubit 2.0 and the Quant-iT dsDNA HS Assay Kit (Thermo Fisher Scientific). Microbial DNA was sequenced at the Michigan State University (MSU) Genomics Core Facility (East Lansing, MI) 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 HiSeq 2000 and MiSeq. The V4 region 16S gene amplicon 2×250 bp paired-end reads

were generated with region specific primers (515 f [5' GTGCCAGCMGCCGCGGTAA) and 806 r (5' GGACTACHVGGGTWTCTAAT]) that included Illumina flowcell adapters (31). PCR products were normalized and pooled with Invitrogen SequelPrep DNA Normalization Plates. A combination of Qubit dsDNA HS, Caliper LabChip GX HS DNA, and Kapa Illumina Library Quantification qPCR assays was 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. (32). Pooled sequences were loaded on an Illumina MiSeq standard flowcell (v2) and sequenced using a 500-cycle reagent cartridge. Filtering parameters were optimized for detecting low-abundance phylogenetic diversity (33,34). 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 (17), QIIME2 v2018.11 (16), and MG-RAST v4.0.3 (19) to determine how the microbial community composition and diversity metrics (alpha- and beta-diversity) varied among pipelines. The SOP for mothur (32), the QIIME2 moving pictures tutorial (35), and the MG-RAST manual (36) were used for reference to analyze the samples. For mothur and QIIME2, the SILVA small subunit database v132 (37) 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, is represented in Fig. 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 (Fig. 1). For QIIME2, DADA2 v1.8.0 (38) corrected Illumina amplicon sequencing data, including removing phiX reads. Sequences were aligned using MAFFT v7.397 (39), and FastTree v2.1 (40) constructed the phylogenetic tree. BIOM tables were created using biomformat package v2.1 (41) 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 (42). 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 (43) 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” (44), 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, 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.

Data Analysis and Bioinformatics

Pipeline outputs were quantitatively compared using BIOM files from QIIME2, and OTUs and taxonomy files from mothur and MG-RAST which were combined with metadata as phyloseq v1.24.0 (45) objects in R v3.5.1 (46) 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 taxon 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 compared either pipeline output or *in silico* data (14,19). Relative abundance was determined by combining all samples analyzed within each pipeline and sample area, removing taxa that were <1% abundance, and determining the proportional contribution of each taxon to the total community. Differentially abundant taxa among pipelines were determined from relative abundances at the phylum and family taxonomic level using analysis of composition of microbiomes (ANCOM) (47). 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 (48) and Nemenyi (49) *post hoc* tests with the R stats and PMCMR packages (50). Beta-diversity metrics, evaluated using principal coordinate 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 (51) package confirmed beta-diversity and dispersion differences among pipelines. In addition to diversity metrics, a measure of divergence was assessed (Kullback–Leibler divergence) (52) 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) (53) 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. (29).

We also used “mockrobiota” (44) 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 were labeled “correct taxa.” Taxa present only in 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 pipeline comparison results (see *Pipeline Comparison* in Results), QIIME2 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*). Minimum library sizes used for comparison were 1,000

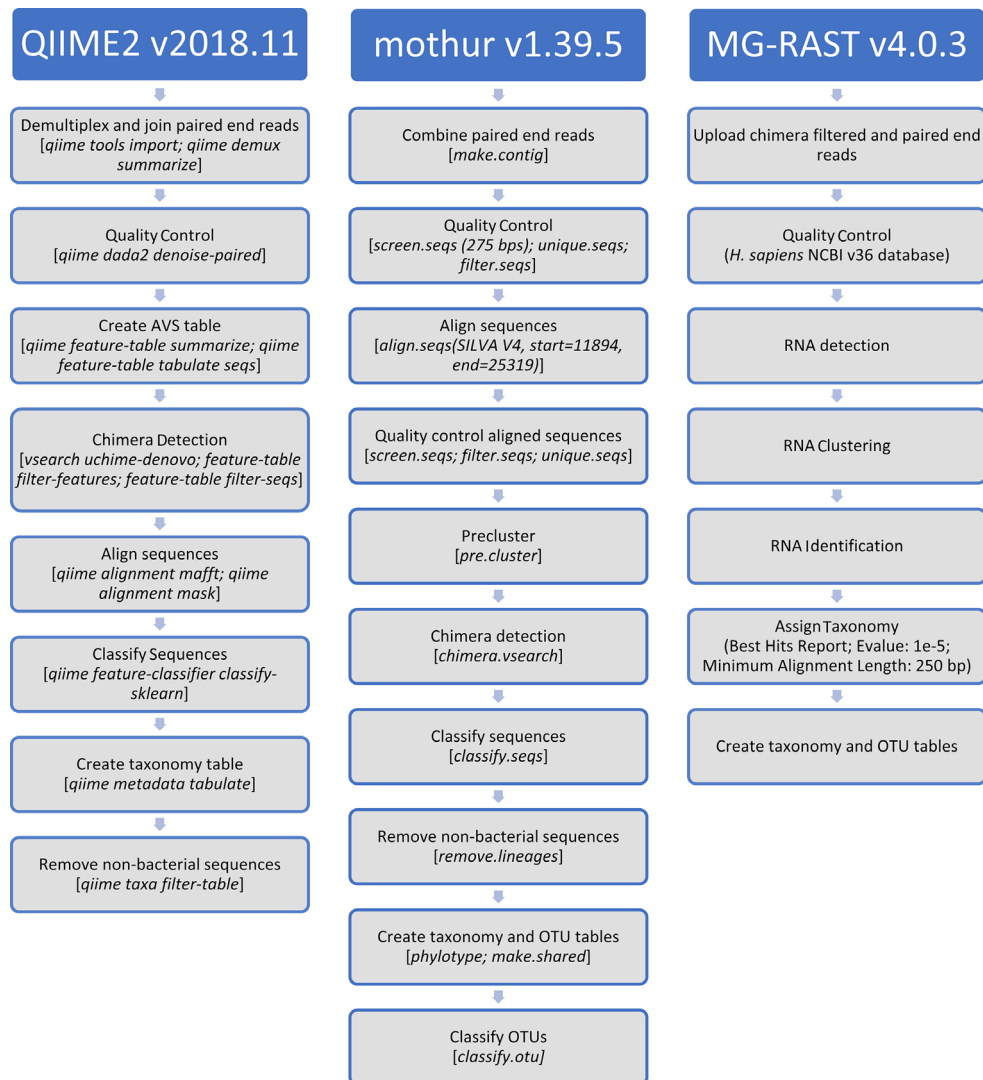


FIG. 1—Workflow for bioinformatic analysis. Several of the steps were shared among pipelines including quality control and aligning sequences. Commands from the code are in square brackets, while parameters are in parentheses. ASV, amplicon sequence variant; OTU, operational taxonomic unit. [Color figure can be viewed at wileyonlinelibrary.com]

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 sample sizes and minimum library sizes 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) (54) 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) (53), and the model error rate was compared among sample sizes and minimum library sizes.

A Priori Power Analysis

To determine statistically significant sample size, an *a priori* power analysis was completed using G*Power 3 v3.0.5 (55). Body locations (mouth and rectum) were compared using the

mean and standard deviation of Faith's phylogenetic diversity from previous bioinformatic analysis (2) calculated using R. An independent mean two-tailed *t*-test was used to determine sample size needed for significant power. For each sample area (mouth and rectum), 27 samples were required for significant power ($\alpha = 0.05$; $1 - \beta = 0.80$). Three random sample sizes were chosen that all have significant power: 60, 120, and 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. (2) 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. Pipeline parameters and microbial community analyses are available 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 (Fig. 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 $\sim 20\times$ smaller than QIIME2 and mothur, tenfold reduction in sequences after filtering, and the highest numbers of unclassified sequences at the family and phylum levels (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 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 levels were reduced for both mouth and rectum sample areas (Fig. 3A,B, Tables S7–S8). MG-RAST had the most differentially abundant taxa (taxa for which relative abundances significantly differed from each other based on statistical tests) at both the phylum (9) and family levels (39), which were represented by about $20\times$ more sequences than mothur and QIIME2 (ANCOM; Tables S9–S11). When considering all samples, MG-RAST produced around 50% lower alpha-diversity for all four metrics compared to mothur and QIIME2 (Kruskal–Wallis $p < 0.05$; *post hoc* Nemenyi $p < 0.05$; Tables S12B and S13A, B). MG-RAST samples clustered closely together (Fig. 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 analyses. Richness and diversity (evenness and richness) metrics from mothur and QIIME2 quantified similar levels of diversity (Fig. 3C,

Table S12A,B). However, QIIME2 had around 20% lower observed richness than mothur (Tables S12B and S13B) in contrast to previous studies (15). 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, Tables S9–S10), which artificially inflated richness measured by mothur. For random forest classification, mothur had slightly lower error rates than QIIME2 (Table 3), and indicator taxa differed among mothur and QIIME2 possibly due to the inflated richness from mothur (Table S15). Despite mothur and QIIME2 samples' admixture (Fig. 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 (Fig. 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 (44) taxonomy reference dataset ($R^2 = 0.27$), while mothur had the lowest ($R^2 = 0.08$) (Fig. 2C,D).

Based on the pipeline comparison results above, QIIME2 was chosen as the best pipeline to examine the effect of minimum library sizes and sample sizes on microbial communities 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 among minimum library sizes indicated the importance of considering minimum library size during forensic microbiome analysis. The number of unclassified sequences at phylum and family levels 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 and 7,000) showed low-abundant taxa (not the top five) phyla as a larger component of the relative abundances due to the random subsampling to a specified library size (Fig. 4A, Table S17). The differences in microbial

TABLE 1—Summary of parameter differences among pipelines compared in this study. Adapted from Plummer et al. (19).

	QIIME2	mothur	MG-RAST
License	Open-Source	Open-Source	Open-Source
Language	Python	C++	Perl
Current version	2019.7	1.43.0	4.0.3
Cited (Google Scholar)	$\sim 1,190$	$\sim 17,800$	$\sim 5,720$
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

Bolded text indicates parameters chosen for running the pipelines.

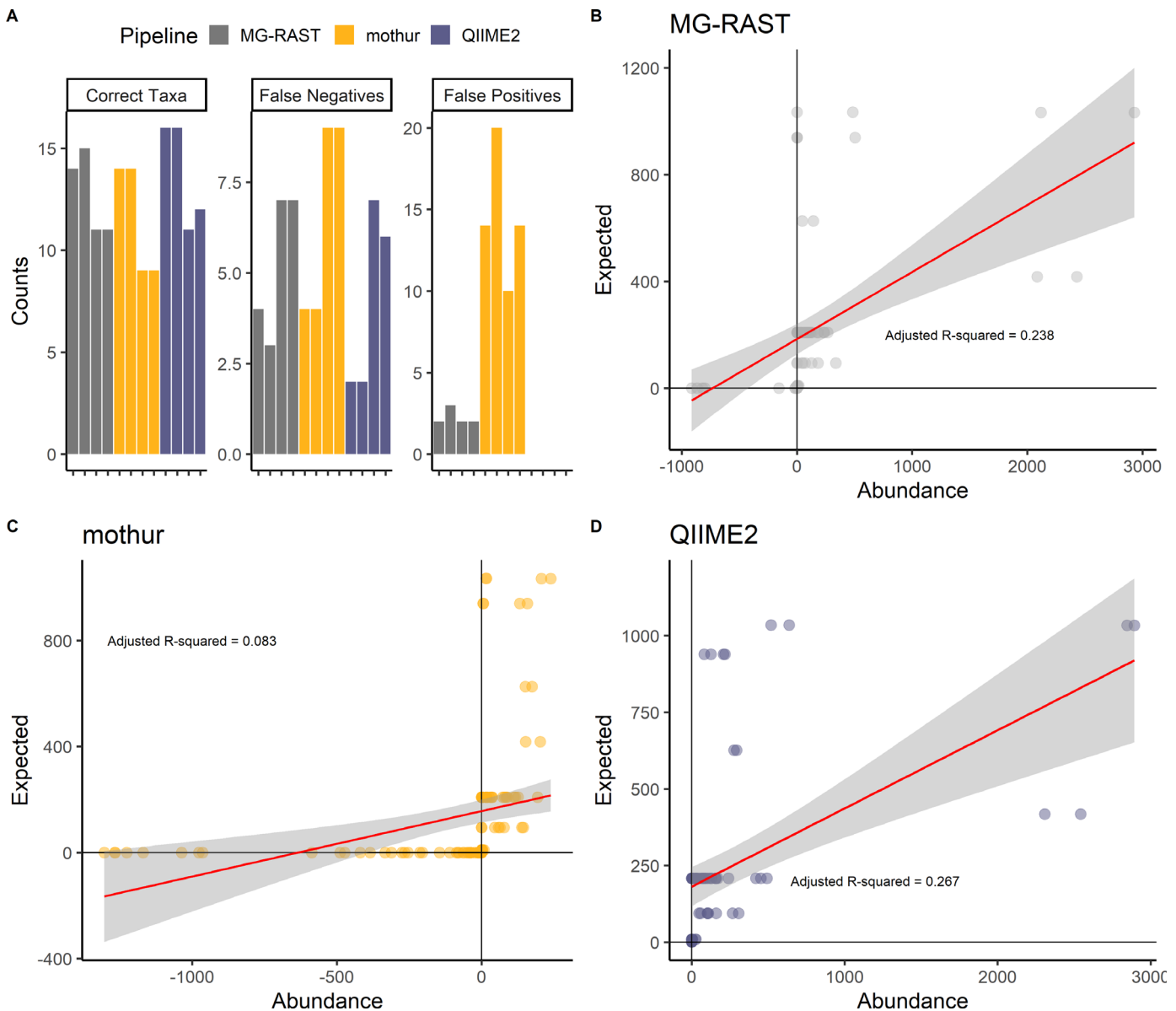


FIG. 2—In silico data error among pipelines. In silico raw sequence data were obtained from mockrobiota (44) and processed through each pipeline using the same methods as the postmortem microbiome dataset. Four samples were processed: two “even” and two “staggered” samples representing different microbial community compositions. (A) Taxon output from each pipeline was compared to the taxonomic reference dataset available on mockrobiota. Taxa that were present in both the mockrobiota taxonomic reference dataset and the pipeline output were 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, and 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 represents the actual concordance. (B) MG-RAST, (C) mothur, and (D) QIIME2. Please see Table S16 for a summary of plot A. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2—Summary of sample number and sequence read differences among pipelines.

Pipeline	Total Number of Samples	Number of Mouth Samples	Number of Rectum Samples	Total Number of Sequence Reads after Filtering	Number of Unclassified Phylum Sequence Reads	Number of Unclassified Family Sequence 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

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.

TABLE 3—Random forest classification error among pipelines.

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)

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 levels. Error rate percentages and number of misclassifications from the total number of samples were included.

community structure among minimum library sizes were evident in the very distinct clustering of minimum library sizes (Fig. 4F). Minimum library size had a large effect on 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 (Kruskal–Wallis, *post hoc* Nemenyi; $p < 0.05$; Fig. 4C, Tables S19–S20). When comparing core microbiome taxa among minimum library sizes, 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 (Fig. 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 (Fig. 5B, Table S21). A majority of taxa were still represented across minimum library sizes. 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 around 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 minimum library sizes of no rarefaction 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 microbial community metrics in a significant way. The total number of sequences, as well

as the number of unclassified sequences at phylum and family taxonomic level, increased 1.5–2× with increasing sample size from 60, 120, to 188 (Table S24), but there were no differentially abundant taxa among subsamples (Fig. 4B). Alpha-diversity metrics did not significantly differ among sample sizes (Kruskal–Wallis; $p > 0.05$; Fig. 4D, Tables S19–S20), but observed richness did increase with increasing sample size, as larger sample sizes captured more low-abundant and 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 (Fig. 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 (Fig. 5A, Table S25), although there were a few OTUs shared among subsamples and unique to the subsample 188 (Fig. 5A, Table S25). Much like minimum library size, error rates for predicting sample area among sample sizes were around 14% for 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 in low-abundant and infrequent taxa among the sample sizes is 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 post-mortem microbiome dataset shows 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-abundant 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 (19). To accurately compare pipelines, we standardized as many steps as possible, such as 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 (20).

Overall, we show that MG-RAST produced 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 (25). However, there are few studies that did not find the same reduction in diversity metrics and increased unclassified reads (15,19), although those studies focused on a smaller sample size of similar microbial communities (35 infant gut microbiome samples) or *in silico* generated data that do not approximate forensic data. Due to the results from this postmortem dataset, the reduction in

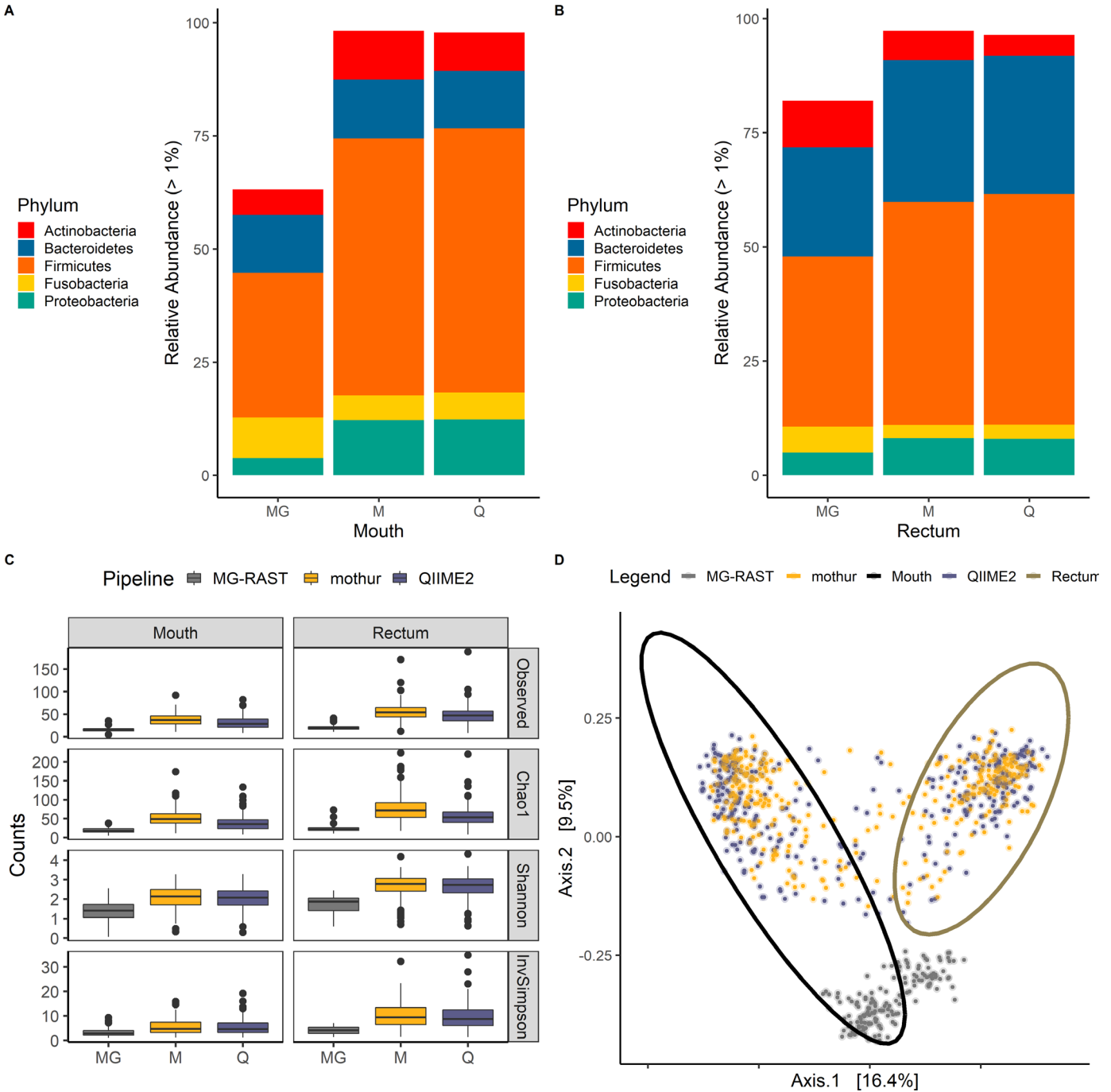


FIG. 3—Microbial community composition among pipelines. (A) Relative abundances of the five most predominate phyla for mouth samples among pipelines (MG: MG-RAST; M: mothur; Q: QIIME2). (B) Relative abundances of the five most predominate phyla for rectum samples among pipelines (MG: MG-RAST; M: mothur; 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 ($p < 0.05$). Please see Tables S7–S14 for a complete summary of relative abundance, analysis of composition of microbiomes (ANCOM), alpha-diversity, and beta-diversity results. [Color figure can be viewed at wileyonlinelibrary.com]

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 (15,19). 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

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

Pipeline	Sample Area	Sample Size	Minimum Library Size	Total Number of Sequence Reads after Filtering	Number of Unclassified Phylum Sequence Reads	Number of Unclassified Family Sequence Reads
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

among mothur and QIIME2. This study included comparison of bioinformatic pipelines using machine learning outcomes. Rather than an exhaustive search of machine learning algorithms (29), 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 <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 laboratories 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 (2,3) or based on minimum library size of the samples (56). We chose three levels of rarefying for the postmortem data, based on these different approaches: no rarefaction, 7,000 sequences (based on alpha-rarefaction curves), and 1,000 sequences (minimum library size of nondefective samples). Previous research determined that in large enough datasets of very different microbial communities (such as postmortem communities across body site) (2), rarefying should not negatively impact downstream analysis (24). However, we found differences among the normalization strategies, indicating that selecting an appropriate minimum library size should be taken into consideration for forensic microbiome analysis. While we showed that more low-abundant 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 (24). Demonstrated by the PCoA plot (Fig. 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 (24). 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 (of both 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 (2), the classifications of manner of death in this study are generally lower than those of previously reported studies with the same dataset (29). However, a previous study provided a comprehensive evaluation of machine learning algorithms, which was not the major focus of this study (29). 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 random forest classifications changed among minimum library sizes, which has implications for forensic cases, as indicator taxa among classifications have potential downstream applications in casework.

Sample Size

There is a trade-off 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 (2). 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

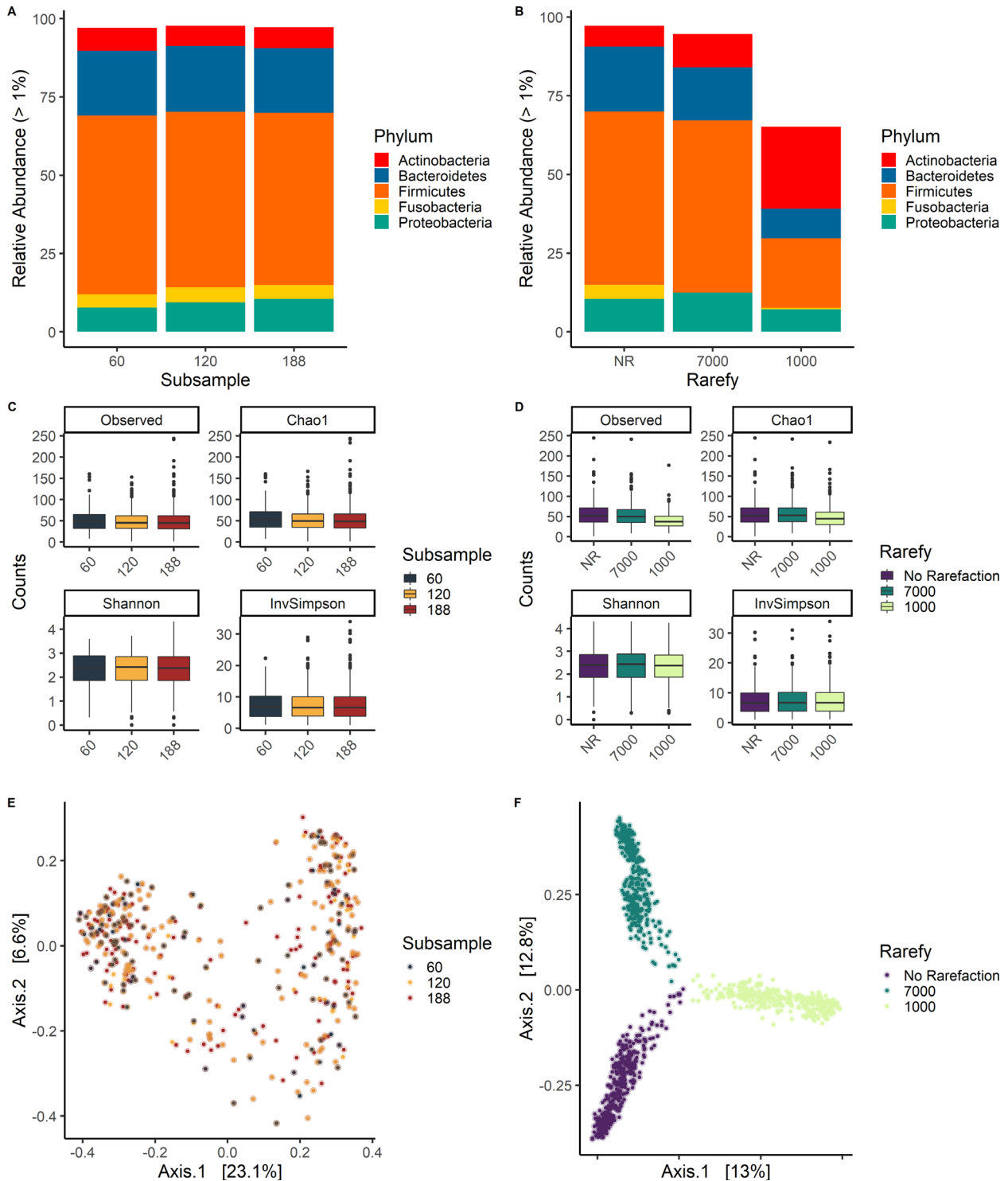


FIG. 4—Microbial community composition among sample sizes and minimum library sizes. (A) Relative abundances of the five most predominate phyla among subsamples with no rarefaction. (B) Relative abundances of the five most predominate phyla among minimum library sizes of subsample 188. (C) Alpha-diversity 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 ($p < 0.05$). Please see Tables S17–S20 for a complete summary of relative abundance, alpha-diversity, and beta-diversity results. [Color figure can be viewed at wileyonlinelibrary.com]

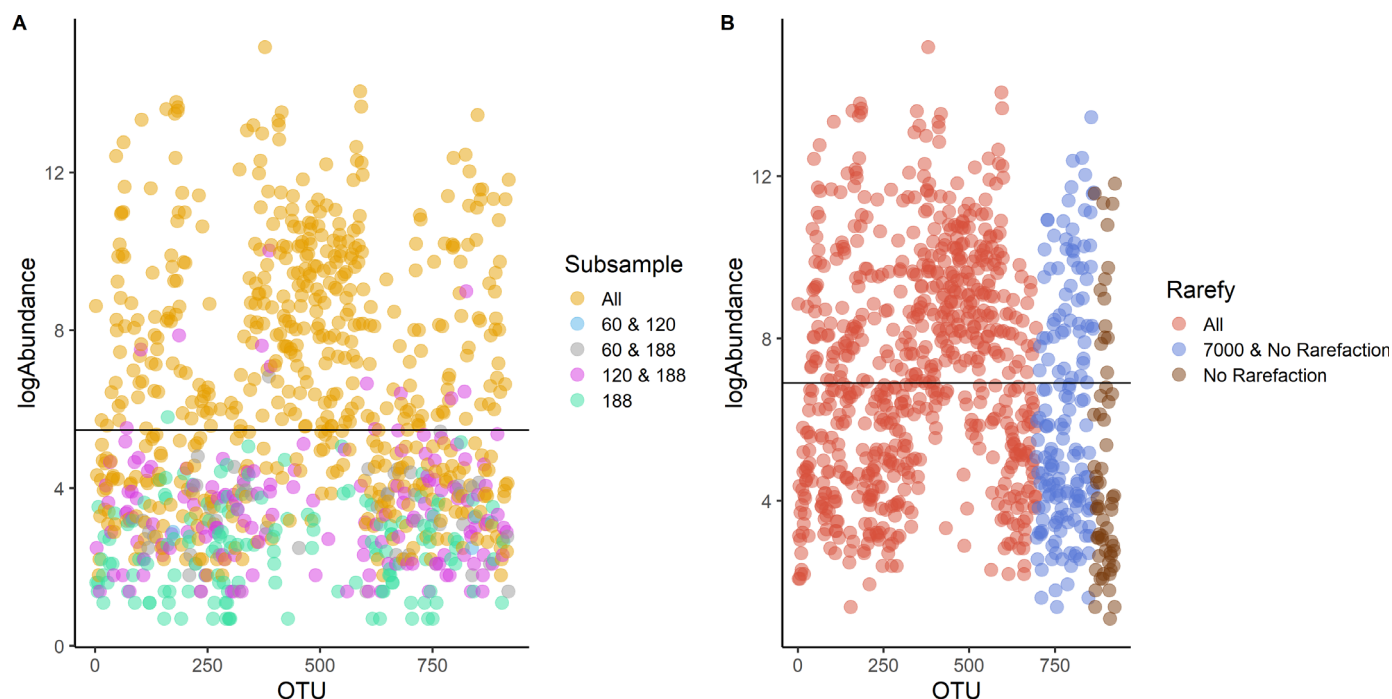


FIG. 5—Core microbiome among sample sizes and minimum library sizes. In this case, core microbiome was considered as OTU (operational taxonomic unit) membership among defined groups (sample size or minimum library size). (A) Occupancy vs. log abundance of shared OTUs among sample sizes at no rarefaction. The horizontal line indicates mean log abundance. (B) Occupancy vs. log abundance of shared OTUs among minimum library sizes for subsample 188. The horizontal line indicates mean log abundance. Please see Tables S21 and S25 for a complete summary of core microbiome results. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 5—Random forest classification error among sample sizes and minimum library sizes.

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)

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, minimum library size, and subsample. Error rate percentages and number of misclassifications from the total number of samples were included.

observed richness did increase with increasing sample size, as larger sample sizes captured more low-abundant and infrequent taxa. As expected, increasing subsample size increased error rate for random forest classifications (sample area and manner of death). However, those random forest models are arguably more robust for downstream forensic applications (29); therefore, including as many samples as possible in forensic microbiology studies is best practice moving forward.

Conclusions

This work was the first to compare pipeline and associated 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: (i) The QIIME2 pipeline is the most suitable pipeline for this type of postmortem microbiome dataset; (ii) rarefying data is the

best normalization practice for downstream statistical analyses (24). 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 (iii) sample size should be maximized to capture low-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 is essential. 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.

Acknowledgments

The authors thank M.M. Brewer and C.R. Weatherbee for help with sample processing. The authors also thank S.E. Evans for comments on the study design and methodology.

References

- Budowle B, Schutzer SE, Einseln A, Kelley LC, Walsh AC, Smith JAL, et al. Building microbial forensics as a response to bioterrorism. *Science* 2003;301(5641):1852–3.
- Pechal JL, Schmidt CJ, Jordan HR, Benbow ME. A large-scale survey of the postmortem human microbiome, and its potential to provide insight into the living health condition. *Sci Rep* 2018;8(1):5724.
- Metcalf JL, Xu ZZ, Weiss S, Lax S, Treuren WV, Hyde ER, et al. Microbial community assembly and metabolic function during mammalian corpse decomposition. *Science* 2016;351(6269):158–62.
- Schmedes SE, Sajantila A, Budowle B. Expansion of microbial forensics. *J Clin Microbiol* 2016;54(8):1964–74.
- Dobay A, Haas C, Fucile G, Downey N, Morrison HG, Kratzer A, et al. Microbiome-based body fluid identification of samples exposed to indoor conditions. *Forensic Sci Int Genet* 2019;40:105–13.
- Schmedes SE, Woerner AE, Budowle B. Forensic human identification using skin microbiomes. *Appl Environ Microbiol* 2017;83(22):1672–17.
- Benbow ME, Pechal JL, Lang JM, Wallace JR. The potential of high-throughput metagenomic sequencing of aquatic bacterial communities to estimate the postmortem submersion interval. *J Forensic Sci* 2015;60(6):1500–10.
- Pechal JL, Crippen TL, Benbow ME, Tarone AM, Dowd S, Tomberlin JK. The potential use of bacterial community succession in forensics as described by high throughput metagenomic sequencing. *Int J Legal Med* 2014;128(1):193–205.
- Johnson HR, Trinidad DD, Guzman S, Khan Z, Parziale JV, DeBruyn JM, et al. A machine learning approach for using the postmortem skin microbiome to estimate the postmortem interval. *PLoS ONE* 2016;11(12):e0167370.
- Metcalf JL, Parfrey LW, Gonzalez A, Lauber CL, Knights D, Ackermann, et al. A microbial clock provides an accurate estimate of the postmortem interval in a mouse model system. *Elife* 2013;2:1104.
- Carter DO, Tomberlin JK, Benbow ME, Metcalf JL, editors. *Forensic microbiology*. Hoboken, NJ: John Wiley & Sons Ltd, 2017.
- Metcalf JL, Xu ZZ, Bouslimani A, Dorrestein P, Carter DO, Knight R. Microbiome tools for forensic science. *Trends Biotech* 2017;35(9):814–23.
- Leipzig J. A review of bioinformatic pipeline frameworks. *Brief Bioinform* 2017;18(3):530–6.
- Sivarajah U, Kamal MM, Irani Z, Weerakkody V. Critical analysis of big data challenges and analytical methods. *J Bus Res* 2017;70:263–86.
- Golob JL, Margolis A, Hoffman NG, Fredricks DN. Evaluating the accuracy of amplicon-based microbiome computational pipelines on simulated human gut microbial communities. *BMC Bioinformatics* 2017;18(1):283.
- Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019;37(8):852–7.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;75(23):7537–41.
- Keegan KP, Glass EM, Meyer F. MG-RAST, a metagenomics service for analysis of microbial community structure and function. *Microbial Env Genet* 2016;1399:207–33.
- Plummer E, Twin J, Bulach DM, Garland SM, Tabrizi SN. A comparison of three bioinformatics pipelines for the analysis of preterm gut microbiota using 16S rRNA gene sequencing data. *J Proteomics Bioinform* 2015;8(12):283–91.
- Siegwald L, Touzet H, Lemoine Y, Hot D, Audebert C, Caboche S. Assessment of common and emerging bioinformatics pipelines for targeted metagenomics. *PLoS ONE* 2017;12(1):e0169563.
- Mysara M, Njima M, Leys N, Raes J, Monsieus P. From reads to operational taxonomic units: an ensemble processing pipeline for MiSeq amplicon sequencing data. *Gigascience* 2017;6(2):1–10.
- Nilakanta H, Drews KL, Firrell S, Foulkes MA, Jablonski KA. A review of software for analyzing molecular sequences. *BMC Res Notes* 2014;7:830.
- McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol* 2014;10(4):e1003531.
- Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 2017;27(5):27.
- D'Argenio V, Casaburi G, Precone V, Salvatore F. Comparative metagenomic analysis of human gut microbiome composition using two different bioinformatic pipelines. *Biomed Res Int* 2014;2014:325340.
- Hyde ER, Haarmann DP, Lynne AM, Bucheli SR, Petrosino JF. The living dead: bacterial community structure of a cadaver at the onset and end of the bloat stage of decomposition. *PLoS ONE* 2013;8(10):e77733.
- Ioannidis JP. Why most published research findings are false. *PLoS Medicine* 2005;2(8):696–701.
- Clarke TH, Gomez A, Singh H, Nelson KE, Brinkac LM. Integrating the microbiome as a resource in the forensics toolkit. *Forensic Sci Int Genet* 2017;30:141–7.
- Zhang Y, Pechal JL, Schmidt CJ, Jordan HR, Wang WW, Benbow ME, et al. Machine learning performance in a microbial molecular autopsy context: a cross-sectional postmortem human population study. *PLoS ONE* 2019;14(4):e0213829.
- Pechal JL, Schmidt CJ, Jordan HR, Benbow ME. Frozen: thawing and its effect on the postmortem microbiome in two pediatric cases. *J Forensic Sci* 2017;62(5):1399–405.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7(5):335–6.
- Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 2013;79(17):5112–20.
- Caporaso JG, Knight R, Kelley ST. Host-associated and free-living phage communities differ profoundly in phylogenetic composition. *PLoS ONE* 2011;6(2):e16900.
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 2012;6(8):1621–4.
- Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, et al. Moving pictures of the human microbiome. *Genome Biol* 2011;12(5):R50.
- Glass EM, Wilkening J, Wilke A, Antonopoulos D, Meyer F. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb Protoc* 2010;2010(1):pdb.prot5368.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;41(Database issue):D590–D596.

38. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high resolution sample inference from Illumina amplicon data. *Nat Methods* 2016;13(7):581–3.
39. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30(4):772–80.
40. Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* 2010;5(3):e9490.
41. McDonald D, Clemente JC, Kuczynski J, Rideout JR, Stombaugh J, Wendel D, et al. The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. *Gigascience* 2012;1(1):7.
42. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;26(19):2460–1.
43. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 2016;4:e2584.
44. Bokulich NA, Rideout JR, Mercurio WG, Shiffer A, Wolfe B, Maurice CF, et al. mockrobiota: a public resource for microbiome bioinformatics benchmarking. *mSystems* 2016;1(5):e00062-16.
45. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 2013;8(4):e61217.
46. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2018.
47. Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R, Peddada SD. Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb Ecol Health Dis* 2015;26:27663.
48. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *J Am Stat Assoc* 1952;47(260):583–621.
49. Nemenyi P. Distribution-free multiple comparisons [dissertation]. Princeton, NJ: Princeton University, 1963.
50. Pohlert T. The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR), 2014. <http://CRAN.R-project.org/package=PMCMR> (accessed September 19, 2019).
51. Oksanen J, Guillaume Blanchet F, Friendly M, Kindt R, Legendre P, McGlenn D, et al. vegan: community ecology package. R package version 2.5-2, 2018. <https://CRAN.R-project.org/package=vegan> (accessed September 19, 2019).
52. Rosa PLS, Deych E, Carter S, Shands B, Yang D, Shannon WD. HMP: hypothesis testing and power calculations for comparing metagenomic samples from HMP. R package version 2.0, 2019. <https://CRAN.R-project.org/package=HMP> (accessed September 19, 2019).
53. Liaw A, Wiener M. Classification and regression by random forest. *R News* 2002;2(3):18–22.
54. Shade A, Handelsman J. Beyond the Venn diagram: the hunt for a core microbiome. *Environ Microbiol* 2012;14(1):4–12.
55. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007;39(2):175–91.
56. DeBruyn JM, Hauther KA. Postmortem succession of gut microbial communities in deceased human subjects. *PeerJ* 2017;5:e3437.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Summary of case metadata among autopsy cases reported in Pechal et al. (2), stratified by metadata category. All information was provided by autopsy reports.

Table S2. Summary of taxonomic name changes after pooling pipeline outputs.

Table S3. Kullback-Leibler Divergence matrices among pipelines, library sizes, and sample sizes.

Table S4. List of sample names that remained after filtering for both sample areas.

Table S5. Summary of library sizes among pipelines.

Table S6. Confusion matrices of random forest classification of 1,000 decision trees. Classification out-of-box error rate of sample area and manner of death were determined among pipelines and taxonomic levels.

Table S7. Mean relative abundances of phyla greater than 1% relative abundance among pipelines and sample areas.

Table S8. Mean relative abundances of families greater than 1% relative abundance among pipelines and sample areas.

Table S9. Analysis of the composition of microbiomes (ANCOM) result for differentially abundant phyla among pipelines and sample areas.

Table S10. Analysis of the composition of microbiomes (ANCOM) results for differentially abundant families among pipelines and sample areas.

Table S11. Average number of samples and sequences representative of differentially abundant families identified by Analysis of the composition of microbiomes (ANCOM) among pipelines.

Table S12. Alpha-diversity metrics for each sample among pipeline and sample area including: observed richness, Chao1, Shannon diversity (1-D), and Inverse Simpson diversity (1/D).

Table S13. Results of A) Kruskal-Wallis and B) pairwise *post hoc* Nemenyi tests of alpha-diversity metrics among pipeline and sample area.

Table S14. Results from permutational multivariate analysis of variance (PERMANOVA) tests on weighted Jaccard distance matrix including A) beta-diversity and B) beta-dispersion for 999 permutations.

Table S15. Top twenty predictor taxa based on mean decrease in Gini for random forest classification of manner of death and sample area.

Table S16. The number of correct taxa, false negatives, and false positives for *in silico* data among pipelines and community types (even and staggered).

Table S17. Mean relative abundances of phyla greater than 1% relative abundance among sample sizes, minimum library sizes, and sample areas.

Table S18. Results from permutational multivariate analysis of variance (PERMANOVA) tests on weighted Jaccard distance matrix including A) beta-diversity and B) beta-dispersion for 999 permutations

Table S19. Alpha-diversity metrics for each sample among sample size and minimum library size including: observed richness, Chao1, Shannon diversity (1-D), and Inverse Simpson diversity (1/D).

Table S20. Results of A) Kruskal-Wallis and B) pairwise *post hoc* Nemenyi tests of alpha-diversity metrics among sample size and minimum library size.

Table S21. Core microbiome among minimum library size. The core microbiome in this case was considered shared OTUs among minimum library size.

Table S22. Confusion matrices of random forest classification with 1,000 decision trees.

Table S23. Top twenty predictor taxa based on mean decrease in Gini for random forest classification manner of death and sample area. The top twenty predictor taxa among minimum library size, sample size, and taxonomic levels were all included.

Table S24. Number of samples that remained after filtering (including chimera detection and rarefaction) among sample area, sample size and minimum library size.

Table S25. Core microbiome among sample size. The core microbiome in this case was considered shared OTUs among subsamples.