H28 Evaluating Bioinformatic Pipeline Performance for Forensic Microbiome Analysis

Sierra Kaszubinski, BS, Michigan State University, East Lansing, MI 48824; Jennifer L. Pechal, PhD, Michigan State University, East Lansing, MI 48824; Heather R. Jordan, PhD, Mississippi State University, Mississippi, MS 39762; Carl J. Schmidt, MD, Wayne County Medical Examiner's Office, University of Michigan, Detroit, MI 48207; M. Eric Benbow, PhD, Michigan State University, East Lansing, MI 48824*

Learning Overview: After attending this presentation, attendees will understand how available bioinformatic pipelines are applied to forensic microbiology and postmortem microbiome research. Attendees will see that downstream output of multiple pipelines differs based on statistical evaluation, including taxon abundance, diversity metrics, and machine learning model building.

Impact on the Forensic Science Community: This presentation will impact the forensic scientific community by informing bioinformatic analysis for further forensic microbiology research and future casework.

While recent research suggests excellent potential for microbial community use in forensics, additional foundational work is needed before forensic microbiology can be applied in the criminal justice system. Downstream bioinformatic analysis needs to be optimized of forensic microbiological data created by high-throughput sequencing platforms in research and case work.[1] To accomplish this goal, a better understanding of performance among bioinformatic pipelines is needed to reveal potentially significant differences in downstream analysis and data interpretation. The ultimate outcome of such evaluation will be the identification of analytical pipelines optimized for use in forensic microbiology.

One aspect of forensic microbiology is examining the human microbiome, which consists of 10-100 trillion microbial cells per person.[2] The composition of these microbial communities can be specific to groups of living individuals and varies by area of the body (e.g., gut vs. skin).[2] Current research on forensic microbiology includes human microbial fingerprinting and estimating time since death using the postmortem microbiome.[3] Advances in sequencing technology have made studying microbes at the genomic level much more affordable. Large-scale microbial identification can be accomplished using 16S rRNA gene sequences obtained from targeted amplicon high-throughput sequencing. These microbial taxon identifications provide a microbiome profile that has potential use in forensics.

Despite increasing research in the use of microbiomes in forensics, there is still a limited understanding of the human postmortem microbiome beyond studies that take place in controlled anthropological facilities.[4] One of the first studies to investigate large-scale (n=188 cases) postmortem microbial changes using high-throughput sequencing and statistical analyses during decomposition revealed dynamic changes for the postmortem microbiome.[5] This research was the largest dataset to provide empirical evidence that microbial communities of the eyes, ears, nose, mouth, and rectum could predict time since death during routine death investigation using high-throughput technology and *in silico* computational tools.[5]

In silico tools are needed to analyze forensic microbiology data. For bioinformatic analysis, raw data files undergo a series of transformations using executable command line software known as pipelines.[6] Pipelines most commonly cited are QIIME, mothur, and MG-RAST.[7] Current literature is not sufficient to justify which pipeline would be most useful in analyzing forensic microbial data. Previous studies comparing pipelines either used simulated data, small sample sizes (n<40) composed of the same sample type (i.e., human gut microbial data), or *in silico* data.[7-12] Studies analyzing samples of limited sample type and number do not extrapolate to forensic microbiology studies, which often included swabs from multiple body sites.[4,5,13]

Two anatomic areas, the mouth and rectum, were selected from the large and variable dataset from Pechal et al. to quantitively evaluate each pipeline.[5] Random subsamples were made from the dataset to include 30 cases, 60 cases, 120 cases, and 188 cases. Microbial sequence reads were analyzed with QIIME2 (2017.8), mothur (v.1.39.5), and MG-RAST (4.0.2). The reference database (SILVA) and operational taxonomic unit generation (*de novo*) were controlled. Phylum and family level taxon were used for comparison and revealed a significant difference ($p < 0.01$) in the number of unclassified reads between pipelines. MG-RAST also had fewer taxa present at 1% relative abundance than mothur or QIIME2. The number and identity of taxa shared between anatomic areas, or core taxa, was distinct among pipelines. Alpha diversity metrics, including Shannon and Inverse Simpson diversity indices, were compared among the pipeline outputs. Alpha diversity metrics were significantly different ($p < 0.01$) between MG-RAST and the other pipelines. The pipelines were also compared using random forest, a machine learning algorithm. While predicting anatomic area, the error rates for the model were relatively similar, predicting the correct anatomic area about 95% of the time. Yet, important predictor taxa differed among pipelines. Overall, mothur and QIIME2 had similar results, while MG-RAST was distinct. QIIME2 and mothur can be systematically used for further forensic microbiome analysis.

Reference(s):

1. Carter, David O. et al. *Forensic Microbiology.* Hoboken: John Wiley & Sons Ltd, 2017.
2. Ursell, Luke K. et al. Defining the Human Microbiome. *Nutrition Reviews* 70 (2012). doi:10.1111/j.1753-4887.2012.00493.x.
3. Metcalf, Jessica L. et al. Microbiome Tools for Forensic Science. *Trends in Biotechnology* 35, no. 9 (2017): 814–823. doi:10.1016/j.tibtech.2017.03.006.
4. Pechal, Jennifer L. et al. The Potential Use of Bacterial Community Succession in Forensics as Described by High Throughput Metagenomic Sequencing. *International Journal of Legal Medicine* 128, no. 1 (2013): 193–205. doi:10.1007/s00414-013-0872-1.
5. Pechal, Jennifer L. et al. A Large-Scale Survey of the Postmortem Human Microbiome, and Its Potential to Provide Insight into the Living Health Condition. *Scientific Reports* 8, no. 1 (2018). doi:10.1038/s41598-018-23989-w.
6. Leipzig, Jeremy. A Review of Bioinformatic Pipeline Frameworks. *Briefings in Bioinformatics* (2016). doi:10.1093/bib/bbw020.

7. Plummer, Erica, and Jimmy Twin. A Comparison of Three Bioinformatics Pipelines for the Analysis of Preterm Gut Microbiota Using 16S RRNA Gene Sequencing Data. *Journal of Proteomics & Bioinformatics* 8, no. 12 (2015). doi:10.4172/jpb.1000381.
8. Golob, Jonathan L. et al. Evaluating the Accuracy of Amplicon-Based Microbiome Computational Pipelines on Simulated Human Gut Microbial Communities. *BMC Bioinformatics* 18, no. 1 (2017). doi:10.1186/s12859-017-1690-0.
9. Siegwald, Léa et al. Assessment of Common and Emerging Bioinformatics Pipelines for Targeted Metagenomics. *Plos One* 12, no. 1 (2017). doi:10.1371/journal.pone.0169563.
10. Nilakanta, Haema et al. A Review of Software for Analyzing Molecular Sequences. *BMC Research Notes* 7, no. 1 (2014): 830. doi:10.1186/1756-0500-7-830.
11. Mysara, Mohamed et al. From Reads to Operational Taxonomic Units: An Ensemble Processing Pipeline for MiSeq Amplicon Sequencing Data. *GigaScience* 6, no. 2 (2017): 1–10. doi:10.1093/gigascience/giw017.
12. D'Argenio, Valeria et al. Comparative Metagenomic Analysis of Human Gut Microbiome Composition Using Two Different Bioinformatic Pipelines. *BioMed Research International* (2014): 1–10. doi:10.1155/2014/325340.
13. Metcalf, Jessica L. et al. A Microbial Clock Provides an Accurate Estimate of the Postmortem Interval in a Mouse Model System. *ELife* 2 (2013). doi:10.7554/elife.01104.

Forensic Microbiology, Postmortem Microbiome, Bioinformatics