Cover Page

Award number: 2016-DN-BX-4194

Project Title: Improving estimates of the postmortem interval with metagenomics and metabolomics

Principal Investigator: Jessica Lynne Metcalf, Associate Professor, Department of Animal Sciences. 1171 W. Pitkin Ave, Animal Sciences Building, Colorado State University, Fort Collins, CO, 80521. (970)-491-0626. Jessica.Metcalf@colostate.edu

Co-PIs:
Rob Knight, Professor, Department of Pediatrics, University of California, San Diego, 9500 Gilman Drive, MC 0763, La Jolla, CA 92093-0763, Phone: (858) 246-1184, Email: robknight@ucsd.edu

Pieter Dorrestein, Professor, Skaggs School of Pharmacy & Pharmaceutical Sciences, University of California - San Diego, 9500 Gilman Drive La Jolla, CA 92093. Phone: (858) 822-5437. Email: pdorrestein@ucsd.edu

David O. Carter, Professor, Laboratory of Forensic Taphonomy, Forensic Sciences Unit, Division of Natural Sciences and Mathematics, Chaminade University of Honolulu, 115 Wesselkamper Science Center, Honolulu, HI 96816. Phone: (808) 739-8352. Email: david.carter@chaminade.edu

Project period: 01/2017 - 12/31/2018
Colorado State University
Award Amount: $684,273

Summary of the project

The use of molecular data has radically transformed the forensic sciences by providing the ability to identify individuals, unknown materials, and metabolic processes. These molecular approaches have had a significant impact on medicolegal death investigation because they have led to the identification of people associated with death as well as the factors contributing to and causing death. Recently postmortem microbiome studies have shown that amplicon-based (e.g. 16S rRNA gene) next-generation sequencing data can be used to estimate the postmortem interval (PMI) and locate clandestine graves. These data have significant potential for forensic science, but their power may be increased even further by incorporating metagenomic and metabolomic data. Comparing these three datasets (amplicon, shotgun metagenomics, and metabolites) will allow us to determine (1) what microbes are present, (2) what the microbes are doing, and (3) what chemical compounds they are generating, respectively. This will ultimately generate a three-pronged approach to investigating postmortem microbial ecology and improve our understanding of the accuracy, reliability, and measurement validity of postmortem microbiome data. The proposed project will achieve this goal by incorporating shotgun metagenomic data and metabolomics data into the workflow of a currently funded project (NIJ-2015-DN-BX-K016) that utilizes thousands of samples to investigate the decomposition of human corpses in three contrasting environments: Colorado, Tennessee, and Texas. Furthermore, the design of the proposed project will allow for the achievement of both fundamental and applied research goals: the proposed datasets will not only be applied to a forensic context but they also represent one of the first studies to investigate these components in disparate terrestrial ecosystems. Thus, our proposed project represents an innovative research opportunity in forensic science.

Major goals and objectives

The purpose of this project is to assess the utility of complimentary multi-omics data types for estimating PMI. We propose to cost-effectively generate shotgun metagenomic and metabolomics data sets by leveraging skin and gravesoil samples being collected at three anthropological research facilities (Texas, Colorado, and Tennessee) over four seasons (36 bodies total) as part of NIJ-2015-DN-BX-K016. As part of this funded research project, we are generating next-generation 16S rRNA and 18S rRNA gene data sets for 2,700 skin and soil samples.

Goal 1: Determine whether metagenomic data and metabolomics data provide accurate estimates of PMI during decomposition of human corpses.

We hypothesize that multi-omic data sets will provide additional power for accurately estimating PMI of corpses. Taxonomic-based predictions suggest that the functional capacity of microbial decomposer communities changes in a clock-like manner (Metcalf et al. 2016). Therefore, we hypothesize that functional gene content (via shotgun metagenomics) of corpse-associated

samples will provide additional power to estimate PMI and allow adjustments for details of the season and site that are difficult to reveal from the taxon profiles alone. To test these hypotheses, we reconstructed functional gene content and high resolution taxonomy (e.g. species, strain-level) for microbial data sets using the shotgun metagenomics data, and tested their utility for estimating PMI. Additionally, we tracked the abundance of small molecules associated with decomposition. Because microbial communities change in response to nutrient availability, we hypothesize that the by-products of nutrient consumption by microbes result in predictable changes in the abundance of metabolites, and these temporal metabolite profiles are useful for accurately estimating PMI. An overall goal is to develop the most accurate and cost-effective molecular clock for estimating PMI. Therefore, we compared amplicon-based PMI estimates resulting from NIJ-2015-DN-BX-K016, to metagenomic and metabolomics-based PMI estimates.

Goal 2: Determine whether metagenomic data and metabolomics data of soils provide high-resolution locality information potentially useful for death scene investigations.
We hypothesize that multi-omic soils data will provide useful locality information for death scene investigators. Recent research demonstrates that gravesoils (corpse-associated soils) contain distinct microbial profiles, which may be useful for locating clandestine graves (Metcalf et al. 2016). Therefore, we hypothesize that both shotgun metagenomics (functional gene profiles and strain-level taxonomic profiles) and metabolomics gravesoils data will exhibit unique signatures of decomposition, suggesting their use for locating unknown (likely buried) gravesites. Recent research also suggests that soils (Young, Weyrich, and Cooper 2014; Young et al. 2015) and dust (Barberan, Ladau, et al. 2015; Barberan, Dunn, et al. 2015) may have distinct microbial signatures that can link them to specific geographic locations. Therefore, we hypothesize that metagenomic and metabolite profiles of both gravesoils (and nearby control soils) will have geographic locality information that is useful for linking trace soils to geographic locations.

Goal 3: Determine temporal stability of common, human-associated metabolites during decomposition.
Human-associated metabolites can be used to understand lifestyle, and identify drugs and medications. We hypothesize that certain metabolites persist for a longer time after death than others. Detection of postmortem human-associated metabolites may provide clues about the time of death (investigated in Phase 1), identify of the deceased, and cause of death. Therefore, determining the time metabolites persist after death is important knowledge for forensic science.

Phase 1: Generate metagenomic and metabolomics data sets for a time series of decomposition samples collected at three anthropological research facilities across four seasons for a total of 36 donor bodies.

Objective 1 - Generate shotgun metagenomic data sets for gravesoil samples collected during decomposition for a total of 36 donor bodies.

Objective 2 - Annotate functional genes and generate per sample gene abundance tables

Objective 3 – Generate strain-level data sets using polymorphisms and by assembling partial and full genomes of microbes from shotgun metagenomic data

Objective 4 - Generate untargeted metabolomics data sets for skin and gravesoil samples collected during decomposition and generate per sample metabolite profiles.

Objective 5 – Using machine learning tools, estimate errors for predicting PMI using each data type alone, and in every combination, and determine which method(s) provide the most power for accurately estimating PMI.

Phase 2: Determine whether metagenomic data and metabolomics data of soils provides high-resolution locality information potentially useful for death scene investigations.

Objective 5 – Determine whether gene functional profiles, strain-level microbial community data, and metabolite profiles of gravesoils have distinct decomposition signatures that would be useful for locating clandestine graves.

Objective 6 – Determine whether gene functional profiles, strain-level microbial community data, and metabolite profiles of soils provide links to specific locations at both the regional and global scales by comparing data to the Earth Microbiome Project (EMP) soils multi-omic database.

Phase 3: Determine whether human-associated metabolites, which may be useful for identifying the deceased or cause of death, degrade during decomposition or persist in the skin or surrounding soil.

Objective 7 –Determine whether human-associated metabolites persist on skin and soil during decomposition as a proof of concept for using postmortem metabolites to help identify the deceased or infer cause of death.

Research questions

1. Do metagenomic data and metabolomics data improve PMI prediction compared to amplicon data sets (16S rRNA and 18S rRNA)?

1a. Do functional microbial genes go beyond existing predictions from 16S rRNA profiles, either in aggregate or in specific carefully selected subsets? Do they yield accurate PMI estimates?

1b. Do strain level microbial data provide accurate estimates of PMI?

1c. Do biologically active small molecules produced during decomposition provide accurate PMI estimates?

1d. Does any combination of amplicon data (16S rRNA, 18S rRNA), metagenomic functional gene content or strain-level data, and metabolomic data improve PMI prediction, compared to each of the data sets alone?

2. Do soil metagenomic data and metabolomics data have potential as a tool for locating clandestine graves or for soil trace forensics?

2a. Do metagenomic functional profiles, strain-level taxonomic profiles, and/or metabolomic profiles show distinct signatures during decomposition, thus showing potential for locating unknown graves?

2b. Are metagenomic functional profiles, strain-level taxonomic profiles, and/or metabolomic profiles highly unique to geographic locations, thus showing potential as a tool as soil trace analysis?

3. How long do metabolite profiles of common natural products or medications (e.g. lotions, hair products) persist on the skin or in the soil after death?

3a. Do metabolites that are useful indicators of lifestyle, identity, or even cause of death persist for days or weeks after death?

Research design, methods, analytical and data analysis techniques

Sample collection

Skin and soil swabs were collected daily from 36 human cadavers during the first 21 days of decomposition. The cadavers were placed at three anthropological facilities (3 cadavers per facility per season over four seasons) as part of award NIJ-2015-DN-BX-K016.

DNA extraction, metagenomic sequencing, and data processing

Skin and soil swabs were immediately frozen upon collection and stored at -20°C at the University of Colorado Boulder. Skin and soil DNA were extracted from swabs using the PowerSoil DNA isolation kit (MoBio Laboratories, Carlsbad, CA, USA), according to the standard Earth Microbiome Project protocols (http://www.earthmicrobiome.org). DNA was initially extracted from one swab of each sample as part of funded project NIJ-2015-DN-BX-K016. For samples that did not return sufficient DNA for shotgun metagenomics sequencing,

second duplicate swabs were re-extracted at Colorado State University following the same protocols with the minor adjustments of heated elution solution and two elution rinses to increase DNA yield.

Shotgun metagenomics sequencing

DNA was shipped to Co-PI Knight at the University of California San Diego for preparation of shotgun metagenomics libraries. Prior to shogun metagenomics library preparation, input DNA is transferred to a 384-well plate and quantified using a PicoGreen fluorescence assay (ThermoFisher, Inc). Input DNA was then normalized to 1 ng in a volume of 3.5 µL of molecular-grade water using an Echo 550 acoustic liquid-handling robot (Labcyte, Inc). Enzyme mixes for fragmentation, end repair and A-tailing, ligation, and PCR were prepared and added in approximately 1:8 scale volumes using a Mosquito HV micropipetting robot (TTP Labtech). Fragmentation is performed at 37 °C for 20 min, followed by end-repair and A-tailing at 65 °C for 30 min.

Sequencing adapters and barcode indices were added in two steps, following the iTru adapter protocol (Glenn et al. 2019). Universal adapter "stub" adapter molecules and ligase mix are first added to the end-repaired DNA using the Mosquito HV robot and ligation performed at 20 °C for 1 h. Unligated adapters and adapter dimers are then removed using AMPure XP magnetic beads and a BlueCat purification robot (BlueCat Bio). 7.5-µL magnetic bead solution is added to the total adapter-ligated sample volume, washed twice with 70% EtOH, and then resuspended in 7 µL molecular-grade water. Next, individual i7 and i5 are added to the adapter-ligated samples using the Echo 550 robot. Because this liquid handler individually addresses wells, and we use the full set of 384 unique error-correcting i7 and i5 indices, we are able to generate each plate of 384 libraries without repeating any barcodes, eliminating the problem of sequence misassignment due to barcode swapping (Sinha et al. 2017; Costello et al. 2018). To ensure that libraries generated on different plates can be pooled if necessary, and to safeguard against the possibility of contamination due to sample carryover between runs, we also iterate the assignment of i7 to i5 indices each run, such that each unique i7:i5 index combination is only repeated once every 147,456 libraries. 4.5 µL of eluted bead-washed ligated samples is added to 5.5 µL of PCR master mix and PCR-amplified for 15 cycles. The amplified and indexed libraries are then purified again using magnetic beads and the BlueCat robot, resuspended in 10 µL water, and 9 µL of final purified library transferred to a 384-well plate using the Mosquito HTS liquid-handling robot for library quantitation, sequencing, and storage. 384 samples are then normalized based on a PicoGreen fluorescence assay, for sequencing on the Illumina Hiseq 4000 or NovaSeq platform.

Taxonomic and functional annotations

Shotgun reads were processed to remove low quality reads and reads which mapped to the human genome using Atropos (Didion, Martin, and Collins 2017). Paired-end reads were merged with Seqtk and converted to FASTA format (https://github.com/lh3/seqtk). Shotgun read counts

per sample ranged from 0 to 30,328,667 with a median of 1,740,022. Samples with less than 500,000 paired reads were removed from the analyses. Shotgun reads were then processed through Shogun (Hillmann et al. 2018) to align the reads to the Web of Life (WoL) (Zhu et al. 2019) phylogenetic taxonomic database using the BURST aligner (Al-Ghalith and Knights 2017). The WoL database is a reference phylogeny of 10,575 bacterial and archaeal genomes based on the RefSeq database. We annotated shotgun reads with functional gene information with Woltka (https://github.com/qiyunzhu/woltka). Woltka maps the WoL annotations to the UniRef protein database then profiles the annotations to EggNOG v4.5 ortholog groups.

Metagenomes
Genome assembly analyses are underway.

Metabolite extraction and data generation
Skin and soil metabolites were extracted off a replicate swab type used for DNA recovery. Replicate swabs were shipped to Co-PI Dorrestein from Colorado State University. Molecules were identified using Ultra-performance liquid chromatography – quadrupole – time of flight (UPLC-Q-TOF) and matrix assisted laser desorption/ionization TOF (MALDI-TOF), which detects higher molecular weight molecules. Skin and soil swabs samples were extracted using a solution of 80% Methanol with 0.5 uM sulfamethazine added to serve as an extraction standard. Briefly, samples were mixed in a 96 well plate for two minutes followed by a 15 minute sonication bath. Next, samples were incubated for 2 hours at 4°C, followed by a 12 hour incubation at -20°C. Swabs were then removed and samples lyophilized and stored at -80°C until analysis. Untargeted metabolomics LC-MS/MS data were collected from each sample. Two types of datasets were generated from each sample: MS1 data for global and statistical analysis and MS/MS data for molecular annotation. Molecular annotation is performed through the GNPS platform https://gnps.ucsd.edu/. Molecules were annotated with the GNPS reference libraries (Wang et al. 2016) using accurate parent mass and MS/MS fragmentation pattern according to level 2 or 3 of annotation defined by the 2007 metabolomics standards initiative (Sumner et al. 2007). If needed and if the authentic chemical standard is available, MS/MS data was collected from the chemical standard and compared to MS/MS spectra of the molecule annotated from the sample (level 1 of annotation).

Diversity analysis and model predictions
Shotgun metagenomic data taxonomic and functional annotations, as well as metabolite annotations were imported into QIIME2 (Bolyen et al. 2019) for further analysis. Low abundance features were filtered from each dataset. Potential contaminants were filtered from the taxonomic dataset using decontam (Davis et al. 2018). The taxonomic dataset was normalized by rarefying to 9,039 features. Beta diversity distances were calculated for each dataset: generalized UniFrac for taxonomy, robust Atchison for functional, and Jaccard for metabolites. Beta diversity comparisons between facilities were performed using PERMANOVA and beta

diversity comparisons over ADD were performed using linear mixed-effects models. Normalized taxonomic and metabolite tables were utilized for random forest regression to predict ADD using scikit-learn v0.22.2 (Pedregosa et al. 2011). The functional dataset was normalized by total sum scaling to 10,000 features before regression analysis. Nested cross-validation random forest regression models were created using parameter hyperparameter tuning. Validation was compared by both leaving one facility out per iteration and by leaving one body out per iteration for training and testing the models.

Expected applicability of the research

Our research greatly advances knowledge required for developing microbiome tools to estimate PMI. We anticipate the outcomes of this research as one of the last steps before validation studies will begin to test how well these tools work in realistic death scene scenarios. Furthermore, our research explores the use of soil microbiomes for forensics purposes beyond PMI estimation, such as an indicator of gravesoils or geographic location.

Participants and other collaborating organizations

This project was led by a team of researchers from Colorado State University (Metcalf, formerly at U Colorado Boulder), U of California San Diego (Knight, Dorrestien), and Chaminade University Honolulu (Carter). Postdoctoral scientist Zachary Burcham (CSU, Metcalf) led data analysis and manuscript preparation. GRA Aeriel Belk contributed to machine learning data analysis. At UC San Diego, Research Associate Greg Humphrey managed the shotgun metagenomics data generation and postdoctoral scientist Amina Bouslimani oversaw metabolomics data generation. Finally, we worked with the Forensic Technology Center of Excellence RTI to share our results with stakeholders and discuss next steps forward with other scientists working on similar topics and the justice practitioners.

Changes in approach from original design and reason for change, if applicable

There were no major changes in approach from the original proposed research.

Outcomes

Activities/accomplishments

During this project, we generated and analyzed microbial shotgun metagenomics and metabolite data sets of human cadaver-associated samples, which allowed us to address our goals as described below in *Results and findings*. These activities included training opportunities for a graduate student (Belk) and two postdoctoral scientists (Burcham, Bouslimani), as well as dissemination activities such as research talks at the American Academy of Forensic Sciences

(AAFS) 2019 and the NIJ R&D 2020 annual meetings. Additionally, we hosted a stakeholder meeting at AAFS 2019.

Results and findings

Goal 1: Determine whether metagenomic data and metabolomics data provide accurate estimates of PMI during decomposition of human corpses.

We utilized soil shotgun metagenomics sequencing data and soil and skin metabolomics data to generate random forest regression models to predict PMI, and compare these results to previous results using 16S rRNA amplicon sequencing data. Shotgun metagenomic data were used in two different ways 1) to generate a species-level taxonomy relative abundance table and 2) to generate a gene function relative abundance table. Overall, a model which incorporates shotgun metagenomics soil species level data, facility, humidity, and precipitation data is the best for predicting Accumulated Degree Day (ADD) (Figure 1A). Metabolites from the skin also performed well, and slightly better than soil metabolite data (Figure 1A). Importantly, both shotgun metagenomics data and metabolite data appear to generate models with lower errors than 16S rRNA amplicon sequencing data. Finally, within the shotgun sequencing data, we assessed models at different taxonomic levels, and discovered that higher resolution taxonomic levels such as species performed better than lower levels such as phylum (Figure 1B). For these highly performing models, we analyzed important features (taxa) contributing to low model errors (Table 1). We are currently still optimizing models and the use of metadata such as humidity and precipitation. Therefore, these results may change slightly.

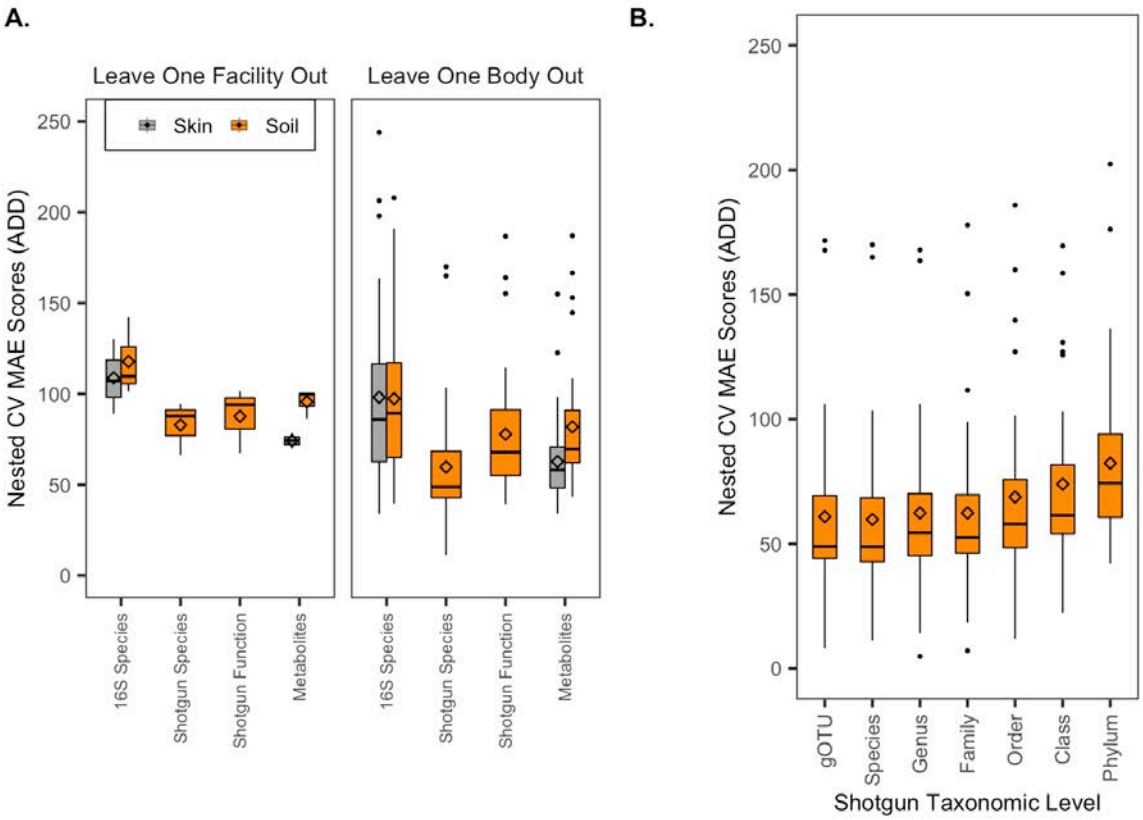

Figure 1. Nested cross-validation scores of random forest regression models estimating ADD from different data types, including species level taxonomy using 16S rRNA data and shotgun metagenomics data, gene function data (shotgun), and metabolites. **A.** An error rate comparison of the models validated by leaving one facility out for training and validated by leaving one body out for training. **B.** An error rate comparison of using different taxonomic levels in the shotgun metagenomics data. Error rates are represented as the mean absolute error (MAE) in ADD; therefore, the lower the error the better the model's performance. The black diamonds represent the generalization error (mean nested cross-validation score) of the model.

Table 1. Top five important shotgun species for predicting ADD.

Importance	Species	Class	Associated Habitats
0.0643	*Vagococcus lutrae*	Bacilli	Human, Swine, Otter, Feces, Water
0.0622	*Bacteroides coprosuis*	Bacteroidia	Human, Swine, Feces
0.0441	*Morganella morganii*	Gammaproteobacteria	Human & Mammal Gut
0.0286	*Proteus mirabilis*	Gammaproteobacteria	Human & Mammal Gut, Soil, Water
0.0215	*Ignatzschineria larvae*	Gammaproteobacteria	Flesh Fly

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Goal 2: Determine whether metagenomic data and metabolomics data of soils provide high-resolution locality information potentially useful for death scene investigations.

In this goal we focus on the potential of decomposition-associated soils, or gravesoils, as a potential forensic tool. Gravesoils are an attractive sample type for outdoor death scenes because it does not require direct sampling of the decedent. Furthermore, skin or other human sample types have the additional challenge of host DNA or metabolites dominating the data. In the case of shotgun metagenomics data, the human DNA sequence reads are filtered out, potentially wasting a large fraction of sequencing effort. From the soil data, we first assess whether decomposition has a measurable influence on gravesoils during decomposition. Previously, we demonstrated that microbial taxa relative abundance, as estimated by amplicon sequencing data, changes substantially during decomposition (Metcalf et al. 2013, 2016). Here we demonstrate that microbial relative abundance and gene function, as estimated by shotgun metagenomics data, and metabolites also change measurably during decomposition (Figure 2, Table 2). These data suggest that soil has both microbial genomic as well as small molecule indicators of decomposition that could potentially be used to locate buried bodies.

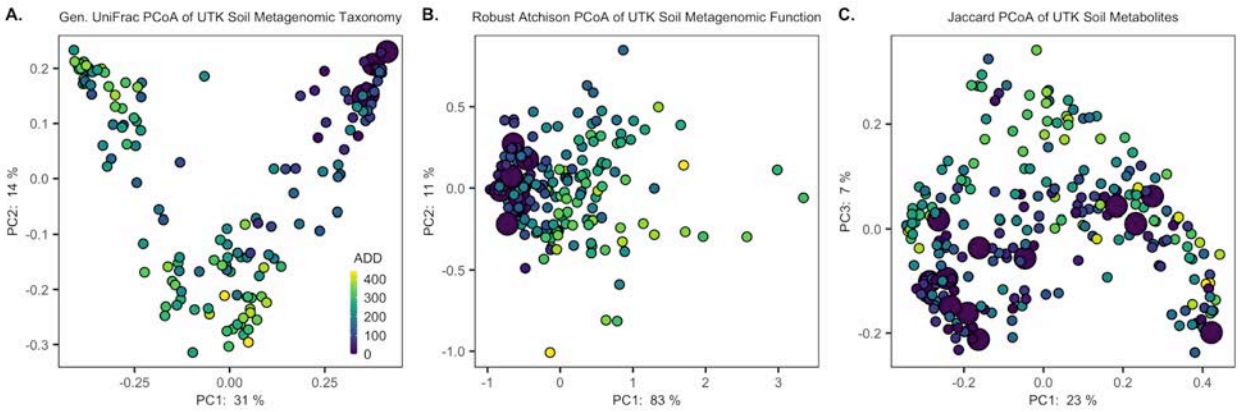

Figure 2. PCoAs of UTK soil beta diversity over ADD and data types of **A.** generalized UniFrac (α = 0.5) shotgun metagenomic taxonomy, **B.** robust Atchison shotgun metagenomic function, and **C.** Jaccard metabolites.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 2. Linear mixed-effects model results for UTK soil beta diversity over ADD. *P*-value < 0.05 indicates significance.

Data Type	Principle Coordinate	Variable	Estimate	SE	Z-score	*P*-value
Shotgun Metagenomic Taxonomy	PC1	Intercept	0.259	0.057	4.536	<0.001
		ADD	-0.001	<0.001	-9.630	<0.001
		Group Var.	0.031	0.106	N/A	N/A
Shotgun Metagenomic Taxonomy	PC1	Intercept	0.752	0.114	6.590	<0.001
		ADD	-0.004	<0.001	-11.715	<0.001
		Group Var.	0.096	0.102	N/A	N/A
Metabolites	PC3	Intercept	0.059	0.026	2.305	0.021
		ADD	<-0.001	<0.001	-6.780	<0.001
		Group Var.	0.007	0.035	N/A	N/A

Second, soils may have potential as trace forensic evidence if they can be used to determine a geographic location. Therefore, we investigated how distinct soil microbial communities were among anthropological research facilities. We discovered that microbial taxonomic signals were very different across geographic locations, with metagenomic gene function slightly less so (Figure 3, Table 2). Metabolite profiles, however, appear more conserved across locations (Figure 3, Table 2). This is likely due to some conservation of microbial function and small molecules in soils, regardless of location or soil type. Functional redundancy in microbial taxa, however, allows microbial communities to vary across locations without functions or small metabolites varying as much (Louca et al. 2018). Therefore, we conclude that microbial taxonomic data (either amplicon or shotgun metagenomic) are likely the most powerful tool for the use of soils to link to locations.

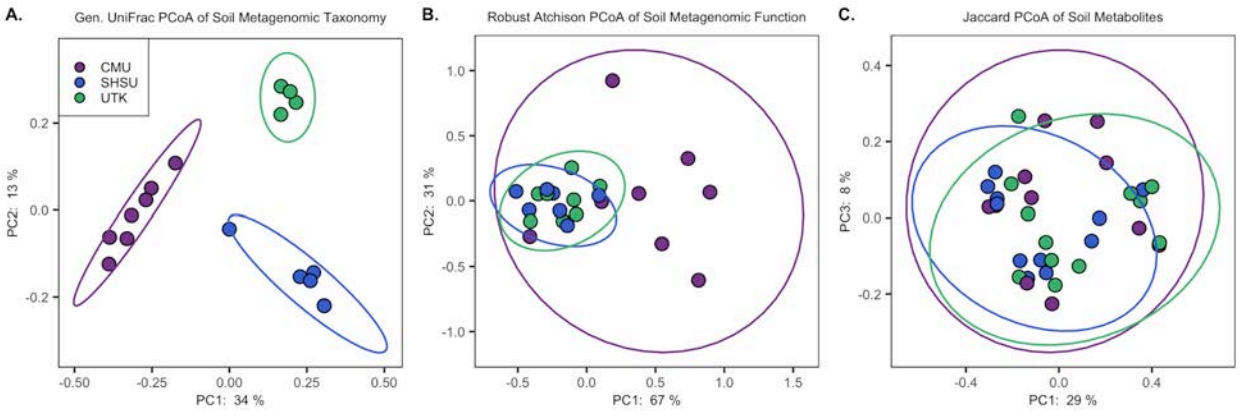

Figure 3. PCoAs of soil beta diversity between facilities at day 0 and data types of **A.** generalized UniFrac (α = 0.5) shotgun metagenomic taxonomy, **B.** robust Atchison shotgun metagenomic function, and **C.** Jaccard metabolites.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 2. PERMANOVA results from beta diversity comparisons between facilities at day 0. *P*-value < 0.05 indicates significance.

Data Type	Sample Size	No. of Groups	Pseudo-F	*P*-value
Shotgun Metagenomic Taxonomy	17	3	5.56	0.001
Shotgun Metagenomic Function	24	3	5.82	0.002
Metabolites	34	3	0.93	0.526

Goal 3: Determine temporal stability of common, human-associated metabolites during decomposition.

To date, we have identified several common human-associated metabolites and confirmed their presence across seasons (Figure 4). Analyses are underway to analyze additional human-associated metabolites and their persistence during decomposition.

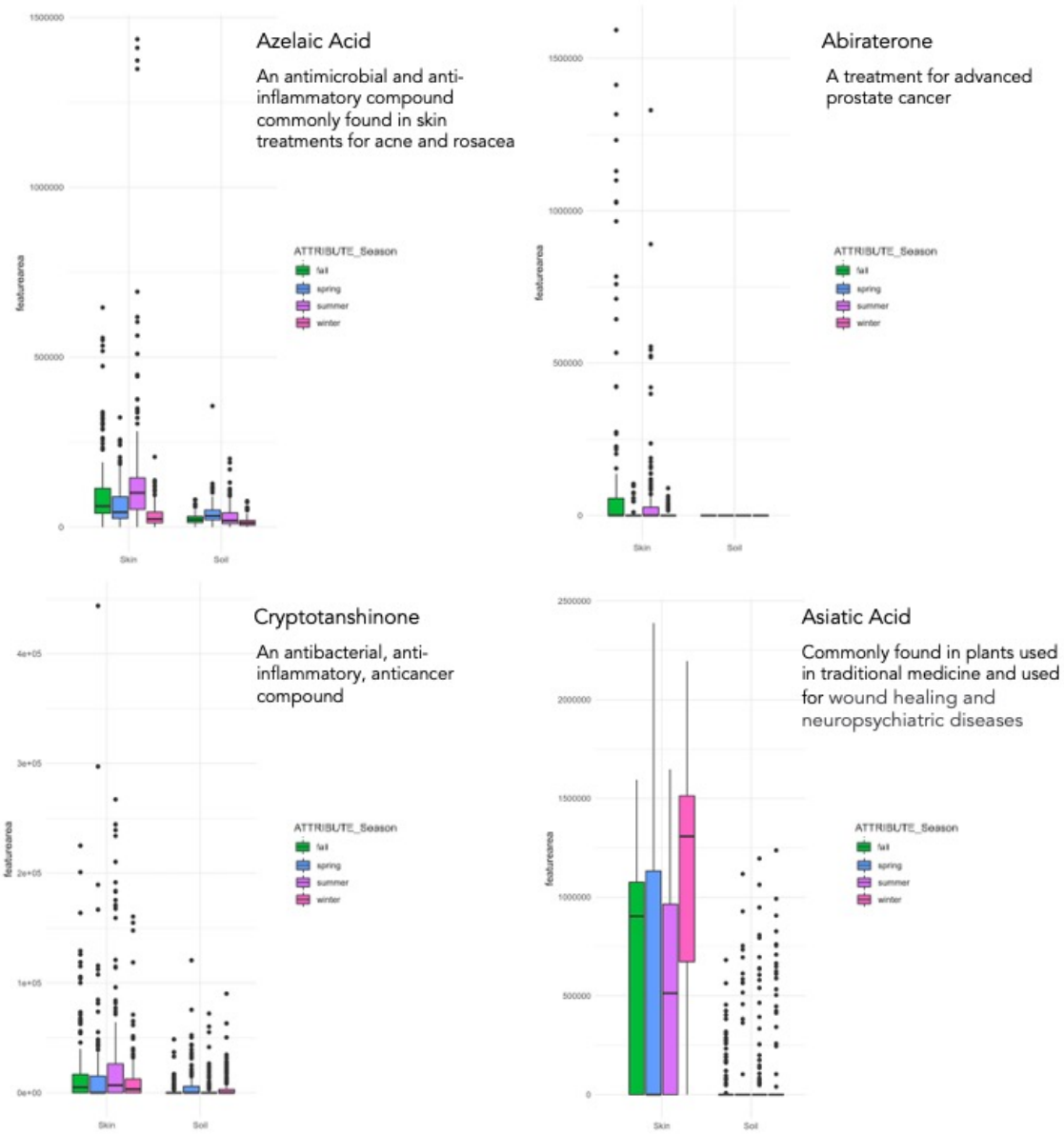

Figure 4. Four common human-associated metabolites that were at detectable levels of human cadaver skin, and sometimes gravesoils as well. Overall, higher signals (area under the detection curve) were detected in skin, and in some cases it varied by season.

Limitations

These data further demonstrate the potential usefulness of microbial data for forensic investigation, and particularly estimating the postmortem interval (Goal 1). These results are limited due to sample size. Even though these data substantially increase microbial forensic human cadaver data, and represent decomposition across 36 human cadavers and across three facilities and four seasons, the reality is that these data are still fairly small for machine learning applications. As more data becomes available, these models will be made more robust.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Artifacts

List of products (e.g., publications, conference papers, technologies, websites, databases), including locations of these products on the Internet or in other archives or databases

Metcalf, JL. 2019 Estimating the Postmortem Interval Using Microbes: Knowledge Gaps and a Path to Technology Adoption. Forensic Science International. Genetics 38 (January): 211–18. https://www.fsigenetics.com/article/S1872-4973(18)30403-4/fulltext

Belk A, Xu ZZ, Carter DO, Lynne A, Bucheli S, Knight R, Metcalf JL 2018 Microbiome Data Accurately Predicts the Postmortem Interval Using Random Forest Regression Models Genes, 9(2), 104 doi:103390/genes9020104

Burcham et al. Microbial metagenome and metabolome predict the postmortem interval. *in prep*

Website: http://www.jessicalmetcalf.com/research-2/#forensics

Video products: Tangled Bank (https://www.tangledbankstudios.org/) https://www.youtube.com/watch?v=B_lHQSxz9GI

Data sets will be released under QIITA (https://qiita.ucsd.edu/) studies 11271 and 11489 upon publication, as well as other national sequence archive databases.

Data sets generated (broad descriptions will suffice)

We generated two large data sets as part of this award. We generated a shotgun metagenomics DNA sequencing data set of 575 human decomposition-associated soil samples, and a metabolomics data set of 795 skin and 806 soil samples. These samples were collected as part of award NIJ-2015-DN-BX-K016, in which skin and soils were collected for the initial 21 days of decomposition of 36 human cadavers across three anthropological facilities and four seasons (3 bodies per facility per season).

These data will be made public upon peer-review publication.

Dissemination activities

In September of 2017, we had the opportunity to be filmed about our work by Tangled Bank (https://www.tangledbankstudios.org/), a film production company of the Howard Hughes Medical Institute (HHMI). The episode is part of a series based on Ed Young's popular book "I Contain Multitudes" and his coverage of microbiome science at the Atlantic Magazine. Watch it here. https://www.youtube.com/watch?v=B_lHQSxz9GI

On February 18, 2019, PI Metcalf led and organized a meeting entitled "Stakeholder meeting for estimating the postmortem interval using microbes." This was in collaboration with Co-PIs David O. Carter, Rob Knight, and the Forensic Technology Center of Excellence (agenda included at end of this document). Belk contributed a talk.

PhD student Belk contributed oral presentations on February 21st, 2019 AAFS. 71st Annual Meeting of the American Academy of Forensic Sciences. February 21, 2019, 4:30pm. Building a Microbial Model to Estimate the Postmortem Interval Using Data Collected During the Spring Season at Three Anthropological Facilities. Aeriel D. Belk, Heather Deel, David O. Carter, Sibyl Bucheli, Aaron Lynne, Melissa Conner, Dawnie Steadman, Giovanna Vidoli, Rob Knight, Jessica L. Metcalf.

Postdoctoral scientist Zachary Burcham contributed oral presentations on February 18th, 2020 AAFS. 72nd Annual Meeting of the American Academy of Forensic Sciences. February 18, 2020, 4:45pm. Using Microbiome Tools to Estimate the Postmortem Interval of Human Remains. Zachary M. Burcham and David O. Carter, Aeriel D. Belk, Heather Deel, Sibyl Bucheli, Aaron Lynne, Melissa Conner, Dawnie Steadman, Giovanna Vidoli, Rob Knight, Jessica L. Metcalf.

Postdoctoral scientist Zachary Burcham or PI Metcalf will present our final results and story at the 2022 AAFS meeting (hopefully post-covid pandemic).

Literature Cited

Al-Ghalith, G., and D. Knights. 2017. "BURST Enables Optimal Exhaustive DNA Alignment for Big Data." *DOI: Doi. org/10. 5281/zenodo* 806850.

Barberan, A., R. R. Dunn, B. J. Reich, K. Pacifici, E. B. Laber, H. L. Menninger, J. M. Morton, et al. 2015. "The Ecology of Microscopic Life in Household Dust." *Proceedings. Biological Sciences / The Royal Society* 282 (1814). https://doi.org/10.1098/rspb.2015.1139.

Barberan, A., J. Ladau, J. W. Leff, K. S. Pollard, H. L. Menninger, R. R. Dunn, and N. Fierer. 2015. "Continental-Scale Distributions of Dust-Associated Bacteria and Fungi." *Proceedings of the National Academy of Sciences of the United States of America* 112 (18): 5756–61.

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.

Costello, Maura, Mark Fleharty, Justin Abreu, Yossi Farjoun, Steven Ferriera, Laurie Holmes, Brian Granger, et al. 2018. "Characterization and Remediation of Sample Index Swaps by Non-Redundant Dual Indexing on Massively Parallel Sequencing Platforms." *BMC Genomics* 19 (1): 332.

Davis, Nicole M., Diana M. Proctor, Susan P. Holmes, David A. Relman, and Benjamin J.

Callahan. 2018. “Simple Statistical Identification and Removal of Contaminant Sequences in Marker-Gene and Metagenomics Data.” *Microbiome* 6 (1): 226.
Didion, John P., Marcel Martin, and Francis S. Collins. 2017. “Atropos: Specific, Sensitive, and Speedy Trimming of Sequencing Reads.” *PeerJ*. https://doi.org/10.7717/peerj.3720.
Glenn, Travis C., Roger A. Nilsen, Troy J. Kieran, Jon G. Sanders, Natalia J. Bayona-Vásquez, John W. Finger, Todd W. Pierson, et al. 2019. “Adapterama I: Universal Stubs and Primers for 384 Unique Dual-Indexed or 147,456 Combinatorially-Indexed Illumina Libraries (iTru & iNext).” *PeerJ*. https://doi.org/10.7717/peerj.7755.
Hillmann, Benjamin, Gabriel A. Al-Ghalith, Robin R. Shields-Cutler, Qiyun Zhu, Daryl M. Gohl, Kenneth B. Beckman, Rob Knight, and Dan Knights. 2018. “Evaluating the Information Content of Shallow Shotgun Metagenomics.” *mSystems* 3 (6). https://doi.org/10.1128/mSystems.00069-18.
Louca, Stilianos, Martin F. Polz, Florent Mazel, Michaeline B. N. Albright, Julie A. Huber, Mary I. O’Connor, Martin Ackermann, et al. 2018. “Function and Functional Redundancy in Microbial Systems.” *Nature Ecology & Evolution* 2 (6): 936–43.
Metcalf, J. L., L. Wegener Parfrey, A. Gonzalez, C. L. Lauber, D. Knights, G. Ackermann, G. C. Humphrey, et al. 2013. “A Microbial Clock Provides an Accurate Estimate of the Postmortem Interval in a Mouse Model System.” *eLife* 2: e01104.
Metcalf, J. L., Z. Z. Xu, S. Weiss, S. Lax, W. Van Treuren, E. R. Hyde, S. J. Song, et al. 2016. “Microbial Community Assembly and Metabolic Function during Mammalian Corpse Decomposition.” *Science* 351 (6269): 158–62.
Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. “Scikit-Learn: Machine Learning in Python.” *Journal of Machine Learning Research: JMLR* 12 (Oct): 2825–30.
Sinha, R., G. Stanley, G. S. Gulati, C. Ezran, and K. J. Travaglini. 2017. “Index Switching Causes ‘spreading-of-Signal’ among Multiplexed Samples in Illumina HiSeq 4000 DNA Sequencing.” *BioRxiv*. https://www.biorxiv.org/content/biorxiv/early/2017/04/09/125724.full.pdf.
Sumner, Lloyd W., Alexander Amberg, Dave Barrett, Michael H. Beale, Richard Beger, Clare A. Daykin, Teresa W-M Fan, et al. 2007. “Proposed Minimum Reporting Standards for Chemical Analysis.” *Metabolomics: Official Journal of the Metabolomic Society* 3 (3): 211–21.
Wang, M. X., J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, et al. 2016. “Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking.” *Nature Biotechnology* 34 (8): 828–37.
Young, J. M., L. S. Weyrich, J. Breen, L. M. Macdonald, and A. Cooper. 2015. “Predicting the Origin of Soil Evidence: High Throughput Eukaryote Sequencing and MIR Spectroscopy Applied to a Crime Scene Scenario.” *Forensic Science International* 251 (June): 22–31.
Young, J. M., L. S. Weyrich, and A. Cooper. 2014. “Forensic Soil DNA Analysis Using High-Throughput Sequencing: A Comparison of Four Molecular Markers.” *Forensic Science International. Genetics* 13 (November): 176–84.
Zhu, Qiyun, Uyen Mai, Wayne Pfeiffer, Stefan Janssen, Francesco Asnicar, Jon G. Sanders, Pedro Belda-Ferre, et al. 2019. “Phylogenomics of 10,575 Genomes Reveals Evolutionary Proximity between Domains Bacteria and Archaea.” *Nature Communications* 10 (1): 5477.