

1 **TECHNICAL NOTE**

2 **CRIMINALISTICS**

3 *Jill Hager Cocking,<sup>1,2</sup> M.S.; Sgt. Ryan Turley,<sup>3</sup> Viacheslav Y. Fofanov,<sup>1</sup> Ph.D.; Kimberly Samuels-Crow,<sup>1</sup>*  
4 *Ph.D.; Bruce Hungate,<sup>4,5</sup> Ph.D.; Rebecca L. Mau,<sup>2</sup> M.S.; Paul S. Keim,<sup>2</sup> Ph.D.; J. Gregory Caporaso,<sup>6</sup>*  
5 *Ph.D.; and Crystal Hepp,<sup>1</sup> Ph.D.*

6 **Forensic analysis of Soil Microbiomes: Linking Evidence to a**  
7 **Geographic Location**<sup>\*</sup>

8

9 <sup>1</sup> School of Informatics, Computing and Cyber Systems, Northern Arizona University, Flagstaff, Arizona

10 <sup>2</sup> Pathogen and Microbiome Institute, Northern Arizona University, Flagstaff, Arizona

11 <sup>3</sup> Flagstaff Police Department, Flagstaff, Arizona

12 <sup>4</sup> Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona

13 <sup>5</sup> Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, Arizona

14 <sup>6</sup> Center for Applied Microbiome Science, Pathogen and Microbiome Institute, Northern Arizona University,  
15 Flagstaff, Arizona

16 \* Presented at the 71<sup>st</sup> Annual Meeting of the American Academy of Forensic Sciences, February 18-23, 2019, in  
17 Baltimore, MD.

18

19 **ABSTRACT:** Over the past two decades, advances in molecular biology have greatly expanded our  
20 understanding of microbiomes – the diverse assemblages of microorganisms that inhabit the human body  
21 as well as the world around us, and applications in microbiome science have become an active area of  
22 research. Differences in the diversity (i.e., richness) and composition of microbiomes has been found to be  
23 informative in varied areas of science, including human health, agronomy, and forensic science. Soil  
24 harbors microbiomes that vary based on many factors, including the geology of the soil (e.g., sand, silt, or  
25 clay), climate, and use of the soil. As a result, the microbiological composition of any two soil samples will  
26 never be exactly alike. This inherent variation between microbiomes of different locations has proven to be  
27 specific enough to be potentially useful in forensic investigations to associate a person or piece of evidence  
28 to a source site.

29           In this study, a soil microbiome was extracted from the sock of a criminal suspect and compared to  
30 the microbiome of soil samples taken from locations traveled to by the suspect. The locations analyzed  
31 varied in their soil microbiome composition, and the microbiome profiled from the sock was found to be  
32 most similar to the location where the suspect was thought to have left the body of a murder victim. These  
33 results provide a case study illustrating that information contained in a soil microbiome may be applied to  
34 link evidence to the location where a crime took place, potentially serving as an investigative tool in law  
35 enforcement.

36 **KEYWORDS:** forensic science, microbiome, soil, QIIME 2, DNA, sock

37           Recent advances in DNA sequencing and bioinformatics analysis methods have led to a better  
38 understanding of microbiomes, communities of different species of microscopic organisms whose  
39 metabolisms are tightly linked to one another, to their environments, or to their plant and animal hosts.  
40 These technological advances have led to our recent recognition that there are orders of magnitude more  
41 microbial species than were previously thought to have existed (1). This has led to a revolution in  
42 microbiology, with new sub-fields of science forming to understand the composition and functional  
43 activities of microbiomes. The microbiome of the skin and inside the human body varies substantially  
44 depending on the body location sampled and has revealed itself as an important indicator of health (2). For  
45 example, individuals with chronic conditions like Crohn's Disease and chronic rhinosinusitis harbor gut (3)  
46 and sinus microbiomes (4), respectively, with lower richness (i.e., fewer different species present) than the  
47 microbiomes of healthy individuals. Similarly, individuals with acute viral gastroenteritis also have less  
48 diverse gut microbiomes (5). Although fecal supplementation following antibiotic treatment has seen fringe  
49 use in medicine since the 1950s (6), we are now beginning to formally develop applications of this  
50 knowledge in human health. For example, the transplantation of a healthy individual's gut microbiome into  
51 a patient suffering from recurring *Clostridium difficile* infections is now becoming common (7).

52           Microbiome analysis techniques extend beyond human health and already have some recognized  
53 forensic applications. For example, skin microbiomes differ between human subjects, and enough of this

54 unique microbial fingerprint is left behind on the objects that we touch that objects, such as keyboards and  
55 computer mice, can be linked to their owner solely based on the microbes found on their surfaces (8).  
56 Similarly, the bacterial community microbiome found in human saliva was shown to be unique to an  
57 individual (9). Even time of death information may be gleaned from the succession of microbial  
58 communities which are involved in decomposition (10). Taken together, these studies suggest a role for  
59 microbiome science in forensics applications.

60 Microbes, and bacteria and fungi in particular, are integral components of soil. The specific  
61 microbial composition of a soil sample is driven by factors including the physicochemical properties of the  
62 soil (e.g., pH, salinity) (11, 12), environmental features (13), and land use (14), such that no two soil samples  
63 will ever be exactly alike. This variation between soils from different locations has proven to be specific  
64 enough to be potentially useful in forensic investigations (15).

65 We therefore hypothesized that soil extracted from a piece of evidence may provide sufficient  
66 information to link that evidence to the soil's source.

67 In October of 2017, our laboratory was contacted by the Flagstaff Police Department. A woman  
68 was missing and presumed dead. A suspect was in custody but was not revealing the location of the missing  
69 woman. Using various investigative tools, the police knew where the suspect had traveled since being  
70 released on bail from the Flagstaff jail a few days earlier. The police were trying to decide whether to focus  
71 their search efforts in the town of Mayer, Arizona (Yavapai County) or Williams, Arizona (Coconino  
72 County), with a distance between the two of 87 to 118 miles depending on the route taken. A sock embedded  
73 with soil, believed to have been worn by the suspect while not wearing a shoe, was in the custody of the  
74 police. Their hope was that the sock could be analyzed and linked to one of the two locations to aid in  
75 finding the body. During the following few days (before the microbiome analysis was completed), the body  
76 was recovered in Mayer, Arizona. Even though the police no longer needed assistance with the recovery of  
77 the body, we attempted to analyze soil embedded in the sock in order to compare it to various locations  
78 around Yavapai and Coconino Counties in the state of Arizona to determine if the soil microbiome could

79 be a useful forensic tool in future investigations. Knowing that there are inherent variations in soil microbial  
80 community depending on location, we compared the bacterial community found on the sock with bacterial  
81 communities found in multiple soil samples collected from locations traveled to by the suspect during the  
82 days prior to his arrest.

83

## 84 **Methods**

85 Reference surface soil samples were obtained from 18 locations around Williams (Coconino  
86 County) and Mayer (Yavapai Country), Arizona (Table 1). Locations were selected because they were  
87 either near the site where the body was found (Fig. 1) or were places the suspect was known to have traveled  
88 in the days after he was last seen with the victim. Five soil samples collected from other locations in Arizona  
89 were also obtained from the Center for Ecosystem Science and Society (EcoSS) at Northern Arizona  
90 University to serve as additional reference samples (the “EcoSS reference samples”). Together this resulted  
91 in 23 reference samples where K1-11 and K13 refer to the Mayer soil samples, K12, K14-16, and K18 refer  
92 to the Williams soil samples, K17 refers to the Chino Valley soil sample and W0.GL.1, W0.PJ.1, W0.MC.1,  
93 W0. PP.1 and 217 refer to the EcoSS reference samples.

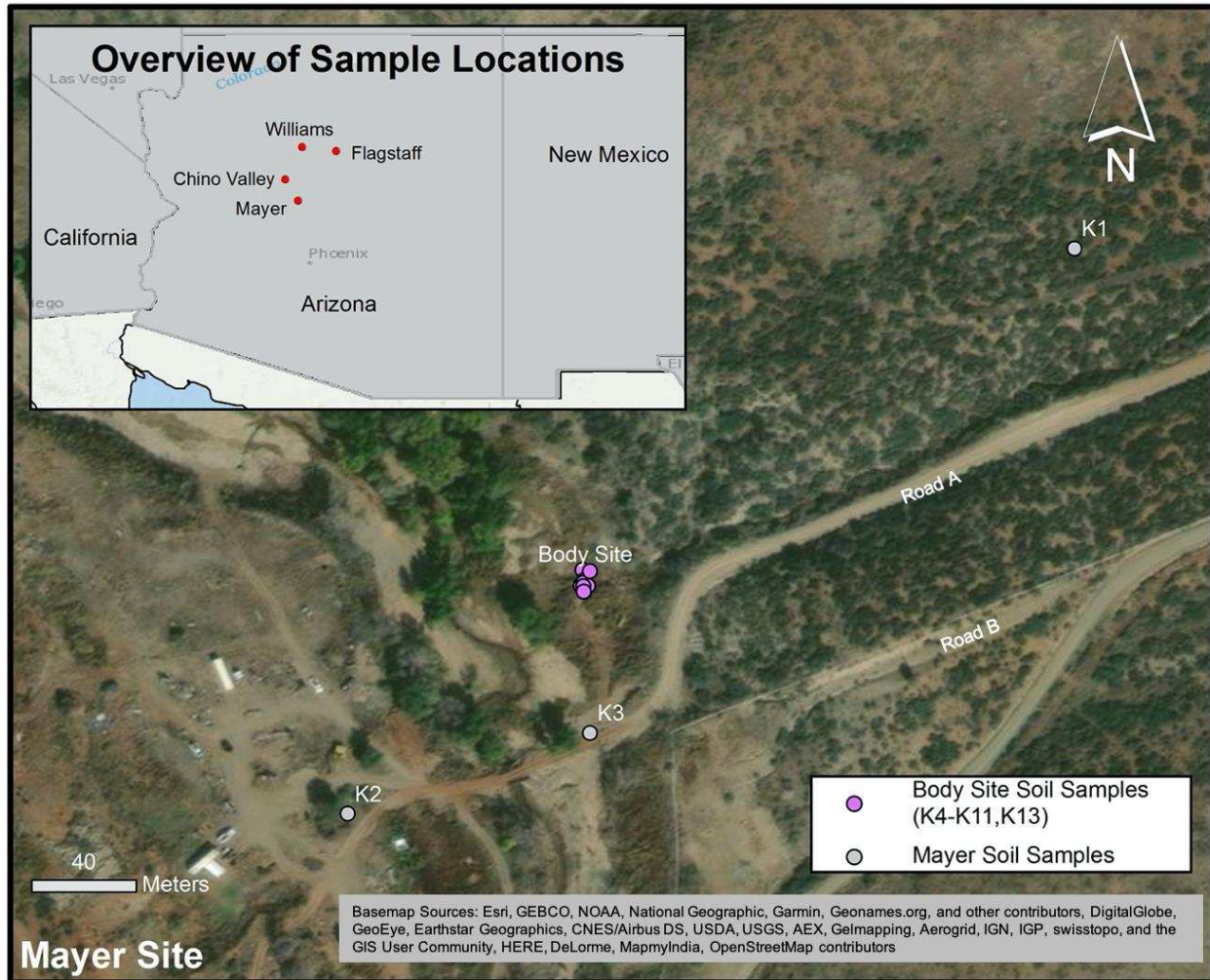
94 TABLE 1- *Information for samples used for analysis.*

Sample Name	Description
Q1	small cutting from sock, ball of foot area
Q2	small cutting from sock, heel area
Q3	swabbing of visibly dirt-covered area of sock for 3 minutes
Q4	swabbing of visibly dirt-covered area of sock for 3 minutes
K1	soil, right off highway Mayer, AZ
K2	soil, Road by trailer of owner of land where body was found, Mayer, AZ
K3	soil, off Road A, Mayer, AZ
K4	soil, dry, cracked ground near where body was found, Mayer, AZ
K5	soil, near body site, Mayer, AZ
K6	soil, 5 ft west of body site decomposition area, Mayer, AZ

K7	soil, 5 ft north of body site decomposition area, Mayer, AZ
K8	soil, 5 ft east of body site decomposition area, Mayer, AZ
K9	soil, body site with visual decomposition residue, Mayer, AZ
K10	soil, body site with visual decomposition residue, Mayer, AZ
K11	soil, behind body site, more plants, Mayer, AZ
K12	soil, Williams, AZ
K13	soil, 5 ft south of body site decomposition area, Mayer, AZ
K14	soil, Williams, AZ
K15	soil, off Forest Service Road, Williams, AZ
K16	soil, Forest Service Road, Williams, AZ
K17	Soil, Chino Valley, AZ
K18	Soil, Williams, AZ
W0.GL.1	Grassland site of C. Hart Merriam Gradient, Northern Arizona
W0.PJ.1	Pinyon-Juniper site of C. Hart Merriam Gradient, Northern Arizona
W0.MC.1	Mixed Conifer site of C. Hart Merriam Gradient, Northern Arizona
W0.PP.1	Ponderosa Pine site of C. Hart Merriam Gradient, Northern Arizona
217	Mixed Conifer site of C. Hart Merriam Gradient, Northern Arizona
Reagent Blank Sock (RBsock)	Reagent Blank extracted alongside sock samples
Reagent Blank Soil (RBsoil)	Reagent Blank extracted alongside soil samples

95

96



97

98 FIG. 1 - Locations of reference soil samples collected at and near the body site in Mayer, Arizona. Inset  
99 map shows geographic relationships between the Mayer site and other locations relevant to the case.

100 Mayer Soil Samples include K1, K2, and K3. Body site soil samples include K4-K11 and K13 (Table 1).

101

102 The sock belonging to the suspect was visibly soiled over most of its surface. DNA extraction  
103 from the sock was performed in two ways with two replicates per extraction approach, yielding four query  
104 samples, referred to here as Q 1-4. Two DNA samples were extracted from cuttings of the sock itself  
105 (approximately 2 cm<sup>2</sup> each). One cutting was taken from the ball of the foot area (Q1) and one cutting was  
106 taken from the heel area (Q2). For the second extraction method, 2 pre-moistened cotton swabs (Q3, Q4)

107 were rubbed over the visibly soiled area of the outside of the sock for 3 minutes each, as described by Goga  
108 (16). The swabs were then removed from the applicator and used for the remainder of the extractions. The  
109 extractions were performed using the Qiagen DNeasy® PowerSoil Kit (Qiagen, Germantown, MD)  
110 according to the manufacturer's protocol, with the following variations. The cuttings and swabs were placed  
111 at 65°C for 10 minutes followed by 2 minutes of horizontal vortexing at the maximum speed of the vortexer.  
112 The final elution volume was 100 µL.

113 For the reference soil samples (K1-18), approximately 0.25 grams of each of the 18 soil samples  
114 (Table 1) were added to a PowerBead tube containing solution C1. This tube was placed at 65°C for 10  
115 minutes followed by 2 minutes of horizontal vortexing at maximum speed. The remainder of the extraction  
116 was performed according to the Qiagen DNeasy® PowerSoil Kit's manufacturer's instructions.

117 The EcoSS reference soil samples differed from the reference samples collected for this study in  
118 that they were collected below the soil surface (0-10 cm) while K1-18 were taken from the surface (as the  
119 surface soil would be the most likely to come into contact with the suspect's sock). There is known variation  
120 in soil microbiome composition depending on sampling depth (17), but these samples were included to  
121 provide additional background soils to which we could compare our query samples.

122 Extractions from Q1-4 and K1-18 were performed at different times in 2017. DNA was extracted  
123 from the EcoSS reference samples in 2014 and 2015 by using a MO BIO PowerSoil™ DNA Isolation Kit  
124 (Qiagen, Germantown, MD) and following the manufacturer's directions. Briefly, approximately 0.25 g of  
125 soil was added to the lysis tube and lysed using a MP Biomedicals FastPrep Homogenizer (MP Biomedicals,  
126 Irvine, CA). The final elution volume was 100 µL.

127 During the extractions for Q1-4 and K1-18, a reagent blank was taken through the entire process to  
128 monitor laboratory and extraction reagent contamination. For the sock samples, the reagent blank (RBsock)  
129 consisted of a cotton tipped swab moistened with UltraPure distilled H<sub>2</sub>O and cut with scissors used for the  
130 sock and handled with tweezers used for the sock extraction. The reagent blank was processed alongside

131 the sock extractions. For the reference soil reagent blank (RBsoil), water was added to a weigh boat and  
132 then placed in a PowerBead tube and processed alongside the known soil samples.

133 The hypervariable V4 region of the 16S rRNA gene was amplified from each of the reference soil  
134 samples and sock samples, as well as the reagent blanks. This amplified DNA was then prepared for  
135 sequencing on the Illumina MiSeq instrument according to the protocol presented by Caporaso *et al* (18).  
136 The resulting sequences were analyzed using QIIME 2 microbiome bioinformatics platform (19). Sequence  
137 quality control was performed using the denoise-paired method of QIIME 2's DADA2 (20) plugin with the  
138 following parameter settings: trunc\_len\_f 293; trunc\_len\_r 208; trim\_left\_f 6, trim\_left\_r 6. The resulting  
139 amplicon sequence variants were assigned taxonomy using q2-feature-classifier's classify-sklearn method  
140 against (21) GreenGenes (22) 13\_8. ASV sequences were aligned using MAFFT (23) (qiime alignment  
141 mafft), highly variable positions were filtered (qiime alignment mask), an unrooted tree was constructed  
142 using FastTree (24) (qiime phylogeny fasttree), and the tree was rooted by midpoint rooting (qiime  
143 phylogeny midpoint-root). Weighted and unweighted UniFrac (25) distances were computed 100 times  
144 each at an even sampling depth of 1000 sequences per sample. This low depth of coverage was used to  
145 retain all samples in the analysis, and 100 iterations were run to confirm that conclusions were robust across  
146 rarefied feature tables. These analyses were performed using the beta-rarefaction visualizer in QIIME 2's  
147 diversity plugin. Sample tree illustrations were generated with ete3 (26).

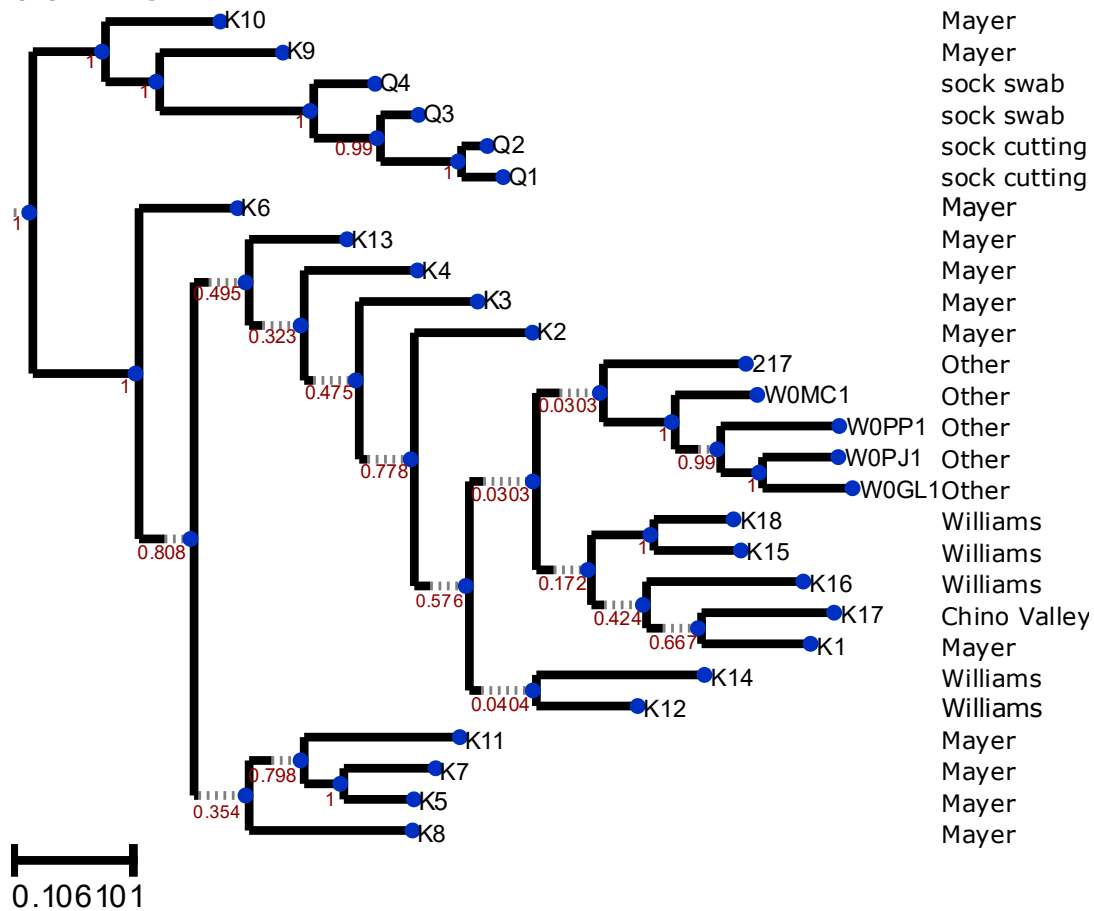
## 148 **Results**

149 DNA was successfully extracted and the V4 region of the 16S rRNA gene was amplified from the  
150 4 sock samples, the 2 reagent blanks, and the 23 reference soil samples. Weighted and unweighted UniFrac  
151 neighbor joining trees were constructed to evaluate the similarity of microbiomes (Fig. 2a and 2b,  
152 respectively). Briefly, the UniFrac metrics provide distances between pairs of microbiome samples. Smaller  
153 values indicate that a pair of samples are similar in their composition, while larger values indicate that a  
154 pair of samples are dissimilar in their composition. The unweighted UniFrac metric is considered a  
155 qualitative metric in that it only compares samples based on which microbes are present, but does not

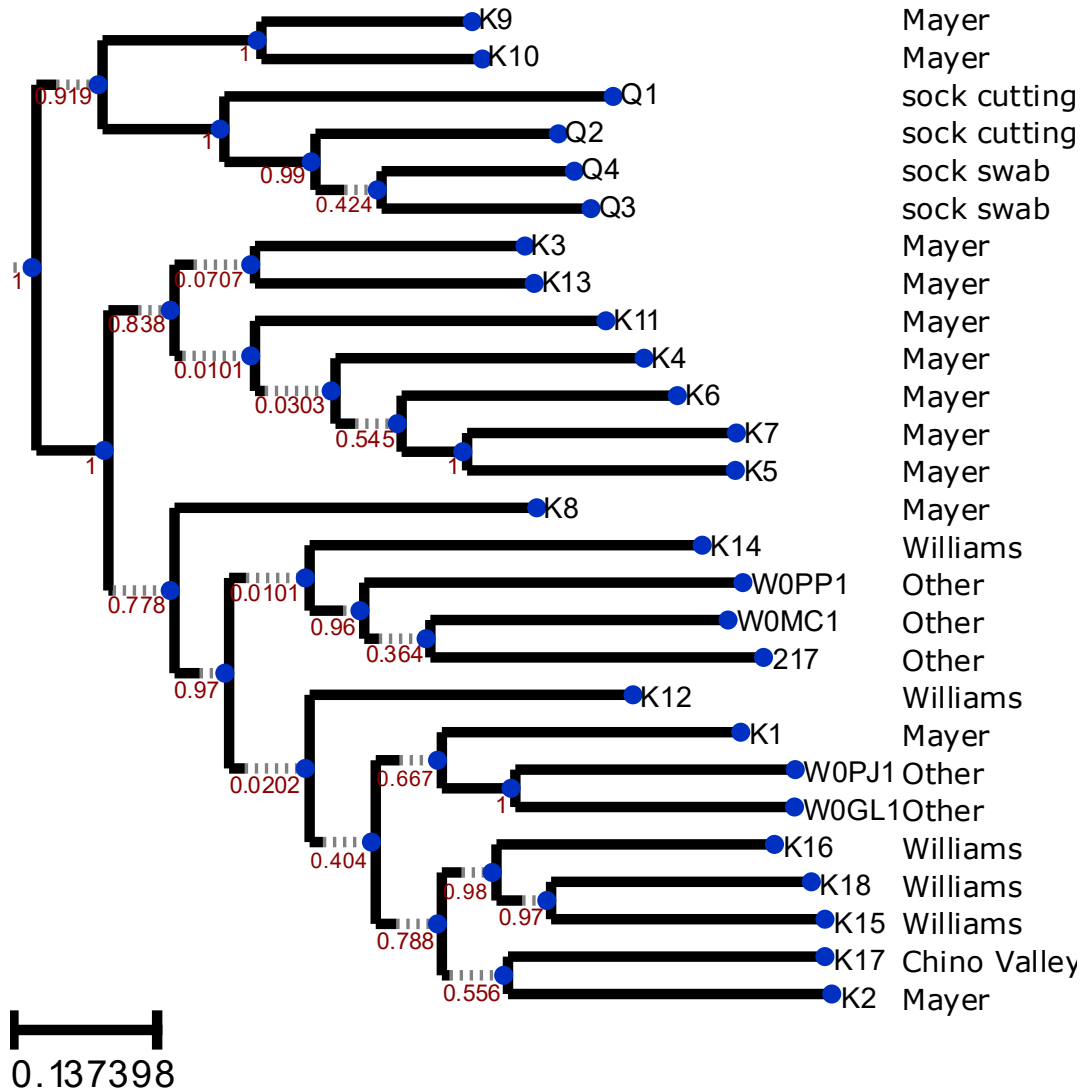


156 consider the abundance of those microbes. The weighted UniFrac metric is considered a quantitative metric  
 157 because it compares the abundances of different microbes in the samples. Because estimation of microbial  
 158 abundances is imperfect using the techniques applied for microbiome profiling, both weighted and  
 159 unweighted UniFrac metrics are often computed and compared. These metrics are applied to compute  
 160 distances between all pairs of microbiome samples, and the resulting distance matrix can be summarized  
 161 by constructing a neighbor joining tree. In this tree, samples are represented as the leaves (or tips), and the  
 162 length of the branches between leaves represents the distance between the samples.

(a) Weighted UniFrac



(b) Unweighted UniFrac



164

165 FIG. 2- Neighbor joining trees illustrating (a) unweighted UniFrac and (b) weighted UniFrac distances

166 between samples. Leaves of the trees represent samples, and the branch length between pairs of leaves

167 represents the dissimilarity between samples. Values above the internal nodes of the tree represent

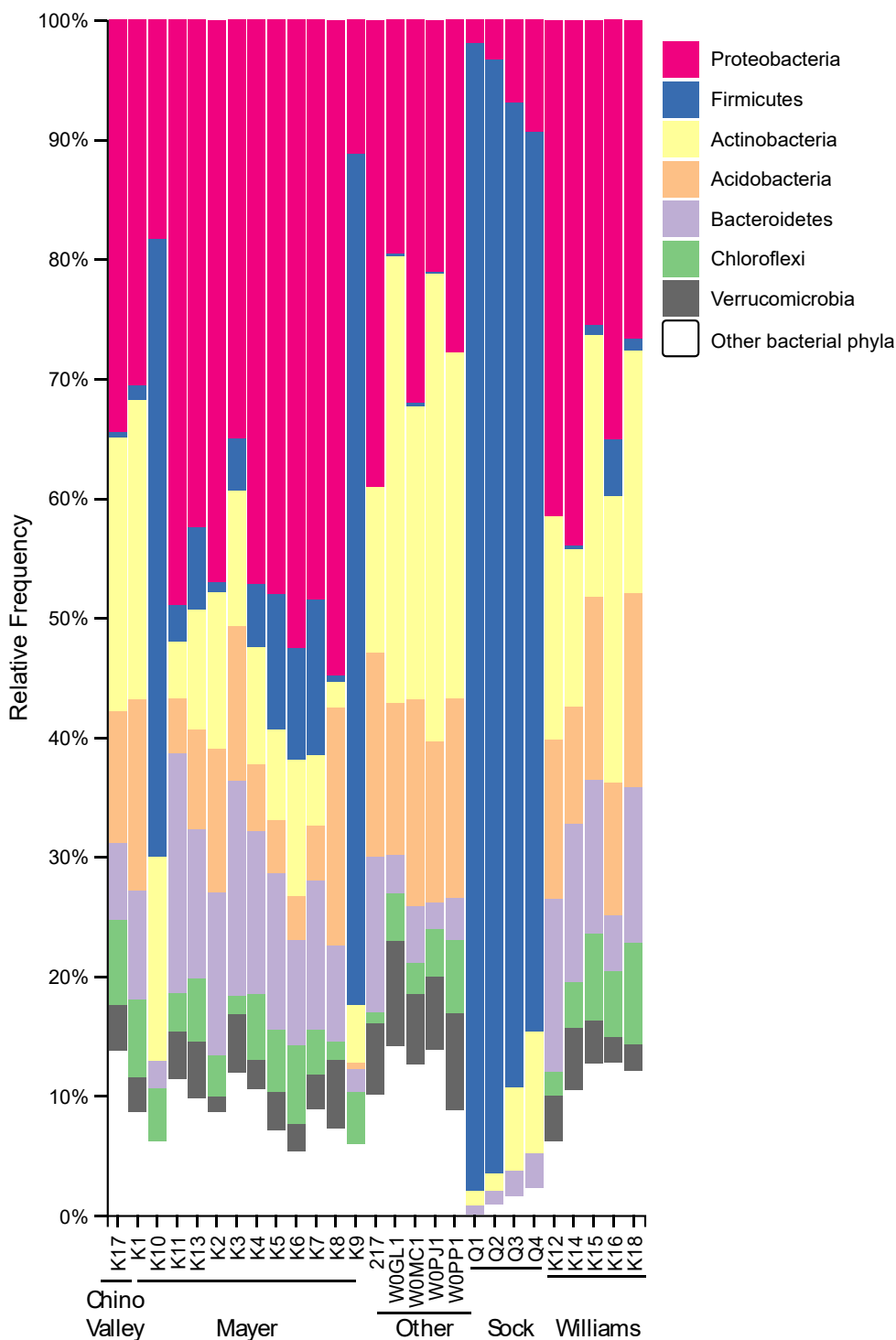
168 jackknife support values, ranging between 0 and 1. Larger values indicate more robust groupings of

169 samples.

170

171 Both the weighted and unweighted UniFrac neighbor joining trees illustrate that the sock samples  
172 are all most similar to each other in composition, and that the closest soil samples are all from Mayer, where  
173 the suspect left the remains of the victim. Clustering of the sock and Mayer soil samples was highly robust,  
174 and suggest that the soil on the suspect's sock could have informed investigators of which cities should be  
175 the focus of search efforts.

176 Analysis of the taxonomic composition revealed typical soil microorganisms for all soil samples  
177 (Fig. 3). As would be expected, the dominant microorganisms in the sock sample were taxa commonly  
178 found on human skin. Because the victim's body was left at sites K9 and K10, we were concerned that skin  
179 microbes found at those sites would link those samples to the sock, irrespective of the soil microbial  
180 composition. We therefore performed parallel analyses to those presented here after filtering the dominant  
181 skin bacterial family found here, Staphylococcaceae, from the sock and soil samples. This resulted in the  
182 taxonomic compositions presented in Fig. S2. The sock samples were still most similar to the Mayer  
183 samples, even after removal of all Staphylococcaceae (Fig. S1).



184

185 FIG 3- Microbiome taxonomic composition at the phylum levels for all samples.

186

187

188           The sock swabbing technique obtained more of the soil profile with less human-associated  
189 microbes compared to the cutting of the sock method (Fig. 3), though all of the sock samples clustered  
190 together in our analyses, suggesting that either approach would have led us to the same conclusion (Fig. 2).  
191 Because the swabbing technique produced less human-associated microbes and was not destructive of the  
192 evidence, this approach is likely a better choice than extraction of DNA from the sock cuttings.

193           Microbial DNA amplification was observed in the reagent blanks. This is to be expected as bacteria  
194 are ubiquitous in any environment so careful monitoring of contamination from laboratory equipment and  
195 reagents is crucial (27). Reagent blanks were processed alongside both the sock and reference soil samples.  
196 Although bacteria were present in both reagent blanks, the composition and abundance varied greatly from  
197 the reference soil samples and the query sock samples (Fig. S3).

## 198 **Discussion**

199           The QIIME 2 platform was applied for analysis of microbiome data in this study. QIIME 2's  
200 retrospective data provenance tracking feature may prove to be helpful in microbiome-based forensics  
201 work. All analysis steps, including versions of software installed on the system when each step was run, are  
202 automatically tracked as metadata associated with its results. This would allow an expert to determine with  
203 complete certainty what computational steps were taken to generate a result. As DNA analysis workflows  
204 can be complex, this automated recording will provide experts with the information they need to be  
205 confident in a given result or to identify potential issues such as the presence of a software bug or suboptimal  
206 analysis step in a workflow that may impact conclusions drawn from the data. Data provenance can be  
207 viewed for the results generated for this paper by loading the QIIME 2 results from Supplementary File 1  
208 with QIIME 2 View (<https://view.qiime2.org>).

209           The ability to associate a piece of evidence to a location is a valuable tool to law enforcement. In  
210 the case presented here, a murder suspect was known to have traveled over a long distance during a few

211 days' time. Other evidence pointed to this individual having committed a murder, and both police and the  
212 victim's family were anxious to discover the remains of the victim. With the police having narrowed down  
213 some possible locations for the body site, we were able to link an item of evidence to the location where  
214 the victim's body was left. This result was possible because we were able to create a small database of  
215 locations known to have been visited by the suspect through police investigative techniques. Although the  
216 victim's body was located with the assistance of the suspect, we believe that had this not happened, we  
217 would have been able to advise law enforcement that the soil embedded in the suspect's sock most likely  
218 came from the Mayer, Arizona area rather than other locations where soil was collected based on the data  
219 presented here. In cases where areas coming in contact with the item of evidence are not known, a database  
220 of known soils from across a county or even a state would be very useful.

221

## 222 Acknowledgements

223 Microbiome analysis and the QIIME 2 platform were partially funded by National Science Foundation  
224 grant number 1565100 to JGC. Laboratory supplies, sequencing, and personnel time were covered by the  
225 Arizona Board of Regents Technology Infrastructure Research Initiative Fund as startup funds to CMH.

226 Ecosystem reference soil samples were collected with support by the National Science Foundation, Division of  
227 Environmental Biology (grant number: DEB-1241094).

228

229

230

231

232

233

234 **References**

- 235 1. Hugenholtz P. Exploring prokaryotic diversity in the genomic era. *Genome Biology* 2002; 3(2):  
236 reviews0003.1–reviews0003.8.
- 237 2. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, et al. Structure,  
238 function and diversity of the healthy human microbiome. *Nature* 2012; 486: 207-214.
- 239 3. Gevers D, Kugathasan S, Denson LA, Vazquez-Baeza Y, Van Treuren W, Ren B, et al. The  
240 treatment-naïve microbiome in new-onset Crohn’s disease. *Cell Host Microbe* 2014; 15(3):  
241 382-392.
- 242 4. Cope EK, Goldberg AN, Pletcher SD, Lynch SV. Compositionally and functionally distinct  
243 sinus microbiota in chronic rhinosinusitis patients have immunological and clinically divergent  
244 consequences. *Microbiome* 2017; 5:53.
- 245 5. Chen SY, Tsai CN, Lee YS, et al. Intestinal microbiome in children with severe and  
246 complicated acute viral gastroenteritis. *Sci Rep* 2017;7:46130.
- 247 6. De Groot PF, Frissen MN, de Clercq NC, Nieuwdorp M. Fecal microbiota transplantation in  
248 metabolic syndrome: History, present and future. *Gut Microbes* 2017; 8 (3): 253-267.
- 249 7. Allegretti JR, Mullish BH, Kelly C, Fischer M. The evolution of the use of faecal microbiota  
250 transplantation and emerging therapeutic indications. *The Lancet Therapeutics* 2019; 394  
251 (10196): 420-431.
- 252 8. Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R. Forensic identification  
253 using skin bacterial communities. *PNAS* 2010; 107 (14): 6477-6481.
- 254 9. Leake SL, Pagni M, Falquet L, Taroni F, Greub G. The salivary microbiome for differentiating  
255 individuals: proof of principle. *Microbes Infect* 2016; 18: 399-405.
- 256 10. Metcalf JL, Xu ZZ, Wiess S, Lax S, Van Treuren W, Hyde ER, et al. Microbial community  
257 assembly and metabolic function during mammalian corpse decomposition. *Science* 2016; 351:  
258 158-162.
- 259 11. Ma J, Ibekwe AM, Yang CH, Crowley DE. Bacterial diversity and composition in major fresh  
260 produce growing soils affected by physiochemical properties and geographic locations. *Sci*  
261 *Total Environ* 2016; 563-564: 199-209.
- 262 12. Pershina EV, Ivanoca EA, Korvigo IO, Chirak EL, Sergaliev JH, Abakumov EV, et al.  
263 Investigation of the core microbiome in main soil types from the East European plain. *Sci Total*  
264 *Environ* 2018; 631-632: 1421-1430.
- 265 13. Neilson JW, Califf K, Cardona C, Copeland A, van Treuren W, Josephson KL, et al. Significant  
266 Impacts of Increasing Aridity of the Arid Soil Microbiome. *mSystems* 2017; 2 (3).
- 267 14. Lacerda-Junior GV, Noronha MF, Cabral L, Delforno TP, Sousa STPD, Fernandes-Junior PI,  
268 et al. Land Use and Seasonal Effects on the Soil Microbiome of a Brazilian Dry Forest. *Front.*  
269 *Microbiol* 2019.
- 270 15. Jesmok EM, Hopkins JM, Foran DR. Next-Generation Sequencing of the Bacterial 16S rRNA  
271 Gene for Forensic Soil Comparison: A Feasibility Study. *Journal of Forensic Science* 2016; 61  
272 (3): 607-617.
- 273 16. Goga H. Comparison of bacterial DNA profiles of footwear insoles and soles of feet for the  
274 forensic discrimination of footwear owners. *Int J Legal Med* 2012; 126: 815-823.
- 275 17. Eilers KG, Debenport S, Anderson S, Fierer N. Digging deeper to find unique microbial  
276 communities: The strong effect of depth on the structure of bacterial and archaeal communities  
277 in soil. *Soil Biology & Biochemistry* 2012; 50: 58-65.
- 278

- 279 18. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-  
280 throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms.  
281 International Society for Microbial Ecology 2012; 6: 1621-1624.
- 282 19. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al.  
283 Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2.  
284 Nature Biotechnology 2019; 37: 852-857.
- 285 20. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AA, Holmes SP. DADA2: High-  
286 resolution sample inference from Illumina amplicon data. Nature Methods 2016; 13: 581-583.
- 287 21. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Knight R, Huttley GA, Caporaso JG.  
288 Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-  
289 feature-classifier plugin. Microbiome 2018; 6 (90).
- 290 22. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved  
291 Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria  
292 and archaea. ISME J. 2001; 6 (3): 610-618.
- 293 23. Katoh K, Standley DM. MAFFT Multiple sequence alignment software version 7 :  
294 Improvements in performance and Usability. Molecular Biology and Evolution 2013; 30 (4):  
295 772-780.
- 296 24. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for  
297 Large Alignments. PLOS One 2010; <https://doi.org/10.1371/journal.pone.0009490>.
- 298 25. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative  $\beta$  Diversity  
299 measures lead to different insights into factors that structure microbial communities. Applied  
300 and Environmental Microbiology 2007; 1576-1585.
- 301 26. Huerta-Cepas J, Serra F, Bork P. ETE 3: Reconstruction, analysis, and visualization of  
302 phylogenomic data. Mol Biol Evol 2016; 33(6): 1635-1638.
- 303 27. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and  
304 laboratory contamination can critically impact sequence-based microbiome analyses. BMC  
305 Biol 2014; 12:87.

306

307

308

309

310

311

312

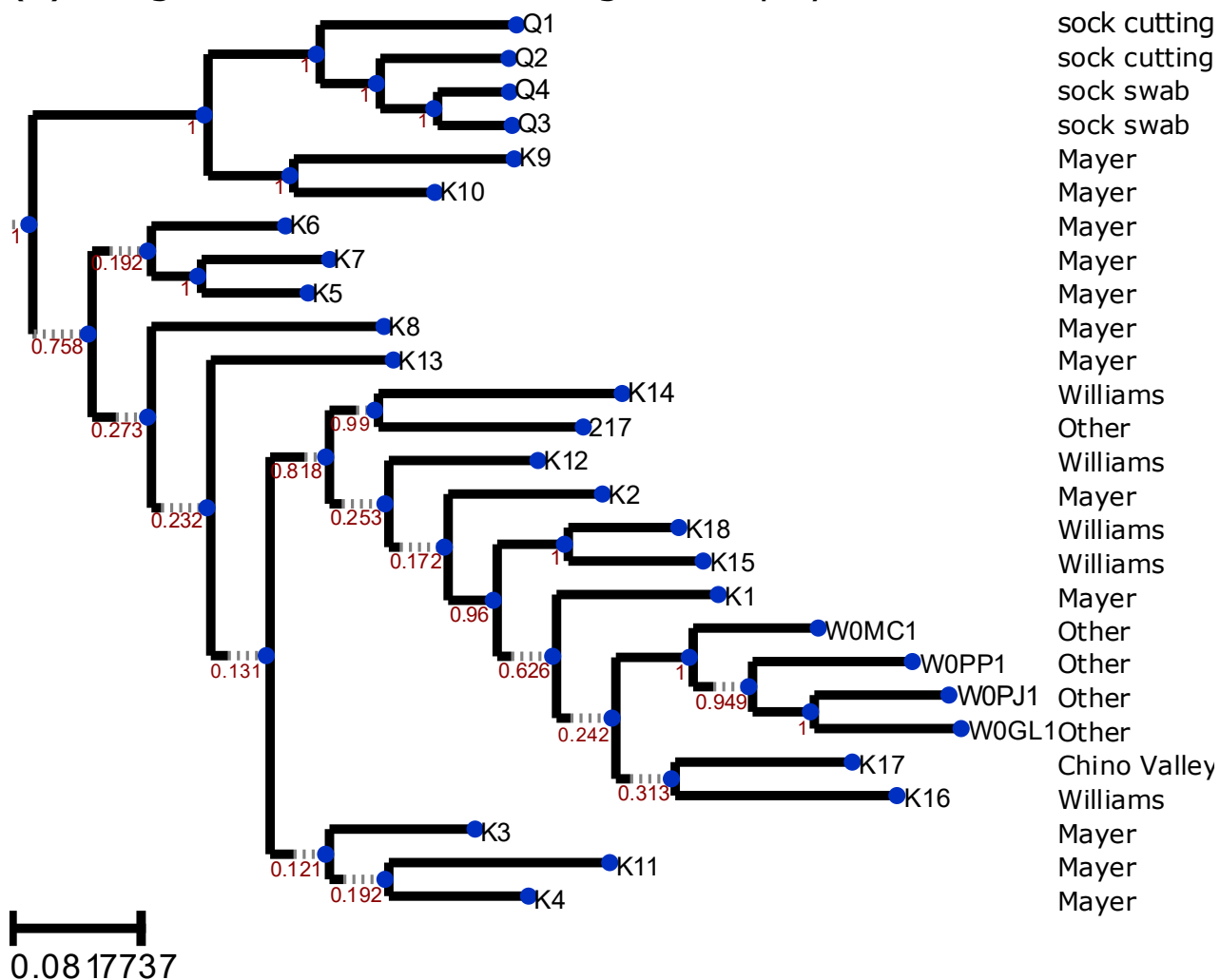
313



314

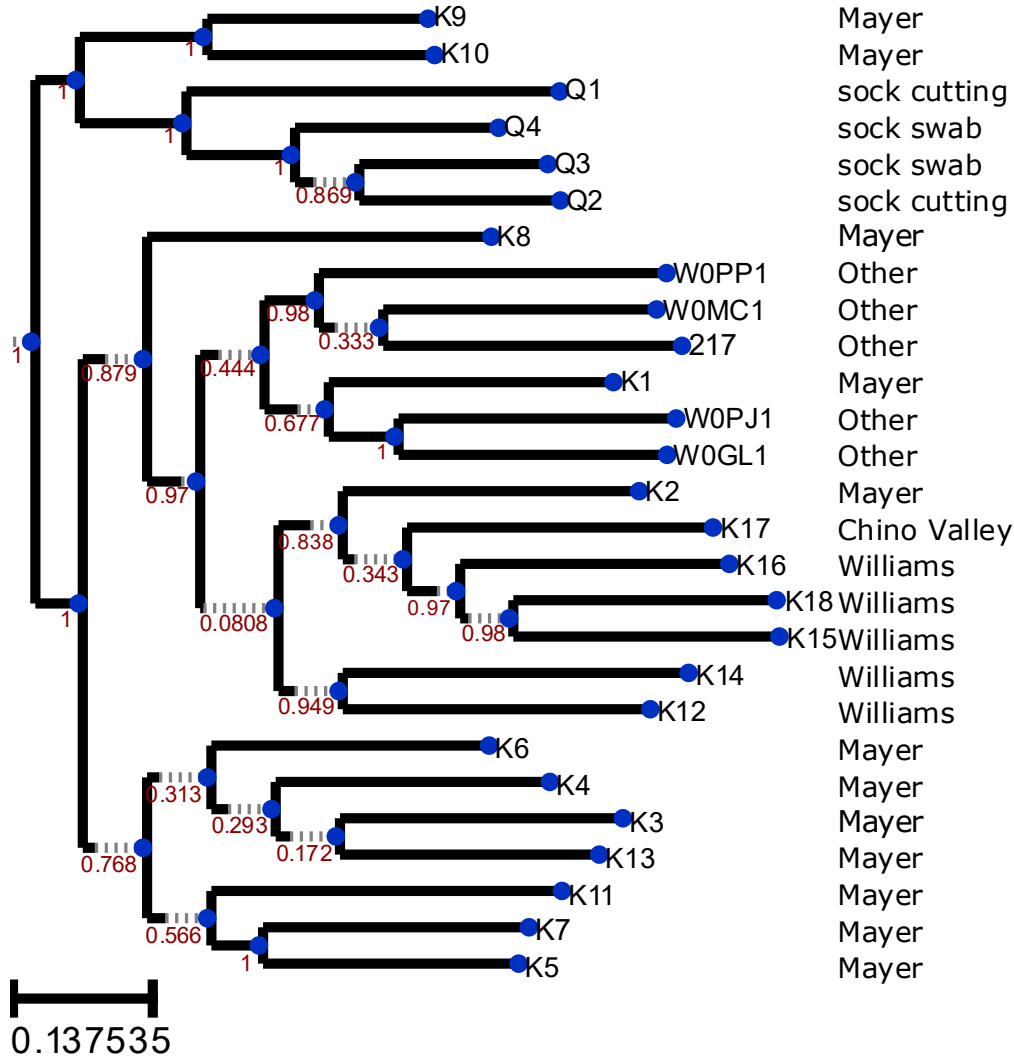
315 **Supporting Information**

(a) Weighted UniFrac excluding all Staphylococcaceae



316

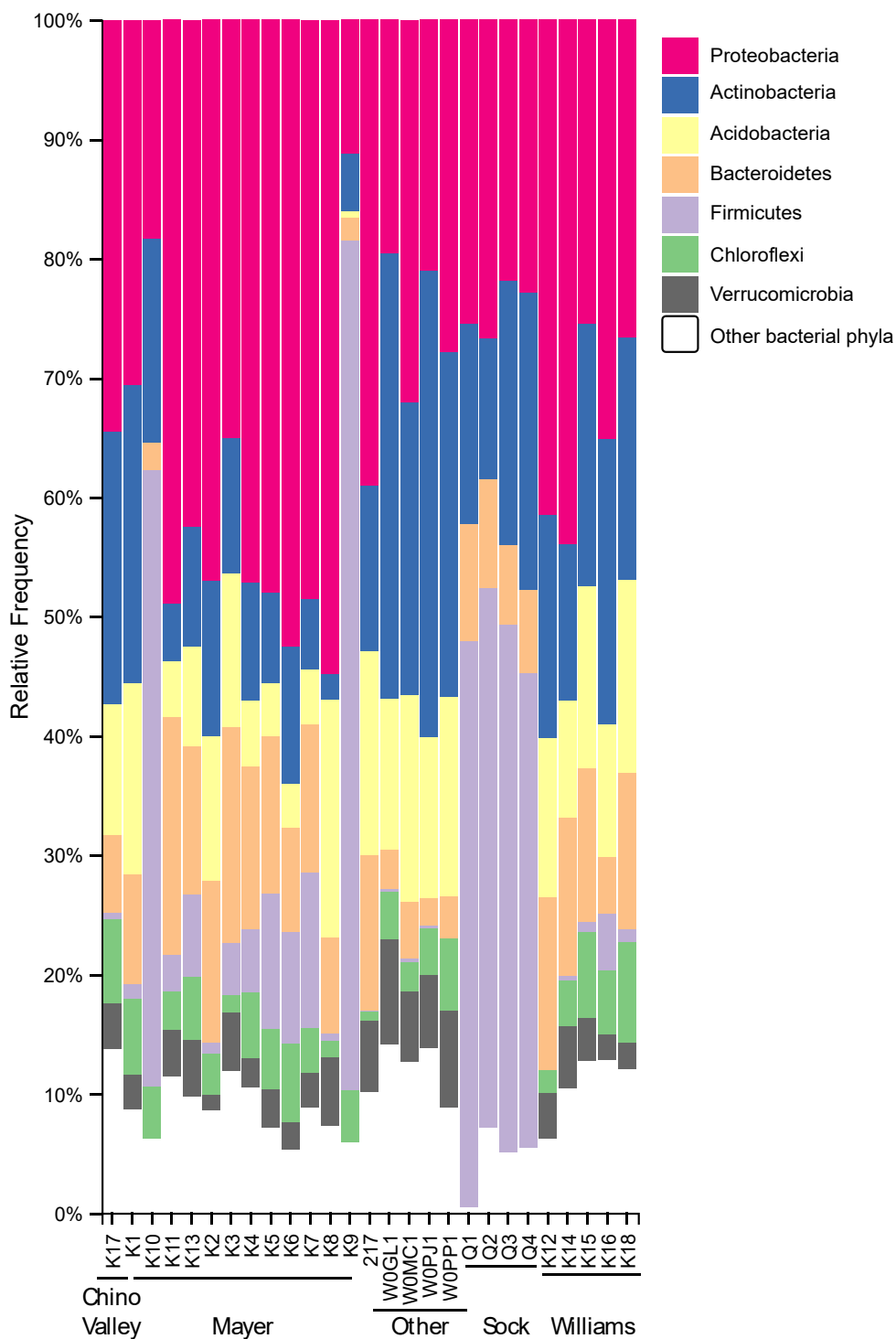
(b) Unweighted UniFrac excluding all Staphylococcaceae



317

318 FIG. S1- Neighbor joining trees illustrating (a) unweighted Unifrac and (b) weighted Unifrac distances

319 between samples after excluding all Staphylococcaceae.

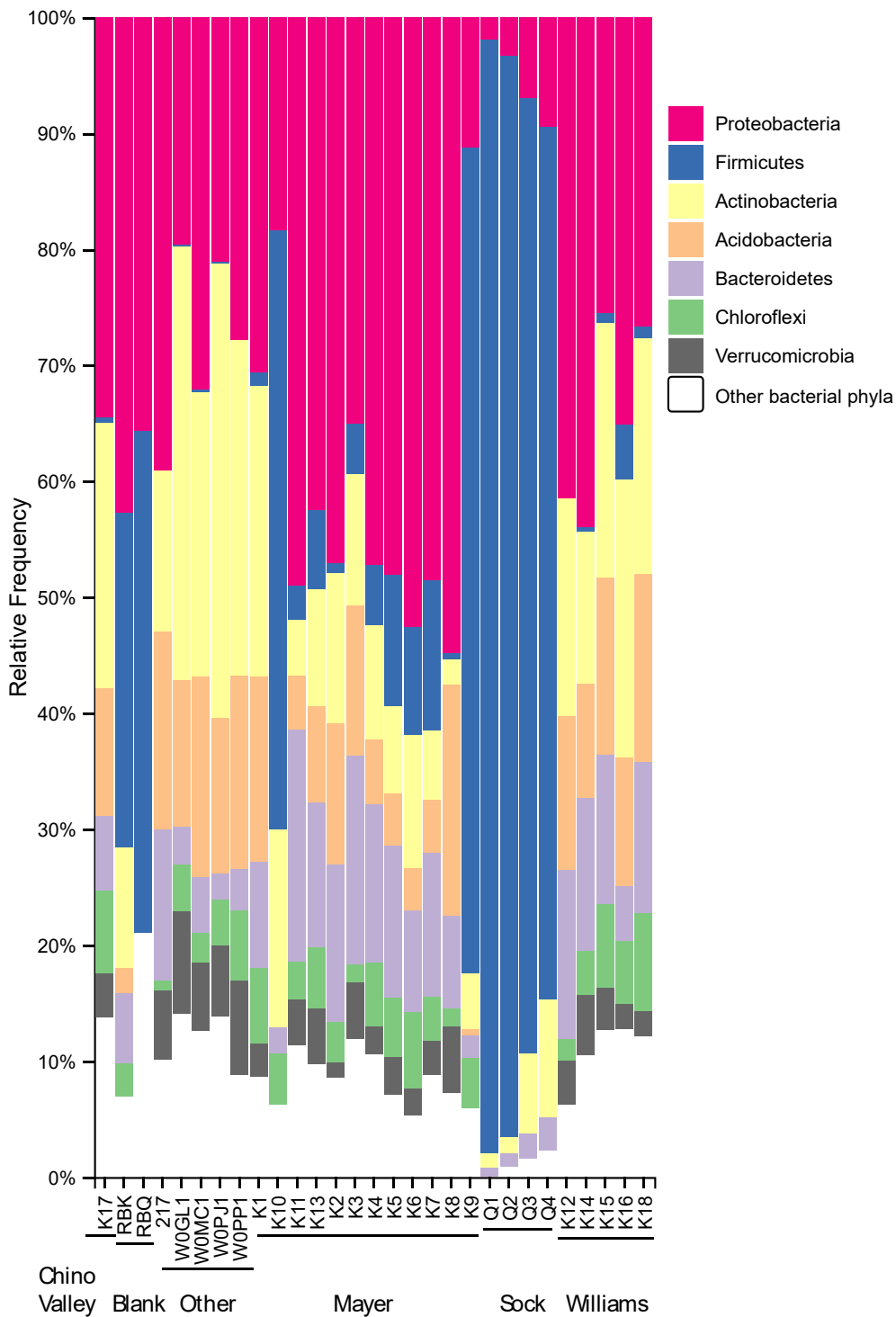


320

FIG. S2-

321 *Microbiome taxonomic composition at the phylum levels for all samples after excluding all*

322 *Staphylococcaceae.*



323

324 FIG. S3- Microbiome taxonomic composition at the phylum levels for all samples including reagent

325 blanks RBQ and RBK.

326

327