A peer-reviewed version of this preprint was published in PeerJ on 9 June 2020.

View the peer-reviewed version (peerj.com/articles/9235), which is the preferred citable publication unless you specifically need to cite this preprint.

Coil DA, Neches RY, Lang JM, Jospin G, Brown WE, Cavalier D, Hampton-Marcell J, Gilbert JA, Eisen JA. 2020. Bacterial communities associated with cell phones and shoes. PeerJ 8:e9235
https://doi.org/10.7717/peerj.9235

Bacterial communities associated with cell phones and shoes

David A Coil Corresp., [1] , **Russell Y Neches** [1] , **Jenna M Lang** [1] , **Guillaume Jospin** [1] , **Wendy E Brown** [2,3] , **Darlene Cavalier** [3,4] , **Jarrad Hampton-Marcell** [5] , **Jack A Gilbert** [6] , **Jonathan A Eisen** Corresp. [7]

[1] Genome Center, UC Davis, Davis, California, United States

[2] Department of Biomedical Engineering, University of California, Irvine, Irvine, California, United States

[3] Science Cheerleaders, Inc, United States

[4] SciStarter.org, United States

[5] Argonne National Laboratory, University of Chicago, Lemont, Illinois, United States

[6] Department of Surgery, University of Chicago, Chicago, Illinois, United States

[7] Genome Center, Department of Evolution and Ecology, Department of Medical Microbiology and Immunology, University of California, Davis, Davis, California, United States

Corresponding Authors: David A Coil, Jonathan A Eisen
Email address: dcoil@ucdavis.edu, jaeisen@ucdavis.edu

Background: Every human being carries with them a collection of microbes, a collection that is likely both unique to that person, but also dynamic as a result of significant flux with the surrounding environment. The interaction of the human microbiome (i.e., the microbes that are found directly in contact with a person in places such as the gut, mouth, and skin) and the microbiome of accessory objects (e.g., shoes, clothing, phones, jewelry) is of potential interest to both epidemiology and the developing field of microbial forensics. Therefore, the microbiome of personal accessories are of interest because they serve as both a microbial source and sink for an individual, they may provide information about the microbial exposure experienced by an individual, and they can be sampled non-invasively.

Findings: We report here a large-scale study of the microbiota found on cell phones and shoes. Cell phones serve as a potential source and sink for skin and oral microbiota, while shoes can act as sampling devices for the microbial environmental experience. Using 16S rRNA gene sequencing, we characterized the microbiota of thousands of paired sets of cell phones and shoes from individuals at sporting events, museums, and other venues around the United States.

Conclusions: We place this data in the context of previous studies and demonstrate that the microbiota of phones and shoes are different. This difference is driven largely by the presence of “environmental” taxa (taxa from groups that tend to be found in places like soil) on shoes and human-associated taxa (taxa from groups that are abundant in the human microbiome) on phones. This large dataset also contains many novel taxa, highlighting the fact that much of microbial diversity remains uncharacterized, even on commonplace objects.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27514v1 | CC BY 4.0 Open Access | rec: 1 Feb 2019, publ: 1 Feb 2019

PeerJ Preprints NOT PEER-REVIEWED

1 **Bacterial communities associated with cell phones and shoes**

2

3 David A. Coil, Russell Y. Neches, Jenna M. Lang, Guillaume Jospin, Wendy E. Brown, Darlene
4 Cavalier, Jarrad Hampton-Marcell, Jack A. Gilbert, Jonathan A. Eisen

5

6 David A. Coil (coil.david@gmail.com). Genome Center, University of California, Davis, CA,
7 USA

8 Russell Y. Neches (russell@vort.org). Genome Center, University of California, Davis, CA,
9 USA

10 Jenna M. Lang (jennomics@gmail.com) Genome Center, University of California, Davis, CA,
11 USA

12 Guillaume Jospin (guillaume.jospin@gmail.com). Genome Center, University of California,
13 Davis, CA, USA

14 Wendy E. Brown (wendy.brown@uci.edu). Science Cheerleaders, Inc., Department of
15 Biomedical Engineering, University of California Irvine, Irvine, CA, USA

16 Darlene Cavalier (darcav1@gmail.com). Science Cheerleaders, Inc., SciStarter.org

17 Jarrad Marcell (jhampton-marcell@anl.gov). Argonne National Laboratory, University of
18 Chicago, Lemont, IL, USA

19 Jack A. Gilbert (gilbert.jack@gmail.com). Department of Surgery, University of Chicago,
20 Chicago, IL, USA

21 Jonathan A. Eisen (jaeisen@ucdavis.edu). UC Davis Genome Center, Department of Evolution
22 and Ecology, Department of Medical Microbiology and Immunology, University of California,
23 Davis, Davis, CA, USA

24

25

26

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27514v1 | CC BY 4.0 Open Access | rec: 1 Feb 2019, publ: 1 Feb 2019

27 Corresponding Authors:
28
29 David A Coil
30 451 Health Sciences Drive, Davis, CA, 95618, USA
31 Email address: coil.david@gmail.com
32
33 Jonathan A Eisen
34 451 Health Sciences Drive, Davis, CA, 95618, USA
35 Email address: jaeisen@ucdavis.edu
36
37 *Running title:* Cell phone and shoe bacteria
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57

PeerJ Preprints NOT PEER-REVIEWED

Abstract

Background: Every human being carries with them a collection of microbes, a collection that is likely both unique to that person, but also dynamic as a result of significant flux with the surrounding environment. The interaction of the human microbiome (i.e., the microbes that are found directly in contact with a person in places such as the gut, mouth, and skin) and the microbiome of accessory objects (e.g., shoes, clothing, phones, jewelry) is of potential interest to both epidemiology and the developing field of microbial forensics. Therefore, the microbiome of personal accessories are of interest because they serve as both a microbial source and sink for an individual, they may provide information about the microbial exposure experienced by an individual, and they can be sampled non-invasively.

Findings: We report here a large-scale study of the microbiota found on cell phones and shoes. Cell phones serve as a potential source and sink for skin and oral microbiota, while shoes can act as sampling devices for microbial environmental experience. Using 16S rRNA gene sequencing, we characterized the microbiota of thousands of paired sets of cell phones and shoes from individuals at sporting events, museums, and other venues around the United States.

***Conclusions**:* We place this data in the context of previous studies and demonstrate that the microbiota of phones and shoes are different. This difference is driven largely by the presence of "environmental" taxa (taxa from groups that tend to be found in places like soil) on shoes and human-associated taxa (taxa from groups that are abundant in the human microbiome) on phones. This large dataset also contains many novel taxa, highlighting the fact that much of microbial diversity remains uncharacterized, even on commonplace objects.

Keywords

16S rRNA gene, cell phones, shoes, citizen science, biogeography, human microbiome, Illumina, taxonomy, microbial dark matter, ASV

Introduction

Recent years have dramatically expanded our understanding of the human microbiome (e.g. (McDonald et al., 2018)), the microbiome of the built environment around us (e.g. (National Academies of Sciences, Engineering, and Medicine et al., 2017)), and the interactions between the two (e.g. (Leung & Lee, 2016)). This understanding has implications for fields ranging from medicine to forensics to architecture. In addition to the millions of microbes that we carry around each day, the majority of people on the planet now possess a cell phone. Previous work on the microbiome associated with phones has shown that people share a much greater percentage of their microbes with their own phone than with the phones of others (Meadow, Altrichter & Green, 2014). As for the environment around us, shoes (or other foot coverings) act in some ways as microbial sampling devices. We have previously described data suggesting this to be the case, as well as demonstrated that the microbiome of cell phones and shoes from the same person are quite distinct (Lax et al., 2015).

Throughout 2013-2014, we organized public events around the United States for the purpose of swabbing surfaces of the built environment and collecting bacteria for isolation via culturing. Cultured isolates from these samples were screened and a subset of them were sent to the International Space Station (ISS) for growth in microgravity (Coil et al., 2016). As part of the public outreach component of this project, we engaged the public in helping collect these swabs, as well as in swabbing their cell phones and shoes for a nationwide microbial biogeography study. Thousands of people participated in this project, and we initially collected ~3500 paired cell phone/shoe samples. The intent of examining bacteria on cell phones and shoes was twofold; firstly to scale up the results of previous studies on shoes and phones and to look for patterns in the biogeography at a national scale. The second was to engage people in thinking about cell phones as being a proxy for sampling the microbes found on a person and their shoes as being a proxy for sampling the microbes found in a person's environment. However, given the logistical constraints, disparate sampling sites/personnel, and Institutional Review Board (IRB) waiver limitations, we were very constrained in what metadata we could collect. In the end, the only information retained for each sample was the physical location (GPS coordinates), rough age of participants, sample object type (cell phone or shoe), and event (basketball game, museum visit,

120 etc.). Swabs from these samples were sent back to the laboratory, DNA was extracted from them,
121 and the DNA was used for 16s rRNA gene PCR amplification and sequencing. To our
122 knowledge, this represents the largest collection of bacterial community sequencing data
123 associated with cell phones or shoes.
124
125 **Materials and Methods**
126
127 *Sample collection*
128
129 Cell phone and shoe samples were collected on sterile cotton swabs (Puritan cotton tipped #25-
130 806) and participants were instructed to "swab for about 15 seconds as if trying to clean the
131 object". Swabs were kept at room temperature by necessity and then sent overnight to the
132 University of Chicago, where they were kept at -80 °C until processing. DNA extractions, library
133 preparation, and Illumina sequencing (paired-end 150 bp) were performed exactly as described
134 in our previous work using swabs from the ISS (Lang et al., 2017). In brief: samples were
135 prepared using Mo BIO UltraClean kits, DNA extracted using Zymo ZR-96 kits, DNA amplified
136 using EMP barcoded primer sets targeting the V4 region of the 16S rRNA gene, amplicons were
137 cleaned and pooled and sequenced on an Illumina MiSeq platform.
138
139 *Data processing/validation*
140
141 Data from our study reported here was combined with comparable data from a few other
142 microbiome studies: a study of swabs of the International Space Station, (Lang et al., 2017), a
143 study examining the microbiomes of both cell phones and their owners (Meadow, Altrichter &
144 Green, 2014), and a study we conducted of the microbiome of cell phones and shoes (Lax et al.,
145 2015).
146
147 All datasets were prepared by following the DADA2 protocols (regular or big data, depending on
148 the size of the dataset) (Callahan et al., 2016a). All four data sets were pre-processed separately,
149 and each lane of our large dataset was also pre-processed individually to account for error
150 patterns from different runs or machines. Reads longer than 150 base pairs (bp) were trimmed

151 down to 150 bp before processing with DADA2. Low quality regions of reads were removed by
152 trimming bases that did not satisfy a Q2 quality score. The reads were also trimmed down to a
153 length of 145 bp. Reads containing Ns were discarded and we used two expected errors to filter
154 the overall quality of the read (rather than averaging quality scores) (Edgar & Flyvbjerg, 2015).
155 Only forward reads were considered for this study in order to have uniform data sets (since some
156 of the data sets only had forward reads). Error models were calculated using one million reads
157 for the three published data sets. Our samples were additionally separated into sequencing lanes.
158 Each lane was dereplicated individually according to the DADA2 "BigData" protocol to generate
159 amplicon sequence variants (ASVs). The ISS samples were pre-processed using the standard
160 workflow using all the reads available and dereplicating all the samples at the same time. All
161 seven sequence tables were merged to generate a single biom-like table for statistical analyses.
162 ASVs were assigned taxonomy using the dada2 function "assignTaxonomy" and the Silva (NR
163 v132) database (Quast et al., 2013; Yilmaz et al., 2014; Glöckner et al., 2017). ASVs that were
164 taxonomically assigned to mitochondria or chloroplast were removed. We excluded the ASVs
165 not represented in 5% of our samples or those with "unidentified" Phyla assignments. Very
166 closely related ASVs were merged using both a phylogenetic tree based approach and the
167 taxonomic labels comparisons (tip_glom and tax_glom functions from phyloseq). Samples were
168 excluded if they did not contain at least one ASV after the filtering. Finally, the resulting ASV
169 table was selected for only those ASVs assigned to the bacterial kingdom using the subset_taxa
170 function.
171

172 Alignment of the observed sequences was performed using Clustal Omega (Goujon et al., 2010;
173 Sievers et al., 2011), and an approximate maximum likelihood phylogeny was constructed using
174 FastTree2 (Price, Dehal & Arkin, 2009, 2010). Metadata was loaded from the mapping files for
175 each of the four studies as tab-delimited tables, and relevant columns were extracted using
176 Pandas (McKinney & Others, 2010) (retained values were: Age, City, Date, Event, Gender,
177 Hand, Module, Run, Sample, Sport, State, Study, Surface, Time, Touches, Type, Wash). OTU
178 filtering, taxonomic agglomeration, and ordination was performed using phyloseq (McMurdie &
179 Holmes, 2013) using Callahan *et al.* as a guide (Callahan et al., 2016b). Variable importance
180 measures were estimated by training a random forest classifier (Breiman, 2001; Geurts, Ernst &
181 Wehenkel, 2006; Pedregosa et al., 2011) on the ASV counts and extracting the attribute

182 importance values from the trained classifiers (Janitza, Strobl & Boulesteix, 2013). The PCoA
183 ordination of the ASV data was generated using the ordinate and plot_ordination functions from
184 Phyloseq. We exported the ordination coordinates and averaged values for cell phones and shoes
185 separately to find the centroid of the two data spreads. We plotted a line bisecting
186 perpendicularly the segment between the two centroids to highlight the separation between the
187 two groups. We used ggplot2 to overlay this line on the sample and taxa (at the phylum level)
188 versions of the PCoA (Wickham, 2010). We ran an ANalysis Of SIMilarity (ANOSIM) test
189 available through the vegan R package to assess the similarities between the phone and shoe
190 samples using Bray-Curtis distances and 999 permutations (Oksanen et al., 2011).
191
192 **Results/Discussion**
193
194 In total, ~3500 swabs were collected for this study at 38 events (see Table 1 for details). Of
195 these, some samples were lost in transit and a further 864 samples were excluded from
196 sequencing due to an irretrievable loss of the sample ID data. Sequencing was done on 2,486
197 samples with 599,386,254 paired end reads generated across four lanes of Illumina HiSeq
198 PE150.
199
200 Following the DADA2 protocol, we combined the data from our 2,486 samples with data from
201 three other microbiome studies (439 samples and 57,864,099 reads) and then carried out quality
202 filtering on the combined data set which resulted in 2,673 samples moving forward for further
203 analysis. For subsequent analysis on this combined data set, we only used the forward reads
204 because some of the comparison studies only reported forward reads. These reads were then used
205 to identify amplicon sequence variants (ASVs). 227,629 unique ASVs were identified and
206 taxonomic assignments were made for these ASVs using the Silva NR v132 database. Using
207 Phyloseq, those ASVs that were assigned to mitochondria or chloroplasts (in total 72,400 or 32%
208 of the ASVs) were excluded from further analysis, resulting in 155,229 remaining ASVs. ASVs
209 present in too few samples (less than 5%) were removed, keeping 1,928 ASVs. We grouped
210 closely-related taxa separated by a cophenetic distance smaller than 0.4, further reducing to 291
211 ASVs. ASVs that were taxonomically assigned to anything that was not bacteria were also
212 excluded (289 ASVs remaining).

213
214 The ASV based filtration reduced the total number of samples to 2,630 (since some samples did
215 not contain any of these final ASVs). In total, these 289 unique ASVs included 64,067,941 of the
216 initial reads. For some analyses, we further reduced this final data set by including only samples
217 from this study. This resulted in 40,432,677 reads representing 223 unique ASVs from 2,185
218 samples.
219
220 In order to examine and visualize differences between samples, we plotted a PCoA ordination of
221 samples based on sample to sample Bray-Curtis distances of the microbial communities in those
222 samples (FIGURE 1). A quick examination of the plot revealed that cell phones (green) and
223 shoes (black) appear to group separately (something seen in prior studies); this is supported by
224 ANOSIM statistical analysis which showed a significance of 0.001 for this separation of shoes
225 and phones. Visual examination suggests that floor samples (light blue) group with shoes (as
226 expected), while spacecraft (yellow) group with phones, presumably because both of these
227 communities have major contributions from human associated taxa. However, we did not test the
228 significance of these groupings.
229
230 As an alternative method for examining the potential importance of the metadata variables
231 (sample type, sport, location, and sequencing run) we utilized variable importance measures
232 (VIMs). These VIMs were estimated by training a random forest classifier (Breiman, 2001;
233 Geurts, Ernst & Wehenkel, 2006; Pedregosa et al., 2011) to assign samples to their metadata
234 categories (sample type, city, state, sequencing run and sport) based on their ASV counts, and
235 extracting the variable importance values from the trained classifiers (Janitza, Strobl &
236 Boulesteix, 2013). Note that variable importance analysis is a distinct application of random
237 forests from the more widely-used classification application. Extracting VIMs not does not
238 include the optimization and benchmarking steps required to use random forests in their
239 predictive capacity. Sample feature importances indicate that the sample type (shoe or phone)
240 was the most predictive of the observed community structure, followed by the geographic
241 location of the sample (Supplemental Figure 1). The sport played at the venue where the sample
242 was collected is less predictive of the community structure than the sequencing run. Overall,
243 these results support and extend our previous findings that the microbiomes of shoes and phones

244 are distinct. Interestingly, the city where an event took place was more predictive of community
245 structure than state, suggesting the possibility that there are local biogeography effects in
246 patterning the microbial community. Further analysis of this large dataset may reveal more
247 detailed patterns, such as the influence of geographic location on microbial communities
248
249
250 To further examine the differences between cell phones and shoes we identified the centroids of
251 the two data spreads, after first removing all the data from previous studies (FIGURE 2). The
252 line in this figure represents the bisection of these two centroids, to highlight their separation.
253 We then used this bisection line to examine in more detail the taxa that contribute to the
254 separation of shoe and phone samples.
255
256 We did this by generating a series of plots showing only the ASVs belonging to each phylum
257 separately (FIGURE 3). The line in each plot is the same as in the sample plot in Figure 2 and
258 those ASVs to the top/left can be considered to be driving the "shoe" portion of the PCoA and
259 the ASVs to the bottom/right can be considered to drive the "phone" portion of the PCoA. These
260 plots (and the underlying data) show some interesting phyla-specific patterns. Some phyla (e.g.,
261 Bacteroides and Firmicutes) have many ASVs on both sides of the line, indicating that there are
262 ASVs from these phyla that are biased towards shoes and others that are biased towards phones.
263
264 Two phyla (Tenericutes and Fusobacteria) contain only ASVs that are skewed towards phones.
265 We believe this is likely due to these ASVs being human associated taxa. For example, the
266 taxonomic assignments of the Fusobacteria ASVs were *Leptotrichia* (n=2) and *Fusobacterium*
267 (n=1); these two genera are generally found in animal microbiomes including the oral
268 microbiome of humans and other mammals. The two Tenericutes ASVs were both taxonomically
269 assigned to the *Mycoplasma* genus; many members of this genus are animal associated.
270
271 In contrast, there are many phyla (Acidobacteria, Cyanobacteria, Deinococcus-Thermus,
272 Planctomycetes, Fibrobacteres, Nitrospirae, Chloroflexi, Armatimonadetes, and
273 Gemmatimonadetes) which include only ASVs that are skewed towards shoes. We presume that
274 these ASVs from these phyla represent taxa from the broader environment (e.g., soil) that would

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27514v1 | CC BY 4.0 Open Access | rec: 1 Feb 2019, publ: 1 Feb 2019

275 be picked up by shoes. Examination of the taxonomic assignments for these ASVs supports this
276 possibility, with genera assignments including taxa commonly found in water or soil such as
277 *Chroococcidiopsis, Oscillatoria, Chroococcidiopsis, Truepera, Deinococcus, Longimicrobium,*
278 *Gemmatirosa, Gemmatimonas, Nitrospira,* and *Planctomyces*.
279
280 **Novel evolutionary lineages**
281
282 One of the reasons we chose to sample cell phones and shoes is that they are such
283 commonplace objects used by so many people all around the world. The fact that they are so
284 commonplace makes them useful in the context of crowdsourcing and participatory
285 microbiology projects: many people have both of them, one can use them as a way to get people
286 to think about microbes hidden in the world around them, and they have potential for various
287 forensic types of analyses.
288 In relation to this, we examined how many (if any) of these microbes present in such
289 everyday objects were from any of the so-called "microbial dark matter" branches in the tree of
290 life. The term "microbial dark matter" or MDM for short is used in this context to refer to major
291 evolutionary lineages for which few or no representatives have ever been grown in the lab or
292 studied in detail (Rinke et al., 2013). To examine the MDM in these samples, we examined the
293 taxonomic annotation of ASVs and identified those that were assigned to phyla or candidate
294 phyla that are generally viewed as MDM lineages. The phyla we focused on were:
295 Aegiribacteria, AncK6, Armatimonadetes, Atribacteria, BRC1, Caldiserica, Calditrichaeota,
296 Chrysiogenetes, Cloacimonetes, Coprothermobacteraeota, Dadabacteria, Dependentiae,
297 Diapherotrites, Edwardsbacteria, Elusimicrobia, Entotheonellaeota, Fervidibacteria, FCPU426,
298 GAL15, Hydrogenedentes, Latescibacteria, Margulisbacteria, Nanoarchaeaeota, Nitrospinae,
299 Omnitrophicaeota, Patescibacteria, PAUC34f, Rokubacteria, RsaHf231, WOR-1, WPS-2, WS1,
300 WS2, WS4, and Zixibacteria. We also then examined the distribution patterns of these ASVs
301 across samples and the whether they showed any skew between phones and shoes (Supplemental
302 Table 1).
303 This analysis of ASVs assigned to MDM lineages revealed that in fact quite a large
304 number of ASVs found in our study were from such MDM groups. In some cases, these ASVs
305 assigned to these groups are quite rare - for example ASVs from WOR-1, Edwardsbacteria and

306 Diapherotrites was found to be present in one sample each. However, some were present in a
307 much wider range of samples, and we focused most of our attention on those. Of the nine MDM
308 phyla for which ASVs were found to be present in at least 10% of samples (Armatimonadetes,
309 Patescibacteriam, WPS-2, Entotheonellaeota, Dependentiae, BRC1, Rokubacteria,
310 Latescibacteria, Elusimicrobia), all were found more often in shoe samples than phone samples.
311 This is not surprising given that (1) phone samples tend to be enriched for human associated
312 microbes, only a few of which are in current MDM groups and (2) many MDM lineages are
313 known to be found in soil, which is presumably abundant on shoes. Two of these widespread
314 MDM phyla (Armatimonadetes, Patescibacteriam) were found to have ASVs present in almost
315 50% of samples. Twelve classes and thirteen orders were found to be present in more than 10%
316 of samples. Of these, all were skewed towards shoe samples except two taxa (Gracilibacteria
317 within Patescibacteria, and Absconditabacteriales within Gracilibacteria).
318 Overall these results show that though MDM is frequently portrayed as mostly coming
319 from remote, isolated, or extreme environments, a remarkable fraction of people are traveling
320 around with representatives from these groups on commonplace objects.
321
322 **Conclusion.**
323
324 These data support previous work by ourselves and others demonstrating that the microbiome of
325 cell phones and shoes are distinct, even when belonging to the same person. In this analysis, we
326 also highlight which phyla are most responsible for the observed differences in microbial
327 communities between phones and shoes. This difference is driven largely by the presence of
328 "environmental" taxa (taxa from groups that tend to be found in places like soil) on shoes and
329 human-associated taxa (taxa from groups that are abundant in the human microbiome) on
330 phones. Lastly, we show that a number of "microbial dark matter" taxa are present, even
331 abundant, on these commonplace objects.
332
333
334
335
336

337 **Availability of Supporting Data**

338

339 All raw sequencing data has been deposited at NCBI under BioProject PRJNA470730
340 (https://www.ncbi.nlm.nih.gov/sra/SRP145522). All data analysis, supporting files and
341 intermediate analysis files are available at Zenodo:
342 (https://zenodo.org/record/1419350#.W6Uy5PlRdEY). An interactive visualization of this data is
343 available at www.phinch.org.

344

345

346 **Acknowledgements**

347

348 The authors would like to thank the many Science Cheerleaders who ran the sampling events. In
349 particular, we'd like to thank Bart Leahy who coordinated the coordinators. We would also like
350 to thank SciStarter and Science Cheerleaders, Inc. for being partners in this work. And finally,
351 we would like to thank the thousands of people who gave willingly their time and their samples
352 as part of participating in this study.

353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372

373 **References**

374

375

376 Breiman L. 2001. Random Forests. *Machine learning* 45:5–32.

377 Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016a. DADA2:
378 High-resolution sample inference from Illumina amplicon data. *Nature methods* 13:581–
379 583.

380 Callahan BJ, Sankaran K, Fukuyama JA, McMurdie PJ, Holmes SP. 2016b. Bioconductor
381 Workflow for Microbiome Data Analysis: from raw reads to community analyses.
382 *F1000Research* 5:1492.

383 Coil DA, Neches RY, Lang JM, Brown WE, Severance M, Cavalier D, Eisen JA. 2016. Growth
384 of 48 built environment bacterial isolates on board the International Space Station (ISS).
385 *PeerJ* 4:e1842.

386 Edgar RC, Flyvbjerg H. 2015. Error filtering, pair assembly and error correction for next-
387 generation sequencing reads. *Bioinformatics* 31:3476–3482.

388 Geurts P, Ernst D, Wehenkel L. 2006. Extremely randomized trees. *Machine learning* 63:3–42.

389 Glöckner FO, Yilmaz P, Quast C, Gerken J, Beccati A, Ciuprina A, Bruns G, Yarza P, Peplies J,
390 Westram R, Ludwig W. 2017. 25 years of serving the community with ribosomal RNA
391 gene reference databases and tools. *Journal of biotechnology* 261:169–176.

392 Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R. 2010. A new
393 bioinformatics analysis tools framework at EMBL–EBI. *Nucleic acids research* 38:W695–
394 W699.

395 Janitza S, Strobl C, Boulesteix A-L. 2013. An AUC-based permutation variable importance
396 measure for random forests. *BMC bioinformatics* 14:119.

397 Lang JM, Coil DA, Neches RY, Brown WE, Cavalier D, Severance M, Hampton-Marcell JT,
398 Gilbert JA, Eisen JA. 2017. A microbial survey of the International Space Station (ISS).
399 *PeerJ* 5:e4029.

400 Lax S, Hampton-Marcell JT, Gibbons SM, Colares GB, Smith D, Eisen JA, Gilbert JA. 2015.
401 Forensic analysis of the microbiome of phones and shoes. *Microbiome* 3:21.

402 Leung MHY, Lee PKH. 2016. The roles of the outdoors and occupants in contributing to a
403 potential pan-microbiome of the built environment: a review. *Microbiome* 4:21.

404 McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, Aksenov AA,
405 Behsaz B, Brennan C, Chen Y, DeRight Goldasich L, Dorrestein PC, Dunn RR, Fahimipour
406 AK, Gaffney J, Gilbert JA, Gogul G, Green JL, Hugenholtz P, Humphrey G, Huttenhower
407 C, Jackson MA, Janssen S, Jeste DV, Jiang L, Kelley ST, Knights D, Kosciolek T, Ladau J,
408 Leach J, Marotz C, Meleshko D, Melnik AV, Metcalf JL, Mohimani H, Montassier E,
409 Navas-Molina J, Nguyen TT, Peddada S, Pevzner P, Pollard KS, Rahnavard G, Robbins-
410 Pianka A, Sangwan N, Shorenstein J, Smarr L, Song SJ, Spector T, Swafford AD, Thackray
411 VG, Thompson LR, Tripathi A, Vázquez-Baeza Y, Vrbanac A, Wischmeyer P, Wolfe E,
412 Zhu Q, American Gut Consortium, Knight R. 2018. American Gut: an Open Platform for
413 Citizen Science Microbiome Research. *mSystems* 3. DOI: 10.1128/mSystems.00031-18.

414 McKinney W, Others. 2010. Data structures for statistical computing in python. In: *Proceedings*
415 *of the 9th Python in Science Conference*. Austin, TX, 51–56.

416 McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and
417 graphics of microbiome census data. *PloS one* 8:e61217.

418 Meadow JF, Altrichter AE, Green JL. 2014. Mobile phones carry the personal microbiome of
419 their owners. *PeerJ* 2:e447.

420 National Academies of Sciences, Engineering, and Medicine, National Academy of Engineering,
421 Division on Engineering and Physical Sciences, Health and Medicine Division, Division on
422 Earth and Life Studies, Board on Infrastructure and the Constructed Environment, Board on
423 Environmental Studies and Toxicology, Board on Life Sciences, Committee on
424 Microbiomes of the Built Environment: From Research to Application. 2017. *Microbiomes*
425 *of the Built Environment: A Research Agenda for Indoor Microbiology, Human Health, and*
426 *Buildings*. National Academies Press.
427 Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara RB, Simpson GL, Solymos
428 P, Stevens MHH, Wagner H. 2011. vegan: Community ecology package. *R package*
429 *version*:117–118.
430 Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer
431 P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M,
432 Duchesnay É. 2011. Scikit-learn: Machine Learning in Python. *Journal of machine learning*
433 *research: JMLR* 12:2825–2830.
434 Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with
435 profiles instead of a distance matrix. *Molecular biology and evolution* 26:1641–1650.
436 Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for
437 large alignments. *PloS one* 5:e9490.
438 Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The
439 SILVA ribosomal RNA gene database project: improved data processing and web-based
440 tools. *Nucleic acids research* 41:D590–6.
441 Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S,
442 Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu W-T, Eisen

443 JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T. 2013.
444 Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499:431–
445 437.
446 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert
447 M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality
448 protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*
449 7:539.
450 Wickham H. 2010. *ggplot2: Elegant Graphics for Data Analysis*. Springer New York.
451 Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W,
452 Glöckner FO. 2014. The SILVA and "All-species Living Tree Project (LTP)" taxonomic
453 frameworks. *Nucleic acids research* 42:D643–8.

454
455

PeerJ Preprints NOT PEER-REVIEWED

Table 1 (on next page)

Sample Collection Information

Table 1: Sample Collection Information. “Age” is a rough approximation based on attendees of the event (A=Adult, K=Kid, M=Mixed). “n=” refers to the number of samples that were actually sequenced. “Event title or location” is how the samples are referenced in the data files.

PeerJ Preprints NOT PEER-REVIEWED

Age	City	State	n=	Date	Event title or Location	Description
A	Palmdale	CA	19	7/19/2013	TeachersInSpace	Teachers in Space summer workshop at Aero Institute
K	San Diego	CA	14	8/24/2013	PWCoronado	Pop Warner Cheer Organization: Coronado Islanders
K	Monrovia	CA	31	9/24/2013	Wildrose	Wildrose Elementary School
K	Castro Valley	CA	12	9/29/2013	PWGladiators	Pop Warner Cheer Organization: Castro Valley Gladiators
M	San Francisco	CA	147	11/2/2013	BASF	Bay Area Science Festival
A	Denver	CO	33	5/8/2013	DMNS	Denver Museum of Nature and Science
K	Fountain	CO	37	10/10/2013	ColeMiddle	Cole Middle School
A	Washington	DC	13	4/12/2013	YNDC	Yuri's Night party at Science Club in Washington D.C.
M	Washington	DC	50	9/14/2013	SmithsonianAirSpace	Women in Space Day/Smithsonian Museum of Air and Space
M	Washington	DC	280	4/25/2014	SciEngFest	USA Science and Engineering Festival
A	Fort Lauderdale	FL	16	8/14/2013	Broward	STEM Teacher Event
K	Orlando	FL	40	9/7/2013	PWBrantley	Pop Warner Cheer Organization: Lake Brantley Patriots
A	Miami	FL	28	9/25/2013	MiamiDolphins	Miami Dolphins NFL football game
K	Atlanta	GA	33	4/27/2013	Girl Scouts	Girl Scouts at Atlanta Science Festival
K	Potlatch	ID	25	10/10/2013	Potlatch	Potlatch Junior High School
A	Longmeadow	MA	10	9/26/2013	Tufts	Tufts University Pediatric Infectious Diseases Hospital
M	Baltimore	MA	24	5/4/2014	KidneyFoundation	Kidney Foundation Walk at the Baltimore Zoo
A	Columbia	MD	69	6/9/2013	HowardCCC	Howard County Community Challenge
A	Landover	MD	6	10/29/2013	Redskins	Washington D.C. NFL football game
A	Durham	NC	36	4/12/2013	YNNC	Yuri's Night party at Museum of Life and Science in Durham, NC
A	Durham	NC	246	2/17/2014	ScienceOnline	Science Online scientific conference - NC State University
A	New York	NY	40	4/16/2013	YNNY	Yuri's Night party at National Arts Club in New York, NY
K	Chittenango	NY	35	9/4/2013	PWBears	Pop Warner Cheer Organization: Chittenango Bears
A	Tulsa	OK	78	9/11/2013	TulsaCCBio	Tulsa Community College Bio Class

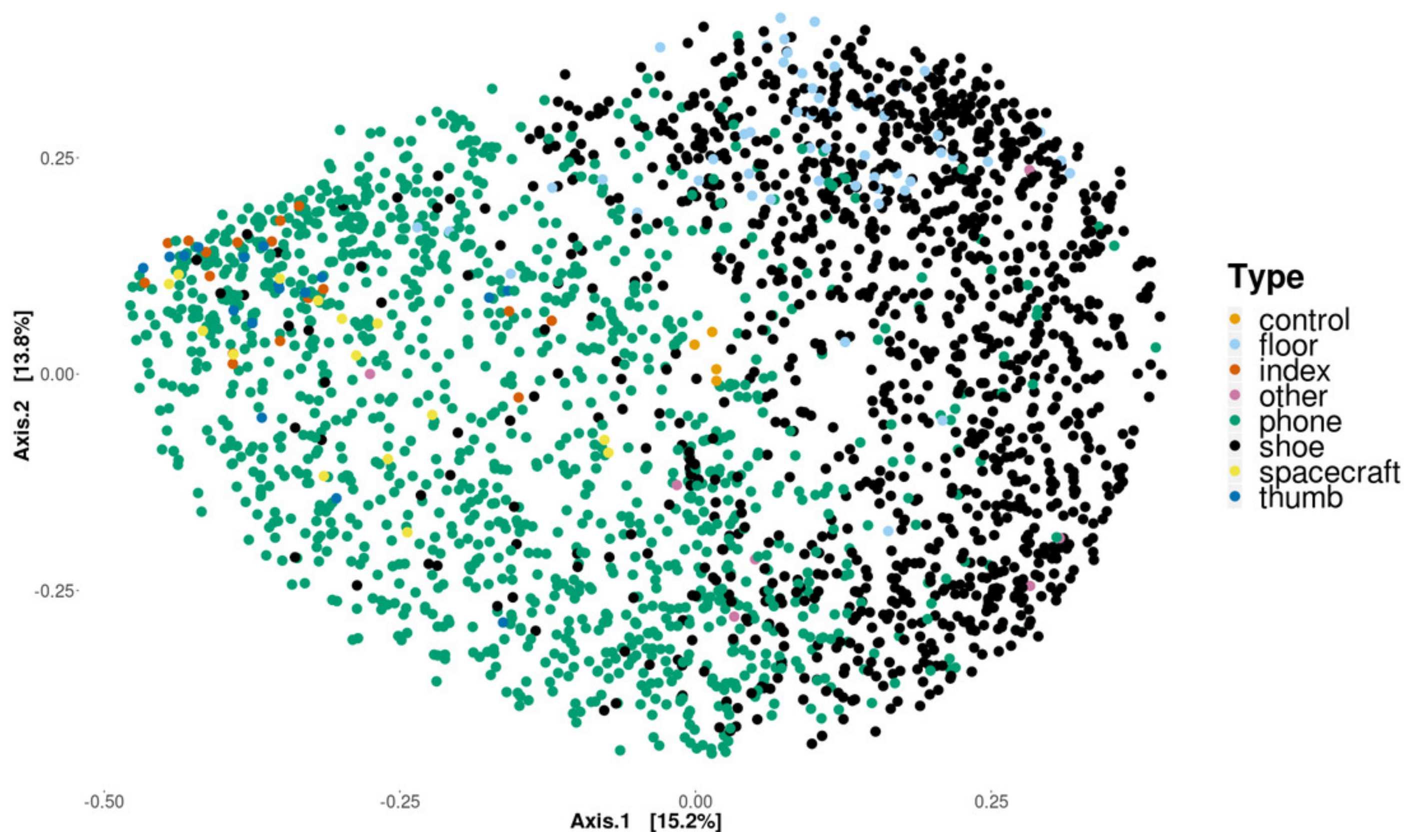
						Project
K	Salem	OR	20	10/4/2013	ChapmanHill	Chapman Hill Elementary School
M	Philadelphia	PA	5	4/20/2013	PhillyScienceFest	Philadelphia Science Festival 2013
M	Philadelphia	PA	72	4/25/2013	PhilliesGame	Philadelphia Phillies MLB baseball game
A	Philadelphia	PA	10	5/23/2013	CHF	Chemical Heritage Foundation
A	Philadelphia	PA	3	5/30/2013	FranklinInstitute	The Franklin Institute
A	Philadelphia	PA	17	6/4/2013	PhillyANS	The Academy of Natural Sciences at Drexel University
A	Philadelphia	PA	72	2/18/2014	76ers	Science at the Sixers - Philadelphia 76ers NBA basketball game
M	Philadelphia	PA	33	4/26/2014	DiscoveryDays	NaturePalooza - at The Schuylkill Center for Environmental Education
M	Philadelphia	PA	23	4/30/2014	DrexelLibrary	Philadelphia Science Festival: Katharine Drexel Library
M	Philadelphia	PA	171	5/3/2014	PhillySciFest	Philadelphia Science Festival 2014
M	San Antonio	TX	84	4/12/2013	SPURS	San Antiono Spurs NBA basketball game
M	Houston	TX	171	4/14/2014	YYCPA	Young Women's College Perparatory Academy
K	Unknown		13	4/23/2014	KidScoop	Nationwide competition through KidScoop magazine
M	Dulles	VA	70	9/28/2013	NSFSTEM	National Science Foundation, STEM Careers Fair; Dulles Town Center

1

Figure 1

Principal coordinate (PCoA) plot of all samples

FIGURE 1: Principal coordinate (PCoA) analysis plot of the Bray-Curtis distances of 16S rRNA gene sequence based ASVs for all samples, colored by the type of sample.

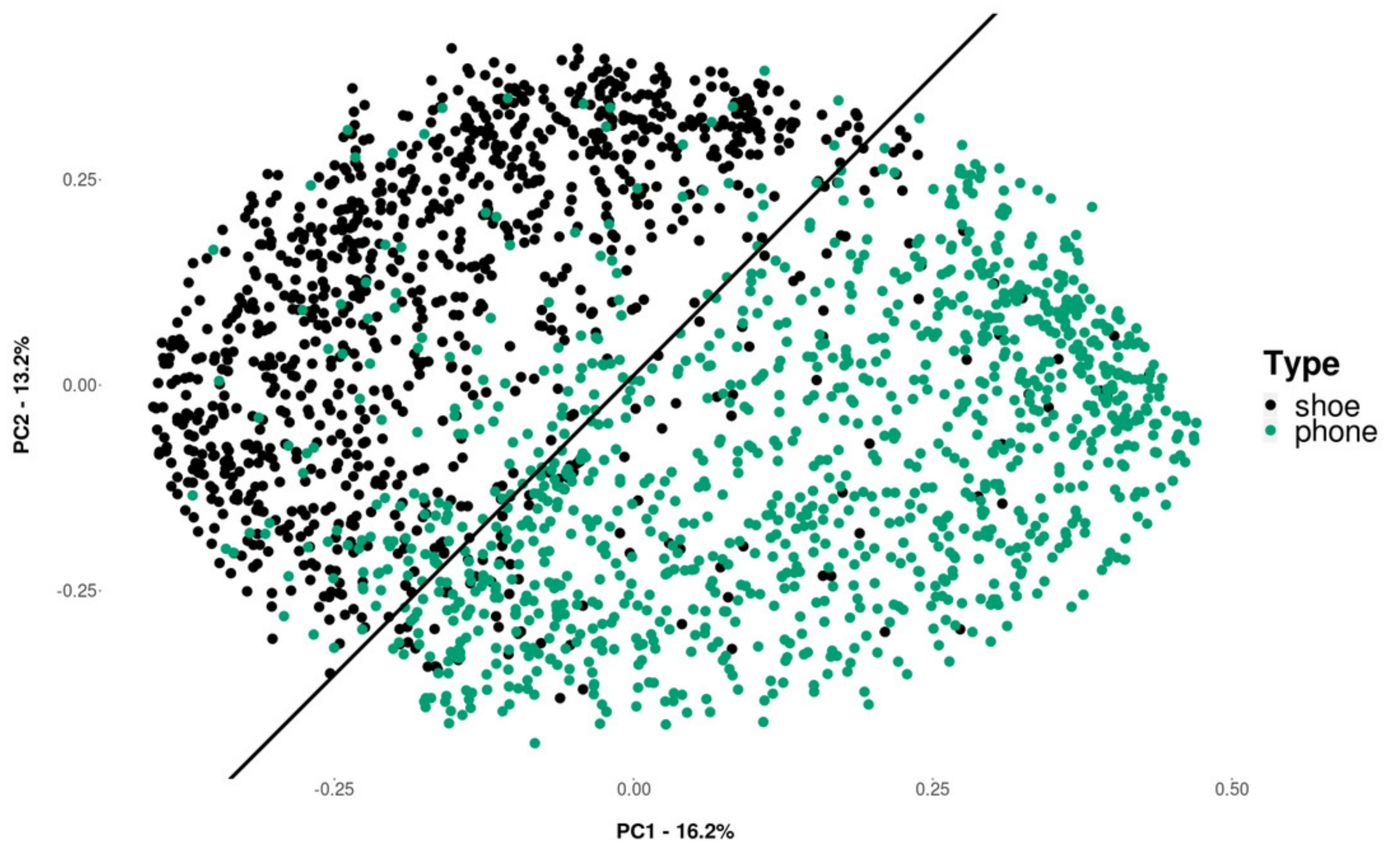


PeerJ Preprints NOT PEER-REVIEWED

Figure 2

Principal coordinate (PCoA) plot of samples in this study

FIGURE 2: Principal coordinate (PCoA) analysis plot of the Bray-Curtis distances of 16S rRNA gene sequence based ASVs for cell phone and shoe samples from only this study, colored by sample origin. The line is the bisection of the centroids of the two sample types (phones and shoes).



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27514v1 | CC BY 4.0 Open Access | rec: 1 Feb 2019, publ: 1 Feb 2019

PeerJ Preprints NOT PEER-REVIEWED

Figure 3

Principal coordinate (PCoA) plot of the ASVs for Phyla identified from this study.

FIGURE 3: Principal coordinate (PCoA) analysis plot of the Bray-Curtis distances of 16S rRNA gene sequence based ASVs for Phyla identified from this study (Taxa version of Figure 2). This is showing a split representation of individual Phyla to prevent overlapping points. The line represents the split between cell phone and shoe samples from Figure 2.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27514v1 | CC BY 4.0 Open Access | rec: 1 Feb 2019, publ: 1 Feb 2019

PeerJ Preprints NOT PEER-REVIEWED

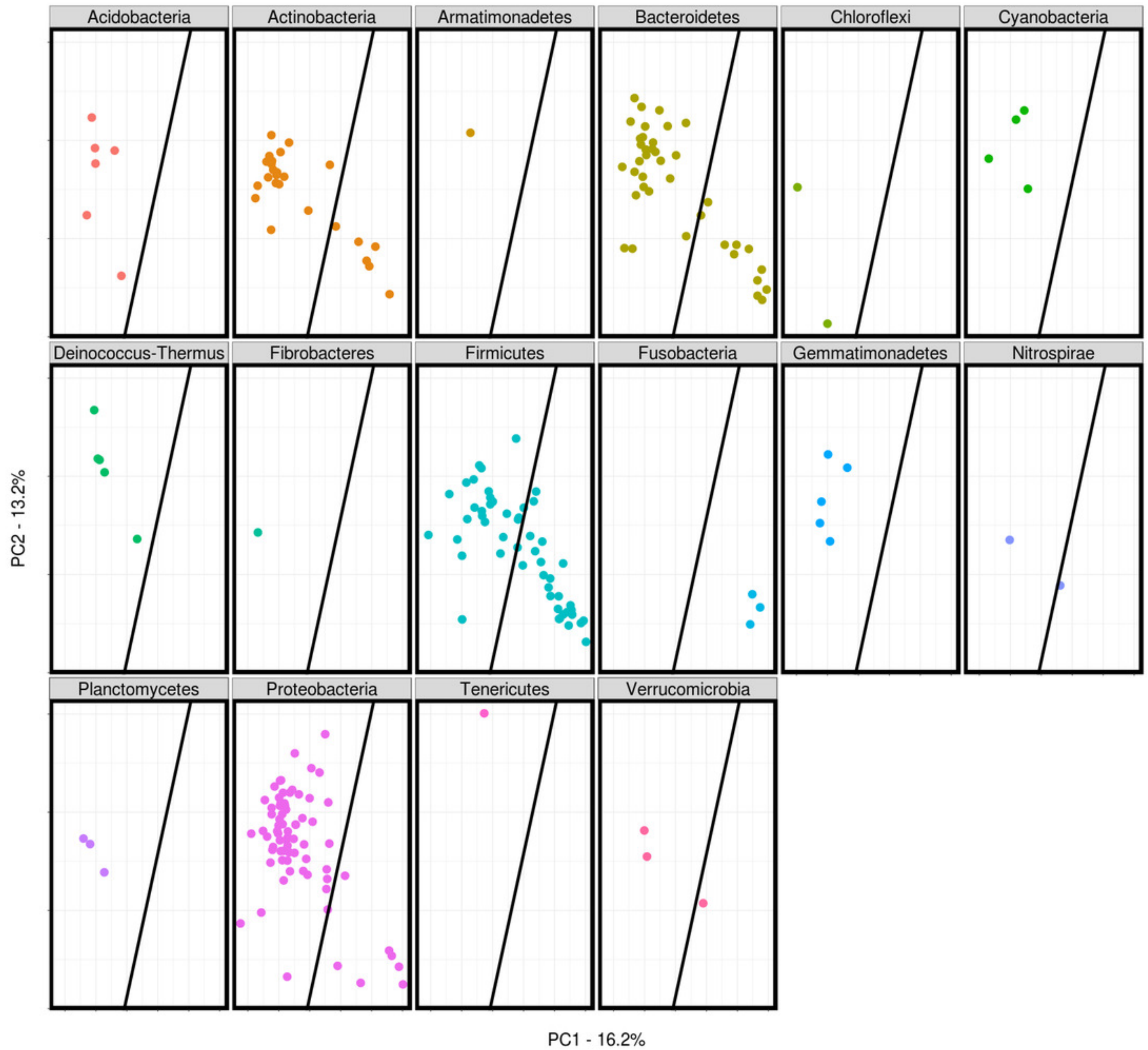
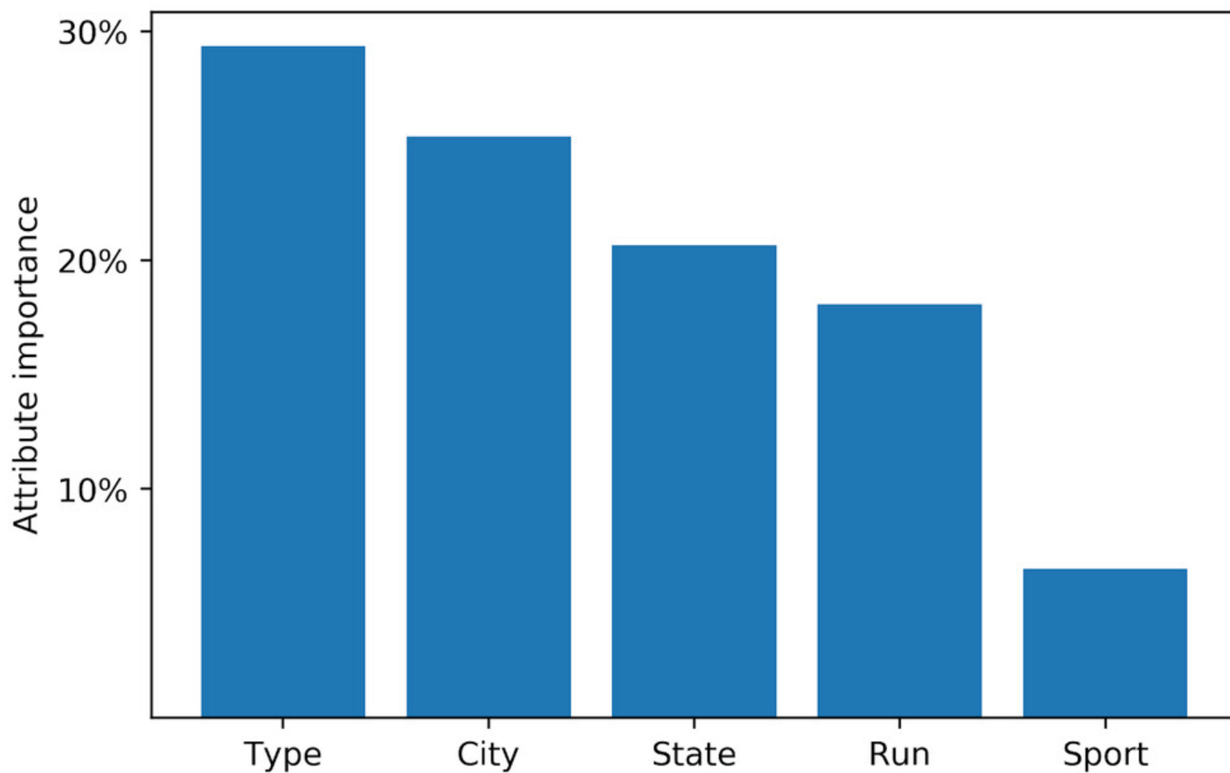


Figure 4

Supplemental Figure 1: Importance of metadata variables (attribute importance analysis)



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27514v1 | CC BY 4.0 Open Access | rec: 1 Feb 2019, publ: 1 Feb 2019

Table 2(on next page)

Supplemental Table 1. MDM (Microbial dark matter) phyla distribution summarized for shoes vs. phones

PeerJ Preprints NOT PEER-REVIEWED

Phylum	%cell	%shoe	#samples	% of Samples
Armatimonadetes	26.3	73.7	1068	47.8
Patescibacteria	36.7	63.3	1041	46.6
WPS-2	15.8	84.2	404	18.1
Entotheonellaeota	25.3	74.7	360	16.1
Dependentiae	19.7	80.3	356	15.9
BRC1	11.1	88.9	352	15.8
Rokubacteria	29.0	71.0	352	15.8
Latescibacteria	29.1	70.9	278	12.4
Elusimicrobia	25.9	74.1	259	11.6
RsaHf231	10.7	89.3	103	4.6
Nanoarchaeaeota	32.4	67.6	71	3.2
Omnitrophicaeota	36.0	64.0	50	2.2
Hydrogenedentes	23.3	76.7	43	1.9
WS4	26.5	73.5	34	1.5
Zixibacteria	39.1	60.9	23	1.0
FCPU426	21.7	78.3	23	1.0
WS2	31.3	68.8	16	0.7
Nitrospinae	30.8	69.2	13	0.6
GAL15	63.6	36.4	11	0.5
Dadabacteria	10.0	90.0	10	0.4
Atribacteria	50.0	50.0	6	0.3
Margulisbacteria	33.3	66.7	6	0.3
Coprothermobacteraeota	33.3	66.7	6	0.3

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27514v1 | CC BY 4.0 Open Access | rec: 1 Feb 2019, publ: 1 Feb 2019

Caldiserica	50.0	50.0	4	0.2
Calditrichaeota	0.0	100.0	4	0.2
Cloacimonetes	25.0	75.0	4	0.2
WS1	0.0	100.0	3	0.1
PAUC34f	50.0	50.0	2	0.1
AncK6	100.0	0.0	2	0.1
Acetothermia	0.0	100.0	2	0.1
Diapherotrites	0.0	100.0	1	0.0
Edwardsbacteria	100.0	0.0	1	0.0
WOR-1	0.0	100.0	1	0.0

1