1 **Title: Ecological succession and viability of human-associated microbiota on**
2 **restroom surfaces**
3
4 **Running Title: Longitudinal Analysis of the Restroom Microbiome**
5
6

7 Sean M. Gibbons[1,2*], Tara Schwartz[3*], Jennifer Fouquier[3,4], Michelle Mitchell[3], Naseer
8 Sangwan[2], Jack A. Gilbert[2,5,6,7,8], Scott T. Kelley[3]
9
10 [1]Graduate Program in Biophysical Sciences, University of Chicago, Chicago, IL 60637,
11 U.S.A.
12 [2]Institute for Genomic and Systems Biology, Argonne National Laboratory, 9700 South
13 Cass Avenue, Argonne, IL 60439, U.S.A.
14 [3]Department of Biology, San Diego State University, 5500 Campanile Drive, San Diego,
15 CA 92182, U.S.A.
16 [4]Graduate Program in Bioinformatics and Medical Informatics, San Diego State
17 University, 5500 Campanile Drive, San Diego, CA 92182, U.S.A.
18 [5]Department of Ecology and Evolution, University of Chicago, 1101 E 57th Street,
19 Chicago, IL 60637
20 [6]Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543, USA
21 [7]College of Environmental and Resource Sciences, Zhejiang University, Hangzhou,
22 310058, China
23 [8]Corresponding Author: gilbertjack@anl.gov; (630) 915-2383
24 [*]These authors contributed equally to the work presented in this manuscript
25
26

**Abstract**

Human-associated bacteria dominate the built environment (BE). Following decontamination of floors, toilet seats, and soap dispensers in 4 public restrooms, *in situ* bacterial communities were characterized hourly, daily, and weekly to determine their successional ecology. The viability of cultivable bacteria, following the removal of dispersal agents (humans), was also assessed hourly. A late successional community developed within 5-8 hours on restroom floors, and showed remarkable stability over weeks to months. Despite late successional dominance by skin- and outdoor-associated bacteria, the most ubiquitous organisms were predominantly gut-associated taxa, which persisted following exclusion of humans. *Staphylococcus* represented the majority of the cultivable community, even after several hours of human-exclusion. MRSA-associated virulence genes were found on floors, but were not present in assembled *Staphylococcus* pan-genomes. Viral abundances, which were predominantly enterophage, human papilloma and herpes viruses, were significantly correlated with bacteria abundances, and showed an unexpectedly low virus-to-bacteria ratio in surface-associated samples, suggesting that bacterial hosts are mostly dormant on BE surfaces.

**Importance**

We present a detailed longitudinal study of bacterial and viral communities residing on restroom surfaces. Human-associated microbes residing on these surfaces follow rapid and predictable ecological succession from fecal-dominated to skin-dominated communities, and remain viable for many hours. In addition, we show that common opportunistic pathogens, from the genus *Staphylococcus*, dominate the cultivable community for hours after the exclusion of humans.

## Introduction

The analysis of microbial diversity of indoor environments, collectively termed the Built Environment (BE), is important because of its potential impact on human health. It is estimated that humans in industrialized countries spend as much as 90% of their lives indoors (1, 2). Indeed, for billions of humans, the "great indoors" comprises the new human ecosystem. BEs contain an enormous variety of potential microhabitats for microorganisms, and are continually colonized by human and outdoor-associated microbiota (3-5). Understanding the ecological dynamics of the microbiota in BEs may help develop strategies to define and promote an indoor microbiome that minimizes disease risk (2).

While it has long been known that viable bacteria can be cultured from virtually any surface in an indoor ecosystem, we know relatively little about the true diversity and viability of the indoor microbiome. In the past, studies of microbial diversity relied mainly on culture-based techniques (3, 6). However, the application of culture-independent sequencing techniques to the study of BE microbiology has already greatly expanded our understanding of the origin and diversity BE microbes (2).

Comparisons of sequence data collected in one location to other existing datasets generated by the same approaches allows inference of the likely environmental origins of BE communities (e.g., human skin, soil, etc.) (7). Additionally, the impact of season and geographic location on bacterial community composition (5, 8) has revealed an extraordinary variability in BE-associated microbial diversity. However, with a few exceptions, most studies have involved single time-point samplings of surfaces. While this allows for a characterization of microbial diversity, and comparative analysis

74  between surfaces, replicated time-series studies need to be undertaken in order to

75  understand the formation, stability, and dynamics of BE communities (9). In addition,

76  most BE work has focused on bacterial communities, and there are few studies looking at

77  viral community diversity (2). These are significant gaps that need to be filled in order to

78  understand the distribution and behavior of the microbes that inhabit our BEs.

79        Restrooms are a shared public space with clear disease transmission potential (4).

80  However, the potential for disease transmission from a surface fomite relies on the

81  accumulation and continued viability of pathogenic taxa. A prior amplicon-sequencing

82  study investigating the biogeography of restroom surfaces established putative

83  colonization sources, gender-specific microbial signatures, and surface-specific

84  community structure for restroom microbial communities (4). This spatial study revealed

85  the dominance of human-associated microbes on restroom surfaces, but did not approach

86  the questions of community assembly dynamics, temporal stability, or viability. In

87  addition, this study focused on bacterial diversity, and did not investigate patterns in viral

88  abundance and diversity. Using a combination of 16S rRNA amplicon sequencing,

89  shotgun metagenomics, culturing, culture-independent bacterial and viral abundance

90  estimates, and building science measurements, we addressed the following questions:

91  What are the successional dynamics of BE microbial communities? How stable are these

92  communities over different timescales? Do we see reproducible assembly of the same

93  microbial community? How and when do different source environments contribute to BE

94  microbial communities? How long do microbes remain viable on BE surfaces, and do we

95  see persistently viable human pathogens? What is the structure and diversity of the

96  surface-associated viral communities, and how does viral abundance relate to bacterial

97    abundance? And finally, what effects do environmental factors (e.g., temp, humidity,

98    occupancy) have on the diversity and abundance of microbiota? To answer these

99    questions, we characterized the microbial community structure, function, and abundance

100   on floors, toilet seats and soap dispensers over time in 4 restrooms (high use and low use,

101   male and female) following decontamination of each surface. Initially, surfaces were

102   analyzed hourly, and then daily for up to 8 weeks. To determine the influence of humans

103   as a dispersal source, surfaces were sterilized again, and following 4 hours of use,

104   humans were excluded from the restrooms and longitudinal changes in community

105   structure and viability of floor-associated bacterial communities were determined.

106

107   **Materials and Methods**

108   *Sample Collection:*

109   A visual summary of the experimental design, sampling locations, and deconmatination

110   results are included in **Figures 1 and S1**. For detailed information about each study, see

111   below.

112   *Eight-Week Study*

113   Samples were collected from three surfaces (the toilet seat, the floor in front of the toilet,

114   and the soap dispenser pump) in two public restrooms on the third floor of the North Life

115   Sciences building at San Diego State University. Samples were collected once a week for

116   an eight-week period between November 22, 2011 and January 31, 2012. For all studies

117   presented in this manuscript, the floor samples were collected from an area beginning

118   directly below the edge of the toilet bowl and extending outwards away from the toilet.

119   Eight hours prior to sample collection each surface was decontaminated using a 10%

120  bleach solution that was allowed to sit for approximately twenty minutes. After the

121  twenty-minute bleach treatment, the surfaces were rinsed with sterile water (DNA and

122  RNA free molecular grade water). The treated surfaces were shown to be DNA and RNA

123  free by epifluroescence microscopy using SYBR Gold (Invitrogen) staining.

124  Approximately eight hours post decontamination, sterile rayon tipped swabs

125  (MacroPur™ Swab P) that had been moistened with sterile 1X phosphate buffered saline

126  (PBS) were used to obtain surface samples from the targeted restroom surfaces. 50.8 cm

127  x 50.8 cm areas were swabbed for each sample. The entire surface was passed over once

128  with the swab, which took approximately 45 seconds. After surface swabbing, the tips of

129  the swabs were broken off into 1.5 ml microtubes containing 500 µl of sterile 1X PBS

130  solution. The swab tips were then immediately vortexed for ten seconds. 100 µl of

131  solution was removed and fixed in 100 µl of 4% paraformaldehyde for later analysis. The

132  remaining samples were then stored at -20°C until further processing.

133  *Eight-Hour Study*

134  Samples were collected every hour over an eight-hour period from the floor in front of

135  the toilet in two female and two male restrooms in the North Life Sciences Building at

136  SDSU. For sampling purposes, the floor in front of each toilet was partitioned into eight

137  equally sized rectangles to ensure that we did not re-sample the same surface more than

138  once in the eight-hour period. This experiment was conducted on two different days:

139  November 30, 2012 and December 5, 2012 and on two different floors (first and third).

140  The female and male restrooms on the third floor are open to everyone and are used

141  frequently throughout the day. The restrooms on the first floor are locked and reserved

142  for faculty and staff use only, and are used much less frequently. An hour before sample

143  collection began, the floor in front of each toilet was soaked in 10% bleach for twenty

144  minutes as described above. Once an hour for eight time points, a sterile rayon tipped

145  swab that was moistened with sterile 1X PBS was used to swab one of the eight randomly

146  selected rectangles (25.4 cm x 25.4 cm) on the floor. The entire floor surface was passed

147  over once with the swab, which took approximately 30 seconds. The soap pump handles,

148  and the surfaces of the toilet seats were swabbed, making sure the entire surface was

149  passed over once. Sampling and storage for molecular work and microscopy was

150  performed as described previously. The time-series samples were completed in the four

151  restrooms on the two sampling days.

152  *Month-Long Study*

153  Samples were collected from three restroom surfaces (the toilet seat, the floor in front of

154  the toilet, and the soap dispenser pump) in the same four restrooms used in the eight-hour

155  study every other day beginning January 30, 2013 and ending February 27, 2013. All

156  sampled surfaces were bleach treated as described above the morning of January 30,

157  2013. This was the only time during the course of this experiment that the restroom

158  surfaces were treated with bleach. The first samples were collected in the afternoon of

159  January 30, 2013. Subsequent samples were collected every other afternoon until

160  February 27, 2013, yielding 15 sampling time points. 50.8 cm x 50.8 cm floor areas were

161  swabbed for each sample. The entire surface was passed over once with the swab, which

162  took approximately 45 seconds. The soap pump handles, and the surfaces of the toilet

163  seats were swabbed, making sure the entire surface was passed over once. Sampling and

164  storage for molecular work and microscopy was performed as described previously.

165  *Human-free study*

166    Male and Female, high-use public restroom floors (in front of the toilet seat in two stalls)

167    at San Diego State University were decontaminated as stated above. The restroom floor

168    was cleaned at 8:00 AM, and the restrooms were opened for use as needed for four hours.

169    At 12:00, the restrooms were locked for the remainder of the day to allow for sample

170    collection without restroom use. For sampling purposes, the floor in front of each toilet

171    was partitioned into five equally sized rectangles (25.4 cm x 25.4 cm) to ensure that we

172    did not re-sample the same surface more than once in the eight-hour period. The entire

173    surface was passed over once with the swab, which took approximately 30 seconds.

174    Using aseptic technique, a double tipped CultureSwab (BD, Franklin Lakes, NJ) was

175    dipped in 0.02 μm filtered 1X PBS and used to collect floor surface samples at 12:00,

176    14:00, 16:00, 18:00, and 20:00 hours. One of the two swabs was broken off into

177    microtubes filled with 500 μL of 1X PBS to be used for 16S rRNA gene sequencing and

178    for microscopy (as described above). The other swab was broken off into 600 μL of 0.02

179    μm filtered 1X PBS to be used later for culturing. After the two microtubes were

180    vortexed, 100 μL were removed from each 500 μL tube and added to 100 μL of 4%

181    paraformaldehyde and later analyzed by microscopy. From each 600 μL tube intended for

182    culturing, 80 μL were added to 5 mL of Tryptic Soy Broth (TSB) (BD, Franklin Lakes,

183    NJ) in 15 mL capacity Falcon tubes for culturing at 4 conditions, to mimic room

184    temperature and human body temperature: aerobic 25°C and 37°C, and anaerobic 25°C

185    and 37°C. For the anaerobic conditions, AnaeroGen (Oxoid, Lenexa, KS) packets were

186    used to remove oxygen inside of an airtight jar containing anaerobic samples, and oxygen

187    indicator strips were used to verify the absence of oxygen for the entire culture period.

188    For all samples, including negative and positive controls, the lids were cracked slightly to

189    ensure air exchange as typical for culturing. Aerobic samples were incubated for 2 days,

190    and anaerobic samples were incubated for 4 days. After the incubation period, the

191    samples were stored in a -20°C freezer until they were sent to Argonne National

192    Laboratory for DNA extraction, PCR, and sequencing.

193    *Microscopy*

194    Epifluorescence microscopy was used to ensure that all samples collected contained

195    microbes and VLPs as well as to estimate the abundance of both bacteria and VLPs in

196    each sample. We used methods developed by the Rohwer laboratory (10). Briefly, 100 µl

197    of each of the paraformaldehyde fixed samples collected from restrooms was suspended

198    in 5ml of sterile water, and then filtered onto 0.02 µm Whatman Anodisc filter

199    membranes. The filters were then stained with 1X SYBR Gold for 10 min in the dark

200    before being rinsed and mounted onto slides. The slides were visualized using the Leica

201    microscope at the Electron Microscope Facility at San Diego State University. Image-Pro

202    Plus software was used to record digital images of the slides as well as generate estimates

203    of both bacterial and viral abundance.

204    *Bacterial DNA Extraction*

205    Bacterial DNA was extracted directly from swab tips and residual 1X PBS collection

206    buffer using the PowerSoil DNA Isolation Kit (Mo Bio Laboratories) following the

207    protocol in Flores et al. (2011). The Eight-Hour Study and the One-Month time period

208    used the same protocol, but DNA was extracted using the PowerSoil-htp 96 Well Soil

209    DNA Isolation Kit (Mo Bio Laboratories). Extracted DNA was sent to Argonne National

210    Laboratory for sequencing.

211    *DNA Amplicon and Metagenome Sequencing*

212   Amplicon sequencing was performed using primers designed to be massively multiplexed

213   and cover the V4-V5 hypervariable region of the 16S rRNA gene using the standard

214   methods outlined by the Earth Microbiome Project (http://www.earthmicrobiome.org/emp-

215   standard-protocols/16s/) (11).  Samples were sequenced on the Illumina MiSeq platform, at

216   the Argonne National Laboratory core sequencing facility (11). Shotgun-metagenomic

217   sequencing for single-read annotation was performed on 7 surface swab samples

218   (F1FH811302012, F1MH211302012, F3FH412052012, F3FH711302012,

219   F3MH611302012, F3MH811302012, F3MH412052012). Metagenomic libraries were

220   prepared using 1 ng of genomic DNA and the Nextera XT protocol according to

221   manufacturer's instructions (Illumina). Metagenomes were run through the MG-RAST

222   annotation pipeline (12). Shotgun-metagenomes were sequenced for an additional 8

223   samples (S1Post25A12pmD3, S1Post25A4pmD3, S1Post37A12pmD1,

224   S1Post37A8pmD1, S1Post37A8pmD3, S1Post37An12pmD1, S1Post37An4pmD1,

225   S1Post37An6pmD1) from cultured swabs for genome assemblies. Libraries were

226   prepared as described above.

227   *Sequence analysis*

228   QIIME (v. 1.7.0, Quantitative Insights into Microbial Ecology; www.qiime.org) was used

229   to filter reads and cluster OTUs as described previously (11, 13). Approximately 18% of

230   the ~17 million raw amplicon reads were removed during quality filtering, leaving ~14

231   million reads for downstream analysis. Briefly, the open reference OTU picking script

232   (pick_open_reference_otus.py) (14) was employed, where sequences were first clustered

233   with the Greengenes (May 2013) reference database (15); OTUs that did not cluster with

234   known taxa (at 97% identity) in the database were then clustered *de novo*.  Singleton

235  sequences were removed prior to downstream analyses. Representative sequences for

236  each OTU were aligned using PyNast, with a minimum alignment overlap of 75 bp (16).

237  Alignments were used to build a phylogenetic tree (FastTree (17)). We computed alpha

238  diversity metrics among substrates using the alpha_diversity.py script in QIIME

239  (Shannon entropy, species richness, phylogenetic diversity), using the same sequence

240  depth for all samples (3700 sequences per sample). The beta_diversity_through_plots.py

241  script was used to compute beta diversity distances between samples (weighted UniFrac),

242  and to construct principal coordinate (PCoA) plots, and account for both the phylogenetic

243  composition (18) and the relative abundance of taxa. Beta-diversity comparisons were

244  done using ANOSIM (compare_categories.py; QIIME). We tested whether the

245  abundances of particular OTUs differed significantly between surfaces or sampling times

246  using ANOVA analyses (Bonferonni corrected) with the otu_category_significance.py

247  script. ANOVAs and linear regressions were run using the R software package, to

248  compare Shannon diversity to the metadata (19). QIIME was used to calculate the core

249  microbial communities for different surfaces, times, and dates. Taxonomic distributions

250  across sample categories were calculated (from phylum to genus levels) using the

251  summarize_taxa_through_plots.py script in QIIME. The 2D histogram of PCoA space

252  was generated using Matplotlib (20).

253  *Quality trimming and de novo metagenome assembly*

254  Raw shot gun sequence reads (paired end; average insert size = 180bp) were quality

255  trimmed using high-throughput sequence analysis toolkit, Nesoni

256  (shttp://www.vicbioinformatics.com/software.nesoni.shtml) at following parameters; --

257  adaptor-clip, --match 10, --max-errors 1, --clip-ambiguous yes, --quality 10 and --length

258      70. One of the samples (S1Post25A12pmD3) failed to sequence well, and 94.6% of

259      sequences were lost during quality trimming. The remaining 7 assembly samples were

260      high quality, and only lost 5.6% of their raw reads, on average, during quality filtering.

261      Quality trimmed data from 8 human free samples were assembled into scaffolds using

262      velvet assembler (version 1.2.10) (21) set at following parameters, k=51, -exp_cov=55, -

263      cov_cutoff=5, -ins_length 180 -ins_length_sd 20, -min_contig_lgth= 250 and

264      scaffolding-yes. Taxonomical status was assigned to the metagenomes using MetaPhlAn

265      (22).

266      *Metagenome based recovery of genomes*

267      Scaffolds (minimum length = 300bp) from human free metagenome assemblies were

268      clustered into bins using tetra-nucleotide frequency usage anomaly Z-statistics, read

269      depth and %G+C profiles. Briefly, tetra-nucleotide frequency (TNF) based matrix was

270      constructed for each assembly (scaffolds, minimum length = 3kb) using custom R script.

271      TNF values were arcsine square root-transformed before performing clustering via

272      hierarchical agglomerative clustering with squared Euclidean distance and ward criterion

273      in pv-clust package (23) A correlation based (minimum pair-wise $R^2$ value = 0.9) sub-

274      graph was constructed for each cluster. Each sub-graph was manually checked for read

275      depth and %G+C profiles and outliers (SEM ± 1) were excluded from further analysis.

276      Taxonomic status was assigned to the reconstructed bins using BLAST2LCA program

277      (https://github.com/emepyc/Blast2lca) and phylogenetically divergent contigs were also

278      removed. The completeness of the reconstructed draft genomes (n=6) was estimated

279      based on the single copy marker gene profiles (24). Paired end read information was used

280      to iteratively increase the length of draft genomes using the PRICE assembler (25).

281    Metagenome raw data, assemblies, and reconstructed population genomes were

282    compared against (26) *mecA* (broad-spectrum beta-lactam resistance) and staphylococcal

283    chromosome cassette mec (SCCmec, mobile genetic element that carries encoded by the

284    *mecA* gene)  reference gene sequences using BLASTX. Automated genome annotations

285    of reconstructed *Staphylococcus* population genomes (**Table S1**) were performed using

286    RAST (27) and KAAS (28) servers. A whole genome based maximum likelihood

287    phylogenetic tree was constructed for the reconstructed *Staphylococcus* genomes (**Table**

288    **S1**) and reference genomes using PhyoPhlAn (29).

289    *Data availability*

290    Raw amplicon data is available in the SRA database under accession number

291    SRP049338. 16S and genome assembly raw data, along with sample metadata, can be

292    accessed on FigShare (http://dx.doi.org/10.6084/m9.figshare.899218) and MG-RAST-

293    annotated samples can be accessed on the MG-RAST webserver under project number

294    8313: http://metagenomics.anl.gov/linkin.cgi?project=8313.

295

296    **Results and Discussion**

297    Approximately 14 million high-quality 16S rRNA V4 amplicons representing 77,990

298    distinct operational taxonomic units (OTUs) were generated from 602 samples, along

299    with bacterial and viral abundance counts. 4.5 million metagenomic reads were

300    sequenced from 7 floor samples for single-read annotation, and 34 million reads were

301    generated from 8 cultured enrichments for genome reassembly.

302    *Longitudinal analysis surface-associated microbial succession*

303      The communities associated with each surface (floor, soap dispenser, and toilet

304      seats) converged upon a confined region within PCoA space within 5 hours following

305      decontamination, and the resulting late-successional surface community structure did not

306      differ significantly across 8 weeks of continued sampling (ANOSIM, $p > 0.8$; **Fig. 2A,**

307      **Fig. 3**). Floor communities showed a rapid reduction in the relative abundances of

308      *Firmicutes* and *Bacteroidetes*, concomitant with a relative increase in the abundances of

309      *Proteobacteria*, *Cyanobacteria* (78.3 % of which are annotated as 'Chloroplast', which

310      are likely to be derived from dietary plant biomass (30) or from plant material tracked in

311      from outdoors) and *Actinobacteria* over the course of a day (**Fig. 2B**). Succession was

312      highly reproducible for floor-associated communities, showing equivalent trajectories in

313      four different restrooms over two separate days (**Fig. 3A-C**), and beta-diversity distance

314      was significantly correlated with time (mantel $r = 0.226$, $p < 0.0001$). Floor-associated

315      communities quickly develop toward a meta-stable region within PCoA space (the

316      smaller of the two peaks, **Fig. 2A**; **Fig. 3C**). However, over a slightly longer time frame a

317      more stable optimum is defined by a further reduction in *Firmicutes* and expansion in

318      *Proteobacteria* and *Cyanobacteria* (**Fig. 2**). A Bayesian classifier known as

319      SourceTracker (7), trained on the Earth Microbiome Project (EMP) database (includes

320      both human-associated and environmental samples), showed that the early-successional

321      community is dominated by fecal-associated taxa that are likely aerosolized by toilet

322      flushing (**Fig. 4**), and are largely displaced by skin- and outdoor-associated taxa within 8

323      hours (**Fig. 2B; Fig. 4**). The stable communities present on toilet seats and soap

324      dispensers were comprised of ~45% fecal- and ~45% skin-associated taxa (**Fig. 4**).

325      *Environmental factors influencing surface-associated microbial community structure*

326     Microbial communities clustered significantly based on sample surface (ANOSIM

327     R = 0.7428, p < 0.001). Only toilet seat samples clustered based on restroom gender

328     (ANOSIM R = 0.199, p = 0.001; **Fig. 5A**), with *Lactobacillus* and *Anaerococcus*

329     dominating female toilet seats (31-33), and the gut-associated *Roseburia* and *Blautia*

330     being more abundant on male toilet seats (ANOVA, FDR-corrected p < 0.05; **Fig. 5B**).

331     High-use and low-use restrooms had significantly different toilet seat and soap dispenser

332     microbial communities (ANOSIM R = 0.142, p = 0.001, and R = 0.091, p = 0.001,

333     respectively; **Fig. 5C**), but floor-associated communities were not significantly different.

334     Fecal-associated OTUs were found at a greater relative abundance on high-use toilet seats

335     (e.g. *Bacteroidetes* and *Coprococcus*), while skin-associated OTU were more prevalent

336     on low-usage seats (e.g. *Corynebacterium*; **Fig. 5D**).

337     *Changes in community structure after human-exclusion*

338     In the eight hours following the exclusion of humans there was no significant

339     reduction in the prevalence of fecal microbiota (Pearson's R = -0.1093; p = 0.45). This

340     persistence of fecal microbiota, as in the late-successional floor community, is potentially

341     due to dormant or endospore forming taxa (e.g. *Firmicutes*; **Fig. 2B**). We suggest that

342     active fecal microbiota would experience a significant shock when transiting between the

343     warm, moist, anaerobic conditions of the host to the colder, drier, aerobic floor. Thus, we

344     propose that fecal taxa able to enter into a dormant phase can persist for longer periods of

345     time on restroom surfaces, although further work needs to be done to quantify dormancy

346     in these BE systems. The floor community structure, following human exclusion,

347     resembled the late-successional floor communities, while the viable organisms that grew

348     in culture from these same samples resembled a subset of soap dispenser and toilet-seat

349    associated communities (weighted Unifrac; **Fig. S2**). These viable communities were not

350    significantly different across time or between culture conditions (i.e. anaerobic or

351    aerobic, at 25° or 37° C; weighted UniFrac; ADONIS, $p > 0.8$). Despite the lack of

352    statistical difference in beta-diversity between culture conditions, we did see notable

353    differences in the source environments. Skin- and outdoor-associated taxa comprised 68-

354    98%, and fecal taxa represented 0-15% of the cultured communities. On average, both

355    pre- and post-culture communities had a larger percentage of outdoor-associated taxa

356    than other samples from the prior experiments, which suggests that, over time, human-

357    associated taxa are displaced in the absence of dispersal (**Fig. 4**). Looking specifically at

358    different culture conditions, ambient temperature (25°C), under both aerobic and

359    anaerobic culture conditions, showed equal proportions of environmental- and skin-

360    associated taxa in the cultures (~1:1). However, at human body-temperature (37°C), skin-

361    associated microbiota were dominant (~70% of the community), regardless of oxygen

362    potential. Several *Staphylococcus* taxa were positively correlated with incubation

363    temperature (Bonferroni-corrected Pearson's $p < 0.05$), and several OTUs shared 100%

364    nucleotide similarity across the 16S rRNA V4 region with *Staphylococcus aureus*, which

365    is the most common cause of skin and soft tissue infections (34). The presence of

366    *Staphylococcus* isolates was verified by metagenome assembly of several pan-genomes

367    from cultured samples (**Fig. 6**, **Table S1**), which shows that opportunistic pathogens

368    remain viable on surfaces for many hours following human exclusion. The assembled

369    pan-genomes made up large fractions of the total number of shotgun reads from most of

370    sequenced culture metagenomes (**Table S2**). **Table S3** shows the distribution of

371    annotated genes across SEED subsystem functional categories for each pan-genome.

372   *Composition of the surface-associated core microbial community*

373         Of the taxa that were ubiquitous across samples, most were associated with phyla

374   that are known to dominate the human gut (except for *Corynebacterium*). While skin-

375   associated OTUs dominated overall, they were not represented in this core community,

376   because they were likely derived from a more diverse seed bank, as the human skin

377   microbiome is more variable through time than the gut (35). Concordantly, Shannon

378   diversity rises significantly from morning to afternoon (in the 8-hour study), which

379   supports the idea that late-successional taxa are derived from more diverse sources ($R^2 =$

380   0.282, p < 0.0001; **Fig. S3A**).

381   *Viral:bacterial ratios and the composition of the viral metagenome*

382         Bacterial and viral abundances, as determined by epifluorescence microscopy,

383   were significantly positively correlated (R = 0.764, p < 0.0001; **Fig. S3B**), and were

384   negatively correlated with temperature ($R^2 = 0.048$ and $R^2 = 0.153$, respectively; p <

385   0.03). Bacterial abundance was positively, albeit weakly, correlated with humidity ($R^2 =$

386   0.0412, p < 0.02). Viral abundance was significantly lower in high-usage restrooms

387   (ANOVA, p < 0.002), but no significant relationship existed between either viral or

388   bacterial abundance and gender. Absolute viral and bacterial abundance show no

389   discernable trend across time, as the bacterial community reaches an abundance plateau

390   ($\sim 6.2 \times 10^3$ cells cm$^{-2}$, as determined by microscopy) within 1 hour after decontamination,

391   suggesting limited growth and a high rate of dispersal. The viral:bacterial (v:b) ratio was

392   ~1:1, which is around 10-20 times lower than considered typical for environmental

393   samples (36). We speculate that this lowered ratio suggests microbial dormancy (e.g.,

394   sporulation, persistor cells, or lowered metabolic rates) as phage lytic cycles cannot occur

395    in dormant cells. Even if lytic cycles could occur, the low bacterial density observed on

396    these surfaces may further limit the spread of phage (37).

397         In the human-exclusion experiment, v:b ratios tended to be lower (0.2-0.3). The

398    ratios differed across culture conditions as well (aerobic at 25° C = 0.204 ± 0.253 SD;

399    anaerobic at 25° C = 0.040 ± 0.006 SD; aerobic at 37° C = 0.211 ± 0.131 SD; and

400    anaerobic at 37° C = 0.338 ± 0.151 SD). Most of the phage detected on restroom surfaces

401    were enterophages (within *Microviridae*; **Fig. S4**), and because there were few gut

402    microbes that persisted on surfaces, or that grew up in culture, it follows that phage

403    abundance would remain low in the absence of their hosts. The fact that the 25° C,

404    anaerobic condition consistently yielded the lowest v:p ratio may indicate that most

405    phage/hosts are unlikely to have encountered this condition in their original

406    environmental context (e.g. the anaerobic gut environment is at 37° C, and non-host

407    associated environments that are likely to be dispersal sources for restrooms, like soils,

408    tend to be aerobic). The most abundant viral group within *Microviridae* was enterophage

409     X174, indicating that most of the viral sequences detected on restroom surfaces were

410    derived from the gut. Human papilloma viruses (within the *Papilomaviridae* family) and

411    human herpes viruses (within *Herpesvirales*) were also detected in high abundance,

412    which are skin- and epithelium-associated (38, 39).

413    *Metagenome data show the prevalence of bacterial virulence factors*

414         In addition to standard cellular functions such as carbohydrate and protein

415    metabolism (**Fig. S5A**), virulence-associated functions were present, including

416    fluoroquinolone resistance, heavy metal efflux and multidrug efflux pumps (**Fig. S5B**).

417    *Staphylococcus* methicillin-resistance pathways were the 6th most abundant category

418 (representing 3% of 'Bacterial' sequences; **Fig. S5C**). *S. aureus* is increasing in

419 prevalence outside of the hospital environment (34), and these results suggest that it

420 might be a common constituent of public restroom surfaces. Culturing work showed a

421 high prevalence of Staphylococcal species, but the assembled pan-genomes of cultured

422 organisms did not contain methicillin-resistance genes.

423 **Conclusion**

424       When compared to host-associated environments, restroom surfaces are dry,

425 barren, and resource-poor. As such, these surfaces probably do not support considerable

426 microbial growth, as evidenced by low cell densities. Continual dispersal, dormancy, and

427 cell death appear to be the dominant forces shaping community structure through time,

428 with minor contributions from cell growth and competition. The prevalence of skin-

429 associated, rather than fecal-associated taxa, in the late-successional community suggests

430 that organisms are selected for their ability to persist in a dry, aerobic environment, which

431 is a very different environment from the gut. Human-associated microbiota, including

432 *Staphylococcus* strains, can remain viable on BE surfaces for many hours after their

433 dispersal agents are removed. This suggests that common BE surfaces may be significant

434 fomites for viable human pathogens.

435 **Acknowledgements**

446    **References**

447    1.    **Custovic A, Taggart S, Woodcock A.** 1994. House dust mite and cat allergen in
448          different indoor environments. Clin Exp Allergy **24:**1164.
449    2.    **Kelley ST, Gilbert JA.** 2013. Studying the microbiology of the indoor
450          environment. Genome Biol **14:**1-9.
451    3.    **Kembel SW, Jones E, Kline J, Northcutt D, Stenson J, Womack AM,**
452          **Bohannan BJ, Brown G, Green JL.** 2012. Architectural design influences the
453          diversity and structure of the built environment microbiome. ISME J **6:**1469-
454          1479.
455    4.    **Flores GE, Bates ST, Knights D, Lauber CL, Stombaugh J, Knight R, Fierer**
456          **N.** 2011. Microbial Biogeography of Public Restroom Surfaces. PLoS One
457          **6:**e28132.
458    5.    **Rintala H, Pitkäranta M, Toivola M, Paulin L, Nevalainen A.** 2008. Diversity
459          and seasonal dynamics of bacterial community in indoor environment. BMC
460          Microbiol **8:**56.
461    6.    **Angenent LT, Kelley ST, Amand AS, Pace NR, Hernandez MT.** 2005.
462          Molecular identification of potential pathogens in water and air of a hospital
463          therapy pool. P Natl Acad Sci USA **102:**4860-4865.
464    7.    **Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG,**
465          **Bushman FD, Knight R, Kelley ST.** 2011. Bayesian community-wide culture-
466          independent microbial source tracking. Nat Methods **8:**761-763.
467    8.    **Hewitt KM, Gerba CP, Maxwell SL, Kelley ST.** 2012. Office space bacterial
468          abundance and diversity in three metropolitan areas. PLoS One **7:**e37849.
469    9.    **Knight R, Jansson J, Field D, Fierer N, Desai N, Fuhrman JA, Hugenholtz P,**
470          **van der Lelie D, Meyer F, Stevens R.** 2012. Unlocking the potential of
471          metagenomics through replicated experimental design. Nat Biotechnol **30:**513-
472          520.
473    10.   **Sano E, Carlson S, Wegley L, Rohwer F.** 2004. Movement of viruses between
474          biomes. Appl Environ Microb **70:**5842-5846.
475    11.   **Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N,**
476          **Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G,**
477          **Knight R.** 2012. Ultra-high-throughput microbial community analysis on the
478          Illumina HiSeq and MiSeq platforms. ISME J **6:**1621-1624.

479    12.    **Glass EM, Meyer F.** 2011. The Metagenomics RAST Server: A Public Resource
480          for the Automatic Phylogenetic and Functional Analysis of Metagenomes, p. 325-
481          331, Handbook of Molecular Microbial Ecology I. John Wiley & Sons, Inc.

482    13.    **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello**
483          **EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST,**
484          **Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD,**
485          **Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann**
486          **J, Yatsunenko T, Zaneveld J, Knight R.** 2010. QIIME allows analysis of high-
487          throughput community sequencing data. Nat Meth **7:**335-336.

488    14.    **Rideout JR, He Y, Navas-Molina JA, Walters WA, Ursell LK, Gibbons SM,**
489          **Chase J, McDonald D, Gonzalez A, Robbins-Pianka A.** 2014. Subsampled
490          open-reference clustering creates consistent, comprehensive OTU definitions and
491          scales to billions of sequences. PeerJ **2:**e545.

492    15.    **McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A,**
493          **Andersen GL, Knight R, Hugenholtz P.** 2012. An improved Greengenes
494          taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria
495          and archaea. ISME J **6:**610-618.

496    16.    **Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight**
497          **R.** 2010. PyNAST: a flexible tool for aligning sequences to a template alignment.
498          Bioinformatics **26:**266-267.

499    17.    **Price MN, Dehal PS, Arkin AP.** 2010. FastTree 2 – Approximately Maximum-
500          Likelihood Trees for Large Alignments. PLoS One **5:**e9490.

501    18.    **Lozupone C, Knight R.** 2005. UniFrac: a New Phylogenetic Method for
502          Comparing Microbial Communities. Appl Environ Microb **71:**8228-8235.

503    19.    **R Core Development Team.** 2008. R: A language and environment for statistical
504          computing. R Foundation for Statistical Computing, Vienna, Austria.

505    20.    **Hunter JD.** 2007. Matplotlib: A 2D graphics environment. Comput Sci Eng 90-
506          95.

507    21.    **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly
508          using de Bruijn graphs. Genome Res **18:**821-829.

509    22.    **Haft DH, Tovchigrechko A.** 2012. High-speed microbial community profiling.
510          Nat Methods **9:**793-794.

511    23.    **Suzuki R, Shimodaira H.** 2006. Pvclust: an R package for assessing the
512          uncertainty in hierarchical clustering. Bioinformatics **22:**1540-1542.

513    24.    **Wu M, Scott AJ.** 2012. Phylogenomic analysis of bacterial and archaeal
514          sequences with AMPHORA2. Bioinformatics **28:**1033-1034.

515    25.    **Ruby JG, Bellare P, DeRisi JL.** 2013. PRICE: software for the targeted
516          assembly of components of (meta) genomic sequence data. G3 **3:**865-880.

517    26.    **Elements IWGotCoSCC.** 2009. Classification of staphylococcal cassette
518          chromosome mec (SCCmec): guidelines for reporting novel SCCmec elements.
519          Antimicrob Agents Ch **53:**4961-4967.

520    27.    **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K,**
521          **Gerdes S, Glass EM, Kubal M.** 2008. The RAST Server: rapid annotations using
522          subsystems technology. BMC Genomics **9:**75.

523    28.    **Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M.** 2007. KAAS: an
524            automatic genome annotation and pathway reconstruction server. Nucleic Acids
525            Res **35:**W182-W185.
526    29.    **Segata N, Börnigen D, Morgan XC, Huttenhower C.** 2013. PhyloPhlAn is a
527            new method for improved phylogenetic and taxonomic placement of microbes.
528            Nat Commun **4**.
529    30.    **Kelley ST, Dobler S.** 2011. Comparative analysis of microbial diversity in
530            Longitarsus flea beetles (Coleoptera: Chrysomelidae). Genetica **139:**541-550.
531    31.    **Vásquez A, Jakobsson T, Ahrné S, Forsum U, Molin G.** 2002. Vaginal
532            Lactobacillus flora of healthy Swedish women. J Clin Microbiol **40:**2746-2749.
533    32.    **Ma B, Forney LJ, Ravel J.** 2012. Vaginal microbiome: rethinking health and
534            disease. Annu Rev Microbiol **66:**371-389.
535    33.    **Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL,**
536            **Karlebach S, Gorle R, Russell J, Tacket CO.** 2011. Vaginal microbiome of
537            reproductive-age women. P Natl Acad Sci USA **108:**4680-4687.
538    34.    **LeBlanc DM, Reece EM, Horton JB, Janis JE.** 2007. Increasing incidence of
539            methicillin-resistant Staphylococcus aureus in hand infections: a 3-year county
540            hospital experience. Plast Reconstr Surg **119:**935-940.
541    35.    **Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R.** 2009.
542            Bacterial community variation in human body habitats across space and time.
543            Science **326:**1694-1697.
544    36.    **Weinbauer MG.** 2004. Ecology of prokaryotic viruses. FEMS Microbiol Rev
545            **28:**127-181.
546    37.    **Lewis K.** 2010. Persister cells. Annu. Rev. Microbiol. **64:**357-372.
547    38.    **Davison AJ, Eberle R, Ehlers B, Hayward GS, McGeoch DJ, Minson AC,**
548            **Pellett PE, Roizman B, Studdert MJ, Thiry E.** 2009. The order herpesvirales.
549            Arch Virol **154:**171-177.
550    39.    **Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah**
551            **KV, Snijders PJ, Peto J, Meijer CJ, Munoz N.** 1999. Human papillomavirus is
552            a necessary cause of invasive cervical cancer worldwide. J Pathol **189:**12-19.

**Figure Legends**

**Figure 1.** (A) Samples were collected from three surfaces in both female and male
restrooms at San Diego State University. The surfaces analyzed were the toilet seat, the
floor in front of the toilet, and the soap dispenser pump. Epifluorescence microscopy
confirmed that bacteria and virus like particles (VLPs) are present on all three surfaces.
(Restroom cartoon modified from Flores et al., 2011). (B) Epifluorescence microscopy
images show selected restroom surfaces are DNA and RNA free after twenty minutes of
treatment with 10% bleach; T is the length of time the surface was soaked in bleach.

**Figure 2.** (A) A 2D histogram of floor samples (including 8-hour, 8-week, and month-
long experiments) in principal coordinate space (weighted UniFrac). Peaks denote areas
within principal coordinate space where samples are found most frequently (regions of
stability). The smaller peak corresponds to the later time points in the 8-hour study. The

569 larger peak shows the stable community state that remains relatively fixed in the 9-week
570 and month-long samplings (corresponding to the community structures highlighted in red
571 in panel B). We only observed early-successional community composition in the 8-hour
572 time series. Over longer timescales, the community was consistently found in the late-
573 successional state. (B) Composition of the microbial community along its successional
574 trajectory. The asterisk above the 8AM time point denotes the sample taken directly
575 following rigorous decontamination of the floor surface with bleach. The orange and red
576 boxes surrounding time points refer to the average community states characteristic of the
577 two peaks (labeled with corresponding colors: orange and red) seen in panel A.
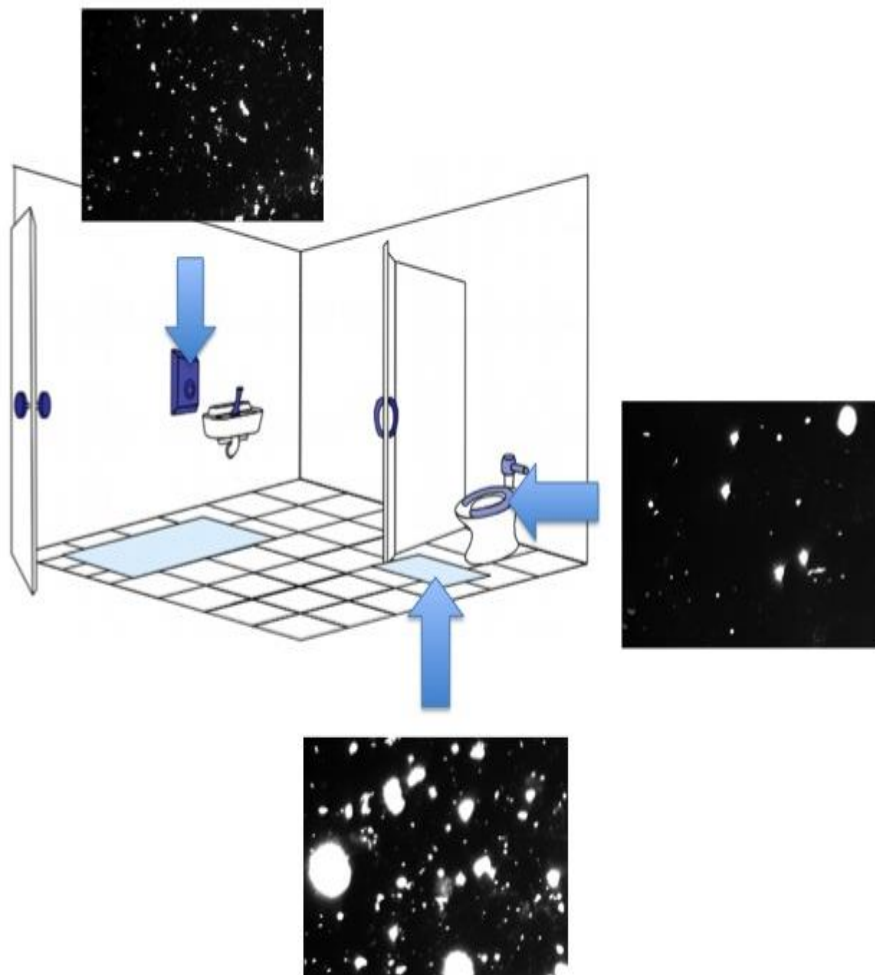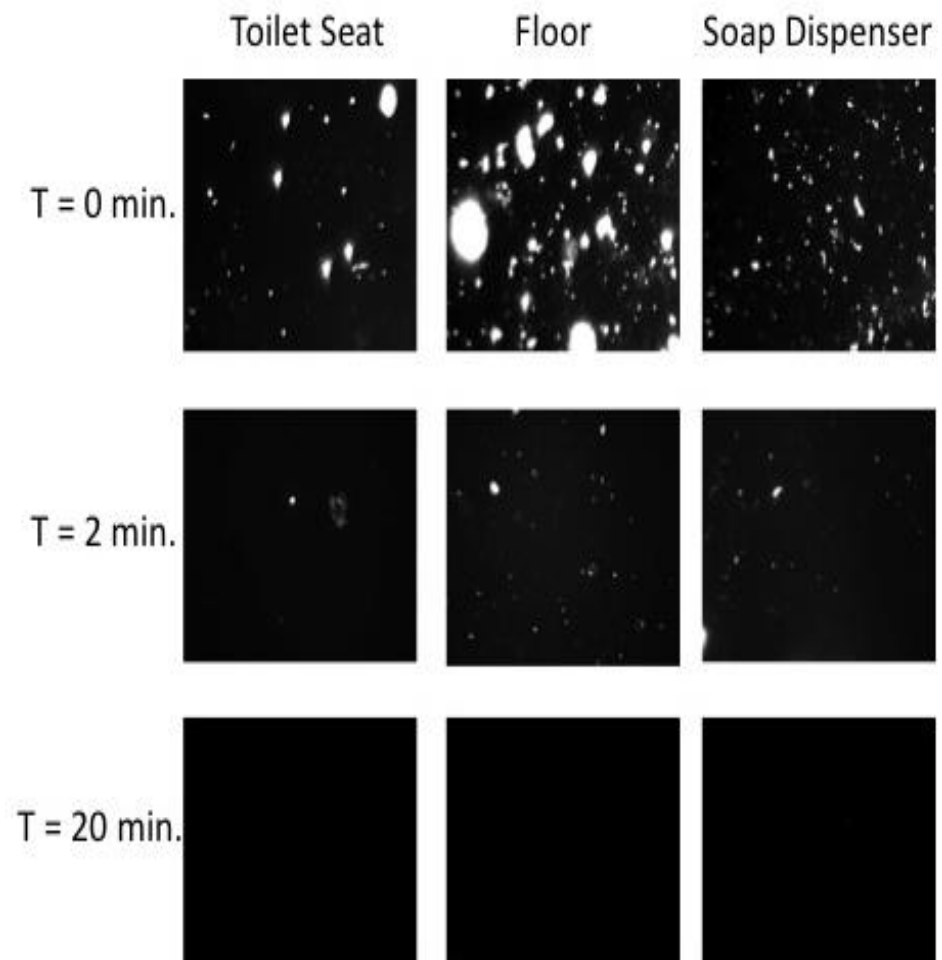578
579 **Figure 3.** PCoA plots of floor microbial communities over different timescales. (A)
580 Replicate 8-hour time series experiments cluster on top of one another. (B) Same plot as
581 in (A), but with samples colored by time point, and time point replicates encapsulated by
582 convex hulls. (C) Samples from the 8-hr experiment (rainbow colors) show a larger
583 spread than samples taken from longer-term studies (8-week and 1-month with daily
584 sampling; dark blue), showing that succession is rapid and occurs within 5-8 hours. Black
585 arrow in panel C shows the successional trajectory.
586
587 **Figure 4.** The relative proportion of OTUs derived from particular source environments
588 (as determined using SourceTracker). The first eight samples represent the average of
589 replicates taken at each hour for the 8-hour study. The 8-week and month bars represent
590 the average of all replicates for floor samples only. Soap and seat bars show the average
591 of replicates for those surfaces across the 8-week and month-long studies. The pre- and
592 post-culture bars represent the averages from the human-exclusion study (floor samples
593 prior to culturing, and after culturing). The asterisk above the 8AM bar indicates that the
594 floor was decontaminated prior to taking this sample. The source environment database
595 was constructed using Earth Microbiome Project (EMP) data (closed reference OTUs;
596 Greengenes release from May, 2013). The 'outdoor' category includes database samples
597 from many outdoor environments: freshwater, freshwater microbial mat, freshwater
598 sediment, bird nest, hotsprings water, hotsprings microbial mat, ice, marine biofilm,
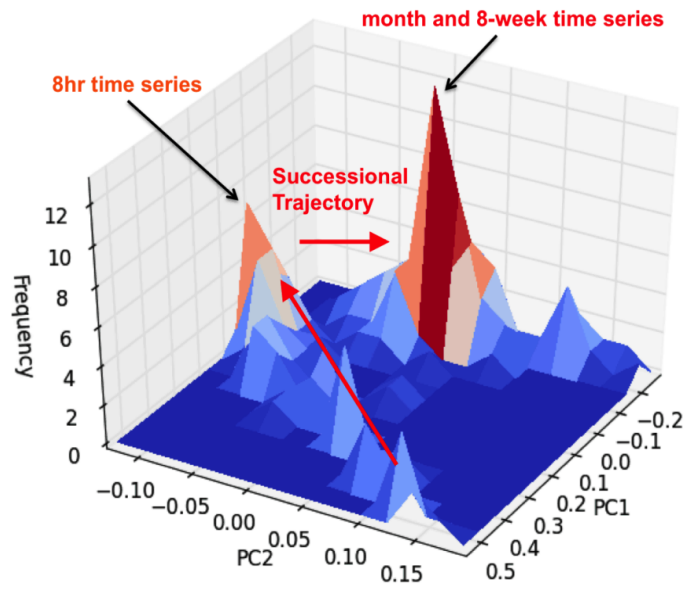599 marine water, marine sediment, hypersaline water, sand, sandstone, soil.
600
601 **Figure 5**. PCoA (weighted UniFrac) in (A) shows a clear separation of toilet seat
602 associated microbial communities based on gender. (B) depicts the top 10 most abundant
603 OTUs that show significant differences in abundance between male and female toilet
604 seats (most resolved taxonomic annotation shown on x-axis). (C) shows seat samples
605 cluster separately based on restroom usage frequency (high vs. low). (D) displays 7
606 OTUs that exhibit significantly different abundances between high and low usage
607 frequencies (most resolved taxonomic annotations are listed along x-axis).
608 **Figure 6.** Species-level metagenomic diversity from metagenomic assemblies of post-
609 culture swab samples, from the human free study. (A) species-level (SK1-SK8) relative
610 abundance patterns across 8 metagenomes, with hierarchical clustering of both rows and
611 columns (average linkage clusters, using Bray-Curtis distance), reporting only the 25
612 most abundant species annotations (according to the 90th percentile of the abundances).
613 The heatmap key shows percent relative abundance. (B) Rooted tree representing the

614    phylogenetic position of 6 *Staphylococcus* population genomes (highlighted in blue),

615    along with reference strains and *Bdellovibrio bacteriovorus HD100* as the out-group. The

616    tree was constructed using PhyloPhlAn (29) with concatenated amino acid sequences

617    from ~400 conserved proteins. Values assigned to internal nodes within the phylogeny

618    represent bootstrap support (bootstrap values < 0.50 are not reported). The scale bar

619    represents 0.2 changes per amino acid position.

620

**A.**



**B.**

| | Toilet Seat | Floor | Soap Dispenser |
|---|---|---|---|
| T = 0 min. | | | |
| T = 2 min. | | | |
| T = 20 min. | | | |

**A.**

**B.**

**A.**

PCoA - PC1 vs PC2

● Dec. 5, 2012
▶ Nov. 30, 2012

PC1 - Percent variation explained 53.97%
PC2 - Percent variation explained 4.78%

**B.**

PCoA - PC1 vs PC2

9AM
8AM
10AM
11AM
12PM
2PM
1PM
3PM

PC1 - Percent variation explained 53.97%
PC2 - Percent variation explained 4.78%

**C.**

PC2 (11 %)

■ 8-week and month-long study floor samples
■ 8 AM
■ 9 AM
■ 10 AM
■ 11 AM
■ 12 PM
■ 1 PM
■ 2 PM
■ 3 PM

PC1 (29 %)

PC3 (8 %)

**A.** PCoA - PC1 vs PC2

- Female
- Male

PC2 - Percent variation explained 13.30%
PC1 - Percent variation explained 31.19%

**B.**

Relative Abundance

- Female
- Male

Lactobacillus, Lactobacillus iners, Roseburia, Lactobacillus, Anaerococcus, Anaerococcus, Lactobacillus, Blautia, Anaerococcus, Corynebacterium

**C.** PCoA - PC1 vs PC2

- low usage
- high usage

PC2 - Percent variation explained 15.50%
PC1 - Percent variation explained 25.87%

**D.**

Relative Abundance

- high usage
- low usage

Bacteroides, Erysipelotrichaceae, Coprococcus eutactus, Phascolarctobacterium, Sphingomonadaceae, Corynebacterium, Cloacibacterium

**A.**



```
10  20  30  40  50  60  70  80  90
```

s__Staphylococcus_hominis
s__Staphylococcus_warneri
s__Methylobacterium_unclassified
s__Rhodopseudomonas_palustris
s__Bradyrhizobium_japonicum
s__Staphylococcus_epidermidis
s__Bacillus_subtilis
s__Staphylococcus_capitis
s__Pseudomonas_unclassified
s__Staphylococcus_haemolyticus
s__Mycobacterium_unclassified
s__Frankia_unclassified
s__Staphylococcus_lugdunensis
s__Faecalibacterium_unclassified
s__Bacteroides_coprocola
s__Akkermansia_muciniphila
s__Bacteroides_plebeius
s__Escherichia_coli
s__Faecalibacterium_prausnitzii
s__Bacteroides_vulgatus
s__Bifidobacterium_adolescentis
s__Alistipes_shahii
s__Sutterella_wadsworthensis
s__Escherichia_unclassified
s__Alistipes_putredinis

S1Post25A4pmD3 = SK7
S1Post37A8pmD3 = SK2
S1Post37An6pmD1 = SK5
S1Post25A12pmD3 = SK8
S1Post37A12pmD = SK3
S1Post37An12pmD1 = SK6
S1Post37A8pmD1 = SK1
S1Post37An4pmD1 = SK4

**B.**

*Bdellovibrio bacteriovorus* HD100
*Staphylococcus hominis subsp hominis* C80
*Staphylococcus* SK1
*Staphylococcus* sp 693-2
*Staphylococcus haemolyticus* JCSC 1435.BRC
*Staphylococcus lugdunensis* HKU09-01
*Staphylococcus* SK5
*Staphylococcus warneri* L37603
*Staphylococcus epidermis* A487
*Staphylococcus* SK6
*Staphylococcus* SK3
*Staphylococcus epidermis* RP62A.BRC
*Staphylococcus* SK4
*Staphylococcus epidermis* M23864-W1
*Staphylococcus caprae* C87
*Staphylococcus* SK2
*Staphylococcus capitis* CR01

99
100
100
100
100
100

0.2