AEM Accepts, published online ahead of print on 29 March 2013
Appl. Environ. Microbiol. doi:10.1128/AEM.00331-13
Copyright © 2013, American Society for Microbiology. All Rights Reserved.

AEM Accepts published online ahead of print

1 **Revised**

2

3 **Culture-Independent Analysis of Aerosol Microbiology in a Metropolitan Subway**
4 **System**

5

6 **Running title:** Subway bioaerosol microbiology

7

8 Charles E. Robertson[1#], Laura K. Baumgartner[1#§], Jonathan K. Harris[2], Kristen L.
9 Peterson[1], Mark J. Stevens[2], Daniel N. Frank[3], and Norman R. Pace[1*]

10

11 [1] Department of Molecular, Cellular, and Developmental Biology. University of Colorado,
12 Boulder, CO USA.
13 [2] Department of Pediatrics, University of Colorado School of Medicine, Aurora, CO USA.
14 [3] Division of Infectious Diseases, University of Colorado School of Medicine, Aurora, CO
15 USA.

16

17 * **Corresponding author.** Mailing address: Department of Molecular, Cellular, and
18 Developmental Biology, University of Colorado, Boulder, CO, USA 80309-0347. Phone:
19 (303) 735-1864. Fax: (303) 492-7744. Email: norman.pace@colorado.edu.

20

21 [#] Contributed equally to study.

22

23 [§] Current address: Front Range Community College, Longmont, CO 80501

24

25 **Keywords:** Pyrosequences / Sanger sequences / rRNA phylogeny / aerosol bacteria /
26 subway bacteria

27

AEM Accepts published online ahead of print

ABSTRACT

The goal of this study was to determine the composition and diversity of microorganisms associated with bioaerosols in a heavily trafficked metropolitan subway environment. We collected bioaerosols by fluid impingement on several New York City subway platforms and associated sites in three sampling sessions over a 1½ year period. The types and quantities of aerosolized microorganisms were determined by culture-independent phylogenetic analysis of small-subunit ribosomal RNA gene sequences, using both Sanger (universal) and pyrosequencing (bacterial) technologies. Overall the subway bacterial composition was relatively simple; only 26 taxonomic families made up ~75% of sequences determined. The microbiology was more or less similar throughout the system and with time, and most similar to outdoor air, consistent with highly efficient air mixing in the system. Identifiable bacterial sequences indicated that the subway aerosol assemblage was composed of a mix of genera and species characteristic of soil, environmental water, and human skin commensal bacteria. Eukaryotic diversity was mainly fungal, dominated by organisms of types associated with wood rot. Human skin bacterial species (at 99% rRNA sequence identity) included the *Staphylococcus* spp. *S. epidermidis* (the most abundant and prevalent commensal of the human integument), *S. hominus*, *S. cohnii*, *S. caprae* and *S. haemoliticus*, all well-documented human commensal bacteria. We encountered no organisms of public health concern. This study is the most extensive culture-independent survey of subway microbiota so far and puts in place pre-event information required for any bioterror surveillance activities or monitoring of the microbiological impact of recent subway flooding events.

AEM Accepts published online ahead of print

59 **Introduction**

60

61 The microbiological quality of the air we encounter daily, and depend upon
62 absolutely, is a significant yet little-addressed societal concern. Regulatory focus on air
63 quality has been on chemical and particulate materials, which are readily measured.
64 However, there is comparably little knowledge of the nature of the aerosolized
65 microorganisms and microbial products that occur in different public settings and to
66 which the public is exposed daily. Because many public places concentrate large
67 numbers of humans and therefore may be key locales for transmission of natural
68 pathogens or deliberately released agents, understanding the microbial ecology of
69 bioaerosols in public settings is critical for public health, occupational health, and
70 biodefense. For instance, prior knowledge of composition, sources, and temporal and
71 spatial dynamics of bioaerosol microbes is essential for tracking pathogen dispersal in
72 public settings.

73

74 One public arena through which large numbers of people pass daily is the
75 municipal subway system. Subway facilities have been established by major
76 metropolitan areas throughout the world. New York City (NYC) subways, for instance,
77 had a ridership of 1.6 billion in 2011 (1). The huge number of people exposed to the
78 subway environment underscores the importance of developing some understanding of
79 the microbiological air quality of subway platforms.

80

81 The results of previous studies emphasize that airborne particulate materials in
82 subways are different from what is found on city streets or in other indoor environments.
83 This is particularly due to aerosolized metallic dust, which most likely is generated by
84 the action of iron train wheels on tracks (2)(3)(4)(5)(6)(7). Chillrud and colleagues have
85 examined the effects of NYC subway exposure on children (8) and subway workers
86 (9)(10), and saw no health effects. Birenzvinge et al. showed that the load of particulate
87 matter in the air on the subway platforms was correlated to the frequency of train traffic
88 (11). The unusual particulate quality of the subway environment may indicate also that

AEM Accepts published online ahead of print

89 the microbial contents of subway aerosols differ from those detected in other indoor or
90 outdoor environments.
91
92 Thus far, studies of aerosol microbiology in the subway environment have
93 focused primarily on culture-dependent techniques, most notably viable counts of
94 bacteria and fungi, sometimes with biochemical or molecular identification of cultivars
95 (3)(12)(13)(14)(15). However, studies that focus on colony forming units likely identify
96 only a small fraction of actual microbial contents because most environmental microbes
97 are not cultured using standard techniques (16). Consequently, there is little general
98 perspective on the nature of microorganisms that might be encountered in subway air.
99
100 In order to gain some overview of the nature of the microbial load of a subway
101 system independently of culture, we examined the contents of aerosols collected on
102 several New York City subway platforms and associated sites (terminal, park and
103 unused platform) during three sampling sessions over a 1 ½ year period. The
104 compositions of aerosolized microbial loads were determined by culture-independent
105 phylogenetic analysis of small-subunit rRNA gene sequences, using both Sanger and
106 454 sequencing technologies. The longer Sanger sequences provide for more accurate
107 phylogenetic identification, whereas pyrosequences provide more comprehensive
108 sampling of aerosol microbiology.
109

110 Materials and Methods

111

112 Air sampler setup, cleaning and blanks.

113 Air was sampled with a custom-modified Omni 3000 fluid impinger (Innovaprep
114 LLC, Drexel, MO), modified to minimize tubing and allow full removal of tubing for
115 cleaning and sterilization. Omni particle capture efficiency ranges from 40% for 0.5 µm
116 particles to >90% for >3 µm particles (17) The device was chosen because of its
117 relatively high capture rate of bacteria-sized particles, high volume of air passage (250-
118 300 l/min in this study) to minimize sampling time, and relatively quiet function, to
119 minimize conspicuity during this public sampling campaign.

AEM Accepts published online ahead of print

120
121 Empty sampling cartridges were sterilized by irradiation in a UV Stratalinker 1800
122 (Stratagene, La Jolla, CA), then molded shut with chloroform. Cartridges were filled
123 with a filter-sterilized impinging solution, which consisted of phosphate buffered saline
124 (PBS; 137 mM NaCl, 10 mM Phosphate, 2.7 mM KCl, pH 7.4) and 0.005% Tween in
125 diethylpyrocarbonate (DEPC) treated water. Blank cartridges were prepared at the
126 same time as the sample cartridges, carried during sampling, and processed in the
127 same manner as the air samples. None of the cartridge blanks produced amplifiable
128 DNA. The average air-sampling rate was ~300 l/min. During sampling, liquid volume lost
129 to evaporation was replaced with DEPC-treated water. The air sampler was cleaned
130 between stations by replacing all of the tubing with sterile tubing and cleaning the
131 contactor and ports with DEPC-treated water and alcohol (isopropanol wipe then
132 washed with 70% ethanol) and air-dried. Cleaning blank samples were taken by filling
133 the sampler (contactor and tubing) with 5 mL DEPC-treated water, which was allowed to
134 sit in the sampler for 5 minutes and then extracted by sterile syringe from the sampling
135 port. The cleaning blanks were filtered and extracted in the same fashion as the air
136 samples; none of the blanks produced amplifiable DNA.
137
138 **DAPI Counts.**
139 Aliquots of sample fluid were adjusted immediately following sampling to 4%
140 formaldehyde for direct cell counts, and held and shipped overnight on wet ice to the
141 laboratory. Aerosolized iron in most subway air samples interfered with microscopic
142 counts and was removed before counting: samples were vortexed gently for 30s, then
143 placed against a magnetic rack for ~1 min. Sample fluids were carefully removed with a
144 pipette and a known volume filtered through a 0.2 micron black polycarbonate filter.
145 The filter was placed on a slide, flooded with 10 µg/ml diamino-2-phenylindole (DAPI)
146 for 5-10 minutes, washed with two, 2-min washes of 10 mM Tris-Cl, 1 mM EDTA, pH7.4
147 (TE) and slides were counted at 400x.
148
149 **DNA Extraction, Amplification, and Sequencing (Sanger and 454).**

AEM Accepts published online ahead of print

150 Most (~85%) of the sample fluid from each collection was filtered through 0.2 µm
151 polycarbonate filters (Millipore Isopore, Billerica, MA)), which then were placed in sterile
152 microcentrifuge tubes, immediately frozen in liquid nitrogen and shipped frozen to the
153 laboratory, where they were stored at -80°C until processed.
154
155 Samples were processed to remove particulate iron before DNA extraction by
156 dissolving the filter in phenol, chloroform, and Buffer B (200 mM NaCl, 200 mM Tris-Cl
157 pH 8.0, 20 mM EDTA, 5% SDS) while vortex mixing. Samples were centrifuged at 100
158 x g for 2 minutes and then placed against a strong magnet. All liquids were removed to
159 a new sterile tube, leaving the magnetic iron behind. This was repeated and the sample
160 was then placed in a tube with zirconium beads for DNA extraction by bead beating with
161 two volumes of buffer-saturated phenol and ethanol precipitation of nucleic acids.
162
163 Extracted genomic DNA was amplified for cloning by PCR with the (nominally)
164 universal small subunit rRNA gene primers 515F and 1391R (18). PCR reactions were
165 conducted at 94 °C for 2 min, followed by 30 cycles at 94 °C for 20 s, 52 °C for 20 s,
166 and 65 °C for 1.5 min, followed by a 65 °C elongation step for 10 min. Each 50-µL
167 reaction contained 10 µL Eppendorf 2.5× HotMasterMix (Eppendorf, New York, NY), 10
168 µL water, 0.05% BSA (Sigma-Aldrich, St. Louis, MO), 100 ng of each oligonucleotide
169 primer, and 1–5 ng of DNA template. Triplicate PCR reactions were conducted for each
170 sample and pooled before purification with the Montage gel purification system
171 (Millipore). In some (9/27) gel purifications of PCR products 16S and 18S bands were
172 processed and analyzed separately. For most of the universal libraries, the mixture of
173 the two was processed. Some samples were not effective as templates and were not
174 analyzed further. PCR-amplified rRNA genes were cloned with Topo-TA according to
175 manufacturer's instructions (Life Technologies, Carlsbad, CA) and Sanger sequencing
176 was conducted on an Amersham MegaBACE 1000 capillary sequencer following the
177 manufacturer’s protocols.
178
179 DNA samples also were analyzed by pyrosequencing on a Roche 454 GS-FLX
180 platform. DNAs were amplified in three independent reactions using barcoded primers

AEM Accepts published online ahead of print

181 (27F-338R) (19). Negative PCR controls for each primer were assayed in parallel and
182 did not exhibit bands in agarose gels. The three independent reactions were pooled
183 and amplicons were confirmed by agarose gel electrophoresis. DNA contents of pools
184 were normalized using the SequalPrep Normalization plate (Life Technologies), and
185 equal amounts mixed to construct the amplicon pool (20). The amplicon pool was
186 concentrated by evaporation, size selected by electrophoresis on a 1.5% agarose gel
187 (Tris-acetate EDTA buffer) and gel purified by Montage kit (Millipore) prior to
188 sequencing. Pyrosequencing was conducted per manufacturers protocols using Roche
189 454 titanium chemistry.

190

191 **Sequence Analysis.**

192 Sanger sequences were quality filtered and assembled with XplorSeq (21). Raw
193 pyrosequences were quality filtered and sorted into their respective barcoded libraries
194 with BARTAB (22). Filtering for both Sanger and pyrosequence data removed
195 nucleotides with mean Q < 20 at 5' and 3' ends and over a 10 nt window; sequences
196 with >1 ambiguous base were discarded; and all sequences of length < 200 nt were
197 discarded. Infernal (23) and ChimeraSlayer (24) were used to screen bacterial
198 sequences as described previously (25). Taxonomic classification of all sequences was
199 done with the classifier functionality of standalone SINA using the Silva108 non-
200 redundant database as reference (376,437 sequences which are the 99% cluster
201 representatives of Silva108, (26)). Ecological statistics (such as Species Observed
202 (27)), pie charts, and heatmaps were prepared with the Explicet software package (CER
203 manuscript in preparation. Software available upon request to author.) Sanger
204 sequences have been deposited in GenBank under the accession numbers JX394222-
205 JX397762 and pyrosequences have been deposited in the Short Read Archive with
206 accession number SRA055336 (BioProject Accession number PRJNA169671).

207

208 Determination of potential sources of DNA was done by BLAST analysis of all
209 pyrosequences against three separate databases composed of long sequences
210 (≥1200nt) associated with the Human Skin Microbiome study (28), sequences in Silva
211 108 whose isolation source metadata tag contained the word "soil", and sequences in

AEM Accepts published online ahead of print

212 Silva 108 whose isolation source metadata tag contained the word "water". To be
213 considered a BLAST database match, pyrosequences were required to have a
214 minimum of 95% overlap with the BLAST database hit sequence, a minimum bit score
215 of 50 and either 95% or 99% identity with the respective BLAST database hit sequence.
216 We did not use tools that compare fingerprints of ecosystems with bioaerosol samples
217 because air is an assemblage, not a specific microbiome.
218
219 **Results**
220
221 The microbiological contents of nominally similar environmental samples tend to
222 vary site-to-site and over time. Variation is particularly expected in the subway setting,
223 a scattered collection of potentially more or less sequestered spaces connected by
224 kilometers of track tunnels. In order to explore both spatial and temporal aspects of the
225 microbiology of the NYC subway system, we collected air samples at seven subway
226 stations (specified in figures and located as shown in Figure S1) and three adjacent
227 sites in lower Manhattan over a 1 ½ year period (2007-2008). Two, 5-6 m^3 samples
228 were collected at each subway station for each sampling timepoint, one from each end
229 of the platform, using a metered, high-volume fluid impinger. The impinger was
230 modified so that all internal surfaces could be exposed for cleaning or disposal following
231 their use (Materials and Methods).
232
233 Sampling parameters and other metadata are summarized in Table S1.
234 Microbial cell counts (Materials and Methods) of typically 1-4x10^4 cells/m^3 were
235 observed throughout the system (average, 2.2x10^4 cells/m^3, Table S1). These directly
236 counted microbial loads were typical of outdoor air samples taken at the same time and
237 at the low end of loads typically encountered in indoor environments such as occupied
238 buildings (10^4-10^5 cells/m^3) (29)(30)(31). On the other hand, direct counts of
239 aerosolized subway bacteria are much higher, ~100-fold higher, than colony-forming
240 counts encountered in subway environments, which are typically several hundred
241 CFU/m^3 in different subway systems (14)(15)(13). We see no clear patterns of variation

AEM Accepts published online ahead of print

242 in cell counts by season, station or extent of local traffic (light vs. heavy, Table S1) in
243 this sampling.
244
245 The phylogenetic distribution of aerosolized microbes in each specimen was
246 determined by analysis of small-subunit (SSU) rRNA gene sequences. Total nucleic
247 acids were extracted from impinger fluids using a bead-beating protocol and processed
248 for Sanger or for pyrosequence analysis (Materials and Methods). Clone libraries for
249 Sanger sequencing were constructed from PCR products obtained using the (nominally)
250 universal primer set 515F-1391R; pyrosequencing libraries were developed using
251 bacteria-specific barcoded primers targeting the V1-V2 region of the small subunit rRNA
252 gene. The longer Sanger sequences (700-800 nt in this case) provide for relatively
253 accurate phylogenetic assignments and a three-domain census. The more numerous,
254 but shorter (~200 nt) pyrosequences provide a broader census assessment of bacteria.
255 Overall 3,541 Sanger sequences and 60,707 pyrosequences were determined and
256 deposited in GenBank.
257
258 **Phylum-level diversity of bioaerosols**
259 The phylum-level diversities encountered among Sanger (universal) and
260 pyrosequences (bacteria-only) were similar throughout the examined subway sites and
261 similar to outdoor air samples (Union Square Park, as summarized in Figure 1.
262 Qualitatively similar subway and outdoor bacterial diversity was found in the Sanger (Fig.
263 1A and 1B) and pyrosequence analyses (Fig. 1C and 1D), although there was some
264 quantitative variation. Few of the sequences (<1%) were identical to sequences in the
265 databases, but many were related to database sequences at the genus or even species
266 taxonomic levels (below). Based on the results from PCR amplification with universal
267 primers, bacterial (65%) and fungal (34%) sequences dominated the overall subway
268 rRNA assemblage (Fig. 1B). Archaeal sequences, consisting of a few crenarchaeotes
269 and methanogens with no specific known relatives, were rarely encountered (<1% of
270 total). Collectively the main bacterial sequence diversity observed in the subway
271 environment was remarkably simple, mainly comprising only four of the ~100 known
272 bacterial phyla: Actinobacteria, Firmicutes, Bacteroidetes and Proteobacteria, with only

AEM Accepts published online ahead of print

273 minor contribution from other phyla. Only a few sequences fell into candidate phyla
274 (OP10 and TM7) with no cultured representation. A full abundance-ranked list of
275 bacterial taxa detected in the study is compiled as Table S2.
276
277 Although simple at the phylum level, the diversity of subway air samples was
278 more complex at lower taxonomic levels. Comparison of taxon collectors curves (Fig. 2)
279 generated from pyrosequence and Sanger datasets indicates that the Sanger sequence
280 set, at 700-800 nucleotides in length, substantially under-samples the microbial diversity
281 relative to the pyrosequencing data set (~200 nucleotide length), although the most
282 abundant sequences are expected to occur in the Sanger data set (32). However, even
283 the pyrosequencing dataset did not exhaustively sample the microbes present in the
284 system as a whole, as indicated by the continuous increase in observed taxa (Sobs)
285 with ongoing sampling of sequences (Fig. 2). This is a common theme in environmental
286 sequence surveys; the pool of rare sequences, the "rare biosphere," is always extensive
287 (33). Nonetheless, pyrosequencing was sufficiently extensive to cover the majority of
288 the microbial diversity encountered. Consequently, we focus on the pyrosequences for
289 site-to-site statistical comparisons.
290
291 **Bacterial Diversity**
292 Despite the considerable diversity of rRNA sequences encountered in subway air
293 samples, only 26 family-level taxa comprised most (~75%) of the sequence types
294 observed; their distributions in the different stations, abundance-ranked by heat map,
295 are shown in Figure 3. (Table S1 lists all taxa encountered to 0.01% of total.) At this
296 taxonomic level of analysis, the makeup of the subway microbiology was generally
297 similar to that observed outdoors (Union Square Park) in an area with considerable
298 human traffic. Most of the abundant taxa occurred throughout the sample set at
299 consistently higher levels than other taxa, but the distributions of the abundant taxa
300 varied between different subway stations and times of sampling. For instance,
301 taxonomic families that were particularly abundant in one or more samples occurred in
302 other samples as well, but in lesser relative abundance (Fig. 3).
303

AEM Accepts published online ahead of print

304 We explored potential correlations in the distributions of the bacterial diversity
305 between different subway sites using standard ecological beta-diversity indices (i.e.,
306 Morisita-Horn and UniFrac), but identified few associations. As shown in Figure 4, a
307 Morisita-Horn analysis of similarities between the compositions of the different samples,
308 microbiotas tended to be similar in general (M-H index 0.6-0.8; identical communities
309 would have an index = 1.0). The Morisita-Horn analysis does indicate a few pairs of
310 stations with exceptionally high similarities (M-H index >0.9), which in some instances
311 can be rationalized by proximity. However, potential correlations such as station
312 proximity tend to be inconsistent in subsequent samplings.
313
314 A few samples in Figure 4 also are exceptionally different from the others. These
315 can be explained largely by idiosyncratic abundances of particular phyla. For instance,
316 dissimilarity between the GC456:1:2008-08 and other samples (Fig. 4; M-H index <0.3)
317 was driven by an unusual abundance of *Enterobacteriaceae* sequences (Fig. 4);
318 dissimilarity of the BowlingGreen:1: 2008-08 sample from others resulted from
319 idiosyncratic abundances of *Microbacteriaceae*, *Nocardioidaceae* and 480-2 (an
320 environmental actinobacterial sequence clade with no named representative) in that
321 sample (Fig. 3). Moreover, the somewhat unusual microbial compositions observed for
322 the GC456:2:2008-08 and BowlingGreen:1:2008-08 samples was not reproduced with
323 samples taken at the same sites but a year or more earlier (GC456:March2007, Bowling
324 Green:August2007), when the microbiology of those sites was generally similar to that
325 of other stations sampled. In general, however, any apparent correlations have no
326 particular and consistent explanations in terms of the sites or timepoints sampled.
327 Samples taken at the same sites months or a year apart usually had no specific
328 relationships; even samples taken on the same dates at opposite ends of particular
329 platforms (e.g. TimesSquare:1:2007 and TimesSquare:2:2007, and GC456:1:2008-08
330 and GC456:2:2008-08) were no more correlated with each other than with other
331 samples (M-H <0.9).
332
333 **Sources of Bacterial Diversity**

AEM Accepts published online ahead of print

334 At low taxonomic levels, genera and species, the sequence diversity seen
335 throughout the subway system was complex. However, most of the sequences fall into
336 taxonomic groups that also contain described and named species (Fig. 3): respectively,
337 65% and 70% of pyrosequences and Sanger sequences were classifiable to the genus
338 level. This is an unusually detailed taxonomic result with environmental samples, which
339 typically yield many sequence types not in reference databases (34). The result
340 indicates that the diversity we detected generally belongs to a relatively intensively
341 documented portion of the bacterial SSU rRNA tree. Many of the sequences
342 determined were nearly identical to rRNA gene sequences of named species and, in
343 some cases, the nature of those species points to the sources of the microbiology
344 encountered in the system.
345
346 In order to gain some further perspective on the sources of the sequences, we
347 used BLAST analysis to compare the subway sequences to those of databases
348 representing other environments. As general reference environments, we considered
349 human skin, soil, and water (Materials and Methods). Figure 5A shows comparisons of
350 the subway dataset with those environmental datasets at 95% sequence overlap and
351 two levels: 95% sequence identity, approximately the genus-level; and 99% identity, a
352 stringent species-level relationship. The results of the comparisons establish the
353 general propensity of identifiable sequences to sort with both environmental (soil and
354 water) and human skin microbiota, as indicated. Approximately 20% of identifiable
355 species are associated with human skin. Additional detail is provided in Figure 5B,
356 which indicates that sequences of identifiable human skin microbiota in the subway
357 station air seem derived mainly from the foot, hands, arms, and head. Details of body
358 site representation for the different stations are given in Supplementary Figure S2.
359
360 One conspicuous example of putative human skin microbes in subway aerosols
361 is the most abundant taxonomic family throughout the system (Fig. 3), the
362 *Staphylococcaceae*. Overall, 85% of the sequences representing *Staphylococcaceae*
363 were members of the genus *Staphylococcus* in the Silva 108 database. Most of the
364 observed *Staphylococcus* spp. sequences were not affiliated with named species, but

AEM Accepts published online ahead of print

365 rather represented a broad phylogenetic distribution within the genus. However, 475 of
366 the ~3000 *Staphylococcus* spp. sequences were identified (99% sequence identity) with
367 5 well-studied species: *S. epidermidis* (the most abundant by ~4-fold and the most
368 abundant and prevalent commensal on the human integument), *S. hominis*, *S. cohnii*, *S.*
369 *caprae* and *S. haemoliticus*. These are all well documented human commensal
370 bacteria, associated mainly with skin. *Staphylococcus* spp. were particularly abundant in
371 areas of human habitation, for instance Grand Central Station mezzanine
372 (GCMez:1:2007-08) and Times Square stations (Fig. 3). Similar human–related
373 microbiological signatures also were represented in other of the family groups cited in
374 Figure 3. The *Micrococcaceae* sequence collection, for instance, includes *Kocuria* spp.,
375 *Micrococcus* spp. and others known as human skin microbiota.
376
377 Of course, many organisms related to known environmental microbes also were
378 identifiable: for instance ~20% of the *Micrococcaceae* sequences correspond to those of
379 *Arthrobacter* spp., which are related to common soil organisms. The *Moraxellaceae*
380 sequences, as further example, include many *Acinetobacter* spp. and *Psychrobacter*
381 spp. sequences related to known soil and water microbes. Overall, as summarized in
382 Figure 5A, the bacterial diversity of the subway system aerosol that can be identified as
383 to source was comprised mainly of soil and water microbes with a significant overlay of
384 human skin microorganisms. No pathogens beyond those associated with human
385 commensal organisms were observed.
386
387 **Eukaryotic diversity**
388 The Sanger sequences based on universal primers provide the only perspective
389 on aerosolized eukaryotic diversity detected in the study. Nearly all (96%) of the
390 eukaryotic sequences corresponded to diverse fungi. The few eukaryotic rRNA
391 sequences beyond fungal sequences included insect, plant, rodent, and a few protist
392 sequences. Fungal sequences comprised ~35% and 40% of Sanger sequences of
393 subway and outdoor rRNA genes, respectively (Fig. 2). These relatively high
394 abundances of fungal rRNA genes compared to bacterial genes cannot be taken as
395 cellular (or spore) frequencies relative to bacteria, however, due to the common

AEM Accepts published online ahead of print

396 occurrence of large numbers of rRNA genes per genome in fungi, typically 100 or more
397 (35)(36). In contrast, bacteria are expected to have only a few rRNA genes (37).
398 Consequently, the abundance of fungi relative to bacteria in cell numbers probably is
399 overestimated in these sequence sets by 10-100 fold. A low frequency of fungus spores
400 observed in direct counts was consistent with that expectation.
401
402 The taxonomic breakdown of the fungal sequences detected was complex and,
403 as with bacterial sequences, few sequences corresponded exactly to a named organism
404 (Fig. 6). All, however, fall into more or less well-described groups including mushrooms
405 (e.g. representatives of *Agaricomycetes*), mildews (e.g. members of *Eurotiales*),
406 saprophytes and plant pathogens (e.g. *Capnodiales*), yeasts (e.g. *Saccharomycetales*)
407 and wood-rotting fungi (most taxa listed in Fig.6). No sequences of known human
408 pathogens were present in the datasets. Although most fungal sequences were
409 comparable between outside and subway samples, approximately 20% of the fungal
410 sequences in the subway samples had no counterparts in the outside libraries (Fig. 6).
411 Wood rot fungi were common among all the sequences detected, so the increased
412 fungal diversity among the subway samples likely is due to the presence of wooden
413 track structure throughout the system.
414
415 **Discussion**
416
417 The subway environment, underground and away from light, might seem remote
418 from our usual environment and potentially occupied by distinct or novel kinds of
419 microorganisms. Rather than unusual however, our survey finds that the microbiota
420 encountered in the NYC subway is fairly mundane, essentially a mix of outdoor air with
421 an overlay of human-associated microorganisms typical of the skin. No significant
422 evidence of pathogens or other organisms of concern was obtained, beyond what might
423 be encountered in any human-occupied indoor setting. Thus, this survey provides the
424 pre-event information necessary for surveillance activities for pathogens that might
425 occur or be introduced into the system. The results also provide pre-event information

AEM Accepts published online ahead of print

426 necessary for interpretation of the microbiological consequences of the recent flooding
427 of the NYC subway system during Hurricane Sandy in 2012.
428
429 Although subways might be considered confined environments, the similarity of
430 subway air microbiota to that of outside air suggests that the subway air significantly
431 equilibrates with outside air on relatively short timescales. There is little or no local air
432 conditioning in the NYC subway. Instead, air movement in the system is driven by
433 passive train-pumping, with air taken in and exhausted through street-level ports, the
434 NYC sidewalk grillwork. The general uniformity of microbial assemblages throughout
435 the system indicates good air mixing, a testimony to the efficiency of the train-pumping
436 process.
437
438 We have no information as to the viability of any of the organisms detected by
439 gene sequences. Indeed, direct cell counts as reported here are ~100-fold higher than
440 reported for colony-forming units in other studies of subway aerosols (14), certainly
441 indicating that most of what we identify is not cultured using standard methods. It is
442 possible, perhaps likely, that many or most aerosolized organisms are non-viable
443 because uncontrolled desiccation of bacteria can be lethal (38)(15)(13). Although
444 bacterial and fungal spores are long-lived in desiccated or partially desiccated
445 conditions, in the case of bacteria we see little evidence for the specific selection of
446 phylotypes known for spore formation. For instance, rRNA sequences representative of
447 the spore-forming *Bacillaceae* family, common soil constituents, are only a minor
448 constituent of the subway assemblage (~11% of total bacteria, Fig. 3). We did not
449 detect any potential bioterror agents, such as the spore-forming organism *Bacillus*
450 *anthracis*. On the other hand, health-related concerns are not restricted to virulent
451 microbes, since allergic and hypersensitive responses to bacteria and fungi do not
452 require viable organisms.
453
454 No standing-crop aerosol microorganisms are known, so the microbes detected
455 in this survey are not expected to be a "community" of interacting organisms. Rather,
456 we consider the aerosol microbiota as an assemblage of microbes derived from

AEM Accepts published online ahead of print

457 microenvironment-associated communities the air has contacted. Although most of the
458 microbiology is of unknown origins beyond vague descriptions such as "soil" or "water,"
459 the human skin is likely to be a significant microenvironmental source of identifiable
460 aerosolized microbes. This identification is possible because the microbial diversity
461 associated with skin has been characterized intensively by culture and molecularly, and
462 the rRNA sequence database of named skin microbes is large (28). Consequently, we
463 are able to use a stringent criterion, 99% sequence identity, for speciation of subway
464 microbes based on sequence. (Most environmental studies use a more relaxed 97%-
465 identity rRNA sequence bin for species-level taxonomic calls, which incorporates a
466 much more diverse collection of organisms than our conservative estimate.) The 99%
467 identity of rRNA gene sequences of many subway microbes with the corresponding
468 sequences of specific human skin commensal microbes is strong evidence for our
469 conclusion that a significant component of subway aerosol microbiology is human-
470 derived. This is not surprising; much of the microbiology of human-occupied indoor
471 environments has long been identified with human skin microbiology (39)(15)(40).
472
473 Overall, approximately 5% of subway aerosol sequences correspond specifically
474 to human skin bacterial sequences (Fig. 5). Although our sampling campaign focused
475 on occupied stations, similar results were seen with aerosols collected at an unused
476 station (CityHall, Fig. 3). This indicates that human-associated microorganisms are
477 dispersed throughout the system. This perhaps is surprising considering the large
478 spatial volume of hundreds of kilometers of subway, even in the face of the
479 considerable human traffic. Although shedding of skin flakes carrying microbes is often
480 invoked as sources of human skin microbes, we suggest that convection driven by
481 body-temperature may also be a major factor. That is, we humans all have body
482 temperatures of ~37°C, but when exposed to ambient air we are surrounded by a lower
483 temperature, typically ~20°C. Thus, we all continuously and actively emit a convective
484 plume of warm air – carrying our skin microbiology selectively (41).
485
486 Each sampling event and DNA library analysis was conducted only once, so we
487 acknowledge that any specific variation in microbial contents between the samples

AEM Accepts published online ahead of print

488 might result from statistical flux or intrinsic sampling variation. Nonetheless, although
489 each of the samples is a unique snapshot, the census results are broadly consistent in
490 most samples and they collectively identify the microbiology likely to be encountered in
491 the NYC subway system. We see no consistent evidence for local pockets of specific
492 diversity. Rather, the diversity observed at different sites seems to reflect random
493 draws on a complex assemblage distributed throughout the system.
494
495 The development of "next-generation" sequencing technology has dramatically
496 changed the depth to which we can explore the natural microbial world and the results
497 have provided entirely new perspective on environmental microbial diversity. Early
498 molecular studies of natural microbial diversity tended to be limited by technology and
499 cost to a few hundred sequences, and described only a limited extent of environmental
500 diversity; now far deeper coverage is readily achieved. Sequence surveys such as the
501 current one and many others have revealed that environmental microbiotas consist not
502 simply of collections of specific microbes, "species," but rather complex collections of
503 more or less closely related phylotypes, with little real demarcation between formal
504 taxonomic grades such as genera and species. The results of high-volume sequencing
505 of environmental microbiotas also begin to capture sequences that occur only rarely, a
506 "rare biosphere" that sometimes contains far more phylogenetic diversity than spanned
507 by sequences of the more abundant organisms in the particular sampling (39). For
508 instance, in the current study only four bacterial phyla were the major contributors to the
509 subway aerosol microbiology. However, inclusion of more rare sequences expands the
510 detection to ~12 phyla (Table S2). Although not abundant, these rare sequences
511 potentially represent significant environmental diversity and contribution to the local
512 pangenome pool.
513

514 **Acknowledgments**

515
516 The authors thank Michael Metz, Mike Gemelli, Charles Burrus and Frank
517 Klimasz of the NYC Metropolitan Transit Authority for their assistance in conducting the
518 sampling campaign. We also thank Kimberly Ross, Piret Koll, and Leah Feazel for

AEM Accepts published online ahead of print

519 assistance in sampling, and Dr. Paula Olsiewski of the Alfred P. Sloan Foundation for
520 encouragement and support. This study was supported by grants from the Alfred P.
521 Sloan Foundation (NRP) and NIH HG005964 (DNF).
522
523 **Author Contributions**
524
525 Conceived and designed the experiments: NRP, LKB
526 Performed the experiments: LKB, KLP
527 Analyzed the data: CER, DNF, JKH, NRP, LKB
528 Wrote the paper: NRP, DNF, CER, JKH, LKB
529
530

531

References

533

534 1. http://www.mta.info/nyct/facts/ridership/index.htm#atGlance_s.
535 2. **Awad A.** Environmental study in subway metro stations in Cairo, Egypt. J. Occup.
536 Health **44**:112–118.
537 3. **Cho JH, Hee Min K, Paik NW**. 2006. Temporal variation of airborne fungi
538 concentrations and related factors in subway stations in Seoul, Korea. Int J Hyg
539 Environ Health **209**:249–255.
540 4. **Jung H-J, Kim B, Malek MA, Koo YS, Jung JH, Son Y-S, Kim J-C, Kim H, Ro C-U**. 2012.
541 Chemical speciation of size-segregated floor dusts and airborne magnetic particles
542 collected at underground subway stations in Seoul, Korea. J. Hazard. Mater. **213-**
543 **214**:331–340.
544 5. **Salma I, Posfai M, Kovacs K, Kuzmann E, Homonnay Z, Posta J**. 2009. Properties
545 and sources of individual particles and some chemical species in the aerosol of a
546 metropolitan underground railway station. Atmospheric Environment **43**:3460–3466.
547 6. **Nieuwenhuijsen M**. 2007. Levels of particulate air pollution, its elemental
548 composition, determinants and health effects in metro systems. Atmospheric
549 Environment **41**:7995–8006.
550 7. **Midander K, Elihn K, Wallén A, Belova L, Karlsson A-KB, Wallinder IO**. 2012.
551 Characterisation of nano- and micron-sized airborne and collected subway particles, a
552 multi-analytical approach. Sci. Total Environ. **427-428**:390–400.
553 8. **Chillrud SN, Epstein D, Ross JM, Sax SN, Pederson D, Spengler JD, Kinney PL**. 2004.
554 Elevated airborne exposures of teenagers to manganese, chromium, and iron from
555 steel dust and New York City's subway system. Environ. Sci. Technol. **38**:732–737.
556 9. **Chillrud SN, Grass D, Ross JM, Coulibaly D, Slavkovich V, Epstein D, Sax SN,**
557 **Pederson D, Johnson D, Spengler JD, Kinney PL, Simpson HJ, Brandt-Rauf P**. 2005.
558 Steel dust in the New York City subway system as a source of manganese, chromium,
559 and iron exposures for transit workers. J Urban Health **82**:33–42.
560 10. **Grass DS, Ross JM, Family F, Barbour J, James Simpson H, Coulibaly D, Hernandez**
561 **J, Chen Y, Slavkovich V, Li Y, Graziano J, Santella RM, Brandt-Rauf P, Chillrud SN**.
562 2010. Airborne particulate metals in the New York City subway: a pilot study to assess
563 the potential for health impacts. Environ. Res. **110**:1–11.
564 11. **Birenzvige A, Eversole J, Seaver M, Francesconi S, Valdes E, Kulaga H**. 2003.
565 Aerosol characteristics in a subway environment. Aerosol Science and Technoogy
566 **37**:210–220.
567 12. **Gilleberg S, Faull JL, Graeme-Cook KA**. 1998. A preliminary survey of aerial
568 biocontaminants at six London Underground stations. International Biodeterioration
569 & Biodegradation **41**:149–152.
570 13. **Seino K, Takano T, Nakamura K, Watanabe M**. 2005. An evidential example of
571 airborne bacteria in a crowded, underground public concourse in Tokyo. Atmospheric
572 Environment **39**:337–341.

AEM Accepts published online ahead of print

AEM Accepts published online ahead of print

573 14. **Hwang S, Yoon C, Ryu K, Paik S, Cho J**. 2010. Assessment of airborne environmental
574 bacteria and related factors in 25 underground railway stations in Seoul, Korea.
575 Atmospheric Environment **44**:1658–1662.
576 15. **Dybwad M, Granum PE, Bruheim P, Blatny JM**. 2012. Characterization of airborne
577 bacteria at an underground subway station. Appl. Environ. Microbiol. **78**:1917–1929.
578 16. **Pace NR**. 1997. A Molecular View of Microbial Diversity and the Biosphere. Science
579 **276**:734 –740.
580 17. **Kesavan JS, Schepers DR**. 2006. Characteristics and sampling efficiencies of Omni
581 3000 aerosol samplers. ECBC-TN-28; Edgewood Chemical Biological Center:
582 Aberdeen, MD.
583 18. **Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR**. 1985. Rapid determination
584 of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. U.S.A.
585 **82**:6955–6959.
586 19. **Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R**. 2008. Error-correcting barcoded
587 primers for pyrosequencing hundreds of samples in multiplex. Nat. Methods **5**:235–
588 237.
589 20. **Harris JK, Sahl JW, Castoe TA, Wagner BD, Pollock DD, Spear JR**. 2010. Comparison
590 of Normalization Methods for Construction of Large, Multiplex Amplicon Pools for
591 Next-Generation Sequencing. Appl Environ Microbiol **76**:3863–3868.
592 21. **Frank DN**. 2008. XplorSeq: a software environment for integrated management and
593 phylogenetic analysis of metagenomic sequence data. BMC Bioinformatics **9**:420.
594 22. **Frank DN**. 2009. BARCRAWL and BARTAB: software tools for the design and
595 implementation of barcoded primers for highly multiplexed DNA sequencing. BMC
596 Bioinformatics **10**:362.
597 23. **Nawrocki EP, Kolbe DL, Eddy SR**. 2009. Infernal 1.0: inference of RNA alignments.
598 Bioinformatics **25**:1335–1337.
599 24. **Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D,**
600 **Tabbaa D, Highlander SK, Sodergren E, Methé B, DeSantis TZ, Petrosino JF,**
601 **Knight R, Birren BW**. 2011. Chimeric 16S rRNA sequence formation and detection in
602 Sanger and 454-pyrosequenced PCR amplicons. Genome Res. **21**:494–504.
603 25. **Li E, Hamm CM, Gulati AS, Sartor RB, Chen H, Wu X, Zhang T, Rohlf FJ, Zhu W, Gu C,**
604 **Robertson CE, Pace NR, Boedeker EC, Harpaz N, Yuan J, Weinstock GM,**
605 **Sodergren E, Frank DN**. 2012. Inflammatory Bowel Diseases Phenotype, C. difficile
606 and NOD2 Genotype Are Associated with Shifts in Human Ileum Associated Microbial
607 Composition. PLoS ONE **7**:e26284.
608 26. **Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO**. 2007.
609 SILVA: a comprehensive online resource for quality checked and aligned ribosomal
610 RNA sequence data compatible with ARB. Nucleic Acids Res. **35**:7188–7196.
611 27. **Magurran AE**. Measuring Biological Diversity. Blackwell Publishing Company, Malden,
612 MA.
613 28. **Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, Bouffard GG,**
614 **Blakesley RW, Murray PR, Green ED, Turner ML, Segre JA**. 2009. Topographical
615 and temporal diversity of the human skin microbiome. Science **324**:1190–1192.
616 29. **Toivola M, Alm S, Reponen T, Kolari S, Nevalainen A**. 2002. Personal exposures and
617 microenvironmental concentrations of particles and bioaerosols. J Environ Monit
618 **4**:166–174.

AEM Accepts published online ahead of print

619 30. **Kujundzic E, Zander DA, Hernandez M, Angenent LT, Henderson DE, Miller SL**.
620 2005. Effects of ceiling-mounted HEPA-UV air filters on airborne bacteria
621 concentrations in an indoor therapy pool building. J Air Waste Manag Assoc **55**:210–
622 218.
623 31. **Fabian M, Miller S, Reponen T, Hernandez M**. 2005. Ambient bioaerosol indices for
624 indoor air quality assessments of flood reclamation. J. Aerosol Sci. **36**:763–783.
625 32. **Kuczynski J, Liu Z, Lozupone C, McDonald D, Fierer N, Knight R**. 2010. Microbial
626 community resemblance methods differ in their ability to detect biologically relevant
627 patterns. Nat. Methods **7**:813–819.
628 33. **Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM,**
629 **Herndl GJ**. 2006. Microbial diversity in the deep sea and the underexplored "rare
630 biosphere". Proc. Natl. Acad. Sci. U.S.A. **103**:12115–12120.
631 34. **Kirk Harris J, Gregory Caporaso J, Walker JJ, Spear JR, Gold NJ, Robertson CE,**
632 **Hugenholtz P, Goodrich J, McDonald D, Knights D, Marshall P, Tufo H, Knight R,**
633 **Pace NR**. 2013. Phylogenetic stratigraphy in the Guerrero Negro hypersaline
634 microbial mat. ISME J **7**:50–60.
635 35. **Ganley ARD, Kobayashi T**. 2007. Highly efficient concerted evolution in the
636 ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome
637 shotgun sequence data. Genome Res. **17**:184–191.
638 36. **Maleszka R, Clark-Walker G**. 1990. Magnification of the rDNA cluster in
639 Kluyveromyces lactis. Molecular and General Genetics **223**:342–344.
640 37. **Lee ZM-P, Bussema C 3rd, Schmidt TM**. 2009. rrnDB: documenting the number of
641 rRNA and tRNA genes in bacteria and archaea. Nucleic Acids Res. **37**:D489–493.
642 38. **Potts M**. 2001. Desiccation tolerance: a simple process? Trends Microbiol. **9**:553–559.
643 39. **Rintala H, Pitkäranta M, Toivola M, Paulin L, Nevalainen A**. 2008. Diversity and
644 seasonal dynamics of bacterial community in indoor environment. BMC Microbiol.
645 **8**:56.
646 40. **Hospodsky D, Qian J, Nazaroff WW, Yamamoto N, Bibby K, Rismani-Yazdi H,**
647 **Peccia J**. 2012. Human Occupancy as a Source of Indoor Airborne Bacteria. PLoS One **7**.
648 41. **Gao N, Niu J**. 2005. CFD study of the thermal environment around a human body: a
649 review. Indoor Built Environment **14**:5–16.
650

651

652 **Figure Legends**

653

654 **Figure 1. Phylum-level distribution of aerosolized microbiota observed at NYC**
655 **subway stations and Union Square Park (outside).**

656

657 The percent abundance is shown for phyla observed among the Sanger and the
658 pyrosequence sequences extracted from impinged air sampled at Union Square Park, a

AEM Accepts published online ahead of print

659 proxy for outdoor air (2A, 2C), and eight subway stations 2C, 2D). 2A, 2B: Sanger data
660 (bacteria, archaea, and eucarya). 2C, 2D: Pyrosequence data (bacteria only).
661
662 **Figure 2. Comparison of species observed (Sobs) for Sanger and Pyrosequence**
663 **datasets.**
664
665 Collectors curves of taxonomic classifications seen with sequence sample size. Sanger
666 sequence values are in black. Pyrosequence values are in red.
667
668 **Figure 3. Abundance of bacterial family-level aerosol taxa in NYC subway**
669 **stations and Union Square Park (outside).**
670
671 Family-level taxonomic categories greater than or equal to 1% abundance in the
672 particular sample are shown ranked in abundance from top to bottom. The sample
673 naming convention is: station name, followed by the station sample number (samples
674 taken at both ends of platforms), followed by the date the sample was obtained.
675 Complete taxonomic classifications are available in Table S1. The percent of each
676 category is indicated by the number shown over the colored squares of the heatmap.
677 Numbers do not add to 100% due to rounding.
678 Abbreviations used:
679 GC: Grand Central Station (456 and 7 Train platforms indicated)
680 GCMezz: Grand Central Station Mezzanine
681 Union14: Union Square and 14th Street Station
682 UnionSqPark: Union Square Park (proxy for out-of-doors).
683
684 **Figure 4. Morisita-Horn diversity comparison of aerosol taxa in NYC subway**
685 **stations and Union Square Park (outside).**
686
687 Family level taxonomic classifications of pyrosequence data were used to compute the
688 Morisita-Horn beta diversity index for all sample pairs. Sample pairs with most overlap
689 in all taxonomic categories are shown as red squares, while sample-pairs with least

690 overlap are shown as white squares. Columns are oriented from most similar (top) to
691 least similar to other sites.
692 Abbreviations used:
693 GC: Grand Central Station (456 and 7 Train platforms indicated)
694 GCMezz: Grand Central Station Mezzanine
695 Union14: Union Square Park and 14th Street Station
696 UnionSqPark: Union Square Park (proxy for out-doors).
697
698 **Figure 5. Comparison of aerosol DNA sequences from NYC subway and Union**
699 **Square Park with DNA sequences obtained from human skin, water, and soil.**
700
701 **5A:** Bacterial pyrosequence data were compared by BLAST to databases containing
702 sequences from the Human Skin Microbiome (HSM) (28) and sequences obtained from
703 ribosomal sequence database Silva 108 with source indicated as either soil or water. A
704 match between aerosol sequence and reference was identified whenever the
705 pyrosequence matched a database sequence with at least 95% overlap and by at least
706 (columns in figure) 95% identity (genus-species level relationship) or at least 99%
707 identity (close species relationship). Sequence matches were sorted into eight (non-
708 mutually exclusive) categories: those that did not match any of the three databases
709 went into “none”, those sequences that matched both human skin and soil sequences
710 were assigned to “Skin, Soil”, etc.
711
712 **5B.** Potential human skin sources of aerosol DNA sequences from NYC subway
713 stations were estimated by comparison of pyrosequence data by BLAST to a set of
714 Human Skin Microbiome databases created by random resampling of skin sites to the
715 size of the smallest skin site sequence set (~3,800 sequences/site). A match between
716 pyrosequence and skin site sequence was called when the pyrosequence had at least
717 95% overlap with the site sequence and 99% identity. The percent of these high
718 stringency matches are shown as a spectrum from red (> 25% of pyrosequences
719 match) to white (0% matches). Skin sites are defined by metadata in HSM GenBank
720 deposited sequences. The percent of skin-related sequences in each category is

AEM Accepts published online ahead of print

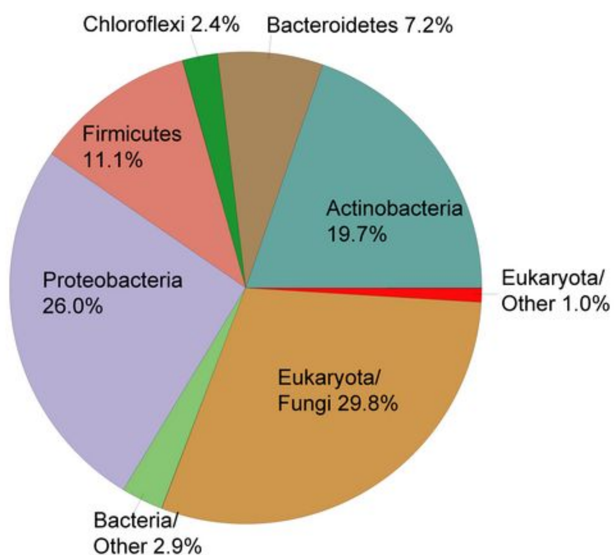
AEM Accepts published online ahead of print

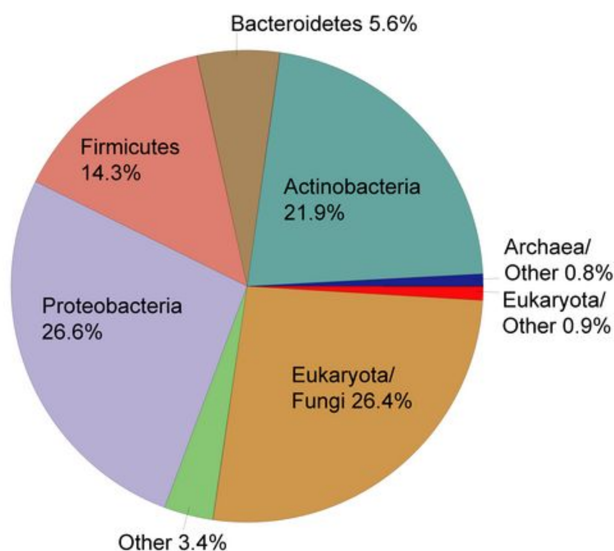
721 indicated by the number imposed on the colored squares. Numbers do not add to
722 100% due to rounding.
723
724 **Figure 6. Comparison of aerosolized fungal DNA sequences from NYC subway**
725 **stations and Union Square Park.**
726
727 Subway data were pooled by averaging the percent abundances of fungal sequences
728 for each taxonomic category shown. The percent of each category is indicated by the
729 number imposed on the colored squares. Numbers do not add to 100% due to rounding.
730
731 **Figure S1. Schematic of sample sites in and associated with the NYC subway**
732 **system.**
733
734 **Figure S2: Human skin sites of aerosol DNA sequences from NYC subway**
735 **stations.**
736
737 Skin sites were estimated by comparison of pyrosequence data by BLAST to a
738 set of Human Skin Microbiome databases created by random resampling of skin sites to
739 the size of the smallest skin site sequence set (~3,800 sequences/site). A match
740 between pyrosequence and skin site sequence was called when the pyrosequence had
741 at least 95% overlap with the site sequence and 99% identity. The percent of these
742 high stringency matches in the different station samples are shown as a spectrum from
743 red (> 25% of pyrosequences match in the sample) to white (0% matches). Skin sites
744 are defined by metadata in HSM GenBank deposited sequences. The percent of each
745 category is indicated by the number imposed on the colored squares. Numbers do not
746 add to 100% due to rounding.
747
748 **Table S1: Metadata associated with samples.**
749
750 **Table S2: Family-level taxonomic categories greater than or equal to 0.01%**
751 **abundance overall shown ranked in abundance from top to bottom.** The sample

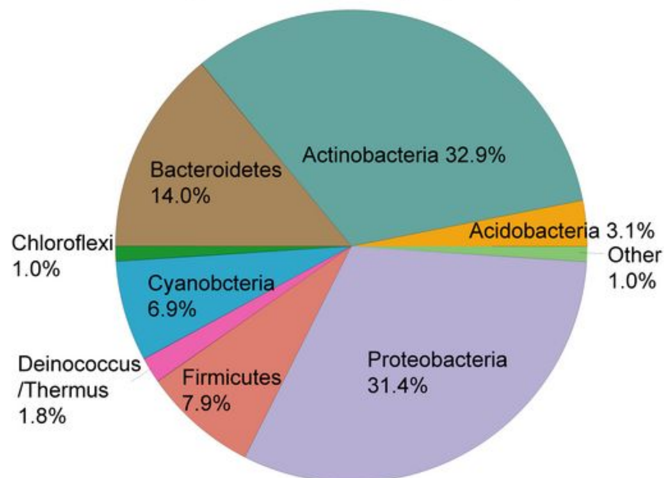
AEM Accepts published online ahead of print

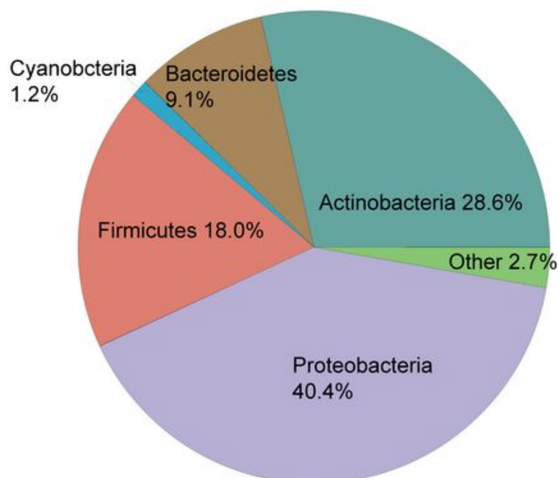
752 naming convention is: station name, followed by the station sample number (samples
753 taken at both ends of platforms), followed by the date the sample was obtained.
754 Numbers do not add to 100% due to rounding.
755

AEM Accepts published online ahead of print

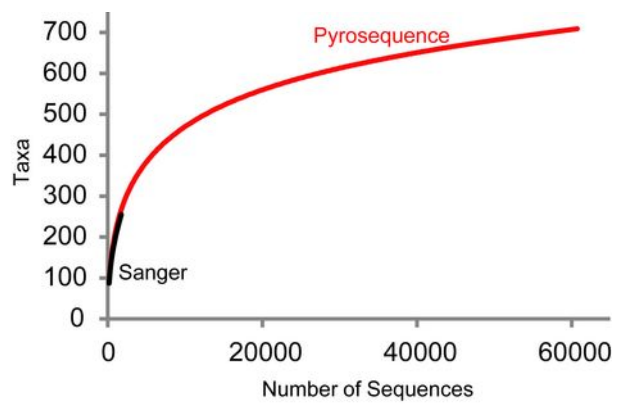




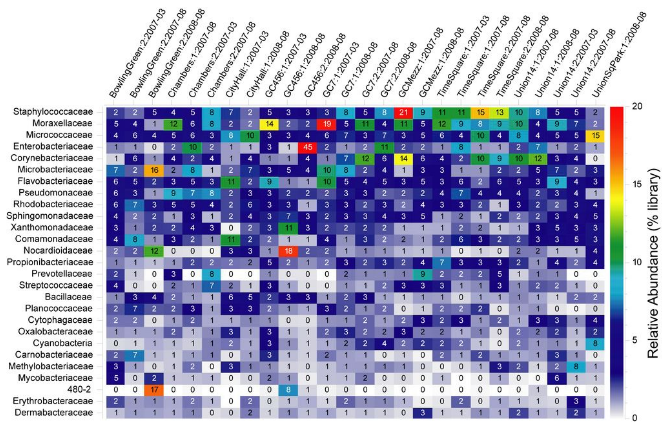




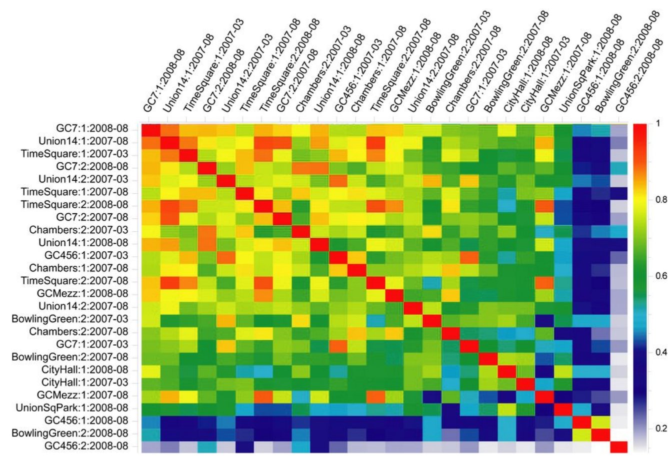
AEM Accepts published online ahead of print



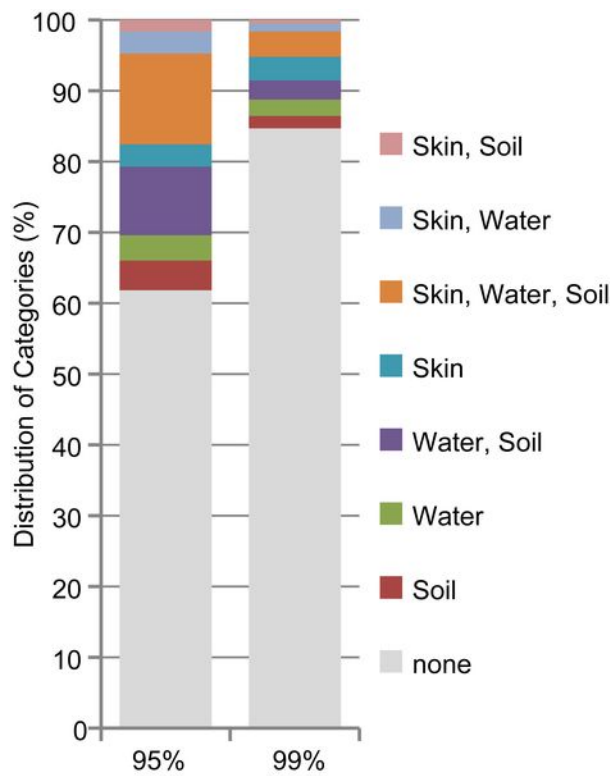
AEM Accepts published online ahead of print



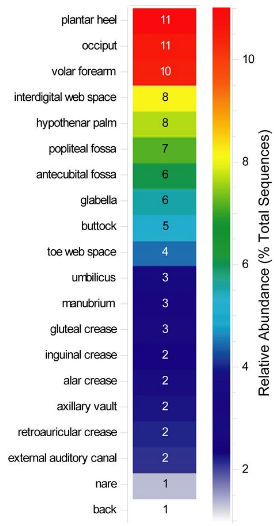
AEM Accepts published online ahead of print



AEM Accepts published online ahead of print



AEM Accepts published online ahead of print



AEM Accepts published online ahead of print