


RESEARCH

Open Access



The subway microbiome: seasonal dynamics and direct comparison of air and surface bacterial communities

Jostein Gohli^{1*} , Kari Oline Bøifot^{1,2}, Line Victoria Moen¹, Paulina Pastuszek³, Gunnar Skogan¹, Klas I. Udekwu⁴ and Marius Dybwad^{1,2}

Abstract

Background: Mass transit environments, such as subways, are uniquely important for transmission of microbes among humans and built environments, and for their ability to spread pathogens and impact large numbers of people. In order to gain a deeper understanding of microbiome dynamics in subways, we must identify variables that affect microbial composition and those microorganisms that are unique to specific habitats.

Methods: We performed high-throughput 16S rRNA gene sequencing of air and surface samples from 16 subway stations in Oslo, Norway, across all four seasons. Distinguishing features across seasons and between air and surface were identified using random forest classification analyses, followed by in-depth diversity analyses.

Results: There were significant differences between the air and surface bacterial communities, and across seasons. Highly abundant groups were generally ubiquitous; however, a large number of taxa with low prevalence and abundance were exclusively present in only one sample matrix or one season. Among the highly abundant families and genera, we found that some were uniquely so in air samples. In surface samples, all highly abundant groups were also well represented in air samples. This is congruent with a pattern observed for the entire dataset, namely that air samples had significantly higher within-sample diversity. We also observed a seasonal pattern: diversity was higher during spring and summer. Temperature had a strong effect on diversity in air but not on surface diversity. Among-sample diversity was also significantly associated with air/surface, season, and temperature.

Conclusions: The results presented here provide the first direct comparison of air and surface bacterial microbiomes, and the first assessment of seasonal variation in subways using culture-independent methods. While there were strong similarities between air and surface and across seasons, we found both diversity and the abundances of certain taxa to differ. This constitutes a significant step towards understanding the composition and dynamics of bacterial communities in subways, a highly important environment in our increasingly urbanized and interconnect world.

Keywords: 16S rRNA gene, Aerosol, Air, Amplicon sequencing, Microbiome, Seasonal variation, Subway

* Correspondence: Jostein.Gohli@ffi.no

¹Norwegian Defence Research Establishment, P.O. Box 25, NO-2027 Kjeller, Norway

Full list of author information is available at the end of the article



Background

Microorganisms are ubiquitous in the truest sense of the word; regardless of where humans reside, they are subjected to a plethora of microbes, some of which have profound impacts on our health [1, 2], and our individual microbiomes [3]. The built environment (BE) is the primary environment of modern humans [4], and hence where we mainly encounter microorganisms. Mass transit environments such as subways facilitate a constant flow of microbes among humans and among different BEs [5]. They are thus particularly important for human health due to their potential for spreading pathogens [6] and impact large numbers of people.

The subway surface microbiome has already been characterized using both 16S rRNA gene amplicon sequencing [5, 7] and shotgun metagenomics [5, 8, 9]. Notably, the MetaSUB project [10] has produced a microbiome and antimicrobial resistance atlas from mass transit surface samples spanning 58 cities across the world [11]. Subway air—which is of particular interest with regard to bioterrorism [12] and infectious diseases [13]—has also been studied using both culture-based [14–17] and more recently culture-independent approaches [18–21]. Robertson et al. [21] described the composition and diversity of subway air microbiomes in New York. Leung et al. [20] found extensive bacterial diversity in the air of the Hong Kong subway system, showing that changes in microbial communities were governed by temperature, humidity, and the number of commuters. Triadó-Margarit et al. [19] investigated air microbiomes in the Barcelona subway system and found significant overlap among different environments within the subway and dominance of a few widespread groups of microorganisms. Fan et al. [18] observed variation in the fungal and bacterial air microbiomes of the Beijing Subway between peak and non-peak commuting hours. While many studies have addressed subway air or surface microbiomes separately, to our knowledge, no study has yet provided a direct comparison of these two important and probably closely interacting sample matrices.

Seasonality is a time-dependent, fundamental shift in environmental conditions that is expected to vary greatly across geographical scales. It is well known that outdoor microbiomes show significant variation across seasons [22–24] and that outdoor air strongly contributes to indoor microbiome composition [25]. Hence, seasonal effects on BE microbiomes are to be expected. However, two BE studies of seasonal microbiome variation, one in Finnish office buildings [26] and another at a children's daycare center in Virginia, USA [27], found no significant seasonal trends. Patel et al. [28] cultivated bacteria from dust collected at railway stations in England and Scotland and found seasonal trends in bacterial

abundance, and Heo et al. [14] found concentrations of culturable bacteria in underground subway air to vary across seasonal transitions; however, culture-independent methods have not been utilized to evaluate seasonal microbial diversity in subways or similar environments.

In the present study, we analyzed surface and air samples collected at 16 subway stations in Oslo, Norway, a relatively small capital city—compared with cities previously studied in this context—at the northern boundary of the temperate region. The aim of this work is to provide a direct comparison of surface and air bacterial microbiomes—to identify unique and ubiquitous taxa and to quantify differences in diversity among these sample matrices. Furthermore, we address an important knowledge gap, namely that of seasonal dynamics in subway air and surface bacterial microbiomes. The main hypotheses tested here are (1) that bacterial microbiome composition and diversity varies significantly across seasons, and (2) that bacterial microbiomes found on surfaces and in air differ with regard to composition and diversity.

Methods

Air (69) and surface (177) samples were collected at 16 subway stations in Oslo, Norway across four seasons from November 2016–June 2017 (Additional file 1: Tables S1 and S2). At each sampling location, one air sample and three surface samples were collected. An Aerotrak 8220 (TSI, Shoreview, MN, US) optical particle counter fitted with an external probe (Model: 1300102) was used to record temperature and humidity.

Air sampling

Air samples were collected and air filters extracted as previously described in Bøifot et al. [29]. Briefly, particulates in air were collected on an electret filter with a SASS3100 air sampler for 30 min, and at 300 L of air per minute (Research International, Monroe, WA, USA). Filters were placed in 50-mL centrifuge tubes and stored in a transport cooler with ice packs before they were transferred to -80°C upon return to the laboratory. Particulates were extracted from the filter with NucliSENS lysis buffer (10 mL, BioMérieux, Marcy-l'Étoile, France), and the filter extract was centrifuged at 7000g for 30 min. The supernatant was transferred to a new 50-mL tube, while the centrifuged pellet was resuspended with PowerBead Solution (550 μL , Qiagen, Hilden, Germany) and transferred to autoclaved (121°C , 45 min) bead tubes (2 mL, Sarstedt, Nümbrecht, Germany) filled with 600 mg, 0.1-mm zirconia/silica beads (BioSpec Products, Bartlesville, OK, USA). PowerSoil Solution C1 (60 μL , Qiagen) was added and bead beating (1 min, maximum intensity) was performed in a Mini-Beadbeater-8 (BioSpec Products). Bead tubes were centrifuged at 13,000g

for 2 min and potential inhibitors removed according to the DNeasy PowerSoil Kit with Solution C2 (250 μ L) and C3 (200 μ L). Following inhibitor removal, we combined the supernatant with that from the filter extract and isolated DNA according to the manufacturer's protocol of the NucliSENS Magnetic Extraction Reagents Kit (BioMérieux). Negative controls (reagents) were prepared by processing blank samples (unexposed filters) along with the air samples.

Surface sampling

Surface samples were collected for three surface types at each station: kiosks, railings, and benches. A nylon, flocked swab (Copan eSwab 490 CE.A, Copan Diagnostics, CA, USA) wetted in Amies liquid medium, was used to swab the surface for 3 min, covering an area as large as possible. The swab was placed in a 15-mL centrifuge tube and stored in a transport cooler with ice packs before they were transferred to -80°C upon return to the laboratory. DNA was isolated according to the DNeasy PowerSoil Kit (Qiagen) protocol, except that the standard PowerBead Tubes were replaced with the customized bead tubes described above for air samples. Swabs were cut with sterilized scissors into bead tubes filled with PowerBead Solution (550 μ L, Qiagen) and Solution C1 (60 μ L, Qiagen) before tubes were bead beaten (1 min, maximum intensity) in a Mini-Beadbeater-8 (BioSpec Products). Negative controls (reagents) were prepared by cutting clean swabs into bead tubes and performing DNA isolation.

ZymoBIOMICS Microbial Community Standard (10 μ L, Zymo Research) was added to one bead tube before DNA isolation (isolated according the protocol described above), which served as a positive control.

Quantification of total DNA and bacterial 16S rRNA gene copies

DNA yield was measured with Qubit 3.0 Fluorimeter (Life Technologies, Carlsbad, CA, USA) using the Qubit dsDNA HS assay (Life Technologies). Bacterial 16S rRNA gene copy yield was determined with a qPCR assay performed according to Liu et al. [22] on a LightCycler 480 (Roche Diagnostics, Oslo, Norway). A standard curve was generated with serial dilutions of *Escherichia coli* DNA (seven 16S rRNA gene copies per genome). Bacterial 16S rRNA gene copy yields were analyzed with linear models in R [30]. Given that air and surface samples were collected with different sampling protocols, the data was grouped by air and surface prior to analysis. Surface type, surface material, surface treatment, season, indoor/outdoor station, time of day, temperature (mean and standard deviation), humidity (mean and standard deviation), and sequence run were included as predictors in these models, which were subsequently subjected

to a stepwise model (predictor variable) selection with the stepAIC R function [31].

16S rRNA gene amplicon sequencing

The 16S rRNA gene was amplified by PCR (Additional file 1: Table S3), using forward primer 341F, 5'-CCTACGGGNGGCWGCAG-3' and reverse primer 805R, 5'-GGACTACHVGGGTWTCTAAT-3', targeting the V3 and V4 regions of the gene. Sequence libraries were prepared following the 16S Metagenomic Sequencing Library Preparation protocol [32] and sequenced on Illumina MiSeq in four separate runs. Four swab negative controls, three air negative controls, and one ZymoBIOMICS Microbial Community Standard positive control were included as study controls.

Sequence analysis

Primers and adapters were removed from demultiplexed sequence reads using TrimGalore [33], a perl wrapper for Cutadapt [34] and FastQC [35]. A big data pipeline, i.e., forward reads only, was used to infer amplicon sequence variants (ASV) using the dada2 R package [36]. Filtering was performed with the filterAndTrim function in dada2; reads that mapped to the phiX genome were removed, all reads were truncated to 250 bp, and reads of < 250 bp, that contained any unassigned bases or bases of quality score < 2 , were discarded, and the maximum number of expected read errors per read was set to 2. Learned error rates were used for inferring ASV before removing chimeras (dada2 functionality). Dada2 analyses were run separately for the four sequence runs, before merging the feature tables. SILVA SSU v.132 [37], which is the largest dedicated 16S taxonomy database, was used for assigning taxonomy to the ASVs. The ASV table, taxonomy table, and metadata were imported into the phyloseq R package [30] for analyses.

Reads not assigned to the phylum level were removed before rarefaction. All samples were rarified to the lowest read depth after assessing rarefaction curves with observed diversity and Shannon's diversity index. The data set was split into air and surface samples, and into surface types, before summarizing the most abundant phyla, families, and genera in both subsets.

Three random forest classification analyses were performed with 10,001 trees, using air/surface, the four seasons, and surface type as classification bins. ASVs not assigned to genus were discarded before conglomerating all ASVs to the level of genus. A prevalence filter of < 10 and a total abundance filter of < 20 were implemented prior to calculating Z-scores from abundances for the remaining 817 genera. The most important features (genera) for correctly assigning samples to their correct bin (air/surface, season, or surface type) was identified using mean decrease in model accuracy (MDA), i.e., the

negative impact on model accuracy by excluding a feature.

Shannon's diversity index scores were analyzed using linear models in R [31]. Firstly, variables only relevant for surface samples (surface type, material, and treatment (painted/not painted)) were analyzed. Second, air/surface, season, indoor/outdoor station, time of day, temperature (mean and standard deviance), humidity (mean and standard deviance), and sequence run, along with all possible two-way interactions, were included as predictors of Shannon's diversity index scores in a separate model. This model was subjected to a stepwise model selection with the stepAIC R function [38].

Prior to analyses of among-sample diversity, ASVs with prevalence < 5 were removed. A Bray–Curtis dissimilarity matrix was ordinated using PCoA, and PERMANOVA tests were performed using the same predictor variables (including all two-way interactions) mentioned above. Manual AIC model selection was performed by dropping the least significant variable in a step-wise fashion, until further removals no longer improved the model's AIC score.

Results

All negative controls (four swabs and three air samples) failed to generate sequenceable libraries in the library preparation step due to insufficient DNA yields. The positive controls showed no sign of contamination and yielded the correct genera. Analyses of 16S rRNA gene copy yields found that bacterial numbers decreased with increasing humidity, peaked during spring for air samples (Additional file 1: Table S4; Figure S1), and were highest during summer, at outdoor stations, and on kiosks for surface samples (Additional file 1: Table S4; Figure S2). For surface samples, the number of 16S rRNA gene copies was also significantly higher in one of the sequence runs (Additional file 1: Table S4; Figure S2).

After QC filtering, 41 M forward reads remained (Additional file 1: Figure S3). A total of 12.6% were lost in the ASV inference step (dada2, with error modeling), and a further 15.1% were removed as chimeras, leaving 30.7 M forward reads. From this material, dada2 identified 328, 615 ASVs. A total of 13,788 of these were not assigned to phylum and removed. Rarefaction curves for observed diversity and Shannon's diversity index (Additional file 1: Figure S4) were evaluated before rarefying all samples to a common read depth of 6358, which removed only three samples.

Taxonomy and community composition

In both air and surface samples, the phyla *Actinobacteria* and *Proteobacteria* dominated, with abundances of 42.9% and 23.9%, and 31.3% and 27.5% respectively (Table 1; Fig. 1a). The top 20 phyla were the same in

both air and surface samples, with the top five also showing identical ordering by abundance. At the family level, *Micrococcaceae* was most abundant in both air and surface samples (10.5% and 7.2% respectively; Table 1). Notably, *Rubrobacteriaceae* and *Pseudonocardiaceae*, who were highly abundant in air samples (ranking as the 5th (3.5%) and 12th (2.1%) most abundant), were not found among the surface sample top 20 families, ranking as the 53rd (0.3%) and 48th (0.4%). The two unique families in the surface top 20 (*Lactobacillaceae*, *Deinococcaceae*) were both in the air top 25. At the genus level, similarities between the air and surface top 20 were still pronounced (Table 1). Most notably, *Staphylococcus* was the 2nd most abundant in air (3.8%) and most abundant in surface samples (4.7%). In line with the theme from the family level results, *Rubrobacter* (*Rubrobacteriaceae*), which ranked as the third most abundant in air (3.5%), was the 49th (0.3 %) most abundant in the surface samples. Of the other genera that appeared exclusively in the air top 20, only *Pseudonocardia* (1.1%) and *Nesterenkonia* (0.9%) had a substantially lower abundance ranking in surface (76th (0.2%) and 68th (0.2 %), respectively). Of the five genera that were exclusively in the surface top 20, only *Pseudomonas* ranked outside the air top 30 (37th).

Abundance plots of phyla across seasons, and air and surface (Fig. 2a), showed a relatively stable distribution; however, *Firmicutes* exhibited higher relative abundance in winter and spring, while *Cyanobacteria* appeared to be more abundant during summer and autumn. As also seen in the top 20 abundance table (Table 1), *Actinobacteria* had a higher relative abundance in air samples, and *Proteobacteria* was more abundant in surface samples. We observed notable seasonal differentiation in three *Verrucomicrobia* families (*Verrucomicrobiaceae*, *Rubritaleaceae*, and *Chthoniobacteraceae*; Fig. 1b), who were all most abundant during summer, especially in surface samples. *Streptomycetaceae*, *Pseudonocardiaceae*, *Rubrobacteriaceae*, and *Halococcaceae* all showed higher relative abundance in air samples with no strong seasonal patterns (Fig. 1b).

The comparison of highly abundant taxa in surface samples taken from kiosks, benches, and railings revealed a high degree of similarity across surface types (Additional file 1: Table S5; Figure S5).

Indicator genera: random forest classification

Random forest classification analyses, using genera as classification features, showed a high level of success in assigning samples to their correct bins (air or surface and correct season). The season classification had an out-of-bag error rate of 8.9%, with the highest class error found for summer samples, where nine samples were incorrectly classified as autumn samples (Table 2).

Table 1 Top 20 phyla, families, and genera in air samples ($N = 69$) and surface samples ($N = 177$). Dots indicate that a group is also represented in the top 20 set from the other sample matrix. Prevalence is the sum of prevalence of all ASVs within a taxonomic group

Air samples				Surface samples			
Phylum		Prevalence	Abundance (%)	Phylum		Prevalence	Abundance (%)
<i>Actinobacteria</i>	•	32638	42.92	<i>Actinobacteria</i>	•	59693	31.32
<i>Proteobacteria</i>	•	27216	23.88	<i>Proteobacteria</i>	•	61648	27.47
<i>Firmicutes</i>	•	12494	11.97	<i>Firmicutes</i>	•	28790	16.65
<i>Bacteroidetes</i>	•	13390	8.10	<i>Bacteroidetes</i>	•	34972	10.99
<i>Cyanobacteria</i>	•	3988	6.29	<i>Cyanobacteria</i>	•	10775	7.88
<i>Chloroflexi</i>	•	2731	1.53	<i>Deinococcus-Thermus</i>	•	3184	1.37
<i>Acidobacteria</i>	•	2451	1.10	<i>Acidobacteria</i>	•	4904	1.05
<i>Deinococcus-Thermus</i>	•	1445	1.06	<i>Chloroflexi</i>	•	3591	0.67
<i>Planctomycetes</i>	•	1976	0.68	<i>Fusobacteria</i>	•	1557	0.59
<i>Euryarchaeota</i>	•	294	0.66	<i>Planctomycetes</i>	•	3321	0.51
<i>Gemmatimonadetes</i>	•	1335	0.53	<i>Verrucomicrobia</i>	•	2281	0.37
<i>Patescibacteria</i>	•	998	0.34	<i>Patescibacteria</i>	•	2086	0.33
<i>Verrucomicrobia</i>	•	910	0.33	<i>Gemmatimonadetes</i>	•	1667	0.26
<i>FBP</i>	•	410	0.17	<i>FBP</i>	•	958	0.18
<i>Armatimonadetes</i>	•	381	0.15	<i>Armatimonadetes</i>	•	773	0.12
<i>Fusobacteria</i>	•	250	0.10	<i>Euryarchaeota</i>	•	162	0.11
<i>Chlamydiae</i>	•	185	0.06	<i>Epsilonbacteraeota</i>	•	155	0.03
<i>Epsilonbacteraeota</i>	•	65	0.03	<i>Chlamydiae</i>	•	172	0.02
<i>Nitrospirae</i>	•	56	0.02	<i>Spirochaetes</i>	•	111	0.02
<i>Spirochaetes</i>	•	28	0.01	<i>Nitrospirae</i>	•	69	0.01
Family		Prevalence	Abundance (%)	Family		Prevalence	Abundance (%)
<i>Unassigned</i>	•	19657	15.01	<i>Unassigned</i>	•	42005	14.23
<i>Micrococcaceae</i>	•	3092	10.49	<i>Micrococcaceae</i>	•	5787	7.21
<i>Sphingomonadaceae</i>	•	3354	4.78	<i>Sphingomonadaceae</i>	•	8367	5.84
<i>Staphylococcaceae</i>	•	1965	4.26	<i>Staphylococcaceae</i>	•	4280	5.00
<i>Burkholderiaceae</i>	•	5034	4.06	<i>Streptococcaceae</i>	•	3810	4.53
<i>Rubrobacteriaceae</i>		1143	3.52	<i>Burkholderiaceae</i>	•	10578	4.42
<i>Hymenobacteraceae</i>	•	4187	3.35	<i>Hymenobacteraceae</i>	•	9178	3.84
<i>Nocardioideaceae</i>	•	2899	3.24	<i>Moraxellaceae</i>	•	3982	3.35
<i>Moraxellaceae</i>	•	1689	2.96	<i>Corynebacteriaceae</i>	•	4365	3.34
<i>Acetobacteraceae</i>	•	2555	2.60	<i>Acetobacteraceae</i>	•	5957	2.64
<i>Corynebacteriaceae</i>	•	1578	2.49	<i>Cardioideaceae</i>	•	4933	2.29
<i>Intrasporangiaceae</i>	•	1498	2.30	<i>Propionibacteriaceae</i>	•	2947	2.20
<i>Pseudonocardiaceae</i>		1755	2.14	<i>Beijerinckiaceae</i>	•	3413	2.06
<i>Beijerinckiaceae</i>	•	1558	2.07	<i>Microbacteriaceae</i>	•	3654	1.93
<i>Geodermatophilaceae</i>	•	969	2.06	<i>Flavobacteriaceae</i>	•	4970	1.89
<i>Microbacteriaceae</i>	•	1604	1.80	<i>Intrasporangiaceae</i>	•	2498	1.52
<i>Propionibacteriaceae</i>	•	1224	1.71	<i>Geodermatophilaceae</i>	•	1747	1.46
<i>Streptococcaceae</i>	•	955	1.43	<i>Rhodobacteraceae</i>	•	3227	1.28
<i>Rhodobacteraceae</i>	•	1543	1.27	<i>Lactobacillaceae</i>		1474	1.16
<i>Flavobacteriaceae</i>	•	1476	1.13	<i>Deinococcaceae</i>		2604	1.15

Table 1 Top 20 phyla, families, and genera in air samples ($N = 69$) and surface samples ($N = 177$). Dots indicate that a group is also represented in the top 20 set from the other sample matrix. Prevalence is the sum of prevalence of all ASVs within a taxonomic group (Continued)

Air samples				Surface samples			
Genus		Prevalence	Abundance (%)	Genus		Prevalence	Abundance (%)
<i>Unassigned</i>	•	38724	26.51	<i>Unassigned</i>	•	82052	24.72
<i>Micrococcus</i>	•	377	3.97	<i>Staphylococcus</i>	•	3208	4.57
<i>Staphylococcus</i>	•	1445	3.83	<i>Sphingomonas</i>	•	3184	4.23
<i>Rubrobacter</i>		1143	3.52	<i>Streptococcus</i>	•	2430	3.96
<i>Sphingomonas</i>	•	1260	3.37	<i>Hymenobacter</i>	•	8696	3.75
<i>Hymenobacter</i>	•	3850	3.19	<i>Corynebacterium</i>	•	3278	2.70
<i>Arthrobacter</i>	•	999	2.99	<i>Arthrobacter</i>	•	1904	2.07
<i>Corynebacterium</i>	•	1248	2.13	<i>Kocuria</i>	•	915	1.99
<i>Nocardioiodes</i>	•	1867	2.06	<i>Micrococcus</i>	•	594	1.82
<i>Psychrobacter</i>	•	453	1.49	<i>Psychrobacter</i>	•	1106	1.65
<i>Blastococcus</i>		494	1.38	<i>Flavobacterium</i>	•	3355	1.54
<i>Kocuria</i>	•	383	1.26	<i>Nocardioiodes</i>	•	2985	1.43
<i>Streptococcus</i>	•	658	1.22	<i>Cutibacterium</i>		523	1.23
<i>Pseudonocardia</i>		753	1.10	<i>Lactobacillus</i>	•	1459	1.16
<i>Nesterenkonia</i>		378	0.92	<i>Deinococcus</i>	•	2528	1.14
<i>Flavobacterium</i>	•	907	0.92	<i>Massilia</i>		1738	1.01
<i>Methylobacterium</i>		315	0.84	<i>Pseudomonas</i>		1545	0.92
<i>Deinococcus</i>	•	1025	0.80	<i>Acinetobacter</i>		1596	0.84
<i>Lactobacillus</i>	•	629	0.80	<i>1174-901-12</i>		1076	0.83
<i>Acidiphilium</i>	•	593	0.80	<i>Acidiphilium</i>	•	1432	0.78

Psychrobacter was the most important genus for correct season classification (MDA = 0.043, Table 2; Additional file 1: Figure S6).

The classification of samples as either air or surface had an out-of-bag error rate of 6.1%. The model was highly successful in correctly classifying surface samples, with only two samples being wrongly assigned as air samples (class error = 1.1%). Air samples, on the other hand, had a substantial class error (18.8%), with 13 of 69 air samples being classified as surface samples (Table 3). *Ralstonia* was the most important genus for correct classification (MDA = 0.015; Table 3; Additional file 1: Figure S7).

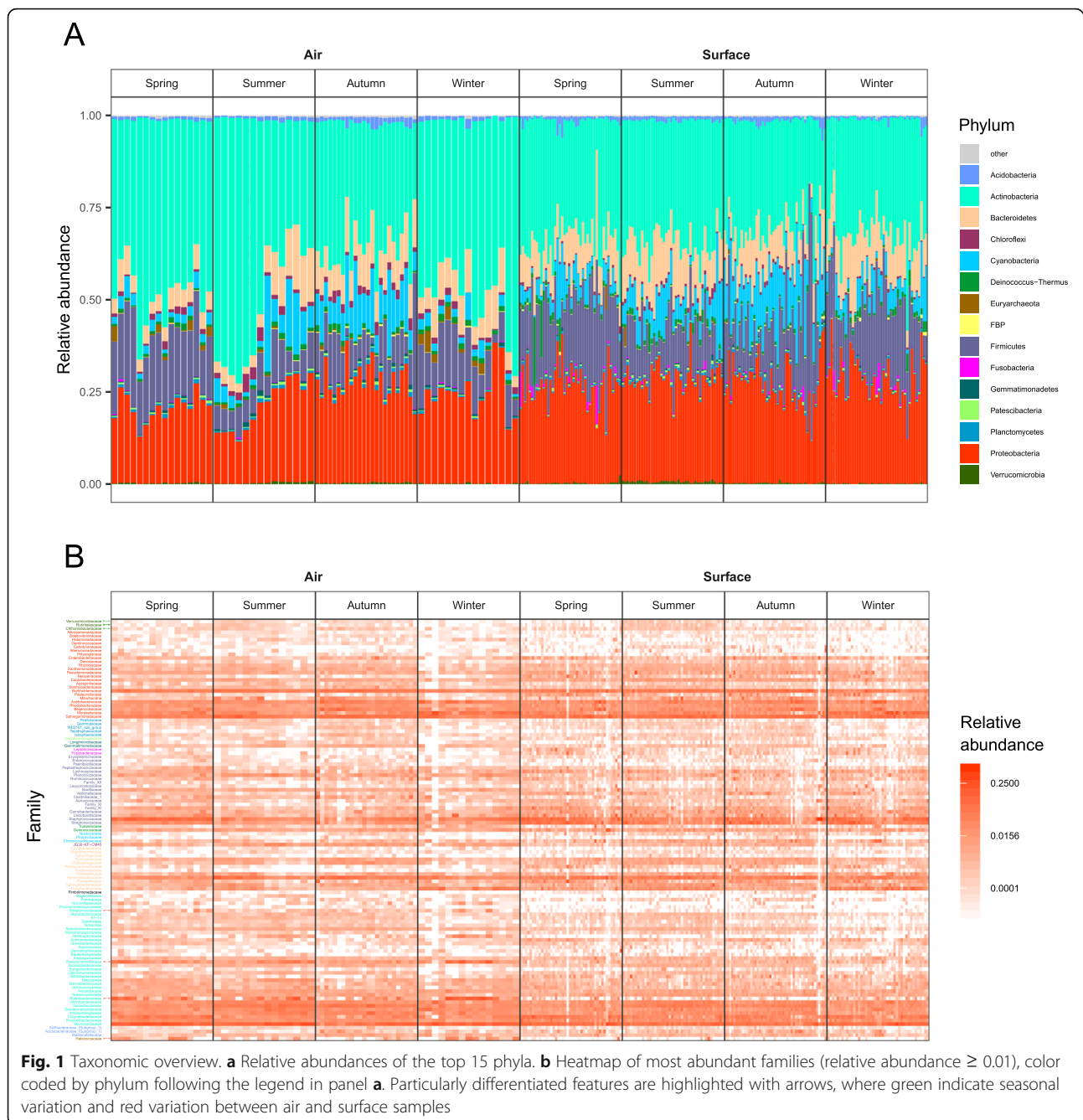
The classification analysis of samples by surface type had a substantially higher out-of-bag error rate (42.37%; Additional file 1: Table S6), with a large number of samples being misclassified for all surface types. The error was particularly pronounced for railing samples, where 41 out of 56 samples were not binned correctly (class error of 73.2%).

Diversity

When assessing all ASVs without prevalence or abundance filtering, we found the majority to be exclusive to

one season, and either surface or air (Fig. 2a). On the other hand, ASVs that had prevalence > 4 and abundance > 10 were largely present across all seasons and in both surface and air samples (Fig. 2b).

For Shannon's diversity index scores, the models that assessed variables specific to surface (surface type, material, and treatment) were all non-significant (all $p > 0.16$). The step-wise AIC model selection scheme on a model with the remaining predictors—air/surface, season, indoor/outdoor station, time of day, temperature, humidity, sequence run, and all possible two-way interactions—returned a model which contained four significant predictors (temperature, $p < 0.001$; air/surface, $p = 0.005$; season, $p < 0.001$; and humidity, $p = 0.017$; Fig. 3) and four significant interactions, which together explained 27% of the variance in Shannon's diversity index scores and had an overall p value of 1.04×10^{-09} (Table 4). Diversity was higher during spring and summer, in air samples, and at higher temperatures and lower levels of humidity (Fig. 3). Of the significant interaction effects, temperature: air/surface ($p = 0.002$) was most notable; closer inspection indicated that surface samples had higher diversity than air samples at low temperatures, and lower diversity at higher temperatures (Additional file 1: Figure S8).



In a multivariate PERMANOVA model of among-sample diversity (ordinated Bray–Curtis dissimilarity) with predictors specific to surface, we found only surface type to be significant ($F = 2.03$, $R^2 = 0.02$, $p = 0.001$; Additional file 1: Figure S9). A PERMANOVA model with the remaining predictors (air/surface, season, indoor/outdoor station, time of day, temperature, humidity, sequence run, and all possible two-way interactions) was subjected to a step-wise AIC model selection scheme, which produced a model that explained 56% of

among-sample diversity. This model included six predictors, and three two-way interactions (Table 5). Whether samples were taken from air or surface was a highly significant predictor ($p = 0.001$), explaining 4% of the total variance. Season ($p = 0.001$) and subway station ($p = 0.001$) explained 11%, and 15% of the variance respectively. Sequence run was also a significant predictor of among-sample diversity ($p = 0.001$) and explained 2% of the variance. Ordination plots revealed clear differentiation for air/surface, season, and sequence run (Fig. 4);

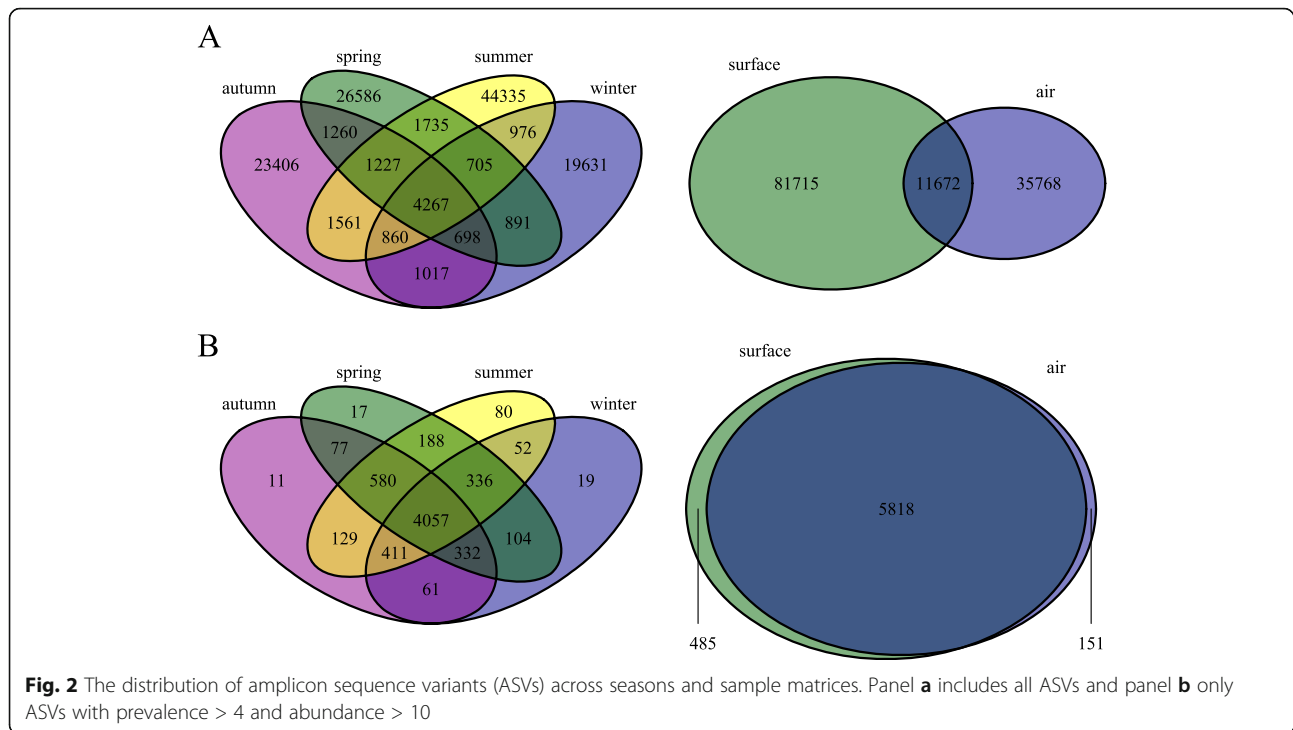


Fig. 2 The distribution of amplicon sequence variants (ASVs) across seasons and sample matrices. Panel **a** includes all ASVs and panel **b** only ASVs with prevalence > 4 and abundance > 10

however, subway stations, which explained the largest amount of variance, showed no discernable clustering.

Discussion

Mass transit systems are BEs critical to the everyday lives of a vast number of people, and its potential role in the transmission of infectious diseases as well as bioterrorism risk cannot be understated [6, 12, 13]. With the advent of molecular assay techniques—more recently high-throughput sequencing—we are no longer restricted to culture-based techniques and can better unravel microbial diversity in mass transit systems. Characterization of microbial diversity in such environments is vital to understand the dynamics of antimicrobial resistance and enables the detection and monitoring of potential pathogens and bioterrorism threat agents. Furthermore, it is essential for the understanding of how our own microbiomes interact with the microbiomes that surround us, and how this ultimately may affect our health and wellbeing [3, 9]. A vital step in this effort is to explore the variability of mass transit microbiomes across sample matrices and temporal scales, and identify important drivers of such variation. In this study, we described the biological background in both air and surfaces from 16 subway stations in Oslo, Norway—a smaller and more northerly city compared with other cities where subway microbiomes have previously been mapped. We provide a direct comparison of surface and air communities, and an assessment of seasonal variation in subway microbiomes.

Taxonomy, relative abundances, and ecology

In the entire dataset, over 300 K unique ASVs were identified. This is substantially higher than comparable studies [5, 7, 20, 21]; however, direct comparisons of studies are not feasible since differences in sampling and wet lab protocols, and sequencing depths may strongly influence results. Further, the use of different taxonomic classifiers with different sensitivities will have substantial effects on the number of OTUs/ASVs reported [39].

The top five most abundant phyla in both surface and air samples (Table 1; Fig. 1a) matched the top five phyla in the Mexico City subway (station and train surfaces) [7] perfectly, the only difference being their ordering by relative abundance. Further, three genera in the top five overlapped between our surface samples and the Mexico City study: *Staphylococcus*, *Streptococcus*, *Corynebacterium*. Major phyla identified in the subway studies from Hong Kong subway [20] and New York [21] were also the same as those identified in the present study.

We found that many less abundant ASVs were unique to specific seasons or sample matrices, while abundant groups were, for the most part, ubiquitous across seasons, and surface and air samples; importantly, a very low filtering cutoff was sufficient to remove almost all taxa only present in single seasons or a single sample matrix (Fig. 2). In surface and air samples, the top 20 most abundant phyla were the same and ordered identically. Two families were highly abundant in air but not in surface samples: *Rubrobacteriaceae* and *Pseudonocardiaceae* with relative abundances of 3.52% and 2.14% in

Table 2 Random forest classification models of season

Season					
Out-of-bag estimate of error rate: 8.94%					
Confusion matrix	Autumn	Spring	Summer	Winter	Class error (%)
Autumn	53	1	5	1	11.7
Spring	0	54	1	1	3.6
Summer	9	0	63	0	12.5
Winter	2	2	0	54	6.9
Most important genera in sample classification					
Family:genera	Autumn	Spring	Summer	Winter	MDA
<i>Moraxellaceae: Psychrobacter</i>	0.026	0.039	0.040	0.069	0.043
<i>Microbacteriaceae: Cryobacterium</i>	0.022	0.057	0.011	0.004	0.022
<i>Flavobacteriaceae: Flavobacterium</i>	0.009	0.012	0.021	0.040	0.020
<i>Nocardioideae: Nocardioidea</i>	0.020	0.024	0.011	0.013	0.016
<i>Flavobacteriaceae: Gillisia</i>	0.024	0.003	0.002	0.014	0.010
<i>Chitinophagaceae: Ferruginibacter</i>	0.009	0.012	0.007	0.009	0.009
<i>Gaiellaceae: Gaiella</i>	0.000	0.011	0.002	0.018	0.008
<i>Ilumatobacteraceae: CL500-29_marine_group</i>	0.001	0.015	0.003	0.012	0.007
<i>Burkholderiaceae: Polaromonas</i>	0.008	0.015	0.001	0.002	0.006
<i>Rubritaleae: Luteolibacter</i>	0.000	0.018	0.003	0.004	0.006
<i>Sphingomonadaceae: Qipengyuania</i>	0.001	0.001	0.009	0.010	0.005
<i>Clostridiaceae_1: Clostridium_sensu_stricto_13</i>	0.000	0.016	0.001	0.006	0.005
<i>Xanthomonadaceae: Thermomonas</i>	0.001	0.006	0.003	0.012	0.005
<i>Chthoniobacteraceae: Candidatus_Udaebacter</i>	0.001	0.009	0.002	0.009	0.005
<i>Staphylococcaceae: Staphylococcus</i>	0.012	0.007	0.002	0.000	0.005
<i>Microbacteriaceae: Galbitalea</i>	0.003	0.017	0.000	0.001	0.005
<i>Pseudoalteromonadaceae: Pseudoalteromonas</i>	0.003	0.000	0.010	0.004	0.005
<i>Phormidiaceae: Tychonema_CCAP_1459: 11B</i>	0.001	0.000	0.006	0.011	0.004
<i>Ilumatobacteraceae: Ilumatobacter</i>	0.001	0.006	0.002	0.009	0.004
<i>Demequinaceae: Demequina</i>	0.000	0.010	0.002	0.004	0.004

Confusion matrices show the classification of samples and the associated class error. The mean decrease in model accuracy (MDA; from removing the genus in question) and mean Z-scores are given for the 20 most important genera for classifying samples

Table 3 Random forest classification models of air/surface

Air/surface			
Out-of-bag estimate of error rate: 6.1%			
Confusion matrix	Air	Surface	Class error (%)
Air	56	13	18.8
Surface	2	175	1.1
Most important genera in sample classification			
Family:Genera	Air	Surface	MDA
<i>Burkholderiaceae: Ralstonia</i>	0.027	0.010	0.015
<i>Streptomycetaceae: Streptomyces</i>	0.020	0.006	0.010
<i>Pseudonocardiaceae: Pseudonocardia</i>	0.018	0.006	0.009
<i>Streptococcaceae: Streptococcus</i>	0.015	0.004	0.007
<i>Pseudonocardiaceae: Saccharopolyspora</i>	0.012	0.004	0.006
<i>Neisseriaceae: Neisseria</i>	0.013	0.003	0.006
<i>Nocardiopsaceae: Nocardioopsis</i>	0.011	0.004	0.006
<i>Rubrobacteriaceae: Rubrobacter</i>	0.011	0.004	0.006
<i>Micrococcaceae: Micrococcus</i>	0.008	0.004	0.005
<i>Carnobacteriaceae: Granulicatella</i>	0.012	0.003	0.005
<i>Pasteurellaceae: Haemophilus</i>	0.012	0.002	0.005
<i>Peptostreptococcaceae: Terrisporobacter</i>	0.008	0.004	0.005
<i>Micrococcaceae: Pseudarthrobacter</i>	0.009	0.003	0.005
<i>Planococcaceae: Planomicrobium</i>	0.009	0.003	0.005
<i>Halococcaceae: Halococcus</i>	0.007	0.003	0.004
<i>Micrococcaceae: Rothia</i>	0.008	0.002	0.004
<i>Halococcaceae: Halalkalicoccus</i>	0.007	0.002	0.003
<i>Porphyromonadaceae: Porphyromonas</i>	0.007	0.002	0.003
<i>Planococcaceae: Planococcus</i>	0.007	0.002	0.003
<i>Pseudonocardiaceae: Actinomycetospora</i>	0.006	0.002	0.003

Confusion matrices show the classification of samples and the associated class error. The mean decrease in model accuracy (MDA; from removing the genus in question) and mean Z-scores are given for the 20 most important genera for classifying samples

air samples. Our findings were similar at the genus level: the radiotolerant [40] *Rubrobacter* (which constituted the *Rubrobacteriaceae* contribution in its entirety; 3.52%) had a uniquely high abundance in air, along with *Pseudonocardia* (most notably producers of antibiotics for use in pest control by fungus farming ants [41]; 1.10%), and *Nesterenkonia* (ubiquitous, and in extreme environments, opportunistically pathogenic [42] 0.92%). We have no explanation for why these particular bacteria were so abundant in air, but not on surfaces. One might expect that all bacteria in air eventually settle on surfaces; however, the chemical and biological properties, and size of bacteria [43], along with environmental variables in air, can affect both deposition and

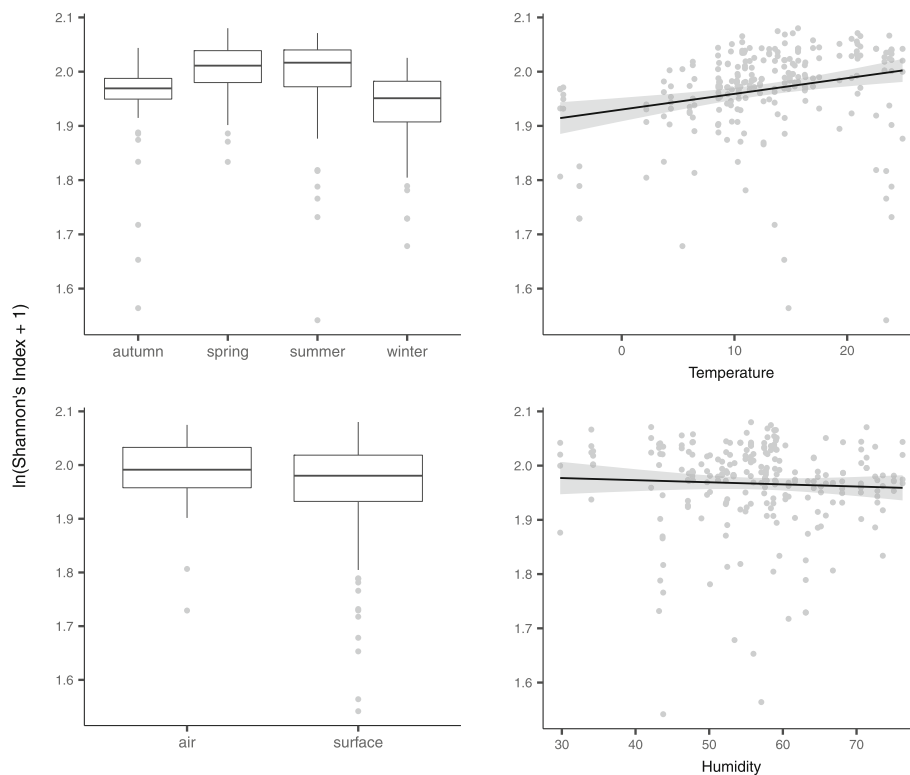


Fig. 3 Analysis of Shannon's diversity index. The four significant predictors of within-sample diversity (see Table 4)

resuspension rates, which introduces a high level of complexity in the relationship between air and surface microbiomes.

We observed that three *Verrucomicrobia* families (*Verrucomicrobiaceae*, *Rubritaleaceae*, and *Chthoniobacteraceae*) varied in abundance across seasons, showing the highest abundance during summer (Fig. 1b). *Verrucomicrobia*, which is part of the PVC superphylum, is ecologically diverse, often highly abundant and present in a range of different environments [44].

Among the three investigated surface types—kiosks, benches, and railings—we found more congruency among the highly abundant taxa (Additional file 1: Table S5), as compared with the level of differentiation observed between air and surface (Table 1).

To identify genera that were highly divergent among seasons, surface and air, and surface types, we performed random forest classification analyses, where genera were scored by their ability to bin samples in their correct category (season/sample matrix/surface type). The two genera with the highest importance for classifying samples by season, namely *Psychrobacter*, and *Cryobacterium* (Table 2) are both psychrophilic (cold tolerance or preference towards colder temperatures) [45, 46]. *Psychrobacter* was most abundant during winter and *Cryobacterium* during spring (Table 2; Additional file 1: Figure S6). For correctly binning surface and air samples,

Ralstonia and *Streptomyces* were the most important genera, both being more abundant in air samples (Table 3; Additional file 1: Figure S7). *Ralstonia* are environmental opportunistically pathogenic bacilli [47], while *Streptomyces* is a species-rich genus, highly abundant in soil where they play an important role in carbon cycling [48]. We note that *Ralstonia* has been identified as a common contaminant in sequence library preparation steps [49] and that such contaminants may introduce stronger bias in sequence data from low-biomass samples, such as air [50]. The random forest classification of samples by surface type performed very poorly (Additional file 1: Table S6), which indicated that genus level taxonomic composition is not strongly diverged among surface types. Thus, we conclude that taxonomic representation is much more similar across surface types, than across air/surface or different seasons.

Diversity

Analyses of within-sample diversity (Shannon's diversity index) and among-sample diversity (ordinated Bray–Curtis dissimilarity distances) revealed several interesting patterns. We analyzed diversity with some hitherto untried predictors (season, surface/air, indoor/outdoor stations), and some that have been included in previous subway studies (temperature, humidity, time of day, surface types).

Table 4 The best-fit model for Shannon's diversity index score, which explained 27% of within-sample diversity variance and had a p value of 1.04×10^{-09} . Slopes are given for continuous predictor variables and interactions between continuous and categorical predictors with two levels. Observed trends, from low to high average Shannon's diversity scores, are given for the categorical predictors

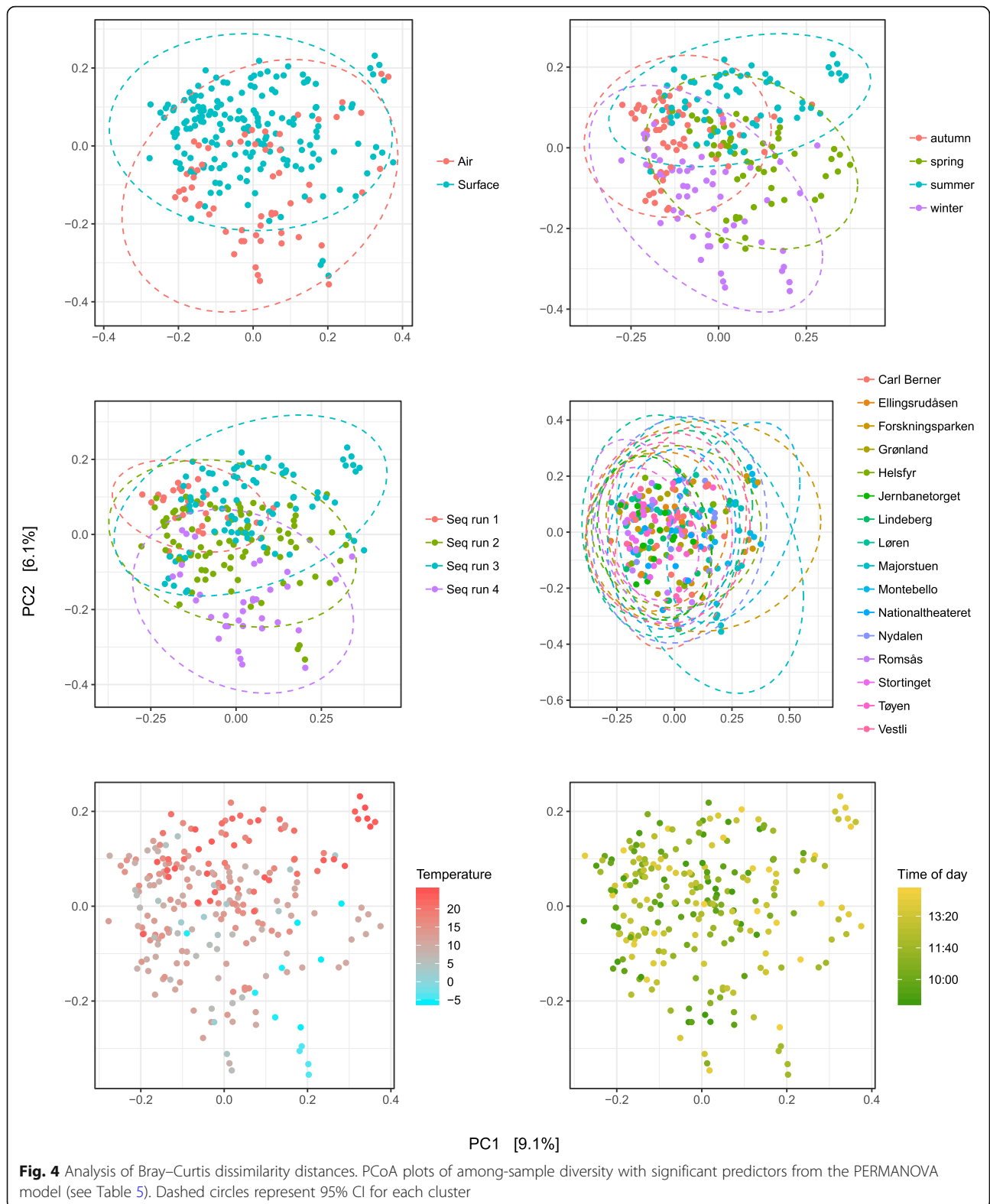
Predictor	DF	Sum Sq.	Mean Sq.	F	p	Slope/trend
Temperature	1	0.089	0.089	17.62	< 0.001	0.0006
Air/surface	1	0.041	0.041	8.20	0.005	Surface > air
Season	3	0.102	0.034	6.75	< 0.001	Winter > autumn > summer > spring
Humidity	1	0.029	0.029	5.74	0.017	- 0.0017
Humidity SD	1	0.005	0.005	1.04	0.309	- 0.1415
Temperature SD	1	0.002	0.002	0.47	0.493	- 0.1018
Time of day	1	0.002	0.002	0.43	0.514	- 0.0001
Indoor/Outdoor	1	0.000	0.000	0.000	1.000	Outdoor > indoor
Temperature SD: temperature	1	0.097	0.097	19.11	< 0.001	- 0.0130
Temperature: air/surface	1	0.051	0.051	10.02	0.002	- 0.0047
Time: indoor/outdoor	1	0.029	0.029	5.75	0.017	0.0005
Humidity SD: season	3	0.061	0.020	4.04	0.008	
Time: season	3	0.034	0.011	2.24	0.085	
Season: indoor/outdoor	3	0.029	0.010	1.91	0.130	
Humidity SD: humidity	1	0.005	0.005	1.01	0.316	0.0011
Temperature SD: season	3	0.006	0.002	0.37	0.773	
Humidity SD: temperature	1	0.000	0.000	0.08	0.772	0.0060
Temperature SD: time of day	1	0.000	0.000	0.00	0.990	0.0004
Residuals	209	1.056	0.005			

We found no evidence for within-sample diversity differing among surface types (kiosks, railings, and benches). Analysis of among-sample diversity *did* reveal a significant association (Additional file 1: Figure S9), although only a relatively small proportion of the variance in among-sample diversity was explained by surface type (~ 2%). This latter finding is congruent with that of Hsu et al. [5], who found that microbial community structure varied significantly across surface types in the Boston metropolitan transit system.

Previous studies have found time of day to be a highly important variable for understanding subway microbiome fluctuations, with peak and non-peak commuting hours showing marked differences [15, 17]. We found time of day to be a significant predictor of among-sample diversity (Table 5; Fig. 4), but that it explained a relatively small proportion of the total variance in diversity, as compared with the other predictors. This may partly be due to the huge difference in number of commuters between Oslo, and

Table 5 The best-fit PERMANOVA model, which explained 56% of among-sample diversity (Bray–Curtis dissimilarity)

Predictor	DF	Sum Sq.	Mean Sq.	F	R^2	p
Air/surface	1	2.484	2.484	14.16	0.04	0.001
Season	3	6.593	2.198	12.52	0.11	0.001
Subway station	15	9.134	0.609	3.47	0.15	0.001
Temperature	1	0.475	0.475	2.71	0.01	0.001
Sequence run	3	1.063	0.354	2.02	0.02	0.001
Time of day	1	0.239	0.239	1.36	< 0.01	0.050
Season: air/surface	3	0.853	0.284	1.620	0.01	0.001
Subway station: air/surface	15	3.382	0.225	1.28	0.06	0.001
Season: subway station	44	9.792	0.223	1.27	0.16	0.001
Residuals	151	26.500	0.175		0.44	



Hong Kong and Beijing, and that the present study sampled outside peak commuting hours. Furthermore, the study design used here is not suited to properly

gauge the importance of time of day—since this would require within-day repeated sampling for single locations.

Temperature was a highly significant predictor of both within-sample (Table 4; Fig. 3) and among-sample diversity (Table 5; Fig. 4), whereas humidity was only significant for within-sample diversity (Table 4; Fig. 3). Note that a very small proportion of the among-sample diversity total variance was explained by temperature (Table 5), while the effect size on within-sample diversity was pronounced (Fig. 3). Leung et al. [20] also found temperature and humidity to influence microbial diversity in the Hong Kong subway; note however, these associations were only significant when including outdoor stations. Our results show that humidity had a weak negative impact on diversity (Fig. 3), which is not congruent with Leung et al. who found a positive association. This incongruity may be explained by the variability and non-linear nature of the association between humidity and bacterial survival rates [51], which may give rise to different results across geographical areas and temporal scales. Humidity ranged from approximately 50 to 80% in the Leung et al. study, while our data ranged from 29.8 to 76.3%. Leung et al. found a negative association between temperature and diversity, the opposite of what we observe. Again, this is perhaps explained by the lack of overlap in the temperature range in the two studies (Leung et al.: approximately 24–30 °C; our study: – 5.45–24.91 °C).

Three of the 16 stations included in this study were outdoor subway stations. Indoor/outdoor was borderline significant in a univariate test ($p = 0.08$) of within-sample diversity; however, there was no significant association in the final multivariate model. The temperatures at outdoor stations will vary significantly throughout the seasons and even throughout the day, which may drive the (nearly significant) association between indoor/outdoor and within-sample diversity. When removing temperature from the final model of within-sample diversity, indoor/outdoor was again borderline significant ($p = 0.07$), which leads us to conclude that temperature outcompetes indoor/outdoor in our model (Table 4). Much like for temperature, we found indoor/outdoor to be a significant predictor of diversity in air samples (univariate; $p = 0.04$), but not in surface samples (univariate; $p = 0.29$). Reiterating the observed dynamic between indoor/outdoor and temperature mentioned above, a model with indoor/outdoor and temperature as predictors of air sample diversity only supported temperature ($p = 0.23$, $p = 5 \times 10^{-10}$, respectively). Although outdoor air is known to be a major source for indoor microbiomes [25], one would expect commuters, another important source [20], to be a more significant contributor in indoor environments. Hence, the lack of significance in univariate tests of indoor/outdoor as a predictor of diversity is an unexpected finding. One possible explanation is that there are relatively few commuters in Oslo,

making human sources less dominant, or that effective air exchange reduces the differences between indoor and outdoor air.

A major aim of this study was to compare subway air and surface microbiomes, and we found air/surface to be a highly significant predictor of both within-sample and among-sample diversity (Tables 4 and 5; Figs. 3 and 4). Importantly, the effect of this association was dependent on temperature; we found air to have lower within-sample diversity at low temperatures, and higher diversity at high temperatures (Additional file 1: Figure S8). This can be explained by microbial diversity in air being more sensitive to temperature, as compared with surface. To evaluate this hypothesis, we ran post hoc univariate analyses of Shannon's diversity index scores and temperature on air and surface samples separately, which found temperature to be a non-significant predictor for surface samples, ($R^2 = 0.01$; $p = 0.08$), but highly significant for air samples ($R^2 = 0.52$; $p = 4.05 \times 10^{-11}$). It appears that the diversity differences in air and surface microbiomes to a large extent are driven by differential effects of temperature. One explanation for this observation is the association between temperature and air circulation regimes, which can strongly influence air microbiome composition [52].

We found significant differences in within-sample and among-sample diversity across seasons (Tables 4 and 5; Figs. 3 and 4). Within-sample diversity was highest during spring and summer (Fig. 3). Apart from subway station, seasons explained the largest amount of among-sample diversity of all included predictors ($R^2 = 0.11$; Table 5). Seasonal variation has not previously been evaluated in subways using culture-independent methods; however, Patel et al. [28] cultured bacteria and fungi from dust collected at railway stations in England and Scotland, and Heo et al. [14] measured concentrations of culturable bacteria in underground subway stations through spring and autumn. Both studies are congruent with the results presented here; bacterial numbers increased from spring through summer and decrease towards winter. Several studies have observed seasonality in atmospheric microbiome composition [22–24]. With the outdoors being an important source for BE microbiomes [25], this suggests that seasonal variations in subway microbiomes may be influenced, at least partly, by seasonal changes in atmospheric microbial communities.

Subway station was a highly significant predictor of among-sample diversity, explaining 15% of the total variance (Table 5). However, when inspecting the clustering of PCoA ordinated values in Fig. 4, there are no clear patterns. We suspect that this result is mainly a consequence of including a categorical predictor with too many levels. Hence, we must refrain from concluding on

the importance of subway station as a predictor of microbiome composition in our study. Sequence run was also a significant predictor of among-sample diversity and explained 2% of the total variance. We propose that this stems from an unbalanced partitioning of samples from different seasons, sample matrices, or other variables into the four sequence runs. Alternatively, the association with sequence run may be explained by predictors not included. Both these explanations are congruent with the qPCR results, which show higher bacterial load in the samples from sequence run 3 (Additional file 1: Table S4 and Figure S2).

Caveats

In our study, seasonality was assessed by sampling on single days within seasons without accounting for the variation in shorter time periods (e.g., weekly variation) or repeatability across years. While patterns such as differential abundance of certain taxa in spring and summer, compared with autumn and winter are convincing, a higher resolution sampling scheme should be implemented in the future to distinguish between variations that occur on different timescales. Although we provide a relatively high level of geographical resolution in the present study, we recommend that future studies address seasonal and air/surface variability across cities, countries, and continents using standardized methods.

Conclusions

Understanding the composition and dynamics of air and surface microbiomes in mass transit environments—given their role in facilitating interactions among human and other BE microbiomes as well as infectious disease transmission and bioterrorism risk—is important for future developments in human health and security. Here we provide increased resolution to the state-of-knowledge regarding subway microbiomes by showing that there are significant differences between air and surface microbiomes, identifying seasonality as a central driver of subway microbiome variability, and confirming patterns previously observed in different geographical and demographical contexts. These results imply that biological detection, identification, and monitoring efforts in BEs should take into consideration the different characteristics/properties of air and surfaces, and that studies of microbial community composition should include seasonal sampling.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s40168-019-0772-9>.

Additional file 1: Table S1. Type of environment, latitude and longitude for all sampled stations. **Table S2.** Overview of all samples included in the analyses. **Table S3.** PCR program for 16S rRNA gene

amplicon sequencing. **Table S4.** The best-fit models of qPCR 16S rRNA gene copies for air samples and surface samples. **Table S5.** Top 20 phyla, families, and genera and species in surface samples collected on kiosks, benches, and railings. **Table S6.** Random forest classification models of samples collected from different surface types. **Figure S1.** The significant predictors of qPCR 16S rRNA gene copy yields in air samples. **Figure S2.** The significant predictors of qPCR 16S rRNA gene copy yields in surface samples. **Figure S3.** Quality profile of filtered reads. **Figure S4.** Rarefaction curves with observed diversity and Shannon's Diversity Index. **Figure S5. A)** Relative abundances of the top 15 phyla across the three surface types and seasons. **B)** Heatmap of most abundant families. **Figure S6.** Top 20 most important genera in random forest classification analysis of samples collected in different seasons. **Figure S7.** Top 20 most important genera in random forest classification analysis of air and surface samples. **Figure S8.** Interaction effect between temperature (°C) and air/surface in the linear model of Shannon's diversity index. **Figure S9.** PCoA plot of Bray Curtis dissimilarity distances with the only significant predictor (surface type) from the PERMANOVA model that included only surface-specific predictors.

Acknowledgements

We thank Vegard Osa Lie and Emilie Willoch Olstad for contributing to sampling and sample extraction, and Else Marie Fyske and Kristian Franer for their contribution to the sampling effort.

Authors' contributions

MD conceived, designed, and led the study. JG performed the data analysis and drafted the paper. KO-B performed the sampling and wet lab work. LV-M and PP contributed to the wet lab work and data processing. KIU contributed to study design. All authors contributed to the writing of the manuscript and approved the final version.

Funding

The study was funded by the Norwegian Defence Research Establishment (FF). Additional funding was obtained from the Stockholm Health Authority (Region Stockholm) # SLL 20160933 (KIU).

Availability of data and materials

The sequence data have been deposited in the NCBI Sequence Read Archive under accession PRJNA566330. (<https://www.ncbi.nlm.nih.gov/sra/PRJNA566330>).

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

¹Norwegian Defence Research Establishment, P.O. Box 25, NO-2027 Kjeller, Norway. ²Department of Analytics, Environmental & Forensic Sciences, King's College London, 150 Stamford Street, London SE1 9NH, UK. ³Department of Molecular Biosciences, Stockholm University, SE 10691 Stockholm, Sweden. ⁴SoS BIOs Sweden, Tiundagatan 41, SE 752 30 Uppsala, Sweden.

Received: 26 September 2019 Accepted: 20 November 2019

Published online: 19 December 2019

References

- Kim K-H, Kabir E, Jahan SA. Airborne bioaerosols and their impact on human health. *Journal of Environmental Sciences*. 2018;67:23–35 <https://doi.org/10.1016/j.jes.2017.08.027>.
- Humbal C, Gautam S, Trivedi U. A review on recent progress in observations, and health effects of bioaerosols. *Environment International*. 2018;118:189–93 <https://doi.org/10.1016/j.envint.2018.05.053>.
- Tasnim N, Abulizi N, Pither J, Hart MM, Gibson DL. Linking the gut microbial ecosystem with the environment: does gut health depend on

- where we live? *Frontiers in microbiology*. 2017;8:1935 <https://doi.org/10.3389/fmicb.2017.01935>.
4. Klepeis NE, Nelson WC, Ott WR, Robinson JP, Tsang AM, Switzer P, Behar JV, Hern SC, Engelmann WH. The national human activity pattern survey (NHAPS): a resource for assessing exposure to environmental pollutants. *Journal of Exposure Science and Environmental Epidemiology*. 2001;11(3):231.
 5. Hsu T, Joice R, Vallarino J, Abu-Ali G, Hartmann EM, Shafquat A, DuLong C, Baranowski C, Gevers D, Green JL, et al. Urban transit system microbial communities differ by surface type and interaction with humans and the environment. *mSystems*. 2016;1(3):e00018–6 <https://doi.org/10.1128/mSystems.00018-16>.
 6. Zhang N, Huang H, Duarte M, Zhang J. Dynamic population flow based risk analysis of infectious disease propagation in a metropolis. *Environment International*. 2016;94:369–79 <https://doi.org/10.1016/j.envint.2016.03.038>.
 7. Hernández AM, Vargas-Robles D, Alcaraz LD, Peimbert M. Station and train surface microbiomes of Mexico City's metro (subway/underground). *bioRxiv*. 2019;735027 <https://doi.org/10.1101/735027>.
 8. Afshinnekoo E, Meydan C, Chowdhury S, Jaroudi D, Boyer C, Bernstein N, Maritz Julia M, Reeves D, Gandara J, Chhangawala S, et al. Geospatial resolution of human and bacterial diversity with city-scale metagenomics. *Cell Systems*. 2015;1(1):72–87 <https://doi.org/10.1016/j.cels.2015.01.001>.
 9. Kang K, Ni Y, Li J, Imamovic L, Sarkar C, Kobler MD, Heshiki Y, Zheng T, Kumari S, Wong JCY, et al. The environmental exposures and inner- and intercity traffic flows of the metro system may contribute to the skin microbiome and resistome. *Cell Reports*. 2018;24(5):1190–1202.e1195 <https://doi.org/10.1016/j.celrep.2018.06.109>.
 10. Mason C, Afshinnekoo E, Ahsannudin S, Ghedin E, Read T, Fraser C, Dudley J, Hernandez M, Bowler C, Stolovitzky G, et al. The metagenomics and metadesign of the subways and urban biomes (MetaSUB) international consortium inaugural meeting report. *Microbiome*. 2016;4(1):24 <https://doi.org/10.1186/s40168-016-0168-z>.
 11. Danko DC, Bezdan D, Afshinnekoo E, Ahsanuddin S, Alicea J, Bhattacharya C, Bhattacharyya M, Blekhaman R, Butler DJ, Castro-Nallar E, et al. Global genetic cartography of urban metagenomes and anti-microbial resistance. *bioRxiv*. 2019;724526 <https://doi.org/10.1101/724526>.
 12. Anderson PD, Bokor G. Bioterrorism: pathogens as weapons. *Journal of pharmacy practice*. 2012;25(5):521–9.
 13. Herfst S, Böhringer M, Karo B, Lawrence P, Lewis NS, Mina MJ, Russell CJ, Steel J, de Swart RL, Mence C. Drivers of airborne human-to-human pathogen transmission. *Current Opinion in Virology*. 2017;22:22–9 <https://doi.org/10.1016/j.coviro.2016.11.006>.
 14. Heo KJ, Lee BU. Seasonal variation in the concentrations of culturable bacterial and fungal aerosols in underground subway systems. *Journal of Aerosol Science*. 2016;92:122–9 <https://doi.org/10.1016/j.jaerosci.2015.11.004>.
 15. Hwang SH, Park JB. Comparison of culturable airborne bacteria and related environmental factors at underground subway stations between 2006 and 2013. *Atmospheric Environment*. 2014;84:289–93 <https://doi.org/10.1016/j.atmosenv.2013.11.064>.
 16. Hwang SH, Park WM, Ahn JK, Lee KJ, Min KB, Park JB. Relationship between culturable airborne bacteria concentrations and ventilation systems in underground subway stations in Seoul, South Korea. *Air Quality, Atmosphere & Health*. 2016;9(2):173–8 <https://doi.org/10.1007/s11869-015-0316-9>.
 17. Dybwad M, Granum PE, Bruheim P, Blatny JM. Characterization of airborne bacteria at an underground subway station. *Applied and Environmental Microbiology*. 2012;78(6):1917 <https://doi.org/10.1128/AEM.07212-11>.
 18. Fan H, Li X, Deng J, Da G, Gehin E, Yao M. Time-dependent size-resolved bacterial and fungal aerosols in Beijing subway. *Aerosol and Air Quality Research*. 2017;17(3):799–809.
 19. Triadó-Margarit X, Veillette M, Duchaine C, Talbot M, Amato F, Minguiñón MC, Martins V, de Miguel E, Casamayor EO, Moreno T. Bioaerosols in the Barcelona subway system. *Indoor Air*. 2017;27(3):564–75 <https://doi.org/10.1111/ina.12343>.
 20. Leung MHY, Wilkins D, Li EKT, Kong FKF, Lee PKH. Indoor-air microbiome in an urban subway network: diversity and dynamics. *Applied and Environmental Microbiology*. 2014;80(21):6760 <https://doi.org/10.1128/AEM.02244-14>.
 21. Robertson CE, Baumgartner LK, Harris JK, Peterson KL, Stevens MJ, Frank DN, Pace NR. Culture-independent analysis of aerosol microbiology in a metropolitan subway system. *Applied and Environmental Microbiology*. 2013;79(11):3485 <https://doi.org/10.1128/AEM.00331-13>.
 22. Caliz J, Triadó-Margarit X, Camarero L, Casamayor EO. A long-term survey unveils strong seasonal patterns in the airborne microbiome coupled to general and regional atmospheric circulations. *Proceedings of the National Academy of Sciences, USA*. 2018;115(48):12229–34 <https://doi.org/10.1073/pnas.1812826115>.
 23. Uetake J, Tobo Y, Uji Y, Hill TCJ, DeMott PJ, Kreidenweis SM, Misumi R. Seasonal changes of airborne bacterial communities over Tokyo and influence of local meteorology. *Frontiers in microbiology*. 2019;10(1572) <https://doi.org/10.3389/fmicb.2019.01572>.
 24. Li H, Zhou XY, Yang XR, Zhu YG, Hong YW, Su JQ. Spatial and seasonal variation of the airborne microbiome in a rapidly developing city of China. *Science of the Total Environment*. 2019;665:61–8 <https://doi.org/10.1016/j.scitotenv.2019.01.367>.
 25. Leung MHY, Lee PKH. The roles of the outdoors and occupants in contributing to a potential pan-microbiome of the built environment: a review. *Microbiome*. 2016;4(1):21 <https://doi.org/10.1186/s40168-016-0165-2>.
 26. Rintala H, Pitkäranta M, Toivola M, Paulin L, Nevalainen A. Diversity and seasonal dynamics of bacterial community in indoor environment. *BMC Microbiology*. 2008;8(1):56 <https://doi.org/10.1186/1471-2180-8-56>.
 27. Prussin AJ II, Vikram A, Bibby KJ, Marr LC. Seasonal dynamics of the airborne bacterial community and selected viruses in a children's daycare center. *PLOS ONE*. 2016;11(3):e0151004 <https://doi.org/10.1371/journal.pone.0151004>.
 28. Patel KV, Bailey CL, Harding A-H, Biggin M, Crook B. Background levels of micro-organisms in the busy urban environment of transport hubs. *Journal of Applied Microbiology*. 2018;125(5):1541–51 <https://doi.org/10.1111/jam.14063>.
 29. Bøfot KO, Gohli J, Moen LV, Dybwad M. Performance evaluation of a new custom, multi-component DNA isolation method optimized for use in shotgun metagenomic sequencing-based aerosol microbiome research. *bioRxiv*. 2019;744334 <https://doi.org/10.1101/744334>.
 30. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLOS ONE*. 2013;8(4):e61217.
 31. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria [<https://www.R-project.org>]
 32. 16 s metagenomic sequencing library preparation: Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System: Illumina, Inc; 2013
 33. Krueger F. Trim Galore: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files; 2019, [https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/].
 34. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*. 2011;17(1):10–2.
 35. Andrews S: FastQC: a quality control tool for high throughput sequence data; 2019, [<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>].
 36. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*. 2016;13:581 <https://doi.org/10.1038/nmeth.3869>.
 37. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Research*. 2013;42(D1):D643–8 <https://doi.org/10.1093/nar/gkt1209>.
 38. Venables WN, Ripley BD. *Modern Applied Statistics with S*. New York: Springer; 2002. <https://doi.org/10.1007/978-0-387-21706-2>
 39. Nearing JT, Douglas GM, Comeau AM, Langille MGI. Denoising the denoisers: an independent evaluation of microbiome sequence error-correction approaches. *PeerJ*. 2018;6:e5364 <https://doi.org/10.7717/peerj.5364>.
 40. Chen M-Y, Wu S-H, Lin G-H, Lu C-P, Lin Y-T, Chang W-C, Tsay S-S. *Rubrobacter taiwanensis* sp. nov., a novel thermophilic, radiation-resistant species isolated from hot springs. *International Journal of Systematic and Evolutionary Microbiology*. 2004;54(5):1849–55 <https://doi.org/10.1099/ijs.0.63109-0>.
 41. Cafaro MJ, Poulsen M, Little AEF, Price SL, Gerardo NM, Wong B, Stuart AE, Larget B, Abbot P, Currie CR. Specificity in the symbiotic association between fungus-growing ants and protective *Pseudonocardia* bacteria. *Proceedings of the Royal Society of London, Series B: Biological Sciences*. 2011;278(1713):1814–22 <https://doi.org/10.1098/rspb.2010.2118>.
 42. Chander AM, Nair RG, Kaur G, Kochhar R, Dhawan DK, Bhadada SK, Mayilraj S. Genome insight and comparative pathogenomic analysis of *nesterenkonia jeotgali* strain cd08_7 isolated from duodenal mucosa of celiac disease patient. *Frontiers in microbiology*. 2017;8:129 <https://doi.org/10.3389/fmicb.2017.00129>.

43. Qian J, Hospodsky D, Yamamoto N, Nazaroff WW, Peccia J. Size-resolved emission rates of airborne bacteria and fungi in an occupied classroom. *Indoor Air*. 2012;22(4):339–51 <https://doi.org/10.1111/j.1600-0668.2012.00769.x>.
44. Op den Camp HJM, Islam T, Stott MB, Harhangi HR, Hynes A, Schouten S, Jetten MSM, Birkeland N-K, Pol A, Dunfield PF. Environmental, genomic and taxonomic perspectives on methanotrophic Verrucomicrobia. *Environmental Microbiology Reports*. 2009;1(5):293–306 <https://doi.org/10.1111/j.1758-2229.2009.00022.x>.
45. Kim SJ, Shin SC, Hong SG, Lee YM, Choi I-G, Park H. Genome sequence of a novel member of the genus *Psychrobacter* isolated from Antarctic soil. *Journal of Bacteriology*. 2012;194(9):2403 <https://doi.org/10.1128/JB.00234-12>.
46. Liu Q, Xin Y-H, Chen X-L, Liu H-C, Zhou Y-G, Chen W-X. *Cryobacterium aureum* sp. nov., a psychrophilic bacterium isolated from glacier ice collected from the ice tongue surface. *International Journal of Systematic and Evolutionary Microbiology*. 2018;68(4):1173–6 <https://doi.org/10.1099/ijsem.0.002647>.
47. Steinberg JP. Other gram-negative and gram-variable bacilli. In: Mandell, Douglas, and Bennett's principles and practice of infectious diseases, vol. 3023. Eighth ed; 2015.
48. Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, Klenk H-P, Clément C, Ouhdouch Y, van Wezel GP. Taxonomy, physiology, and natural products of Actinobacteria. *Microbiology and Molecular Biology Reviews*. 2016;80(1):1–43 <https://doi.org/10.1128/mmb.00019-15>.
49. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*. 2014;12(1):87 <https://doi.org/10.1186/s12915-014-0087-z>.
50. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends in Microbiology*. 2019;27(2):105–17 <https://doi.org/10.1016/j.tim.2018.11.003>.
51. Tang JW. The effect of environmental parameters on the survival of airborne infectious agents, 6. *Journal of the Royal Society, Interface, Suppl*. 2009;6(Suppl 6):S737–46 <https://doi.org/10.1098/rsif.2009.0227.focus>.
52. Cáliz J, Triadó-Margarit X, Camarero L, Casamayor EO. A long-term survey unveils strong seasonal patterns in the airborne microbiome coupled to general and regional atmospheric circulations. *Proceedings of the National Academy of Sciences*. 2018;115(48):12229–34 <https://doi.org/10.1073/pnas.1812826115>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

