Current Research in Microbial Sciences 8 (2025) 100374

Contents lists available at ScienceDirect

Current Research in Microbial Sciences

journal homepage: www.sciencedirect.com/journal/current-research-in-microbial-sciences

Bacterial diversity and biomarkers screening of station and carriage surface in Shanghai metro system, China

Lijun Zhang [a,b,1], Xiaojing Li [a,b,1], Lisha Shi [a,b], Yi Zheng [c], Yichen Ding [a,b], Tao Yuan [d], Shuangqing Hu [e], Jian Chen [a,b,*], Ping Xiao [a,b,*]

[a] Division of Public Health Service and Safety Assessment, Shanghai Municipal Center for Disease Control and Prevention, Shanghai 201107, China
[b] State Environmental Protection Key Laboratory of Environmental Health Impact Assessment of Emerging Contaminants, Shanghai 201107, China
[c] Shanghai Shentong Metro Group Co.,Ltd, Shanghai 201103, China
[d] Shanghai Jiao Tong University, Shanghai 200030, China
[e] Shanghai Academy of Environmental Sciences, Shanghai 200233,China

ARTICLE INFO

Keywords:
16S rRNA gene sequencing
Subway microbiome
Surface
Amplicon sequencing
Seasonal variation

ABSTRACT

Background: Mass transit environments, such as the metro, can facilitate the spread of bacteria between humans and their surroundings. These environments are particularly important for human health due to their potential for spreading pathogens and their impact on large populations. To gain a deeper understanding of bacterial distribution in subways, it is essential to identify variables that affect bacterial composition and microorganisms that are probably harmful to human heath.

Methods: We conducted high-throughput 16S rRNA gene sequencing on surface samples from 5 subway stations in Shanghai, China, during the warm(summer), cold(winter) and transition(autumn) seasons. Bacteria community features across the three seasons were distinguished using random forest classification analyses, followed by in-depth diversity analyses.

Results: Significant differences were observed in surface bacterial communities across seasons. Highly abundant bacterial groups were generally ubiquitous. Among these highly abundant families and genera, some were unique to surface samples. Notably, the phyla *Firmicutes, Proteobacteria,* and *Actinobacteria* were predominant, with total abundances of 32.87 %, 29.41 %, and 16.31 %, respectively. Alpha diversity indices were statistically significant ($P < 0.05$) among different seasons, with autumn exhibiting significantly higher alpha diversity metrics compared to summer and winter. Beta diversity analysis revealed significant compositional dissimilarities and distinct clustering patterns among the three seasons ($P < 0.05$). An analysis of similarities (ANOSIM) test results indicated significant differences in bacterial patterns at the phylum, class, order, family, genus levels among the seasons ($P < 0.05$). Random forest classification analyses identified the top 24 bacterial taxa at the genus level across seasons in the metro system.

Conclusions: We provided a direct comparison of surface bacterial microbiomes, and a comprehensive survey of seasonal variation in subways using culture-independent methods. Our findings reveal differences in both diversity and abundance of certain taxa across seasons, with 24 top indicator bacterial genera identified. This work serves as a reference for understanding the composition and dynamics of bacterial communities and for biomarker screening in subways, a crucial public space in our increasingly urbanized and interconnected world.

1. Introduction

Microorganisms are ubiquitous on Earth and can have profound and lasting impacts on human health (Kim et al., 2018). The global COVID-19 pandemic has heightened awareness of microorganisms in public environments and their implications for human health (Kim et al., 2018; Van Leuken et al., 2016; Herfst and al., 2017). Mass transit environments, such as the metro, are particularly noteworthy due to their potential to spread microbes among humans and environments, affecting large populations (Hsu et al., 2016; Zhang et al., 2016).

* Corresponding authors at: 1399 Shenhong Road, Shanghai, China.
E-mail addresses: chenjian_3@scdc.sh.cn (J. Chen), xiaoping@scdc.sh.cn (P. Xiao).
[1] Equal contribution to the study.

https://doi.org/10.1016/j.crmicr.2025.100374

Available online 8 March 2025

Table 1
Monitoring station locations of five **metro** stations.

Station	Environment	Type	Transfer Station	Latitude	Longitude
QB	underground	island platform	No	31.16	121.36
CHJ	underground	island platform	No	31.18	121.40
XJH	underground	island platform	Yes	31.20	121.44
SJDD	underground	island platform	Yes	31.23	121.53
MD	underground	island platform	No	31.22	121.48

Metro systems are major modes of transportation in many metropolitan areas worldwide, valued for their convenience, safety, speed, and environmental benefits. The development of these systems often symbolizes a city's economic and transportation advancement. With rapid expansion of mass transit systems, microbial transmission issues have garnered significant research attention during recently years (Hsu et al., 2016; Xu and Hao, 2017; Minguillón et al., 2018). Understanding microbial diversity in metro environments is crucial for grasping the dynamics of microorganism transmission and for detecting and monitoring potential pathogens and bioterrorism threats. Furthermore, it is essential to identify key microorganisms that interact with the transmit environment and human activity, as these may ultimately affect public health and wellbeing (Gohli et al., 2019; Kang et al., 2018).

Next-generation sequencing (NGS) technologies offer unprecedented opportunities to study microorganisms and their interactions with hosts, providing a comprehensive view of microbial dynamics in urban settings for researchers and policymakers (Danko et al., 2020). Leveraging advances in NGS and culture-independent methods, recent research has explored subway microbiomes in various cities, including New York City, Hong Kong, Mexico City, Moscow and Bangkok (Robertson et al., 2013; Leung et al., 2014; Hernández et al., 2020; Klimenko et al., 2020; Siriarchawatana et al., 2023). These studies have characterized subway surface microbiome using both 16S rRNA gene amplicon sequencing and shotgun metagenomics, revealing that subway microbiomes are predominantly composed of a limited number of microbial taxa at various taxonomic levels. Moreover, the microbial compositions in subway environments are most likely derived from human beings and outdoor environments (Robertson et al., 2013; Klimenko et al., 2020; Guevarra et al., 2022). These baseline microbiome profiles may help to perform long-term surveillance for pathogens, health risks, and bioterrorism threat mitigation within city transit systems (Afshinnekoo et al., 2015). Exploring the variability of mass transit surface microbiomes across different matrices and temporal scales, and identifying key microbiome biomarkers in composition, function, and resources, is crucial.

Shanghai, a prosperous and densely populated city in China, boasts one of the largest urban metro systems in the world (Gong et al., 2017). The system includes subway, light rail lines and maglev trains. Up to 2023, there are 18 metro lines (including the Shanghai Maglev Train), 508 stops and 831 km of tracks in operation, making it the longest metro system in Asia and the third-longest in the world. The system is continuously expanding, with plans to reach 20 lines and 877 km by 2025. Thus, understanding subway microbiome, including both pathogens and non-pathogens, is vital for long-term microbiome surveillance in Shanghai. To our knowledge, few studies have reported on the sampling matrices, temporal scales, and key microbiome biomarkers in Shanghai's metro system, particularly regarding surface microbiomes.

On the other hand, with the development of machine learning, the random forest classifier model has been employed to assess the classification power of biomarkers, especially in clinical medicine (Papoutsoglou et al., 2023; Yang et al., 2021). Some studies have identified fecal microbial markers for early diagnosis of clinical diseases (Liang et al., 2020; Yu et al., 2017; Zhu et al., 2023; Zou et al., 2022) and environmental health issues (Ghannam and Techtmann, 2021) using machine learning techniques. However, the presence of different microbiome biomarkers in various public places, such as subways or metros, remains unclear. Furthermore, current knowledge about seasonal microbiome biomarkers in subway systems is limited, with few studies investigating microbial alterations and diversities across seasons (Leung et al., 2014; Klimenko et al., 2020; Guevarra et al., 2022; Ryon and al., 2022). Research using machine learning to identify key microorganisms across seasons in subway or metro system is also scarce.

In this study, we conducted monitoring across warm, transitional and cold seasons at typical subway stations in Shanghai. We explored bacterial diversity and microorganism characterization using 16S rRNA gene amplicon sequencing and bioanalytical technologies. To identify microbiome biomarkers for different seasons, we utilized random forest classification analyses to highlight key bacterial taxa in the metro system. This research provides a direct comparison of surface bacterial microbiomes and an overall survey of seasonal variations in subways using culture-independent methods. It also identifies microbiome biomarkers for seasonal classification, representing a significant step towards understanding the composition and dynamics of bacterial communities in subways and offering a case study on exploring indicator bacterial communities across different seasons using machining learning.

2. Material and methods

2.1. Setting

Shanghai is located in the east of China with an urban area of over 6340.5 km^2 and a population of 24.8709 million by the end of 2020 (2021-ShanghaiBasicFacts.pdf (shio.gov.cn)). Metro line 9 is the sixth completed subway line in Shanghai, China. The first phase of line 9 opened on December 29, 2007. By the end of 2021, there are 35 stations in the operation of the metro line 9 in Shanghai. It starts from Songjiang South Railway Station in Songjiang District, passes through Minhang, Xuhui and Huangpu Districts, and ends at Caolu Station in Pudong New Area. In our research, five metro stations (QB, CHJ, XJH, SJDD, MD) of metro line 9 were selected to perform the monitoring. Information in detailed was shown in Table 1.

2.2. Sample collection

Surface samples were collected from five surface types at each station and train: Hanging Grip, Vertical Pole, Doorknob, Escalator Handrail, and Ticket Machine. A sterile cotton swab moistened with sterile saline was used to swab the surface. For the surface of regularly planar objects, a sampling specification plate of 10 cm x 10 cm was used to standardize the area size. For the surface of irregular objects, such as the surface of a pull ring which is relatively small in area itself, the sampled area should be as large as possible and close to 10 x 10 cm to facilitate the comparison of the analysis results. The swab was then placed in a 5-mL centrifuge tube and stored in a transport cooler with ice packs until it could be transferred to a − 80 °C freezer upon returning to the laboratory. The sampling was performed during warm, transition and cold seasons over the course of 1 year, both inside the stations in the presence of passengers during normal operational hours and outside the stations.

2.3. Genomic DNA extraction

Samples were brought out of the −80 °C freezer and allowed to thaw to room temperature. DNA was extracted using the TIANamp Micro DNA Kit (DP316, TIANGEN BIOTECH). The integrity of the extracted DNA was assessed using 1 % agarose gel electrophoresis, and its concentration was measured with an ultra-micro spectrophotometer (Thermo NanoDrop2000). Samples passed the quality control test if the DNA mass

Table 2
Sample characteristic and basic information.

Sample name	Season	Surface type	Station	Temperature	Humidity	Trasferstation
QL01	Transition	hanging grip	carriage	23.4	62.6	—
QL02	Transition	hanging grip	carriage	23.4	62.6	—
QL03	Transition	vertical pole	carriage	23.4	62.8	—
QL04	Transition	vertical pole	carriage	23.4	62.8	—
QL05	Transition	doorknob	QB	26.6	53.6	No
QL06	Transition	escalator handrail	QB	24.8	54.7	No
QL07	Transition	touchscreen	QB	24.5	55.5	No
QL08	Transition	doorknob	CHJ	23.2	65.7	No
QL09	Transition	escalator handrail	CHJ	22.3	65.0	No
QL10	Transition	touchscreen	CHJ	22.6	65.2	No
QL11	Transition	doorknob	XJH	25.7	66.3	Yes
QL12	Transition	escalator handrail	XJH	22.1	71.2	Yes
QL13	Transition	touchscreen	XJH	23.5	68.1	Yes
QL14	Transition	ticket machine	MD	24.9	74.9	Yes
QL15	Transition	escalator handrail	MD	24.8	73.6	Yes
QL16	Transition	touchscreen	MD	23.1	76.5	Yes
QL17	Transition	doorknob	SJDD	24.5	74.5	Yes
QL18	Transition	escalator handrail	SJDD	25.4	69.5	Yes
QL19	Transition	touchscreen	SJDD	24.7	69.3	Yes
SL01	Summer	hanging grip	carriage	24.4	75.9	—
SL02	Summer	hanging grip	carriage	24.4	76.1	—
SL03	Summer	vertical pole	carriage	24.4	75.5	—
SL04	Summer	vertical pole	carriage	24.4	75.4	—
SL05	Summer	doorknob	QB	29.0	74.5	No
SL06	Summer	escalator handrail	QB	26.4	75.9	No
SL07	Summer	touchscreen	QB	26.7	76.8	No
SL08	Summer	doorknob	CHJ	30.3	79.1	No
SL09	Summer	escalator handrail	CHJ	27.4	77.0	No
SL10	Summer	touchscreen	CHJ	30.3	77.9	No
SL12	Summer	escalator handrail	XJH	28.7	88.8	Yes
SL13	Summer	touchscreen	XJH	28.3	88.0	Yes
SL17	Summer	doorknob	SJDD	29.5	84.7	Yes
SL18	Summer	escalator handrail	SJDD	29.1	80.6	Yes
SL19	Summer	touchscreen	SJDD	28.1	82.7	Yes
WL01	Winter	hanging grip	carriage	19.7	66.1	—
WL02	Winter	hanging grip	carriage	19.7	66.1	—
WL03	Winter	vertical pole	carriage	19.7	66.1	—
WL04	Winter	vertical pole	carriage	19.7	66.1	—
WL05	Winter	doorknob	QB	19.1	73.6	No
WL06	Winter	escalator handrail	QB	17.1	77.8	No
WL07	Winter	touchscreen	QB	14.7	81.6	No
WL08	Winter	doorknob	CHJ	16.8	70.1	No
WL09	Winter	escalator handrail	CHJ	15.4	70.6	No
WL10	Winter	touchscreen	CHJ	15.0	68.6	No
WL11	Winter	doorknob	XJH	12.4	68.8	Yes
WL12	Winter	escalator handrail	XJH	15.0	74.9	Yes
WL13	Winter	touchscreen	XJH	14.6	76.1	Yes
WL14	Winter	ticket machine	MD	11.8	72.2	Yes
WL15	Winter	escalator handrail	MD	11.8	72.2	Yes
WL16	Winter	touchscreen	MD	11.8	72.2	Yes
WL17	Winter	doorknob	SJDD	14.3	80.2	Yes
WL18	Winter	escalator handrail	SJDD	12.8	73.6	Yes
WL19	Winter	touchscreen	SJDD	11.9	73.5	Yes

concentration was ≥30 ng/μL and the total DNA mass was ≥3 μg. The DNA was then stored at −20 °C for subsequent experiments.

2.4. Genome library construction and sequencing

PCR amplification of the 16S rRNA gene sequences was performed, targeting the V3-V4 regions using a specific set of forward and reverse primers based on a previous study (Hsu et al., 2016; Klimenko et al., 2020; Chen et al., 2022). The forward primer 341F (5′-CCTACGGGNGGCWGCAG-3′) and reverse primer 805R (5′-GGACTACHVGGGTWTCTAAT-3′) were utilized for amplifying the V3 and V4 regions. The PCR was conducted in a total volume of 25 μL, including 12.5 μL of 2X KAPA HiFi HotStart Ready Mix, 400 nM of each primer, 7.5 μL of ddH2O and 3 μL of template DNA. Amplification was carried out in a thermocycler (Px2 Thermal Cycler, Thermo, USA) under the following conditions: initial denaturation at 95 °C for 30 s followed by 28 cycles of denaturation at 95 °C for 30 s, annealing at 55 °C for 30 s and elongation at 72 °C for 30 s.

The amplicons were verified by electrophoresis using 1.5 % agarose gel run at 110 V for 30 min. PCR products of the expected sizes were purified from the gel matrix. The purified amplicons were sequenced using the pair-end method on the MiSeq Illumina platform (Illumina Inc., San Diego, CA, USA), following the manufacturer's instructions. The Nextera XT DNA Library Preparation Kit (Illumina) was employed to construct libraries from the isolated DNA. Illumina MiSeq sequences were processed using the Quantitative Insights Into Microbial Ecology (QIIME2) pipeline, where chimeric sequences, marginal sequence errors, and noisy sequences were filtered. Amplicon sequence variants (ASVs were identified using DADA2, and taxonomy classification was performed using the SILVA reference database (Hsu et al., 2016). The relative abundance of the microbial community associated with each experimental group was visualized using the QIIME2 view.

For genome library construction, DNA was fragmented to approximately 150 bp sizes (Long et al., 2016). The construction involved end

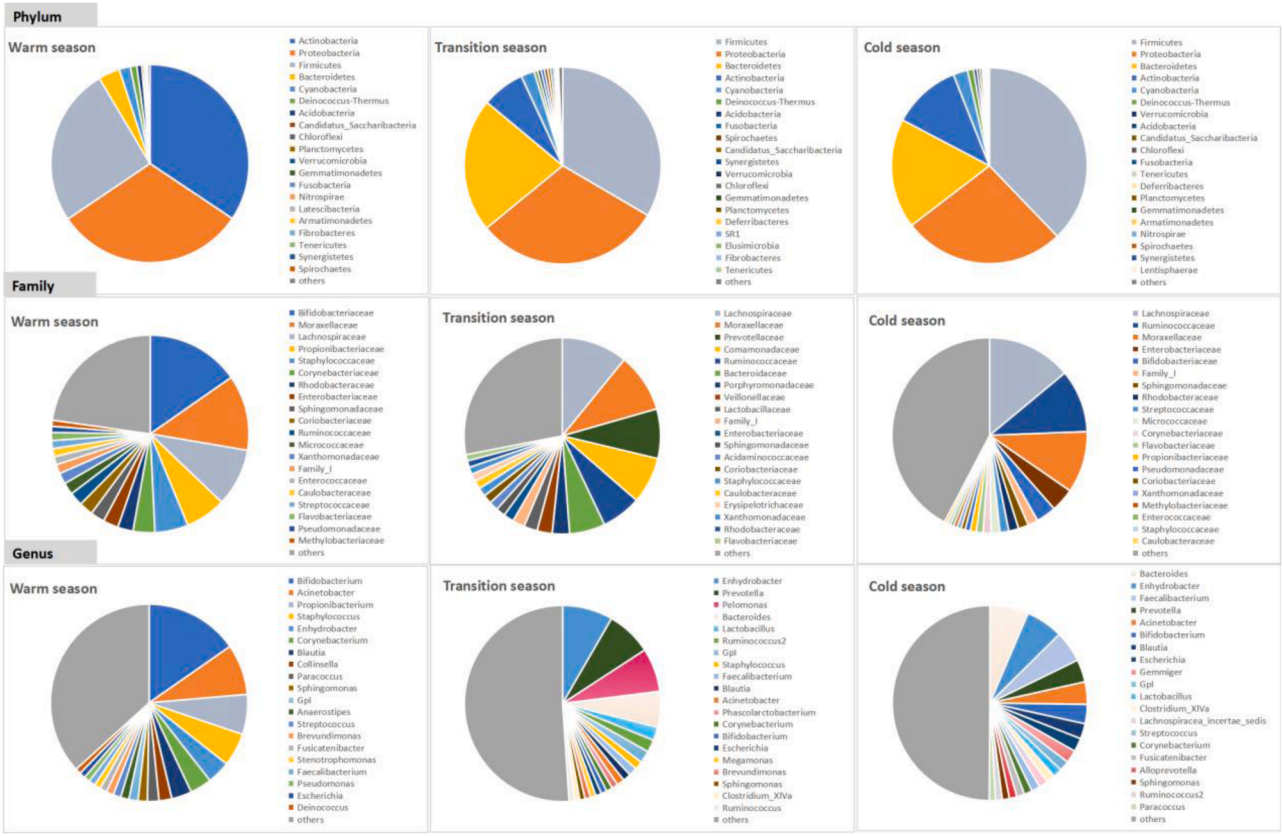

Fig. 1. Taxonomy and community composition across different seasons (top 20).

repair, splice ligation, and PCR amplification, followed by sequencing on the MGISEQ-2000 platform (Jeon et al., 2014).

2.5. Statistical analysis

Statistical analysis of bacterial alpha diversity was conducted on experimental samples collected in summer, autumn, and winter. The analysis considered various sample types (Hanging grip, Vertical pole, Doorknob, Escalator handrail, and Ticket machine) and different stations (QB, CHJ, XJH, SJDD, MD). Diversity indices, including Chao1, ACE, Shannon, Simpson, and coverage, were evaluated using the Kruskal–Wallis test with significance set at a p-value < 0.05. Beta diversity was assessed to determine differences in the microbiome across samples, often combined with dimensionality reduction techniques such as principal Coordinate Analysis (PCoA), Non-metric Multidimensional Scaling (NMDS), or Constrained Principal Component Analysis(CPCA) for visualization. These analyses were performed using the R vegan package(version 2.5–6), and the inter-sample distances were visualized as scatterplots. Random forest classification analysis was used to differentiate samples by season, identifying the most important genera for accurate seasonal classification through Mean Decrease Accuracy (MDA). Key parameters, including the number of trees (ntree) and the number of candidate splitting features per node (mtry), were set based on previous studies (Zhang et al., 2018). According to the reference (Gohli et al., 2019; Pellegrino et al., 2021), to avoid overfitting, the common practice is to optimize a tuning parameter that governed the number of features that were randomly chosen to grow each tree from the bootstrapped data. The optimal model k-fold cross-validation can be determined by choosing tuning parameters (mtry, ntree) that minimized test sample prediction error. when the number of trees reaches 501 and mtry reaches 4, the model performance tends to stabilize. The Out-of-bag estimate (OOB) display the error rate and represent the accuracy of the classifier model, to identify three season groups.

3. Results

3.1. Sample description

Surface samples ($n = 60$) were collected from five subway stations and train carriages over three seasons, from June 2022 to December 2022. These samples included five surface types: Hanging grip, Vertical pole, Doorknob, Escalator Handrail, and Ticket Machine. Station samples were specifically taken from touchscreens and the sides of fare ticketing machines. For each sample, metadata was collected detailing the built environment type, surface type, material composition and collection date (refer to Table 2).

16S rRNA amplicon sequencing was conducted to investigate bacterial community diversity and composition within each experimental group. All negative controls (six swab samples) failed to produce sequence able libraries during the library preparation step due to insufficient DNA yields. The samples showed no signs of contamination and produced the correct genera. Out of the total samples, 53 were successful in 16S rRNA gene amplicon sequencing; however, 7 samples failed in genome extraction.

3.2. Taxonomy and community composition

After the demultiplexing of paired-end reads and the implementation of quality control measures to eliminate chimeric sequences, Operational Taxonomic Unit (OTU) counts were normalized for each sample. A rarefaction curve was generated using the lowest sequencing depth among upstream, midstream, geothermal valley, and downstream samples, as well as biofilm and water samples.

3.2.1. Taxonomy and community composition across different seasons

In the surface samples, the phyla *Firmicutes, Proteobacteria*, and *Actinobacteria* predominated, accounting for total abundances of 32.87

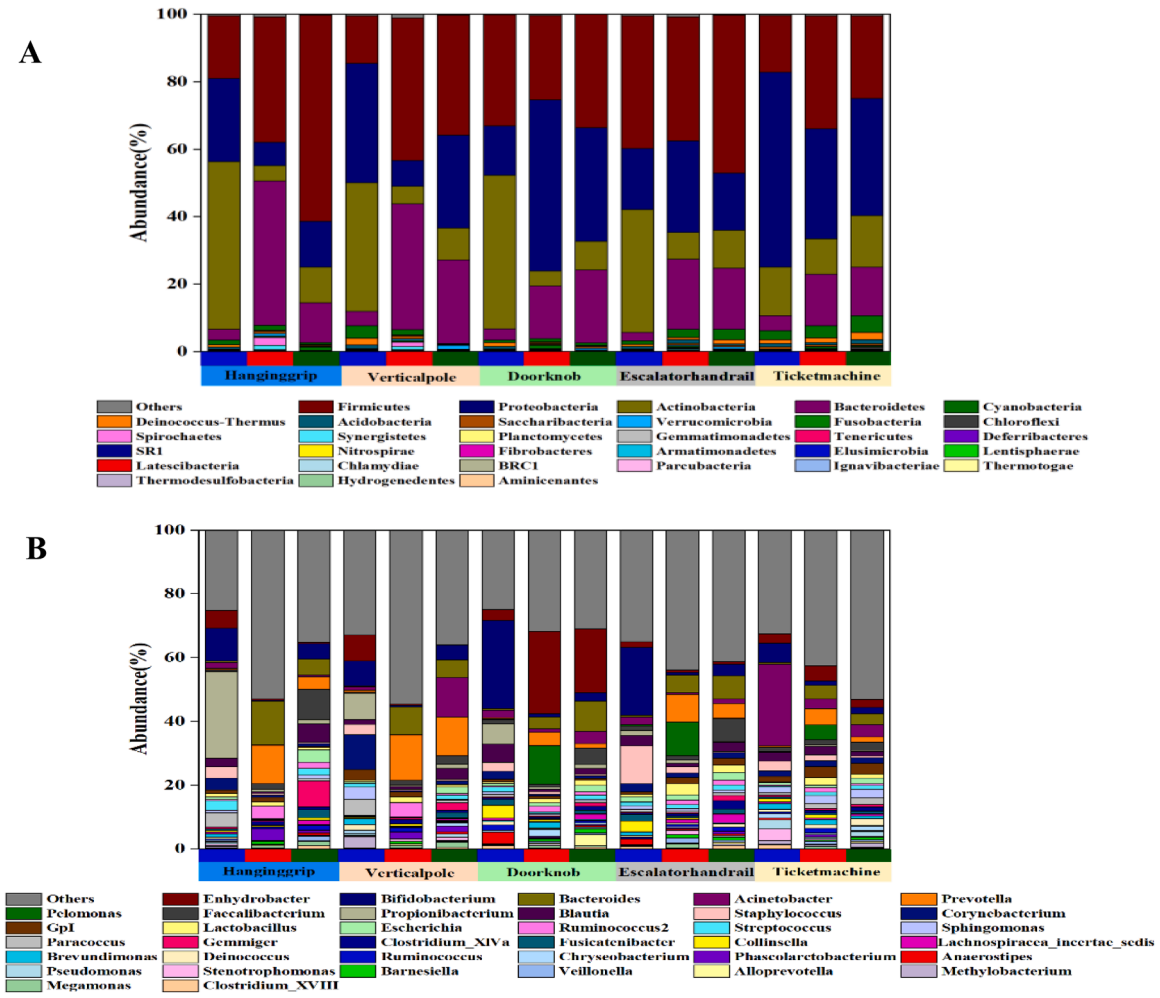

Fig. 2. Taxonomy and community composition across different surfaces (A: phylm level, B: genus level).

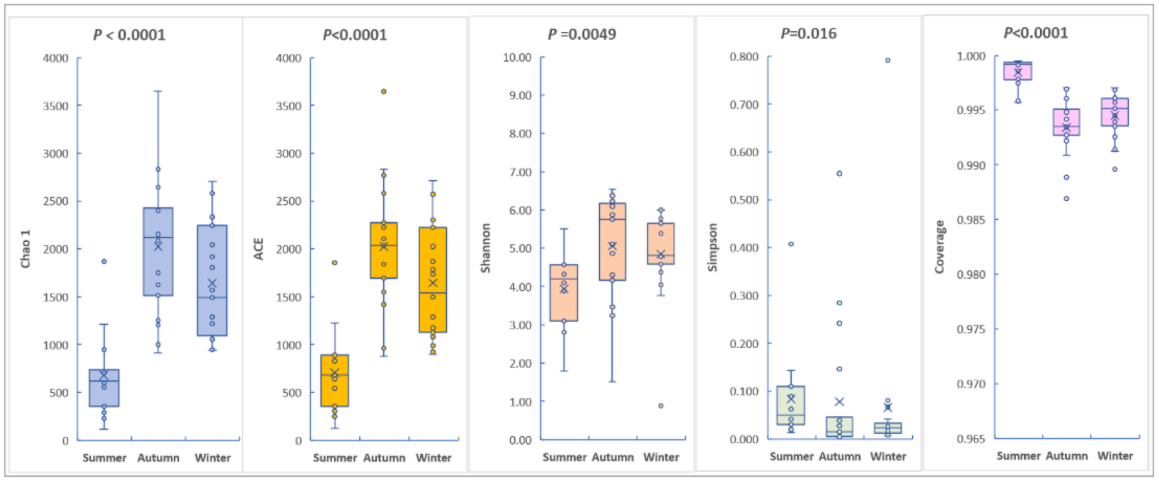

Fig. 3. Alpha diversity of bacterial community across different seasons.

%, 29.41 %, and 16.31 %, respectively (Fig. 1). The top 20 phyla were consistent across both autumn and winter samples, with the top six categorized by abundance appearing in identical order. While *Actinobacteria* was the predominant phylum in summer samples, the top six phyla were also consistent with those observed in the other two seasons. Notably, *Latescibacteria, Elusimicroba, and Letisphaerae* were highly abundant in samples from summer, autumn, and winter, ranking 15th (0.01 %), 18th(0.03 %) and 20th (0.01 %), respectively.

At the family level, the similarities among the top 20 families in the three seasonal samples remained pronounced, with *Lachnospiraceae, Moraxellaceae, and Ruminococcaceae* dominating, showing total abundances of 11.58 %, 10.65 %, and 6.71 %, respectively. However,

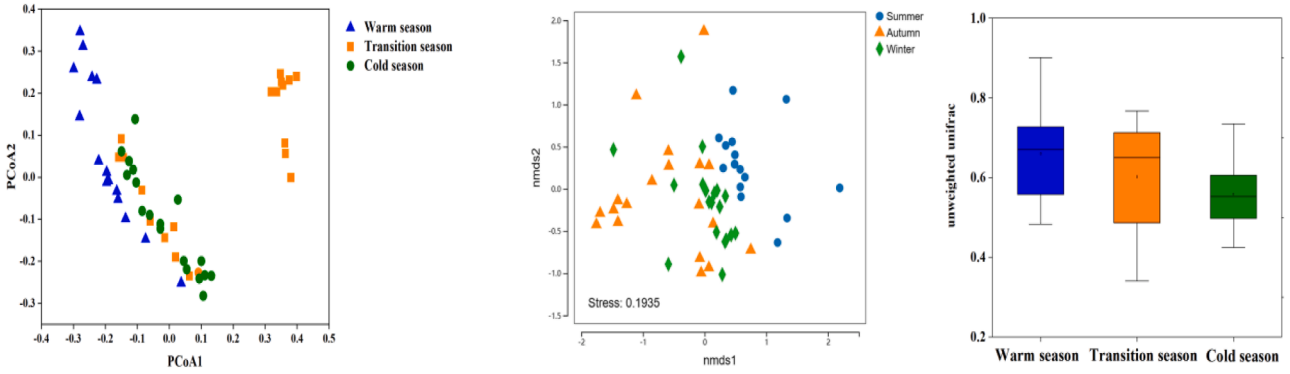

Fig. 4. Beta diversity of bacterial community across different seasons.

Table 3
ANOSIM test results of bacterial patterns in phylum, class, order, family, genus levels between different seasons.

Levels	Group	ANOSIM R	*P*	Group	ANOSIM R	*P*
Phylum	Warm	0.788	0.001	Warm-Transition	0.774	0.001
	Transition			Warm-Cold	0.794	0.001
	Cold			Transition-Cold	0.692	0.001
Class	Warm	0.785	0.001	Warm-Transition	0.773	0.001
	Transition			Warm-Cold	0.791	0.001
	Cold			Transition-Cold	0.805	0.001
Order	Warm	0.780	0.001	Warm-Transition	0.764	0.001
	Transition			Warm-Cold	0.794	0.001
	Cold			Transition-Cold	0.686	0.001
Family	Warm	0.779	0.001	Warm-Transition	0.763	0.001
	Transition			Warm-Cold	0.783	0.001
	Cold			Transition-Cold	0.686	0.001
Genera	Warm	0.544	0.001	Warm-Transition	0.654	0.001
	Transition			Warm-Cold	0.705	0.001
	Cold			Transition-Cold	0.375	0.001
Species	Warm	0.788	0.001	Warm-Transition	0.770	0.001
	Transition			Warm-Cold	0.795	0.001
	Cold			Transition-Cold	0.702	0.001

Acidaminococcaceae was exclusively present in autumn samples, with an abundance of 1.5 %, ranking 13th.

At the genus level, differences among the top 20 genera across the three seasons were observed (Fig. 1). In summer, Bifidobacterium was the most abundant genus, comprising 15.3 % of the total. Four unique genera were identified, including *Anaerostipes*(1.35 %, ranking 12th), *Stenotrophomonas* (1.07 %, ranking 16th), *Pseudomonas*(0.98 %, ranking 18th), and *Deinococcus* (0.92 %, ranking 20th). In autumn, E*nhydrobacter, Prevotella,* and *Pelomonas* were the top three genera, with abundances of 8.34 %, 7.47 %, and 7.21 %, respectively. Additionally, *Phascolarctobacterium* and *Megamonas* were uniquely present in autumn, ranking 12th (1.02 %) and 16th (0.91 %), respectively. In winter, the dominant genera differed from those in summer and autumn, with *Bacteroides, Enhydrobacter,* and *Faecalibacterium* leading the ranks, each with abundances of approximately 6.0 %. Notably, compared to the other two seasons, *Lachnospiracea_incertae_sedis* and *Alloprevotella* were unique to winter, exhibiting abundances of 1.46 % and 1.24 %, respectively.

3.2.2. *Taxonomy and community composition across different surfaces*

Microbial communities from subway surfaces, including Hanging Grip, Vertical Pole, Doorknob, Escalator Handrail, and Ticket Machine, were dominated by members of the phyla *Firmicutes, Proteobacteria, and Actinobacteria*, each accounting for over 20 % of the microbial community. These phyla are commonly associated with human-associated microbiota (Fig. 2A and 2B). The next most abundant taxa were *Bacteroidetes*, which included the genera *Micrococcus* (found in hair and skin) and *Rothia* (found in the oral cavity), along with *Streptococcaceae* (found in the oral cavity) and *Pseudomonadaceae* (Hsu et al., 2016). *Bacteroidetes* are prevalent in soil ecosystems and have been shown to associate with various eukaryotic hosts, including plants, animals, and humans (Pan et al., 2023).

At the genus level, *Enhydrobacer, Bifidobacterium, Bacteroides,* and *Acinetobacter* were the most prominent, collectively comprising approximately 5 % to 20 % of the microbial community, and they are also recognized as typical members of the human microbiota.

3.3. *Diversity*

3.3.1. *Alpha diversity*

16S rRNA amplicon sequencing was conducted to assess bacterial community diversity and composition within each experimental group. The alpha diversity indices-Chao1, ACE, Shannon, Simpson, and coverage-were analyzed for each group (Fig. 3). The results indicated statistically significant differences in alpha diversity indices among the seasonal groups ($P < 0.05$), with autumn showing the highest diversity metrics, significantly surpassing those of summer and winter. Conversely, the alpha diversity indices did not display statistically significant differences among the various surface and station site groups ($P > 0.05$) (supplementary files).

As illustrated in Fig. 3, the genus-level alpha diversity indices for Chao1, ACE, Shannon, Simpson, and coverage were computed and compared across the seasonal samples. The findings revealed that the Chao1 and Simpson indices were significantly different between the seasons ($P < 0.05$).

3.3.2. *Beta diversity*

Beta diversity analysis, based on Bray–Curtis distances, was employed to compare bacterial communities across different experimental groups. The results were visualized using Principal Coordinate Analysis (PCoA) along two axes (Fig. 4). This analysis revealed a distinct clustering pattern, indicating variability in bacterial communities among the seasonal groups. Significant compositional differences and distinct clustering patterns were observed between the three seasons ($P = 0.025$). In contrast, the results further demonstrated significant differences in bacterial community composition at the phylum, class,

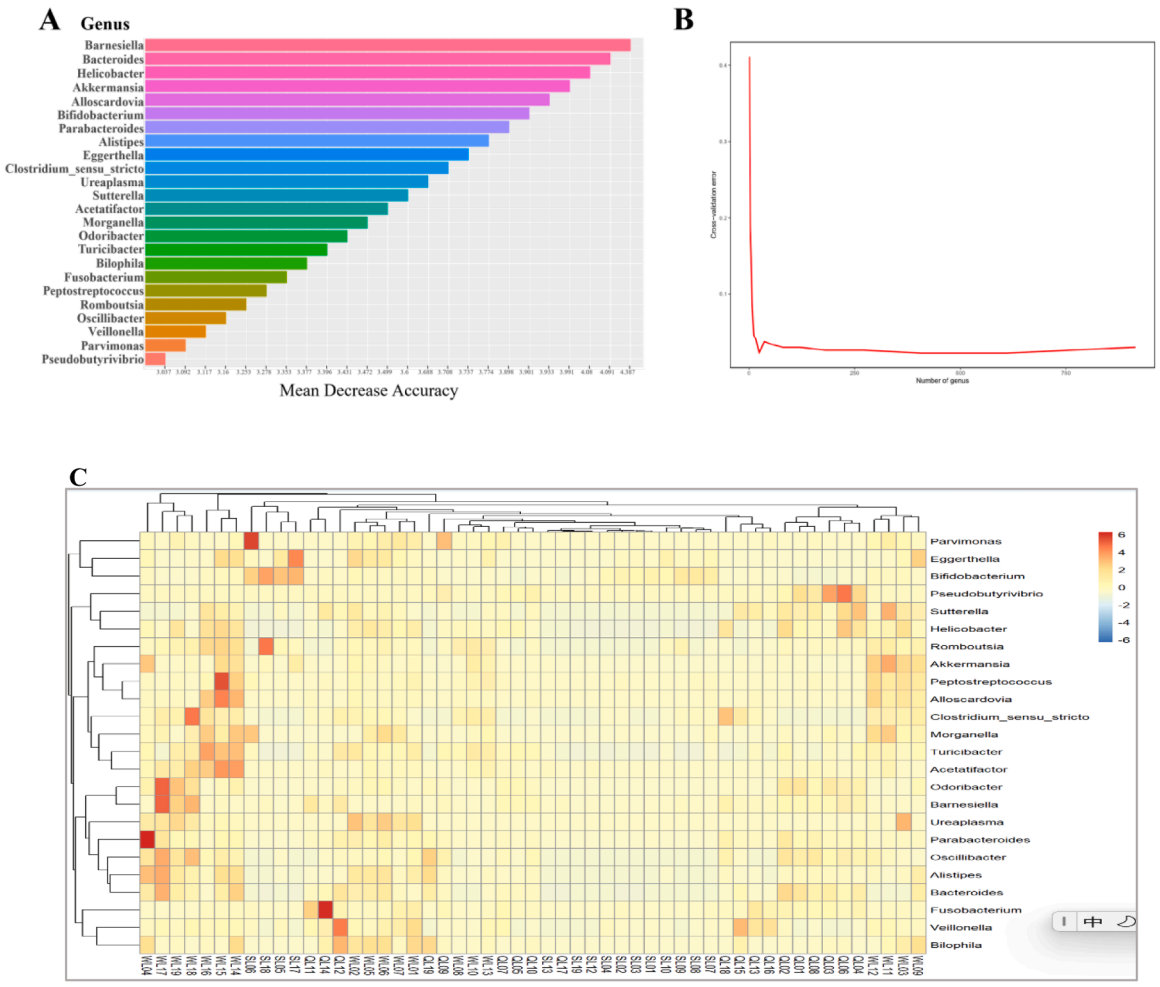

Fig. 5. Bacterial taxonomic biomarkers in different seasons.

order, family, and genus levels across the three seasons (Table 3).

3.3.3. A model to correlate surface bacterial taxonomic biomarkers with seasons

To identify surface bacterial taxonomic biomarkers at the genus level, we employed the Random Forests machine learning algorithm to establish a model correlating these biomarkers with the three seasons (Zhang et al., 2018; Zhang et al., 2024). The model accounted for 94.34 % of the variance in the root microbiota associated with rice residence time in the field. To identify key bacterial classes as biomarkers for correlating with rice residence time, we conducted a 10-fold cross-validation with five repeats. The minimum cross-validation error was achieved using 24 important genera (Fig. 5A). Consequently, these 24 genera were defined as biomarker taxa in the model (Table 4). The list of the top 24 bacterial genera, ranked by their time-discriminatory importance across seasons in the metro system, is presented in Fig. 5B and 5C. Most of the biomarker taxa exhibited high relative abundances in their respective seasons.

Confusion matrices provide insight into the classification of samples and the corresponding classification errors. For the 24 most important genera used in classifying samples, the mean decrease in MDA - reflecting the impact of removing each specific genus - and the mean Z-scores are presented.

(**A:** The top 24 bacterial genera identified as biomarkers were determined through Random Forests regression analysis of their relative abundances across the three seasons. These biomarker taxa are ranked in descending order of their importance to the model's accuracy. **B:** The inset illustrates the 10-fold cross-validation error as a function of the number of input classes used for regression against the seasons, ordered by variable importance. **C:** A heatmap displaying the relative abundances of the top 24 age-predictive bacterial genera is presented, correlated with rice residence time in the field).

4. Discussion

The metro system, is particularly significant concerning infectious diseases and urban public security (Anderson and Bokor, 2012), as evidenced by studies employing both culture-based and culture-independent methodologies (Robertson et al., 2013; Heo and Lee, 2016; Triadó-Margarit et al., 2017). Through metagenomics sequencing, the MetSUB project reveals that urban microbiome across various cities exhibit distinct geographic variation, which may reflect epidemiological differences and enhance future forensic capabilities (Danko et al., 2020).

Shanghai, located at 31°41′ north latitude and 121°29′ east longitude, experiences a northern subtropical maritime monsoon climate characterized by four distinct seasons, ample sunshine, and abundant rainfall. Its spring and autumn are relatively brief compared to the summer and winter (Municipality, 2021). These unique climatic conditions influence the microbial community in Shanghai's environment. Analysis of surface samples showed that the phyla *Firmicutes, Proteobacteria*, and *Actinobacteria* dominated, with respective abundances of 32.87 %, 29.41 %, and 16.31 %. Notably, the top 20 phyla remained consistent, especially in autumn and winter. The results indicated statistically significant differences in alpha diversity and beta diversity indices among various seasonal groups. ANOSIM revealed significant

Table 4
The results of Random Forests machine learning algorithm model.

Season					
Out-of-bag estimate of error rate:5.66 %					
Confusion matrix:	Warm	Tansition	Cold	class.error (%)	
Warm	15	0	0	0.000	
Transition	1	17	1	0.105	
Cold	1	0	18	0.053	
Most important genera in sample classification					
	Autumn	Summer	Winter	MDA	Mean Decrease Gini
Barnesiella	4.042	3.681	0.561	4.387	0.514
Bacteroides	3.719	4.133	0.361	4.091	0.384
Helicobacter	2.335	3.931	3.541	4.080	0.430
Akkermansia	3.061	1.868	3.403	3.991	0.371
Alloscardovia	1.655	2.855	4.001	3.933	0.309
Bifidobacterium	2.999	3.780	2.269	3.901	0.373
Parabacteroides	2.927	3.816	3.230	3.898	0.372
Alistipes	1.965	3.643	2.872	3.774	0.308
Eggerthella	3.244	0.962	3.396	3.737	0.270
Clostridium_sensu_stricto	1.336	3.360	3.953	3.708	0.472
Ureaplasma	2.218	1.792	3.539	3.688	0.193
Sutterella	2.226	3.627	2.810	3.600	0.341
Acetatifactor	0.231	2.826	3.241	3.499	0.277
Morganella	2.666	1.273	2.962	3.472	0.211
Odoribacter	2.474	3.682	1.240	3.431	0.288
Turicibacter	0.803	2.681	2.755	3.396	0.241
Bilophila	0.283	3.263	3.146	3.377	0.312
Fusobacterium	3.322	2.382	0.758	3.353	0.253
Peptostreptococcus	2.432	1.595	3.301	3.278	0.249
Romboutsia	3.030	−0.369	3.008	3.253	0.247
Oscillibacter	1.351	2.376	2.511	3.160	0.237
Veillonella	1.244	2.447	2.412	3.117	0.156
Parvimonas	1.576	2.464	2.634	3.092	0.319
Pseudobutyrivibrio	2.599	1.968	2.365	3.037	0.163

differences in bacterial patterns across phylum, class, order, family, and genus among the three seasons. These findings further underscore that seasonal and meteorological conditions are primary factors influencing microbiome communities.

In addition to seasonal variations, we also investigated the differences in microbiome communities among various subway stations and surface types. Microbial communities on subway surfaces, including hanging grips, vertical poles, doorknobs, escalator handrails, and ticket machines, were primarily composed of members of the phyla *Firmicutes, Proteobacteria,* and *Actinobacteria*, all of which are commonly associated with human microbiota. The next most abundant taxa were *Bacteroidetes*, prevalent in soil ecosystems and associated with various eukaryotic hosts, including plants, animals, and humans (Hsu et al., 2016; Pan et al., 2023). This suggests that human activities are a significant source of environmental microorganisms within the subway system.

To identify genera that displayed considerable seasonal variation, we conducted random forest classification analyses, scoring genera based on their ability to correctly classify samples by season. The 24 genera with the highest importance for seasonal classification included *Barnesiella, Bacteroides, Helicobacter, Akkermansia, Alloscardovia* and *Bifidobacterium*. A study conducted in Oslo, Norway, by Jostein Gohli et al. identified 20 important genera for classifying seasonal samples using random forest classification (Gohli et al., 2019). The genera reported in their study differed slightly from those in our research, likely due to the geographical (Oslo vs. Shanghai) and seasonal (four seasons vs. three seasons) differences between the two studies. In addition, we employed 10-fold cross-validation with five repetitions to evaluate the importance of bacterial classes, a method not mentioned in Gohli et al.'s study or other research on subway microbiomes.

However, the study has certain limitations. The data were obtained from only five subway stations over three seasons, which may not fully capture the complexity of the microbiome within the Shanghai subway environment. Despite this, our results provided valuable insight into the bacterial composition of the subway's micro-environment. Moreover, we demonstrated that random forests screening, combined with 10-fold cross-validation, is a robust method for identifying key microorganisms in environmental monitoring, particularly for public health surveillance in shared spaces. Future research should expand the sample size and incorporate complementary analytical approaches to further validate and enhance these findings.

5. Conclusion

In this study, we conducted a seasonal survey and comparison of bacterial communities using surface samples collected from five representative subway micro-environments in Shanghai. The analysis employed random forests screening combined with 10-fold cross-validation and five repetitions to identify key seasonal microbiomes, with 24 bacterial taxa at the genus level being reported. This work provides not only a direct comparison of surface bacterial microbiomes and an overall survey of seasonal variation in subways using culture-independent methods, but also identifies microbiome biomarkers relevant to season classification. The findings lay the groundwork for future research on subway environments and support the use of 16S rRNA gene sequencing and related analyses as bio-surveillance techniques for monitoring pathogens in urban mass transit systems. This study represents a significant advancement in understanding the composition, dynamics and ecological significance of bacterial communities in subways, offering a case study in the application of machine learning to identify indicator bacterial communities across different seasons.

CRediT authorship contribution statement

Lijun Zhang: Conceptualization, Methodology, Validation, Formal analysis, Funding acquisition, Project administration, Resources, Software, Investigation, Data curation, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Xiaojing Li:** Methodology, Validation, Investigation, Data curation, Funding acquisition, Visualization, Writing – review & editing. **Lisha Shi:** Methodology, Data curation, Software. **Yi Zheng:** Investigation, Methodology, Resources, Data curation, Supervision. **Yichen Ding:** Investigation, Data curation. **Tao Yuan:** Data curation, Project administration. **Shuangqing Hu:** Data curation, Project administration. **Jian Chen:** Conceptualization, Methodology, Formal analysis, Funding acquisition, Data curation, Validation, Supervision. **Ping Xiao:** Conceptualization, Methodology, Funding acquisition, Visualization, Data curation, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical standards

This research did not involve Human Participants and/or Animals.

Funding

This study was sponsored by the Shanghai Municipal Health Commission (Grant No 202240327, GWVI-11.1-38), and the State Environmental Protection Key Laboratory of Environmental Health Impact Assessment of Emerging Contaminants (Grant No SEPKL-EHIAEC-2022).

Acknowledgments

This study was sponsored by Shanghai Municipal Health Commission (Grant No 202240327, GWVI-11.1-38), and the State Environmental Protection Key Laboratory of Environmental Health Impact Assessment of Emerging Contaminants (Grant No SEPKL-EHIAEC-2022). We also acknowledge the cooperation of Shanghai Shentong Metro Group Co., Ltd. and Shanghai An-He Environmental Testing Technology Co., Ltd. during sampling process.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.crmicr.2025.100374.

Data availability

The authors do not have permission to share data.

References

Afshinnekoo, E., et al., 2015. Geospatial resolution of Human and bacterial diversity with city-scale metagenomics. Cell Syst. 1 (1), 97, 97.e3.
Anderson, P.D., Bokor, G., 2012. Bioterrorism: pathogens as weapons. J. Pharm. Pract. 25 (5), 521–529.
Chen, J.S., et al., 2022. Comprehensive assessment of bacterial communities and their functional profiles in the Huang Gang Creek in the Tatun Volcano Group basin, Taiwan using 16S rRNA amplicon sequencing. Ecotoxicol. Environ. Saf. 234, 113375.
Danko, D., et al., 2020. Global Genetic Cartography of Urban Metagenomes and Anti-Microbial Resistance, 724526.
Ghannam, R.B., Techtmann, S.M., 2021. Machine learning applications in microbial ecology, human microbiome studies, and environmental monitoring. Comput. Struct. Biotechnol. J. 19, 1092–1107.
Gohli, J., et al., 2019. The subway microbiome: seasonal dynamics and direct comparison of air and surface bacterial communities. Microbiome 7 (1), 160.
Gong, Y., et al., 2017. Health risk assessment and personal exposure to volatile organic Compounds (VOCs) in metro carriages - a case study in Shanghai, China. Sci. Total. Environ. 574, 1432–1438.
Guevarra, R.B., et al., 2022. Metagenomic characterization of bacterial community and antibiotic resistance genes found in the mass transit system in Seoul, South Korea. Ecotoxicol. Environ. Saf. 246, 114176.
Heo, K.J., Lee, B.U., 2016. Seasonal variation in the concentrations of culturable bacterial and fungal aerosols in underground subway systems. J. Aerosol. Sci. 92, 122–129.
Herfst, S., et al., 2017. Drivers of airborne human-to-human pathogen transmission. Curr. Opin. Virol. 22, 22–29.
Hernández, A.M., et al., 2020. Station and train surface microbiomes of Mexico City's metro (subway/underground). Sci. Rep. 10 (1), 8798.
Hsu, T., et al., 2016. Urban transit system microbial communities differ by surface type and interaction with humans and the environment. mSystems 1 (3).
Jeon, Y.J., et al., 2014. The feasibility study of non-invasive fetal trisomy 18 and 21 detection with semiconductor sequencing platform. PLoS ONE 9 (10), e110240.
Kang, K., et al., 2018. The environmental exposures and inner- and intercity traffic flows of the metro system may contribute to the skin microbiome and resistome. Cell Rep. 24 (5), 1190–1202 e5.
Kim, K.H., Kabir, E., Jahan, S.A., 2018. Airborne bioaerosols and their impact on human health. J. Environ. Sci. 67, 23–35.
Klimenko, N.S., et al., 2020. Co-occurrence patterns of bacteria within microbiome of Moscow subway. Comput. Struct. Biotechnol. J. 18, 314–322.
Leung, M.H., et al., 2014. Indoor-air microbiome in an urban subway network: diversity and dynamics. Appl. Environ. Microbiol. 80 (21), 6760–6770.
Liang, J.Q., et al., 2020. A novel faecal lachnoclostridium marker for the non-invasive diagnosis of colorectal adenoma and cancer. Gut 69 (7), 1248–1257.
Long, Y., et al., 2016. Diagnosis of sepsis with cell-free DNA by next-generation sequencing technology in ICU patients. Arch. Med. Res. 47 (5), 365–371.
Minguillón, M.C., et al., 2018. Aerosol sources in subway environments. Environ. Res. 167, 314–328.
Municipality, I.O.o.S., 2021. Shanghai Basic Facts.
Pan, X., Raaijmakers, J.M., Carrión, V.J., 2023. Importance of bacteroidetes in host–microbe interactions and ecosystem functioning. Trends. Microbiol. 31 (9), 959–971.
Papoutsoglou, G., et al., 2023. Machine learning approaches in microbiome research: challenges and best practices. Front. Microbiol. 14, 1261889.
Pellegrino, E., Jacques, C., Beaufils, Ne., et al., 2021. Machine learning random forest for predicting oncosomatic variant NGS analysis. Sci. Rep. 11, 21820.
Robertson, C.E., et al., 2013. Culture-independent analysis of aerosol microbiology in a metropolitan subway system. Appl. Environ. Microbiol. 79 (11), 3485–3493.
Ryon, K.A., et al., 2022. A history of the MetaSUB consortium: tracking urban microbes around the globe. iScience 25 (11), 104993.
Siriarchawatana, P., et al., 2023. Temporal, compositional, and functional differences in the microbiome of Bangkok subway air environment. Environ. Res. 219, 115065.
Triadó-Margarit, X., et al., 2017. Bioaerosols in the Barcelona subway system. Indoor Air 27 (3), 564–575.
Van Leuken, J.P.G., et al., 2016. Atmospheric dispersion modelling of bioaerosols that are pathogenic to humans and livestock – A review to inform risk assessment studies. Microb. Risk. Anal. 1, 19–39.
Xu, B., Hao, J., 2017. Air quality inside subway metro indoor environment worldwide: a review. Environ. Int. 107, 33–46.
Yang, Y., et al., 2021. Dysbiosis of human gut microbiome in young-onset colorectal cancer. Nat. Commun. 12 (1), 6757.
Yu, J., et al., 2017. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut 66 (1), 70–78.
Zhang, J., et al., 2018. Root microbiota shift in rice correlates with resident time in the field and developmental stage. Sci. China Life Sci. 61 (6), 613–621.
Zhang, L., et al., 2024. Bacterial structures and their associated ARGs in Shanghai subway air, China. J. Aerosol. Sci. 179, 106383.
Zhang, N., et al., 2016. Dynamic population flow based risk analysis of infectious disease propagation in a metropolis. Environ. Int. 94, 369–379.
Zhu, L., et al., 2023. Endocrine disrupting chemicals in indoor dust: a review of temporal and spatial trends, and human exposure. Sci. Total. Environ. 874, 162374.
Zou, M., et al., 2022. Gut microbiota on admission as predictive biomarker for acute necrotizing pancreatitis. Front. Immunol. 13, 988326.