Microbial communities on station and train surfaces in Chennai Metro: Insights into urban transit microbiome

Vijaya Yuvaram Singh[1,2,3,5,*], Veerendra P. Gadekar[2,3,*], Srijith Sasikumar[1,2,3], R. M. Rajeeva Lokshanan[1,2,3], Bharath Prithiviraj[4], Himanshu Sinha[1,2,3,#], Karthik Raman[1,2,3,5,#]

[1] Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, IIT Madras, Chennai, India
[2] Center for Integrative Biology and Systems Medicine, IIT Madras, Chennai, India
[3] Wadhwani School of Data Science and Artificial Intelligence, IIT Madras, Chennai, India
[4] Reckitt Health US, Microbiome Science Platforms, Montvale, New Jersey, USA
[5] Department of Data Science and Artificial Intelligence, IIT Madras, Chennai, India

*These authors contributed equally to the work.

[#]Corresponding authors: Karthik Raman kraman@dsai.iitm.ac.in, Himanshu Sinha sinha@iitm.ac.in

ABSTRACT

Urban public transport systems, particularly metro networks, serve as key hubs for microbial transmission, yet the urban microbiome in densely populated regions like India remains poorly characterized. These environments harbor diverse microbial communities, including both beneficial and pathogenic species, which can influence public health. The COVID-19 pandemic has further highlighted the need to monitor microbial ecosystems, particularly with respect to antimicrobial resistance (AMR) genes that may have escalated due to increased antibiotic use during health crises. In a first-of-its-kind study in India, we comprehensively characterized microbial communities and the prevalence of AMR genes in the Chennai Metro system. Of the 96 surface swab samples collected from 12 stations across two metro lines, 47 samples passed quality control and were subsequently analyzed using whole genome metagenomic sequencing. Comparative analysis with global urban microbiome datasets revealed distinct microbial profiles, including eight core species unique to Chennai. Surface type significantly influenced microbial diversity, with kiosks displaying the highest diversity levels. While AMR gene presence was minimal overall, genes associated with Sulfonamide and Rifamycin resistance were detected. These findings highlight unique microbial signatures and emphasize the need for ongoing surveillance and targeted interventions to mitigate microbial transmission risks in densely populated urban areas.

KEYWORDS

Metagenomics; Whole genome sequencing; Urban microbiome; Urban transit systems; Surface-associated microbial communities; Pangenome analysis; Antimicrobial resistance (AMR)

INTRODUCTION

Public transport systems represent a dynamic environment for microbial transmission, facilitating daily interactions among diverse populations and their microbiomes. The metro system, in particular, acts as a hub for a wide range of microorganisms, including beneficial commensals and symbionts, as well as potentially pathogenic bacteria, making it a potential vector for spreading infectious disease [1, 2]. As critical infrastructure in urban settings, public transportation systems significantly influence the microbial ecology of cities. Global initiatives, such as the MetaSUB consortium [3, 4], have advanced our understanding of urban microbiomes by characterizing the unique microbial profiles of cities around the world, providing crucial data for urban planning and public health interventions. Despite the increasing studies of urban microbiomes, densely populated Indian cities remain unexplored. As one of the fastest-growing urban centers in the world, Chennai, with a population exceeding 12 million [5], exemplifies the dramatic transformation of Indian cities in the 21st century. This study represents the first comprehensive analysis of urban microbiomes in public transportation systems in India, offering critical insights into the unseen biological ecosystems that thrive in these densely populated spaces.

Understanding the microbial communities that inhabit public transit is essential not only for urban public health but also for global efforts to monitor AMR—a growing problem and hidden pandemic [6] made worse by increasing antibiotic usage during health emergencies like the COVID-19 pandemic. In a metropolis like Chennai, where environmental factors, human activity, and rapid urbanization intersect, the dynamics of microbial communities, including their potential to harbor AMR genes, is still poorly understood. Unlike subway systems in more extensively studied cities, Chennai's unique socio-environmental landscape presents a distinct microbial signature, making it a key region for global comparative microbiome studies.

Recent advancements in next-generation sequencing (NGS) technologies have revolutionized our understanding of microbial communities, particularly those that are difficult to cultivate in the laboratory. The increasing availability of NGS data in public repositories has facilitated unprecedented opportunities to track the dissemination of microorganisms within urban environments [7–9]. Investigating bacterial communities in regions severely affected by the COVID-19 pandemic is crucial for understanding the dynamics of AMR gene dissemination, which may have been amplified by the increased use of antibiotics during the pandemic [10, 11].

Urban microbiome research has exploded in recent years, driven by initiatives such as the MetaSUB consortium, which aims to map microbial ecosystems in cities worldwide. However, India—a country where over a billion people live in increasingly interconnected urban areas—has largely remained absent from this global discourse. Our study not only fills this significant geographic gap but also reveals unique microbial patterns, including

the identification of eight core species not found in other global cities. These discoveries raise urgent questions about the role of urban environments in shaping microbial diversity and AMR profiles in developing nations.

By characterizing the microbial landscape of Chennai's Metro system, we provide the first window into how such environments in India may serve as reservoirs for both beneficial and pathogenic microorganisms. Moreover, this work is a call to action for the implementation of continuous surveillance systems in urban transit environments, where rapid microbial transmission could pose heightened risks in future public health crises.

MATERIALS AND METHODS

Sampling microbial communities in the Chennai Metro stations

To explore the microbial communities in the Chennai Metro system, we collected samples (n = 96) from 12 metro trains and stations along the Blue and Green Lines managed by Chennai Metro Rail Limited (CMRL) (Figure 1A). The Blue Line spans 32.65 km, connecting Wimco Nagar Depot to Chennai International Airport, while the Green Line covers 22 km from Chennai Central to St. Thomas. Sampling occurred in two batches: December 2021 and January 2022.

Using the MetaSUB protocol [12], samples were collected with individually packaged Isohelix Buccal Mini Swabs (MS Mini DNA/RNA Swab, Isohelix, Cat.: MS-02) and stored in barcoded 2D Matrix V-Bottom ScrewTop Tubes (Thermo Scientific, Cat.: 3741-WP1D-BR/1.0mL) pre-filled with 400 µL of Zymo Shield transport and storage solution (Zymo Research, Cat.: R1100) to preserve DNA and RNA. Immediately after collection, swabs were placed in the matrix tubes and stored at -80°C until DNA extraction.

The samples were taken from frequently touched surfaces, including banisters (handrails inside metro stations), kiosks (ticket vending machines), rods (handrails inside trains), and ticket counters (Figure 1B). Four samples were collected per station, amounting to 48 samples per batch. Additionally, one negative control (an air sample) was collected at the Chennai Airport metro station. This comprehensive sampling strategy targeted diverse interaction points in the transit environment, providing a robust representation of microbial diversity. Out of 96 collected samples, 47 passed quality control (QC) criteria and were subjected to shotgun metagenomic sequencing for downstream analysis (Figure 1C). A detailed list of all collected samples is provided in Supplementary Tables S1.

DNA Extraction, Library Preparation, and Sequencing

DNA extraction was performed using the Microlab STAR Liquid Handling System (Hamilton, Cat.: Microlab STAR) and the ZymoBIOMICS 96 MagBead DNA Kit (Zymo Research, Cat.: D4308) to ensure high-quality recovery of both abundant and low-

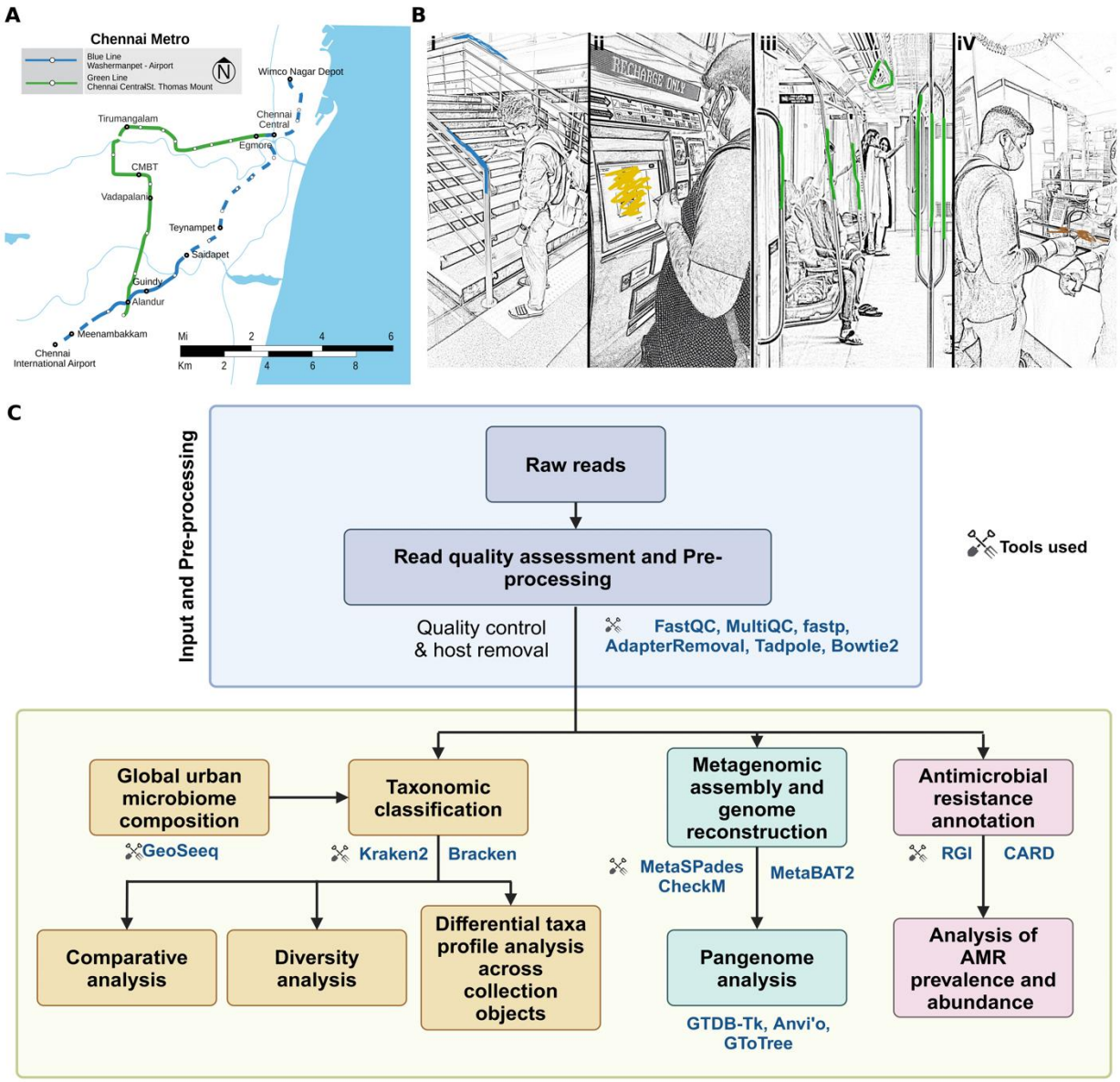

Figure 1: *(A) Chennai Metro Line Map. Highlighting sample collection stations* [13] *(B) Samples were collected from various surfaces, including – (i) handrails in metro stations (banisters), (ii) ticket vending machines (kiosks), (iii) handrails in trains (rods), and (iv) ticket counters. The swabbing areas on these surfaces are highlighted in blue, yellow, green, and red, respectively; (C) Workflow for data processing and analysis.*

abundance microbial species. An initial QC was conducted using the Qubit dsDNA High-Sensitivity Assay. A minimum DNA concentration of 0.9 ng was required for samples to pass QC, both after extraction and library preparation. Libraries were prepared with the KAPA HyperPrep Kit (Roche), and sequencing was carried out using the Illumina NovaSeq 6000 platform, which was chosen for its ability to generate high-throughput, paired-end reads suitable for metagenomic analysis. These methods enabled us to maximize the coverage of the microbial diversity present in the samples.

Preprocessing of the raw reads and mapping

To characterize the bacterial communities present in Chennai's metro stations, bacterial taxa were identified through DNA sequencing. The raw Illumina sequencing reads were first assessed for quality using FastQC v0.11.9, and MultiQC v1.13 was employed to aggregate and visualize the quality reports. Preprocessing and assembly of the reads were conducted using modules from the MetaSUB Core Modular Analysis Pipeline (CAMP; https://github.com/MetaSUB-CAMP) [14]. The preprocessing steps included filtering out low-quality bases, low-complexity regions, and short reads using fastp v0.22.0. Adapter sequences were removed from the filtered reads with AdapterRemoval v2.3.3. Error correction was performed on the filtered reads using the Tadpole tool from the BBTools v39.01 aligner package. Tadpole corrected sequencing errors by comparing overlapping reads, identifying mismatches or gaps, and generating a consensus sequence based on majority bases. This process reduced noise and improved accuracy by correcting base-calling errors. After preprocessing, the quality of the corrected reads was reassessed using FastQC and MultiQC to ensure dataset integrity. Finally, the processed reads were mapped to host reference genomes, including the human genome assembly (GRCh38) and the mouse genome assembly (MM39), using Bowtie2 v2.4.5 with the 'very-sensitive' flag (Table S2). Reads that mapped to the host genomes were subsequently removed to focus on microbial sequences.

Taxonomic classification and identification of the core and sub-core species

For taxonomical classification and determination of the relative abundance of species in all metagenomic samples, we employed Kraken2 v2.1.3 and Bracken v2.7 from the CAMP pipeline. Kraken2 matched individual k-mers within query sequences to the lowest common ancestor (LCA) among all genomes containing the respective k-mer in the kraken2 database v0.1.1 made available by CAMP. To estimate the species-level relative abundance, we used Bracken v2.7, which redistributed reads probabilistically from higher taxonomic levels to refine abundance estimates. We defined the core and sub-core species based on the distribution of taxa prevalence across our dataset. Here, prevalence refers to the fraction of samples in which a particular taxon was found at any abundance. This combined Kraken2-Bracken approach was chosen for its ability to accurately balance computational efficiency with taxonomic depth, which was critical when analyzing large, diverse datasets like those generated in this study. By leveraging

these tools, we were able to ensure both speed and accuracy in taxonomic classification, which was essential for subsequent comparative analyses with global urban microbiome datasets.

Comparative analysis against the global urban microbiome composition

To compare the microbiome profile of Chennai with those of other major global cities, we utilized data from the global metagenomic map of urban microbiomes provided by the MetaSUB consortium [4]. We performed a batch download of clean read FASTQ files from the MetaSUB central data repository using Geoseeq v0.2.8. These clean reads were preprocessed through the CAP2 pipeline, which included filtering out low-quality bases and removing host reads. In total, 2,515 samples from 38 cities were initially obtained based on data availability at the time of download. Using the associated metadata for each sample, we excluded those originating from air or biological sources, focusing on environmental samples. This filtering step resulted in a final dataset of 2,461 samples from 31 cities. The city with the fewest samples contributed 12, and the total number of samples used per city in our analysis is listed in Table S4. We conducted taxonomic classification analysis on the selected samples using Kraken2 and Bracken from the CAMP pipeline, similar to our approach with the Chennai city samples. This allowed us to identify the core and sub-core species. Subsequently, we compared these species across all cities with Chennai to detect any unique signatures specific to Chennai and to identify sample origin locations (Table S3).

Diversity analysis

To assess species richness (the variety of species) and evenness (their relative abundance), providing insights into microbial ecosystem differences across environments, we conducted diversity analysis using the phyloseq v1.44.0 [15] package in R. The phyloseq package integrates with the vegan package for calculating alpha diversity metrics, while ggpubr v0.6.0 [16] was used for visualization. Alpha Diversity (within-sample diversity) was quantified using the Shannon Index, which measures species distribution entropy. Higher values indicate greater diversity by accounting for both richness and evenness. We used the Wilcoxon test for pairwise comparisons and the Kruskal-Wallis test for group comparisons. The p-values from the pairwise comparisons were adjusted (P_{adj}) using the Benjamini-Hochberg (BH) method in R. Beta Diversity (between-sample diversity) was analyzed using the “ordinate” function in phyloseq, applying Bray–Curtis dissimilarity and Principal Coordinate Analysis (PCoA). Bray–Curtis dissimilarity accounts for taxonomic abundance, providing a sensitive measure for detecting differences between samples.

Differential taxonomic profiles across sample collection objects

To examine variations in microbial community composition across different sample collection objects, we performed a differential abundance analysis using MaAsLin2

v1.15.1 [17]. This analysis aimed to identify taxa that exhibit significant differences in abundance between samples collected from kiosks and other collection objects. MaAsLin2 applied a generalized linear model to assess differential abundance by fitting the linear model on relative abundance data. We set a significant threshold of log-fold change |LFC| > 1 and a $P_{adj} < 0.05$ to identify taxa with meaningful differences. This approach quantified and compared microbial profiles, highlighting key taxa potentially influenced by the collection environment.

Metagenomic assembly and genome reconstruction

To reconstruct microbial genomes from the metagenomic data, we first assembled the processed sequencing reads using MetaSPAdes v3.15.5 [18]. This produced a collection of contigs and scaffolds representing genomic fragments recovered from the samples. Following the assembly, MetaBAT2 v2.2.7 [19] was employed to bin metagenome-assembled genomes (MAGs). MetaBAT2 utilized two primary metrics to accurately cluster contigs— coverage depth per contig and tetranucleotide frequency (TNF) patterns. Coverage depth reflected the frequency of nucleotide sequencing, which was determined by aligning raw reads to the assembled contigs. TNF patterns, which represent the frequency of all possible four-nucleotide sequences within a contig, helped distinguish contigs belonging to different microbial genomes based on genomic composition. To evaluate the quality of the reconstructed MAGs, we used CheckM v1.0.1 [20]. This tool assesses genome completeness and contamination by analyzing lineage-specific marker genes, providing crucial insights into the reliability of the MAGs for downstream biological inferences.

Pangenome analysis

To explore the genetic diversity within the microbial population, we conducted a pangenome analysis, which compared multiple genomes to identify core genes shared across genomes and variable genes unique to individual genomes or subsets of genomes. We selected MAGs with completeness greater than 90% and contamination levels below 5%. Taxonomic annotations of the selected MAGs were performed using GTDB-Tk v2.1.1 [21], and reference genomes for identified organisms were retrieved from NCBI. We used Anvi'o v8 [18], a tool that facilitates visualization, gene annotation, functional annotation, and enrichment analysis across genomes using the database of Clusters of Orthologous Genes (COG2020 [22]). We additionally constructed a phylogenomic tree using GToTree v1.8.4 [18]. This tree was based on 15 single-copy gene Hidden Markov Models (HMMs) derived from the Pfam database and NCBI, providing insights into the evolutionary relationships among the genomes under study.

Accessing the gene abundance profile of the AMR gene family

To assess the antibiotic resistance potential of the microbial communities, present in the metro stations, we generated the antibiotic resistance profile of the samples using the Resistome Gene Identifier (RGI) tool v6.0.2 from the Comprehensive Antibiotic

Resistance Database (CARD) [23] database v3.2.6. RGI-bwt function from the RGI tool aligned the short read to CARD's protein homolog models using the k-mer alignment (KMA) read aligner and provided a table of identified drug classes, the annotation for the drug resistance mechanisms, AMR gene family, the number of reads mapped and the coverage. The read counts from RGI-bwt were normalized by the total reads per sample and gene length to calculate RPKM values, estimating the gene abundance levels of AMR genes across different classes of antibiotics based on CARD. We calculated the prevalence of AMR gene families in our sample set by identifying any AMR gene family present across various drug classes and resistance mechanisms. To reduce potential noise and focus on the gene abundance of AMR genes, we set a threshold at the 25th percentile (Q1) of RPKM. For a given drug class or drug resistance mechanism, the prevalence was 100% if at least one AMR gene family was present in all samples. Further, we also used the 2019 WHO AWaRe (Access, Watch, Reserve) classification framework to categorize antibiotic classes associated with the identified AMR genes.

RESULTS

Taxonomic classification and comparative analysis of distinct microbial signatures

We first investigated how microbial species are distributed throughout the environment of the Chennai metro stations (Figure 1). Specifically, we aimed to determine whether the urban environment exhibited a homogeneous microbial ecosystem or comprised distinct yet interconnected communities, particularly in terms of biodiversity. To achieve this, we examined the prevalence of species, defined as the fraction of samples in which a given taxon was found at different thresholds of relative abundance. We identified a right-skewed distribution characterized by a prominent primary peak on the left, representing the majority of species with low prevalence. Additionally, a smaller secondary peak on the right highlighted a subset of species with distinct characteristics or higher prevalence within the sampled ecosystem (Figure 2A). This allowed us to categorize the taxa into three distinct groups: 'core' taxa, present in the majority of samples; 'sub-core' taxa, present in 80-97% of samples; and 'peripheral' species, observed in fewer than 15% of samples. Table S3 summarizes the total number of identified species in Chennai samples across these distinct groups based on their relative abundance.

At a relative abundance of 0.001, which corresponded to more than 400 reads supporting a given species (based on the minimum library size shown in Table S2), we identified 27 core species, 33 sub-core species, and 395 peripheral species in the Chennai samples (Supplementary File 2, Sheet 1). Most samples contained species from three classes - *Actinobacteria, Alphaproteobacteria,* and *Gammaproteobacteria,* although the relative abundance of these classes varied (Figure 2B).

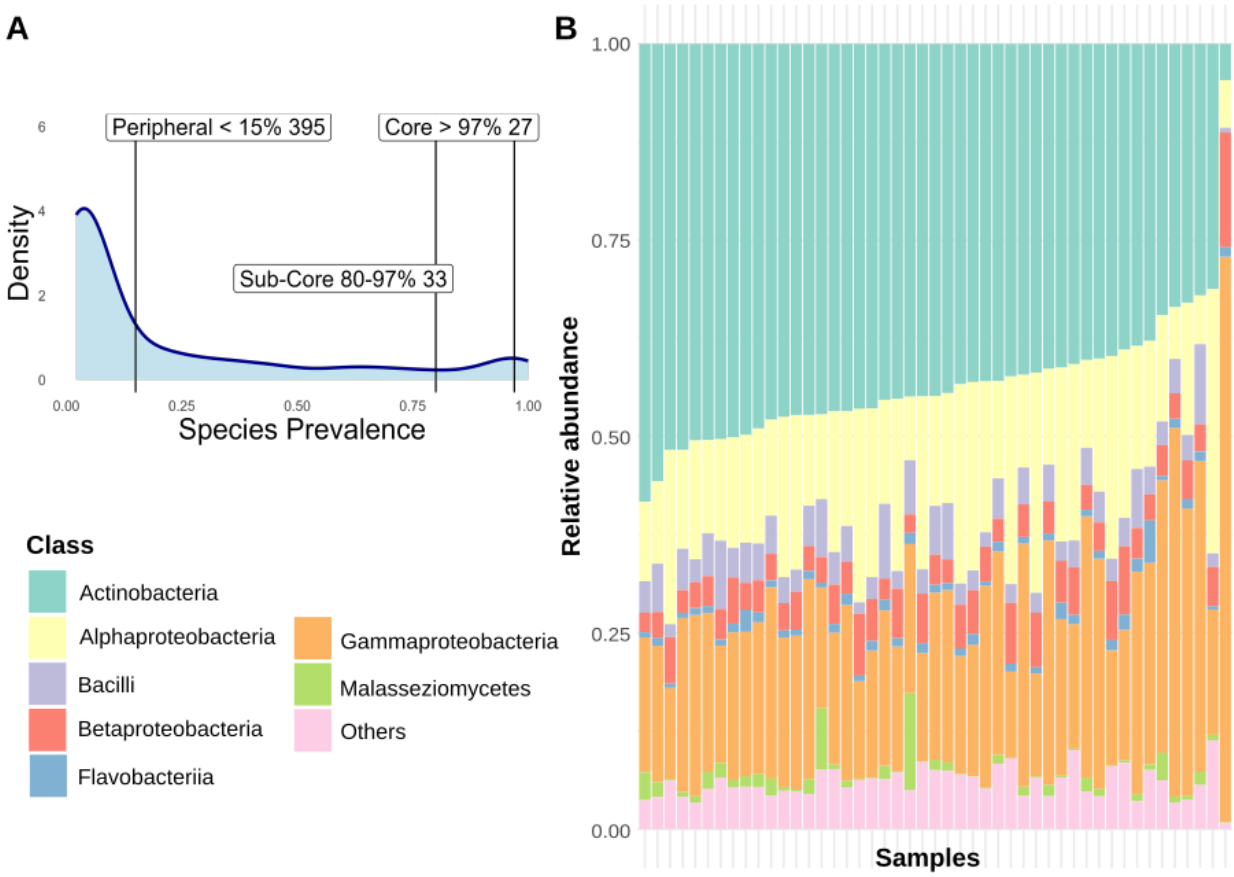

Figure 2: *(A) Prevalence of species at relative abundance 0.001 and (B) Relative abundance of bacterial classes across 47 Chennai samples.*

Geographical regions and urban environments significantly influence the composition and diversity of microbial communities, shaped by factors such as climate, air quality, human activities, and the built environment [36]. To explore these effects, we compared samples from Chennai with those from 31 other global cities (Table S4), aiming to identify common and unique patterns in species distribution. We identified microbial signatures unique to Chennai. To emphasize these, we created a category called 'unique-core,' which included core species specific to Chennai that were not core or sub-core in any other city at the same relative abundance. We found eight such species that were part of the core in Chennai but not in other global cities (Table 1, Table S3).

Among these, *Brachybacterium saurashtrense*, which is adapted to high salinity, was first isolated from the roots of the halophytic plant *Salicornia brachiate* [24]. Its presence indicated that the coastal climate and humidity of Chennai support its growth. Similarly, *Brachybacterium sp. SGAir0954*, previously isolated from tropical air in Singapore [25], thrived in Chennai's tropical and humid conditions. *Brachybacterium faecium* [25], typically linked to poultry environments, suggested agricultural activities near Chennai or the transportation of animal products. The presence of human-associated species like *Staphylococcus arlettae* [26–28], commonly found on the skin, highlighted the role of metro passengers and urban density in the microbial spread. Soil and water-associated species such as *Pseudomonas mendocina* [29], *Pseudomonas alcaligenes* [30], *Sphingobium yanoikuyae* [31] and *Comamonas aquatica* [32] likely reflected the local soil and water microbiome, demonstrating resilience to pollutants and potential for bioremediation in urban ecosystems. Overall, these findings illustrated how Chennai's tropical climate, coastal characteristics, dense human activity, and proximity to agricultural areas shaped the unique microbiome of its metro stations.

Microbial diversity and composition across various surface types

We evaluated the quantity and diversity of microbial communities across different surface types by analyzing both alpha and beta diversity. For alpha diversity, we used the Shannon-Wiener species diversity index, which accounts for species richness (the number of species present) and species evenness (the distribution of individuals among species). The samples collected from the kiosk exhibited a significantly high Shannon index (median = 5.97), indicating greater species diversity compared to other surfaces. In contrast, the samples from the rods showed the lowest Shannon index (median = 5.51), suggesting lower diversity (Figure 3A). For beta diversity, we performed Principal Coordinate Analysis (PCoA) using the Bray-Curtis dissimilarity measure to assess differences in species composition among samples from various surfaces. This analysis considered both the presence or absence and the relative abundances of species. As expected, the samples from the kiosk, which demonstrated the highest species diversity, clustered separately, indicating a distinct species composition compared to samples from other surfaces (Figure 3B).

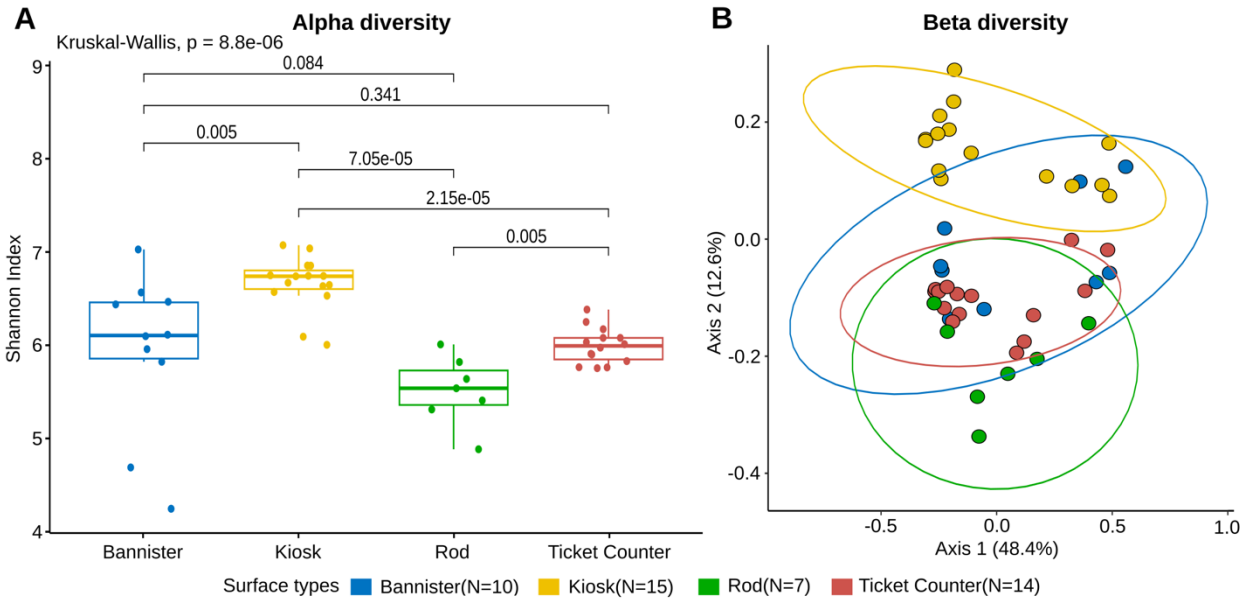

Figure 3: *(A) Alpha diversity (B) Beta diversity of Chennai samples collected from different surface types.*

We further examined the variations in microbial species across different surface types. We found that 19 species were significantly enriched on kiosk surfaces ($|\log_2 FC| > 1$ and $P_{adj} < 0.05$) compared to all other surfaces, while five species were significantly enriched on the rod surfaces (Table S5, S6). This finding aligned with our diversity analysis, where kiosk surfaces, which exhibited greater species diversity, also had more species significantly overrepresented. In contrast, no species were found to be significantly over- or underrepresented on the banister or ticket counter surfaces when compared to the other surface types (Figure 4).

Pangenome profiling of Metagenome-assembled Genomes (MAGs)

We investigated the genetic diversity and functional capabilities of microbial communities in the Chennai metro system through pangenome analysis of metagenome-assembled genomes (MAGs). From the samples, we reconstructed a total of 47 MAGs, each meeting the criteria of over 90% completeness and containing less than 5% contamination, qualifying them for pangenome analysis. All MAGs were mapped to the genus level, identifying 11 genera (Figure 5). Of these, 44 MAGs were further classified at the species level (Supplementary File 2, Sheet 2). We identified a total of 11 species with an average nucleotide identity (ANI) greater than 97% compared to reference genomes. The identified species included *Acinetobacter fasciculus, Acinetobacter junii, Brevibacterium paucivorans, Corynebacterium senegalense, Cutibacterium acnes, Epilithonimonas bovis, Kocuria marina, Micrococcus luteus, Moraxella osloensis, Stutzerimonas stutzeri,* and *Weissella confusa*. Among the identified species, *Cutibacterium acnes* was represented by 18 MAGs, *Stutzerimonas stutzeri* by 11 MAGs, and *Micrococcus luteus* by 7 MAGs, while the remaining species were represented by a single MAG each (Figure 5; Supplementary File 2, Sheet 2).

Our functional enrichment analysis using the COG2020 database identified a significant overrepresentation of certain functions in the MAGs from Chennai. Specifically, for *C. acnes*, we examined 18 MAGs from Chennai alongside 39 reference strains and found ten functions that were significantly overrepresented (Table 2). In the case of *S. stutzeri*, we analyzed 11 MAGs from Chennai and 20 reference strains, revealing four functionalities that were significantly overrepresented in the Chennai MAGs (Table 3). We did not identify any significant functional enrichment in the case of *M. luteus* MAGs.

The enrichment of multiple transposases among the *C. acnes* MAGs indicated that it possessed mobile genetic elements, which could facilitate genetic diversity and adaptability. These elements could help the organism acquire new traits or resistances, particularly in changing environments [34].

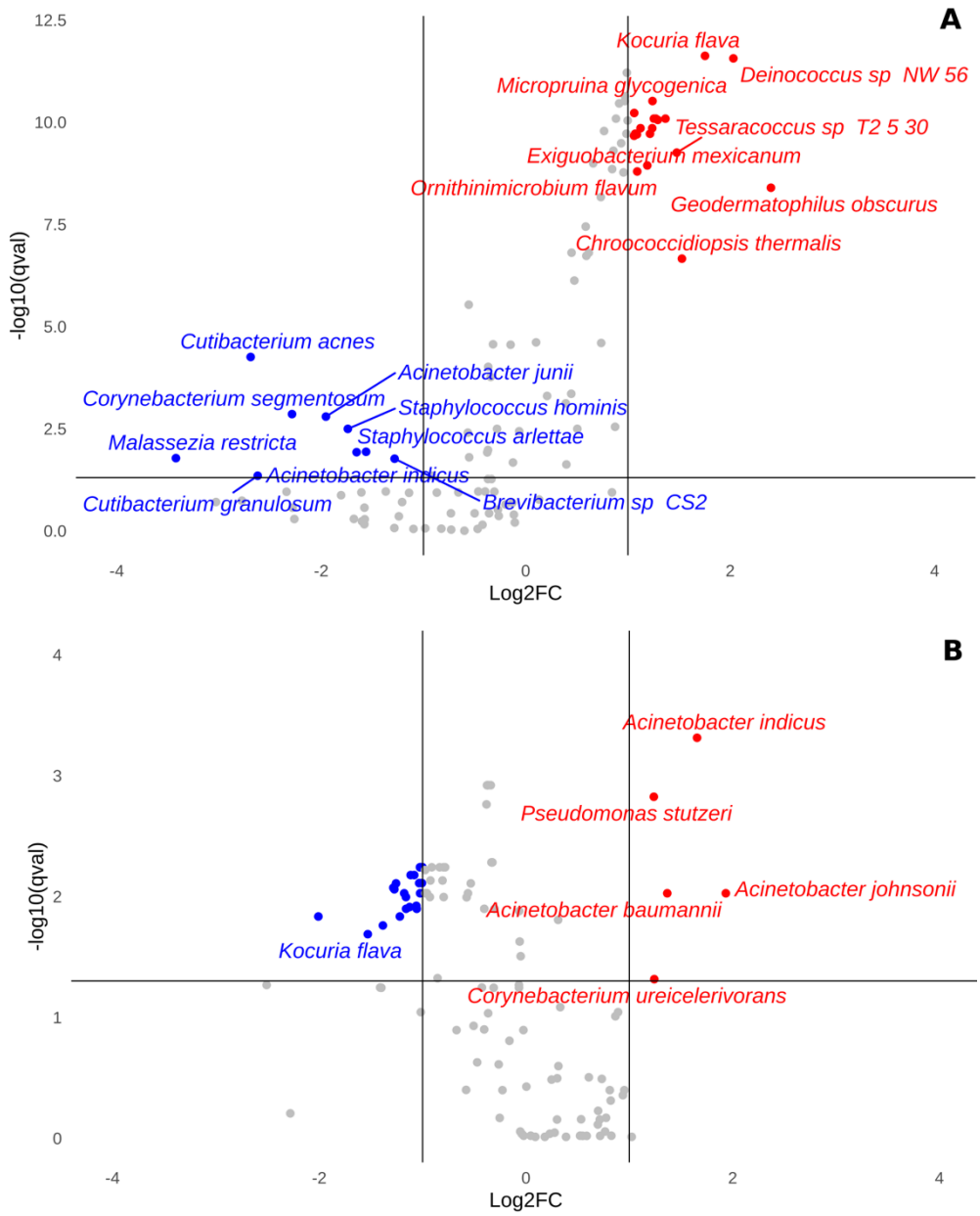

Figure 4: *Differential representation of microbial species across surface types. (A) Comparison of species diversity on kiosk surfaces versus other surfaces. (B) Comparison of species diversity on rod surfaces versus other surfaces. Species with* $log_2FC > 1$ *or* $log_2FC < -1$ *and* $P_{adj} < 0.05$ *are shown in red and blue, respectively.*

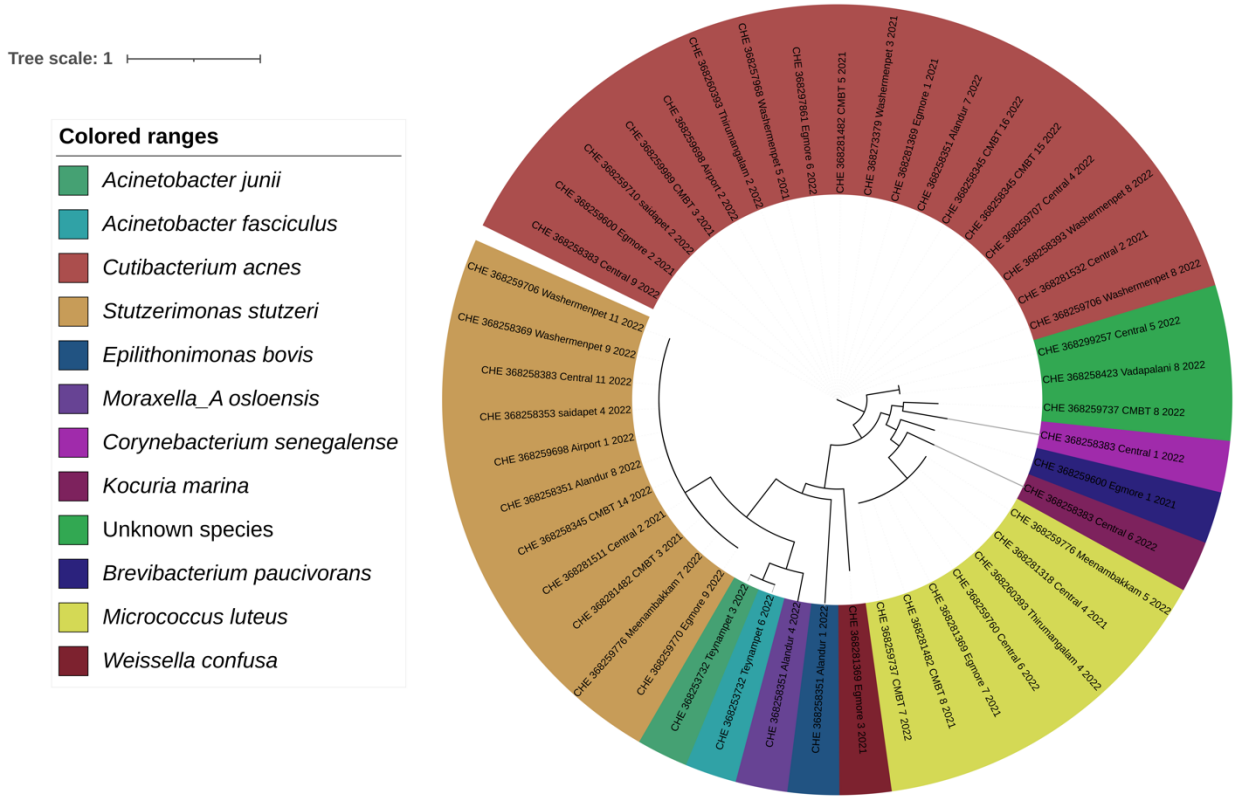

Figure 5: *Phylogenetic tree of the microbial species mapped to the MAGs.*

Assessment of antibiotic resistance gene abundance

We annotated 206 AMR gene families in our sample set, associated with 25 drug classes and 9 resistance mechanisms (Supplementary File 2; Sheet 3). We introduced a new class, "Multi-drug resistance," for genes that confer resistance to more than one antibiotic class. Overall, the AMR gene families exhibited a low gene abundance pattern (Average RPKM – mean: 4.18, median: 1.22) across different drug classes and the resistance mechanisms (Supplementary File 2; Sheet 4). However, certain antibiotic classes, namely, Sulfonamides and Rifamycins, had mean RPKM values of 11 and 19, respectively, exhibiting higher gene abundance compared to other drug classes (Figure 6A). According to the 2019 WHO AWaRe (Access, Watch, Reserve) classification framework [35], the Sulfonamide drug class is categorized as "Access," while Rifamycin is classified as "Watch." Interestingly, although the abundance of AMR classes was low, they were prevalent throughout our sample set, with only Bicyclomycin-like antibiotics, antimicrobial-free fatty acids, and Eifamycin antibiotics showing a prevalence of less than 40% (Figure 6B).

Furthermore, the gene abundance pattern of the AMR gene family associated with the "antibiotic target alteration and antibiotic target replacement" resistance mechanism was notably high compared to others (Figure 6C). Overall, the prevalence of all AMR gene families corresponding to the annotated resistance mechanisms in our sample set exceeded 50% (Figure 6D).

DISCUSSION

Urban public transit systems, such as metro networks, serve as critical hubs for microbial transmission, shaping the urban microbiome through constant human activity and high population density [36]. In densely populated regions like India, where millions rely on metro systems daily, these environments act as dynamic ecosystems where microbial communities interact, adapt, and spread. The Chennai Metro, operated by CMRL, has rapidly become a vital component of the city's urban transit infrastructure. Since its launch on June 29, 2015, the metro has transported an impressive 355.3 million passengers, highlighting its critical role in urban mobility [37]. In 2024 alone, CMRL recorded 105.2 million passengers, a substantial increase from 91.1 million in 2023. As metro ridership continues to grow, so does the importance of understanding its impact on microbial transmission and urban health.

This study presents the first comprehensive metagenomic analysis of an Indian transit system, revealing a unique microbial signature with eight core species distinct to Chennai. The presence of species like *Brachybacterium saurashtrense* and *Brachybacterium sp. SGAir0954* suggests a strong influence of Chennai's coastal climate and humidity, potentially indicating marine spillover effects. Given Chennai's location on the Coromandel Coast of the Bay of Bengal, this region has been historically affected by extreme weather events like the 2004 Indian Ocean tsunami [38]. Therefore,

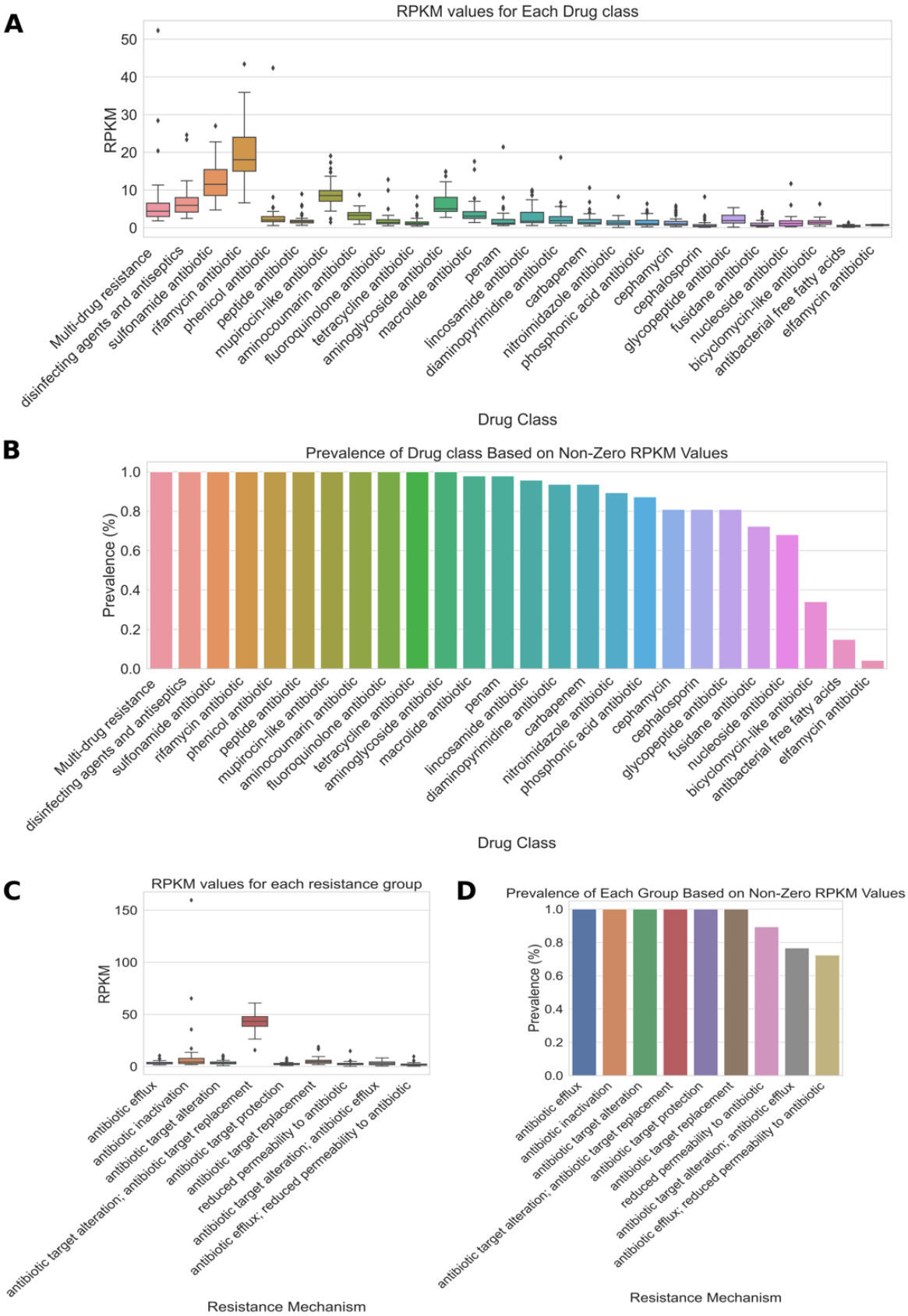

Figure 6: *Gene abundance profiles and prevalence of AMR gene families across drug classes (A, B) and resistance mechanisms (C, D).*

monitoring these microbial communities in cities like Chennai is crucial. The risk of increased flooding from sea-level rise [39] further underscores the need for continued surveillance. Our dataset provides insights into future urban microbiome shifts, emphasizing the importance of continued research.

Surface type significantly influenced microbial diversity, with kiosks exhibiting the highest alpha diversity, consistent with previous studies [40], likely due to frequent human contact. The concentrated usage of kiosks as focal points for user interaction by diverse individuals can create unique microbial communities, potentially resulting in higher alpha diversity despite their smaller size. In contrast, the larger surface areas of handrails may support a more homogenous microbial community due to constant exposure to human contact. Understanding these dynamics is crucial for developing effective cleaning protocols and public health strategies in urban environments.

We investigated the genetic diversity and functional potential of microbial communities in the Chennai metro system by reconstructing 47 high-quality MAGs. Among these, *Cutibacterium acnes*, *Stutzerimonas stutzeri*, and *Micrococcus luteus* were the most frequently represented species, suggesting their ecological dominance or adaptation to the urban transit environment. In contrast, several species were represented by a single MAG, indicating their potential niche specialization or low abundance within the ecosystem. Notably, *Moraxella osloensis*, an opportunistic pathogen commonly found in household environments, has been previously reported as a dominant species on kitchen surfaces and sponges, with its prevalence influenced by cleaning methods [41]. Similarly, *Acinetobacter junii*, frequently identified in house microbiomes in India, highlighting its persistence in built environments [41]. Among *C. acnes* MAGs, we observed an enrichment of transposases, suggesting the presence of mobile genetic elements that could promote genetic diversity and adaptability [34]. These elements may facilitate the acquisition of new traits or resistance mechanisms, particularly under environmental stress. Additionally, the overrepresentation of the multidrug transporter EmrE in *C. acnes* implies a potential mechanism for exporting toxic compounds, including antibiotics, thereby enhancing its survival under antimicrobial pressures. In *S. stutzeri* MAGs, we observed an overrepresentation of Dephospho-CoA Kinase (CoaE), an enzyme critical for coenzyme A (CoA) biosynthesis [42]. CoA is essential for various metabolic pathways, including fatty acid metabolism and acyl-CoA synthesis, underscoring the metabolic versatility of *S. stutzeri* in energy production and adaptation to environmental conditions.

Our AMR gene analysis revealed low overall abundance across most drug classes, yet their widespread presence indicates latent resistance potential. Genes associated with Sulfonamide and Rifamycin resistance exhibited relatively higher abundance, suggesting selective pressure. Sulfonamides act as broad-spectrum antibiotics by inhibiting dihydropteroate synthase [43], while Rifamycins target bacterial RNA polymerase [44]. Notably, the high prevalence of Rifamycin resistance aligns with reports of its presence

in Chennai hospital isolates, indicating possible overlap between clinical and urban microbial populations [45]. Furthermore, AMR genes were predominantly linked to "antibiotic target alteration" and "antibiotic target replacement," suggesting microbial adaptation to counteract the effects of antibiotics. By altering or replacing antibiotic targets, these organisms can evade drug actions, potentially leading to treatment failures.

Our findings underscore the importance of continuous microbial surveillance in urban public transportation systems, particularly in dense and rapidly growing cities like Chennai. As urbanization accelerates and public transport networks expand, the dynamics of microbial transmission in these densely populated environments will become increasingly relevant to public health. This aligns directly with the UN Sustainable Development Goals (SDGs), particularly SDG 3 (Good Health and Well-Being) and SDG 11 (Sustainable Cities and Communities), which emphasize the need for resilient urban infrastructures that promote health and sustainability. The integration of microbial surveillance into public health strategies is essential for achieving these goals, as it enables real-time monitoring of potential risks, such as AMR and the spread of infectious diseases.

Considering the growing emphasis on building smart cities, where technology-driven solutions are employed to enhance urban living, we posit that the use of real-time metagenomic analysis can play a pivotal role in making urban spaces healthier and more resilient. Smart cities are increasingly integrating IoT-based systems for air quality, traffic monitoring, and waste management, but microbial surveillance should also be a core component of these ecosystems. Future research should prioritize long-term monitoring of microbial communities in urban environments, enabling city planners to design more adaptive infrastructures that mitigate public health risks while promoting sustainable development.

The unique microbial signatures identified in Chennai's Metro further emphasize the need for region-specific strategies to address microbial transmission in public spaces. Comparing urban microbiomes across global cities could help policymakers develop targeted cleaning protocols, optimize public health interventions, and manage AMR more effectively. Future research should examine how human behavior, environmental factors, and urban design influence microbial community composition and use this knowledge to build more resilient and health-conscious urban infrastructure within the broader framework of sustainable and smart city development.

AVAILABILITY OF DATA AND MATERIALS

The raw reads are available on SRA under the accession number PRJNA1201944. The codes used in this paper are available in the following GitHub repository: https://github.com/IBSE-IITM/MetaSUB-CMRL. The rest of the data supporting this article is provided within the article and its online supplementary materials.

ACKNOWLEDGMENTS

We thank Chennai Metro Rail Limited (CMRL) and its employees at the respective stations for granting us the opportunity to collect samples from Chennai metro stations and trains for our research on the microbial landscape of the city. We also sincerely thank Mr. Santhanam Moorthyraj from the Dhan Foundation for facilitating our connection with CMRL. We thank Christopher Mason, Ph.D., David Danko, Ph.D., Braden Tierney, Ph.D., and Krista Ryon, Weill Medical College of Cornell University, New York, for their advice, support in the planning, and providing kits for sample collection.

COMPETING INTERESTS

The authors declare no competing interests.

FUNDING

The authors acknowledge funding from the Robert Bosch Centre for Data Science and AI (SB/20-21/0602/BTRBCX/008481), IIT Madras to K.R. and H.S., and the Wadhwani School of Data Science and AI, IIT Madras.

AUTHOR CONTRIBUTIONS

K.R., H.S., B.P. conceived the study; K.R., H.S. acquired funding and supervised the study; V.Y.S., V.P.G. pre-processed the data and established the analysis pipeline; V.Y.S., V.P.G., S.S., and R.L. collected samples and performed the analysis; V.P.G., V.Y.S., S.S. wrote the first draft of the manuscript; V.P.G., V.Y.S., S.S., K.R., H.S. edited the manuscript; all authors reviewed and commented the manuscript.

Table 1: *Eight unique-core species of Chennai city and the habitats in which they are typically found.*

Species	Sources
Brachybacterium sp. SGAir0954	Isolated from tropical air collected in Singapore [33]
Brachybacterium faecium	Isolated from poultry deep litter [25]
Brachybacterium saurashtrense	Isolated from the roots of *Salicornia brachiata*, an extreme halophyte [24]
Pseudomonas mendocina	Found in water and soil samples [29]
Pseudomonas alcaligenes	Commonly found in soil and water [30]
Staphylococcus arlettae	Found in both environment and human source [26–28]
Sphingobium yanoikuyae	Found in environments, ranging from terrestrial to aqueous habitats including sea water [31]
Comamonas aquatica	Found in water source [32]

Table 2: *Significantly overrepresented functions in Cutibacterium acnes MAGs.*

COG2020 FUNCTION	P_{adj}	log_2FC
Ribosomal protein L36 (RpmJ) (PDB:1DFE) (PUBMED:21627333)	1.5E-07	4.98
Transposase InsO and inactivated derivatives (Tra5)	4.6E-07	3.98
Transposase and inactivated derivatives, IS30 family (Tra8)	6.7E-03	3.66
Transposase InsE and inactivated derivatives (InsE) (PDB:2JN6) (PUBMED:23818957)	6.7E-03	3.66
Translation initiation factor IF-1 (InfA) (PDB:5FIM)	4.1E-07	4.88
Acyl-CoA thioesterase PaaI, contains HGG motif (PaaI) (PDB:1PSU)	6.7E-03	3.66
CRISPR/Cas system-associated endoribonuclease Cas2 (Cas2) (PDB:2I0X)	1.5E-02	3.40
Multidrug transporter EmrE and related cation transporters (EmrE) (PDB:2I68)	4.0E-02	3.07
Ca2+/Na+ antiporter (ECM27) (PDB:3V5S)	4.0E-02	3.07
AraC-type DNA-binding domain and AraC-containing proteins (AraC) (PDB:1BL0)	4.0E-02	3.07

Table 3: *Significantly overrepresented functions in Stutzerimonas stutzeri MAGs.*

COG2020 FUNCTION	P_{adj}	log_2FC
Dephospho-CoA kinase (CoaE) (PDB:1JJV)	5.0E-05	4.27
Protein translocase subunit SecG (SecG) (PDB:2AKH)	1.9E-04	4.13
Ribosomal protein S27E (RPS27A) (PDB:1QXF)	4.2E-02	3.13
TolB amino-terminal domain (function unknown) (TolBN) (PDB:1C5K) (PUBMED:10673426;24821816)	4.2E-02	3.13
Ribosomal protein L36 (RpmJ) (PDB:1DFE) (PUBMED:21627333)	2.5E-03	2.98

REFERENCES

1. Hernández AM, Vargas-Robles D, Alcaraz LD, Peimbert M. Station and train surface microbiomes of Mexico City's metro (subway/underground). Sci Rep. 2020;10:8798.
2. Gohli J, Bøifot KO, Moen LV, Pastuszek P, Skogan G, Udekwu KI, et al. The subway microbiome: seasonal dynamics and direct comparison of air and surface bacterial communities. Microbiome. 2019;7:160.
3. Ryon KA, Tierney BT, Frolova A, Kahles A, Desnues C, Ouzounis C, et al. A history of the MetaSUB consortium: Tracking urban microbes around the globe. iScience. 2022;25:104993.
4. Danko D, Bezdan D, Afshin EE, Ahsanuddin S, Bhattacharya C, Butler DJ, et al. A global metagenomic map of urban microbiomes and antimicrobial resistance. Cell. 2021;184:3376-3393.e17.
5. Chennai, India Population 2024. https://worldpopulationreview.com/cities/india/chennai. Accessed 29 Sep 2024.
6. Moriel DG, Piccioli D, Raso MM, Pizza M. The overlooked bacterial pandemic. Semin Immunopathol. 2024;45:481–91.
7. Wensel CR, Pluznick JL, Salzberg SL, Sears CL. Next-generation sequencing: insights to advance clinical investigations of the microbiome. J Clin Invest. 2022;132.
8. Sengupta P, Sivabalan SKM, Mahesh A, Palanikumar I, Kuppa Baskaran DK, Raman K. Big data for a small world: A review on databases and resources for studying microbiomes. J Indian Inst Sci. 2023;:1–17.
9. MetaSUB International Consortium. The Metagenomics and Metadesign of the Subways and Urban Biomes (MetaSUB) International Consortium inaugural meeting report. Microbiome. 2016;4:24.
10. Rehman S. A parallel and silent emerging pandemic: Antimicrobial resistance (AMR) amid COVID-19 pandemic. J Infect Public Health. 2023;16:611–7.
11. Murray AK. The Novel Coronavirus COVID-19 Outbreak: Global Implications for Antimicrobial Resistance. Front Microbiol. 2020;11:1020.
12. Resources - MetaSUB. https://metasub.org/resources/. Accessed 29 Sep 2024.
13. File:Chennai Metro map.svg - Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Chennai_Metro_map.svg. Accessed 30 Sep 2024.
14. Mak L, Tierney B, Ronkowski C, Brizola Toscan R, Turhan B, Toomey M, et al. CAMP: A modular metagenomics analysis system for integrated multi-step data exploration. BioRxiv. 2024.
15. McMurdie PJ, Holmes S. Phyloseq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data. Pac Symp Biocomput. 2012;:235–46.w

16. ggplot2 Based Publication Ready Plots • ggpubr. https://rpkgs.datanovia.com/ggpubr/. Accessed 6 Oct 2024.
17. Mallick H, Rahnavard A, McIver LJ, Ma S, Zhang Y, Nguyen LH, et al. Multivariable association discovery in population-scale meta-omics studies. PLoS Comput Biol. 2021;17:e1009442.
18. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ. 2015;3:e1319.
19. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7:e7359.
20. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.
21. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2019;36:1925–7.
22. Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, Koonin EV. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Res. 2021;49:D274–81.
23. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother. 2013;57:3348–57.
24. Gontia I, Kavita K, Schmid M, Hartmann A, Jha B. Brachybacterium saurashtrense sp. nov., a halotolerant root-associated bacterium with plant growth-promoting potential. Int J Syst Evol Microbiol. 2011;61 Pt 12:2799–804.
25. Lapidus A, Pukall R, Labuttii K, Copeland A, Del Rio TG, Nolan M, et al. Complete genome sequence of Brachybacterium faecium type strain (Schefferle 6-10). Stand Genomic Sci. 2009;1:3–11.
26. Kherdekar RS, Dixit A, Kothari A, Pandey KP, Advani H, Gaurav A, et al. Unusually isolated Staphylococcus arlettae in intra-oral sutures - Case series. Access Microbiol. 2023;5.
27. Dinakaran V, Shankar M, Jayashree S, Rathinavel A, Gunasekaran P, Rajendhran J. Genome sequence of Staphylococcus arlettae strain CVD059, isolated from the blood of a cardiovascular disease patient. J Bacteriol. 2012;194:6615–6.
28. Nanjani SG, Soni HP. Characterization of an extremely halotolerant Staphylococcus arlettae HPSSN35C isolated from Dwarka Beach, India. J Basic Microbiol. 2014;54:843–50.
29. Gani M, Rao S, Miller M, Scoular S. Pseudomonas mendocina bacteremia: A case study and review of literature. Am J Case Rep. 2019;20:453–8.

30. Suzuki M, Suzuki S, Matsui M, Hiraki Y, Kawano F, Shibayama K. Genome Sequence of a Strain of the Human Pathogenic Bacterium Pseudomonas alcaligenes That Caused Bloodstream Infection. Genome Announc. 2013;1.
31. Gu J-G, Han B, Duan S, Zhao Z, Wang Y. Degradation of the endocrine-disrupting dimethyl phthalate carboxylic ester by Sphingomonas yanoikuyae DOS01 isolated from the South China Sea and the biochemical pathway. Int Biodeterior Biodegradation. 2009;63:450–5.
32. Dai W, Zhu Y, Wang X, Sakenova N, Yang Z, Wang H, et al. Draft Genome Sequence of the Bacterium Comamonas aquatica CJG. Genome Announc. 2016;4.
33. Hlaing PPT, Junqueira ACM, Uchida A, Purbojati RW, Houghton JNI, Chénard C, et al. Complete genome sequence of *Brachybacterium* sp. Strain SGAir0954, isolated from Singapore air. Microbiol Resour Announc. 2019;8.
34. Aoki S, Nakase K, Hayashi N, Noguchi N. Transconjugation of erm(X) conferring high-level resistance of clindamycin for Cutibacterium acnes. J Med Microbiol. 2019;68:26–30.
35. Zanichelli V, Sharland M, Cappello B, Moja L, Getahun H, Pessoa-Silva C, et al. The WHO AWaRe (Access, Watch, Reserve) antibiotic book and prevention of antimicrobial resistance. Bull World Health Organ. 2023;101:290–6.
36. Afshinnekoo E, Meydan C, Chowdhury S, Jaroudi D, Boyer C, Bernstein N, et al. Geospatial Resolution of Human and Bacterial Diversity with City-Scale Metagenomics. Cell Syst. 2015;1:72–87.
37. CMRL - Welcome to Chennai Metro Rail. https://chennaimetrorail.org/press-release/. https://chennaimetrorail.org/wp-content/uploads/2025/01/Press-Release-02.01.2025-English.pdf
38. Srinivasalu S, Thangadurai N, Switzer AD, Ram Mohan V, Ayyamperumal T. Erosion and sedimentation in Kalpakkam (N Tamil Nadu, India) from the 26th December 2004 tsunami. Mar Geol. 2007;240:65–75.
39. Nicholls RJ, Marinova N, Lowe JA, Brown S, Vellinga P, de Gusmão D, et al. Sea-level rise and its possible impacts given a “beyond 4°C world” in the twenty-first century. Philos Transact A Math Phys Eng Sci. 2011;369:161–81.
40. Hsu T, Joice R, Vallarino J, Abu-Ali G, Hartmann EM, Shafquat A, et al. Urban Transit System Microbial Communities Differ by Surface Type and Interaction with Humans and the Environment. mSystems. 2016;1.
41. Awasthi S, Hiremath VM, Nain S, Malik S, Srinivasan V, Rose P, et al. Microbial landscape of Indian homes: the microbial diversity, pathogens and antimicrobial resistome in urban residential spaces. Environ Microbiome. 2025;20:25.
42. Shimosaka T, Makarova KS, Koonin EV, Atomi H. Identification of Dephospho-Coenzyme A (Dephospho-CoA) Kinase in Thermococcus kodakarensis and Elucidation of the Entire CoA Biosynthesis Pathway in Archaea. MBio. 2019;10.
43. Ovung A, Bhattacharyya J. Sulfonamide drugs: structure, antibacterial property, toxicity, and biophysical interactions. Biophys Rev. 2021;13:259–72.

44. Adams RA, Leon G, Miller NM, Reyes SP, Thantrong CH, Thokkadam AM, et al. Rifamycin antibiotics and the mechanisms of their failure. J Antibiot. 2021;74:786–98.
45. Murugan K, Kavitha K, Al-Sohaibani S. Rifampicin resistance among multi-resistant MRSA clinical isolates from Chennai, India, and their molecular characterization. Genet Mol Res. 2015;14:2716–25.