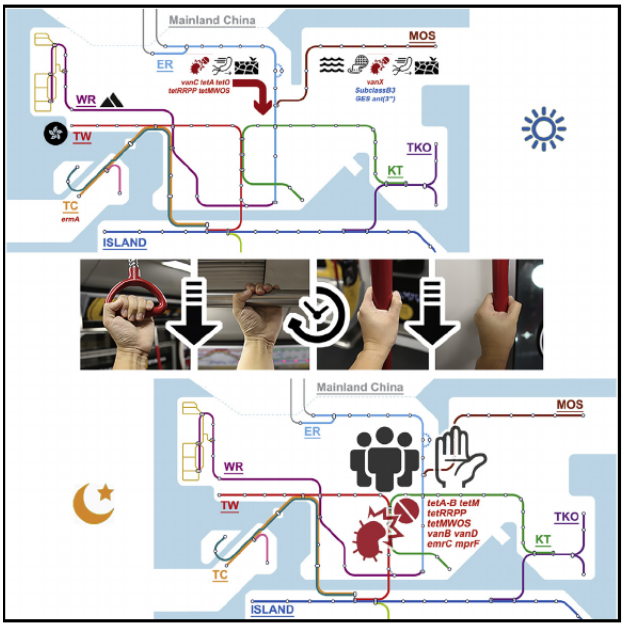
Article

Cell Reports

The Environmental Exposures and Inner- and Intercity Traffic Flows of the Metro System May Contribute to the Skin Microbiome and Resistome

Graphical Abstract



Authors

Kang Kang, Yueqiong Ni, Jun Li, ..., Morten Otto Alexander Sommer, Christopher John Webster, Gianni Panagiotou

Correspondence

gianni.panagiotou@hki-jena.de

In Brief

Kang et al. present a metagenomic analysis of the Hong Kong metro system. They show a reliable and dynamic view of the diurnal flux of microbial transmission and recurrence affected by the traffic flow and establish a baseline for metagenomic studies examining human interactions with public transit.

Highlights

- Intraday time is the primary determinant of metro microbiome and resistome composition
- Human commensals and clinical ARGs are of higher relative abundance in p.m. samples
- Line-specific environmental exposure also modulates the microbial community
- Potential cross-boundary ARG transmission event was captured

Kang et al., 2018, Cell Reports *24*, 1190–1202
July 31, 2018
https://doi.org/10.1016/j.celrep.2018.06.109

CellPress

OPEN ACCESS CellPress

Cell Reports
Article

The Environmental Exposures and Inner- and Intercity Traffic Flows of the Metro System May Contribute to the Skin Microbiome and Resistome

Kang Kang,[1] Yueqiong Ni,[1] Jun Li,[1] Lejla Imamovic,[2] Chinmoy Sarkar,[3] Marie Danielle Kobler,[4] Yoshitaro Heshiki,[1] Tingting Zheng,[1] Sarika Kumari,[3] Jane Ching Yan Wong,[5] Anand Archna,[5] Cheong Wai Martin Wong,[5] Caroline Dingle,[6] Seth Denizen,[7] David Michael Baker,[5] Morten Otto Alexander Sommer,[2] Christopher John Webster,[3] and Gianni Panagiotou[1,8,*]

[1]Systems Biology and Bioinformatics Group, School of Biological Sciences, Faculty of Sciences, The University of Hong Kong, Hong Kong S.A.R., China
[2]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[3]Healthy High Density Cities Lab, HKUrbanLab, The University of Hong Kong, Hong Kong S.A.R., China
[4]School of Public Health, The University of Hong Kong, Hong Kong S.A.R., China
[5]School of Biological Sciences, Faculty of Science, The University of Hong Kong, Hong Kong S.A.R., China
[6]Department of Earth Sciences, The University of Hong Kong, Hong Kong S.A.R., China
[7]Division of Landscape Architecture, The University of Hong Kong, Hong Kong S.A.R., China
[8]Lead Contact
*Correspondence: gianni.panagiotou@hki-jena.de
https://doi.org/10.1016/j.celrep.2018.06.109

SUMMARY

The skin functions as the primary interface between the human body and the external environment. To understand how the microbiome varies within urban mass transit and influences the skin microbiota, we profiled the human palm microbiome after contact with handrails within the Hong Kong Mass Transit Railway (MTR) system. Intraday sampling time was identified as the primary determinant of the variation and recurrence of the community composition, whereas human-associated species and clinically important antibiotic resistance genes (ARGs) were captured as p.m. signatures. Line-specific signatures were notably correlated with line-specific environmental exposures and city characteristics. The sole cross-border line appeared as an outlier in most analyses and showed high relative abundance and a significant intraday increment of clinically important ARGs (24.1%), suggesting potential cross-border ARG transmission, especially for tetracycline and vancomycin resistance. Our study provides an important reference for future public health strategies to mitigate intracity and cross-border pathogen and ARG transmission.

INTRODUCTION

The skin is not merely a physical barrier between the body and the environment but functions as a complex contact interface for a rich and influential microbial community. Indeed, the human skin and its microbial residents play an essential role in health and disease. The human skin microbiome has been shown to have a unique signature and to maintain stable characteristics at the strain level over time (Oh et al., 2016). Despite this resistance to colonization and disruption, the skin is a source of potential infection, especially in vulnerable individuals. For instance, various *Propionibacterium acnes* strains were identified as potential determinants for acne, the most common skin condition (Fitz-Gibbon et al., 2013). Skin from over 90% of atopic dermatitis (AD) patients is highly colonized with the pathogen *Staphylococcus aureus* (Leyden et al., 1974), whereas *Staphylococcus epidermidis* was found to modulate the host immune response to inhibit *S. aureus* (Cogen et al., 2010a, 2010b). The microbial composition of the human skin can be affected by our lifestyle choices, including cohabitants, dwelling environment, and even our pets (Lax et al., 2014). In high-density megacities, mass transit systems also serve as platforms for the exchange of microbes between millions of humans and the built environment (Hong Kong's Mass Transit Railway [MTR] has a daily ridership of around 4.7 million, for example; MTR Corporation, 2017), induced by the high mobility of the population and constant interaction through skin-surface contact. Mass transit systems therefore represent ideal targets for studying the composition and dynamics of urban microbial communities. They constitute passive samplers with unique niches and a potentially higher propensity for microbial accumulation, exposure, and transmission, created through large concentrations of people from diverse urban backgrounds interacting in relatively confined spaces (Be et al., 2015).

To characterize the urban metro system microbiota, several 16S rRNA-based studies have been conducted in high-density cities, including New York and Hong Kong (Leung et al., 2014; Robertson et al., 2013). A recent shotgun metagenomic study performed in the New York metro system characterized the citywide urban microbial geographic distribution using samples collected from station surfaces (Afshinnekoo et al., 2015). This study reported potential health risks associated with the

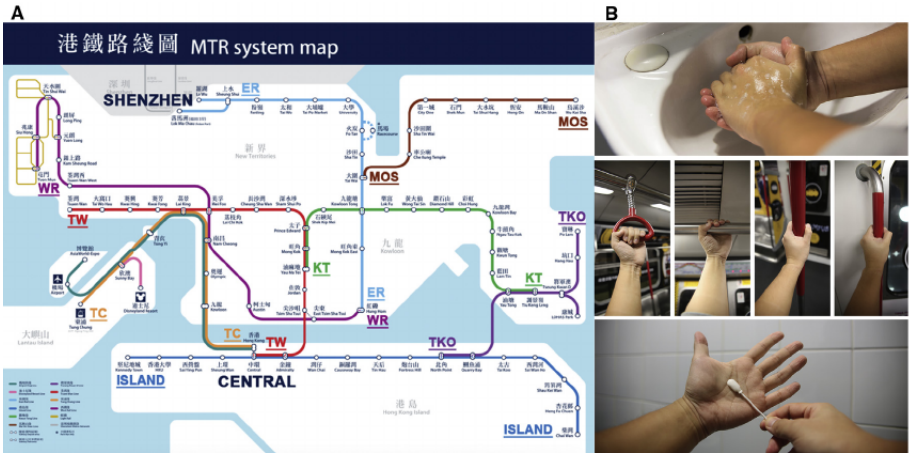


Figure 1. The MTR System and Sampling Procedure

(A) The MTR system and eight urban lines used in this study (excluding the Airport Express line and Disneyland Resort branch). The Central-Hong Kong station and the cross-border rail stations connecting with the MTR and the Shenzhen metro system are labeled.

(B) The sampling procedure included handwashing, handrail touching for 30 min, swabbing, and sample preservation.

microbial communities found in the stations, including the presence of pathogens and antibiotic resistance as well as constant daily recurrence in the vast bacterial ecology. A similar study of the Boston metro system (Massachusetts Bay Transportation Authority [MBTA]) highlighted that different surface types and materials are occupied by different microbes with high variation in functional capacity and pathogenic potential (Hsu et al., 2016). Recently, more studies of metro and urban microbiomes have started under the collaborative framework of the Metagenomics and Metadesign of the Subways and Urban Biomes (MetaSUB) International Consortium (MetaSUB International Consortium, 2016). Previous metagenomic studies of metro systems were conducted by sampling directly from surfaces or indoor air masses. However, to obtain a more focused picture of how microbial communities occupying the surfaces in a metro system may affect human health, studies focusing on the contact interface—by sampling the skin microbiome—are of significant interest. Ultimately, a key objective of performing microbial profiling in a transit system is to identify possible health-related risks, as evidence has accumulated of the transmission of clinically significant disease from surface to hand (Bhalla et al., 2004; Weber et al., 2010) as well as community-acquired antibiotic-resistant infections (Ho et al., 2007; Spellberg et al., 2008; Vandenesch et al., 2003). Therefore, characterizing shifts in the microbial consortia an individual carries after using mass transit is a more direct investigation than intensive sampling of accumulations and communities on built surfaces (Gibbons et al., 2015; Prescott et al., 2017; Zimmerman et al., 2014). Furthermore, because mass transit is characterized by intermittent passenger loads (fluctuating over peak and off-peak hours) and inherent tidal effects between downtown and uptown regions, time series-based studies with distinct sampling times throughout the day and repeated sampling days offer a more reliable and dynamic view of the diurnal flux of microbial transmission and the propensity of recurrence induced by the variation in traffic flow. Finally, investigating the correlations between microbial composition and the underlying geographical or topological characteristics of different metro lines as well as their connectivity to the underlying urban matrix (street networks, buildings, and public spaces) is crucial because these connections govern human mobility and interactions in high-density environments and, hence, influence the rate of microbial contact and transmission (Adams et al., 2015). Several interesting questions that may inform public health surveillance and urban planning were raised. What is the composition of the microbiome and its potential health risk (e.g., antibiotic resistance and pathogenicity) we may acquire when exposed to mass transit? Will the microbiome inherit tidal effects of the traffic flow and vary over traveling time? How might urban density and morphology influence the microbial composition and health risk it carries? To what extent is a city's unique metro microbiome influenced by connecting routes to a neighboring metropolitan region or country?

To answer these questions, we conducted a metagenomic study of the Hong Kong MTR system (Figure 1A) with a cross-border rail connection to the neighboring large city of Shenzhen in mainland China. Shotgun metagenome sequencing followed by taxonomic and functional analysis was performed on samples collected from the surface of human palms after contact with metro railcar handrails (Figure 1B). Our sampling covered different metro lines and was conducted at different time points across the day with weekly replicates. To determine whether the microbiome and resistome vary over the day and differ by distinct metro lines, we looked for time-specific and line-specific signature species and antibiotic resistance genes (ARGs),

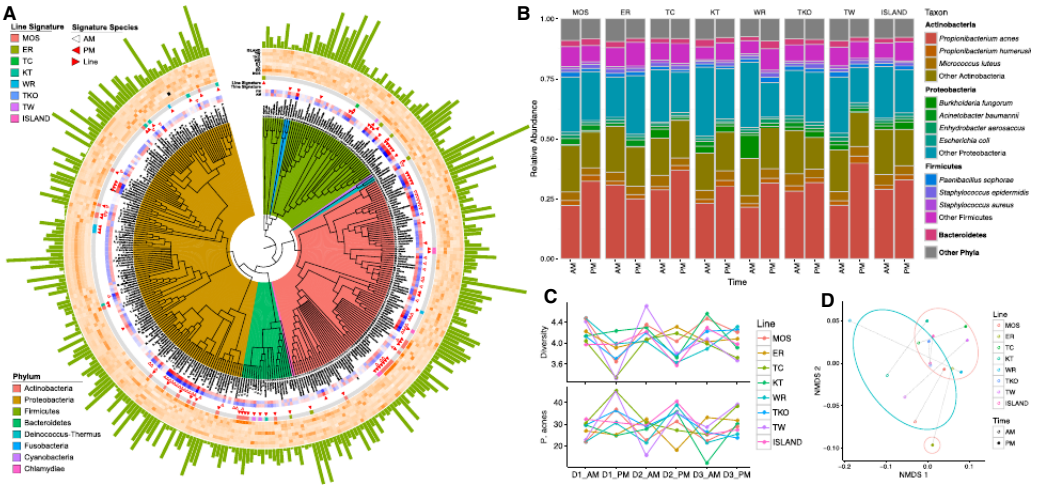


Figure 2. Taxonomy, Phylogeny, Community Diversities, and Dissimilarities of the MTR Microbial Communities

(A) The phylogeny and relative abundance of the most abundant species at species level. The phylogenetic tree was acquired from OpenTree 7.0 (Hinchliff et al., 2015), with phyla indicated by the clade colors. The inner and outer heatmaps illustrate the relative abundance of the species for a.m. or p.m. samples and different MTR lines, respectively. Triangle signs indicate whether the species is a time- or line-specific signature species. Different colors in the colored bar demonstrate different lines for the line-specific signatures. The outermost bars represent the absolute overall abundance of the species after log transformation.

(B) Relative abundances of highly abundant bacterial phyla and their major species across different sampling times and lines.

(C) The variations of community alpha-diversities (Simpson's index) in comparison with the relative abundance of the dominant species *P. acnes* over sampling time points.

(D) The community dissimilarities of the communities from different lines and sampling times, calculated by weighted UniFrac distances, with coordinates calculated by nonmetric multidimensional scaling (NMDS). Means of species abundances from all sampling days were calculated before the dissimilarity calculation.

respectively, and then we attempted to track potential transmission events of ARGs. Further, to reveal how urban design could influence microbial ecology and biosafety issues, we investigated the correlations among metro network topology, ecological features (e.g., community diversity), and health risk levels, such as clinically important ARG abundance, ARG dissemination potential, and pathogenic potential. Our study provides a useful framework for obtaining a comprehensive view of mass transit systems as modulators of human microbial contacts, informing public health planning regarding monitoring and health risk management.

RESULTS

Microbial Communities Acquired by MTR Exposure

Taxonomic profiles of the relative abundance were generated at each taxonomic level for each sample (Data S1A and S1B). Most bacterial reads were derived from four major bacterial phyla: *Actinobacteria*, 51.3% ± 8.5%; *Proteobacteria*, 26.6% ± 9.3%; *Firmicutes*, 11.4% ± 3.0%; *Bacteroidetes*, 2.3% ± 0.7%; and 8.4% ± 1.0% for others (Figure 2A). *P. acnes* was the dominant organism, accounting for 29.1% ± 7.0% of the community (Figure 2B). In the 10 most populated species, 8 were common human and skin commensals, including *P. acnes*, *Micrococcus luteus*, *Propionibacterium humerusii*, *Acinetobacter baumannii*, *S. epidermidis*, *Escherichia coli*, and *S. aureus*. The yeast *Malassezia globosa* (0.72%), *Enterobacteria* phages (0.28%), and *Propionibacterium* phages (0.13% in 27 subtypes, including 8 abundant subtypes) are the most abundant non-bacterial organisms. Not surprisingly, the population of *Propionibacterium* phages is positively correlated with the abundance of *Propionibacterium* (Pearson's correlation coefficient (r) = 0.72, p = 6.7e−9) and 11 *Propionibacterium* species, including *P. acnes* (r = 0.71, q = 1.7e−6). All 8 types of abundant *Propionibacterium* phages co-occurred with their hosts at both genera and species levels. To achieve sufficient resolution for analyses of phage-host interactions and pathogenicity, strain-level taxonomic construction was performed for two major species, *P. acnes* and *S. epidermidis*, with well-documented reference strains using a reference-based approach (Oh et al., 2014). In the 89 strains quantified as *P. acnes*, C1 was identified as the leading strain (making up 12.6% ± 3.2% of the species) in 44 of the 48 samples, followed by PA2 (6.4% ± 0.6%) and P.acn17 (5.5% ± 1.8%) (Figure S1A; Data S1C). Notably, we observed that different *Propionibacterium* phages showed drastic variability in their predilection to multiple *P. acnes* strains. In general, the overall abundance of *Propionibacterium* phages is correlated with 24 *P. acnes* strains (r = 0.47∼0.72, q < 0.1). Specifically, phage ATCC29399B_C shows correlation with 36 strains, followed by PHL037M02, P1.1 and P100_1, correlated with 15, 14, and

13 strains, respectively, whereas ATCC29399B_T, P100_A, PHL010M04, and P100D have only 5, 3, 1, and 0 co-occurring strains, respectively. No strain was significantly correlated with more than one phage subtype, suggesting strong phage-host preferences at the strain level. For *S. epidermidis*, with 91 quantified strains, UC7032 predominates, with a relative abundance of 67.0% ± 16.9% (Figure S1B; Data S1D). To explore the gene content carried by strain diversities, pan-genome profiling was performed for the 10 most abundant species using PanPhlan (Scholz et al., 2016). Gene cluster compositions for most species were commonly shared among samples. By using a pathogen list compiled by different sources (Forsberg et al., 2014; Kembel et al., 2012), we discovered three opportunistic pathogens at very low abundance in the MTR communities: *Helicobacter pylori*, *Acinetobacter* species (sp.) ADP1, and *Photorhabdus asymbiotica* (total relative abundances of 0.004% ± 0.001%). At the strain level, two reported opportunistic pathogenic isolates, *P. acnes* KPA171202 and *S. epidermidis* RP62A, were observed with low relative abundance (2.6% ± 1.2% and 0.08% ± 0.06% of the *P. acnes* and *S. epidermidis* populations, respectively). The virulence factors (VFs) from these species have shown much higher copy numbers than their genome background (Data S2C), especially the VF category Urease (CVF221) from *H. pylori* (350 to the genome). This could be a result of statistical bias introduced by the difference in marker selection between taxonomic profiling and VF annotation but could also potentially suggest a much wider distribution of the virulence genes in the community compared with the abundance of their putative host. The different VF categories from *H. pylori* also showed higher variance (0.03 to 1.89 reads per kilobase per million reads [RPKM], respectively) compared with the VFs from *S. epidermidis* RP62A (1.00 to 1.55 RPKM), suggesting that the strain-level diversity of *H. pylori* might dramatically influence pathogenicity. Nevertheless, lowest common ancestor (LCA) mapping was not suitable for strain-level pathogen quantification for other taxa because of limitations of subspecies-level phylogenetic resolution. Therefore, we marked species containing at least one reported pathogenic strain as potential opportunistic pathogenic species, led by human commensals, including *P. acnes*, *A. baumannii*, *S. epidermidis*, *E. coli*, and *S. aureus*, summing up to 40.13% ± 6.25% of the whole community (and 11.01% ± 2.36% when excluding *P. acnes*) (Data S2B).

Microbial Communities Significantly Differ by Sampling Time

Sampling time (a.m. versus p.m.) was identified as the major determinant of the shifts and recurrences in community composition and diversity. Simpson's alpha-diversity indices notably decreased from a.m. to p.m. (0.90 to 0.85, paired t test, $p = 1.6e{-}4$; Table S1), accompanied by a consistent increase in *P. acnes* abundance (25.7% to 32.5%, paired t test, $p = 2.1e{-}4$; Figures 2B and 2C). The ER line, the only cross-border line connecting to Shenzhen in mainland China, appears as the only outlier. At the species level, the abundances of 140 species significantly decreased from a.m. to p.m. (paired t test, $p < 0.05$), with a median rank of their relative abundance in all analyzed species of 419.5/744 (Figure 3B). We defined these species as a.m. signature species. In only 48 species, the population expanded from a.m. to p.m., including *Enterobacteria* phages and *Propionibacterium* phages in 5 subtypes, with a median rank of 232/744 (Figure 3B). These species were defined as p.m. signature species. The Matthew effect can be used as an analogy for delineating the dynamics of microbial communities "the rich (a.m.) get richer (p.m.) and the poor get poorer" in their abundance (Merton, 1968). By analyzing the pathogenicity of the time-specific signature species, only 3 potential opportunistic pathogenic species of low abundance (<0.3%) were a.m. signatures, whereas 11 were p.m. signature species, led by *P. acnes* and *S. mitis*. The total abundance of potential opportunistic pathogenic species is also dramatically increased (paired t test, $p = 5.9e{-}6$) but becomes insignificant when excluding *P. acnes* (paired t test, $p = 0.25$; Figure 3A). To determine the driver(s) in the niche expansion of *P. acnes*, we performed comparisons at the individual strain level. SK137 and hdn-1 significantly increased from a.m. to p.m., whereas HL035PA1 decreased (Wilcoxon signed-rank test on ranks, $p < 0.05$). Notably, none of these altered strains have been previously reported as pathogenic strains. When assigning the species to ecological categories (including different habitats and physiological traits; Data S3A) using the PATRIC database (Wattam et al., 2014), we detected a significant increase p.m. for the categories of skin-associated, disease and/or human disease-related, hosted and/or human-hosted, non-motile, and salinity-tolerant species and a decrease in marine species (as well as in soil, aquatic, plant-hosted, and motile species when removing the outlier ER line) (paired t test, $p < 0.05$; Figure 3A). Interestingly, for some non-human-associated species with high abundance, we also observed a high fluctuation in abundance regardless of sampling time points. Such examples include the fourth most abundant species, *Burkholderia*, a nitrogen-fixing soil organism, with a relative abundance of 1.97% ± 3.94%, *Bradyrhizobium elkanii* (19^{th}, 0.48% ± 1.64%), and *Labrenzia alexandrii* (20^{th}, 0.48% ± 0.96%). *Burkholderia* and *L. alexandrii* showed remarkable variance in abundance among the three sampling days (ANOVA test, $p < 0.05$, *Burkholderia*: 0.00%, 0.86%, and 5.04%; *L. alexandrii*: 0.91%, 0.53%, and 0.00% for the 3 days). *B. elkanii* had extremely high abundance only in day 1 p.m. samples in the ER and MOS lines (10.47% and 4.53%; the relative numbers were 1.07% and 1.31% in a.m. compared with 0.13% as the mean for other samples). The PATRIC categories of animal-hosted species, aquatic species, marine species, non-antimicrobial resistance (AMR) species, thermo-tolerant species, and species with multiple altitude preferences (1,000 m, 10 m, and 1 m), showed instability among the 3 sampling days, and none of these categories are human-related. One possible explanation is that this variation could be driven by non-human factors, such as humidity and aerosols, because weather conditions during the sampling days differed (Table S1). Based on this evidence, we speculate that daily traffic flow is the driving factor in shaping the microbial communities of a public transit system. Human-associated species are universally more abundant than non-human associated species, and they tend to become more abundant under high traffic density during the day; conversely, species with low relative abundance are unlikely to be human commensals, and they are further outcompeted by the abundant human-associated species during high-traffic hours before

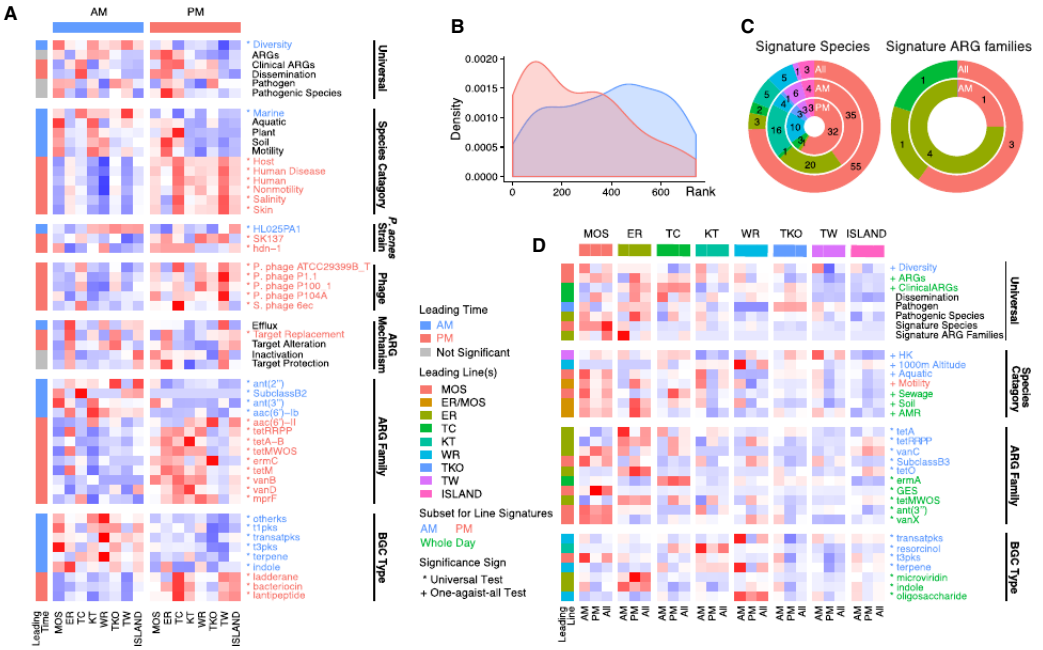


Figure 3. Time-Specific and Line-Specific Signatures and Their Relative Abundance

Relative abundances are scaled by rows to *Z* scores, with blue to red colors representing low to high relative abundance.

(A) Heatmap for time-specific signatures. The colors of the left color bar indicate the leading time point. Row label colors and marks represent the leading time point and statistical significance.

(B) Density plot for the rank distributions of the a.m. (blue) or p.m. (red) signature species among 744 analyzed species. The distribution of p.m. signature species is left-skewed.

(C) Number of line-specific signature species and signature ARG families among a.m. or p.m. subsets, respectively, and among all samples.

(D) Heatmap for line-specific features for a.m. or p.m. subsets and all samples. Row label colors suggest whether the significance was observed in the a.m. or p.m. subset only or among all samples. Colors in the left color bar indicate the leading line of the feature.

recovering their ecological niche during non-traffic hours. Naturally, the abundant species show higher inter-day stability in their composition, whereas, conversely, the less abundant non-commensal species show greater stochastic variations among different sampling days.

Microbial Communities Are Line Dependent

In addition to sampling time, community diversities (Table S1) and dissimilarities revealed that the microbial compositions were also highly influenced by the metro line features (Figure 2D). The distances among communities from different lines become smaller from a.m. to p.m., except for the ER line, the only cross-border line linked to mainland China, which is also the only interchangeable line of the most isolated line in the MTR system, the MOS line. The coordinates of the a.m. and p.m. communities from the MOS line locate closely to the p.m. and a.m. communities from the ER line, respectively. The coordinate swapping may reflect the effect of the traffic tide induced by the unique interchange. IndVal tests (Dufrene and Legendre, 1997) were performed to identify line-specific signature species: species that are relatively abundant only in one line compared with all other lines. Line-specific signatures were investigated among all samples or among only a.m. or p.m. samples. In general, more a.m. signature species (88) could be detected than p.m. signatures (59) or all-day signatures (78) (Figure 3C) because the community dissimilarities dramatically decreased p.m. The most isolated line, MOS, showed a markedly greater number of signature species (55) both a.m. and p.m. as well as the highest community diversity a.m. (one-against-all t test, $p = 8.0e{-}5$; Figure 3D). The ER and KT lines were also characterized by 20 and 16 a.m. signature species, respectively, before losing their discriminatory power p.m. The stability of the signature species is also line-dependent: only 21 of the 88 a.m. signature species could retain specificity p.m., 17 of which are MOS line signatures; only 4 species are signatures for the other 7 lines. When assigning species into PATRIC categories, line-specific signature categories could be observed (one-against-all Kruskal-Wallis test on ranks, $p < 0.05$; Figure 3D). The MOS and ER lines reported the highest abundance of soil species and species with AMR in all samples but motile species only p.m. The MOS

line, which is an entirely aboveground ground line running alongside the Shing Mun Channel, a polluted brackish river, is the most populated line for aquatic species a.m. and sewage species both a.m. and p.m. The WR line, which passes through a mountain region in the New Territories, showed the highest abundance of the species with a habitat preference for altitude around 1,000 m, whereas a low abundance of these environmental categories was observed in underground urban lines such as ISLAND, TW, and TKO. Species that were originally isolated in Hong Kong were most abundant in the TW line, an urban line with the highest traffic volume (Table S2), which connects the densely populated areas in Kowloon with the Central station on Hong Kong Island. All of these observations regarding microbial composition are associated with line-specific environmental exposure. To estimate interline transmissions and track potential sources of the heterogeneously mixed p.m. communities, we employed the Bayesian SourceTracker algorithm (Knights et al., 2011) to estimate the contributions of the a.m. communities, assuming that the p.m. communities (sink communities) originated, transmitted, and were mixed from 8 independent line a.m. communities (source communities), and each line a.m. community is contributing to a particular proportion of the p.m. communities. Not surprisingly, microbial relocation between the ER and MOS lines was detected at a high level. The a.m. community of the ER line contributes a notable proportion to multiple p.m. communities, including the MOS, TC, WR, and TW lines (Figure S2). The two lines with the longest interchange routes to the ER line, the TKO and ISLAND lines, are relatively less colonized by the a.m. community of the ER line. As the most isolated line, the MOS line contributes a very limited percentage to all non-ER lines, supporting its capacity to retain the specificity of signature species p.m.

The MTR Resistome and Functionome

The unique characteristics of the ER line observed in the above analysis triggered our interest in the possibility of cross-border transmission of ARGs via the MTR system. Intensive analysis of the resistome profile in the MTR lines was performed to explore the ARG mechanism (Data S3B), gene family (Gibson et al., 2015; Data S3C), previously annotated clinical relevance (Munck et al., 2015), and dissemination potential (STAR Methods). In general, antibiotic efflux (469.2 ± 182.6 RPM [reads per million reads]) was revealed as the most abundant characterized mechanism in the MTR communities, followed by target protection (146.3 ± 26.7) and inactivation (137.3 ± 35.3), target alteration (69.1 ± 13.5), and replacement (26.7 ± 9.8). Overall, we identified 136 ARG families (the overall abundance was 322.6 ± 62.1 RPM), including 24 clinically important ARG families (14.1 ± 5.1 RPM) (Data S3C). The most abundant ARG families were fluoroquinolone-resistant DNA topoisomerase (42.8 ± 6.6 RPKM), *aph(3′)* double prime (34.5 ± 5.1), RND antibiotic efflux pump (13.4 ± 5.8), ABC antibiotic efflux pump (13.2 ± 3.0), and *macB* (9.2 ± 2.9). The most abundant clinically important ARG families were *ermC* (3.6 ± 2.7 RPKM), *msrA* (3.3 ± 2.0), *tetK* (1.7 ± 0.7), *ant4(4′)-Ia* (1.5 ± 0.6), and *tetM* (1.0 ± 0.4). The abundance of clinically important ARG families showed higher inter-sample variability than non-clinical families, suggesting a potentially higher effect of human activities. Therefore, a systematic analysis of the time- and line-specific signature ARG families was a logical follow-up.

The p.m. Signature of the ARG Families Is Highly Associated with Human Activities and Clinical Antibiotic Practice

When comparing the resistome between a.m. and p.m. communities, no differentiation was captured in the abundance of the ARG mappable reads. Similarly, the abundance of clinically important ARGs and community-wide dissemination potential increased modestly without statistical significance (Figure 3A). The community-wide dissemination potential was estimated from the horizontal gene transfer (HGT) rate of each ARG family (Data S4; STAR Methods). At the ARG family level, the abundance of 4 ARG families significantly decreased from a.m. to p.m. (*ant(2″)*, *ant(3″)*, *aac(6′)-Ib*, and SubclassB2; all non-clinical), whereas 9 families increased, including clinically important families like *tetM*, *ermC*, *vanB*, tetracycline resistance ribosomal protection protein (tetRRPP), *tetA-B*, *tetM-tetW-tetO-tetS* (tetMWOS), *vanD*, and *mprF* (paired t test, $p < 0.05$; Figure 3A). Resistance against the clinically widely used antibiotics tetracycline, vancomycin, erythromycin, and methicillin was highly observed in p.m. communities. These observations led us to pose a new question: is the increase in clinically relevant antibiotic resistance contributed only by the colonization process of human-associated species, which occurs equally in all lines, or it is also accompanied by inter-line ARG dispersion carried by the transmission of line-specific signature species? To answer this question, we investigated the line-specific ARG signatures and taxonomic contributors.

The ER Line's Resistome Signature Suggests Potential Cross-Border Antibiotic Resistance Transmission

Besides sampling time, we also examined the specific characteristics of each line and observed that the MOS line has a significantly higher abundance of all ARG reads. The highest abundance of clinically important ARGs and highest community-wide dissemination potential were observed in the TC and ER lines (one-against-all Kruskal-Wallis test on ranks, $p < 0.05$; Figure 3D). The ER and RC lines showed a high abundance of the ARG family *blaI* beta-lactamase repressor gene as the major contributor to the high dissemination potential. The MOS line is ranked third in terms of community-wide dissemination potential, driven by the high HGT rate of the *aph(3′)* family. These 3 lines are uptown lines with predominantly aboveground tracks. Similar to the line-specific signature species, the ER and MOS lines also lead in terms of number of signature ARG families: 5 for the ER line, 4 for the MOS line, and only one for each of the other lines (*emrA* for the TC line) (Figure 3C). The MOS line has more non-clinical signature ARG families (*ant(3″)*, SubclassB3, and GES) and only one clinically relevant family (*vanX*); in contrast, all ER line signature ARG families are related with clinically used antibiotics (Figure 3D), and 4 of 5 (*tetA*, *tetO*, tetRRPP, and tetMWOS) are involved in resistance against tetracycline, besides one against vancomycin (*vanC*). Among these 5 families, 4 (excluding tetMWOS) are a.m. signatures without specificities p.m.

Busy Lines Do Not Show a Higher Abundance of Clinically Important ARGs or Higher ARG Dissemination Potential

To disclose whether civic characteristics such as urban density and metro network topology are also influencing the

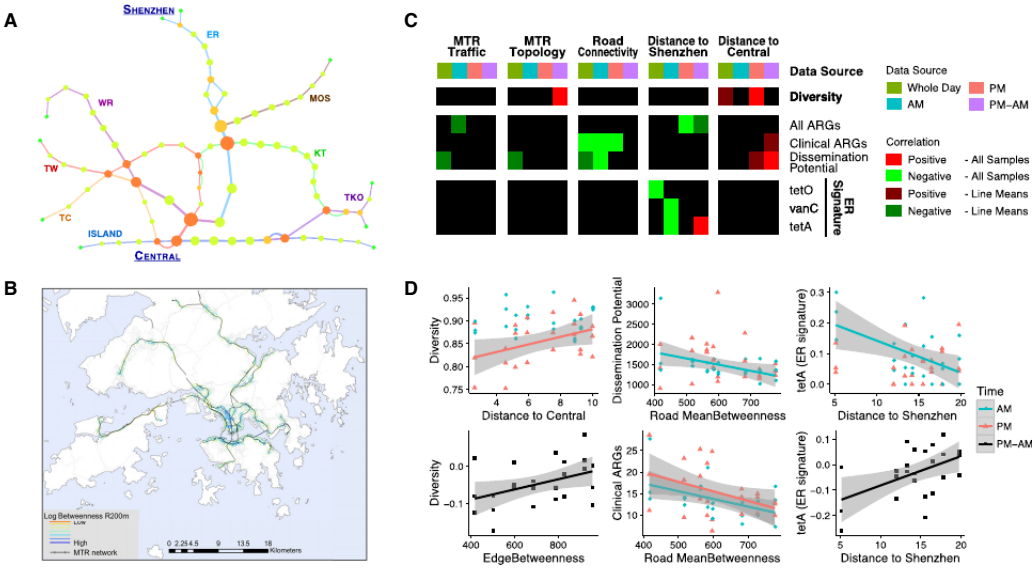


Figure 4. Correlations between Line Biological Observations and Geographical or Topological Characteristics

(A) The topological network of the MTR system, with the node color, node size, edge color, and edge width representing the node degree, betweenness centrality, metro line, and edge betweenness.

(B) Movement potential (betweenness centrality) of the hybrid MTR train-road network at a 200-m spatial scale modeled by sDNA (Cooper and Alain, 2015).

(C) Associations between biological observations and geographical or topological characteristics. A red color indicates a positive correlation and green a negative correlation. A brighter color suggests a significant correlation by using all samples individually in the coefficient calculation, whereas a darker color suggests a significant correlation by using line means. For the MTR topology, a cell is marked as significant when at least one feature (mean node degree, mean betweenness centrality, mean closeness centrality, or mean edge betweenness) is significantly associated with the biological observation. For road connectivity, a cell is marked as significant when the biological observation is significantly correlated with the sDNA movement potential (betweenness centrality) at more than one spatial scale (from 100 m to 500 m).

(D) Some examples of significant correlations between one biological observation and one geographical or topological characteristic, with different dot and fitting line colors indicating different distinct times (a.m., p.m., or the increment from a.m. to p.m.).

microbial diversity, line signatures, resistome composition, and its transmission potential, we performed an additional association analysis using a hybrid urban movement network operationalized by integrating the MTR topological network (Figure 4A; Data S5A), the adjoining street network linked to MTR station entrances (Figure 4B; Figure S3; Data S5B), the traffic volume (The Transport and Housing Bureau of the Government of the Hong Kong Special Administrative Region, 2014; Table S2), as well as the distances to the Central station and Shenzhen (mainland China) (Data S5A). The distance to Central station was used to distinguish downtown and uptown lines, whereas the distance to Shenzhen was introduced to capture intercity transmission events. We correlated these features with community-wide and health-related observations, including community diversity, ARG abundance, clinically important ARG abundance, and community-wide ARG dissemination potential. At the community scale, diversity (in Simpson's index) was observed to be positively correlated with the line's average distances to Central station (Figures 4C and 4D; significant p.m., Pearson's correlation coefficient $r = 0.41$, $p = 0.046$). The diversity differences between p.m. and a.m. (Simpson's index p.m.–a.m.) were positively correlated with the line average edge betweenness ($r = 0.47$, $p = 0.019$; Figures 4C and 4D). These findings indicate that isolated or uptown lines tend to have higher community diversities and greater decreases in the community diversities during the day. In relation to the resistome, strong negative correlations were captured between the clinically important ARG abundance and the mean betweenness of the road connectivity networks ($r = -0.42 \sim -0.37$, $p < 0.05$; Figures 4C and 4D), whereas a negative correlation was also observed for the community-wide ARG dissemination potential in the a.m. samples ($r = -0.48 \sim -0.41$, $p < 0.05$; Figures 4C and 4D). There was also a strong positive correlation between the increase of dissemination potential (p.m.–a.m.) and the distance to Central station ($r = 0.44$, $p = 0.033$; Figure 4C). Based on these correlations, the isolated uptown lines tend to have a higher abundance in clinically important ARGs and ARGs with higher dissemination potential. The distance to Shenzhen (mainland China) was negatively correlated with the overall

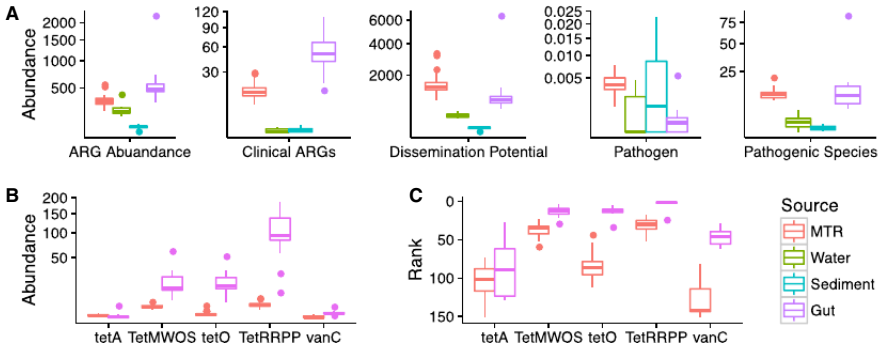


Figure 5. Comparisons of Health-Related Features among Different Microbial Communities

(A) Boxplots for the abundance of health-related features in different microbial communities, including the abundance of all ARGs (in RPMs), clinically important ARG families (in RPMs), community-wide dissemination potential, opportunistic pathogens, and potential opportunistic pathogenic species (excluding *P. acnes*).

(B) The abundance of the ER line signature ARG families for all MTR and Chinese gut samples (in RPKMs).

(C) The ranks of the ER line signature ARG families among all observed ARG families for all MTR and Chinese gut samples.

ARG abundance p.m. ($r = -0.48$, $p = 0.018$; Figure 4C), as well as its increment (weak correlation calculated from line mean values), suggesting that lines closer to Shenzhen tend to have higher ARG input during the day. We also detected the correlations for three individual ER line signature ARG families (*tetO*, *tetA*, and *vanC*), whose abundances were negatively correlated with the distance to Shenzhen, especially in a.m. communities ($r = -0.63 \sim -0.31$, $p < 0.05$; Figures 4C and 4D). The increased abundance of *tetA* was positively correlated with distance ($r = 0.50$, $p = 0.012$; Figures 4C and 4D), supporting that the farther away from Shenzhen, the higher the *tetA* abundance that could be gained over a day, which is consistent with our previous observations: an ER line a.m. signature becomes a p.m.-enriched ARG family in all MTR lines far from Shenzhen.

In addition, we conducted a comparative analysis among diverse environmental samples from Hong Kong (including drinking water and marine sediments) as well as gut microbial communities of Chinese people living on the mainland, collected from several separate studies (Guo et al., 2016; Yu et al., 2017; Figure 5A). Not surprisingly, the microbial communities from Chinese human gut samples have a significantly higher abundance of ARGs and especially of clinically important ARGs compared with the MTR communities (t test, $p = 0.016$ for ARG abundance and 1.7e−5 for clinical ARG abundance). This observation also meets the recent study on the sewage resistome for 15 Chinese cities (Su et al., 2017). Clinically important ARGs were rarely observed in drinking water and marine sediment samples. MTR samples have the highest community-wide dissemination potential. followed by gut samples. Because of the special attributes of the ER line, we also compared the abundance of the ER signature ARG families between the MTR samples and Chinese human gut samples (Figures 6B and 6C). Apart from *tetA* (with low abundance in both communities), the abundance and ranks of four ER line signature families are consistently higher in the Chinese gut samples (t test or Wilcoxon rank test, $p < 0.05$).

Correlations between Species, ARG Families, and Biosynthetic Gene Clusters

To further characterize the functionality of the microbial communities in connection to the observed resistome profile, we quantified and compared the abundance of the biosynthetic gene clusters (BGCs) (Donia et al., 2014; Data S3D), which has also been suggested by the MetaSUB Consortium as a standard component for urban microbiome characterization (MetaSUB International Consortium, 2016). Among the 22 retrieved BGC types, 3 types were identified as p.m. signature types: ladderane, bacteriocin, and lantipeptide. Bacteriocin and lantipeptide are BGC types closely related to antibiotic resistance. Six types were identified as a.m. signature types: type I polyketide (t1pks), trans-AT polyketide (transatpks), type III polyketide (t3pks), terpene, indole, and other types of polyketide (otherks) (Figure 3A). Among the 7 line-specific signature BGC types, 4 were associated with the WR line (Figure 3D). SparCC correlations were calculated among BGC types and ARG families (Data S6). The p.m. signature lantipeptide was detected to be highly associated with multiple ARG families (Figure 6C). The BGC types aryl polyene (arylpolyene) and homoserine lactone (hserlactone) were also highly associated with ARG families. Investigating the ARG co-occurring and co-excluding networks (Figures 6A and 6B), we could observe competing modules in both a.m. and p.m. networks: one human or clinic-associated module with the *msr*, *ble*, *mpr*, *erm*, *ctx-m*, and *tet* families and one non-clinical module with the *ade*, *bae*, *tol*, *sox*, *cfr*, *rnd*, *abc*, and *ant* families. In the a.m. network, stronger co-occurring links could be observed in the non-clinical module, whereas the connections among the clinical families were relatively weak. In contrast, stronger links among clinical links could be identified in the p.m. network, whereas the non-clinical module was losing internal connections. These observations also support our conclusions from the time-specific ARG analysis.

Co-occurring relationships and taxonomic links were observed between the abundances of ARG families, BGC types,

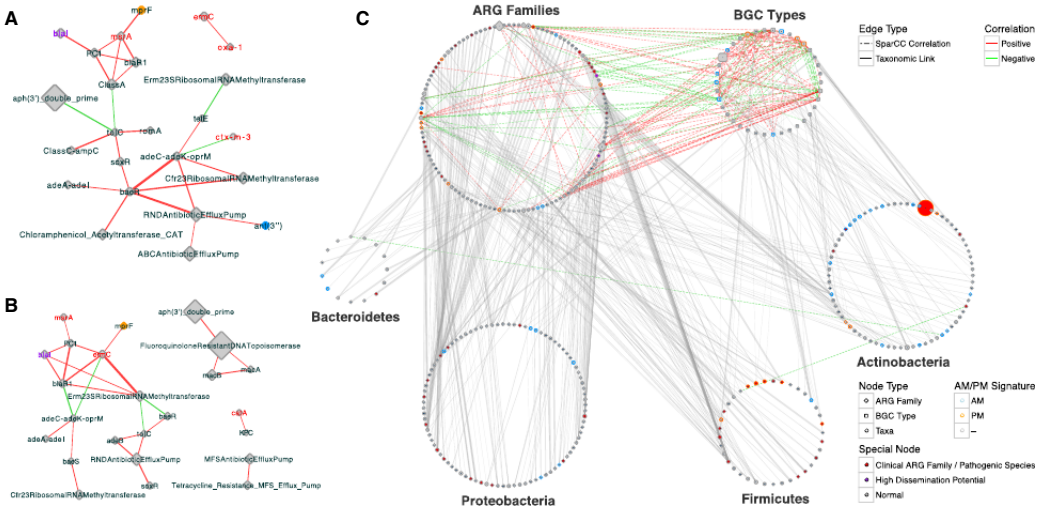


Figure 6. Time-Specific Network Plots for ARG Families, BGC Types, and Their Taxonomic Contributors

(A and B) Co-occurring and co-excluding networks for ARG families for a.m. (A) and p.m. (B) subsets, respectively. Node sizes indicate the relative abundance of the families. Node colors suggest time-specific signature ARG families (blue for a.m. and orange for p.m.). Clinical importance and high dissemination potential (HGT rate > 100) are illustrated by the label colors (red and purple, respectively). Edge widths and colors represent the SparCC coefficients (Friedman and Alm, 2012) and the directions of correlation.

(C) The contribution and correlation networks of taxa, ARG families, and BGC types for the p.m. subset. Species are presented as dots and grouped by phyla, whereas ARG families and BGC types are indicated by diamonds and squares. Different stroke colors indicate time-specific signatures. Filling colors suggest clinical importance (for ARG families, red), dissemination potential (for ARG families, purple), or pathogenic potential (for species, red). Edge width represents the correlation coefficient or the proportion of a taxonomic contribution to ARG or BGC (5% to 100%), whereas edge colors illustrate the directions of the correlations.

and their contributing species (Figure 6C; Data S7). At the phylum level, the most abundant phylum, *Actinobacteria*, contributed to both ARGs and BGCs. *Proteobacteria* and *Firmicutes* contributed more to ARGs but less to BGCs, whereas *Bacteroidetes* showed links only to ARGs. The dominant species, *P. acnes,* was not determined to be a major contributor to any ARG family. *S. aureus*, *E. coli*, *S. suis*, *K. oxytoca*, *S. arlettae*, and *Williamsia* sp. D3 were the top-ranking species contributing to the clinically important ARG families. Notably, we could observe a high abundance of *S. aureus* in the TC, MOS, and ER lines. The ER line was the most populated line for *S. suis*, *K. oxytoca*, and *S. arlettae*, whereas the MOS line showed a higher abundance of *Williamsia* spp. The abundance of the top clinical ARG contributing species was always led by clinically important ARG-enriched lines (the ER and TC lines). Investigating the line signature ARG families, *S. suis* was the major contributor to the ER signature families *tetA* and *tetO*. The ER line was also the most abundant line for *S. suis*. *blaI*, an ARG family with an extremely high HGT rate, had a distinctly higher abundance in the TC and ER lines, and its only contributing species, *S. aureus*, was also abundant in the TC (second) and ER (fourth) lines. For the MOS leading family *aph(3′)*, also with a high HGT rate, multiple contributors (*S. maltophilia*, *A. baumannii*, and *P. aeruginosa*) were identified, and the MOS line was validated to have a high abundance of *A. baumannii*, followed by *S. maltophilia*. On the other hand, BGC-type contributing organisms were more extensively distributed throughout the whole community (4,303 taxa contributing to 40 BGC types) in comparison with the ARG contributors (only 2,116 taxa linked with 136 ARG families). The greatest BGC contribution was from *M. luteus*, followed by *P. acnes*, *Gordonia bronchialis*, *Kytococcus sedentarius*, and *Corynebacterium jeikeium*. For the 4 ARG-associated BGC types mentioned above, *P. acnes* was the dominant contributor to lantipeptide, whereas bacteriocin was evenly linked with multiple species from *Streptococcus*, *Kocuria*, *Burkholderia*, *Haemophilus*, and other genera, and aryl polyene and homoserine lactone shared *A. baumannii* as their leading contributing species.

DISCUSSION

Previous studies have investigated the microbial composition of mass transit aerosols within the New York City and Hong Kong MTR (Leung et al., 2014; Robertson et al., 2013) as well as station and train surfaces in New York City and Boston (Afshinnekoo et al., 2015; Hsu et al., 2016). The skin microbial community should also be appreciated as a pivotal component of such indoor investigations. The skin microbiota is adapted to the human body and may influence human health through its capacity to travel with individuals, shed from the skin, adhere to indoor

surfaces, and recolonize other individuals (Lax et al., 2014). Our study focuses on the characterization of skin microbiota acquired when exposed to a public mass transit environment such as the Hong Kong MTR. Like other studies of train surfaces, station surfaces, and aerosols (Afshinnekoo et al., 2015; Hsu et al., 2016; Leung et al., 2014; Robertson et al., 2013), the metro microbial communities recovered from palm samples in the Hong Kong MTR were largely derived from human skin commensals regardless of the sampling sites. In addition, shotgun sequencing provided significantly higher resolution of the taxonomic profiles and identified *P. acnes* as the dominant species, which was not identified in early 16S rRNA sequencing-based studies, leading to a potential bias in the interpretation of microbial communities (Leung et al., 2014; Robertson et al., 2013). By applying shotgun sequencing, we could delve into the strain-level analysis, which identified putative pathogenicity, interactions with distinct phage subtypes, and strain-specific variation and contribution to morning versus afternoon conditions.

Regarding the resistome, the resistance genes against chloramphenicol, tetracycline, and beta-lactamase reported by previous studies (Afshinnekoo et al., 2015; Hsu et al., 2016) were also observed in our samples. Our investigation of the skin microbial community enabled the precise characterization of the antibiotic resistance, based on the relatively high stability of the microbial communities (Oh et al., 2016). In comparison with the Boston metro metagenomic study (Hsu et al., 2016), more resistance to widely used clinical antibiotics, including vancomycin, erythromycin, and methicillin, could be detected in our samples with daily recurrence. Besides the clinically important ARGs, this study also investigated and estimated the ARG dissemination potential at the family level and community scale. Routine screening for clinically important ARGs and assessment of their dissemination potential would allow improved detection of potential health risks induced by antibiotic resistance outbreaks, geographical transmission, or cross-species migration.

The universal stability of the skin microbiome (Oh et al., 2016) promotes the reliability and the reproducibility of longitudinal analyses across the sampling intervals used in our study: short intervals as in different sampling times in a day and median intervals as in different sampling days in successive weeks. The sampling intervals have successfully assisted with exploring the effect on community diversity and stability from the intraday tidal effect as well as the inter-day variation. Intraday sampling time was identified as the major driver of the variation and recurrence of microbial composition and diversity in conjunction with the ARGs and BGCs carried by the community composition. When evaluating the inter-day stability, the less abundant environmental species (not human-hosted organisms) showed significantly higher instability, further suggesting their potential competitive disadvantages under the pressure of human traffic and the influence of non-human factors, such as indoor climate (Brodie et al., 2007; Kelley and Gilbert, 2013). Taxonomic contributors to the ARG families were also captured for the time- and line-specific signature ARGs, with strong co-occurring trends, serving as evidence of the presence of signature ARGs and their transmission events (Hiramatsu et al., 2014; Lowy, 2003). Tracing ARG-contributing species could serve as a starting point for strategies that might inhibit ARG transmission and dissemination in the future. Furthermore, in our analyses, we considered the taxonomy, resistome, and BGC profiles as interrelated components. Taxonomic contributors and associated BGC types were regarded as direct and implied evidence for the divergence observed in the resistome (Donia et al., 2014).

When exploring the resistome, clinically important ARG families display higher variances, suggesting greater influence from human traffic. Notably, most p.m. signature ARG families are closely affiliated with clinical use of antibiotics. The specificity of line-specific signatures decreased from a.m. to p.m., in accordance with a previous evaluation of the stability of site-specific signatures of skin microbes (Oh et al., 2016). By reviewing the time- and line-specific ARG analyses in conjunction, we captured some evidence for intra- and intercity ARG transmission. As a notable case, we captured 4 tetracycline- and 2 vancomycin-resistant ARG families as universal p.m. signature families: tetMWOS, tetRRPP, *tetM*, *tetA-B*, *vanB*, and *vanD*. All of these families were identical or highly associated with the ER line signature families. The ER line showed a markedly higher abundance in these families a.m., and then these families gained in abundance in all lines and became universal p.m. signature families, whereas the ER line lost specificity in these families (Figure 3A). From these findings, one may hypothesize that the ER line, the only cross-border line linked to mainland China, may serve as a potential source of clinically important ARGs, especially against tetracycline, which is a commonly used antibiotic in China's swine feedlots (Wu et al., 2010) and can be detected in the soil in the Pearl River Delta area (Li et al., 2011), where the cities of Hong Kong and Shenzhen are located. It appears that these ARGs disperse into different lines in the MTR system with the carriage of human traffic and finally become relatively abundant in every line in the later hours of the day.

Furthermore, from our study, robust patterns in microbial community dynamics and clinically relevant ARG transmission were observed. To evaluate the potential real health effect of such transmission events, the population size of the microbial community becomes a key factor. In our preliminary study, we also collected samples by swabbing the train surfaces, following the same procedure as used in the Boston study (Hsu et al., 2016). Our DNA amounts from both palm and surface samples (<1 ng/ L) were orders of magnitude lower than those in the original study (40∼220 ng/ L) (Hsu et al., 2016), which suggests a relatively lower overall biomass of the MTR microbiome and relatively higher safety when using the MTR, as the transmission of pathogens and ARGs is kept at a low level. The low DNA recovered may be attributed to the effectiveness of the antibacterial nano-silver-titanium dioxide coating (NSTDC) applied in the MTR (MTR Corporation, 2006). Assuming that no such antibacterial practice was applied, the robust patterns of the clinically relevant ARG transmission could be amplified, resulting in increased health effects. We propose that public policy makers should continue to pursue and evaluate evidence of the effect of antimicrobial strategies applied to public transit systems and checkpoints (e.g., border control regions, airports and aircraft, cross-border trains and buses, etc.), especially between regions with different norms and regulations in industrial and clinical antibiotic usage. In addition to maintaining and expanding the silver nanoparticle coating prophylaxis, public restroom

OPEN ACCESS CellPress

sanitation could be improved, and hand sanitizer dispensers could be provided at the exits of metro, train train, and bike-sharing stations, airports, and hospitals to reduce the transmission of pathogens and clinically relevant ARGs. At a higher level, the growing understanding of the contribution of microbial communities to human health should be reflected in design of built environments and materials and equipment for public use with attention to the microbiome. Studies have found that architecture and ventilation design affect community diversity (Kembel et al., 2014). Furthermore, given the effect of substrate on community composition, future work on the dynamics of optimizing spaces and surfaces for human health in relation to microbiota is needed.

In summary, our investigation of the microbiome associated with public transit reveal that communities are decidedly shaped by human interactions and traffic flows, and we hope to have established a baseline for further studies by incorporating multidimensional influences from the host, time, environment, geography, and varied civic characteristics. In future studies, more sampling time points could be considered to delineate more exquisite intraday dynamics, especially the time points before and after morning and evening peaks, respectively. A dynamic and surveillance of the microbial communities responsive to and shaped by the public transit systems could serve as a reference for public policy-making regarding the prevention of ARG spread via intercity and cross-border traffic and protect millions of citizens from potential health hazards in daily life.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, two tables, and seven data files and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.06.109.

ACKNOWLEDGMENTS

The authors thank Thrimendra Dissanayake for sampling and Kong Ka Fai for preparing the sampling consumables. We also thank the Centre for Genomic Sciences (CGS), University of Hong Kong, for technical help with sample quality control, library construction, and DNA sequencing.

AUTHOR CONTRIBUTIONS

K.K. performed the analyses and wrote the initial manuscript. G.P. supervised the study. Y.N., J.L., and G.P. designed the study. M.D.K., C.D., S.D., and C.J.W. advised on the civic issues. D.M.B. advised on the analyses of ecology. L.I. and M.O.A.S. advised on the health-related studies. L.I. performed the functional metagenomic study. Y.N., J.L., Y.H., and T.Z. built the databases and computational pipelines for the ARG and BGC analyses. C.S., S.K., and C.J.W. built the spatial design network analysis (sDNA) model. J.C.Y.W., A.A., and C.W.M.W. conducted quality control and experimental optimization. Y.N., Y.H., T.Z., S.K., and C.W.M.W. collected the samples. All authors provided invaluable feedback for the analyses and approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 20, 2017
Revised: January 25, 2018
Accepted: June 27, 2018
Published: July 31, 2018

REFERENCES

Adams, R.I., Bateman, A.C., Bik, H.M., and Meadow, J.F. (2015). Microbiota of the indoor environment: a meta-analysis. Microbiome *3*, 49.

Afshinnekoo, E., Meydan, C., Chowdhury, S., Jaroudi, D., Boyer, C., Bernstein, N., Maritz, J.M., Reeves, D., Gandara, J., Chhangawala, S., et al. (2015). Geospatial Resolution of Human and Bacterial Diversity with City-Scale Metagenomics. Cell Syst *1*, 72–87.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Bansal, M.S., Alm, E.J., and Kellis, M. (2012). Efficient algorithms for the reconciliation problem with gene duplication, horizontal transfer and loss. Bioinformatics *28*, i283–i291.

Be, N.A., Thissen, J.B., Fofanov, V.Y., Allen, J.E., Rojas, M., Golovko, G., Fofanov, Y., Koshinsky, H., and Jaing, C.J. (2015). Metagenomic analysis of the airborne environment in urban spaces. Microb. Ecol. *69*, 346–355.

Bhalla, A., Pultz, N.J., Gries, D.M., Ray, A.J., Eckstein, E.C., Aron, D.C., and Donskey, C.J. (2004). Acquisition of nosocomial pathogens on hands after contact with environmental surfaces near hospitalized patients. Infect. Control Hosp. Epidemiol. *25*, 164–167.

Brodie, E.L., DeSantis, T.Z., Parker, J.P., Zubietta, I.X., Piceno, Y.M., and Andersen, G.L. (2007). Urban aerosols harbor diverse and dynamic bacterial populations. Proc. Natl. Acad. Sci. USA *104*, 299–304.

Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60.

Chen, L., Zheng, D., Liu, B., Yang, J., and Jin, Q. (2016). VFDB 2016: hierarchical and refined dataset for big data analysis–10 years on. Nucleic Acids Res. *44* (*D1*), D694–D697.

Cogen, A.L., Yamasaki, K., Muto, J., Sanchez, K.M., Crotty Alexander, L., Tanios, J., Lai, Y., Kim, J.E., Nizet, V., and Gallo, R.L. (2010a). Staphylococcus epidermidis antimicrobial delta-toxin (phenol-soluble modulin-gamma) cooperates with host antimicrobial peptides to kill group A Streptococcus. PLoS ONE *5*, e8557.

Cogen, A.L., Yamasaki, K., Sanchez, K.M., Dorschner, R.A., Lai, Y., MacLeod, D.T., Torpey, J.W., Otto, M., Nizet, V., Kim, J.E., and Gallo, R.L. (2010b). Selective antimicrobial action is provided by phenol-soluble modulins derived from Staphylococcus epidermidis, a normal resident of the skin. J. Invest. Dermatol. *130*, 192–200.

Cooper, C.H.C., and Alain, J. (2015). sDNA: how and why we reinvented Spatial Network Analysis for health, economics and active modes of transport. N. Malleson, ed. GIS Research UK (GISRUK) 2015 Proceedings, pp. 122–127.

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. J. Veg. Sci. *14*, 927–930.

Donia, M.S., Cimermancic, P., Schulze, C.J., Wieland Brown, L.C., Martin, J., Mitreva, M., Clardy, J., Linington, R.G., and Fischbach, M.A. (2014). A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. Cell *158*, 1402–1414.

Dufrene, M., and Legendre, P. (1997). Species assemblages and indicator species: The need for a flexible asymmetrical approach. Ecol. Monogr. *67*, 345–366.

Fitz-Gibbon, S., Tomida, S., Chiu, B.H., Nguyen, L., Du, C., Liu, M., Elashoff, D., Erfe, M.C., Loncaric, A., Kim, J., et al. (2013). Propionibacterium acnes strain populations in the human skin microbiome associated with acne. J. Invest. Dermatol. *133*, 2152–2160.

Forsberg, K.J., Reyes, A., Wang, B., Selleck, E.M., Sommer, M.O., and Dantas, G. (2012). The shared antibiotic resistome of soil bacteria and human pathogens. Science *337*, 1107–1111.

Forsberg, K.J., Patel, S., Gibson, M.K., Lauber, C.L., Knight, R., Fierer, N., and Dantas, G. (2014). Bacterial phylogeny structures soil resistomes across habitats. Nature *509*, 612–616.

Friedman, J., and Alm, E.J. (2012). Inferring correlation networks from genomic survey data. PLoS Comput. Biol. *8*, e1002687.

Gibbons, S.M., Schwartz, T., Fouquier, J., Mitchell, M., Sangwan, N., Gilbert, J.A., and Kelley, S.T. (2015). Ecological succession and viability of human-associated microbiota on restroom surfaces. Appl. Environ. Microbiol. *81*, 765–773.

Gibson, M.K., Forsberg, K.J., and Dantas, G. (2015). Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. ISME J. *9*, 207–216.

Guo, F., Li, B., Yang, Y., Deng, Y., Qiu, J.W., Li, X., Leung, K.M.Y., and Zhang, T. (2016). Impacts of human activities on distribution of sulfate-reducing prokaryotes and antibiotic resistance genes in marine coastal sediments of Hong Kong. FEMS Microbiol. Ecol. *92*, fiw128.

Heshiki, Y., Dissanayake, T., Zheng, T., Kang, K., Yueqiong, N., Xu, Z., Sarkar, C., Woo, P.C.Y., Chow, B.K.C., Baker, D., et al. (2017). Toward a Metagenomic Understanding on the Bacterial Composition and Resistome in Hong Kong Banknotes. Front. Microbiol. *8*, 632.

Hinchliff, C.E., Smith, S.A., Allman, J.F., Burleigh, J.G., Chaudhary, R., Coghill, L.M., Crandall, K.A., Deng, J., Drew, B.T., Gazis, R., et al. (2015). Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proc. Natl. Acad. Sci. USA *112*, 12764–12769.

Hiramatsu, K., Katayama, Y., Matsuo, M., Sasaki, T., Morimoto, Y., Sekiguchi, A., and Baba, T. (2014). Multi-drug-resistant Staphylococcus aureus and future chemotherapy. J. Infect. Chemother. *20*, 593–601.

Ho, P.L., Cheung, C., Mak, G.C., Tse, C.W., Ng, T.K., Cheung, C.H., Que, T.L., Lam, R., Lai, R.W., Yung, R.W., and Yuen, K.Y. (2007). Molecular epidemiology and household transmission of community-associated methicillin-resistant Staphylococcus aureus in Hong Kong. Diagn. Microbiol. Infect. Dis. *57*, 145–151.

Hsu, T., Joice, R., Vallarino, J., Abu-Ali, G., Hartmann, E.M., Shafquat, A., DuLong, C., Baranowski, C., Gevers, D., Green, J.L., et al. (2016). Urban Transit System Microbial Communities Differ by Surface Type and Interaction with Humans and the Environment. mSystems *1*, e00018–e00016.

Huson, D.H., Mitra, S., Ruscheweyh, H.J., Weber, N., and Schuster, S.C. (2011). Integrative analysis of environmental sequences using MEGAN4. Genome Res. *21*, 1552–1560.

Jeong, H., Sung, S., Kwon, T., Seo, M., Caetano-Anollés, K., Choi, S.H., Cho, S., Nasir, A., and Kim, H. (2016). HGTree: database of horizontally transferred genes determined by tree reconciliation. Nucleic Acids Res. *44* (*D1*), D610–D619.

Kelley, S.T., and Gilbert, J.A. (2013). Studying the microbiology of the indoor environment. Genome Biol. *14*, 202.

Kembel, S.W., Jones, E., Kline, J., Northcutt, D., Stenson, J., Womack, A.M., Bohannan, B.J., Brown, G.Z., and Green, J.L. (2012). Architectural design influences the diversity and structure of the built environment microbiome. ISME J. *6*, 1469–1479.

Kembel, S.W., Meadow, J.F., O'Connor, T.K., Mhuireach, G., Northcutt, D., Kline, J., Moriyama, M., Brown, G.Z., Bohannan, B.J., and Green, J.L. (2014). Architectural design drives the biogeography of indoor bacterial communities. PLoS ONE *9*, e87093.

Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman, F.D., Knight, R., and Kelley, S.T. (2011). Bayesian community-wide culture-independent microbial source tracking. Nat. Methods *8*, 761–763.

Lang, J.M., Darling, A.E., and Eisen, J.A. (2013). Phylogeny of bacterial and archaeal genomes using conserved genes: supertrees and supermatrices. PLoS ONE *8*, e62510.

Lax, S., Smith, D.P., Hampton-Marcell, J., Owens, S.M., Handley, K.M., Scott, N.M., Gibbons, S.M., Larsen, P., Shogan, B.D., Weiss, S., et al. (2014). Longitudinal analysis of microbial interaction between humans and the indoor environment. Science *345*, 1048–1052.

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. *44* (*W1*), W242–W245.

Leung, M.H., Wilkins, D., Li, E.K., Kong, F.K., and Lee, P.K. (2014). Indoor-air microbiome in an urban subway network: diversity and dynamics. Appl. Environ. Microbiol. *80*, 6760–6770.

Leyden, J.J., Marples, R.R., and Kligman, A.M. (1974). Staphylococcus aureus in the lesions of atopic dermatitis. Br. J. Dermatol. *90*, 525–530.

Li, Y.W., Wu, X.L., Mo, C.H., Tai, Y.P., Huang, X.P., and Xiang, L. (2011). Investigation of sulfonamide, tetracycline, and quinolone antibiotics in vegetable farmland soil in the Pearl River Delta area, southern China. J. Agric. Food Chem. *59*, 7268–7276.

Li, J., Sung, C.Y.J., Lee, N., Ni, Y., Pihlajamäki, J., Panagiotou, G., and El-Nezami, H. (2016). Probiotics modulated gut microbiota suppresses hepatocellular carcinoma growth in mice. Proc. Natl. Acad. Sci. USA *113*, E1306–E1315.

Lowy, F.D. (2003). Antimicrobial resistance: the example of Staphylococcus aureus. J. Clin. Invest. *111*, 1265–1273.

McArthur, A.G., Waglechner, N., Nizam, F., Yan, A., Azad, M.A., Baylay, A.J., Bhullar, K., Canova, M.J., De Pascale, G., Ejim, L., et al. (2013). The comprehensive antibiotic resistance database. Antimicrob. Agents Chemother. *57*, 3348–3357.

McMurdie, P.J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE *8*, e61217.

Merton, R.K. (1968). The Matthew Effect in Science: The reward and communication systems of science are considered. Science *159*, 56–63.

OPEN ACCESS CellPress

MetaSUB International Consortium (2016). The Metagenomics and Metadesign of the Subways and Urban Biomes (MetaSUB) International Consortium inaugural meeting report. Microbiome *4*, 24.

MTR Corporation (2006). MTR Uses Nano Technology to Enhance Hygiene Levels. https://thesilveredge.com/study-mtr-uses-silver-nano-technology-to-enhance-hygiene-levels-on-passenger-rail-lines/.

MTR Corporation (2017). MTR Patronage Figures for August 2017. http://www.mtr.com.hk/en/corporate/investor/patronage.php.

Munck, C., Albertsen, M., Telke, A., Ellabaan, M., Nielsen, P.H., and Sommer, M.O. (2015). Limited dissemination of the wastewater treatment plant core resistome. Nat. Commun. *6*, 8452.

Ni, Y., Li, J., and Panagiotou, G. (2016). COMAN: a web server for comprehensive metatranscriptomics analysis. BMC Genomics *17*, 622.

Oh, J., Byrd, A.L., Deming, C., Conlan, S., NISC Comparative Sequencing Program; Kong, H.H., and Segre, J.A. (2014). Biogeography and individuality shape function in the human skin metagenome. Nature *514*, 59–64.

Oh, J., Byrd, A.L., Park, M., NISC Comparative Sequencing Program; Kong, H.H., and Segre, J.A. (2016). Temporal Stability of the Human Skin Microbiome. Cell *165*, 854–866.

Prescott, S.L., Larcombe, D.L., Logan, A.C., West, C., Burks, W., Caraballo, L., Levin, M., Etten, E.V., Horwitz, P., Kozyrskyj, A., and Campbell, D.E. (2017). The skin microbiome: impact of modern environments on skin ecology, barrier integrity, and systemic immune programming. World Allergy Organ. J. *10*, 29.

Pruitt, K.D., Brown, G.R., Hiatt, S.M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., Farrell, C.M., Hart, J., Landrum, M.J., McGarvey, K.M., et al. (2014). RefSeq: an update on mammalian reference sequences. Nucleic Acids Res. *42*, D756–D763.

Robertson, C.E., Baumgartner, L.K., Harris, J.K., Peterson, K.L., Stevens, M.J., Frank, D.N., and Pace, N.R. (2013). Culture-independent analysis of aerosol microbiology in a metropolitan subway system. Appl. Environ. Microbiol. *79*, 3485–3493.

Scholz, M., Ward, D.V., Pasolli, E., Tolio, T., Zolfo, M., Asnicar, F., Truong, D.T., Tett, A., Morrow, A.L., and Segata, N. (2016). Strain-level microbial epidemiology and population genomics from shotgun metagenomics. Nat. Methods *13*, 435–438.

Segata, N., Börnigen, D., Morgan, X.C., and Huttenhower, C. (2013). PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat. Commun. *4*, 2304.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Sommer, M.O.A., Dantas, G., and Church, G.M. (2009). Functional characterization of the antibiotic resistance reservoir in the human microflora. Science *325*, 1128–1131.

Spellberg, B., Guidos, R., Gilbert, D., Bradley, J., Boucher, H.W., Scheld, W.M., Bartlett, J.G., and Edwards, J., Jr.; Infectious Diseases Society of America (2008). The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. Clin. Infect. Dis. *46*, 155–164.

Su, J.Q., An, X.L., Li, B., Chen, Q.L., Gillings, M.R., Chen, H., Zhang, T., and Zhu, Y.G. (2017). Metagenomics of urban sewage identifies an extensively shared antibiotic resistome in China. Microbiome *5*, 84.

The Transport and Housing Bureau of the Government of the Hong Kong Special Administrative Region (2014). The Transport and Housing Bureau of the Government of the Hong Kong Special Administrative Region. http://www.thb.gov.hk/.

van der Helm, E., Imamovic, L., Hashim Ellabaan, M.M., van Schaik, W., Koza, A., and Sommer, M.O.A. (2017). Rapid resistome mapping using nanopore sequencing. Nucleic Acids Res. *45*, e61.

Vandenesch, F., Naimi, T., Enright, M.C., Lina, G., Nimmo, G.R., Heffernan, H., Liassine, N., Bes, M., Greenland, T., Reverdy, M.E., and Etienne, J. (2003). Community-acquired methicillin-resistant Staphylococcus aureus carrying Panton-Valentine leukocidin genes: worldwide emergence. Emerg. Infect. Dis. *9*, 978–984.

Wattam, A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J.L., Gillespie, J.J., Gough, R., Hix, D., Kenyon, R., et al. (2014). PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res. *42*, D581–D591.

Weber, D.J., Rutala, W.A., Miller, M.B., Huslage, K., and Sickbert-Bennett, E. (2010). Role of hospital surfaces in the transmission of emerging health care-associated pathogens: norovirus, Clostridium difficile, and Acinetobacter species. Am. J. Infect. Control *38* (*5, Suppl 1*), S25–S33.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., Lee, S.Y., Fischbach, M.A., Muller, R., Wohlleben, W., et al. (2015). antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. *43* (*W1*), W237–W443.

Wu, N., Qiao, M., Zhang, B., Cheng, W.D., and Zhu, Y.G. (2010). Abundance and diversity of tetracycline resistance genes in soils adjacent to representative swine feedlots in China. Environ. Sci. Technol. *44*, 6933–6939.

Yu, J., Feng, Q., Wong, S.H., Zhang, D., Liang, Q.Y., Qin, Y., Tang, L., Zhao, H., Stenvang, J., Li, Y., et al. (2017). Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut *66*, 70–78.

Zimmerman, A., Bai, L., and Ginty, D.D. (2014). The gentle touch receptors of mammalian skin. Science *346*, 950–954.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
48 palm microbiome samples	This study	https://www.ncbi.nlm.nih.gov/biosample/ SAMN07714726 to SAMN07714761
Critical Commercial Assays		
PowerSoil DNA Isolation kit	MO BIO Laboratories	12888-100
Nextera XT DNA Library Preparation Kit	Illumina	FC-131-1096
ESwab Collection & Transport System	BD company	220246
Deposited Data		
Sequencing data	This study	Sequence Read Archive (SRA): No. SRP119528 under project RJNA413474
The eukaryotic tree of life	Lang et al., 2013; Segata et al., 2013	N/A
PATRIC database	Wattam et al., 2014	https://patricbrc.org/
Lists of potential pathogens	Forsberg et al., 2014; Kembel et al., 2012	N/A
CARD database	McArthur et al., 2013	https://card.mcmaster.ca/
ResFams database	Gibson et al., 2015	www.dantaslab.org/resfams/
Clinically important ARGs	Munck et al., 2015	N/A
VFDB database	Chen et al., 2016	www.mgc.ac.cn/VFs/
HGTree database	Jeong et al., 2016	https://omictools.com/hgtree-tool
NCBI plasmid RefSeq database	Pruitt et al., 2014	https://www.ncbi.nlm.nih.gov/refseq/
Sequencing data of other environmental samples	NCBI Sequence Read Archive	ERP013563, ERP013562, ERP012177, SRP033730, SRP061803
Software and Algorithms		
DIAMOND	Buchfink et al., 2015	https://ab.inf.uni-tuebingen.de/software/diamond
MEGAN5	Huson et al., 2011	https://ab.inf.uni-tuebingen.de/software/megan5/
OpenTree7.0	Hinchliff et al., 2015	https://tree.opentreeoflife.org/opentree/
blast	Altschul et al., 1990	https://blast.ncbi.nlm.nih.gov/Blast.cgi
R 3.2	http://www.r-project.org	http://www.r-project.org
VEGAN	Dixon, 2003	N/A
phyloseq	McMurdie and Holmes, 2013	https://joey711.github.io/phyloseq/
RANGER-DTL 2.0	Bansal et al., 2012	http://compbio.mit.edu/ranger-dtl/
antiSMASH 3.0	Weber et al., 2015	https://antismash.secondarymetabolites.org/
Cytoscape 3.3.0	Shannon et al., 2003	http://www.cytoscape.org/
Spatial Design Network Analysis (sDNA)	Cooper and Alain, 2015	https://www.cardiff.ac.uk/sdna/
SparCC algorithm	Friedman and Alm, 2012	N/A
IndVal test	Dufrene and Legendre, 1997	N/A
iTOL	Letunic and Bork, 2016	https://itol.embl.de/
Bayesian SourceTracker	Knights et al., 2011	N/A
PanPhlan	Scholz et al., 2016	http://segatalab.cibio.unitn.it/tools/panphlan/
Strain-level taxonomic profiling tool	Oh et al., 2014	https://github.com/ohlab/skinmetagenome

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for data and resources and should be directed to and will be fulfilled by the Lead Contact, Gianni Panagiotou (gianni.panagiotou@hki-jena.de).

OPEN ACCESS CellPress

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Sample collection

Each sample was a pooled community from the palms of 6 healthy individuals (without symptoms of skin disease or other illness) who maintained contact with different handrails in the Hong Kong MTR system for 30 minutes (Figures 1A and 1B). For each of the 8 urban lines in the Hong Kong MTR system, samples were collected in the morning (9:00 to 11:30 am) and evening (19:00 to 21:30 pm). Each line was sampled 3 times within 3 consecutive weeks. For each sample, we collected metadata recording the line, sampling date and time, sampling route interval and weather condition of the sampling day (Table S1).

In each sampling section, volunteers followed a standard procedure: 1. Wash hands using normal soap and distilled water for 90 s before boarding the train; 2. Touch different handrail surfaces in a running train of a specific line for approximately 30 minutes; 3. Swab each palm for 2 minutes using the ESwab Collection & Transport System (BD company, pre-moisturized with 0.15M NaCl and 0.1% Tween 20). To provide adequate DNA yield for library preparation of each sample, one swab was used for all 12 palms. Swabs were stored in 15 mL LabServ tubes on ice and transported to −20°C freezer within an hour. DNA extractions were performed within 24 hours after sample collection.

An extra 6 negative controls were prepared (3 from the same palms without touching any surface for 30 minutes, 3 from unused swabs). The DNA yields were undetectable by Qubit (< 0.007 ng/ l in 50 L elution buffer) for all negative controls, suggesting that the DNA that we collected originated from the MTR train surfaces but not from the swabs, buffer solutions or remaining bacteria on palms after washing.

Ethics approval

All participants in the sampling procedure were researchers from the University of Hong Kong, and gave written informed consent. Sequencing reads from human genome were removed before analyses. The study was approved by the Human Research Ethics Committee (HREC) of the University of Hong Kong.

METHOD DETAILS

DNA extraction and metagenomic sequencing

Microbial DNA was extracted with the PowerSoil DNA Isolation kit (MO BIO Laboratories) following the official protocol. 0.95 ± 0.41 ng DNA was obtained for each sample. The extracted DNA was used to construct shotgun metagenomic libraries using the Nextera XT kit following the standard protocol, which is optimized for the library construction for low DNA input (1 ng). The metagenome sequencing was performed with Illumina HiSeq 1500 (101bp PE) at the Centre for Genomic Sciences of The University of Hong Kong. The sequenced reads were processed for quality control to remove the adaptor regions, low quality reads and the reads from human genome following the previously described steps (bwa mapping against GRCh37/hg19 reference genome using mem algorithm, extracting reads with > 95% identity) (Li et al., 2016). 75% to 94% of the reads passed the quality control, with a mean value of 90.4%. The human reads occupied 39% to 78% of the clean reads, with a mean value of 43.8%.

The total DNA amount obtained (0.95 ± 0.41 ng) was much lower than in other studies which collected samples by swabbing surfaces in metro stations directly (Afshinnekoo et al., 2015; Hsu et al., 2016), but comparable to previous human skin metagenomic projects (Oh et al., 2014). Besides the disparity in the microbial density between the handrail and the colonized palm, another possible explanation for the low DNA collected could be the effect of the antibacterial Nano-Silver-Titanium Dioxide Coating (NSTDC) applied to MTR surfaces (MTR Corporation, 2006). The DNA library construction was optimized for the low DNA input and shotgun metagenomic sequencing was performed for all samples. We used our previously developed method to perform functional metagenomics analysis after the metagenomics sequencing (Forsberg et al., 2012; Munck et al., 2015; Sommer et al., 2009; van der Helm et al., 2017), however, the DNA amount of these samples were lower than the minimal requirements for the construction of functional libraries.

Taxonomic profiling

From the metagenomic sequencing data, a million reads were randomly sub-sampled from each sample for taxonomic profiling. These reads were mapped to the NCBI non-redundant (nr) database by DIAMOND (Buchfink et al., 2015) using the default settings. To estimate the taxonomic composition of the samples, the Lowest Common Ancestor (LCA) algorithm was implemented with LCA mapper from mtools, MEGAN5 (Huson et al., 2011) (-f Detect -ms 50 -me 0.01 -tp 50). The relative abundance for each taxonomic level was distille following the previously described pipeline (Heshiki et al., 2017; Huson et al., 2011). The 744 most abundant species (composing 95% of the total reads in all samples) were used in further statistical analyses (Data S1B).

Phylogeny

Synthetic tree of phylogeny and taxonomy for the most abundant bacterial species (composing 90% of the total reads) was acquired from OpenTree7.0 (Hinchliff et al., 2015), integrating the major studies on the phylogenetic trees for bacteria (Lang et al., 2013; Segata et al., 2013).

Reference-based strain mapping

Strain profiling of *P. acnes* and *S. epidermidis* was conducted according to the previously suggested reference-based approach (Oh et al., 2014). The non-core region strain quantifications were used in statistical comparisons (Data S1C and S1D).

Pan-genome analysis

Pan-genome profiling was conducted by PanPhlan using default parameters for the 10 most abundant species and 10 most ARG contributing species (Scholz et al., 2016). Presence/absence and abundance profiles were generated for gene clusters.

Species categorization

The bacterial species were queried against PATRIC database (Wattam et al., 2014) for categorization. The features recorded in genome metadata include: habitat, host, mobility, salinity, altitude/depth, temperature range, oxygen requirement, pathogenicity and antimicrobial resistance. To link the PATRIC features with the taxonomic profiling at the species level, all features in any strain were aggregated and assigned to the corresponding species (Data S3A). For example, 6 strains of *Vibrio ordalii* were marked as aquatic strains in PATRIC, so the abundance of *V. ordalii* at the species level was counted for the category of aquatic species.

Identification of pathogens and potential pathogenic species

To identify opportunistic pathogens, two lists of potential pathogens (Forsberg et al., 2014; Kembel et al., 2012) were combined (Data S2A). The PATRIC database (Wattam et al., 2014) was used as a reference for taxonomic structures and species/strain names. The organisms appearing on the list were marked as opportunistic pathogens, and the species containing at least one opportunistic pathogenic strain were treated as potential opportunistic pathogenic species (Data S2B).

Identification and quantification of ARG mechanism and families

To identify ARGs and assign them to different mechanisms and ARG families, two different procedures were performed. For the ARG mechanism annotation, the one million reads subset used for taxonomic profiling were mapped against CARD database (McArthur et al., 2013) (Data S3B), as described previously (Heshiki et al., 2017). For the ARG family information, all reads were mapped against the extended ResFams database (Gibson et al., 2015). The 27 sequences of clinically important ARGs described previously (Munck et al., 2015) were also included in the extended database, if the sequenced were absent in ResFams (blastp identity < 95%). New clinically important ARG families were built for the absent ARGs in ResFams: all existing ResFams sequences were mapped against those ARGs, and the mappable genes (blastp, -e 1e-5–id 95) were classified into regarding clinically important ARG families. ARG families may have overlaps due to the family ranges defined by ResFams. In both annotation procedures, reads were mapped using DIAMOND blastp with stringent parameters (-e 1e-10–id 95, best hits reserved). The mappable reads were further mapped against the nr database using DIAMOND blastp. If ARGs did not remain the best hits (sorted by e value) in the nr database mapping step, the reads were removed from the ARG mapping results. The coverage rate of each ARG was calculated by using the mappable reads from all samples. Among 136 quantified ARG families, 98 families have a coverage > 50%, 91 families > 70% and 84 families > 90%. The abundances of ARG families were calculated as RPKM (Reads per Kilobase per Million mapped reads) (Data S3C). The overall abundances of all ARG families were calculated as RPM (Reads per Million mapped reads) (Data S4B).

Identification of virulence factors

All sequencing reads were mapped against VFDB (Chen et al., 2016), following the same procedure and parameter of ARG family annotation. The abundances of the virulence factors of the opportunistic pathogens were summarized in RPKM (Data S2C).

Community diversity and dissimilarity

To estimate the alpha-diversity of the microbial community for each sample, the Simpson diversity index was calculated with VEGAN (Dixon, 2003) based on the species-level relative abundance (Table S1).

To estimate the community dissimilarities, weighted UniFrac distances were calculated by phyloseq (McMurdie and Holmes, 2013) based on the relative abundance of species, as well as functional classes (for example, ARG family profiles). Ordinations were calculated by nonmetric multidimensional scaling (NMDS) for the illustration.

Dissemination potential estimation

To estimate the plasmid-mediated horizontal gene transfer (HGT) rate, all ARG sequences were mapped against the NCBI plasmid RefSeq database (Pruitt et al., 2014) using blastp (-e 1e-5–id 95). The protein sequences from all functional ARGs were mapped to the gene families that were acquired from the HGTree database (Jeong et al., 2016). The HGTree gene family of an ARG was defined by the gene family with the highest proportion of valid hits (-e 1e-5 and coverage > 50% in query or subject) in HGTree families. RANGER-DTL (Bansal et al., 2012) was used for phylogenetic reconciliation analysis based on the gene tree and species tree provided by the HGTree. The HGT rate in each family was calculated by dividing the number of HGT events by the total phylogenetic tree length of that family. Finally, the HGT rates were assigned to ResFams ARG families with unique mapping (Data S4A). The ARG families with extremely high HGT rate (> 100) are *blaI* and *aph(3')*.

The community-wide dissemination potential (Data S4B) was estimated for each sample, by as the following equation:

$$\sum \mathrm{HGT}_i * \mathrm{RPKM}_i \quad \text{(Equation 1)}$$

where:

i is an HGTree gene family;
HGT_i is the HGT rate of the family *i*;
RPKM_i is the abundance of the HGTree family *i* in RPKM.

Identification and quantification of biosynthesis gene clusters

BGCs annotation was based on antiSMASH 3.0 (Weber et al., 2015) and modified as previously described (Donia et al., 2014; Ni et al., 2016). All reads were mapped against the database using DIAMOND with the same parameters as ARGs. The abundances of BGCs/BGC types were calculated as RPKM (Data S3D).

Identification of taxonomic links for ARGs and BGCs

For all the reads that could be mapped to ARGs/BGCs, taxonomic profiling was carried out using the same pipeline used for the 1 million read subsamples, with more stringent filtering criteria (bootstrap value > 50 in LCA mapping at species level). Therefore, the major contributing species for each ARG family or BGC type could be identified (**Dataset S7**).

Morphometric modeling of hybrid rail-road urban network

To test the hypothesized correlations between microbial transmission, loads and diversity and underlying urban morphology and connectivity, a series of network models were conducted. As a first step, initial assessment of network topology of the 8 urban lines was conducted by Cytoscape 3.3.0 (Shannon et al., 2003). Averaged edge features (EdgeBetweenness) and node features (including BetweennessCentrality, ClosenessCentrality, Degree, NeighborhoodConnectivity, NumberOfEdges, Radially) were calculated for each line (Data S5A). Line civic features (number of stations/interchange stations, number of interchangeable lines, and average station distance to Shenzhen) were also incorporated in the topology analysis.

The second step involved more detailed network modeling employing spatial Design Network Analysis (sDNA) (Cooper and Alain, 2015), a state-of-the-art urban network analysis technique developed by teams at Cardiff University and The University of Hong Kong to model multiple scale urban morphological metrics (morphometrics). The MTR network database was first coupled with the street network database by digitizing the network of subways linking platforms to station entry-exit points to adjacent street link in ArcGIS 13 and the hybrid network was cleaned, transcribed into an access graph model and subjected to sDNA analysis. Among a suite of various network morphological metrics (morphometrics) generated, betweenness centrality or through-movement potential was employed in the present study. It is the simulated counts of movement through each link in the network, given its position in the network and the geometrical and topological connectivity with other links within the network. sDNA betweenness of x in a graph of N links may be defined as:

$$Bt\,WI(x) = \sum_{y\,N\,z} \sum_{\boldsymbol{R_y}} L(y)L(z)P(z)OD(y\ z\ x) \quad \text{(Equation 2)}$$

where:

y and z are the geodesic end points;
R_y is the set of links within a defined radius from y;
L(y) and L(z) are length of links y and z respectively;
P_z is the proportion of link z within the defined radius.

OD is a function defined as:

$$OD = \begin{cases} 1 & \textit{if } x \textit{ is on the geodesics from } y \textit{ to } z \\ \frac{1}{2} & \textit{if } x \equiv y \neq z \\ \frac{1}{2} & \textit{if } x \equiv z \neq y \\ \frac{1}{2} & \textit{if } x \equiv z \equiv y \\ 0 & \textit{otherwise} \end{cases} \quad \text{(Equation 3)}$$

It was modeled at micro-meso-macro-level urban scales (100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 5000, 7500, 10000, 15000 m street catchments) to capture potential effects of scale on human mobility and microbial transmission (Data S5B). For each MTR station, betweenness values were aggregated at three levels: with respect to all the street links closest to all exits of a

metro station, all street links within a 50 m and 100 m catchment radii. For each MTR line, the street connectivity index was represented by the mean betweenness for all the links for the line as well as sDNA betweenness values of all stations on it. Other urban measures employed in the study included network proximity to two transportation hubs, namely Hong Kong Central station and Shenzhen Railway Station and Luohu metro link.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

All statistical analyses were performed in R. For comparative analysis between AM and PM samples (including taxa, species categories, ARG families, BGC types, etc.), paired t tests (for abundance/proportion) or Wilcoxon signed-rank tests (for rankings) were performed. For comparative analysis among different lines or different sampling days, ANOVA (for abundance) or Kruskal-Wallis tests (for rankings) were performed. For line-specific species categories, Wilcoxon tests with one-against-all comparison were performed. To discover line-specific signatures for ARG families and BGC types, IndVal tests (Dufrene and Legendre, 1997) were performed. Alpha value of 0.05 was used for all statistical tests.

Co-occurring and co-excluding relationships among taxa, ARG families and BGC types were identified by SparCC algorithm (Friedman and Alm, 2012). Correlations with coefficient absolute value greater than 0.3 were reserved (Data S6).

Pearson's correlation coefficients were calculated between topological or geographical features and biological features (overall ARG abundance, clinically important ARG family abundance, community-wide dissemination potential, ER line signature ARG families' abundance, etc.). Correlations with p value < 0.05, or adjusted q value < 0.1 with Bonferroni correction for multiple comparisons, were marked as significant correlations.

Data visualization

R, python and corresponding packages including ggplot2, gplots, heatmap.plus and matplotlib were used for illustration of statistical results. iTOL (Letunic and Bork, 2016) was used for the phylogeny visualization. Cytoscape 3.3.0 (Shannon et al., 2003) was used to visualize the analyses incorporating network and topology.

DATA AND SOFTWARE AVAILABILITY

Data from the other studies

The sequencing data of other environmental samples used from comparison were obtained from NCBI Sequence Read Archive (accession number: ERP013563, ERP013562, ERP012177, SRP033730 and SRP061803).

Availability of data and materials

The sequencing reads files were deposited into the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) with accession number SRP119528 under project PRJNA413474. The datasets supporting the conclusions of this study are included within the article and the additional files.