Inhalable antibiotic resistome from wastewater treatment plants to urban areas: Bacterial hosts, dissemination risks, and source contributions

Jiawen Xie[a,b], Ling Jin[a,b,c], Dong Wu[d], Amy Pruden[e], Xiangdong Li[a,b*]

[a] *Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, China*

[b] *The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen 518057, China*

[c] *Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, China*

[d] *Key Laboratory for Urban Ecological Processes and Eco-Restoration, School of Ecological and Environmental Science, East China Normal University, Shanghai 200241, China*

[e] *Department of Civil & Environmental Engineering, Virginia Tech, Blacksburg, Virginia 24060, United States*

Jiawen Xie (jiawen.xie@connect.polyu.hk)

Ling Jin (ling.jin@polyu.edu.hk)

Dong Wu (dwu@des.ecnu.edu.cn)

Amy Pruden (apruden@vt.edu)

*Corresponding author

Email: cexdli@polyu.edu.hk

Tel: (852) 2766-6041

Fax: (852) 2334-6389

Abstract

Antibiotic resistance genes (ARGs) are commonly detected in the atmosphere, but questions remain regarding their sources and relative contributions, bacteria hosts, and corresponding human health risks. Here we conducted a qPCR- and metagenomics-based investigation of inhalable fine particulate matter ($PM_{2.5}$) at a large wastewater treatment plant (WWTP) and in ambient air of Hong Kong, together with an in-depth analysis of published data of other potential sources in the area. $PM_{2.5}$ was observed with increasing enrichment of total ARGs along the coastal–urban–WWTP gradient and clinically relevant ARGs commonly identified in urban and WWTP sites, illustrating anthropogenic impacts on the atmospheric accumulation of ARGs. With certain kinds of putative antibiotic-resistant pathogens detected in urban and WWTP $PM_{2.5}$, a comparable proportion of ARGs co-occurred with MGEs was found between the atmosphere and WWTP matrices. Despite similar emission rates of bacteria and ARGs within each WWTP matrix, about 11–13% of the bacteria and >57% of the relevant ARGs in urban and WWTP $PM_{2.5}$ were attributable to WWTPs. Our study highlights the importance of WWTPs in disseminating bacteria and ARGs to the ambient air from a quantitative perspective, and thus the need to control potential sources of inhalation exposure to protect the health of urban populations.

Keywords

WWTP; inhalable antibiotic resistome; emission rate; ARG host; pathogen; health risk.

Synopsis: The study highlights the important role of WWTPs in disseminating bacteria and antibiotic resistance genes (ARGs) to the ambient air of urban areas from a quantitative perspective.

1. Introduction

Antibiotic resistance has become a growing threat to public health, especially in the last 2-3 decades. It is estimated that the continued development of antimicrobial resistance will result in 10 million deaths per year and 100 trillion USD in economic losses by 2050 [1]. Some ARGs are intrinsic to some bacteria; however, other ARGs are mobile and can be horizontally transferred to a broad range of host bacteria, a phenomenon that appears to be exacerbated by anthropogenic activities [2]. Inputs of antimicrobials and antibiotic resistance genes (ARGs) to terrestrial and aquatic compartments from various human, agricultural, and industrial sources have now been widely documented [3-10], but only recently is the atmosphere also beginning to be recognized as a potential key reservoir of ARGs [11-14]. Notably, ARGs have been found to be associated with airborne particles. In particular, ARGs associated with fine particulate matter ($PM_{2.5}$) pollution present a concern because $PM_{2.5}$ can penetrate the human lung. While other studies have begun to consider the implications of ARG exposure via water/food ingestion and skin contact [15,16], there is a need to better characterize the range of inhalation exposures to ARGs. Such exposures are of particular concern when it can be confirmed that ARGs are carried within pathogens suspended in the atmosphere, as has been suggested by prior studies [13,17-19].

Recent studies are painting a picture of differential ARG distribution in the atmosphere at global and regional scales [11,16], implying that spatial heterogeneity of ARG sources is likely, including in the numbers and types of ARGs emitted. This is in addition to varying atmospheric conditions (e.g., smog) that can influence the transport of resistant strains and potentially also impose selective pressure on their survival. Various potential sources of airborne ARGs have been recognized, including typical anthropogenic ARG hotspots, such as livestock farms, wastewater treatment plants (WWTPs), hospitals, and solid waste treatment systems [20-22].

Diminishing concentrations of source-indicative bacterial and ARG species in the atmosphere with increasing distance from putative source sites, along with close resemblance of ARG profiles between putative sources and nearby aerosols provide strong evidence of key anthropogenic sources [22,23]. However, there is a need to move towards quantifying ARG contributions from various putative sources to the ambient air. Recently, some quantitative source-predictive models, such as SourceTracker [24], have been applied towards source apportionment of ARGs found in sediment and urban stream samples [25,26]. They could potentially be extended to airborne ARGs.

Among putative airborne sources, WWTPs are prime candidates for consideration. They are fundamental to urban infrastructure and are known to receive high concentrations of antimicrobials, antibiotic-resistant bacteria, pathogens, and ARGs. Recent studies indicate that the treatment processes (e.g., aeration) of WWTPs tend to produce bioaerosols of less than 1–2 μm in size, which is within the respirable range [27,28]. Some studies have correspondingly noted the frequent occurrence of bioaerosol-inducing respiratory symptoms among WWTP operators [29]. In contrast to bacteria, few studies have touched on the field of ARG emissions from wastewater and sludge, especially through metagenomic-based approaches. One recent study provided qualitative comparisons of antibiotic resistome (i.e., total ARGs carried across a sample) profiles between activated/dewatered sludge and bioaerosols, but did not provide a quantitative assessment, such as emission rates [30]. Moreover, antibiotic-resistant pathogens are known to exist in WWTP bioaerosols, as confirmed by culture-dependent methods [31]. The systematic screening of metagenomic data to identify linkages of ARGs with potential pathogenic host bacteria in WWTP aerosols of inhalable range could help to better define the range and extent of associated inhalation exposure and risks, in particular with considerations of World Health Organization (WHO) priority pathogens and clinically-relevant ARGs [32,33].

Further examining the co-occurrence of ARGs with mobile genetic elements (MGEs) could provide additional insight into the extent to which horizontal gene transfer (HGT) occurs among airborne bacteria.

Given the hypothesis that WWTPs are a major source contributor to urban airborne bacteria and ARGs, we conducted a qPCR- and metagenomic-based study of the bacterial and ARG profiles of various compartments and surrounding $PM_{2.5}$ at a large WWTP in Hong Kong, with urban and coastal $PM_{2.5}$ included as well. Based on assembled contigs, we assessed the mobility of ARGs and potential carriage by pathogenic bacteria, i.e., the "resistome risk" [34] of airborne ARGs across urban land-use types. Metagenomes of samples in the current study were further congregated with publicly available sequencing data of other putative emission sources to predict the source profiles of airborne bacteria and ARGs in WWTP and urban areas. A simple one-box model was then applied to estimate the airborne emission potential of bacteria and ARGs in WWTP-relevant matrices to surrounding environments. Overall, we aimed to quantify the contribution of WWTPs to airborne bacteria and ARGs in urban environments, which can help to advance understanding of the potential airborne antibiotic resistance exposure risks of large urban populations.

2. Materials and Methods

2.1 Site description and sampling strategy

A $PM_{2.5}$ sampling campaign was deployed at the Stonecutters Island Sewage Treatment Works (SCISTW), the campus of The Hong Kong Polytechnic University (PU), and the Hok Tsui Regional Air Monitoring Station (HT) in Hong Kong, representing WWTP, urban, and coastal background environments, respectively (Fig. 1A; Table S1). SCISTW is the largest sewage treatment plant in Hong Kong, serving almost half the city's population, with a daily capacity

for treating 2.45 million m^3 of sewage. It applies the process of chemically enhanced primary treatment (CEPT) without biological treatment before discharging the effluent. The sampling period lasted from 24 to 30 October 2016 at SCISTW and HT, and from 28 October to 4 November 2016 at PU. 24-hour $PM_{2.5}$ samples were collected daily on pre-baked (500°C for 5 h) quartz filters (8 × 10 $in.^2$, PALL) using high-volume air samplers (TH-1000C II, Wuhan Tianhong Instruments Co., Ltd., China) at a flow rate of 1 m^3 min^{-1}. Single samplers were set up to collect $PM_{2.5}$ samples at PU and HT, while four sites were selected for simultaneous sampling at SCISTW, namely the flocculation tank (FT), sedimentation tank (ST), primary sludge storage tank (SST), and barge facility for primary sludge transportation (BF) (Figs. 1B and S1). One sampler was placed beside each of the abovementioned sites, except the BF site equipped with two samplers in 2 m intervals as duplicates. A blank filter was placed in a non-operating sampler at the flocculation tank, PU, and HT for quality control. Correspondingly, aqueous sewage samples, including screened influent, flocculation tank sewage, sedimentation tank effluent, and final effluent after de-chlorination, as well as CEPT sludge (primary) before centrifugation and sludge cake (primary) ready for transport to be incinerated or placed in a landfill site, were collected daily from 24 to 28 October 2016 in SCISTW (Fig. 1B). Samples were kept at 4°C and transported immediately to the laboratory for analysis. More detailed information on SCISTW, the meteorological parameters, and the collected samples (sampling period and sample size) is provided in Section S1, Figs. S2 and S3, and Tables S2 and S3 in the Supporting Information.

2.2 Sample pretreatment and DNA extraction

Since quartz filter is not suitable for direct DNA extraction due to its hygroscopicity, each $PM_{2.5}$ sample was cut into segments and repeatedly sonicated with sterilized 1× phosphate-buffered saline (PBS) for three times (20 min each time) in an ice-water bath to get particles

off from quartz filters. PBS extracts from the same sample were combined and filtered through a 0.2-μm polyethersulfone (PES) membrane disc filter (47 mm, PALL). Blank filters were treated simultaneously using the same operation procedure. Influent, flocculation tank sewage, and sedimentation tank sludge collected from the sludge storage tank (~100 mL) were centrifuged at 9840 *g* and 4°C for 10 min upon arrival at the laboratory; the supernatant was carefully discarded, and the pellets were kept. With regard to the sedimentation tank effluent and de-chlorinated discharge, 250 mL solutions were filtered through 0.2-μm PES membranes (PALL). All samples were preserved at −80°C before DNA extraction. All the tools and consumables used were sterilized.

DNA extraction was conducted using the FastDNA SPIN Kit for Soil (MP Biomedicals) according to the manufacturer's instructions, except for a modified purification step involving the use of Agencourt AMPure XP beads (Beckman Coulter) for $PM_{2.5}$ samples [17,35]. The 0.2-μm membrane disc filters of effluent and $PM_{2.5}$ were first cut into pieces, while the pellets from sewage and sludge (around 150 mg) were directly used as inputs for the DNA kit. DNA extraction efficiency of the current method for air samples has been estimated in our prior study [36] and listed in Table S4. For each sewage/sludge/effluent sample, DNA extracted in triplicate were pooled together to reduce heterogeneity and then purified using the DNeasy PowerClean CleanUp Kit (QIAGEN). DNA extracts of $PM_{2.5}$ from every three or four consecutive days in each site were combined to ensure sufficient material for a downstream analysis; DNA extracts of other samples from the same sites were combined in a manner corresponding to the time divisions of the $PM_{2.5}$ samples and based on the daily flow of influent (Table S2). Finally, 13 $PM_{2.5}$ DNA samples covering the entire sampling period, 3 DNA samples of the field blanks, and 3 sewage, 2 effluent, and 2 sludge DNA samples covering around half of the sampling period were used for downstream experiments (Table S3).

176

177 **2.3 qPCR quantification of 16S rRNA genes and shotgun metagenomic sequencing**

178 DNA extracts of $PM_{2.5}$ and sewage/sludge/effluent samples were diluted 10 and 100 times,
179 respectively, to minimize PCR inhibition before quantification of the 16S rRNA gene on a
180 StepOnePlus Real-Time PCR System (Applied Biosystems). The dilution factor was
181 determined by testing a number of randomly selected samples (for details of qPCR, please refer
182 to Section S2 of the Supporting Information). Subsequently, around 25-100 ng undiluted DNA
183 of each sample was used for low-input library construction and shotgun metagenomic
184 sequencing on an Illumina Hiseq X Ten platform with the PE150 strategy. In total, 444 GB of
185 clean data from 20 samples (excluding field blanks that failed in library construction) were
186 obtained after trimming sequencing adaptors and filtering low-quality reads using fastp [37].
187 The sequencing data were uploaded to the NCBI BioProject database
188 (https://www.ncbi.nlm.nih.gov/bioproject/) with the accession number of PRJNA693982.

189

190 **2.4 Bioinformatic and statistical analysis**

191 Taxonomy classification was conducted in Kraken 2 (v2.0.8-beta) [38] and Bracken (v2.5.0)
192 [39] using the standard Kraken 2 database. ARGs were predicted using the DeepARG (v1.0.2)
193 pipeline [40] (--deeparg_identity 70 --gene_coverage 5 --deeparg_probability 0.8) and doubled
194 checked by manual inspection. Alpha diversity (Shannon index) was calculated ("vegan"
195 package [41] in R) after a rarefaction of all sequencing data to 17,754,687 reads per sample,
196 the minimum read number in all of the samples (Table S5), using the seqkit tool [42]. Venn
197 diagrams and Upset plots revealing the intersections of bacterial communities and ARG
198 profiles among different sample types were drawn using the "UpSetR" package [43] in R.
199 Afterwards, a non-metric multidimensional scaling (NMDS) analysis was conducted to
200 differentiate the bacterial and ARG profiles among different samples based on Bray–Curtis

dissimilarity, which was followed by a further identification of the determined biomarkers with a linear discriminant analysis (LDA) effect size (LEfSe) method [44]. Procrustes analysis was then used to check the similarity of bacterial and ARG distribution. Permutational Multivariate Analysis of Variance (PERMANOVA) and Dunn's test for a post-hoc Kruskal-Wallis comparison were conducted using a "vegan" package [41] in R.

To evaluate the co-occurrence patterns of ARGs, MGEs, and bacteria, clean sequencing data were assembled on the MetaStorm platform [45] using the built-in iterative de Bruijn graph *de novo* assembler for short reads sequencing data with highly uneven sequencing depth (IDBA-UD) [46]. The output contigs (summary statistics in Table S5) were then deposited in the NanoARG platform [47] for annotation of ARGs, MGEs, and putative bacterial hosts. Briefly, ARGs were predicted against DeepARG database using DeepARG-LS model (--iden 70 --prob 0.5 --cov 40) [40]; MGEs were aligned to a combined NCBI non-redundant database and integron-integrase (I-VIP) database using DIAMOND (--id 70 --evalue 1e-5 --query-cover 40) [48]. Taxonomy annotation was conducted using Centrifuge with default parameters [49]. Samples with less than 10 ARG-carrying contigs assembled were excluded from downstream contig-based analyses. Human pathogens, including those known to commonly manifest in resistant infections, such as ESKAPE (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp.) and other WHO priority pathogens in urgent need of novel antibiotics for treatment (Table S6), were identified according to the pathogen list (species level) summarized in a previous study [50] with minor updates (Table S7). The genetic location (chromosome or plasmid) of the identified ARGs was analyzed using PlasFlow [51]. ARGs co-localized with integrase/transposase/recombinase-encoding genes were considered potentially mobile. Assembled data were uploaded to the online MetaCompare platform to evaluate the potential

horizontal dissemination risks of ARGs [34], which was based on the co-occurrence patterns of ARGs, MGEs, and pathogen-like sequences on assembled contigs. In addition, although constrained by the total sequencing depth, metagenome-assembled genomes (MAGs) were constructed using metaWRAP (v1.2.1) [52] and annotated for ARGs, MGEs, and bacterial hosts (detailed information in Section S3 in the Supporting Information).

2.5 Source tracking and estimation of bacterial and ARG emissions from WWTPs

SourceTracker (v1.0.1) [24] was adopted in R with default parameters to explore the contributions of diverse putative sources (*e.g.*, soil, freshwater, seawater, sediment, landfill leachate, human skin, oral cavity, human airway, and WWTP sewage/sludge/effluent) to the airborne bacteria and ARGs. About 345 shotgun metagenomic sequencing data in total, including 338 downloaded from the NCBI and MG-RAST databases (listed in Table S8) and 7 from this study (sewage/effluent/primary sludge), were used as source samples in source tracking, while the remaining 13 $PM_{2.5}$ samples were treated as sink samples. To focus on the local situation, most of the downloaded samples were from Hong Kong and South China, except those without suitable metagenomic data from Hong Kong. Considering that WWTPs with secondary treatment systems could also be an important source, sequencing data of activated sludge from the largest secondary WWTP in Hong Kong (retrieved from public database) were also included in source tracking. All data were subjected to the same bioinformatic analysis pipeline in fastp, Kraken 2, Bracken, and DeepARG. Bacterial profiles at the species level and ARG profiles at the ARG subtype level, after removing those presented in less than half of the samples in each source, were used as the input for SourceTracker. The prediction performance of SourceTracker based on the current source data was evaluated using leave-on-out cross-validation strategy as proposed by Li et al. [25]

To further estimate the emission of bacteria and ARGs from different WWTP media to the ambient air ($PM_{2.5}$-based), a simple one-box model was applied for calculation (Fig. 1C) with the following assumptions: 1) the flux of $PM_{2.5}$-associated bacteria and ARGs in and out of the box was from top to bottom; 2) there was no deposition inside the box; and 3) $PM_{2.5}$-associated bacteria and ARGs were well mixed within the box and under a steady state during the sampling period. In sum, SCISTW was regarded as a cubical box of length *L*, width *W*, and height *H*. The emission potential of bacteria (using the 16S rRNA gene as an indicator) and ARGs from different WWTP matrices within the box was calculated based on the law of conservation of mass following Equations 1–4.

$$\frac{d(LWHC)}{dt} = E_{total} + (C_{in} - C)v_L WH + (C_{in} - C)v_w LH; \text{ for a steady state, } \frac{dC}{dt} = 0$$

Equation 1

$$E_i = E_{total} p_i, \sum_i p_i = 1 \text{ and } i = \text{sewage, sludge, and effluent}$$

Equation 2

$$E_i = Q_i \alpha_i, i = \text{sewage, sludge, and effluent}$$

Equation 3

$$Emission\ proportion_i = \frac{\alpha_i}{Conc_i}, i = \text{sewage, sludge, and effluent}$$

Equation 4

In Equation 1–4, C_{in} and C represent the concentration of target genes ($PM_{2.5}$-associated 16S rRNA gene or ARGs [copy m^{-3}]) in the inflow air and air within the box, respectively. t refers to time (s). v_L and v_w are the decomposed wind speeds (m s^{-1}) parallel to the length and width direction of the box, respectively, while E is the emission of target genes per unit of time (copy s^{-1}) from the whole SCISTW (*total*) or different matrices (i) inside. p_i refers to the contribution of the individual matrix (i) to $PM_{2.5}$-associated bacteria or ARGs within SCISTW, which can be converted from the result of SourceTracker. Q_i is the daily flow of sewage/effluent (m^3 d^{-1}) or CEPT sludge (kg d^{-1}). α_i is the emission of target genes (respirable size) per unit volume of

276 sewage/effluent (copy m^{-3}) or per unit weight of sludge cake (copy kg^{-1}) to the atmosphere.
277 $Conc_i$ refers to the concentration of target genes in sewage/effluent (copy m^{-3}) or sludge (copy
278 kg^{-1}). In this study, 350 m in length, 285 m in width, and 12 m in height were adopted according
279 to the occupied area of SCISTW and the height of flocculation/sedimentation tanks plus the
280 height of the air sampler. The concentration of $PM_{2.5}$-associated 16S rRNA genes (or ARGs)
281 in HT with the contribution from WWTP deducted, and that in SCISTW were used as C_{in} and
282 C, respectively. The amount of 16S rRNA genes ($PM_{2.5}$, sewage, sludge, and effluent) was
283 quantified by qPCR, while the total amount of ARGs was the product of the amount of 16S
284 rRNA genes and the relative abundance of ARGs (normalized to the 16S rRNA genes)
285 predicted by DeepARG. v_L and v_w were calculated from the wind speed and wind direction
286 during the sampling period recorded by our weather station (Fig. S2). The daily inflow of
287 influent (Table S2) was used as Q_{sewage} and $Q_{effluent}$, while Q_{sludge} was estimated to be
288 600,000 kg d^{-1} during the sampling period.
289
290 **3. Results and discussion**
291 **3.1 Differences in bacterial community composition in $PM_{2.5}$ compared with aqueous and**
292 **solid WWTP samples**
293 Overall, more than 99% of the annotated bacterial reads could be assigned to the species level
294 (Fig S4). A high number of shared bacterial species between sewage/primary sludge/effluent
295 and $PM_{2.5}$ (97% of all bacterial species identified) was observed (Fig. S5A). This suggests that
296 $PM_{2.5}$ could be substantially influenced by WWTPs as a source, even significantly downwind.
297 However, the $PM_{2.5}$ bacterial community structure at the species level was significantly
298 different from that in sewage/primary sludge/effluent samples in SCISTW ($p < 0.001$) (Figs.
299 2A and S6), even though the bacterial species they contained were similar. As revealed by a
300 LefSe analysis, Actinobacteria, which play an important role in decomposing organic materials

[53], could be indicative taxa in the atmosphere, in contrast to sewage, primary sludge, and effluent in SCISTW rich in *Pseudomonas* (LDA score > 4.5, $p < 0.05$) (Fig. 2B). When making comparisons along the land-use transect, the bacterial community structure in urban and SCISTW $PM_{2.5}$ was more stable than the coastal background $PM_{2.5}$, as reflected by the higher Shannon index scores (Fig. 2C) and was well-separated from them (Fig. 2A). This pattern is consistent with different dominant sources of airborne bacteria across local geographical locations under different levels of disturbance from human activities. Compared with coastal areas, an increase of over one order of magnitude in the concentration of $PM_{2.5}$-associated 16S rRNA gene (indicating bacterial loads) was noted in urban/SCISTW sites (Fig. 2D), which also suggests higher absolute contributions of potential sources and a more bacteria-favorable atmospheric condition in human-impacted areas.

3.2 Broad spectrum of ARGs found in WWTP and urban $PM_{2.5}$

More than eight dominant types of ARGs (their relative abundance normalized to the 16S rRNA gene > 0.05) were identified in most of the samples (Fig. S7). The total relative abundance (Fig. 3A) and richness (subtype number) (Fig. S5B) of ARGs in $PM_{2.5}$ were elevated across the coastal–urban–SCISTW transect, finally becoming comparable to the sewage/primary sludge/effluent samples in SCISTW. This remarkable pattern highlights a highly diverse airborne resistome under potential impacts by human activities, especially from WWTPs. This pattern also mirrors the ranking of major ARG types by relative abundance across the different samples. In general, the descending order of relative abundance was: multidrug resistance (0.62–1.66 copies/16S rRNA gene) > peptide, macrolide-lincosamide-streptogramin (MLS), and aminoglycoside resistance (0.08–0.48 copies/16S rRNA gene) > tetracycline and beta-lactam resistance (0.04–0.26 copy/16S rRNA gene) > rifamycin and glycopeptide resistance (0.01–0.14 copy/16S rRNA gene) (Fig. S7). This pattern is

qualitatively consistent with the view that WWTPs are an important source of ARGs in urban air environments, although we may not confirm the extent to which the proportion of the observed similarity between $PM_{2.5}$ and WWTP matrices was attributed to the inherent baseline of ARGs across environment.

Even though the dominant drug resistance categories appeared to be those shared between $PM_{2.5}$ and other samples, as aforementioned (Fig. S7), $PM_{2.5}$ ARG profiles with a broad composition at the subtype level were clearly separated from those of sewage/primary sludge/effluent samples (Fig. S8). Since the ARG richness in human impacted $PM_{2.5}$ is comparable to that in sewage/primary sludge/effluent (Fig. S5B), the lower Shannon Index scores of the ARG profiles in $PM_{2.5}$ samples (Fig. 3B) indicated a markedly lower evenness of airborne ARGs in urban/WWTP sites compared to the aqueous and solid WWTP samples. The LEfSe analysis further served to identify a series of indicator ARGs that effectively discriminated between the ARG profiles of airborne particles versus other sample types. These discriminatory ARGs were distributed across various drug resistance types and included: *tetA* (*48*), conferring resistance to tetracycline; *bacA* and *rosB*, conferring resistance to peptide antibiotics; and *mtrA*, *mexF*, and *cpxR*, conferring resistance to multiple antibiotics (LDA score > 3.5, $p < 0.05$) (Figs. #c and S9).

Yang et al. [20] observed similar airborne ARG patterns: resistances to multiple drugs and bacitracin (a kind of polypeptide) were dominant in nearby WWTPs and urban areas. Remarkably, some multidrug resistance genes, such as *mexF*, were prevalent and enriched in airborne particles in poultry farms and WWTPs at a higher relative abundance than that observed in animal feces and activated sludge, respectively [20]. Observations in the present study and Yang et al. [20]'s suggest the possibility to identify ARGs that are indicative of

contaminants from different sources. Interestingly, there was a lack of correlation between the airborne antibiotic resistome and the taxonomic composition of the bacterial community (Fig. 3D); thus, such indicators would likely be associated with a limited number of bacterial hosts [16]. In addition, it was of concern to discover a number of clinically-relevant ARGs, which encode resistance to certain kinds of antibiotics in the "watch" and "reserve" groups proposed by the WHO [54]. Such ARGs are prominent airborne biomarkers in human-impacted areas. For example, *bla*$_{SHV}$, which confers resistance to beta-lactam, and *vanR* and *vanA*, which encode resistance to vancomycin, were enriched in urban $PM_{2.5}$. Additionally, *mcr-5*, which encodes resistance to colistin (peptide), was more abundant in SCISTW air in comparison with urban and coastal air (Fig. 3C and Table S10; LDA score > 2.5, $p < 0.05$).

3.3 Co-occurrence patterns of ARGs, MGEs, and potential hosts to indicate the environmental resistome risks

3.3.1 Potential mobility of airborne ARGs across sampling sites

An analysis of the assembled contigs generally indicated that ARGs were widely distributed on both chromosomes and plasmids in bacterial genomes. Notably, plasmids were found to be the dominant genetic location for ARGs detected in $PM_{2.5}$ samples, in contrast to the comparable roles between plasmid and chromosome in other more compact media in SCISTW (Fig. 4A). However, when considering the potential mobility of ARGs, only a small minority of the identified ARGs (at the subtype level) co-occurred with MGEs (i.e., integrons/transposons) on the same contig: <2% in sewage/primary sludge/effluent, around 3% in urban $PM_{2.5}$, and none in coastal and SCISTW $PM_{2.5}$ (Table S11 and Fig. S10A). The potentially mobile ARGs (PM-ARGs) in urban $PM_{2.5}$ were less associated with humans, such as *sul1* (sulfonamide-resistant) and *aadA* (aminoglycoside-resistant), which are commonly found in the environment [32]. A low proportion of ARGs co-occurred with MGEs at

contig/genome level have also been observed previously: in the metagenomic investigation only a small portion of ARGs were found to have been carried physically on MGEs (<20% ARGs on plasmid or integron), despite the fact that ARGs and MGEs were widely detected and frequently found to be correlated in varying kinds of environmental samples, including human and animal feces, influent/effluent and activated/digested sludge from WWTPs, soil, sediment, permafrost, and natural water [55,56]. Moreover, in stark contrast to the diverse range of ARGs encountered across the samples (Fig. S7), the potentially mobile ARGs (PM-ARGs) identified in urban $PM_{2.5}$ mainly appeared to only confer resistance to aminoglycoside and sulfonamide (Figs. S10B and S12). The differences between the total ARG and PM-ARG profiles suggest that dissemination rates will vary across drug resistance types.

3.3.2 Conserved bacterial hosts of airborne antibiotic resistance, including human pathogens

Proteobacteria and Firmicutes were found to be the phyla most highly associated with airborne ARGs encountered in this study, regardless of sampling location, carrying 30%–83% and 13%–46% of the total identified ARGs in the airborne bacterial community, respectively. In SCISTW and urban sites, a substantial portion of airborne ARGs were associated with Actinobacteria (24% and 17%, respectively) (Fig. S10A). This observation also corresponded to the MAG-based result on ARB profile (Fig. S11). At the species level, ARGs in $PM_{2.5}$ near aqueous sources (in coastal areas and sites near flocculation/sedimentation tanks in SCISTW) tended to be associated with potential pathogens, including *Ralstonia picketti* and *Ralstonia insidiosa*, which prefer moist environments. In contrast, in the urban sampling location and near SCISTW sludge sites, ARGs were associated with a broader distribution of both pathogenic and ostensibly benign species (Fig. 4B and Table S12).

401 DNA markers pertaining to putative ESKAPE and priority pathogens identified by the WHO
402 (Table S6) were found to be associated with ARGs on assembled contigs from WWTP and
403 urban $PM_{2.5}$ samples. A lower total relative abundance of these putative antibiotic-resistant
404 pathogens (normalized to the total contig number) was estimated in $PM_{2.5}$ (14.4 ± 18.6 ppm in
405 the urban site and 7.3 ± 3.3 ppm in SCISTW) compared with other potential source matrices
406 in SCISTW (98.0 ± 26.3 ppm); however, different species were dominant in airborne
407 communities (Fig. S13). In contrast to *K. pneumoniae* as the predominant species in
408 sewage/primary sludge/effluent, *A. baumannii* and *S. aureus* were found to be the predominant
409 antibiotic-resistant priority pathogens in SCISTW and urban $PM_{2.5}$, respectively. Antibiotic-
410 resistant *A. baumannii* existed in considerable abundance both in SCISTW air and in sewage,
411 primary sludge, and effluent, but was not detected in urban air in the present study. This
412 observation suggests that antibiotic-resistant *A. baumannii* might possibly be disseminated
413 from aqueous and solid matrices in SCISTW to the nearby air, but that it might not be easy for
414 this species to survive airborne transport to downwind areas. It was also worth noting that there
415 has been a remarkable increase in hospital-associated cases of methicillin-resistant *S. aureus*
416 in Hong Kong over the past few years, as well as a noticeable rate of detection of carbapenem-
417 and multidrug-resistant *Acinetobacter*, extended-spectrum β-lactamase (ESBL)-producing and
418 cephalosporin-resistant *E. coli* and *Klebsiella*, and erythromycin- and penicillin-resistant *S.*
419 *pneumoniae* [57]. While certain kinds of the abovementioned resistant species were identified
420 in SCISTW air, including multidrug-resistant *A. baumannii*, vancomycin-resistant *E. faecium*,
421 and macrolide-resistant *S. pneumoniae*, the ARGs co-occurring with these putative priority
422 pathogens in urban $PM_{2.5}$ in this study were different from the drug resistance types associated
423 with the clinical cases in Hong Kong (Table S13). This suggests that these particular ARGs
424 with clinical relevance are less likely to directly migrate from a clinical environment to urban
425 areas via airborne communication or fail to persist long in air after transmission. However, this

result could have been too limited by the current sequencing depth to have allowed for a comprehensive screening of antibiotic-resistant pathogens, and there was a lack of verification of resistance by culture-based methods.

Compared to the host profile of total ARGs in $PM_{2.5}$, PM-ARGs that were only detected in urban sites were observed to be harbored by fewer taxa, i.e., only Proteobacteria (dominant) and Firmicutes (Fig. S10). PM-ARGs appeared to be similarly constrained taxonomically in sewage/primary sludge/effluent samples as well. Such findings correspond to a previous study on *intI1*, showing that ARG-carrying integrons were highly conserved in Proteobacteria [55]. In general, across this study, the associated range of bacterial taxa appeared to diminish from the whole microbial community to ARG hosts, and then to PM-ARG hosts. Given the uncertainty in the accuracy of the assembly of metagenomic data [58], future studies to more closely examine the postulated relationships identified here with culture-based studies would be beneficial.

3.3.3 Comparable resistome risk score of $PM_{2.5}$ among coastal, urban, and WWTP sites

Based on the association of ARG, MGE, and putative pathogens annotated on assembled contigs, MetaCompare was employed to rank the resistome risk among samples, with a particular focus on critical human pathogens (Table S6). Compared with sewage/primary sludge/effluent, $PM_{2.5}$ generally exhibited a lower resistome risk score ($p < 0.05$; Figs. 4C and S14). However, there were no significant differences in score among the $PM_{2.5}$ samples collected in the different sites. This finding is consistent with the above observation that ARGs were rarely associated with MGEs in the air examined in this study. Putative ESKAPE pathogens carrying ARGs were more abundant in potential emission sources, such as sewage and sludge in WWTPs. Still, it is important to consider that any pathogenic bacteria harboring

ARGs dwelling in the atmosphere could pose exposure and infection risks, particularly in the $PM_{2.5}$ fraction, since it is amenable to inhalation. Thus, there is a need for human health risk assessments of environmental sources to consider the potential for exposure to airborne ARGs.

3.4 Estimating the contributions of WWTP to the ambient airborne bacteria and ARGs

About 345 metagenomes representing different sample types collectively formed a database of putative sources of airborne bacteria and ARGs. These samples tended to be clustered by source category, regardless of the composition of the bacterial community or ARG profile ($p < 0.01$, Fig. S16). It should be noted that activated sludge was also included as a potential source since it was found different from CEPT sludge in both ARG relative abundance and composition (Figs. S15 and S16). The prediction performance of the model using the current datasets is described in Supporting Information (Section S4).

Echoing the comparable bacterial community structure in $PM_{2.5}$ between the urban and SCISTW sites (Fig. 2), their corresponding source profiles of bacteria were relatively consistent (Fig. 5A). Notably, there appeared to be a dominant contribution of bacteria from agricultural soil and plants (26%–29%) among the identified sources. Others have also suggested that terrestrial sources play an important role in shaping the composition of the airborne bacterial community in human-impacted areas [17,59], along with WWTP matrices (11%–13%), including primary sludge, secondary sludge, sewage, and effluent in descending order (4.5%–5.4%, 3.6%–4.9%, 1.9%–2.0%, and 0.8%–1.1%, respectively), and human skin (8%–10%). Although the leading sources of bacteria in coastal areas turned out to be of human origin (55%), the absolute contributions from human airways, human skin, and human oral cavities appeared to be comparable across the three sites ($p > 0.05$) (Fig. 5B). This finding is reasonable since the coastal site was situated downwind of major urban areas (receiving

476 human-related impacts) and ~60 m above sea level (few apparent marine-indicative signals
477 detected). However, it should be noted that nearly 21% of the $PM_{2.5}$-associated bacteria in the
478 coastal site and around half in the other two sites could not be apportioned to the current listed
479 sources. This is probably because potential sources are still missing in the current model due
480 to their lack of representation in public databases.
481
482 In contrast to the source tracking results of bacteria, the ARG source profile was more
483 comparable among the three sampling sites (Fig. 5A), with less than 12% undetermined. 57%–
484 60% of the $PM_{2.5}$-associated ARGs were ascribed to WWTP (26%–27% from primary sludge,
485 15%–18% from sewage, 6%–10% from secondary sludge, and 6%–9% from effluent),
486 followed by agricultural vegetation (10%–14%) and the remaining sources with comparable
487 contributions (2%–4%). The disparities between the source profiles of $PM_{2.5}$-associated
488 bacteria and ARGs could be due to the different enrichment levels of ARGs in bacterial
489 communities from different sources. Specifically, WWTP-originated matrices in this study
490 were responsible for up to 13% of the $PM_{2.5}$-associated bacteria, but for much higher
491 proportions of ARGs at the three sites, as mentioned above. This finding corroborates the view
492 that WWTPs, as an important ARG hotspot in urban areas, could be a major anthropogenic
493 source of ARGs in the atmospheric environment. In addition, the contrasting source profiles
494 could also suggest the dissimilar adaptability to the atmospheric environment among different
495 ARB or between ARB and other general bacteria after their aerosolization from source
496 environments, e.g., many ARG-carrying bacteria have more similar patterns of survival,
497 growth, and decay than the general bacterial community. Based on the current resolved sources,
498 the $PM_{2.5}$-associated ARGs in Hong Kong are more likely attributable to anthropogenic
499 activities in agricultural and engineering systems. Moreover, additional comparisons were
500 made with $PM_{2.5}$ collected during a similar period (unpublished data) in urban Guangzhou, a

megacity located 120 km northwest of Hong Kong. Not surprisingly, Hong Kong-style WWTPs, characterized by mixed sewage (fresh and saline) and dual application of CEPT and secondary treatment, contribute much less to airborne ARGs of urban Guangzhou than its contribution in Hong Kong (Fig. S18). Despite the low proportion in source profile, CEPT-related ARG signals in Guangzhou still highlight the potential impact of Hong Kong's sewage treatment activities on the airborne resistome in nearby cities. These prediction results accord with the regional pollution pattern via air transportation and reflect the feasibility of using ARG signature in source tracking. However, the current results could be biased by the limited sample size in this study and a certain portion of non-local metagenomic data with potentially geographically dependent microbial features. In consideration of this eventuality, it is important to conduct a metagenomic-based survey in various local putative sources, including those not involved in the current study, such as hospitals, to further improve the completeness and accuracy of the results.

Finally, the emission rates of bacteria and ARGs in WWTP, expressed as the proportion of bacteria and ARGs in a certain source to be released to the atmosphere, were estimated. In general, the emission rates of both bacteria and ARGs was significantly higher (by two orders of magnitude) from primary sludge (1E-7 and 6E-8 on average for bacteria and ARG, respectively) than from sewage and effluent ($p < 0.01$) under the current configuration of SCISTW. The comparable levels of the estimated emission rates between bacteria and ARGs from each WWTP matrix would seem to indicate a similar aerosolization potential of general bacteria and ARB. However, whether the antibiotic resistance property benefits the survival of these aerosolized ARB in the harsh atmospheric environment with distinct selective pressures (e.g., UV light, desiccation) still requires investigations. Lastly, as another important source, WWTPs incorporating a range of secondary biological treatments should be incorporated into

526 future studies to fully estimate the emission of bacteria and ARGs from WWTPs to the urban
527 environment.
528
529 **3.5 Environmental Implications**
530 This study highlights the dissemination of ARGs and their bacterial hosts from typical urban
531 sources like WWTPs to the ambient atmosphere based on investigations into $PM_{2.5}$, which can
532 be inhaled into the human respiratory tract and thus has potential health implications. The high
533 contribution of anthropogenic sources (e.g., WWTPs) to airborne ARGs emphasize the need
534 to regulate and manage relevant urban settings to mitigate the dissemination of ARGs to the
535 atmospheric environment. However, the complex interactions between bacterial taxa and how
536 environmental stress alters the bacterial communities after their aerosolization from putative
537 sources does not seem to have been considered extensively in the algorithm. The current simple
538 box model used to estimate the emission rates of bacteria in WWTP-originated matrices is an
539 initial attempt to address this issue from a macro perspective. Future studies should also address
540 the bacterial aerosolization within microenvironments, such as considering the partitioning of
541 bacteria in at the air-liquid/air-soil interface. Further, a combination of the use of aerodynamic
542 transport models and efforts to monitor the dynamic process of aerosolization (coupled with
543 chamber tests) may help to produce a more accurate picture of the dissemination flux of
544 bacteria and ARGs from various sources to the ambient atmosphere. In this study, although
545 only a small portion of bacterially associated ARGs was linked with MGEs in $PM_{2.5}$, the HGT
546 potential of ARGs in the atmosphere should not be overlooked. This evaluation should also
547 consider the extracellular fraction of airborne ARGs [36] and the phage-mediated process, an
548 important HGT pathway in highly turbulent environments [60], such as the atmosphere. In
549 addition, culture-dependent techniques are required to confirm the existence and viability of
550 ARB identified in this study and their expression of antibiotic resistance, as well as to verify

the *de facto* transmission of ARGs via HGT. Such efforts will help to provide solid evidence beyond inferences derived from metagenomic data and further facilitate understanding of ARG dissemination mechanisms, including linkage to specific host species.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI:

Detailed information of the SCISTW and meteorological conditions during the sampling period, qPCR procedure, bacteria/ARG data, and prediction performance of SourceTracker in Supporting Information 1 (Section S1-S4 and Figure S1-S18); description of sampling sites, sample information, statistics of sequencing data, and source data used in source tracking provided in Supporting Information 2 (Table S1-S13).

AUTHOR INFORMATION

Corresponding author

*Phone: +852 2766 6041; fax: +852 2334 6389; e-mail: cexdli@polyu.edu.hk.

ORCID

Jiawen Xie: 0000-0001-6461-4464

Ling Jin: 0000-0003-1267-7396

Dong Wu: 0000-0003-3059-0593

Amy Pruden: 0000-0002-3191-6244

Xiangdong Li: 0000-0002-4044-2888

Author Contributions

J.W.X., L.J., and X.D.L. designed the study. J.W.X. collected samples with L.J. and performed the laboratory experiments. D.W., and A.P. provided guidance on data analysis. J.W.X., L.J., and X.D.L. analyzed the data. J.W.X., A.P, and X.D.L. wrote the paper with input from all of the authors. All of the authors read and approved the final manuscript.

Notes

The authors declare that they have no competing financial interests.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (92043302, 22193063, and 42007393), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB40000000), the National Key R&D Program of China (2017YFC0212000), and the Research Grants Council of Hong Kong (115210618, 15203920, 25210420, and T21-705/20-N). Additional support was provided by the US National Science Foundation Awards OIE 1545756, OAC 2004751, CBET 1936319, and NRT 2125798. The authors thank the Drainage Services Department of the Hong Kong Government for supporting our sampling work in SCISTW, the researchers concerned for sharing their sequencing data in the public database, and The University Research Facility in Chemical and Environmental Analysis (UCEA) and The University Research Facility in Big Data Analytics (UBDA) at The Hong Kong Polytechnic University for providing a data analysis platform. Dr. Ruoyu You from the Department of Building Energy and Environment of The Hong Kong Polytechnic University gave important suggestions on the one-box model. Prof. Tong Zhang and Dr. Xiaole Yin of The University of Hong Kong provided invaluable comments on an earlier version of the manuscript.

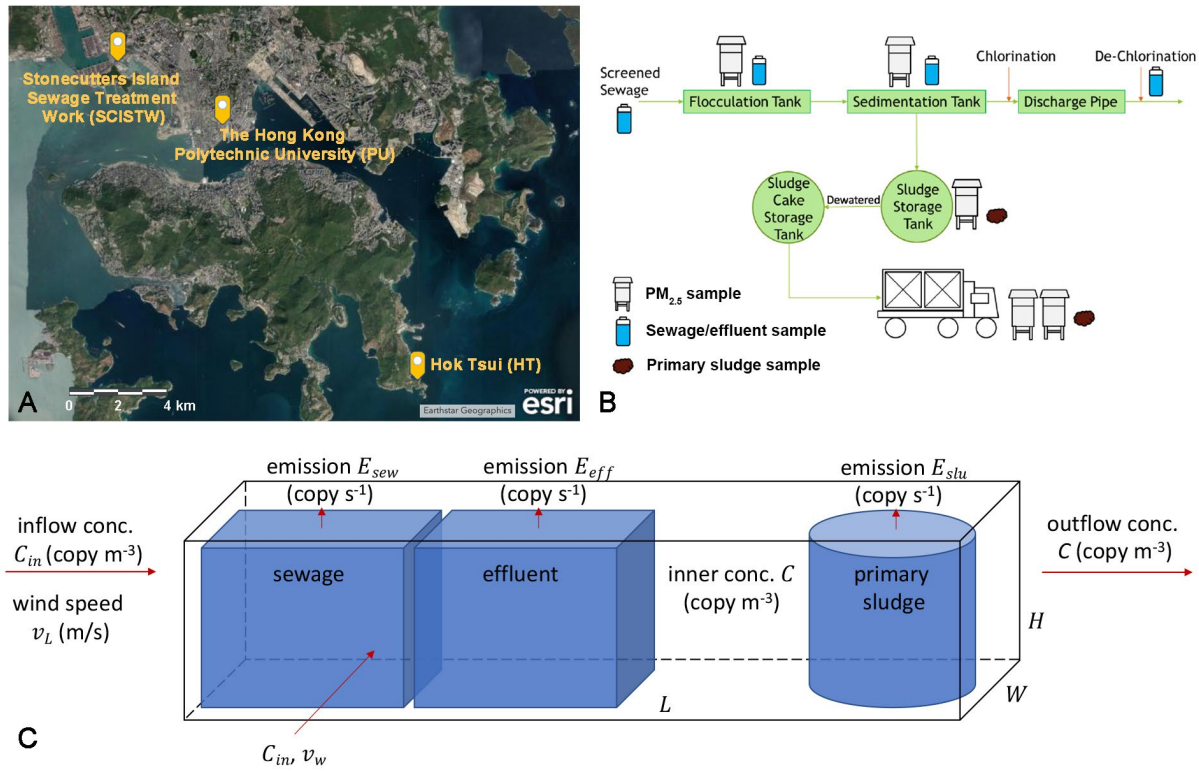
REFERENCES

1. O'Neill J., Tackling drug-resistant infections globally: final report and recommendations. Review on Antimicrobial Resistance. 19 May 2016. https://amr-review.org.
2. Tripathi V. & Cytryn E., Impact of anthropogenic activities on the dissemination of antibiotic resistance across ecological boundaries. *Essays Biochem.* 2017, 61, 11-21.
3. Schages L., Wichern F., Kalscheuer R., & Bockmühl D., Winter is coming – Impact of temperature on the variation of beta-lactamase and mcr genes in a wastewater treatment plant. *Sci. Total Environ.* 2020, 712, 136499.
4. Yang F.X., Mao D.Q., Zhou H., Wang X.L., & Luo Y., Propagation of New Delhi metallo-β-lactamase genes (bla_{NDM-1}) from a wastewater treatment plant to its receiving river. *Environ. Sci. Technol. Lett.* 2016, 3, 138-143.
5. Ji X.L., Shen Q.H., Liu F., Ma J., Xu G., Wang Y.L., & Wu M.H., Antibiotic resistance gene abundances associated with antibiotics and heavy metals in animal manures and agricultural soils adjacent to feedlots in Shanghai; China. *J. Hazard. Mater.* 2012, 235, 178-185.
6. Zhu Y.G., Zhao Y., Li B., Huang C.L., Zhang S.Y., Yu S., Chen Y.S., Zhang T., Gillings M.R., & Su J.Q., Continental-scale pollution of estuaries with antibiotic resistance genes. *Nat. Microbiol.* 2017, 2, 16270.
7. Wu D., Huang X.H., Sun J.Z., Graham D.W., & Xie B., Antibiotic resistance genes and associated microbial community conditions in aging landfill systems. *Environ. Sci. Technol.* 2017, 51, 12859-12867.
8. Chen B.W., Yuan K., Chen X., Yang Y., Zhang T., Wang Y.W., Luan T.G., Zou S.C., & Li X.D., Metagenomic analysis revealing antibiotic resistance genes (ARGs) and their genetic compartments in the tibetan environment. *Environ. Sci. Technol.* 2016, 50, 6670-6679.
9. Chen B.W., Yang Y., Liang X.M., Yu K., Zhang T., & Li X.D., Metagenomic profiles of antibiotic resistance genes (ARGs) between human impacted estuary and deep ocean sediments. *Environ. Sci. Technol.* 2013, 47, 12753-12760.
10. Ouyang W.Y., Huang F.Y., Zhao Y., Li H., & Su J.Q., Increased levels of antibiotic resistance in urban stream of Jiulongjiang River, China. *Appl. Microbiol. Biotechnol.* 2015, 99, 5697-5707.
11. Li J., Cao J.J., Zhu Y.G., Chen Q.L., Shen F.X., Wu Y., Xu S.Y., Fan H.Q., Da G., Huang R.J., Wang J., de Jesus A.L., Morawska L., Chan C.K., Peccia J., & Yao M.S., Global survey of antibiotic resistance genes in air. *Environ. Sci. Technol.* 2018, 52, 10975-10984.
12. Ling A.L., Pace N.R., Hernandez M.T., & LaPara T.M., Tetracycline resistance and class 1 integron genes associated with indoor and outdoor aerosols. *Environ. Sci. Technol.* 2013, 47, 4046-4052.
13. Pal C., Bengtsson-Palme J., Kristiansson E., & Larsson D.G.J., The structure and diversity of human, animal and environmental resistomes. *Microbiome* 2016, 4, 54.
14. Jin L., Xie J.W., He T.T., Wu D., & Li X.D., Airborne transmission as an integral environmental dimension of antimicrobial resistance through the "One Health" lens. *Crit. Rev. Environ. Sci.* 2021, 10.1080/10643389.2021.2006537.
15. Xie J.W., Jin L., Luo X.S., Zhao Z., & Li X.D., Seasonal disparities in airborne bacteria and associated antibiotic resistance genes in $PM_{2.5}$ between urban and rural sites. *Environ. Sci. Tech. Lett.* 2018, 5, 74-79.
16. Xie J.W., Jin L., He T.T., Chen B.W., Luo X.S., Feng B.H., Huang W., Li J., Fu P.Q., & Li X.D., Bacteria and antibiotic resistance genes (ARGs) in $PM_{2.5}$ from China: implications for human exposure. *Environ. Sci. Technol.* 2019, 53, 963-972.

17. Cao C., Jiang W.J., Wang B.Y., Fang J.H., Lang J.D., Tian G., Jiang J.K., & Zhu T.F., Inhalable microorganisms in Beijing's $PM_{2.5}$ and PM_{10} pollutants during a severe smog event. *Environ. Sci. Technol.* 2014, 48, 1499-1507.
18. Hu J.L., Zhao F.Z., Zhang X.X., Li K., Li C.R., Ye L., & Li M., Metagenomic profiling of ARGs in airborne particulate matters during a severe smog event. *Sci. Total Environ.* 2018, 615, 1332-1340.
19. Qin N., Liang P., Wu C.Y., Wang G.Q., Xu Q., Xiong X., Wang T.T., Zolfo M., Segata N., Qin H.L., Knight R., Gilbert J.A., & Zhu T.F., Longitudinal survey of microbiome associated with particulate matter in a megacity. *Genome Biol.* 2020, 21, 55.
20. Yang Y., Zhou R.J., Chen B.W., Zhang T., Hu L.G., & Zou S.C., Characterization of airborne antibiotic resistance genes from typical bioaerosol emission sources in the urban environment using metagenomic approach. *Chemosphere* 2018, 213, 463-471.
21. He P., Wu Y., Huang W.Z., Wu X.W., Lv J.Y., Liu P.D., Bu L., Bai Z.J., Chen S.Y., & Feng W.R., Characteristics of and variation in airborne ARGs among urban hospitals and adjacent urban and suburban communities: A metagenomic approach. *Environ. Int.* 2020, 139, 105625.
22. Li L.Y., Wang Q., Bi W.J., Hou J., Xue Y.G., Mao D.Q., Das R., Luo Y., & Li X.D., Municipal solid waste treatment system increases ambient airborne bacteria and antibiotic resistance genes. *Environ. Sci. Technol.* 2020, 54, 3900-3908.
23. Gao X.L., Shao M.F., Luo Y., Dong Y.F., Ouyang F., Dong W.Y., & Li J., Airborne bacterial contaminations in typical Chinese wet market with live poultry trade. *Sci. Total Environ.* 2016, 572, 681-687.
24. Knights D., Kuczynski J., Charlson E.S., Zaneveld J., Mozer M.C., Collman R.G., Bushman F.D., Knight R., & Kelley S.T., Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods* 2011, 8, 761-763. Package available at https://github.com/danknights/sourcetracker (retrieved on 24 March 2020).
25. Li L.G., Yin X.L., & Zhang T., Tracking antibiotic resistance gene pollution from different sources using machine-learning classification. *Microbiome* 2018, 6, 93.
26. Baral D., Dvorak B.I., Admiraal D., Jia S.G., Zhang C., & Li X., Tracking the sources of antibiotic resistance genes in an urban stream during wet weather using shotgun metagenomic analyses. *Environ. Sci. Technol.* 2018, 52, 9033-9044.
27. Han Y.P., Yang T., Chen T.Z., Li L., & Liu J.X., Characteristics of submicron aerosols produced during aeration in wastewater treatment. *Sci. Total Environ.* 2019, 696, 134019.
28. Bauer H., Fuerhacker M., Zibuschka F., Schmid H., & Puxbaum H., Bacteria and fungi in aerosols generated by two different types of wastewater treatment plants. *Water Res.* 2002, 36, 3965-3970.
29. Melbostad E., Eduard W., Skogstad A., Sandven P., Lassen J., Søstrand P., & Heldal K., Exposure to bacterial aerosols and work-related symptoms in sewage workers. *Am. J. Ind. Med.* 1994, 25, 59-63.
30. Han I. & Yoo K., Metagenomic profiles of antibiotic resistance genes in activated sludge, dewatered sludge and bioaerosols. *Water* 2020, 12, 1516.
31. Zhang M.Y., Zuo J.N., Yu X., Shi X.C., Chen L., & Li Z.X., Quantification of multi-antibiotic resistant opportunistic pathogenic bacteria in bioaerosols in and around a pharmaceutical wastewater treatment plant. *J. Environ. Sci.* 2018, 72, 53-63.
32. Zhang A.N., Gaston J.M., Dai C.L., Zhao S.J., Poyet M., Groussin M., Yin X.L., Li L.G., van Loosdrecht M., & Topp E., An omics-based framework for assessing the health risk of antimicrobial resistance genes. *Nat. Commun.* 2021, 12, 4765.
33. Martínez J.L., Coque T.M., & Baquero F., What is a resistance gene? Ranking risk in resistomes. *Nat. Rev. Microbiol.* 2015, 13, 116-123.

34. Oh M., Pruden A., Chen C.Q., Heath L.S., Xia K., & Zhang L.Q., MetaCompare: a computational pipeline for prioritizing environmental resistome risk. *FEMS Microbiol. Ecol.* 2018, 94, fiy079.
35. Jiang W.J., Liang P., Wang B.Y., Fang J.H., Lang J.D., Tian G., Jiang J.K., & Zhu T.F., Optimized DNA extraction and metagenomic sequencing of airborne microbial communities. *Nat. Protoc.* 2015, 10, 768-779.
36. He T.T., Jin L., Xie J.W., Yue S.Y., Fu P.Q., & Li X.D., Intracellular and extracellular antibiotic resistance genes in airborne $PM_{2.5}$ for respiratory exposure in urban areas. *Environ. Sci. Technol. Lett.* 2021, 8, 128-134.
37. Chen S.F., Zhou Y.Q., Chen Y.R., & Gu J., fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018, 34, i884-i890.
38. Wood D.E., Lu J., & Langmead B., Improved metagenomic analysis with Kraken 2. *Genome Biol.* 2019, 20, 257.
39. Lu J., Breitwieser F.P., Thielen P., & Salzberg S.L., Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* 2017, 3, e104.
40. Arango-Argoty G., Garner E., Pruden A., Heath L.S., Vikesland P., & Zhang L.Q., DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome* 2018, 6, 23.
41. Dixon P., VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* 2003, 14, 927-930.
42. Shen W., Le S., Li Y., & Hu F.Q., SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* 2016, 11, e0163962.
43. Conway J.R., Lex A., & Gehlenborg N., UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 2017, 33, 2938-2940.
44. Segata N., Izard J., Waldron L., Gevers D., Miropolsky L., Garrett W.S., & Huttenhower C., Metagenomic biomarker discovery and explanation. *Genome Biol.* 2011, 12, R60.
45. Arango-Argoty G., Singh G., Heath L.S., Pruden A., Xiao W.D., & Zhang L.Q., MetaStorm: a public resource for customizable metagenomics annotation. *PLoS One* 2016, 11, e0162442.
46. Peng Y., Leung H.C., Yiu S.M., & Chin F.Y., IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 2012, 28, 1420-1428.
47. Arango-Argoty G., Dai D.J., Pruden A., Vikesland P., Heath L.S., & Zhang L.Q., NanoARG: a web service for detecting and contextualizing antimicrobial resistance genes from nanopore-derived metagenomes. *Microbiome* 2019, 7, 88.
48. Buchfink B., Xie C., & Huson D.H., Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 2015, 12, 59-60.
49. Kim D., Song L., Breitwieser F.P., & Salzberg S.L., Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res.* 2016, 26, 1721-1729.
50. Li B., Ju F., Cai L., & Zhang T., Profile and fate of bacterial pathogens in sewage treatment plants revealed by high-throughput metagenomic approach. *Environmental Science & Technology* 2015, 49, 10492-10502.
51. Krawczyk P.S., Lipinski L., & Dziembowski A., PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Res.* 2018, 46, e35-e35.
52. Uritskiy G.V., DiRuggiero J., & Taylor J., MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 2018, 6, 158.

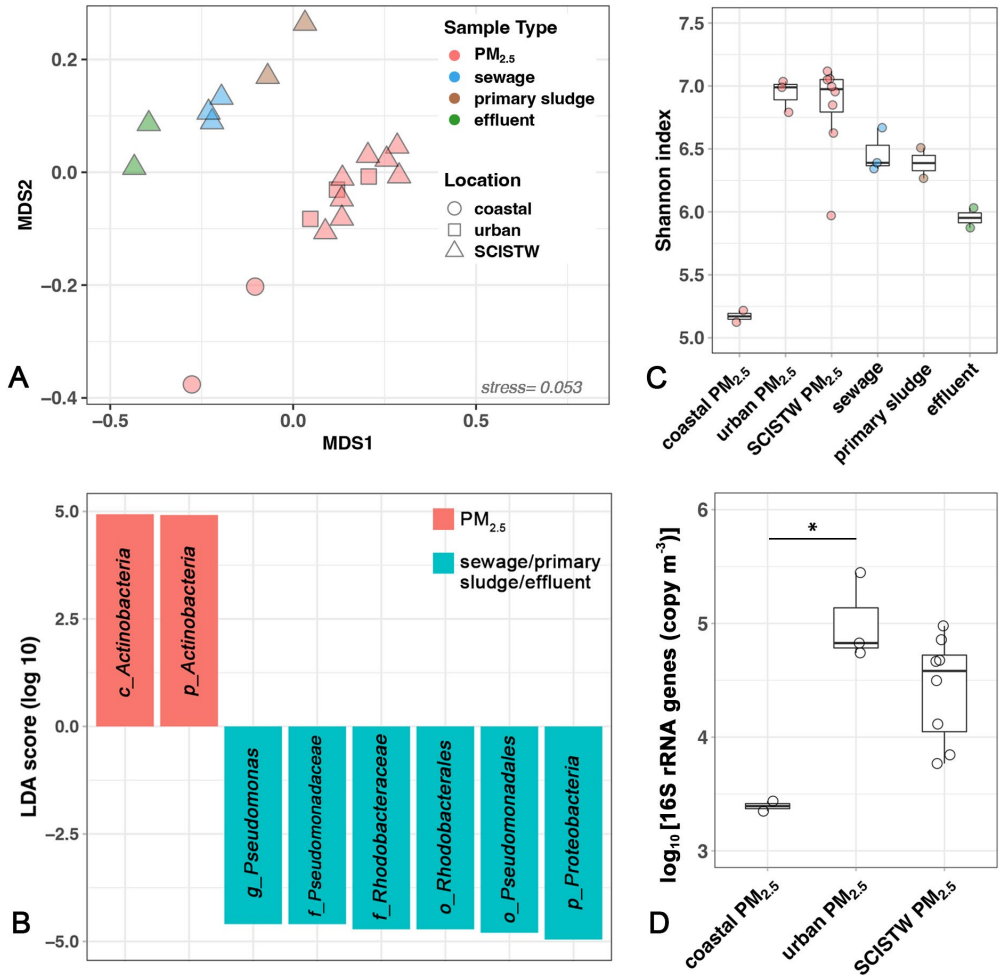
53. Anandan R., Dharumadurai D., & Manogaran G.P., An introduction to actinobacteria. In Actinobacteria-Basics and Biotechnological Applications, Dhanasekaran D. & Jiang Y., Eds. Intechopen: 11 February 2016. https://www.intechopen.com/chapters/49873.
54. World Health Organization, The 2019 WHO AWaRe classification of antibiotics for evaluation and monitoring of use (WHO/EMP/IAU/2019.11). 2019.
55. Zhang A.N., Hou C.J., Negi M., Li L.G., & Zhang T., Online searching platform for the antibiotic resistome in bacterial tree of life and global habitats. *FEMS Microbiol. Ecol.* 2020, 96, fiaa107.
56. Ma L.P., Li A.D., Yin X.L., & Zhang T., The prevalence of integrons as the carrier of antibiotic resistance genes in natural and man-made environments. *Environ. Sci. Technol.* 2017, 51, 5721-5728.
57. Centre for Health Protection, Hong Kong strategy and action plan on antimicrobial resistance 2017-2022. Department of Health, HKSAR Government. 10 July 2022. https://www.chp.gov.hk/files/pdf/amr_action_plan_eng.pdf.
58. Brown C.L., Keenum I.M., Dai D.J., Zhang L.Q., Vikesland P.J., & Pruden A., Critical evaluation of short, long, and hybrid assembly for contextual analysis of antibiotic resistance genes in complex environmental metagenomes. *Sci. Rep.* 2021, 11, 3753.
59. Bertolini V., Gandolfi I., Ambrosini R., Bestetti G., Innocente E., Rampazzo G., & Franzetti A., Temporal variability and effect of environmental variables on airborne bacterial communities in an urban area of Northern Italy. *Appl. Microbiol. Biotechnol.* 2013, 97, 6561-6570.
60. Sun R.N., Yu P.F., Zuo P., & Alvarez P.J., Bacterial concentrations and water turbulence influence the importance of conjugation versus phage-mediated antibiotic resistance gene transfer in suspended growth systems. *ACS Environ. Au* 2021, 10.1021/acsenvironau.1c00027.
61. Esri. World Imagery [basemap]. "World Imagery". 13 December 2009. https://services.arcgisonline.com/ArcGIS/rest/services/World_Imagery/MapServer (retrieved on 30 December 2021).



773

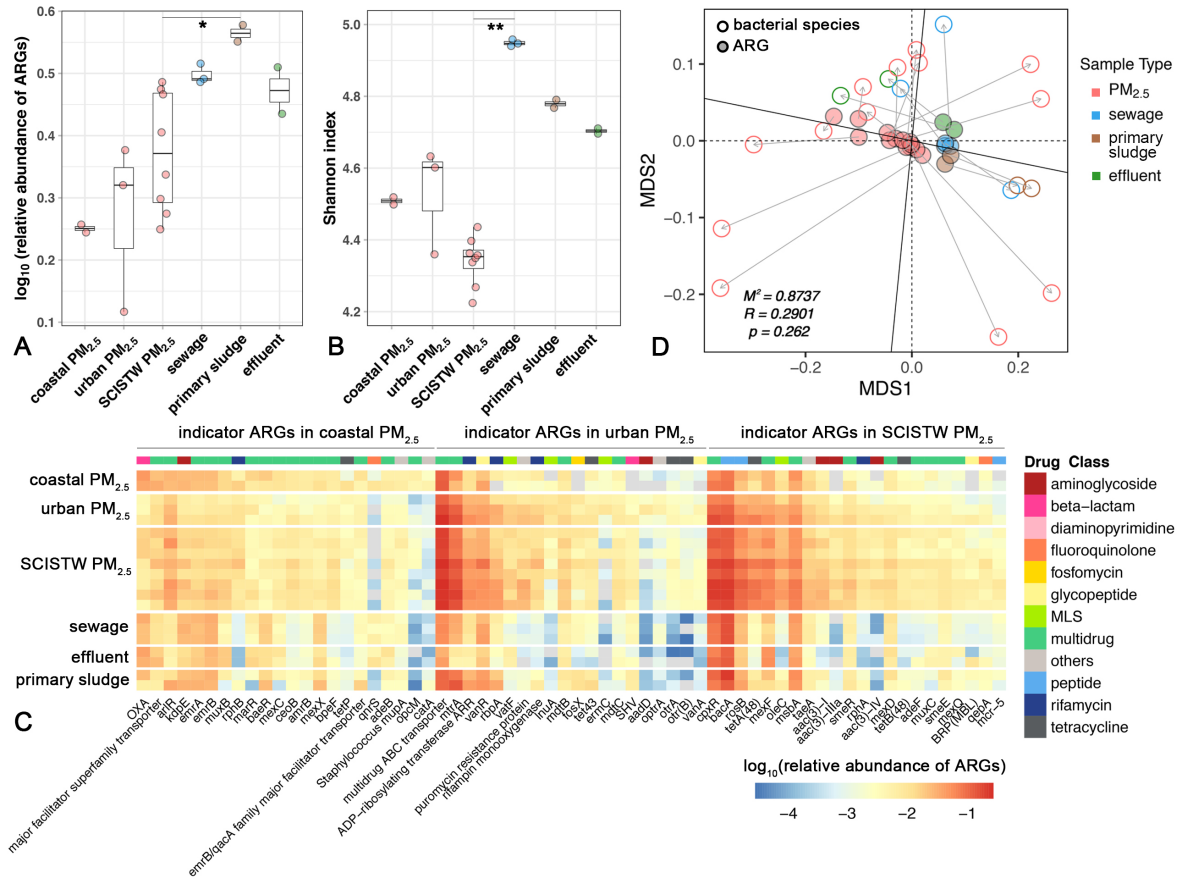
774 **Figure 1.** Map of three sampling locations in Hong Kong, created using ArcGIS online [61]
775 (A) and the general sewage treatment process in SCISTW, with the specific sampling positions
776 and corresponding sample types marked (B). Panel (C) presents the conceptual simple box
777 model used in the calculation of the emission potential of bacteria and ARGs from different
778 SCISTW matrices to the ambient atmosphere.

779



780

781 **Figure 2.** Comparison of bacterial abundance and community composition among sample
782 types and sampling locations. Panel (A) visualizes the NMDS results (Bray–Curtis
783 dissimilarity) of the bacterial community at the species level (PERMANOVA $R^2 = 0.5152$, p
784 < 0.001, permutations = 999), which corresponds to the differentiation of key taxon indicators
785 among different sample types revealed by a LEfSe analysis (LDA score > 4.5, $p < 0.05$) in
786 Panel (B). Panels (C) and (D) show the alpha diversity (bacterial species-based Shannon index)
787 among different sample types and the absolute abundance of $PM_{2.5}$-associated bacteria
788 (indicated by the 16S rRNA gene) across the coastal–urban–SCISTW transect, respectively.
789



790

791 **Figure 3.** Antibiotic resistome across sample types and sampling locations. Panels (A) and (B)
792 show the relative abundance of total identified ARGs (normalized to the 16S rRNA gene, log
793 scale) and the alpha diversity of ARGs (Shannon index based on ARG subtype), respectively.
794 Panel (C) is the relative abundance profile of indicator ARG subtypes (normalized to the 16S
795 rRNA gene, log10 transformed) in $PM_{2.5}$ across sampling locations determined by a LEfSe
796 analysis ($p < 0.05$). Panel (D) tests the correlation of community structure between bacteria
797 and ARGs via a Procrustes analysis (NMDS-based).

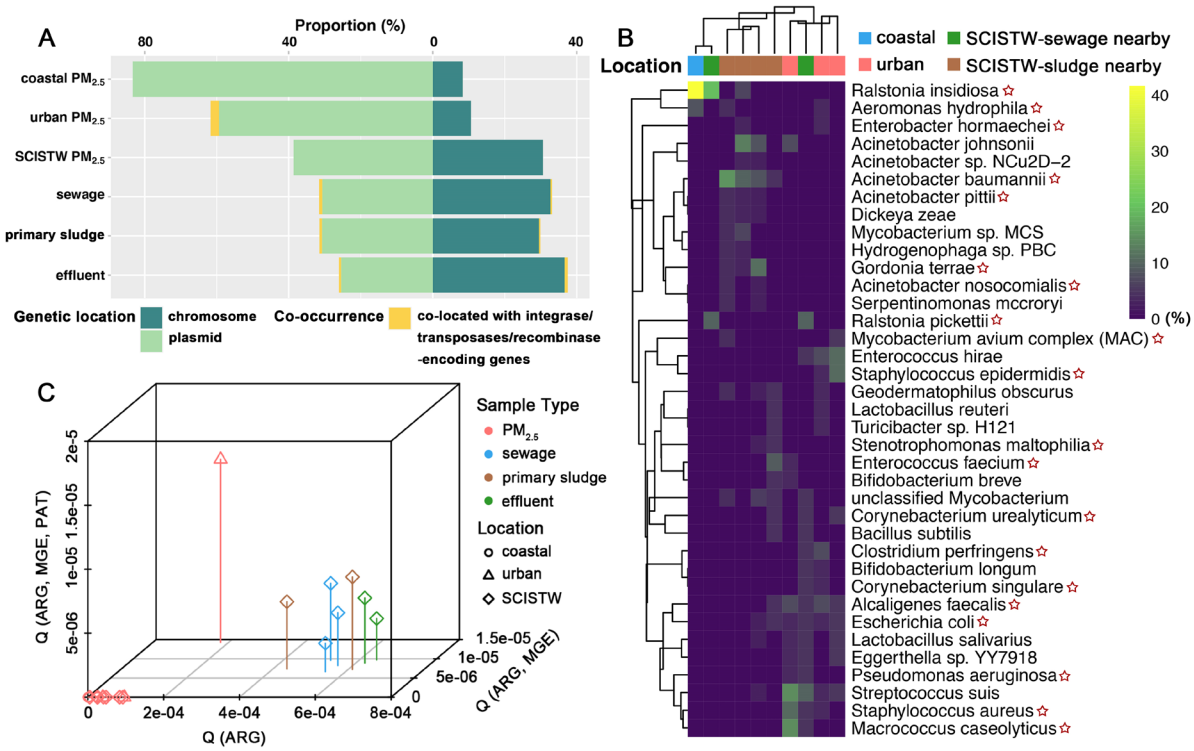


Figure 4. Co-occurrence patterns of ARGs, MGEs, and potential bacterial hosts based on assembled contigs. Panel (A) shows the genetic locations of ARGs in bacterial genomes (unclassified fraction not shown), as well as the proportions of ARGs that are co-localized with integrase/transposes/recombinase-encoding genes. Panel (B) presents the percentage of the identified ARGs in various bacterial species in each $PM_{2.5}$ sample (using package "pheatmap" in R). Here only the antibiotic-resistant species presented in no less than two $PM_{2.5}$ samples are shown. The red stars indicate the (emerging) human pathogens. Panel (C) shows the co-occurrence probability of the targeted components – ARGs, ARGs–MGEs, and ARGs–MGEs–putative pathogens (PAT), respectively – identified in the samples. This information is expressed as the contig number with critical components divided by the total assembled contig number.

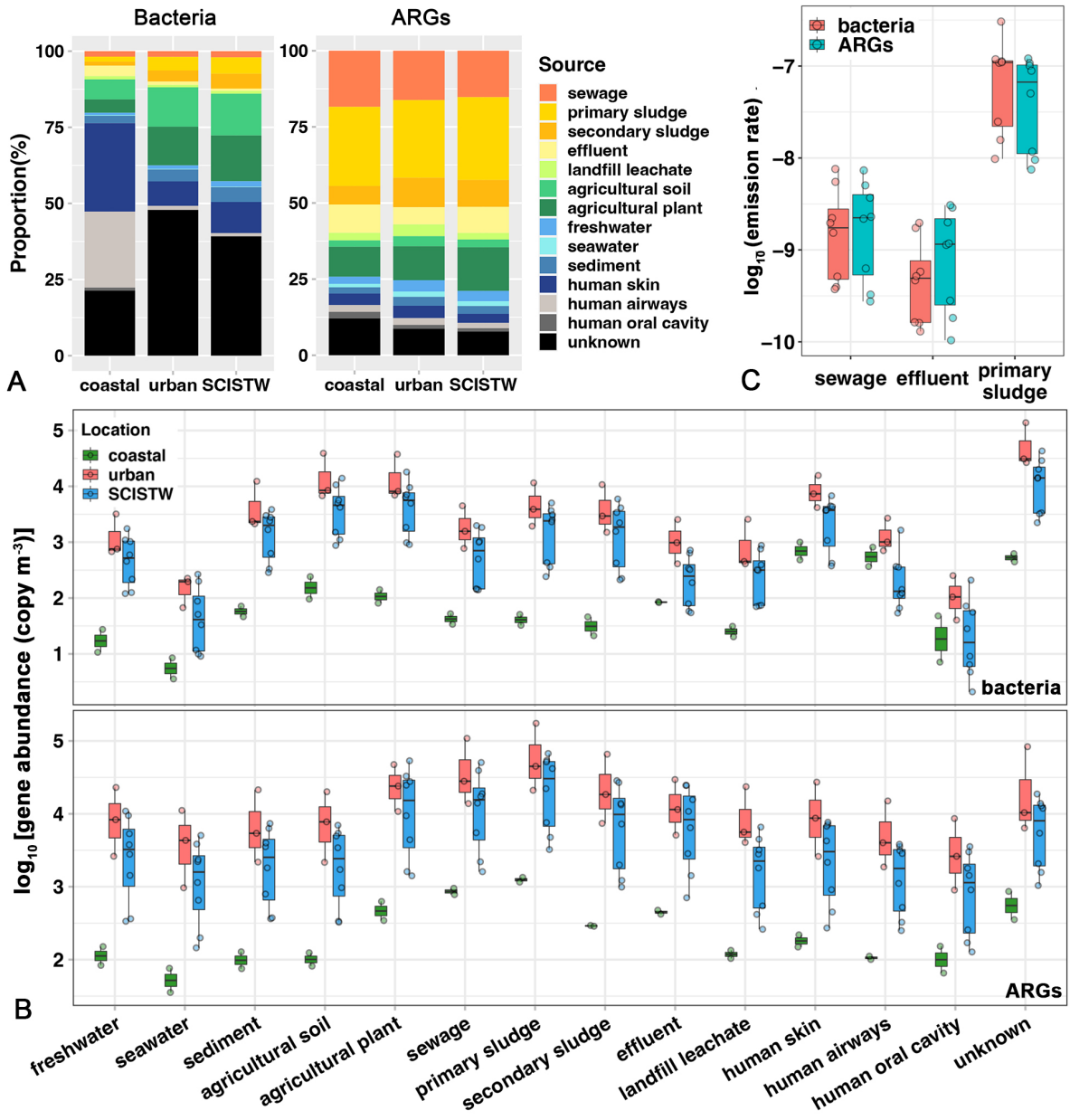


Figure 5. Contributions of various potential sources to $PM_{2.5}$-associated bacteria and ARGs in coastal, urban, and SCISTW sites. Panel (A) is the source profiles of bacteria and ARGs predicted by SourceTracker. Panel (B) shows the estimated absolute abundance of $PM_{2.5}$-associated 16S rRNA genes and ARGs contributed by different sources. The absolute abundance of ARGs was assumed to be the product of the absolute abundance of the 16S rRNA genes and the relative abundance of ARGs normalized to the 16S rRNA gene (predicted by DeepARG). The estimated emission rates of bacteria and ARGs in different SCISTW matrices are displayed in Panel (C), and are expressed as the proportion of bacteria or ARGs emitted to the ambient atmosphere.