

Geographics and bacterial networks differently shape the acquired and latent global sewage resistomes

Received: 5 February 2025

Accepted: 28 October 2025

Published online: 21 November 2025



Hannah-Marie Martiny^{1,123}✉, Patrick Munk^{1,123}✉, Alessandro Fuschi², Ágnes Becsei³, Nikiforos Pyrounakis¹, Christian Brinch¹, Global Sewage Consortium*, D. G. Joakim Larsson⁴, Marion Koopmans⁵, Daniel Remondini², István Csabai³ & Frank M. Aarestrup¹

Antimicrobial resistance genes (ARGs) have rapidly emerged and spread globally, but the pathways driving their spread remain poorly understood. We analyzed 1240 sewage samples from 351 cities across 111 countries, comparing ARGs known to be mobilized with those identified through functional metagenomics (FG). FG ARGs showed stronger associations with bacterial taxa than the acquired ARGs. Network analyses further confirmed this and showed potential for source attribution of both known and novel ARGs. The FG resistome was more evenly dispersed globally, whereas the acquired resistome followed distinct geographical patterns. City-wise distance-decay analyses revealed that the FG ARGs showed significant decay within countries but not across regions or globally. In contrast, acquired ARGs showed decay at both national and regional scales. At the variant level, both ARG groups had significant national and regional distance-decay effects, but only FG ARGs at a global scale. Additionally, we observed stronger distance effects in Sub-Saharan Africa and East Asia compared to North America. Our findings suggest that differential selection and niche competition, rather than dispersal, shape the global resistome patterns. A limited number of bacterial taxa may act as reservoirs of latent FG ARGs, highlighting the need of targeted surveillance to mitigate future resistance threats.

Acquired resistance to antimicrobials among bacteria is a significant and increasing problem. It is difficult to assess the global burden of antimicrobial resistance (AMR), but global mortality has been estimated to be 1.14 million deaths in 2021¹, with African regions being disproportionately affected and estimated to increase to 1.91 M by 2050². Another study predicts an average loss of 1.8 years of life expectancy globally by 2035, along with an additional cost of US\$ 855

billion per year in extra healthcare costs and lost workforce productivity, or -1% of the global economy³.

Antimicrobial resistance genes (ARGs) are ancient and naturally occur in bacteria across diverse environments^{4–9}; however, their rapid emergence in the clinic and global dissemination have been largely driven by human activities. Most resistome studies focus on those so-called acquired ARGs that have been mobilized from their origin and

¹Research Group for Genomic Epidemiology, Technical University of Denmark, Lyngby, Denmark. ²Department of Physics and Astronomy (DIFA), University of Bologna, Bologna, Italy. ³Department of Physics of Complex Systems, ELTE Eötvös Loránd University, Budapest, Hungary. ⁴Department of Infectious Diseases, Institute of Biomedicine, University of Gothenburg, and Centre for Antibiotic Resistance Research in Gothenburg, Gothenburg, Sweden. ⁵Erasmus Medical Centre, Rotterdam, The Netherlands. ¹²³These authors contributed equally: Hannah-Marie Martiny, Patrick Munk.*A list of authors and their affiliations appears at the end of the paper. ✉e-mail: pmun@food.dtu.dk; hanmar@food.dtu.dk

transferred between species^{10–12}, but this likely only represents a small subset of the total resistome¹⁰, where it has been suggested that only between 5 and 15% of all ARGs are being studied, as predicted *in silico*¹¹.

In silico prediction does not prove the functionality of a given gene. However, another method of discovering new ARGs is functional metagenomics (FG), which, based on random cloning and phenotypic selection, has revealed a diverse resistome across many bacterial communities^{13,14}. While many of these novel ARGs, or FG ARGs, identified either *in silico* or *in vitro*, likely are intrinsic genes of environmental bacteria and have not been acquired by other bacteria, they represent a latent reservoir of resistance that might be mobilized in the future^{11,13}. Several studies have shown that ARGs are subjected to dispersal limitations within and between reservoirs^{15–17}, but a systematic comparison of the distributions of acquired ARGs and FG ARGs has, to our knowledge, not yet been performed.

Epidemiological tracking of ARGs could help build a general understanding of how ARGs evolve, change hosts, and disseminate throughout our ecosystem, and inform interventions and policies^{18,19}. Global compatible data, including bacterial genomic data, are, however, scarce. Sewage offers a convenient and ethical way of monitoring AMR, as it integrates waste from humans, their animals, and the surrounding environment²⁰. We and others have recently utilized human sewage to monitor AMR in large, mainly healthy human populations^{15,16,20}. These sewage resistome studies have found systematic differences in the abundance and diversity of acquired ARGs between the different world regions and much variation that can neither be explained by antimicrobial use nor bacterial taxonomic composition, but more strongly correlated with socioeconomic, health, and environmental factors^{16,21,22}. Analysis of flanking sequences of a subset of acquired ARGs has suggested each ARG has unique patterns of dispersal limitation and global transmission, while other studies have identified mobilization differences as a function of resistance mechanism²³.

Dividing the world into regions typically represents a grouping of cultural, economic, historical, political, or other factors that can provide valuable insights into population differences but might be of little relevance concerning the dispersal limitations of ARGs. Thus, the question remains whether it is such regional factors or simply geographical distances that limit dispersal. Distance-decay relationships can reveal the effects of such spatial processes²⁴. Studies have shown that human activities significantly influenced the presence of ARGs on coasts, but their similarity decreased as they dispersed into estuaries and the open sea, indicating strong distance-decay gradients along aquatic environments^{24,25}. Most studies of distance-decay stick to a single region, where it can be difficult to investigate whether events on one side of the globe affect the other. However, a single study looked into the resistomes of globally distributed lake sediments and found that ARG compositions exhibited distance-decay relationships but were largely shaped by bacterial community structure²⁵.

As part of our recent efforts to combine multiple existing collections of ARG references into a single database called PanRes²⁶, we also included two collections of ARGs identified through functional cloning, namely ResFinderFG 2.0¹⁴ and the collection from Daruka et al.¹³. We hypothesize that these ARGs identified through FG would be more strongly associated with environmental bacterial taxa and possibly show more substantial dispersal limitations than those ARGs that have mobilized, in relation to our human-associated bacteria, and could travel globally with us.

In this study, we aimed to systematically compare the abundance, diversity, and bacterial associations of acquired ARGs with those ARGs identified with FG. We expanded our collection of global sewage datasets to cover samples between 2016 and 2021. Using this dataset, we characterized the resistomes of 3131 acquired ARGs from ResFinder¹⁹ and 4990 FG ARGs from ResFinderFG 2.0¹⁴ and Daruka et al.¹³. While the results confirmed the regional clustering of acquired

ARGs as shown in our previous studies^{15,16}, we found, contrary to our hypothesis, that the FG ARGs were much more evenly distributed across the globe. Only in Sub-Saharan Africa were the relative abundances of acquired and FG ARGs equal. Our distance-decay analyses revealed that while the FG ARGs exhibited a continuous decay, distance did not appear to affect dispersal at the inter-regional scale. The network analyses revealed that the FG ARGs were strongly linked with the underlying sewage bacteriomes, suggesting that the majority of the FG ARGs represent a latent reservoir of resistance. Together, these findings highlight the importance of monitoring both the acquired and FG ARGs to address the current and future threats of AMR.

Results

Summary of sewage sample metagenomes

The urban sewage collection included 1240 samples from 351 cities across 111 countries, spanning all seven world regions from 2016 to 2021 (Supplementary Fig. 1). On average, each sample had 32.39 million (M) trimmed sequence fragments (range: 0.03–515.56 (M), std: 21.82 M), totaling more than 8.93×10^{10} read fragments. All trimmed fragments were mapped and aligned against the mOTUs conserved marker genes and the PanRes ARGs. 0.16% of these read fragments were assigned to mOTUs and 0.04% to PanRes ARGs. In our chosen subset of the PanRes collections, 0.019% trimmed read fragments were aligned to acquired ARGs, and 0.024% to FG ARGs. See the Supplementary Notes for the distribution of all PanRes ARGs.

A total of 141.07 M fragments were assigned to mOTUs (sample average: 0.10 M, std: 0.05 M), of these 99.4% were bacteria, 0.002% eukaryotes, and 0.56% archaea. For the bacterial hits, the fragments matched 13324 genera; however, 10,974 of them were placeholder names, and 2350 were latinized bacterial genera. The most common genera were *Pseudomonas* (2.91% of bacterial reads), *Acinetobacter* (2.44%), *Acidovorax* (1.77%), *Neisseria* (1.36%), and *Streptococcus* (0.87%).

We found evidence of 1052 acquired ARGs and 3095 FG ARGs, totaling 17.28 M fragments and 21.75 M fragments, respectively. On average, a sample had 0.015 M fragments to acquired ARGs (range: 0.000214–0.113 M, std: 0.013 M) and 0.019 M to FG ARGs (range: 0.000612–0.239 M, std: 0.016 M) (Supplementary Fig. 4a). 201 acquired ARGs and 527 FG ARGs were present in at least 50% of the samples with resistance fragments, with 1 acquired ARG (*aph(6)-id_2*) and 6 FG ARGs being present in all.

Functionally identified ARGs are more abundant and geographically widespread

In agreement with our previous findings on a subset of the samples, the acquired ARGs were most abundant in Sub-Saharan Africa (SSA), the Middle East & North Africa (MENA), and South Asia (SA) (Fig. 1a). FG ARGs showed a higher and more evenly distributed abundance across the regions and countries than the acquired ARGs, with particularly high abundances in SSA and MENA (Fig. 1a, Supplementary Fig. 5, Supplementary Fig. 6). There were also a few cases where the acquired ARG load was high enough to approach FG loads, as seen in Cambodia and Iran (Supplementary Fig. 5). Alpha diversity analyses showed a closer relationship between the bacteriome and the FG resistomes compared to the acquired resistomes (Supplementary Fig. 4).

We estimated the regional pan- and core-resistomes. Overall, the sizes of the FG pan- and core-resistomes were larger than the acquired resistome (Supplementary Fig. 6), likely due to a combination of more FG references than acquired and an overall higher abundance (Supplementary Fig. 3, Supplementary Fig. 4). The core resistome constituted 12% and 23% of the pan-resistomes for FG and acquired genes, respectively.

Overall, the abundance and beta diversity of the acquired ARGs reflected the world regions, with 12% of the beta diversity explained by regional grouping (permanova, $p = 0.001$). In contrast, world regions

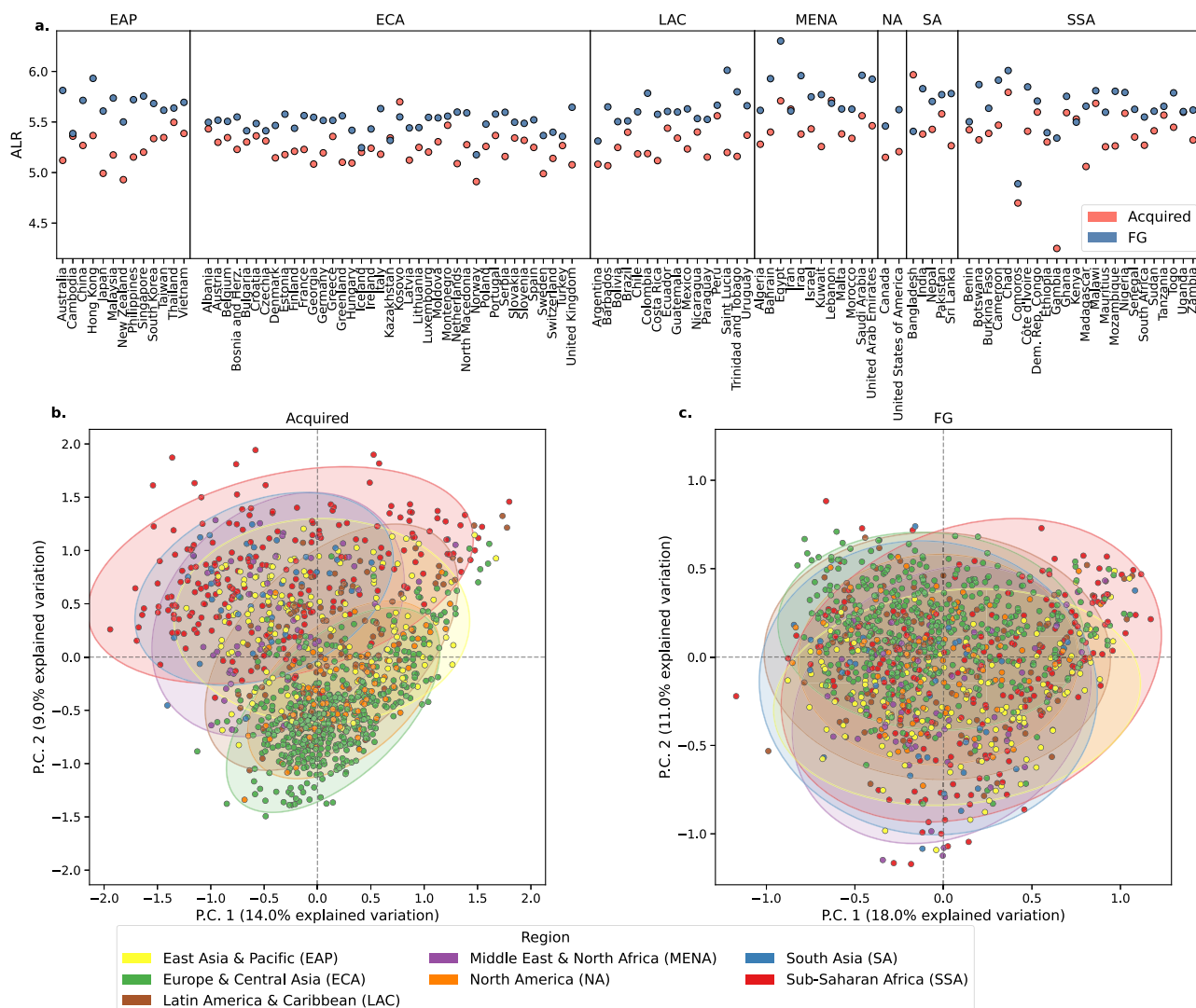


Fig. 1 | Abundance and beta-diversity of ARGs across geographical regions. **a** The country-wise ALR abundances of acquired ARGs and the FG ARGs. **b, c** PCA biplots of resistance genes (98% homology grouping), in which the PCA loadings

were calculated from CLR values. Each marker represents a sewage sample and is colored by the world region for **b** acquired and **c** FG ARGs.

explained only 7.4% of the FG ARG resistome (permanova, $p = 0.001$, Fig. 1b). World regions could better explain the beta diversity of the combined FG collection than the two resources individually, 7.2% and 6.8% for ResFinderFG and Daruka, respectively (permanova, $p = 0.001$, Supplementary Fig. 7). Procrustes analyses indicated that the acquired resistome was slightly more closely related to the bacteriome (0.88, $p = 0.001$) than the FG resistome (0.69, $p = 0.001$).

Analyzing the resistomes at drug class levels showed that the abundances of the acquired ARGs were more consistent with world regions than for the FG ARGs, although there was a lot of unexplained variation (permanova, $p = 0.001$, Supplementary Table 1). An exception was glycopeptide resistance, where regions explained just 2.6% of the variance for acquired ARGs but 7.8% for the FG resistome (permanova, $p = 0.001$). Regions only explained a minor part of the beta-diversity of fluoroquinolone and polymyxin resistances for either ARG collections (Table 1, Supplementary Data 6).

Urban sewage total resistome and ARG variants are subject to distance-decay effects

Exploiting that our sampling sites were taken at distances ranging from less than 1 km to >15,000 km, we wanted to determine the link

between sample-wise distance and metagenomic similarity using linear distance-decay (DD) models (Fig. 2, Supplementary Table 2). We stratified sampling pairs based on whether they included different countries or even different regions. We observed significant distance-decay effects between the pairwise sample distance and resistome and bacteriome similarities, and that the rate of DD varied over local (within country), regional, and global scales (between regions).

However, contrary to our original hypothesis that the FG ARGs would have more dispersal limitations than the acquired ARGs, the DD models suggested that there is a significant difference in the DD slopes for acquired ARGs once outside a country (slope_{within region} = -0.044, Supplementary Table 3). The DD slopes for the FG ARGs did not show a significant difference at the three spatial scales (Supplementary Table 3).

There was no detected effect of physical distance on taxonomic compositions at national or regional scales, but at the interregional scale, we observed a slight increase in similarity (reverse decay; slope = 0.0015; $P = 0.0$). Mantel tests showed that the FG ARGs were slightly more correlated to bacterial dissimilarity ($\rho = 0.76$; $P = 0.001$) than the acquired ARGs ($\rho = 0.7$; $P = 0.001$).

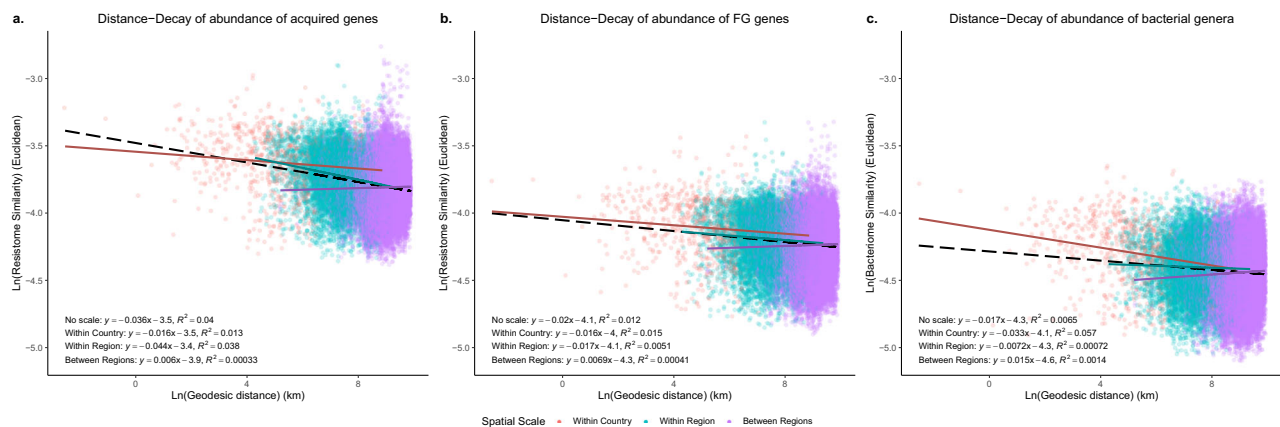


Fig. 2 | Distance-Decay analyses of read abundances for city resistomes and bacteriomes. Distance-Decay relationships for the bacteriome and resistome city communities ($n = 60,030$ pairwise comparisons) for the abundance of **a** Acquired ARGs, **b** FG ARGs, and **c** bacterial genera. The x-axis shows the pairwise city distances in kilometres (km), and the y-axis shows similarities. The dashed line

represents the fit across all spatial scales and individual fits in solid: cities within the same country, cities within the same region, and cities that are in different regions. Model parameters and adjusted R^2 values are listed for each model in its corresponding plot.

We used metaSPAdes to recover contigs from the urban sewage samples to investigate the distance-decay patterns at the gene-variant level. We identified 1633 different ARGs in 83,869 contigs with the Flankophile pipeline. Among these, 28,482 contigs contained 452 different acquired ARGs, and 56,091 contigs contained 1181 different FG ARGs.

We found 27,217 variants of these known reference sequences, including 6098 variants of acquired genes (median: 3 variants/ARG, range: 1–269) and 21,119 FG ARGs (median: 4 variants/ARG, range: 1–325). Only 1060 of the ResFinder ARG variants and 3283 of the functionally identified ARG variants were observed more than once. Most of these were only found in a single or a few cities and countries, despite observing some variants in more than 300 cities (Supplementary Fig. 8).

For presence within and between regions, we observed an initial drop in prevalence in the countries, followed by an uptick in regional prevalence of different ARG variants (Supplementary Fig. S8a, b). Clustering cities based on acquired ARG variants resulted in strong regional grouping (Fig. 3a). However, clustering on FG variants did not (Fig. 3b). For example, cities in SSA and SA were more separated in their acquired variants from the rest of the world, except Pretoria, which clustered more with European and Central Asian cities.

Geographic distances between cities were strongly negatively correlated with ARG variant sharing (Fig. 3c, d; Supplementary Table 2). However, this effect completely disappeared at the inter-regional scale for acquired ARGs (slope = -0.01 ; $P = 0.11$). The FG genes behaved differently with distance, continuing to decay gene variant similarity sharing across regions (slope = -0.084 ; $P = 1.7 \times 10^{-102}$). Mantel tests over all the spatial scales suggested that physical distance was more correlated to acquired variant sharing ($\rho = 0.6$; $P = 0.001$) than to FG variant sharing ($\rho = 0.32$; $P = 0.001$, Supplementary Table 2).

The individual distance-decay relationships for each world region revealed that these effects mirror the global trends (Supplementary Fig. 9) but with greater variation (Supplementary Table 4), likely due to fewer data points available for specific regions (Supplementary Fig. 1). Interestingly, the acquired resistomes for East Asia & Pacific and SSA had some of the steepest declines in similarities between countries with slopes of -0.1 ($P < 0.001$, Supplementary Table 4), suggesting less connectivity between cities in those countries. Oppositely, there were only weak slopes for North American resistomes, indicating that despite large distances, the similarity of cities in North America does not decrease significantly.

Network community detection reveals clustering of fecal bacteria and community clustering of acquired and FG resistomes

To elucidate the global dynamics of the sewage resistome, we constructed a correlation network integrating abundance data on bacterial species and ARGs (Supplementary Data Files 3–5). The network was reconstructed on a subset of elements consistently detected worldwide to effectively manage the typical sparsity of metagenomic data (details in Methods). This subset included 1865 mOTUs, 252 FG ARGs, and 71 acquired ARGs, with the latter tending to show a less uniform global spread.

The resulting network exhibited a sparse correlation structure with an edge density of -1.1% , with -25% of the edges connecting bacterial species and ARGs, and 857 ($\sim 40\%$) of the vertices were isolated, showing no relevant relationships. Furthermore, the network demonstrated a pronounced community structure with a modularity index of 0.7, indicating tightly interconnected distinct clusters of bacteria and ARGs likely originating from diverse ecological, environmental, and human-associated sources (Fig. 4). We observed that the six main communities detected in the network (Fig. 4b) contained 226 FG and 55 acquired ARGs and varied greatly in their number of elements, composition, and connectivity. Detailed plots of each community are in Supplementary Fig. 10.

The prevalence of human gut-associated species was noticeably high in communities 1 and 6, with 70% (115 species) in community 1 and 79% (49 species) in community 6. There were other notable differences in the composition of these two communities. In community 1, 78% of the human-associated microbiome species belonged to the phylum *Bacillota* (formerly known as *Firmicutes*), and other phyla included *Actinomycetota* (11%), *Pseudomonadota* (6%), *Euryarchaeota* (1%), and *Mycoplasmadota* (1%) (Supplementary Fig. 10a, Supplementary Fig. 12). In contrast, in community 6, 67% of the human-associated bacteria were members of the phylum *Bacteroidota* (formerly known as *Bacteroidetes*), with other phyla being *Pseudomonadota* (18%), *Bacillota* (8%), *Fusobacteria* (2%), and *Spirochaetota* (2%) (Supplementary Fig. 10f, Supplementary Fig. 12). The number of ARGs present in the two communities also differed greatly, as two acquired and 46 FG ARGs were part of community 1, and only two acquired and eight FG ARGs in community 6. Interestingly, 16 of the 46 FG genes in community 1 conferred resistance to glycopeptides (33%).

Community 2 showed low connectivity (around 2% internal edge density), especially compared to other communities (Supplementary Fig. 10b). It comprises 278 mOTUs, including 26 functional and 14 acquired ARGs. About 65% of its members are unassigned bacterial

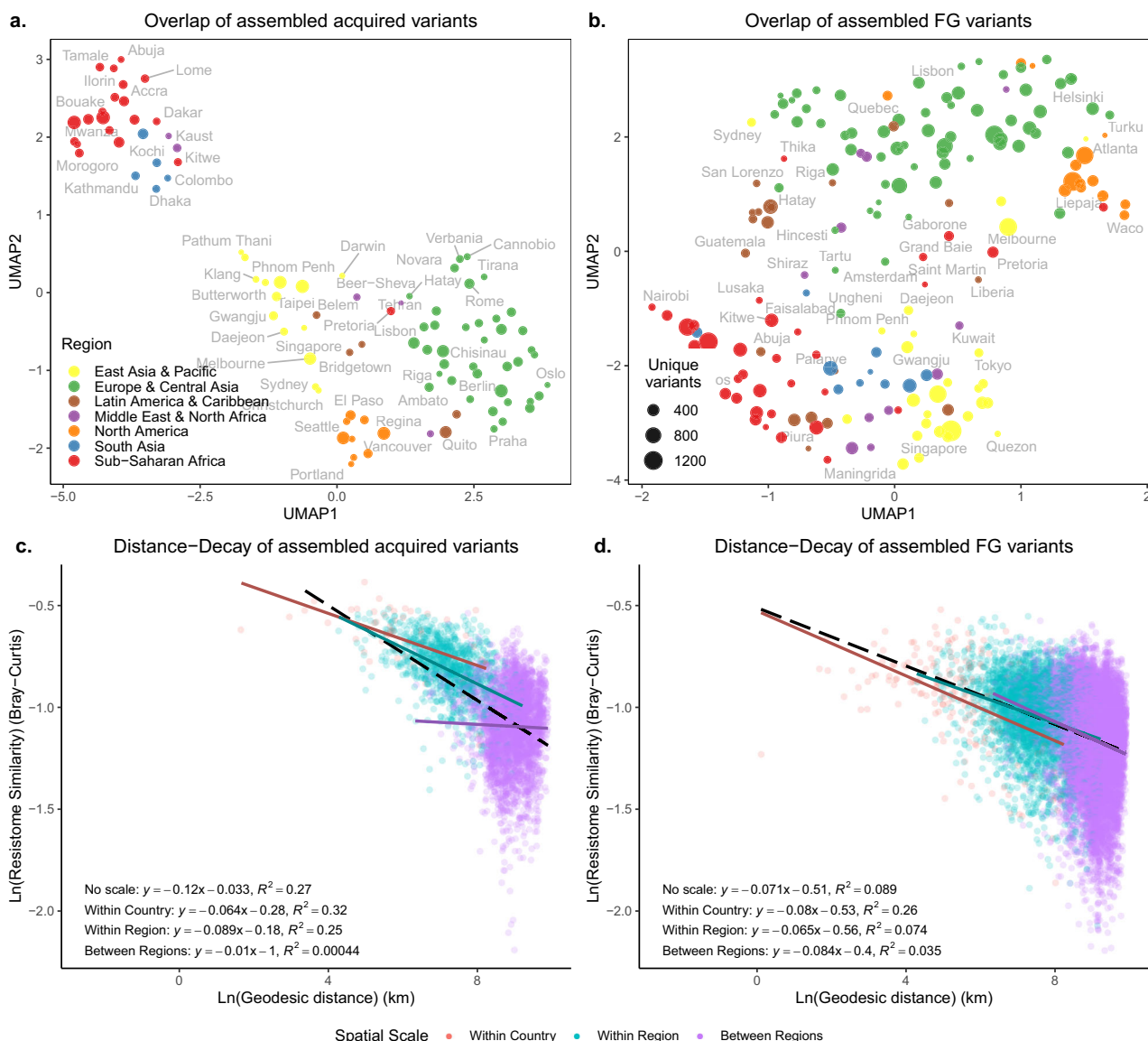


Fig. 3 | UMAP and distance-decay analysis of resistomes. UMAP clustering of shared variants among the cities for **a** acquired ARGs and **b** FG ARGs. Only cities with more than 100 non-singleton alleles were retained, and Hellinger transformed and clustered with the UMAP algorithm. Each marker represents a city, colored by region and sized by the number of unique variants in that city. City labels were optimized using the ggrepel package to minimize overlap. Distance-decay relationships for assembled city resistomes across different spatial scales for **c** acquired

($n = 4656$ pairwise comparisons) and **d** the FG variants ($n = 16,471$ pairwise comparisons). The x-axis shows the pairwise city distances in kilometers (km), and the y-axis shows resistome similarities. The dashed line represents the fit across all spatial scales and individual fits in solid: cities within the same country, cities within the same region, and cities that are in different regions (between regions). Model parameters and adjusted R^2 values are listed for each model in its corresponding plot.

species that serve as both abundant and central nodes. Among these, *Bacteria ext_mOTU_v3_18373* and *Bacteria ext_mOTU_v3_18384* exhibit particularly high connectivity and abundance, though their phyla remain unclassified. Beyond these two species, the other central taxa are also unidentified but span various families, including *Acidaminococcaceae*, *Atopobiaceae*, *Comamonadaceae*, *Eubacteriaceae*, *Moraxellaceae*, *Selenomonadaceae*, *Sporomusaceae*, and *Synergistaceae*. The only central, fully classified species are *Brachymonas denitrificans* and *Acinetobacter touneri*.

Community 3 contained the largest number of different acquired ARGs including macrolide resistance genes (*mef(c)*, *mef(b)*, *mph(e)*, *mph(g)*, and *msr(d)*); lincosamide resistance genes (*lnu(b)* and *lnu(d)*); and aminoglycoside resistance genes (*aadA* variants and *ant(6)-Ia*). This community also encompassed a diverse range of bacterial taxa, including *Acinetobacter*, *Chryseobacterium*, and *Flavobacterium*, as

well as some human-associated species such as *Acinetobacter johnsonii*, *Akkermansia muciniphila*, *Dialister invisus*, and *Lactococcus lactis*. Notably, it contained a large proportion of unassigned taxa (Supplementary Fig. 10c).

Community 4 was distinguished by its uniform composition and a very high prevalence of FG ARGs (Supplementary Fig. 10d). This community exclusively consisted of bacteria from the *Enterobacteriaceae* family (19 species) and comprised 126 nodes, ~85% of which were ARGs. Central to this community were *Escherichia coli* and *Klebsiella pneumoniae*, which, along with 17 other closely related species, formed a tightly knit group. These bacteria were associated with a diverse array of ARGs, specifically 104 FG ARGs.

Community 5 exhibited a coherent structure consisting of 113 elements, where 90 were bacteria and 23 ARGs (Supplementary Fig. 10e). This community predominantly included 65 bacteria from

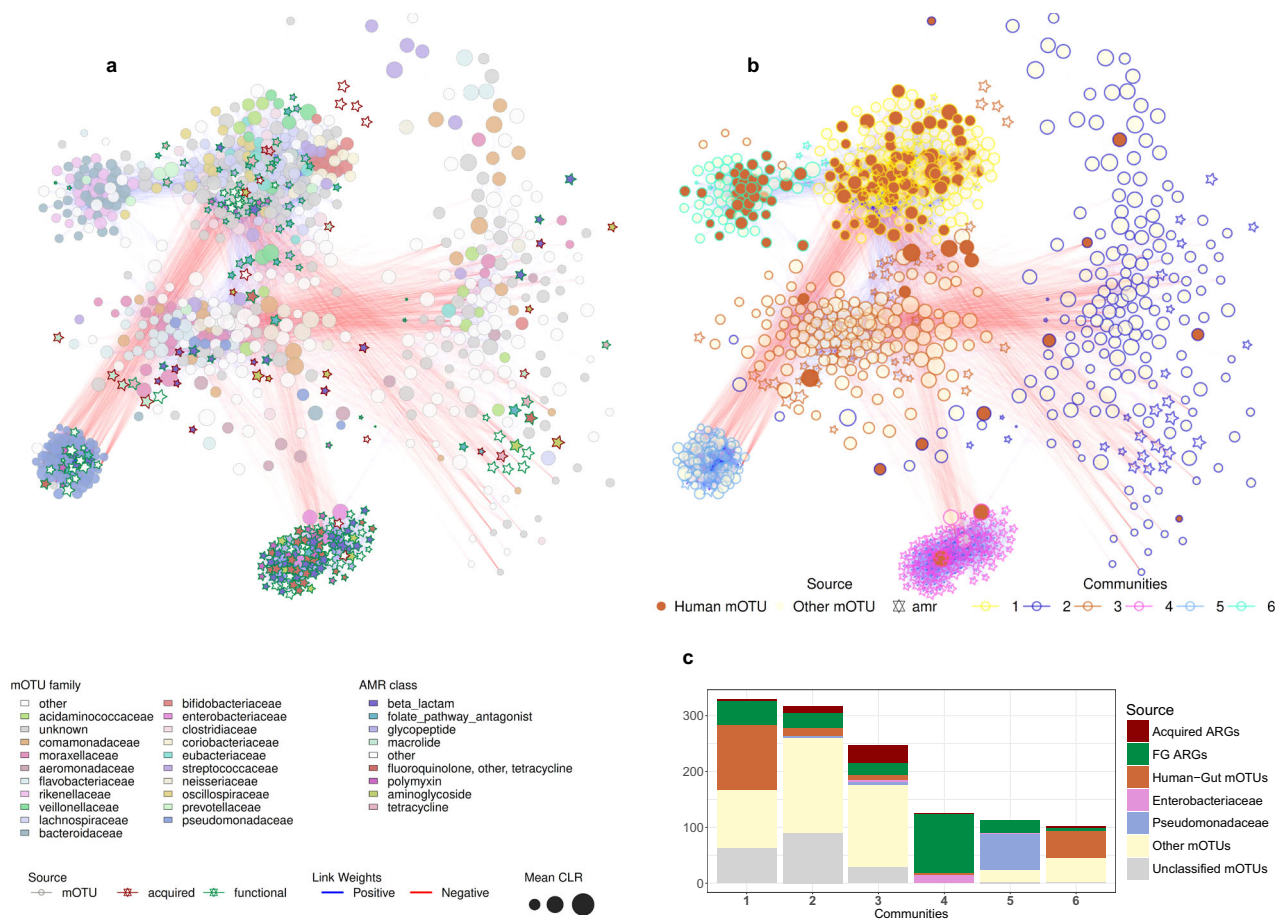


Fig. 4 | The core of the correlation network of the abundances of bacteria and ARGs in the wastewater samples. Each node represents either a species or an ARG, where an edge is the positive (blue) or negative (red) correlation between two nodes. The nodes are shaped by the database source (mOTU, acquired, or FG) and sized by the mean CLR value across samples. **a** Network graph showing the

distribution of species and ARG, where a node is colored by either mOTU family or antimicrobial resistance class. **b** Network graph showing the communities detected. Nodes are here colored by whether they are an ARG or if the bacterial species is related to the human gut. The edge colors of the nodes highlight which community they were part of. **c** The count of the type of nodes in each community.

the *Pseudomonadaceae* family, seven from the *Moraxellaceae* family, and five from the *Yersiniaceae* family. The extraordinary internal connectivity of *Pseudomonadaceae*, with an edge density reaching a surprising 85%, defines the robust structure of this community. Additionally, the ARGs in this community were especially well-connected to the *Pseudomonadaceae* core, far more than to other families. These mainly functional ARGs conferred resistance to beta-lactams, fluoroquinolones, polymyxins, and tetracyclines.

Discussion

Continuous monitoring of the distribution and dissemination of AMR across different environments remains a critical task^{1,27}, and we have previously shown how sewage sampling can effectively support such surveillance activities^{15,16,28}. While multiple studies have shown how populations of bacteria in various environments can be carriers of ARGs^{18,27,29}, existing reference gene databases primarily contain ARGs encountered in clinical settings^{11,14}.

In this study, we sought to investigate the differences in diversity and abundance of the acquired ARGs from the ResFinder¹² database with two collections of ARGs identified with FG^{13,14}. The culture-independent approach of FG offers an opportunity to identify ARGs in both environmental and human-associated microbial populations that have not been characterized before^{30,31}. By analyzing our global collection of 1240 untreated wastewater samples from 111 countries

sampled between 2016 and 2021, we found that functionally identified ARGs (FG ARGs) were more globally dispersed in the sewage resistome than acquired ARGs. This suggests that novel resistance mechanisms might emerge in any geographical location and that possible global hotspots are not dependent on the genomic availability but rather on anthropogenic factors such as selection and transmission.

We confirmed our earlier wastewater studies suggested a regional separation of the acquired ARG abundances^{15,16}, but we also found that this regional signal appeared weaker in this study. Regions could only explain 12% of the beta diversity for the acquired ARGs and even less, at 7.4% for the FG ARGs. We decided to investigate the distance-decay (DD) patterns to elucidate whether the dissemination of ARGs faces significant barriers. Numerous studies have shown that the decay of similarity with geographical distances is a universal pattern observed across all domains of life^{32–37}. Our models, based on both city-wise read abundances and the proportion of shared ARG variants, showed significant DD effects for both the acquired and FG ARGs. However, the slopes of these relationships varied across spatial scales. The acquired ARGs showed a strong negative correlation within a region, but the effect disappeared at inter-regional scales. This pattern could perhaps be due to regional-specific selective pressures related to population differences or competitive exclusion; if an ARG encoding a similar phenotype has already established itself at a specific site, another variant will be excluded. In contrast, the FG ARGs exhibited a more

gradual and continuous global decay, suggesting that while they are more globally dispersed, their dissemination might be shaped by other processes than physical distances, such as environmental conditions or ecological differences^{32,33,38–40}. It should be noted that the observed DD effects might be influenced by the unequal representation of samples from the different regions, with most samples originating from Europe. The spatial patterns of ARGs may also reflect the underlying microbial community structure in the sewage systems, as we observed that FG ARGs were more linked to the bacteriome than the acquired ARGs. This suggests that the majority of FG ARGs were likely still embedded in their original genomic contexts and have not yet been mobilized, representing a latent reservoir of resistance.

To further investigate the bacteriome-resistome relationship, we used network analyses based on read abundances to identify six main communities of bacterial taxa correlated with ARGs. Two of these communities contained primarily taxa associated with human gut samples, confirming our recent study that it might be feasible to source attribute taxa and ARGs¹⁷. Interestingly, one was dominated by *Bacillota* (community 1) and the other one by *Bacteroidota* (community 6), two phyla that together account for over 90% of the abundance of the detected species in human gut microbiota^{41,42}, and their ratio can be a potential indicator of various physiological and pathological states^{43–45}. Community 4 had the largest amount of FG ARGs, and central to this community were *Enterobacteriaceae* nodes, specifically *E. coli* and *K. pneumoniae*. These two in community 4 may partly reflect methodological biases, as both *E. coli* and *K. pneumoniae* have been used as bacterial hosts in functional metagenomic studies^{13,46}. Community 3 had the highest count of acquired ARGs and contained both environmental and human-associated taxa, which could indicate multiple context locations. The presence of ARGs and human-gut-associated taxa in the detected network communities could indicate a potential for faster dissemination and align with the acquired DD slopes, as it has recently been shown that human-associated bacteria displayed faster dispersal rates³⁵.

Whether including FG ARGs in resistome studies and reference databases is clinically relevant remains an open question. On the one hand, FG offers a powerful way to explore the distribution of ARGs in non-pathogenic or non-culturable bacteria that have yet to be mobilized or selected for in clinical settings^{14,47}. On the other hand, their inclusion may introduce too much noise into resistome analyses, as the FG ARGs could be part of the intrinsic resistome and just be a marker for the bacteria they belong to, thus posing low or no threat of becoming an issue to global health⁴⁸. Careful interpretation is therefore required to avoid overestimating the risk associated with FG ARGs, particularly since some of the genes may only be functionally active under experimental settings and may serve other roles or remain inactive in natural microbial communities. While sewage sampling offers a snapshot of the resistome of a human population, it is worth noting that sewage also contains environmental and industrial inputs and is affected by the local climate and seasonality^{15,17,47}. These non-human contributions to sewage, together with unequal sampling coverage of world regions, have likely introduced noise into the results. This noise is particularly relevant to keep in mind for our study of the FG ARGs, as we have seen evidence that some of the FG ARGs appear to be in environmental taxa. Further validation of the observed resistome patterns is necessary to elucidate the ecological and evolutionary contexts of ARG dissemination and their clinical relevance across environments beyond sewage.

Our findings demonstrate that the FG ARGs represent a latent reservoir of resistance located globally. The functional resistome was more associated with the bacteriome, suggesting a strong evolutionary barrier to their mobilization. Our identified distance-decay and network communities suggested that differential selection and niche competition, rather than dispersal, shape the global sewage resistomes. The acquired ARGs likely reflect most of the current burden of

resistance. In contrast, the FG ARGs may become a future problem; thus, including ARGs identified with FG in routine surveillance programs could serve as an early warning system for their mobilization.

Methods

Sampling and sequencing

We started by retrieving the collection of metagenomic reads and assemblies in our previous study¹⁵. We repeated our global calls, recruited partners, and received additional untreated sewage samples from 2018 to 2021. The sampling, DNA extraction, library preparation, and sequencing methods were identical to those used in our previous work^{15,16}.

In brief, bottles of untreated sewage were frozen by the partners and shipped to Denmark. The untreated sewage was thawed, and 250 mL of each sample was centrifuged at 10,000 × g for 10 min to retain the pelleted bacteria. DNA was extracted using our previously published modified DNA extraction protocol⁴⁹, and low-concentration samples were vacuum-concentrated. All samples were shipped on dry ice for sequencing at Admera Health (New Jersey, USA). Here, KAPA Hyper PCR-free library preparation was used, and sequencing was carried out on the Illumina NovaSeq6000, targeting >35 M clusters, with paired-end (2) sequencing and 150 cycles per end. We thus aimed for at least 35 M × 2 ends × 150 bp = 10.5 Gbp of sequence per sample for comparability with previous studies. The cities from which sewage samples were retrieved are marked in Supplementary Fig. 1, with the number of samples from each country and sampling year. A combined list of samples, metadata, sequence data accessions, and other relevant information can be found in Supplementary Data 1.

Bioinformatic processing of reads with ARGprofiler

We processed the raw FASTQ sewage metagenomes using our recently published pipeline, ARGprofiler²⁶ v1.0.0, with default settings. This pipeline was designed with snakemake 7.30.1^{50,51} to profile ARGs both quantitatively and qualitatively in metagenomic datasets. In brief, ARGprofiler takes the raw FASTQ reads and begins by quality checking and trimming them with fastp⁵² 0.23.2. Settings for fastp were set as follows: overlap_diff_limit = 1, average_qual = 20, length_required = 50, --cut_tail. After trimming, ARGprofiler uses KMA⁵³ 1.4.12a to do global alignment of the trimmed reads against the two reference databases, mOTUs⁵⁴ (v3.0.3) and PanRes²⁶ v1.0.1. The following alignment parameters were used: 1, -2, -3, -1 for a match, mismatch, gap opening, and gap extension. For pairing of reads, we used a value of 7 and a minimum relative alignment score of 0.75.

Additionally, all the new sequence runs were subjected to metagenomic assembly, following the same approach as in the previous study^{15,16}. Metagenomic assembly was done using metaSPAdes⁵⁵ (SPAdes v. 3.13.0) with k-mer sizes: 27, 47, 67, 87, 107, 127 and the pre-correction flag set. We filtered away contigs shorter than 1 Kbp due to their large numbers and lack of synteny information, which is important for epidemiological inference¹⁵.

Reference sequence databases. We used the PanRes database as the reference collection for ARGs, as it incorporates multiple existing databases of ARGs into a single non-redundant collection, totaling 14,078 unique ARGs²⁶. This study focused on three of the included source databases: ResFinder¹², ResFinderFG 2.0¹⁴, and those from Daruka et al.⁸ (tagged as csabapal in PanRes). The latter two databases are products of FG. Throughout this study, we group the references based on their presence in ResFinder or the functional collections and refer to them as acquired or FG respectively. The few ARGs in both groups were uniquely counted as acquired for the purpose of this study (Supplementary Fig. 1).

We homology reduced all PanRes reference sequences using USEARCH⁵⁶ (v.11.0.66) to 98% nucleotide identity, with query and target coverage thresholds of 98%. To maintain more alleles in the

acquired and FG groups, we assigned cluster representatives as acquired if they had at least one cluster member tagged as acquired and cluster representatives as FG if they had an FG member and no acquired cluster members (Supplementary Data 2).

ARGprofler uses the mOTUs database^{54,57} (v3.0.3) with the KMA aligner⁵³ (v. 14.12.a) to profile the metagenomes. To deal with many mOTUs having a placeholder name (orphans), we padded the missing names with the highest known taxonomic ranks. All resulting KMA mapstat files with mOTUs sequences were filtered to retain only alignments to bacteria. For analyses of beta-diversities, we filtered away taxonomies that made up less than 0.0001% of the reads assigned to mOTUs across all metagenomes.

Quantifying ARGs using compositional methods

We reported the total number of read fragments aligned to the three reference sequence sets (mOTU, Acquired, and FG), as well as the number of unique reference hits and Shannon⁵⁸ diversity for each sewage sample.

Estimation of the regional pan- and core-resistomes was done as follows: the number of unique ARG references hit in samples from a region was calculated as the size of the regional pan-resistome. The core-resistome was then estimated as the number of unique ARGs that were observed in at least 50% of the regional samples. The pan- and core-resistomes were then stratified by the AMR class for each collection (acquired and FG), and the subset of those classes in both collections was plotted.

KMA reports the count of read fragments aligned to reference sequences, which we treat as parts of a composition⁵⁹. We applied the additive log-ratio transformation (ALR) to length-adjusted ARG read fragment counts with the sum of bacterial mOTUs reads as the reference component.

To perform statistical analysis on ARG abundances, we imputed zeroes following procedures highlighted in previous work⁵⁹. The length-adjusted and zero-replaced abundance counts were then transformed with the centered log-ratio (CLR) transformation, which uses the geometric mean as the reference component^{60–62}.

We compared sample beta diversities using Principal Component Analysis (PCA), following the methodology outlined in previous work^{60–62}. The zero-imputed and length-adjusted counts were centered by the geometric mean, scaled by the total log-ratio variance, transformed into CLR values, and eigen-decomposed to obtain eigenvectors and eigenvalues. These were then used to calculate the principal components and visualized in biplots. In Python 3.12⁶³, the pycodamath⁶⁴ package was used for the compositional analyses, and visualization of abundances and diversities was created using matplotlib 3.8.2⁶⁵, seaborn 0.13.2⁶⁶, and geopandas 0.14.3⁶⁷.

In R 4.3.2⁶⁸, the Procrustes function from the vegan 2.6⁶⁹ was used to calculate the degree to which resistome and bacteriome PCA scores correlated, and the adonis2 function was used to determine the proportion of beta-diversity variance attributable to World Bank Regions.

Metagenomic ARG variants

Using the metagenomic assembled scaffolds produced by metaSPAdes (SPAdes v. 3.13.0), we searched for ARG variants from ResFinder and the functional ARG collections with the Flankophile tool⁶⁸ (v. 0.2.8). Briefly, the pipeline searches assemblies for ARGs, extracts and clusters them, and calls new variants.

Using the annotations of which ARGs and identified variants were found across the assemblies, we created contingency tables summarizing observed occurrences of the ARG sequences (closest known reference and specific variant) across geographical groupings (region, country, and city). ARG sequences that were only observed once and geographical areas with less than 100 ARG copies observed were discarded. Each sample in the contingency tables was normalized using

the Hellinger transformation⁷⁰, as implemented in the package vegan 2.6⁶⁹ for R 4.3.2⁶⁸.

Unimap Manifold Approximation and Projection (UMAP)⁷¹, as implemented in the R package umap 0.2.10.0⁷², was applied to the normalized contingency matrices to reduce their dimensionality, allowing a projection of the number of closest references and called variants shared between the different cities.

Estimating distance-decay relationships

The distance-decay effects were investigated by comparing physical distances with resistome similarities. We calculated the rate of distance-decay as the slope of linear regression on the relationship of ln-transformed similarities with ln-transformed distances, as outlined in Martiny et al.³² and Nekola and White³³. The physical distance between cities was obtained by calculating geodesic distances based on latitudes and longitudes with the R package sf 1.0.15⁷³. Abundance similarities were calculated as Aitchison distances (Euclidean distances on CLR values) with the siml function from the proxyC 0.4.1 package⁷⁴. Similarities of the assembled variant resistomes were calculated using Bray-Curtis dissimilarities⁷⁵ on the contingency tables as implemented in the vegdist function in vegan 2.6⁶⁹.

Distances and dissimilarities were log-transformed, and linear models were fitted both for comparing all samples and within stratified groups of comparisons: those with cities within the same country, cities within the same region, and cities from different world regions. The slopes of the lines for the three different spatial groupings were compared with the emtrends function from emmeans 1.10.5⁷⁶. Mantel tests⁷⁷ were applied to measure the correlation of the physical and resistome distance matrices (not ln-transformed) with 999 permutations, as implemented in the mantel function in vegan 2.6⁶⁹.

Network analyses

A shared correlation network was constructed from mOTU and ARG abundance data. To increase the signal-to-noise ratio and reduce sparseness, we imposed a 25% sample presence prevalence threshold for mOTUs and a median adjusted measure ≥ 1 .

For ARGs, due to variable sample depths (as shown in Supplementary Fig. 11), we excluded samples with total abundances below the 10th percentile. Subsequently, we applied a 25% prevalence threshold and a median of non-zero abundances ≥ 0.01 .

After these filtering steps, we preserved only samples that retained <90% of the abundance for mOTUs and ARGs. This led to 912 samples organized into two matrices: 1865 mOTUs and 323 AMR genes (252 FG, 71 Acquired).

We applied CLR transformations, replacing zeros with a pseudo-count (65% of the detection limit)⁷⁸, merged the normalized data, calculated Spearman correlations, and derived an adjacency matrix using an absolute threshold of ≥ 0.5 . The resulting network is defined as undirected, with weighted and signed links where each node represents an mOTU or an ARG, and the links represent the preserved correlations between them. Finally, we applied a signed version of the modularity algorithm⁷⁹, capable of handling negative link weights, to detect communities within the network.

To identify the human microbiota, we downloaded the Unified Human Gastrointestinal Genome collection (UHGG) (v2.0.2)⁸⁰ and matched the species from this database to those in our classified communities by their clustered names to determine the number of human gut-associated mOTUs. The entire analysis was conducted in R using the mgnet package v0.3-beta (<https://github.com/Fuschi/mgnet>).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw sequencing reads and metagenomic assemblies have been deposited at the European Nucleotide Archive for the different rounds of sampling, which are available under project accession numbers: [PRJEB40798](#), [PRJEB40816](#), [PRJEB40815](#), [PRJEB27621](#), and [PRJEB84064](#). The processed data such as KMA mapstat files, count matrices, and flankophile output, have been deposited on Zenodo at <https://doi.org/10.5281/zenodo.14652832>. The metadata associated with the samples and reference genes are in the Supplementary Data Files 1, 2 and network data in Supplementary Data Files 3, 5. Source data are provided with this paper. This study also utilized the Natural Earth dataset to plot geographical shapes and the publicly available databases PanRes, mOTUs and the UHGG catalog. Source data are provided with this paper.

Code availability

All code used for analysis and data visualizations has been deposited on GitHub at https://github.com/genomicepidemiology/g3_acquired_vs_FG⁸¹.

References

- Murray, C. J. L. et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* **399**, 629–655 (2022).
- Naghavi, M. et al. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *Lancet* **404**, 1199–1226 (2024).
- One Health Global Leaders Group on Antimicrobial Resistance. GLG Report: Towards Specific Commitments and Action in the Response to Antimicrobial Resistance. <https://www.amrleaders.org/about-us/what-we-do/glg-report> (2024).
- D’Costa, V. M. et al. Antibiotic resistance is ancient. *Nature* **477**, 457–461 (2011).
- Van Goethem, M. W. et al. A reservoir of ‘historical’ antibiotic resistance genes in remote pristine Antarctic soils. *Microbiome* **6**, 40 (2018).
- Toth, M., Smith, C., Frase, H., Mobashery, S. & Vakulenko, S. An antibiotic-resistance enzyme from a deep-sea bacterium. *J. Am. Chem. Soc.* **132**, 816–823 (2010).
- Sommer, M. O. A., Dantas, G. & Church, G. M. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* **325**, 1128–1131 (2009).
- Zhu, G. et al. Air pollution could drive global dissemination of antibiotic resistance genes. *ISME J.* **15**, 270–281 (2021).
- Bhullar, K. et al. Antibiotic resistance is prevalent in an isolated cave microbiome. *PLoS ONE* **7**, e34953 (2012).
- Finley, R. L. et al. The scourge of antibiotic resistance: the important role of the environment. *Clin. Infect. Dis.* **57**, 704–710 (2013).
- Inda-Díaz, J. S. et al. Latent antibiotic resistance genes are abundant, diverse, and mobile in human, animal, and environmental microbiomes. *Microbiome* **11**, 44 (2023).
- Bortolaia, V. et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.* **75**, 3491–3500 (2020).
- Daruka, L. et al. ESKAPE pathogens rapidly develop resistance against antibiotics in development in vitro. *Nat. Microbiol.* **10**, 313–331 (2025).
- Gschwind, R. et al. ResFinderFG v2.0: a database of antibiotic resistance genes obtained by functional metagenomics. *Nucleic Acids Res.* **51**, W493–W500 (2023).
- Munk, P. et al. Genomic analysis of sewage from 101 countries reveals global landscape of antimicrobial resistance. *Nat. Commun.* **13**, 7251 (2022).
- Hendriksen, R. S. et al. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nat. Commun.* **10**, 1124 (2019).
- Becsei, Á et al. Time-series sewage metagenomics distinguishes seasonal, human-derived and environmental microbial communities potentially allowing source-attributed surveillance. *Nat. Commun.* **15**, 7551 (2024).
- Larsson, D. G. J. & Flach, C.-F. Antibiotic resistance in the environment. *Nat. Rev. Microbiol.* **20**, 257–269 (2022).
- Berglund, F., Ebmeyer, S., Kristiansson, E. & Larsson, D. G. J. Evidence for wastewaters as environments where mobile antibiotic resistance genes emerge. *Commun. Biol.* **6**, 1–11 (2023).
- Mitobedzka, A. et al. Monitoring antibiotic resistance genes in wastewater environments: the challenges of filling a gap in the One-Health cycle. *J. Hazard. Mater.* **424**, 127407 (2022).
- Njage, P. M. K., van Bunnik, B., Munk, P., Marques, A. R. P. & Aarstrup, F. M. Association of health, nutrition, and socioeconomic variables with global antimicrobial resistance: a modelling study. *Lancet Planet. Health* **7**, e888–e899 (2023).
- Collignon, P., Beggs, J. J., Walsh, T. R., Gandra, S. & Laxminarayan, R. Anthropological and socioeconomic factors contributing to global antimicrobial resistance: a univariate and multivariable analysis. *Lancet Planet. Health* **2**, e398–e405 (2018).
- Nielsen, T. K., Browne, P. D. & Hansen, L. H. Antibiotic resistance genes are differentially mobilized according to resistance mechanism. *GigaScience* **11**, giac072 (2022).
- Morlon, H. et al. A general framework for the distance–decay of similarity in ecological communities. *Ecol. Lett.* **11**, 904–917 (2008).
- Liu, C., Chen, J., Yang, Y., Teng, Y. & Chen, H. Biogeography and diversity patterns of antibiotic resistome in the sediments of global lakes. *J. Environ. Sci.* **127**, 421–430 (2023).
- Martiny, H.-M. et al. ARGprofiler—a pipeline for large-scale analysis of antimicrobial resistance genes and their flanking regions in metagenomic datasets. *Bioinformatics* **40**, btac086 (2024).
- Djordjevic, S. P. et al. Genomic surveillance for antimicrobial resistance—a One Health perspective. *Nat. Rev. Genet.* **25**, 142–157 (2024).
- Jespersen, M. L. et al. Global within-species phylogenetics of sewage microbes suggest that local adaptation shapes geographical bacterial clustering. *Commun. Biol.* **6**, 1–9 (2023).
- Klümper, U. et al. Environmental microbiome diversity and stability is a barrier to antimicrobial resistance gene accumulation. *Commun. Biol.* **7**, 1–13 (2024).
- Pehrsson, E. C., Forsberg, K. J., Gibson, M. K., Ahmadi, S. & Dantas, G. Novel resistance functions uncovered using functional metagenomic investigations of resistance reservoirs. *Front. Microbiol.* **4**, 145 (2013).
- Anthony, W. E., Burnham, C.-A. D., Dantas, G. & Kwon, J. H. The gut microbiome as a reservoir for antimicrobial resistance. *J. Infect. Dis.* **223**, S209–S213 (2021).
- Martiny, J. B. H., Eisen, J. A., Penn, K., Allison, S. D. & Horner-Devine, M. C. Drivers of bacterial β -diversity depend on spatial scale. *Proc. Natl. Acad. Sci.* **108**, 7850–7854 (2011).
- Nekola, J. C. & White, P. S. The distance decay of similarity in biogeography and ecology. *J. Biogeogr.* **26**, 867–878 (1999).
- Horner-Devine, M. C., Lage, M., Hughes, J. B. & Bohannan, B. J. M. A taxa–area relationship for bacteria. *Nature* **432**, 750–753 (2004).
- Louca, S. The rates of global bacterial and archaeal dispersal. *ISME J.* **16**, 159–167 (2022).
- Ramette, A. & Tiedje, J. M. Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proc. Natl. Acad. Sci.* **104**, 2761–2766 (2007).
- Martiny, J. B. H. et al. Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* **4**, 102–112 (2006).
- Bell, G. Neutral macroecology. *Science* **293**, 2413–2418 (2001).
- Garcillán, P. P. & Ezcurra, E. Biogeographic regions and β -diversity of woody dryland legumes in the Baja California peninsula. *J. Veg. Sci.* **14**, 859–868 (2003).

40. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (MPB-32). (Princeton University Press, Princeton, NJ, 2001).
41. Qin, J. et al. A human gut microbial gene catalog established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
42. Eckburg, P. B. et al. Diversity of the human intestinal microbial flora. *Science* **308**, 1635–1638 (2005).
43. Magne, F. et al. The firmicutes/bacteroidetes ratio: a relevant marker of gut dysbiosis in obese patients. *Nutrients* **12**, 1474 (2020).
44. Koliada, A. et al. Association between body mass index and Firmicutes/Bacteroidetes ratio in an adult Ukrainian population. *BMC Microbiol.* **17**, 120 (2017).
45. Turnbaugh, P. J. et al. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
46. Lam, K. N., Cheng, J., Engel, K., Neufeld, J. D. & Charles, T. C. Current and future resources for functional metagenomics. *Front. Microbiol.* **6**, 1196 (2015).
47. Bengtsson-Palme, J. et al. Towards monitoring of antimicrobial resistance in the environment: for what reasons, how to implement it, and what are the data needs. *Environ. Int.* **178**, 108089 (2023).
48. Martínez, J. L., Coque, T. M. & Baquero, F. What is a resistance gene? Ranking risk in resistomes. *Nat. Rev. Microbiol.* **13**, 116–123 (2015).
49. Knudsen, B. E. et al. Impact of sample type and DNA isolation procedure on genomic inference of microbiome composition. *mSystems* **1**, <https://doi.org/10.1128/msystems.00095-16> (2016).
50. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012).
51. Köster, J. et al. Sustainable data analysis with Snakemake. *F1000Research* **10**, 33 (2021).
52. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
53. Clausen, P. T. L. C., Aarestrup, F. M. & Lund, O. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinform.* **19**, 307 (2018).
54. Ruscheweyh, H.-J. et al. Cultivation-independent genomes greatly expand taxonomic-profiling capabilities of mOTUs across various environments. *Microbiome* **10**, 212 (2022).
55. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
56. Edgar, R. C. & Bateman, A. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
57. Milanese, A., Ruscheweyh, H.-J., Mende, D., Zeller, G. & Sunagawa, S. mOTUs database. <https://doi.org/10.5281/zenodo.5140350> (2021).
58. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
59. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome datasets are compositional: and this is not optional. *Front. Microbiol.* **8**, 2224 (2017).
60. Martiny, H.-M. et al. Global distribution of mcr gene variants in 214K metagenomic samples. *mSystems* **7**, e0010522 (2022).
61. Aitchison, J. Principal component analysis of compositional data. *Biometrika* **70**, 57–65 (1983).
62. Calle, M. L. Statistical analysis of metagenomics data. *Genom. Inf.* **17**, e6 (2019).
63. Van Rossum, G. & Drake, F. L. Python 3 Reference Manual. CreateSpace (2009).
64. Brinch, C. pyCoDaMath. Bitbucket https://bitbucket.org/genomic_epidemiology/pycodamath (2025).
65. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
66. Waskom, M. L. seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
67. Jordahl, K. et al. GeoPandas: Python tools for geographic data.
68. Thorn, A. V., Aarestrup, F. M. & Munk, P. Flankophile: a bioinformatic pipeline for prokaryotic genomic synteny analysis. *Microbiol. Spectr.* **0**, e02413–e02423 (2023).
69. Oksanen, J. et al. *Vegan: Community Ecology Package*. (2022).
70. Legendre, P. & Gallagher, E. D. Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**, 271–280 (2001).
71. McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at <https://doi.org/10.48550/arXiv.1802.03426> (2020).
72. Konopka, T. umap: Uniform Manifold Approximation and Projection. (2023).
73. Pebesma, E. Simple features for R: standardized support for spatial vector data. *R. J.* **10**, 439–446 (2018).
74. Watanabe, K. & Cannoodt, R. proxyC: computes proximity in large sparse matrices. 0.4.1 <https://doi.org/10.32614/CRAN.package.proxyC> (2018).
75. Bray, J. R. & Curtis, J. T. An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* **27**, 326–349 (1957).
76. Lenth, R. V. et al. emmeans: estimated marginal means, aka least-squares means. (2024).
77. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res* **27**, 209–220 (1967).
78. Martín-Fernández, J. A., Barceló-Vidal, C. & Pawlowsky-Glahn, V. Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Math. Geol.* **35**, 253–278 (2003).
79. Gómez, S., Jensen, P. & Arenas, A. Analysis of community structure in networks of correlated data. *Phys. Rev. E* **80**, 016114 (2009).
80. Almeida, A. et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat. Biotechnol.* **39**, 105–114 (2021).
81. Martiny, H. M., Fuschi, A., Munk, P., Becsei, A., & Pyrounakis, N. Geographics and bacterial networks differently shape the acquired and latent global sewage resistomes (genomic epidemiology/ gs3_acquired_vs_FG). Zenodo <https://doi.org/10.5281/ZENODO.17249256> (2025).

Acknowledgements

We want to extend our gratitude to everyone who has helped with sampling, shipping, and the logistics of transporting sewage to Denmark. A special thanks to the laboratory technicians at DTU that has organized the sample collection, material transfer, sequencing and logistics. Lastly, we would like to thank the Novo Nordisk Foundation (Grant: NNF16OC0021856: Global Surveillance of Antimicrobial Resistance) and the European Union's Horizon H2020 research and innovation programme (Grant: No. 874735) for funding the work.

Author contributions

F.M.A. and P.M. conceived the study, and F.M.A. and M.K. secured the funding. H.M.M., P.M., and N.P. did quality control and sample selection. Trimming, mapping and assembly were performed by N.P., H.M.M., P.M., A.F., and A.B. carried out data and statistical analyses and visualization. H.M.M. and P.M. drafted the initial manuscript with input from A.F., A.B., C.B., D.G.J.L., M.K., D.R., I.C., and F.M.A. Members of the global sewage consortium carried out sewage sampling, filled in metadata and shipped the samples to DTU. All authors helped to review and improve the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-66070-7>.

Correspondence and requests for materials should be addressed to Hannah-Marie Martiny or Patrick Munk.

Peer review information *Nature Communications* thanks Hyatt Green, Pablo Tsukayamam, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

Global Sewage Consortium

Hannah-Marie Martiny^{1,123}✉, Patrick Munk^{1,123}✉, Nikiforos Pyrounakis¹, Christian Brinch¹, Frank M. Aarestrup¹, Alessandro Fuschi², Daniel Remondini², Ágnes Becsei³, István Csabai³, D. G. Joakim Larsson⁴, Marion Koopmans⁵, Nawel Zaatout⁶, Catherine Rees⁷, Guangming Jiang⁸, Jiahua Shi⁹, Bernhard Benka¹⁰, Franz Allerberger¹⁰, Sandra Koeberl-Jelovcan¹⁰, Khalil Hasan Abdulla¹¹, Ali Bin Thani¹², Anowara Begum¹³, Carlon Worrell¹⁴, Tamegnon Victorien Dougnon¹⁵, Freddy Soria¹⁶, Natasa Mazalica¹⁷, Teddie Rahube¹⁸, Andreza Francisco Martins¹⁹, Carlos Alberto Tagliati²⁰, Larissa Camila Ribeiro de Souza²⁰, Ariane Nzouankeu²¹, Muhammad Attiq Rehman²², Jeff Gauthier²³, Roger C. Levesque²³, Sean D. Workman²⁴, Christopher Yost²⁴, Aiko Adell Nakashima²⁵, Andres Opazo²⁶, Gerardo Gonzales²⁶, Yongjie Yu²⁷, Pilar Donado-Godoy²⁸, Tadjidine Youssouf²⁹, Pablo-César Rivera-Navarro³⁰, Matijana Jergovic³¹, Jasna Hrenovic³², Renata Karpiskova³³, Julien Kalpy Coulibaly³⁴, William Calero-Cáceres³⁵, Mohamed Abouelnaga³⁶, Anna-Maria Hokajärvi³⁷, Annamari Heikinheimo³⁸, Soizick Le Guyader³⁹, Andreas Nitsche⁴⁰, Annika Brinkmann⁴⁰, Sara Schubert⁴¹, Thomas Berendonk⁴¹, Uli Klümper⁴¹, Ernest Bonah⁴², Solomon Asante Sefa⁴³, Andrew Camilleri⁴⁴, Courage Kosi Setsoafia Saba⁴⁴, Victoria Bernice Sedor⁴⁵, Kassiani Mellou⁴⁶, Theologia Sideroglou⁴⁶, Charalampos Kotzamanidis⁴⁷, Jens-Peter B. Henriksen⁴⁸, Mie Møller⁴⁹, Thorunn Rafnar Thorsteinsdottir⁵⁰, A. A. Mohamed Hatha⁵¹, Sima Mohammad⁵², Burhan Shamurad⁵³, Hiwa Ali Faraj⁵³, Dearbháile Morris⁵⁴, Louise O'Connor⁵⁴, Jacob Moran-Gilad⁵⁵, Roberta Orletti⁵⁶, Giuseppina La Rosa⁵⁷, Marcello Iaconelli⁵⁷, Antonio Battisti⁵⁸, Lucia Decastelli⁵⁹, Maira Napoleoni⁶⁰, Alessandra De Cesare⁶¹, Gianluca Corno⁶², Jelena Avsejenko⁶³, Ghassan M. Matar⁶⁴, Christian Penny⁶⁵, Luc Hervé Samison⁶⁶, Daniel L. Banda⁶⁷, Heera Rajandas⁶⁸, Sivachandran Parimannan⁶⁸, Malcolm Vella Haber⁶⁹, Sunita J. Santchurn⁷⁰, Julian Carrillo-Reyes⁷¹, Julian Osvaldo Sanchez-Lara⁷¹, René Arredondo-Hernández⁷¹, Aleksandar Vujacic⁷², Dijana Djurovic⁷², Pushkar Pal⁷³, Carlos J. A. Campos⁷⁴, Isabelle Pattis⁷⁵, Stephen Chambers⁷⁶, Gert-Jan Jeunen⁷⁷, Charles Elikwu⁷⁸, Olayinka Osuolale⁷⁹, Mwapu Ndahi⁸⁰, Oluwadamilola Abiodun-Adewusi⁸¹, Amen Ekhosuehi⁸², Ayorinde Afolayan⁸³, Kayode Fashae⁸³, Mabel Kamweli Aworh⁸³, Akeem Olayiwola Ahmed⁸⁴, Ibrahim Adisa Raufu⁸⁴, Ismail Odetokun⁸⁴, Samaila Musa Chiroma⁸⁵, Claire Mitchell⁸⁶, Rune Holmstad⁸⁷, Natalie Weiler⁸⁸, Dariusz Wasyl⁸⁹, Magdalena Zając⁸⁹, Eugénia Cardoso⁹⁰, Célia Manaia⁹¹, Maria Jorge Campos⁹², Hor-Gil Hur⁹³, Min Joon Song⁹⁴, Sukhwan Yoon⁹⁴, Olga Burduniuc⁹⁵, Peiying Hong⁹⁶, Vladimir Radosavljevic⁹⁷, Angel Anika Cokro⁹⁸, Dagmar Gavacová⁹⁹, Marija Trkov¹⁰⁰, Rukayya Hussain Abubakar¹⁰¹, Karen Keddy¹⁰², Isabel Martínez-Alcalá¹⁰³, Marta Cerdà-Cuellar¹⁰⁴, Ana Cañas¹⁰⁵, Fernando Gonzalez-Candelas¹⁰⁶, Yahia Ali Sabiel¹⁰⁷, Tanja van der Heijden¹⁰⁸, Yiping Hong¹⁰⁹, Julius John Medardus¹¹⁰, Rajeev P. Nagassar¹¹¹, Cemil Kürekci¹¹², Patrick McNamara¹¹³, Lisa Wong¹¹⁴, Erica Fuhrmeister¹¹⁵, John Scott Meschke¹¹⁶, Nicola Beck¹¹⁶, Ayella Maile-Moskowitz¹¹⁷, Jens Thomsen¹¹⁸, Ouidiane Obaidi¹¹⁹, Bruce Paterson¹²⁰, Anne Leonard¹²¹, Lihong Zhang¹²¹ & Kevin Chau¹²²

⁶University of Batna 2, Batna, Algeria. ⁷Melbourne Water Corporation, Melbourne, VIC, Australia. ⁸School of Civil, Mining, Environmental, and Architecture Engineering, University of Wollongong, Wollongong, NSW, Australia. ⁹University of Wollongong, Wollongong, NSW, Australia. ¹⁰Austrian Agency for Health and Food Safety (AGES), Vienna, Austria. ¹¹The Supreme Council for Environment (SCE) in Bahrain, Manama, Bahrain. ¹²University of Bahrain, Sakhir, Bahrain. ¹³University of Dhaka, Dhaka, Bangladesh. ¹⁴Environmental Protection Department, Bridgetown, St. Michael, Barbados. ¹⁵Polytechnic School of Abomey-Calavi, Abomey-Calavi, Benin. ¹⁶Universidad Católica Boliviana San Pablo, La Paz, Bolivia. ¹⁷Public Health Institute of the Republic of Srpska, Banja Luka, Bosnia

and Herzegovina. ¹⁸Botswana International University of Science and Technology, Palapye, Botswana. ¹⁹Federal University of Rio Grande do Sul, Porto Alegre, Brazil. ²⁰Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. ²¹Centre Pasteur du Cameroun, Yaoundé, Cameroon. ²²Research and Productivity Council, Fredericton, NB, Canada. ²³Université Laval, Québec, QC, Canada. ²⁴University of Regina, Regina, SK, Canada. ²⁵Escuela de Medicina Veterinaria, Facultad de Ciencias de la Vida, Universidad Andrés Bello, Santiago, Chile. ²⁶Universidad de Concepción, Concepción, Chile. ²⁷Nanjing University of Information Science and Technology, Nanjing, China. ²⁸Global Health Research Unit on Genomic Surveillance of Antimicrobial Resistance (GHRU), CI Tibaitatá, Corporación Colombiana de Investigación Agropecuaria (AGROSAVIA), Funza, Colombia. ²⁹Laboratoire National de Référence Hospitalier, Moroni, Comoros. ³⁰University of Costa Rica, San José, Costa Rica. ³¹Andrija Stampar Teaching Institute of Public Health, Zagreb, Croatia. ³²University of Zagreb, Zagreb, Croatia. ³³Veterinary Research Institute, Brno, Czechia. ³⁴Institut Pasteur de Côte d'Ivoire, Abidjan, Côte d'Ivoire. ³⁵UTA-RAM-One Health, Group for Universal Advance in bioScience, Department of Food and Biotechnology Science and Engineering, Universidad Técnica de Ambato, Ambato, Ecuador. ³⁶Suez Canal University, Ismailia, Egypt. ³⁷Tampere University, Tampere, Finland. ³⁸University of Helsinki, Helsinki, Finland. ³⁹French Institute Search Pour L'exploitation De LaMer (Ifremer), Nantes, France. ⁴⁰Robert Koch Institute, Berlin, Germany. ⁴¹Institute of Hydrobiology, Technische Universität Dresden, Dresden, Germany. ⁴²Food and Drugs Authority, Accra, Ghana. ⁴³Public Health Laboratory, Ghana Health Service, Takoradi, Ghana. ⁴⁴University for Development Studies, Tamale, Ghana. ⁴⁵Veterinary Services Department, Ministry of Food and Agriculture, National Food Safety Laboratory, Accra, Ghana. ⁴⁶National Public Health Organisation (EODY), Athens, Greece. ⁴⁷Veterinary Research Institute of Thessaloniki, Hellenic Agricultural Organisation-DEMETER, Thessaloniki, Greece. ⁴⁸Kommuneqarfi Sermersooq, Nuuk, Greenland. ⁴⁹Peqqik, Nuuk, Greenland. ⁵⁰University of Iceland, Reykjavik, Iceland. ⁵¹Cochin University of Science and Technology, Cochin, India. ⁵²Shahid Beheshti University, Tehran, Iran. ⁵³The University of Sulaimani, Sulaymaniyah, Iraq. ⁵⁴National University of Ireland, Galway, Ireland. ⁵⁵Ben-Gurion University of the Negev and Ministry of Health, Beer-Sheva, Israel. ⁵⁶Ambiente, Marche, Italy. ⁵⁷Istituto Superiore di Sanità, Rome, Italy. ⁵⁸Istituto Zooprofilattico Sperimentale del Lazio e della Toscana, Rome, Italy. ⁵⁹Istituto Zooprofilattico Sperimentale del Piemonte, Liguria e Valle d'Aosta, Italy. ⁶⁰Istituto Zooprofilattico Sperimentale dell'Umbria e delle Marche "Togo Rosati", Perugia, Italy. ⁶¹University of Bologna, Bologna, Italy. ⁶²Water Research Institute, National Research Council (CNR), Verbania, Italy. ⁶³Institute of Food Safety, Riga, Latvia. ⁶⁴American University of Beirut, Beirut, Lebanon. ⁶⁵Luxembourg Institute of Science and Technology, Belvaux, Luxembourg. ⁶⁶University of Antananarivo, Centre d'Infectiologie Charles Mérieux, Antananarivo, Madagascar. ⁶⁷University of Malawi, Blantyre, Malawi. ⁶⁸AIMST University, COMBio, Kedah, Malaysia. ⁶⁹Environmental Health Directorate, St. Venera, Malta. ⁷⁰University of Mauritius, Reduit, Mauritius. ⁷¹Universidad Nacional Autónoma de México, Mexico City, Mexico. ⁷²Institute for Public Health Montenegro, Podgorica, Montenegro. ⁷³Agriculture and Forestry University, Kathmandu, Nepal. ⁷⁴Cawthron Institute, Nelson, New Zealand. ⁷⁵Institute of Environmental Science and Research Limited (ESR), Christchurch, New Zealand. ⁷⁶University of Otago, Christchurch, New Zealand. ⁷⁷University of Otago, Dunedin, New Zealand. ⁷⁸Babcock University Teaching Hospital, Ilishan-Remo, Nigeria. ⁷⁹Elizade University, Ilara-Mokin, Nigeria. ⁸⁰Federal Ministry of Agriculture and Rural Development, Abuja, Nigeria. ⁸¹Nigeria Field Epidemiology Training Program, Abuja, Nigeria. ⁸²University of Benin, Benin City, Nigeria. ⁸³University of Ibadan, Ibadan, Nigeria. ⁸⁴University of Ilorin, Ilorin, Nigeria. ⁸⁵University of Maiduguri, Maiduguri, Nigeria. ⁸⁶Department of Agriculture, Environment and Rural Affairs, Northern Ireland, UK. ⁸⁷VEAS, Slemmestad, Norway. ⁸⁸Departamento de Bacteriología, Laboratorio Central de Salud Público, Asunción, Paraguay. ⁸⁹National Veterinary Research Institute, Pulawy, Poland. ⁹⁰Águas de Portugal, SGPS, S.A, Lisboa, Portugal. ⁹¹Centre for Biotechnology and Fine Chemistry (CBQF), Universidade Católica Portuguesa, Porto, Portugal. ⁹²Escola Superior de Turismo e Tecnologia do Mar (ESTM), Instituto Politécnico de Leiria (IPLeiria), Peniche, Portugal. ⁹³Gwangju Institute of Science and Technology, Gwangju, Republic of Korea. ⁹⁴Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea. ⁹⁵National Agency for Public Health, Chişinău, Republic of Moldova. ⁹⁶King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. ⁹⁷Institute of Veterinary Medicine of Serbia, Belgrade, Serbia. ⁹⁸Nanyang Technological University, Singapore, Singapore. ⁹⁹Public Health Authority of the Slovak Republic, Bratislava, Slovakia. ¹⁰⁰National Laboratory of Health, Environment and Food, Ljubljana, Slovenia. ¹⁰¹Faculty of Veterinary Sciences, University of Pretoria, South Africa. ¹⁰²Independent consultant, Johannesburg, South Africa. ¹⁰³Catholic University of Murcia, Murcia, Spain. ¹⁰⁴Centre de Recerca en Sanitat Animal (CRESA), IRTA-UAB, Universitat Autònoma de Barcelona, Bellaterra, Spain. ¹⁰⁵Instituto de Salud Carlos III, Madrid, Spain. ¹⁰⁶Universidad de Valencia, Valencia, Spain. ¹⁰⁷Sudan University of Science and Technology, Khartoum, Sudan. ¹⁰⁸Ara Region Bern AG, Herrenschwand, Switzerland. ¹⁰⁹Centers for Disease Control, Taipei, Taiwan, ROC. ¹¹⁰Sokoine University of Agriculture, Morogoro, Tanzania. ¹¹¹Eastern Regional Health Authority (ERHA), Sangre Grande, Trinidad and Tobago. ¹¹²Mustafa Kemal Üniversitesi, Hatay, Turkey. ¹¹³Marquette University, Milwaukee, WI, USA. ¹¹⁴Massachusetts Water Resources Authority, Boston, MA, USA. ¹¹⁵Tufts University, Medford, MA, USA. ¹¹⁶University of Washington, Seattle, WA, USA. ¹¹⁷Virginia Tech, Blacksburg, VA, USA. ¹¹⁸Abu Dhabi Public Health Center, Abu Dhabi, United Arab Emirates. ¹¹⁹Dubai Municipality, WWTP Al Aweer, Dubai, United Arab Emirates. ¹²⁰Scottish Environment Protection Agency, Stirling, UK. ¹²¹University of Exeter Medical School, Cornwall, UK. ¹²²University of Oxford, Oxford, UK.