nature water

Article

https://doi.org/10.1038/s44221-024-00310-w

Global scale exploration of human faecal and sewage resistomes as a function of socio-economic status

Received: 23 July 2023

Accepted: 15 August 2024

Published online: 24 September 2024

Check for updates

Suraj Gupta [1], Xiaowei Wu [2], Amy Pruden [3], Liqing Zhang[4] & Peter Vikesland [3]

Prior studies have shown that socio-economic indicators collectively explain most of the variance in sewage resistomes. However, the relationship between human faecal and sewage resistomes has not been well characterized. We investigated common and discriminating features between human faecal and sewage microbiomes and resistomes by analysing 451 publicly available metagenomic samples from 69 countries (240 human faecal samples from 23 countries and 211 urban sewage samples from 60 countries) representing different socio-economic statuses. We found that sewage and human faecal resistome compositions were distinct, with sewage exhibiting higher relative antibiotic resistance gene abundance and total diversity than human faeces. The ANOSIM test revealed stronger separation by socio-economic status in sewage samples ($R = 0.47$) compared to faecal samples ($R = 0.17$). The distinctions between human faecal and sewage resistomes revealed in this study are key considerations in the advancement of sewage surveillance efforts aimed at informing the antibiotic resistance status of human populations.

The continued rise in antibiotic resistance is a global challenge that reflects the influence of multiple interconnected drivers[1]. The World Health Organization (WHO) Global Action Plan[2] is centred on a One Health framework that is correspondingly reflected in National Action Plans developed and implemented globally[3]. These action plans generally emphasize prudent antibiotic use and stewardship, innovation to minimize antibiotic use and some form of surveillance to track changes in antibiotic resistance over time and in response to mitigation efforts. It is increasingly being recognized that coordinated global environmental surveillance is required to support a One Health approach to stem the spread of antibiotic resistance[4,5].

With growing awareness of the importance of the human gut microbiome to health and well being, its role as a reservoir of antibiotic resistance is of particular concern[6–9]. It is challenging, however, to collect individual human faecal samples because of ethical concerns, legal requirements, the need for informed consent and logistical constraints hindering collection of enough samples to support statistically valid conclusions[10]. As an alternative, Aarestrup and Woolhouse[11] and others have advocated for sewage-based antibiotic resistance surveillance via shotgun metagenomic sequencing. The primary purpose of such surveillance is to first profile the antibiotic resistance genes (ARGs) circulating within the human populations that contribute to sewage. Wastewater treatment plant influent sewage serves as a composite sample that captures and reflects the collective ARGs and antibiotic-resistant bacteria (ARB) circulating within populations served by a given sewershed[12]. Consistent with this understanding, Karkman et al.[13] demonstrated that the normalized total ARG abundances in anthropogenically affected environments (such as sewage, hospital wastewater effluent, river and lake sediments) reflect the extent of faecal pollution. Ideally, surveillance systems

[1]Interdisciplinary PhD Program in Genetics, Bioinformatics and Computational Biology, Virginia Tech, Blacksburg, VA, USA. [2]Department of Statistics, Virginia Tech, Blacksburg, VA, USA. [3]Via Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA, USA. [4]Department of Computer Science, Virginia Tech, Blacksburg, VA, USA. e-mail: pvikes@vt.edu

need to provide an early warning for outbreaks of antibiotic-resistant and other infections while also allowing changes in antibiotic resistance patterns to be tracked with time. Culture-based approaches are the focus of traditional clinical and livestock surveillance[14,15] and are appropriate for precise monitoring of specific targets but can only scratch the surface of the broader microbial ecological factors dictating resistance evolution. On the other hand, shotgun metagenomic sequencing has the capacity to broadly capture key trends, including the potential emergence of new forms of resistance[11,16,17]. While shotgun metagenomics lacks sensitivity in detecting low-abundance species relative to cultivation-based or PCR-based methods, it provides target breadth[18].

Whereas human faeces are a fundamental contributor to the sewage microbiome and resistome (that is, the collection of ARGs carried by a microbial community[19]), the harsh physico-chemical conditions imposed in sewers (for example, variable dissolved oxygen levels, extreme pH and potentially high concentrations of antibiotics and antimicrobials) most certainly create conditions conducive to selection pressure and horizontal gene transfer[20,21]. In addition to human excreta, untreated sewage receives a vast array of microbiological and chemical constituents, including antibiotics and antimicrobials[20,22]. Faecal samples are often collected as representative of the human gut, which is an anaerobic environment[23], whereas sewage is typically conveyed in an aerobic, open-channel flow environment, resulting in a distinct ecosystem. Sewage conveyance systems also receive other non-human microbiome sources and are often plagued by leaks as well as inflow and infiltration. Thus, we hypothesized that faecal microbiomes and resistomes will shift following passage through the sewage collection network. We note three prior published reports assessing the validity of a similar hypothesis. Newton et al.[24] examined taxonomic composition in human faecal and sewage samples using 16S rRNA amplicon sequencing data and demonstrated that sewage represents the faecal microbial community. However, the study made no characterization of the respective resistomes and was focused only on the US population, hence lacking a global dimension. Pal et al.[25] compared resistomes from multiple distinct environments (skin, oral, gastrointenstinal, wastewater, smog and so on) but did not provide detailed comparison of human faecal and sewage resistomes. A study conducted by Su and colleagues[26] in China's major cities demonstrated parity in the bacterial taxonomic makeup of both sewage and human gut samples, but no analysis was done to compare the resistomes. Hence, an in-depth global-scale characterization of faecal and sewage microbiomes and resistomes is warranted.

If the human faecal microbiome changes during sewage conveyance, it has important implications on the development of sewage-based antibiotic resistance surveillance (that is, wastewater surveillance). In addition to tracking trends in ARGs excreted by a community over time, ideally such a system can also provide the resolution required to identify drivers of changes in antibiotic resistance and to assess the efficacy of mitigation strategies[4,5,27]. Recent research has sought to identify pivotal factors shaping both gut[6–9] and sewage resistomes[28–30]. It is often assumed that indiscriminate antibiotic use is the primary driver of the emergence and maintenance of antibiotic resistance within microbial populations[31–34]. However, striking disconnects between antibiotic sales data and antibiotic resistance measures have been identified[35,36]. This finding underscores the importance of holistically considering other potential drivers of antibiotic resistance. Notably, using data collected from WHO and Resistance Map, Collignon et al.[36] found that socio-economic indicators correlate with the observed levels of *Escherichia. coli* and *Klebsiella* spp. resistance in clinical isolates. Similarly, Hendriksen et al.[37] showed that socio-economic indicators collectively explained much of the variance in sewage resistomes collected from across the world. The correlation of socio-economic factors with the gut taxonomic composition has been documented[38,39]. However, little is known with respect to how socio-economics reflects faecal resistomes. The correlation of resistance levels with socio-economic factors such as income, gross domestic product (GDP), health, infrastructure and governance[40–43] suggests that such factors may similarly correlate with human faecal resistomes and the extent to which they are reflected in sewage. Therefore, it is imperative to evaluate the relationship between socio-economic factors and the faecal resistome and to compare the strength of associations, if any, with the sewage resistome.

Faeces and sewage are environmental niches that reflect the extreme high end of microbial density and it is not expected that the total microbial density will change in these environments over time. Hence, it is of importance to examine the relative abundances of bacteria and ARGs within faeces and sewage. In this Article, we carried out large-scale analysis of globally sourced human faecal and sewage metagenomes. We comprehensively compared bacterial resistomes and microbiomes to characterize shifts that occur during sewage conveyance. To determine the potential effects of social and environmental conditions as drivers in the dissemination of antibiotic resistance, a systematic comparison of ARG abundance and diversity was performed across and within resistomes grouped based on country-scale metrics of socio-economic status (https://data.worldbank.org/indicator). Core resistome and discriminatory resistome analyses were performed to identify globally spread ARGs versus those that differentiate resistomes. Finally, regression analyses were performed to elucidate how broad socio-economic indicators correlate with the total relative abundance of ARGs in human faecal and sewage resistomes. This study provides important insight needed to inform the development of sewage surveillance efforts aimed at informing antibiotic resistance status of human populations and at assessing and informing potential interventions.

Data included in this study

During the exploration of public repositories, it was observed that a majority of published gut metagenome datasets are skewed by large cohorts that are available for specific countries or studies. Similarly, the sewage dataset obtained from Hendriksen et al. was also limited with small sample size *N* for several countries, but at the time this study was done, this was the best dataset available that was collected via a uniform sample collection and processing strategy. To address the challenge of limited samples, we sought to compare broad-level differences in resistance profiles across socio-economic groups. The following approach was taken to ensure a sufficient sample size for such comparison. First, the countries were clustered into low, middle and high socio-economic groups based on socio-economic data using K-means clustering. Next, a balanced dataset avoiding over representation of any specific study or country was created.

In total, 275 human faecal samples from 23 countries and 234 urban sewage samples from 62 countries were retrieved from public repositories (Supplementary Table 1). After annotating all 509 samples against the CARD database (v3.0.7), samples with 'unique ARG counts' (the total number of unique ARGs in a sample) and 'total ARG counts' (the total number of ARG hits or counts in a sample) in the bottom fifth percentile were considered as compromised and not included in subsequent analyses. This curation resulted in 240 faecal samples from 23 countries and 211 sewage samples from 60 countries. By design, the samples reflect geographically and socio-economically diverse locations (Supplementary Fig. 1 and Supplemental Table 2). There was less variance in resistome dissimilarities among faecal samples sourced from within the same country than among samples obtained from different countries (permutation test, $P < 0.0001$), which suggests that the samples were representative of the observed antibiotic resistance patterns within a given country. A similar conclusion was reached for the sewage dataset[37]. Lastly, the samples were grouped based on the respective country's socio-economic bin into low, middle and high socio-economic status categories. This resulted in 57, 79 and 104 faecal

https://doi.org/10.1038/s44221-024-00310-w

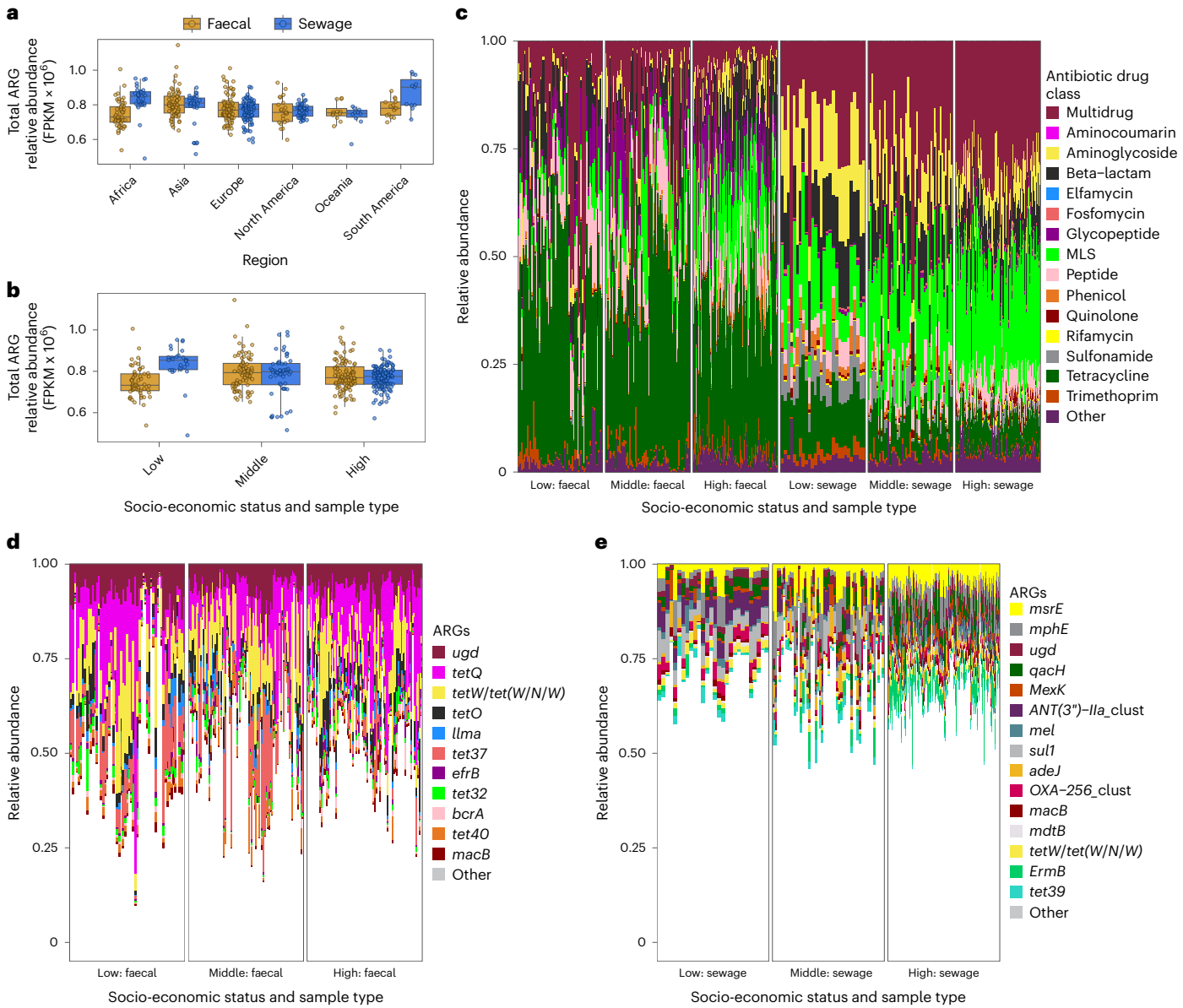

Fig. 1 | Resistome composition and ARG abundance across globally distributed human faecal and sewage samples. a–e, Distribution of total relative abundance of ARGs in human faeces and sewage across continents (**a**) and as a function of socio-economic status (**b**) in units of fragments per kilobase million (FPKM). **c**, Relative abundance of ARGs grouped by drug class in human faeces and sewage as a function of socio-economic status. Relative abundance of the most abundant ARGs in human faeces (**d**) and sewage (**e**), respectively, as classified by socio-economic status. Human faeces data reflect 240 samples across 23 countries collected from publicly available data in the National Center for Biotechnology Information (NCBI) database. Sewage data represent 211 samples across 60 countries collected from publicly available data in NCBI.

samples, and 28, 53 and 130 sewage samples in low, middle and high socio-economic bins, respectively.

Resistome composition patterns by socio-economic status

A total of 617 different ARGs were detected among the faecal samples (with a median of 165 ARGs per sample) and 857 ARGs were detected in the sewage samples (with a median of 394 per sample). The relative abundance of total ARGs varied across locations, socio-economic status categories and sample types (Fig. 1a,b). For the faecal samples, the highest ARG abundances were observed in Asian countries, whereas the lowest abundances were found in North America and Africa. A similar trend was observed when faecal samples were grouped based on socio-economic status (that is, lower total ARG relative abundances for low socio-economic status faecal samples). The trend was somewhat opposite for the sewage samples where the lowest ARG abundances were observed in samples from Oceania and North America and the highest ARG abundances were observed in South America, followed by Africa. Similarly, sewage samples from low socio-economic status countries generally harboured the highest total ARG relative abundances, whereas high socio-economic status countries harboured the lowest levels. Collectively, it was observed that the total relative abundance of ARGs in faecal samples was lower than that observed in sewage samples.

ARGs conferring resistance to tetracyclines reflected the most abundant resistance category across all faecal samples, whereas macrolide-lincosamide-streptogramin (MLS) and multidrug (that is, genes conferring resistance to two or more drug classes) resistance constituted the highest relative abundance across the sewage samples (Fig. 1c). The Mann–Whitney U test and a subsequent

https://doi.org/10.1038/s44221-024-00310-w

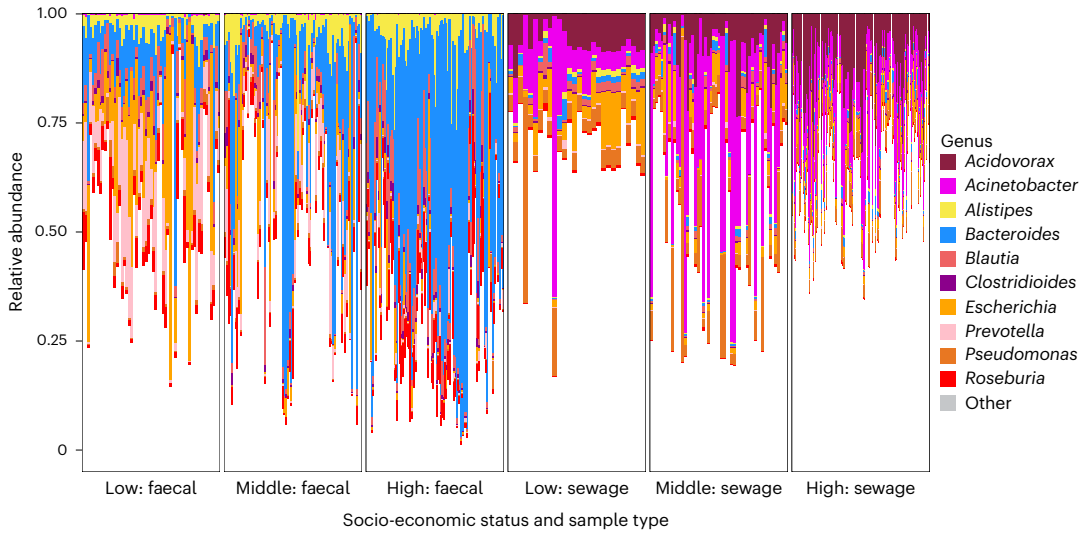

Fig. 2 | Microbiome composition across human faecal and sewage samples. Taxonomic composition of faecal and sewage samples at the genus level as a function of socio-economic status defined as high, middle and low according to socio-economic data collected from World Bank Database. Human faeces data represented 240 samples across 23 countries collected from publicly available data in NCBI. Sewage data represented 211 samples across 60 countries collected from publicly available data in NCBI.

Benjamini–Hochberg correction for the false discovery rate (FDR) were utilized to analyse the variations in the log transformed relative abundance of ARGs, with the significance level set at an adjusted *P* value of 0.05. In the faecal resistome, MLS resistance increased by two- and four-fold in samples in the high socio-economics bin as compared to samples in the middle and low socio-economic bins, respectively. Phenicol resistance was found to be 3.6-fold higher in samples binned in high socio-economic groups than in samples in low socio-economic bins, but there was no significant difference between samples in high and middle socio-economics groups (Supplementary Table 10). In sewage resistomes, the abundance of trimethoprim, phenicol and sulfonamide ARG classes were approximately five- and three-fold higher in samples within low socio-economics groups as compared to samples in high and middle socio-economics groups, respectively (Supplementary Table 10).

The dominant ARGs in human faecal samples belonged to the tetracycline (*tetQ*, *tetW*/*tet*(*W*/*N*/*W*), *tetO*, *tet37*, *tet32*, *tet40*), MLS (*macB*), multidrug (*efrB*), lincosamide (*llma*) and peptide (*ugd*, *bcrA*) classes (Fig. 1d). In sewage samples, genes conferring resistance to aminoglycoside (*ANT*(*3″*)-IIa_clust), beta-lactam (*OXA-256*_clust), MLS (*msrE*, *mphE*, *mel*, *macB*), multidrug (*mexK*, *adeJ*, *mdtB*), quaternary ammonium compounds (*qacH*) and sulfonamide (*sul1*) were most abundant (Fig. 1e). Three dominant ARGs common in both sewage and faecal samples were *macB* (MLS), *ugd* (Peptide) and *tetW*/*tet*(*W*/*N*/*W*) (tetracycline).

Taxonomic composition patterns by socio-economic status

A total of 676 genera were detected across all samples. The dominant genera (Fig. 2) in faecal samples were *Bacteroides, Blautia, Roseburia, Alistipes, Escherichia, Bifidobacterium, Bacillus, Enterococcus* and *Klebsiella*. In sewage samples, the dominant genera were *Acinetobacter, Acidovorax, Pseudomonas, Escherichia, Klebsiella, Enterobacter, Streptococcus* and *Bifidobacterium*.

When clustered by socio-economic status, *Prevotella* predominated in low socio-economic status faecal samples, whereas *Bacteriodes* predominated in high socio-economic status faecal samples. This observation is consistent with previous studies[44]. On comparing the sewage samples across socio-economic status classifications, no discerning pattern was observed, as no particular genus was more prevalent. Many prevalent genera in sewage samples; such as *Acidovorax, Acinetobacter, Chryseobacterium* and *Comamonas* are primarily representative of environmental bacteria[37,45,46]. Unsurprising, this suggests that sewage, along with containing human faecal bacteria, is comprised of other genera that reflect changes occurring in the sewage collection system.

Resistome diversity patterns by socio-economic status

To compare faecal and sewage resistomes, a similarity/dissimilarity analysis was performed via nonmetric multi-dimensional scaling (NMDS) ordination derived from Bray–Curtis dissimilarity matrices (Fig. 3a). Clear clustering was observed for faecal versus sewage samples (ANOSIM R Statistic 0.89, *P* value < 0.001). With regard to dissimilarities in the sewage resistomes, ANOSIM indicated that sewage samples were well separated when grouped by either continent ($R = 0.45$, *P* value < 0.001) or socio-economic status (ANOSIM R Statistic 0.47, *P* value < 0.001) (Supplementary Fig. 2a). In contrast, the separation for human faecal resistomes (Supplementary Fig. 2b) was not as strong as that observed for sewage samples when grouped by either continent (ANOSIM R Statistic 0.15, *P* value < 0.001) or socio-economic status (ANOSIM R Statistic 0.17, *P* value < 0.001). This observation suggests that although the differences are significant, there is relatively higher within group variation compared to between group variation in the case of faecal samples ($R = 0.17$) when compared with sewage samples ($R = 0.47$) across different socio-economic statuses.

Alpha diversity indices such as Simpson diversity index, Chao1 richness and Pielou's evenness were estimated to further compare resistomes (Fig. 3b and Supplementary Fig. 3). Mann–Whitney U test was applied to perform pairwise comparisons of samples grouped according to socio-economic status. It was noted that sewage samples had higher alpha diversity relative to faecal samples (*P* value < 0.05). Among faecal samples, no significant difference was observed in the diversity indices as a function of socio-economic status. Among sewage samples, a significant difference in alpha diversity was observed across the socio-economic status categories (*P* value < 0.05), with the low category having the highest overall alpha diversity. Higher variability in the ANOSIM statistic, which is derived from dissimilarity matrices, was noted among faecal samples within each socio-economic status

https://doi.org/10.1038/s44221-024-00310-w

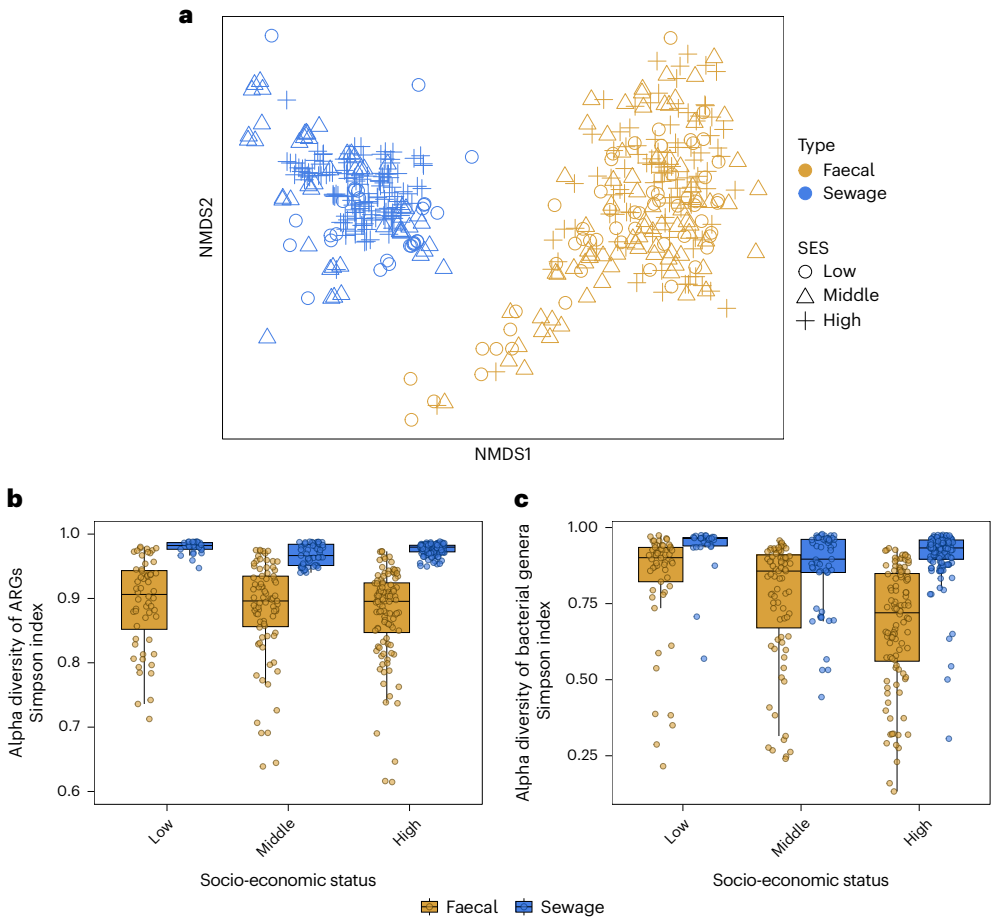

Fig. 3 | Resistome diversity patterns across socio-economic status categories. **a**, Beta diversity—non-metric multi-dimensional scaling analysis of resistome Bray–Curtis dissimilarity matrix as a function of socio-economic status (SES) and sample type. **b**, Alpha diversity—Simpson index of ARGs across faecal and sewage samples of different socio-economic status. **c**, Alpha diversity—Simpson index of genera across faecal and sewage samples of different socio-economic status. In box plots, each sample is represented by a dot with horizontal jitter for visibility. The horizontal box lines represent the first quartile, the median and the third quartile. Whiskers denote the range of points within the first quartile−1.5× the interquartile range and the third quartile + 1.5× the interquartile range. Socio-economic status defined as high, middle and low according to socio-economic data collected from World Bank Database. Human faeces data represent 240 samples categorized into 57 low, 79 middle and 104 high SES samples spanning across 23 countries, collected from publicly available data in NCBI. Sewage data represent 211 samples categorized into 28 low, 53 middle and 130 high SES samples spanning across 60 countries, collected from publicly available data in NCBI.

category than among sewage samples, which was consistent with sewage more strongly indicating socio-economic status than faeces.

Taxonomic diversity patterns by socio-economic status

Microbiome alpha diversity estimated at genus level was compared between sample types and across socio-economic status categories (Fig. 3c and Supplementary Fig. 4). On the basis of the Mann–Whitney U test, a significant difference was observed in the diversity indices (that is, Simpson, Chao1 richness, evenness) in faecal samples across the three socio-economic status categories. The highest diversity was observed in the Low category. For sewage samples, the Simpson diversity index and Chao1 richness median diversity were also highest in the low socio-economic status samples, with no significant difference between high and middle categories. There were no noted differences in richness across the different socio-economic categories.

Relationship between the taxonomy and resistome

The association between the taxonomy and resistome was evaluated using the Mantel test. A Pearson correlation coefficient was estimated between the Bray–Curtis dissimilarity matrices representing the taxonomy and resistomes. A significant correlation ($P < 0.001$) was observed between the ARG and genus-based taxonomy distance matrices for both faecal and sewage samples. The strength of the correlation trended stronger for sewage samples ($r = 0.67$) than for faecal samples ($r = 0.59$).

Core resistome analysis

The core resistome was operationally defined as those ARGs detected in ≥80% of the samples pertaining to a given category (Fig. 4). The core resistome of the human faecal samples constituted 73 ARGs (Supplementary Table 3). In contrast, 253 ARGs composed the core sewage resistome (Supplementary Table 3). There were 64 ARGs shared between the faecal and sewage core resistomes. Notably, a substantial portion of the faecal core resistome was found within the core sewage resistome, with the common ARGs reflecting multidrug (22), tetracycline (12), peptide (6) and glycopeptide (6) resistance. Other core ARGs conferring resistance to specific drug classes of interest were one aminocoumarin, one quinolone, one rifamycin, one trimethoprim, two beta-lactam, four aminoglycoside and five MLS (Supplementary Table 3).

The core resistomes were further compared across socio-economic categories. For the faecal resistome, 87, 78 and 108 ARGs were found as

https://doi.org/10.1038/s44221-024-00310-w

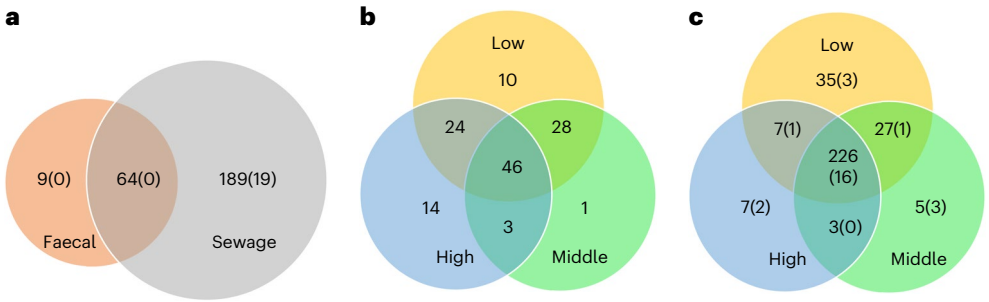

Fig. 4 | Core resistome analysis. a, The number of core ARGs identified in the human faecal and sewage samples. **b**, Number of core faecal ARGs identified across low, middle and high socio-economic status. **c**, Number of core sewage ARGs identified across low, middle and high socio-economic status. Parenthetical numbers reflect the number of clinically relevant genes. Socio-economic status defined as high, middle and low according to socio-economic data collected from World Bank Database. Human faeces data represent 240 samples categorized into 57 low, 79 middle and 104 high SES samples spanning across 23 countries, collected from publicly available data in NCBI. Sewage data represent 211 samples categorized into 28 low, 53 middle and 130 high SES samples spanning across 60 countries, collected from publicly available data in NCBI.

core ARGs across high, middle and low categories, respectively, with 46 ARGs found irrespective of socio-economic status (Fig. 4b). Across sewage resistomes, 243, 261 and 295 ARGs comprised the core across high, middle and low socio-economic categories, with 226 ARGs in common (>75%) (Fig. 4c).

The core resistome analysis was further refined to focus on clinically relevant ARGs. As noted by the bracketed numbers in Fig. 4, none of the ARGs in the core faecal resistomes were of clinical relevance. In contrast, 19 clinically relevant ARGs were found within the core sewage resistome. These 19 ARGs were *TEM-126*_clust, *CARB-5*_clust, *OXA-256*_clust, *qnrS6*_clust, *MCR-9.1*, *SHV-100*_clust, *OXA-46*_clust, *OXA-347*, *OXA-164*_clust, *OXA-280*_clust, *GES-21*_clust, *OXA-296*, *OXA-31*_clust, *OXA-37*/*OXA-20*, *OXA-226*_clust, *OXA-209*, *MCR-5.2*/*MCR-5.1*, *OXA-5*/*OXA-129* and *OXA-464*. Upon comparing clinically relevant ARGs found in the sewage core resistomes across socio-economic status, 16 out of 19 ARGs were found in all three categories. Interestingly, *TEM-126*_clust was consistently found only in the low and high categories, whereas *OXA-296* and *MCR-9.1* were only consistently found in high and middle categories, respectively.

Discriminatory resistome analysis

Resistomes were characterized using the ExtrARG machine learning-based algorithm[47] to identify ARGs that discriminated samples according to categories of interest. All discriminatory resistome analyses were performed on the rarified and normalized resistance count matrix.

The top 50 discriminatory ARGs for each sample type (that is, faecal, sewage) were visualized using a heat map (Fig. 5a). Among the discriminatory ARGs identified, one beta-lactam (*s_mupB*), one lincosamide (*llma*) and two tetracycline (*tetQ* and *tetW*/*tet*(*W*/*N*/*W*)) ARGs were abundant in the faecal samples, whereas the remaining discriminatory ARGs were abundant in the sewage samples. A substantial proportion of the discriminatory ARGs belonged to the multidrug class. Similar analyses were performed as a function of the socio-economic categories (Supplementary Figs. 5–7). It was observed that the peptide (*bcrA* (dominant in faecal samples) and *rosB* (dominant in sewage samples)) ARGs were only discriminatory in high socio-economic status samples (Supplementary Fig. 5). Several discriminatory clinically relevant ARGs were observed to occur at high abundances in the sewage resistomes. For example, beta-lactam (*GES-21*_clust, *OXA-226*_clust, *OXA-256*_clust) ARGs were commonly discriminatory across all sample type-based analyses (that is, comparing faecal and sewage resistomes within each socio-economic group; Supplementary Figs. 5–7). CARB-type (*CARB-3* and *CARB-5*) ARGs specifically discriminated low/middle socio-economic categories (Supplementary Figs. 6 and 7).

Discriminatory analysis was further performed on faecal and sewage samples grouped by socio-economic status (Fig. 5b,c). The major pattern observed among faecal samples was distinct clustering where high socio-economic samples clustered together and several middle/low samples clustered together. Further, aminoglycoside (*cblA*-1, *cfxA6*) and glycopeptide (*vanRA, vanRC, vanRI, vanSA*) ARGs were predominantly found to be associated with high socio-economic status, whereas several multidrug (*tolC*, *acrB, baeS, msbA, mdtB, mdtC, mdtM*) and peptide (*yojI, arnA, eptA, rosB*) ARGs were predominant in middle/low socio-economic status samples. Similar clustering of high versus middle/low was observed for the discriminatory analysis of the sewage samples. It was observed that many multidrug ARGs were predominant in high socio-economic status sewage samples. On the other hand, aminoglycoside, beta-lactam, phenicol, tetracycline, trimethoprim and sulfonamide ARGs were predominant in most of the middle/low category sewage samples. These predominant ARGs in middle/low category sewage samples also included three ARGs of clinical relevance (*CARB-3*_clust, *OXA-164*_clust and *OXA-256*_clust).

Finally, discriminatory analysis was performed on faecal samples grouped based on subject age (in years) into three categories: Group 1 (0-18), Group 2 (19-44), Group 3 (45 and above) (Supplementary Fig. 8) and sex (that is, female versus male) (Supplementary Fig. 9). There were no major patterns observed in the discriminatory analysis based on age or sex.

Socio-economic variables associated with ARG levels

To better identify variables that are likely drivers of the human faecal and sewage resistome patterns identified here, representative socio-economic indicators from the preprocessed World Bank data were used to construct six broad indices (climate, education, GDP, governance, health and infrastructure) (Supplementary Table 8b). Similar to Collignon et al.[36], these indices were estimated by normalizing (mean of 0 and standard deviation of 1) and averaging individual indicators that were representative of the broader index. A univariate analysis was framed to evaluate the association of these broad socio-economic indices with total ARG relative abundance levels in the faecal and sewage resistomes. To address cross-sample dependence, a variance component model was implemented, where the response variable was the total ARG relative abundance in each sample and the independent variables were the broad socio-economic indices corresponding to the country of sample origin.

For faecal samples, univariate analysis revealed that none of the socio-economic indicators individually had a significant impact on the total ARG relative abundance (Supplementary Table 8a). For sewage samples, governance, health and infrastructure were found to be the most significant factors associated with total ARG relative abundance (P value < 0.05).

https://doi.org/10.1038/s44221-024-00310-w

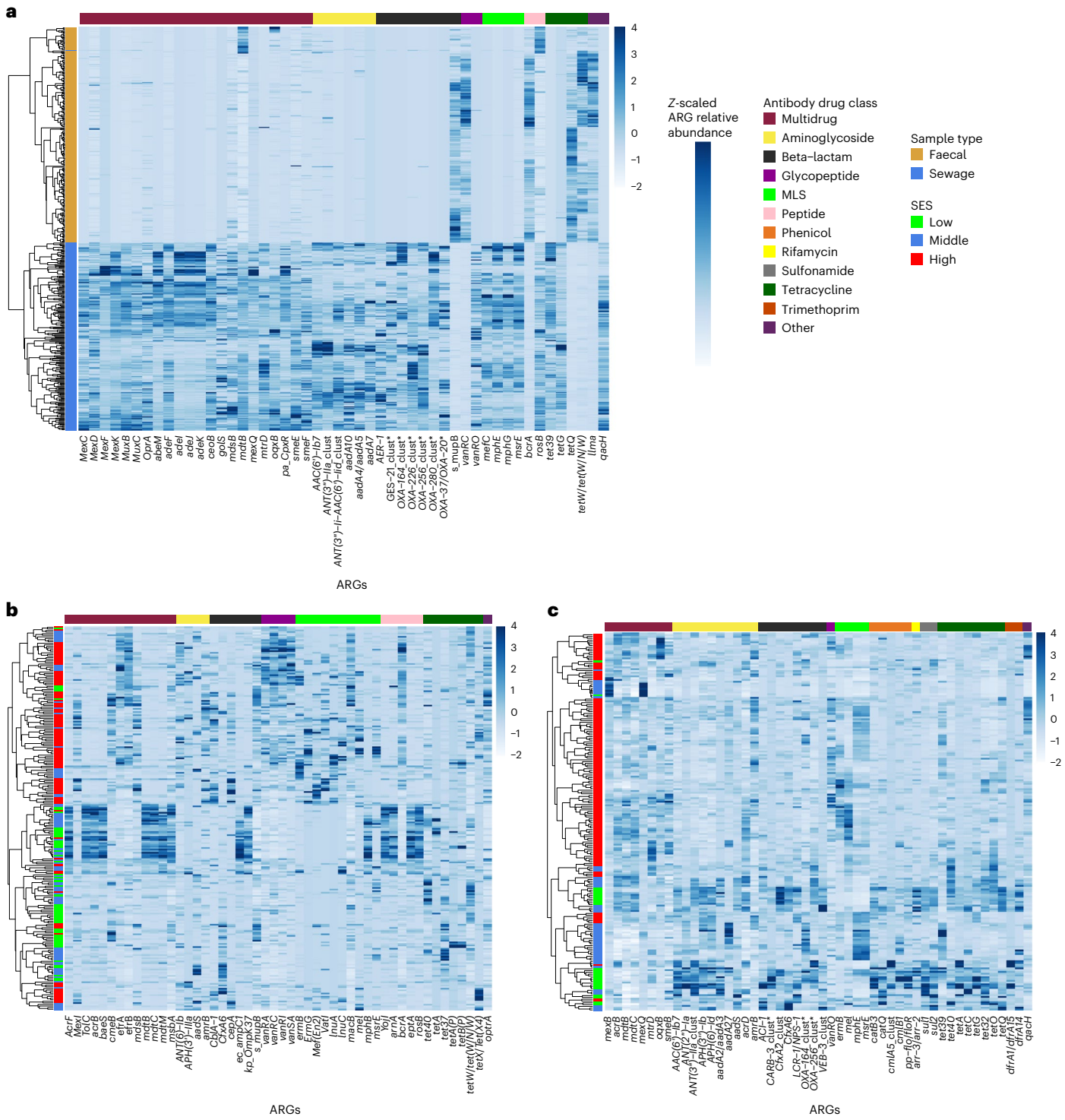

Fig. 5 | Discriminatory resistome analysis. a–c, Faecal and sewage resistomes (**a**), faecal resistomes categorized by socio-economic status (**b**) and sewage resistomes categorized by socio-economic status (**c**). Asterisks indicate ARGs classified as clinically relevant. Human faeces data represent 240 samples categorized into 57 low, 79 middle and 104 high SES samples spanning across 23 countries, collected from publicly available data in NCBI. Sewage data represent 211 samples categorized into 28 low, 53 middle and 130 high SES samples spanning across 60 countries, collected from publicly available data in NCBI.

Discussion

There is growing momentum towards the use of sewage as a key environmental surveillance point for antibiotic resistance[11,35,37,48] and this study addresses critical knowledge gaps towards realizing this aim. The extent to which sewage resistomes represent human faecal resistomes and which sample type best reflects the socio-economic factors thought to associate with the spread of antibiotic resistance, have remained open questions. This study sheds light on these issues by conducting a systematic assessment of various dimensions of microbiomes (for example, total ARG relative abundance, ARG diversity, core ARG analysis, discriminatory ARG analysis, taxonomic composition) in 451 publicly available metagenomic samples from 69 countries

(240 human faecal from 23 countries and 211 urban sewage samples from 60 countries) and assessing the association of socio-economic status/indicators with faecal and sewage resistomes using statistical analysis. It is imperative to note that the interplay between socio-economics and the dissemination of antibiotic resistance is highly complex and none of the statistical analyses are mature enough to infer causation. Nonetheless, these analyses do offer a path forward towards generating hypotheses and understanding various drivers that could contribute to the spread of antibiotic resistance. If strategically applied and globally coordinated, sewage surveillance could provide an early warning system to identify forms of resistance circulating in a community, inform optimal prescription practices, identify new resistance determinants of concern or assess the efficacy of community-scale interventions[4,11,48].

A clear outcome from this study is that sewage resistomes and microbiomes are largely distinct from human faecal resistomes and microbiomes. Whereas the validity of such a hypothesis has been previously shown in small-scale studies[24–26], this study makes an important contribution by using globally sourced metagenomic data to definitely demonstrate that they differ. Sewage resistomes clustered separately from human faecal resistomes based upon NMDS analysis and also exhibited higher relative abundances and higher total ARG diversity than human faeces. The dominant ARG classes were distinct in sewage (MLS/multidrug) relative to human faeces (tetracycline) and, unexpectedly, sewage was notably more highly enriched in clinically relevant ARGs. The dominant genera (*Bacteroides, Prevotella)* in faeces were also found to be in very low abundance in sewage samples. This is expected as these genera are anaerobes and therefore would not survive well in an aerobic sewage environment. This finding suggests that corresponding resistome signals could be lost from anaerobes during sewage conveyance. On top of the differences in ecological niches, sewage is often a composite of faeces from many individuals and other sources such as stormwater, hospital wastewater and partially treated industrial wastewater, which all contribute to the measured differences. Nonetheless, it was clear that sewage was highly influenced by human faeces, with 64 core ARGs and 466 core genera in common (Supplementary Tables 3 and 9). The genera that were found to be highly abundant in sewage were mostly environmental bacteria, thus suggesting that they either uniquely inhabit and are released from sewage biofilms and sediments or are introduced through inflow and infiltration into compromised pipes. The fact that most dimensions of the resistome, including clinically relevant ARGs, gave a stronger signal in the sewage suggests that the sewage collection environment may amplify antibiotic resistance (for example, through imposing selective pressure or facilitating gene exchange in biofilms[49]).

Remarkably, sewage was found to more strongly reflect socio-economic status and clinical resistance indicators than human faeces. In particular, in the NMDS analysis, sewage resistomes were more sharply separated by continent and socio-economic status. The diversity of the sewage resistomes, but not the human faecal resistomes, were differentiated by socio-economic status. To further investigate the role of specific socio-economic indicators, univariate analyses were performed to correlate broad socio-economic factors with total ARG relative abundance in the faecal and sewage resistomes. The analysis revealed that none of the socio-economic indicators individually were significantly associated with total ARG relative abundance in the faecal resistomes. This finding suggests that the influence of socio-economic factors is probably more intricate and collective in nature and extends beyond simple one-on-one associations. The observation that governance, health and infrastructure are associated with the sewage resistome was in line with the conclusions made in prior studies by Collignon et al.[36] and Hendriksen et al.[37]. It is remarkable that even after the differences in our analysis approaches, the same conclusions remain. This commonality underscores the potential importance of socio-economic drivers in dissemination of antibiotic resistance and thus the need to take such factors into account when developing intervention strategies (for example, water sanitation and hygiene) that target low and middle income countries[50,51]. Further, the sewage core resistome contained a wide array of ARGs classified as clinically relevant, whereas no core human faecal resistome ARGs were classified as such. These findings were contrary to the expectation that human faeces would more closely capture indicators of antibiotic resistance at its source (for example, human gut resistomes influenced by antibiotic treatment). On the other hand, human faecal sampling is much more logistically challenging, whereas also probably requiring much larger sample sizes to capture key trends, because of the large degree of individual variation. For example, there can be wide person-to-person variation in faecal resistomes as a function of multiple factors such as diet, lifestyle, past diseases and antibiotic courses[49,52], not just socio-economic status. This was apparent in the greater degree of variance encountered in the human faecal data, whereas the composite nature of the sewage samples reduced sample variance. In addition to the benefits of ease of sewage sampling relative to human faecal sampling pointed out by Aerustrup and Woolhouse[11], the findings of this study illustrate the statistical benefits of targeting sewage as a monitoring point.

Many measures of the sewage resistomes were elevated in countries of low socio-economic status. Relative abundances of total ARGs were highest in Africa and South America and more broadly in sewage from low socio-economic status countries. Sewage alpha and beta diversity were also highest in low socio-economic status locations. Trends revealed in the sewage resistome analysis were largely consistent with those recently reported by Hendriksen et al.[37] and Munk et al.[53] at the continent scale, where systemic differences in ARG abundances were noted between Africa/Asia/South America samples and Europe/North America/Oceania samples. These differences become more apparent upon categorizing samples according to socio-economic status. There were some minor deviations from the specific trends noted in Hendriksen et al.[37]. For example, we observed a difference in distribution of total ARG relative abundance at the specific continent level (for example, Africa and South America). These differences could be related to differences in data processing and annotation parameters in the two studies (for example, CARD versus ResFinder[54] as the annotation database). CARD contains both intrinsic and mobile resistance genes whereas ResFinder is curated to contain only mobile resistance genes. Further, the annotation parameters such as minimum length, e-value and Bitscore thresholds required to get a reliable hit were also different in the two studies and that could lead to some variation in the trends noted, even though the big picture conclusions remain the same. These comparisons further emphasize the need to standardize annotation parameters and pipelines as sewage surveillance expands to global scale[4,55,56].

Interestingly the trends in the faecal resistome differed somewhat from those observed for sewage. For example, the lowest total ARG relative abundance was found in African faecal samples. Low socio-economic status faecal samples (primarily samples from Africa) exhibited the lowest abundance of total ARGs. The highest total ARG relative abundances were found in the middle socio-economic status faecal samples, primarily samples from Asia. The low socio-economic status samples were largely from rural populations. Considering limited access to many antibiotics in low-income countries and wide polarity in drug regulation, it is possible that the observed lower ARG abundances in low socio-economic human faecal resistomes and higher ARG abundances in middle and high socio-economic status locations reflect such differences in usage[57]. However, due to the lack of credible antibiotic usage data for many countries[57], it is challenging to correlate the observed resistance levels with the antibiotic usage pattern and thus this requires further investigation. Upon comparing the geographic pattern in resistome composition between our study and a similar analysis conducted by Fuhrmeister et al.[58], congruent trends were observed across most regions except for Africa.

Fuhrmeister et al.[58] documented the highest total ARG relative abundance in Africa, whereas in our investigation, we observed lowest levels in African samples. This discrepancy may stem primarily from differences in sampling strategies: Fuhrmeister et al.[58] examined a vast dataset comprising 1,589 samples, whereas our study focused on a subsampled dataset of 240 samples. Notably, the observed imbalance in public datasets, featuring larger cohorts from specific countries or studies, prompted us to strive for a balanced dataset. Furthermore, the normalization and ARG annotation parameters also differed between the two studies.

Discriminatory resistome analysis delineated specific dissimilarities between faecal and sewage sample types and across samples grouped based on socio-economic status and could help to identify sewage monitoring indicators of specific interest. For example, a majority of discriminatory ARGs belonged to the multidrug class and were highly abundant in sewage. Clinically relevant ARGs such as *GES-21*_clust, *OXA-226*_clust, *OXA-256*_clust were highly abundant in all the sewage samples, whereas *CARB-3* and *CARB-5* were only abundant in middle/low socio-economic status samples. Interestingly, the overarching feature of the discriminatory analysis was that high socio-economic samples clustered together and several middle/low socio-economic status samples clustered together. This clustering was observed for both human faecal and sewage samples. This finding supports the contention that there could be common drivers of antibiotic resistance in low and middle socio-economic status countries.

Whereas the current study provides an in-depth view into the human faecal and sewage resistomes, it should be noted that there are additional factors and inherent limitations associated with the study design and with metagenomic analysis that could impact interpretation of the results.

First, we note that the socio-economic status analysis performed in this study was necessarily aggregated at the country level. It is thus reasonably likely that the socio-economic status for each specific sample, whether faecal or sewage, could differ from that at the country level. Ideally our study design would have entailed comparison of urban sewage samples to faecal samples from individuals residing in urban environments as a means to isolate the effect of urbanicity, as delineated by Fuhrmeister et al.[58]. However, scrutinization of the metadata for the samples from Hendriksen et al.[37] and the original publications utilized in this study (Supplementary Table 1), indicated that whereas the sewage metagenomes were derived from urban areas, the faecal sample cohort encompasses samples collected from individuals residing in both urban and rural areas. Particularly noteworthy was the fact that all faecal samples from low SES countries were collected from denizens of rural areas (Supplementary Table 1). The limited availability of urban faecal samples from low SES countries thus constrained our capacity to segregate urban from rural samples and perform direct comparisons. This difference could potentially explain the modest association between the relative abundance of antibiotic resistance in faecal resistomes as compared to that for sewage resistomes, which reflect the collective population contributing to the sewershed sample. Nonetheless, our study highlights the significant correlation between the total relative abundances of ARGs in faecal and sewage resistomes, thus underscoring the relevance of socio-economic factors. These observations are crucial from a One Health perspective as they highlight that there is no one single factor responsible for the shaping of resistomes. Our findings emphasize the need to identify the major factors that shape specific resistomes. Such an understanding could help attain a more holistic understanding that can frame precise intervention strategies to curb the spread of antibiotic resistance.

Second, some of the countries included in both the faecal and sewage datasets had a low number of samples available. Whereas comparing resistance profiles across socio-economic status groups yielded important findings, the limited sample size restricts the generalizability of the results. Accordingly, future sampling efforts should focus on addressing this gap to strengthen the study's findings, as suggested by Cai et al.[59]. Because the samples were retrieved from a range of previously published studies, there are inevitable variations resulting from factors such as DNA extraction protocol, sample processing steps, sequencing depth and so on. To address such concerns, samples were screened to maintain uniformity (for example, only healthy cohorts were chosen for faecal samples and only samples sequenced on Illumina platforms were used).

Third, multiple databases are available for ARG annotation (for example, CARD, Resfinder). In this study, CARD was selected as it is well-curated and extensively cited in the antibiotic resistance literature. However, no two databases are identical and none are exhaustive in coverage. When homology-based best hit annotation is applied, as it was in this study, the choice of stringency in annotation parameters (for example, e-value, bitscore, min length of amino acid, % identity) will correspondingly dictate the ratio of false positives and false negatives[60]. Here we selected annotation parameters consistent with previous studies[55,61], but we recognize that appropriate annotation parameters will vary depending on the study question. Further, it is important to note that there is a very high sequence similarity amongst many of the entries in the CARD database (in some cases more than 97%). To avoid this leading to an overestimation in sequence diversity, CARD database sequences were clustered at 90% similarity and the annotations were performed across the representative sequences. Finally, there are various inherent limitations to metagenomics analysis, for example, live/dead organisms cannot be directly distinguished, point mutations cannot be distinguished from sequencing error/misalignment, and there is limited capacity to identify host bacteria carrying ARGs because of limitations to short-read assembly. Moreover, metagenomics is not ideally suited to obtain quantitative absolute measurements. However, the study's primary objective was to establish qualitative baselines for antibiotic resistance determinants in faeces or sewage, which does not necessitate precise quantitative measurements.

Overall, the findings of this study support the development of sewage surveillance as a robust and sensitive means to identify drivers and trends in antibiotic resistance in local populations. To further validate the approach, it would be useful to sample faecal and sewage resistomes from the same network and to assess the capacity of various metadata, including specific consideration of variables reflective of socio-economic status, to predict resistome composition.

Methods

Metagenomic data sources

Metagenomic studies of faecal samples were searched in the previously published literature and 20 NCBI-SRA Bioprojects were identified wherein the sequencing data and corresponding metadata were publicly available[7,37,62–80]. The sewage samples, which were collected from urban areas connected to a centralized wastewater collection system, were obtained from a prior study by Hendriksen et al.[37]. From these, 275 human faecal samples from 23 countries and 234 urban sewage samples from 62 countries were retrieved from EMBL EBI (https://www.ebi.ac.uk/) and NCBI-SRA (https://www.ncbi.nlm.nih.gov/sra) public repositories (Supplementary Table 1). Inclusion criteria were imposed in the selection of samples to minimize the influence of extraneous factors. Only faecal samples from healthy individuals and datasets generated on Illumina shotgun sequencing platforms were selected. The average number of reads per human faecal sample and per sewage sample was 17 (0.5–66.5) and 20 (0.6–94.5) million reads, respectively. SRA accession and associated metadata about the metagenomes retrieved from the databases is included in Supplementary Table 1.

Socio-economic data collection and processing

Socio-economic data were extracted from five World Bank databases (World Development indicators; Health Nutrition and Population

statistics; Worldwide Governance; Poverty and Equity; Environmental Social and Governance data). Supplementary Table 4 reports the number of indicators that are documented in the corresponding database. Databases for 2011–2019 were downloaded to coincide with the collection dates of the samples from which metagenomic sequence data were derived.

The socio-economic data were preprocessed by removing indicators with more than 30% of missingness in the dataset and the remaining missing values were imputed using the missForest R package, a random forest-based imputation method consisting of both categorical and continuous predictors that is computationally efficient for high-dimensional data (more details in Supplementary Information). After imputation, the dataset was standardized (mean = 0 and variance = 1) and transformed using principal component analysis. Principal components (PCs) explaining 80% of the variance in the data were retained. The PCs were used to cluster the countries using K-means clustering. The number of clusters were determined using the elbow method (Supplementary Fig. 10a). Additionally, clusters were visually inspected for human validation. The obtained socio-economic clusters ($n = 3$) were designated 'high, middle and low', as indicated in Supplementary Fig. 10b. Each country was labelled according to its respective cluster and the samples were grouped in high, middle and low socio-economic bins according to the country label (Supplementary Table 2). This grouping information was used in performing socio-economics-based analyses.

ARG annotation database

A modified version of the Comprehensive Antibiotic Resistance Database (CARD) v.3.0.7 was applied in the annotation of ARGs, employing a similar strategy reported by Lee et al.[81]. Known global regulators were removed and the remaining ARGs were clustered at 90% global identity using CD-HIT[82]. The obtained non-redundant representative ARGs were renamed if multiple ARGs belonged to the same cluster to indicate the association of other similar ARGs. For example, if a cluster contained 10 ARGs and 'GeneA' was the representative ARG of that cluster. Then, the 'GeneA' was renamed as 'GeneA_clust'. In case, there are only two genes in a cluster (that is, 'GeneA' and 'GeneB') then the new name was formed by concatenating the two genes (that is, 'GeneA/GeneB'). All analyses were performed with the new naming representing non-redundant ARGs. The list and classification of ARGs in the modified CARD database are provided in Supplementary Table 5-6.

Metagenomic analysis and read annotation

Fastq files obtained from the public databases were processed through FastQC[83] to assess the quality of the metagenomic samples. Further, the samples were processed through Trimmomatic[84] to trim the low-quality reads and adaptors that could pose problems in downstream analysis. Host (human) DNA contamination was removed by aligning the reads to human reference genome (hg38) using Bowtie2[85]. The quality assessed samples were then merged using VSEARCH[86]. The merged files were annotated against the manually curated CARD (3.0.7) database. Due to the varied read lengths (75–150 bp) encountered in our dataset, we established a uniform filtering criterion applicable across all samples. DIAMOND[87] BlastX was used for mapping the reads with the set threshold of e-value < 10^{-10}, identity> 80% and minimum length of 17 amino acids[88–90]. The parameters were chosen to be less conservative to identify ARGs that might not span the entire read. It is acknowledged that this approach will probably increase the false-positive rate but served to keep the annotation parameters consistent as the collected samples had variable read length distribution (75–150 bp). In case of paired-end sequencing, the best hit was considered among the first and second reads hit. The obtained hits were normalized using Fragments Per Kilobase Million. Taxonomy was annotated using kraken2[91]. Clinically relevant ARGs were manually curated using the list in Supplementary Table 7. Clinically relevant ARGs were defined as conveying resistance to World Health Organization classified 'Reserve' antibiotics[92,93].

Statistical and diversity analysis

All tests were performed in R version 3.4.1 and Python[94]. Mann–Whitney U test, a non-parametric statistical test, was used to perform pairwise comparisons. NMDS plots were generated using the "Vegan" package (version 2.5-5) with the Bray–Curtis ordinations[95]. Analysis of similarity (ANOSIM) was used to determine differences in bacterial community beta diversity as well as when assessing resistome profiles. Microbial communities and resistome profiles were compared using Mantel test[96]. Diversity and richness were calculated on rarified versions of the resistance and bacterial count matrices. Count matrices were subsampled to the lowest samples' depth, using the Vegan rarefy function. The Simpson diversity index (1-D), Pielou's evenness, and the Chao1 richness estimates were calculated using the diversity function in the vegan package. The Simpson Diversity Index (1-D) is a measure of biodiversity that takes into account the number of species present as well as their relative abundance. Pielou's Evenness is a measure of how evenly the individuals in a community are distributed among the different species. Chao1 Richness is an estimator of the total number of species in a community based on the number of rare species observed.

Within-country representativeness

To assess whether samples from individual sites were representative of other sites in that country, we compared the Bray–Curtis dissimilarities for pairs of sites within the same country and in different countries for resistome compositions. The significance of these differences was assessed using permutation tests. The country labels for each sample were permuted and the dissimilarities were reassessed for pairs of sites within the same country and in different countries with permuted labels. This procedure was repeated 10^6 times to build up a null distribution of the differences in dissimilarity within and among countries. It was found that resistome dissimilarities were on average 19% higher for pairs of sites in different countries than for pairs of sites within the same country (permutation, $P < 0.0001$). These results demonstrated that there was less variance across samples collected within countries than samples collected from different countries. Thus, individual sites in this study were considered to be generally representative of a given country. It should be noted, however, that this does not imply that the variability within countries is negligible or even low.

Core resistome & discriminatory resistome analyses

The core resistome is defined as the collection of ARGs that are ubiquitous to a given environment. The core resistome was operationally defined as ARGs detected in at least 80% of the samples defined within a given category. The discriminatory resistome is the collection of ARGs that display differential abundance across different sample categories. ARGs that discriminate samples based on *a priori* selected groups were systematically identified using ExtrARG, a machine learning-based algorithm[47]. To reduce the bias that may arise in the discriminatory resistome analysis due to uneven sequencing depth among the samples, the resistance count matrix was first rarified to the lowest total ARG count among the included sample, and was then normalized to Fragments Per Kilobase Million scale.

Regression model to explore the link between ARG abundance and socio-economic factors

For regression analyses, independent variables were constructed by selecting representative socio-economic indicators from the preprocessed World Bank data to construct six broad indices (climate, education, GDP, governance, health and infrastructure). These indices were estimated by normalizing (mean of 0 and standard deviation of 1) and averaging individual indicators that were representative of the broader index (Supplementary Table 8b).

The variance component model (that is, random effects model) employed assumes that the response variable vector **Y**, representing the total ARG relative abundance in collected samples, follows a multivariate normal (MVN) distribution conditional on a matrix X, with its column representing the broad socio-economic indices,

$$\mathbf{Y}|X \sim MVN(\boldsymbol{\mu}, \Sigma),$$

where $\boldsymbol{\mu} = X\boldsymbol{\beta}$ and $\Sigma = \sigma_e^2 I + \sigma_a^2 J$. Here $\boldsymbol{\beta}$ is an unknown vector of fixed effects, I is the identity matrix, J is a block-diagonal matrix such that each diagonal block of 1s corresponds to one country, and the off diagonal blocks are 0 matrices. Correspondingly, σ_e^2 and σ_a^2 are two variance components to be estimated with the former representing variance due to random error at the sample level and the latter representing variance due to random effects at the country level. In the inference procedure, these two parameters σ_e^2 and σ_a^2 are estimated by the maximum likelihood method and estimation of the coefficient vector $\boldsymbol{\beta}$ is obtained by the generalized least square estimator

$$\hat{\boldsymbol{\beta}} = (X^T \hat{\Sigma}^{-1} X)^{-1} (X^T \hat{\Sigma}^{-1} \mathbf{Y}).$$

Identification of important indicators is then done by a test on the generalized regression t-statistic $\hat{\boldsymbol{\beta}}/\mathrm{SE}(\hat{\boldsymbol{\beta}})$, where SE is the standard error.

Data availability

All data sources are available in the main text or the supplementary materials.

References

1. *Integrated Surveillance of Antimicrobial Resistance in Foodborne Bacteria: Application of a One Health Approach: Guidance from the WHO Advisory Group on Integrated Surveillance of Antimicrobial Resistance (AGISAR)* (WHO, 2017).
2. *Global Antimicrobial Resistance Surveillance System (GLASS) Report: Early Implementation 2020* (WHO, 2020).
3. *UK Five Year Antimicrobial Resistance Strategy 2013 to 2018* (Department of Health and Social Care, 2013).
4. Pruden, A., Vikesland, P. J., Davis, B. C. & de Roda Husman, A. M. Seizing the moment: now is the time for integrated global surveillance of antimicrobial resistance in wastewater environments. *Curr. Opin. Microbiol.* **64**, 91–99 (2021).
5. National Academies of Sciences, Engineering & Medicine *Combating Antimicrobial Resistance and Protecting the Miracle of Modern Medicine* (National Academies Press, 2022).
6. Forslund, K., Sunagawa, S., Coelho, L. P. & Bork, P. Metagenomic insights into the human gut resistome and the forces that shape it. *Bioessays* **36**, 316–329 (2014).
7. Rampelli, S. et al. Metagenome sequencing of the Hadza hunter–gatherer gut microbiota. *Curr. Biol.* **25**, 1682–1693 (2015).
8. Van Schaik, W. The human gut resistome. *Philos. Trans. R. Soc. B* **370**, 20140087 (2015).
9. Yang, Z. et al. Preliminary analysis showed country-specific gut resistome based on 1267 feces samples. *Gene* **581**, 178–182 (2016).
10. McGuire, A. L. et al. Ethical, legal, and social considerations in conducting the Human Microbiome Project. *Genome Res.* **18**, 1861–1864 (2008).
11. Aarestrup, F. M. & Woolhouse, M. E. J. Using sewage for surveillance of antimicrobial resistance. *Science* **367**, 630–632 (2020).
12. Li, W. et al. Population-based variations of a core resistome revealed by urban sewage metagenome surveillance. *Environ. Int.* **163**, 107185 (2022).
13. Karkman, A., Pärnänen, K. & Larsson, D. G. J. Fecal pollution can explain antibiotic resistance gene abundances in anthropogenically impacted environments. *Nat. Commun.* **10**, 80 (2019).
14. Koike, S. et al. Monitoring and source tracking of tetracycline resistance genes in lagoons and groundwater adjacent to swine production facilities over a 3-year period. *Appl. Environ. Microbiol.* **73**, 4813–4823 (2007).
15. Munir, M., Wong, K. & Xagoraraki, I. Release of antibiotic resistant bacteria and genes in the effluent and biosolids of five wastewater utilities in Michigan. *Water Res.* **45**, 681–693 (2011).
16. Duarte, A. S. R. et al. Addressing learning needs on the use of metagenomics in antimicrobial resistance surveillance. *Front. Public Health* **8**, 38 (2020).
17. Hendriksen, R. S. et al. Using genomics to track global antimicrobial resistance. *Front. Public Health* **7**, 242 (2019).
18. Pust, M.-M. & Tümmler, B. Identification of core and rare species in metagenome samples based on shotgun metagenomic sequencing, Fourier transforms and spectral comparisons. *ISME Commun.* **1**, 2 (2021).
19. Wright, G. D. Antibiotic resistance in the environment: a link to the clinic? *Curr. Opin. Microbiol.* **13**, 589–594 (2010).
20. Karkman, A., Do, T. T., Walsh, F. & Virta, M. P. J. Antibiotic-resistance genes in waste water. *Trends Microbiol.* **26**, 220–228 (2018).
21. Bengtsson-Palme, J. & Larsson, D. G. J. Concentrations of antibiotics predicted to select for resistant bacteria: proposed limits for environmental regulation. *Environ. Int.* **86**, 140–149 (2016).
22. Kraemer, S. A., Ramachandran, A. & Perron, G. G. Antibiotic pollution in the environment: from microbial ecology to public policy. *Microorganisms* **7**, 180 (2019).
23. Bellali, S., Lagier, J.-C., Raoult, D. & Bou Khalil, J. Among live and dead bacteria, the optimization of sample collection and processing remains essential in recovering gut microbiota components. *Front. Microbiol.* **10**, 1606 (2019).
24. Newton, R. J. et al. Sewage reflects the microbiomes of human populations. *MBio* **6**, e02574–02514 (2015).
25. Pal, C., Bengtsson-Palme, J., Kristiansson, E. & Larsson, D. G. The structure and diversity of human, animal and environmental resistomes. *Microbiome* **4**, 54 (2016).
26. Su, J.-Q. et al. Metagenomics of urban sewage identifies an extensively shared antibiotic resistome in China. *Microbiome* **5**, 84 (2017).
27. Vikesland, P. J. et al. Toward a comprehensive strategy to mitigate dissemination of environmental sources of antibiotic resistance. *Environ. Sci. Technol.* **51**, 13061–13069 (2017).
28. Bengtsson-Palme, J., Kristiansson, E. & Larsson, D. G. J. Environmental factors influencing the development and spread of antibiotic resistance. *FEMS Microbiol. Rev.* **42**, fux053 (2018).
29. Ju, F. et al. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. *ISME J.* **13**, 346–360 (2019).
30. Yin, X. et al. Exploration of the antibiotic resistome in a wastewater treatment plant by a nine-year longitudinal metagenomic study. *Environ. Int.* **133**, 105270 (2019).
31. Kumarasamy, K. K. et al. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. *Lancet Infect. Dis.* **10**, 597–602 (2010).
32. Wang, Y. et al. Comprehensive resistome analysis reveals the prevalence of NDM and MCR-1 in Chinese poultry production. *Nat. Microbiol.* **2**, 16260 (2017).
33. Laxminarayan, R., Sridhar, D., Blaser, M., Wang, M. & Woolhouse, M. Achieving global targets for antimicrobial resistance. *Science* **353**, 874–875 (2016).
34. Collignon, P. Antibiotic resistance: are we all doomed? *Intern. Med. J.* **45**, 1109–1115 (2015).

35. Riquelme, M. V. et al. Demonstrating a comprehensive wastewater-based surveillance approach that differentiates globally sourced resistomes. *Environ. Sci. Technol.* **56**, 14982–14993 (2022).
36. Collignon, P., Beggs, J. J., Walsh, T. R., Gandra, S. & Laxminarayan, R. Anthropological and socio-economic factors contributing to global antimicrobial resistance: a univariate and multivariable analysis. *Lancet Planet. Health* **2**, e398–e405 (2018).
37. Hendriksen, R. S. et al. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nat. Commun.* **10**, 1124 (2019).
38. Bowyer, R. C. E. et al. Socioeconomic status and the gut microbiome: a TwinsUK cohort study. *Microorganisms* **7**, 17 (2019).
39. He, Y. et al. Linking gut microbiota, metabolic syndrome and economic status based on a population-level analysis. *Microbiome* **6**, 172 (2018).
40. Makuta, I. & O'Hare, B. Quality of governance, public spending on health and health status in Sub Saharan Africa: a panel data regression analysis. *BMC Public Health* **15**, 932 (2015).
41. Alividza, V. et al. Investigating the impact of poverty on colonization and infection with drug-resistant organisms in humans: a systematic review. *Infect. Dis. Poverty* **7**, 76 (2018).
42. Okeke, I. N., Lamikanra, A. & Edelman, R. Socio-economic and behavioral factors leading to acquired bacterial resistance to antibiotics in developing countries. *Emerging Infect. Dis.* **5**, 18–27 (1999).
43. Bich, V. T. N. et al. An exploration of the gut and environmental resistome in a community in northern Vietnam in relation to antibiotic use. *Antimicrob. Resist. Infect. Control* **8**, 194 (2019).
44. Wu, G. D. et al. Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108 (2011).
45. O'Rourke, D. P. & Rosenbaum, M. D. in *Laboratory Animal Medicine* 3rd edn (eds Fox, J. G. et al.) 931–965 (Academic Press, 2015).
46. Nakipoglu, Y. et al. Evaluation of the contaminant organisms of humidifier reservoir water and investigation of the source of contamination in a university hospital in Turkey. *Am. J. Infect. Control* **33**, 62–63 (2005).
47. Gupta, S., Arango-Argoty, G., Zhang, L., Pruden, A. & Vikesland, P. Identification of discriminatory antibiotic resistance genes among environmental resistomes using extremely randomized tree algorithm. *Microbiome* **7**, 123 (2019).
48. Huijbers, P. M. C., Flach, C.-F. & Larsson, D. G. J. A conceptual framework for the environmental surveillance of antibiotics and antibiotic resistance. *Environ. Int.* **130**, 104880 (2019).
49. da Silva, S. F. et al. Influence of human eating habits on antimicrobial resistance phenomenon: aspects of clinical resistome of gut microbiota in omnivores, ovolactovegetarians, and strict vegetarians. *Antibiotics* **10**, 276 (2021).
50. Macintyre, A., Wilson-Jones, M. & Velleman, Y. Prevention first: tackling AMR through water, sanitation and hygiene. *AMR Control* http://resistancecontrol.info/2017/prevention-first-tackling-amr-through-water-sanitation-and-hygiene/ (2017).
51. Wuijts, S. et al. Towards a research agenda for water, sanitation and antimicrobial resistance. *J. Water Health* **15**, 175–184 (2017).
52. Sun, J. et al. Environmental remodeling of human gut microbiota and antibiotic resistome in livestock farms. *Nat. Commun.* **11**, 1427 (2020).
53. Munk, P. et al. Genomic analysis of sewage from 101 countries reveals global landscape of antimicrobial resistance. *Nat. Commun.* **13**, 7251 (2022).
54. Zankari, E. et al. Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **67**, 2640–2644 (2012).
55. Bengtsson-Palme, J., Larsson, D. G. J. & Kristiansson, E. Using metagenomics to investigate human and environmental resistomes. *J. Antimicrob. Chemother.* **72**, 2690–2703 (2017).
56. Angers-Loustau, A. et al. The challenges of designing a benchmark strategy for bioinformatics pipelines in the identification of antimicrobial resistance determinants using next generation sequencing technologies. *F1000Res.* https://doi.org/10.12688/f1000research.14509.2 (2018).
57. *WHO Report on Surveillance of Antibiotic Consumption: 2016–2018 Early Implementation* (WHO, 2018).
58. Fuhrmeister, E. R. et al. Evaluating the relationship between community water and sanitation access and the global burden of antibiotic resistance: an ecological study. *Lancet Microbe* **4**, e591–e600 (2023).
59. Chen, C. et al. Characterising global antimicrobial resistance research explains why One Health solutions are slow in development: an application of AI-based gap analysis. *Environ. Int.* **187**, 108680 (2024).
60. Garner, E. et al. Next generation sequencing approaches to evaluate water and wastewater quality. *Water Res.* **194**, 116907 (2021).
61. Keenum, I. et al. Combined effects of composting and antibiotic administration on cattle manure-borne antibiotic resistance genes. *Microbiome* **9**, 81 (2021).
62. Costea, P. I. et al. Subspecies in the global human gut microbiome. *Molecular Syst. Biol.* **13**, 960 (2017).
63. Dhakan, D. B. et al. The unique composition of Indian gut microbiome, gene catalogue, and associated fecal metabolome deciphered using multi-omics approaches. *Gigascience* **8**, giz004 (2019).
64. Karlsson, F. H. et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* **498**, 99–103 (2013).
65. Kristoffer, F. et al. Disentangling the effects of type 2 diabetes and metformin on the human gut microbiota. *Nature* **528**, 262–266 (2015).
66. Le Chatelier, E. et al. Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546 (2013).
67. Lim, M. Y. et al. Stability of gut enterotypes in Korean monozygotic twins and their association with biomarkers and diet. *Sci. Rep.* **4**, 7348 (2014).
68. Liu, W. et al. Unique features of ethnic Mongolian gut microbiome revealed by metagenomic analysis. *Sci. Rep.* **6**, 34826 (2016).
69. Lloyd-Price, J. et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* **550**, 61–66 (2017).
70. Lokmer, A. et al. Use of shotgun metagenomics for the identification of protozoa in the gut microbiota of healthy individuals from worldwide populations with various industrialization levels. *PLoS ONE* **14**, e0211139 (2019).
71. Obregon-Tito, A. J. et al. Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat. Commun.* **6**, 6505 (2015).
72. Pasolli, E. et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* **176**, 649–662 (2019).
73. Pehrsson, E. C. et al. Interconnected microbiomes and resistomes in low-income human habitats. *Nature* **533**, 212–216 (2016).
74. Qin, J. et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60 (2012).
75. Tisza, M. J. & Buck, C. B. A catalog of tens of thousands of viruses from human metagenomes reveals hidden associations with chronic diseases. *Proc. Natl Acad. Sci. USA* **118**, e2023202118 (2021).
76. Tyakht, A. V. et al. Human gut microbiota community structures in urban and rural populations in Russia. *Nat. Commun.* **4**, 2469 (2013).
77. Vatanen, T. et al. Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. *Cell* **165**, 842–853 (2016).

78. Voigt, A. Y. et al. Temporal and technical variability of human gut metagenomes. *Genome Biol.* **16**, 73 (2015).
79. Yatsunenko, T. et al. Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).
80. Zeller, G. et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol. Syst. Biol.* **10**, 766 (2014).
81. Lee, K. et al. Mobile resistome of human gut and pathogen drives anthropogenic bloom of antibiotic resistance. *Microbiome* **8**, 2 (2020).
82. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
83. Andrews, S. FastQC (Babraham Bioinformatics, Babraham Institute, 2010).
84. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
85. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
86. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
87. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
88. Arango-Argoty, G. et al. MetaStorm: a public resource for customizable Metagenomics annotation. *PLoS ONE* **11**, e0162442 (2016).
89. Keenum, I. et al. Metagenomic tracking of antibiotic resistance genes through a pre-harvest vegetable production system: an integrated lab-, microcosm- and greenhouse-scale analysis. *Environ. Microbiol.* **24**, 3705–3721 (2022).
90. Yin, X. et al. ARGs-OAP v2. 0 with an expanded SARG database and Hidden Markov Models for enhancement characterization and quantification of antibiotic resistance genes in environmental metagenomes. *Bioinformatics* **34**, 2263–2270 (2018).
91. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
92. Majeed, H. J. et al. Evaluation of metagenomic-enabled antibiotic resistance surveillance at a conventional wastewater treatment plant. *Front. Microbiol.* **12**, 657954 (2021).
93. *The 2019 WHO AWaRe Classification of Antibiotics for Evaluation and Monitoring of Use* (WHO, 2019).
94. R Core Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2013).
95. Beals, E. W. Bray-Curtis Ordination: An Effective Strategy for Analysis of Multivariate Ecological Data. *Adv. Ecol. Res.* **14**, 1–55 (1984).
96. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220 (1967).

Acknowledgements

This study was supported by National Science Foundation (NSF) award OISE (number 1545756; P.V., A.P., L.Z.), NSF award ECCSS NNCI (number 1542100; P.V., A.P.), NSF award CSSI (number 2004751; L.Z., A.P., P.V.), NSF award NRT (number 2125798; L.Z., A.P., P.V.), Center for Science and Engineering of the Exposome at the Virginia Tech Institute for Critical Technology and Applied Science (ICTAS; P.V., A.P.), Virginia Tech Graduate School supported Genetics, Bioinformatics and Computational Biology (GBCB) and Sustainable Nanotechnology (VTSuN) programmes (S.G., L.Z., A.P., P.V.). We also thank the Advanced Research Computing (ARC) Center and the Fralin Life Sciences Institute of Virginia Tech for technical support.

Author contributions

S.G.: formal analysis, writing—original draft preparation. S.G., X.W., A.P., L.Z., P.V.: conceptualization. S.G., X.W., A.P., L.Z., P.V.: methodology, writing—review and editing. A.P., L.Z., P.V.: resources, project administration, supervision. S.G., A.P., L.Z., P.V.: writing–revising and editing. All authors contributed to the article and approved the final paper.

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/s44221-024-00310-w.

Correspondence and requests for materials should be addressed to Peter Vikesland.

Peer review information *Nature Water* thanks the anonymous reviewers for their contribution to the peer review of this work.