Microbiome

RESEARCH

Open Access

Regional antimicrobial resistance gene flow among the One Health sectors in China

Yuqing Feng[1†], Xin Lu[2*†], Jiayong Zhao[3†], Hongmin Li[4†], Jialiang Xu[5†], Zhenpeng Li[2], Mengyu Wang[2,6], Yao Peng[2], Tian Tian[4], Gailing Yuan[4], Yuan Zhang[2], Jiaqi Liu[2,5], Meihong Zhang[1], A La Teng Zhu La[1], Geruo Qu[2,5], Yujiao Mu[3], Wanshen Guo[3], Yongning Wu[7], Yuyu Zhang[5*], Dexiang Wang[4*], Yongfei Hu[1*] and Biao Kan[2,6*]

Abstract

Background Antimicrobial resistance poses a significant threat to global health, with its spread intricately linked across human, animal, and environmental sectors. Revealing the antimicrobial resistance gene (ARG) flow among the One Health sectors is essential for better control of antimicrobial resistance.

Results In this study, we investigated regional ARG transmission among humans, food, and the environment in Dengfeng, Henan Province, China by combining large-scale metagenomic sequencing with culturing of resistant bacterial isolates in 592 samples. A total of 40 ARG types and 743 ARG subtypes were identified, with a predominance of multidrug resistance genes. Compared with microbes from human fecal samples, those from food and environmental samples showed a significantly higher load of ARGs. We revealed that dietary habits and occupational exposure significantly affect ARG abundance. *Pseudomonadota*, particularly *Enterobacteriaceae*, were identified as the main ARG carriers shaping the resistome. The resistome in food samples was found more affected by mobile genetic elements (MGEs), whereas in environmental samples, it was more associated with the microbial composition. We evidenced that horizontal gene transfer (HGT) mediated by plasmids and phages, together with strain transmission, particularly those associated with the *Enterobacteriaceae* members, drive regional ARG flow. Lifestyle, dietary habits, and occupational exposure are all correlated with ARG dissemination and flies and food are important potential sources of ARGs to humans. The widespread mobile carbapenemase gene, *OXA-347*, carried by non-*Enterobacteriaceae* bacteria in the human gut microbiota, requires particular attention. Finally, we showed that machine learning models based on microbiome profiles were effective in predicting the presence of carbapenem-resistant strains, suggesting a valuable approach for AMR surveillance.

†Yuqing Feng, Xin Lu, Jiayong Zhao, Hongmin Li, and Jialiang Xu contributed equally to this work.

*Correspondence:
Xin Lu
luxin@icdc.cn
Yuyu Zhang
zhangyuyu@btbu.edu.cn
Dexiang Wang
17513332866@163.com
Yongfei Hu
huyongfei@cau.edu.cn
Biao Kan
kanbiao@icdc.cn
Full list of author information is available at the end of the article

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/.

Conclusions Our study provides a full picture of regional ARG transmission among the One Health sectors in a county-level city in China, which facilitates a better understanding of the complex routes of ARG transmission and highlights new points of focus for AMR surveillance and control.

Keywords One Health, Antimicrobial resistance gene, Mobile genetic element, Transmission pathway, Carbapenemase

Background

The increase in antimicrobial resistance (AMR) poses a significant challenge to healthcare systems, raising concerns regarding global public health, and food safety [1, 2]. The dissemination and spread of AMR is complex, involving antimicrobial resistance gene (ARG) flow across all three One Health sectors (animals, humans, and environments). Understanding the emergence and transmission of ARGs and/or antimicrobial-resistant bacteria (ARB) among these sectors as well as their interfaces such as food and water, is a prerequisite for better control of AMR.

Through culture-dependent and -independent approaches, specific ARG dissemination pathways especially those from animals to humans have been proposed. For example, plasmid-mediated colistin resistance was first identified mainly in isolates from animals and then was found widely spread across animals, humans, and environments [3]. The dissemination of AMR among the One Health sectors can be largely attributed to the mobility of ARGs carried by mobile genetic elements (MGEs), which can then be vertically propagated through clonal spread. This process has occurred for many clinically important ARGs, including the carbapenemase gene $bla_{\text{NDM-1}}$ [4], the colistin resistance gene *mcr-1* [5], and the tigecycline resistance genes *tet*(X3) and *tet*(X4) [6, 7]. Diverse MGEs, such as plasmid and foodborne pathogens (especially members of the *Enterobacteriaceae*), are frequently evidenced to be involved in the dissemination of these ARGs [8].

With the development of high-throughput sequencing, genomics, and metagenomics have been increasingly incorporated into One Health studies, enabling investigation of the abundance, transmission, and distribution of ARGs across different sources [9, 10]. Through whole-genome sequencing analyses of *Escherichia coli* isolates, the Australian silver gull was found to host different carbapenemase genes that may be anthropogenic sources [11], while the KPC-2-producing *Serratia marcescens* clone found in farm animals could potentially contribute to the clinical isolates present in nosocomial settings (i.e., healthcare environments where infections are acquired during hospital stays) [12]. Recently, the global distribution of ARGs was profiled by collecting thousands of metagenomic sequencing datasets from six types of habitats, and nearly 24% of the detected ARGs were concluded to pose a health risk to humans [13]. Although these findings enhance our understanding of AMR dynamics, most of the studies were based on large areas and/or time scales, which may not reveal the ecology of AMR at high resolution. Additionally, relatively fewer AMR studies have focused on the dissemination and spread of AMR at One Health interfaces in geographically proximate ecosystems. Assessment of the risk of AMR acquisition in humans, as a result of human behaviors such as dietary habits and occupational exposure, is continually needed.

In this study, we investigated regional AMR transmission in a Chinese city by analyzing the metagenomic sequencing data and the genomes of carbapenem-resistant isolates collected from humans, food, and the environment. We identified key MGEs and bacteria in mediating ARG transmission and proposed potential new transmission pathways for ARG transmission from food and the environment to humans.

Results

Habitat-specific profiles of the antibiotic resistome

We collected 592 samples from nine different human subgroups, three food subgroups, and six environmental subgroups in Dengfeng, Henan Province, China for antibiotic resistome and ARG flow analyses (Fig. 1a and Additional file 2: Table S1). We identified 40 ARG types and 743 ARG subtypes. The most abundant ARG types identified were multidrug resistance genes (151 ARG subtypes; accounting for 27.5% of the total ARG abundance), followed by the macrolide-lincosamide-streptogramin (MLS; 90 subtypes; 24.6%), tetracycline (52 subtypes; 14.2%), aminoglycoside (78 subtypes; 7.7%), and beta-lactam resistance genes (138 subtypes; 6.3%) (Fig. 1b and Additional file 2: Table S2). Samples of different origins (human, food, and the environment) displayed different resistome profiles (Additional file 1: Fig. S1a and Additional file 2: Table S3). The most obvious difference was that compared with human samples, food (pork, chicken, and vegetable/fruit), soil, surface water, and fly microbes showed a very high load of multidrug resistance genes (from 1.460 to 5.236 ARG copies per 16S rRNA gene copy). Interestingly, microbes

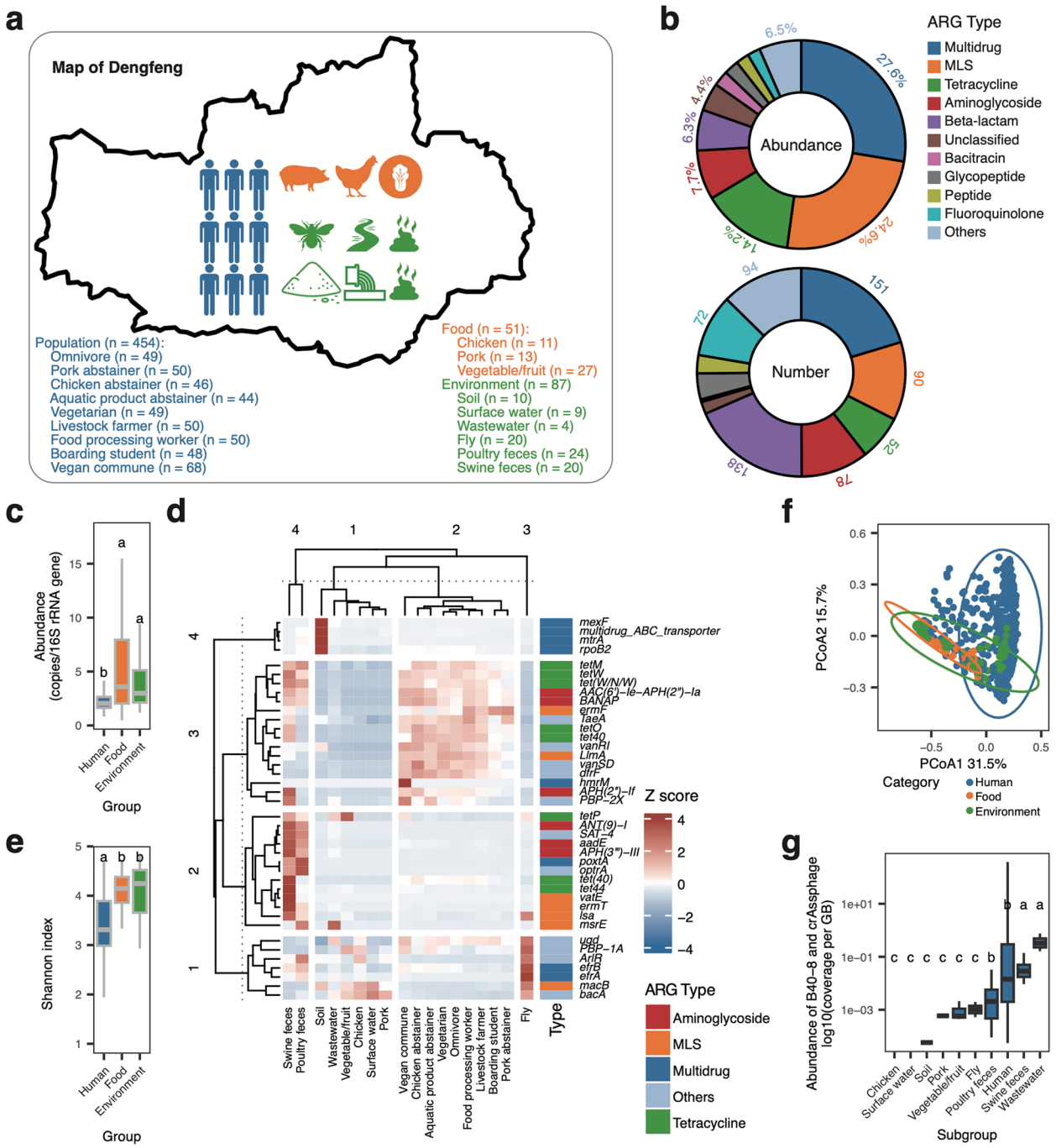

Fig. 1 The resistome profiles of the 592 samples. **a** Design of the sample collection in Dengfeng City, Henan Province, China, involving three groups, 18 subgroups, and 592 samples. **b** The proportion of antimicrobial resistance gene (ARG) types (top panel) and the number of ARG subtypes in each ARG type (bottom panel). **c** The ARG abundance in the three groups. The abundance of ARGs was transformed as a "copy of ARG per copy of 16S rRNA gene". **d** Biomarker ARGs in samples from various subgroups, were identified using the extremely randomized tree algorithm. Heatmap visualizes the *Z* score distribution according to the abundance of each ARG. Clustering was performed using K-means clustering via the R package "ComplexHeatmap". The right annotation shows the type of each ARG. **e** The Shannon indices of ARGs in the three groups. **f** Principal coordinate analysis (PCoA) of the resistome profiles for each sample in the three groups (PERMANOVA, $P < 2.2 \times 10^{-16}$, and $R^2 = 10.9\%$). **g** The abundance of the bacteriophages B40-8 and crAssphage in samples from various subgroups (measured in terms of coverage per Gb)

from food and the environment contained a significantly higher load of ARGs than those from human feces (Kruskal–Wallis test, $P = 2.9 \times 10^{-14}$; Fig. 1c), especially for multidrug ($P < 2.2 \times 10^{-16}$), aminoglycoside ($P = 7.0 \times 10^{-16}$), unclassified ($P < 2.2 \times 10^{-16}$), and bacitracin ($P < 2.2 \times 10^{-16}$) resistance genes (Additional file 1: Fig. S1b and Additional file 2: Table S4).

In humans, dietary habits and occupational exposure were found to affect ARG abundance, especially for the top 20 most abundant ARG types (Kruskal–Wallis tests, $P < 0.05$; 90.0%, 18/20; Additional file 1: Fig. S1c and Additional file 2: Table S5). For example, samples from boarding students showed a higher load of beta-lactam resistance genes; samples from pork abstainers exhibited a lower load of glycopeptide resistance genes; samples from livestock farmers carried a higher load of phenicol resistance genes. Specific aminoglycoside and tetracycline resistance genes were detected at higher levels in the human fecal samples, including *tetM*, *tetW*, *tet(W/NW)*, *tetO*, *tet40*, *AAC(6')-Ie-APH(2")-Ia*, *BANAP*, and *APH(2")-If* (Fig. 1d).

Diversity analyses indicated that ARG alpha diversity, as measured by the Shannon index, was lower in human fecal samples than in samples from other groups (Kruskal–Wallis test, $P < 2.2 \times 10^{-16}$; Fig. 1e), while at the subgroup level, samples from poultry feces, flies, and wastewater had higher Shannon indices (Kruskal–Wallis test, $P < 2.2 \times 10^{-16}$; Additional file 1: Fig. S2a); among the human samples, compared with omnivores, vegan communes exhibited higher ARG alpha diversity ($P = 1.9 \times 10^{-4}$). Beta diversity analysis revealed the clustering of samples from human, food, and environmental sources into three distinct groups (adonis, $P < 0.001$, $R^2 = 36.4\%$; Fig. 1f and Additional file 2: Table S6), highlighting the influence of habitat on the ARG composition. Principal coordinate analysis (PCoA) revealed significant differences in ARG subtypes according to the Axis1 values between the human and animal fecal samples and the other samples (Kruskal–Wallis test, $P < 2.2 \times 10^{-16}$; Additional file 1: Fig. S2b), with the exception of wastewater samples. We then compared the Bray–Curtis distances between the resistome profiles of the human fecal samples and other samples. Our PCoA analyses were supported by the comparison of Bray–Curtis distances, which indicated human fecal samples were more similar to animal feces and wastewater samples compared to other samples (Additional file 1: Fig. S2c). A further analysis of the abundance of the *Bacteroides* bacteriophages B40-8 and crAssphage, indicators of fecal contamination [14], were highly represented in the fecal and wastewater samples (Kruskal–Wallis test, $P < 2.2 \times 10^{-16}$; Fig. 1g and Additional file 2: Table S7).

Pseudomonadota members are the primary drivers that shape the resistome

We next analyzed the microbiome profiles from the metagenomic sequencing data to establish the connections between microbial and ARG compositions. At the phylum level, a total of 31 phyla with varied abundance among samples from different sources was identified (Fig. 2a). *Bacillota* (with a relative abundance of 53.0%) and *Bacteroidota* (34.7%) were the dominant phyla in human fecal samples, whereas the *Pseudomonadota* occupied a higher proportion in food (43.0%) and environmental samples (31.3%). By binning the metagenomic contigs, we generated 6067 strain-level metagenome-assembled genomes (MAGs, 99.0% average nucleotide identity [ANI]) and 1302 species-level MAGs (95.0% ANI) from a total of 14,787 MAGs. Among these species, 230 species were identified as putative novel species (Additional file 1: Fig. S3a and Additional file 2: Table S8). Diversity analyses revealed significantly lower Shannon indices for food samples than for human and environmental samples (Kruskal–Wallis test, $P < 2.2 \times 10^{-16}$; Fig. 2b), and the microbial compositions differed significantly among the three groups of samples (PERMANOVA, $P = 0.001$, $R^2 = 10.9\%$; Fig. 2c and Additional file 2: Table S9). We then performed Procrustes analysis to assess the extent to which the microbial composition influenced the resistome profiles and found a significant correlation between the resistome profiles and the microbial composition ($P = 0.001$, $R^2 = 74.0\%$; Fig. 2d). Further, we applied a random forest model to identify the main contributors (at different taxonomic levels) influencing the resistome profiles. The results showed that *E. coli*, a typical member of *Pseudomonadota* at the species level, was determined to be the main taxon distinguishing the resistome profiles (Fig. 2e, f and Additional file 2: Tables S10–S11). Differences in abundance between the various groups of samples revealed the following: the food samples were higher in the abundance of *Gammaproteobacteria* ($P < 2.2 \times 10^{-16}$), *Enterobacterales* ($P = 2.5 \times 10^{-12}$), and *Enterobacteriaceae* ($P = 1.2 \times 10^{-6}$) than the other two groups of samples, while the environmental samples exhibited higher abundances of *Pseudomonadota* ($P < 2.2 \times 10^{-16}$) and *Gammaproteobacteria* ($P < 2.2 \times 10^{-16}$) than the human fecal samples, and *E. coli* ($P = 1.5 \times 10^{-7}$) and *Escherichia* ($P = 6.3 \times 10^{-7}$) were more abundant at the species and genus levels in the human fecal samples (Fig. 2f and Additional file 2: Table S11). Spearman's correlation analyses further verified the close associations between the abundance of these taxa and the abundance and diversity of the ARGs ($P < 0.05$; Fig. 2g and Additional file 2: Table S12). We compared the relative contributions of vertical and horizontal transfer of ARGs, and found that in food samples, the resistome was

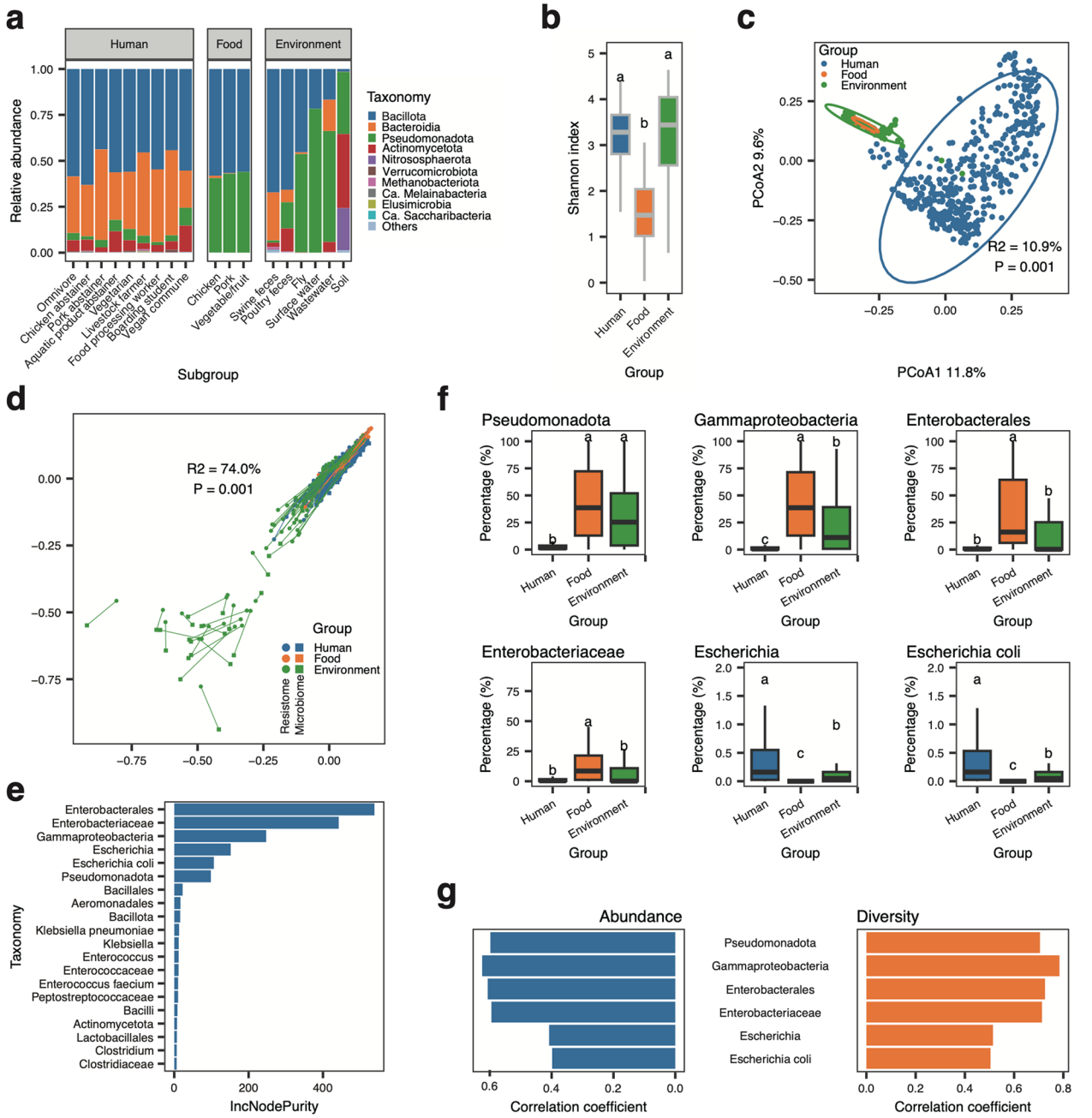

Fig. 2 Relationship between the microbiome and the resistome. **a** The relative abundance of taxa at the phylum level within the microbial communities of samples from various subgroups. **b** The Shannon diversity of the microbial communities in samples from various groups. **c** PCoA of samples from various groups. **d** The correlation between the microbiome and the resistome revealed by Procrustes analysis. Different colored nodes represent different groups; squares and circles represent the microbial profiles and the resistome profiles, respectively. **e** The variable importance of the top 20 taxa was determined by the random forest model using the mean decrease Gini value. A higher value indicates a stronger connection between the taxon and the resistome profile. **f** The relative abundance of six taxa in samples from the three groups. **g** Spearman's correlations between the abundance of the six taxa and the abundance (left panel) and Shannon index (right panel) of the resistome profiles ($P < 0.05$). The length of each bar represents the correlation coefficient

more affected by MGEs, whereas in environmental samples, it was more associated with the microbiome (Additional file 1: Fig. S3b).

To reveal the putative bacterial host of ARGs, we assembled the sequence reads into contigs and used these contigs for taxonomic assignment. After sequence

assembly and ARG annotation, we identified a total of 162,001 ARG-carrying contigs (ACCs) containing 183,720 ARGs from all 592 samples. Taxonomic assignment based on sequence information classified 20.4% of the ACCs at the genus or species level; however, one-third of the ACCs could not be classified, even at the phylum level (Additional file 1: Fig. S3c). A total of 54,831 ACCs, accounting for 33.8% of the total ACCs, could be binned into MAGs, of which, 14,367 had accurate taxonomic assignments according to the taxonomic information of the associated MAGs (Additional file 1: Fig. S3d). Overall, we finally assigned ~ 25.0% of the ACCs to the species or genus level (Additional file 1: Fig. S3e and Additional file 2: Table S13). We then computed the ARG load index at different taxonomic levels, i.e., the proportion of ARGs assigned to each taxon divided by the average relative abundance of the corresponding taxon. At the phylum level, *Pseudomonadota* was found to be the main ARG carrier (load index was 1.5), followed by *Bacillota* (0.8), *Actinomycetota* (0.3), and *Bacteroidota* (0.2) (Additional file 2: Table S14). At the class and order levels, *Gammaproteobacteria* (1.6) and *Enterobacterales* (1.1), respectively, both belonging to *Pseudomonadota*, contributed to the highest load of ARGs.

The enrichment of ARGs in *Enterobacteriaceae* plasmids and prophages/phages

MGEs such as plasmids and prophages/phages are known as a reservoir for ARGs. A total of 13.0% (21,134/162,001) of our assembled ACCs were predicted to be plasmid sequences, and 2.0% (3234/162,001) of ACCs contained prophage/phage sequences (Fig. 3a, b). These sequences together carried 12.3% (22,571/183,720) of the total ARGs. The proportions of plasmid sequences differed among human, food, and environmental samples (Kruskal–Wallis test, $P < 2.2 \times 10^{-16}$), with food samples exhibiting the highest ratio of plasmid sequences compared with human fecal and environmental samples (Fig. 3c). Seen from the ARG composition, the plasmid sequences in food, fly, and surface water samples were found to carry more multidrug resistance genes (Additional file 1: Fig. S4a). Replicon-based typing of the plasmid sequences identified 113 plasmid types that were associated with 102 ARG subtypes (Additional file 1: Fig. S4b), with IncQ1 ($n = 20$), and rep_cluster_1232 ($n = 10$) were the top frequently encountered replicons. To investigate the potential of these plasmid types in disseminating ARGs, we searched the PLSDB database for the replicons we identified in our samples and annotated the ARGs carried by the known plasmid sequences (Additional file 1: Fig. S4c). In total, 291 ARG subtypes were found to be carried by the 113 plasmid types, and plasmids with replicon of IncHI2A/rep_cluster_1088 carried the highest number of ARG subtypes ($n = 106$), followed by IncFIB ($n = 101$) and IncR ($n = 77$). Among these 291 subtypes in the database, 24.4% ($n = 71$) were found carried by the plasmids found in our samples. Interestingly, we found eight ARG subtypes carried by our plasmids that have not been reported previously (Fig. 3d and Additional file 2: Table S15). For example, we showed that *tetL* gene encoding tetracycline efflux protein can be carried by rep_cluster_1018/rep_cluster_1118, rep_clsuter_1118, and rep_cluster_2013 type plasmids.

The prophage/phage fragments found in our samples were predominantly from the family *Peduoviridae* (11.3% in human fecal samples, 33.8% in food samples, and 21.5% in environmental samples; Additional file 1: Fig. S5a–c). Glycopeptide and peptide resistance genes were found to be more abundant in prophages/phages from both human and fecal samples, while the unclassified DNA binding protein H-NS was highly represented in phages from food samples (Additional file 1: Fig. S5d and Additional file 2: Table S16). The glycopeptide resistance gene *vanU* and the peptide resistance gene *ugd* were the two common phage/prophage ARGs found in all three groups of samples. Taking human fecal samples as an example, we found that the frequency of occurrence of each ARG among our samples was highly consistent with that in the Gut Phage Database (GPD) of human gut phages (Spearman's correlation, $R^2 = 0.852$, $P = 5.8 \times 10^{-12}$; Fig. 3e). We then predicted the hosts of the prophages/phages and found that *Bacillota* and *Pseudomonadota* were found to be the main hosts of the phages at the phylum level (Additional file 1: Fig. S5e), and *Enterobacteriaceae* was the main host at the family level (Fig. 3f), especially for prophages/phages in samples from chickens, flies, and surface water (Additional file 1: Fig. S5f). A total of 24 ARG subtypes were found carried by the *Enterobacteriaceae* prophages/phages, including genes encoding multidrug and tetracycline resistance.

Regional AMR dissemination mediated by horizontal gene transfer (HGT) and strain transmission

We then investigated ARG transmission among different habitats by clustering the ARG nucleotide acid sequences from various sources at 100.0% identity and constructed ARG-sharing networks (Additional file 1: Fig. S6a and Additional file 2: Tables S17–S19). According to the ARG risk ranks classified previously [13], we categorized the 328 ARG subtypes into six groups (from high- to low-risk ranks): Q1 ($n = 88$), Q2 ($n = 45$), Q3 ($n = 35$), Q4 ($n = 42$), RI = 0 ($n = 79$), and Unknown ($n = 41$) (Additional file 2: Table S20). We found that ARG subtypes in Q1 had higher sharing frequencies than ARG subtypes in other categories (Kruskal–Wallis test, $P < 2.8 \times 10^{-13}$; Additional file 1: Fig. S6b and Additional file 2: Table S21).

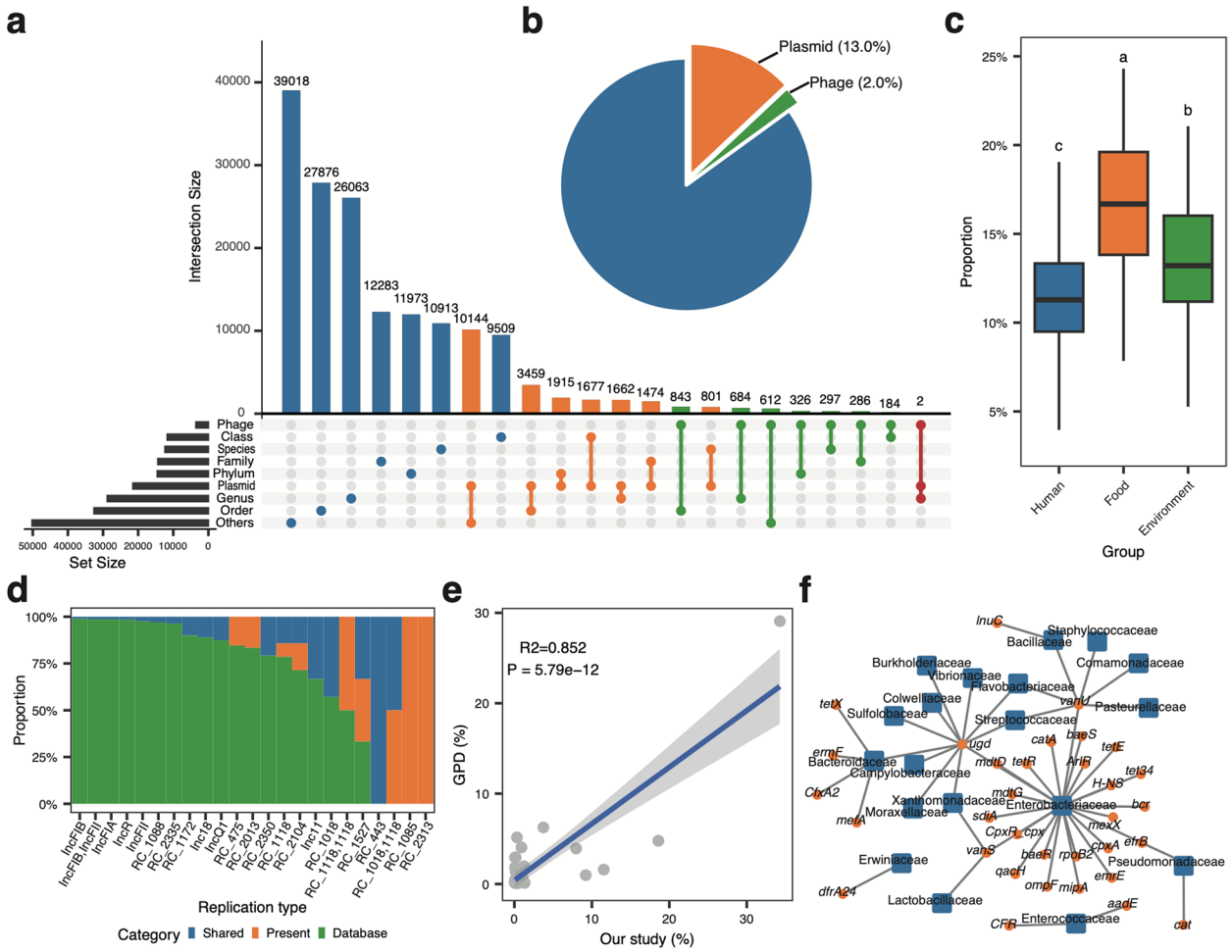

Fig. 3 ARGs carried by plasmids and prophages/phages. **a** The source of ACCs (chromosome, plasmid, or phage). **b** Proportion of plasmids and prophages/phages in all ACCs. **c** Proportion of plasmid sequences in all ACCs of samples from the three groups. **d** Comparisons of the plasmid replicon-related ARGs discovered in our study with those discovered in the PLSDB database. RC, an abbreviation of rep_cluster; different colors represent the outcomes of the comparison (shared, appeared in both our study and the PLSDB database; present, appeared only in our study; database, appeared only in the PLSDB database). **e** Spearman's correlation between the ratios of ARGs carried by prophages/phages in the human samples of this study and in the Gut Phage Database. **f** Co-occurrence network of the ARGs carried by the prophages/phages sequences and the predicted phage hosts. The blue squares represent the families of bacteria; the orange circles represent the ARGs; and the gray lines represent the associations between the ARGs and the phage hosts

Additionally, there was a higher ratio of ARGs shared between human fecal samples and food, animal fecal, wastewater, and fly samples (Fig. 4a), especially those in the Q1 category (Additional file 1: Fig. S6c). This was also the case when different human subgroups were separately compared with the food and environmental subgroups (Fig. 4b). Typically, compared with other human subgroups, the human fecal samples from the vegetarian subgroup exhibited the highest ratio of ARGs shared with vegetables or fruits (41.9%), while the lowest ratio with pork (4.2%). A total of 78 ARGs (Q1) were found to be shared in at least one comparison pair between fly, pork, vegetables/fruits, and different human subgroups (Fig. 4c and Additional file 2: Table S22), and interestingly, flies shared a relatively high ratio of ARGs with food processing workers. Taken together, these results highlight the important role of the high-risk ARGs in AMR spread.

We then used StrainPhlAn to create maps of strain-sharing events between samples to reveal regional strain transmission. We showed that 257 species-level genome bins (SGBs) were present in at least 20 samples (Additional file 2: Table S23). The most prevalent SGB was *Agathobacter rectalis* ($n=328$) from the family *Lachnospiraceae*, followed by *Blautia wexlerae* ($n=311$), *Faecalibacterium prausnitzii* ($n=299$), *Prevotella copri* ($n=298$), and *Blautia_A faecis* ($n=298$). Among six

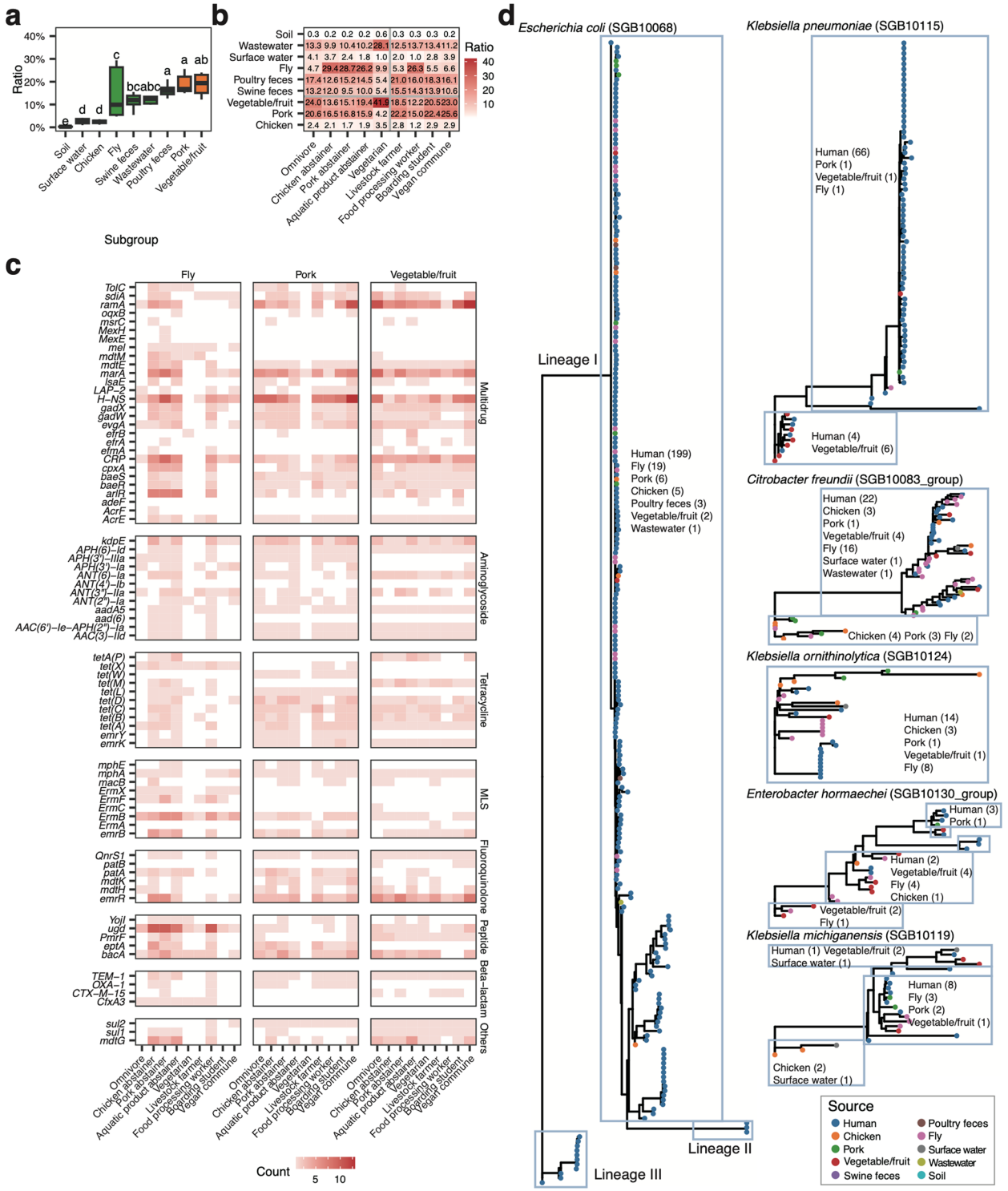

Fig. 4 ARG-sharing network and clonal spread. **a** The ratio of ARGs that are shared between human subgroups and food/environmental subgroups, relative to the total shared ARGs between human fecal samples and food/environmental samples. **b** The proportion of ARGs that are shared with different food/environmental subgroups in different human subgroups. **c** ARG subtypes are shared between samples from different human subgroups and samples from vegetables, fruits, pork, and flies. Only ARG subtypes from the risk rank Q1 were considered. **d** Strain-level phylogenies of seven species of the family *Enterobacteriaceae*. Each node represents a sample containing the corresponding strain. Different colors represent samples from different sources. Blue lines indicate that these samples possess the same strain (normalized phylogenetic distance, nGD < 0.1). The annotations are the number of samples containing the corresponding strain

species of *Enterobacteriaceae*, *E. coli* (SGB10068) was the most prevalent species in our samples ($n=249$), consisting of three different strain lineages (I to III). Lineage I was the largest group present in 235 samples from multiple sources, across human, fly, food (pork, chicken, vegetables, and fruits), and wastewater. For other species in *Enterobacteriaceae*, the same strain lineage of *Klebsiella pneumoniae*, *K. michiganensis*, *K. ornithinolytica*, *Clostridium freundii*, and *Enterobacter hormaechei* was also found shared among samples of different origins, mainly human, food (chicken, pork, vegetables, and fruits) and fly. Notably, flies may play an important role in the transmission of *Enterobacteriaceae* strains, as samples from flies appeared in almost all of the main strain lineages in the phylogenetic trees (Fig. 4d). Apart from the *Enterobacteriaceae*, strain transmission events between human and other samples were also detected for species from other families, including *C. perfringens*, *Enterococcus faecium*, *Weissella confusa*, and *Streptococcus pasteurianus* (Additional file 1: Fig. S7). For example, one *C. perfringens* strain was found in 51 samples across human, vegetable/fruit, fly, pork, and chicken, representing a typical regional strain transmission event of opportunistic pathogen.

The regional dissemination of carbapenemase genes

To further reveal regional ARG transmission, we selected carbapenemase genes as representative ARGs. We identified a total of 111 subtypes of carbapenemase genes in the ACCs from all of our samples (Additional file 2: Table S24). *OXA-347* ($n=308$) was found to be the most prevalent gene carried by the ACCs (Additional file 1: Fig. S8a), being predominantly present in human and animal fecal samples, and wastewater (Additional file 1: Fig. S8b, c).

To uncover the potential transmission of carbapenemase genes, we analyzed the sharing events of these genes across various sources. A total of 200 unique sequence types were identified from the top 20 frequently encountered carbapenemase genes (1 to 30 distinct sequence types per gene), clustering with a cut-off of 100.0% nucleotide identity (Additional file 1: Fig. S9). Among these 20 genes, 11 genes were found to be shared among samples of different sources, suggesting complex dissemination pathways of these genes. For example, *OXA-347*, *OXA-1*, *OXA-85*, and *CfiA2* showed a high sharing ratio between feces and wastewater. Interestingly, *ACT-12* and *ORN-1* showed a high sharing ratio between food and fly samples, suggesting that flies may serve as mediators for ARG cross-transmission. Of note, a sequence type of *OXA-50* was detected in poultry feces, livestock farmers, and food processing workers. There were rare instances of sharing events between chicken or pork and poultry or pork abstainers. Interestingly, vegans who live together were more likely to carry specific carbapenemase genes, such as *SHV-27*, *ORN-1*, and *ACT-28*, possibly due to person-to-person transmission.

We next isolated carbapenem-resistant *Enterobacteriaceae* strains from our 592 samples. In total, 45 resistant isolates affiliated with nine bacterial species were identified from 42 samples (Fig. 5a and Additional file 2: Table S25). The annotation results indicated that 17 carbapenemase gene subtypes were carried by these isolates, among which six subtypes, namely, *IMP-4*, *NDM-3*, *NDM-5*, *OXA-805*, *SFO-1*, and *SHV-61*, were only found in the pure culture isolates, not in the metagenomic data. The carbapenemase genes recovered from these isolates accounted for only a small fraction of the subtypes found in the metagenomic data ($n=111$), indicating the presence of other carriers besides *Enterobacteriaceae* members.

We then clustered each species by including the corresponding genomes of the isolates and MAGs that were assembled from the metagenomic sequencing data. We observed potential transmission of the same strain (ANI ≥ 99.0%) harboring different carbapenemase genes across different sample sources (Fig. 5b and Additional file 2: Table S26). For example, the *E. coli* strain cluster 1 carrying different carbapenemase genes (such as *NDM-1*, *NDM-5*, *OXA-1*, *OXA-10*, and *CTX-M-27*) was widely distributed among 141 samples covering human fecal samples and another seven subgroups (except for samples from surface water and wastewater). Similar situations were also found for *E. coli* strain clusters 2 and 3 containing *NDM-1* or *NDM-5*, and the *K. pneumoniae* strain cluster carrying *SHV-61* or *SHV-33*. A strain of *K. ornithinolytica* harboring three carbapenemase genes (*SFO-1*, *ORN-1*, and *NDM-1*) was found to only be present in fly and human samples, again suggesting the significant role of fly-mediated transmission of ARGs.

Our metagenomic data showed that carbapenemase gene *OXA-347* was prevalent in human and animal fecal samples, however, we did not recover isolates carrying carbapenemase gene *OXA-347*, probably because the SuperCARBA medium used mainly targets *Enterobacteriaceae* bacteria. To determine the host range and potential dissemination of *OXA-347*, we searched for this gene among the complete genomes in the database (Additional file 1: Fig. S10). The *OXA-347* gene was present in 12 bacterial species associated with two known types of plasmid replicons (rep_cluster_663 and rep_cluster_1097) in three *Bacteroides* species (*B. fragilis*, *B. thetaiotaomicron*, and *B. xylanisolvens*) (Fig. 5c and Additional file 2: Table S27) and two unknown types of plasmids in *Myroides albus* and *Sphingobacterium faecium*. Additionally, the chromosome-located

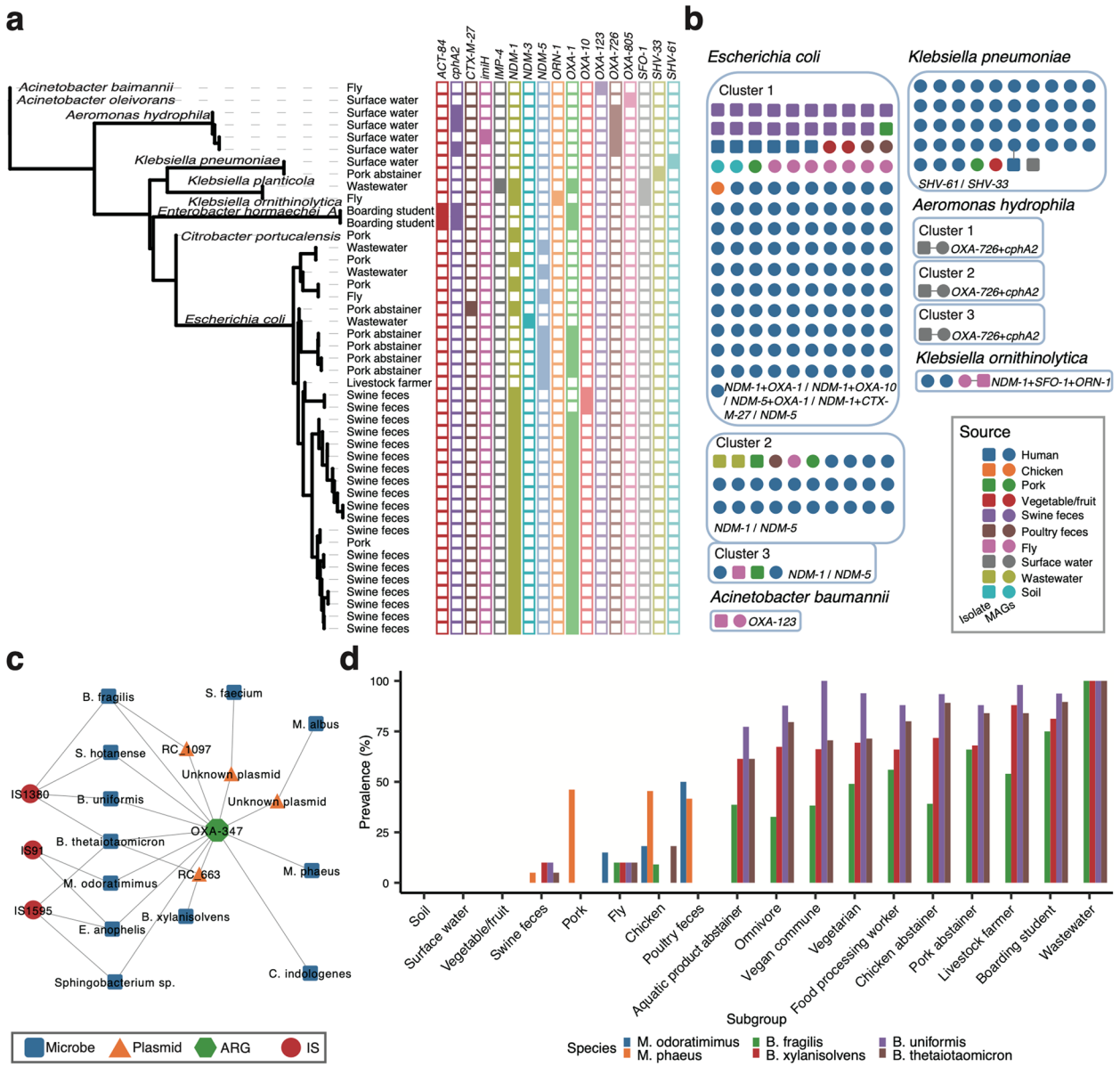

Fig. 5 Distribution and transmission of carbapenemase genes. **a** Phylogenetic tree of 45 carbapenem-resistant isolates was collected in this study. The right panel shows the presence of specific carbapenemase genes. **b** Clusters of MAGs and the isolate genomes in the seven species. We clustered each species by including the corresponding genomes of the isolates and the MAGs with a cut-off of 99.0% ANI. Only the species found both in MAGs and isolates are present. Different colored circles represent the various sources of the MAGs. Blue squares represent the source of the chromosomes. The gray line indicates that the connected MAG and isolate were obtained from the same sample. The rounded blue rectangles indicate groups containing the same strain as the origin of the genomes (ANI ≥ 99.0%). The carbapenemase genes found in the isolate genomes are labeled. **c** The co-occurrence network depicts the relationships between the *OXA-347* gene and the associated microbes, plasmids, and insertion sequences. The blue rectangles represent the microbes; the green hexagon represents the *OXA-347* gene; the orange triangles represent the plasmid replicons; and the red circles represent the insertion sequences. **d** Prevalence of the potential *OXA-347*-carrying microbial species revealed in **c**. Only six species were detected in our data

OXA-347 gene was found to be frequently flanked by insertion sequences such as IS1380, IS91, and IS1595. These results suggest a high spread risk of *OXA-347* via horizontal gene transfer. Out of the 12 bacterial species possessing *OXA-347*, four were found to be common inhabitants in both human feces and wastewater samples (Fig. 5d).

Prediction of carbapenem-resistant strains using a machine learning model

We finally constructed random forest models to evaluate the potential of utilizing metagenomic sequencing data for predicting carbapenem-resistant strains. The results showed that all prediction models could achieve an area under the curve (AUC) exceeding 0.90 when constructed based on microbiome profiles, resistome profiles, or a combination of both (Fig. 6a–c). Notably, microbiome profiles proved to be robust indicators with an AUC of 0.939 and a specificity of 0.595. Among the identified predictors, *Aeromonas hydrophila*, *Bacilli bacterium*, *Alistipes putredinis*, and *Blautia wexlerae* were considered strong contributors (Fig. 6d).

Discussion

In this study, we revealed the regional ARG flow among humans, food, and the environment in a county-level city in China by applying a One Health sampling approach. The key findings from this study were (1) antibiotic resistomes are habitat-specific, and human or animal fecal contamination is an important factor that influences the wastewater ARG composition; (2) *Enterobacteriaceae* bacteria together with their plasmids and bacteriophages are the main ARG carriers; (3) flies and food may be important mediators for the regional spread of ARGs; (4) HGT and strain transmission independently or jointly contribute to regional AMR dissemination; (5) the carbapenemase gene *OXA-347* is widely present among human and animal gut microbiomes; and (6) a microbiome profile-based machine learning model can predict the presence of carbapenem-resistant strains in metagenomic samples.

We found that *Pseudomonadota* bacteria, especially those from the *Enterobacteriaceae*, are the primary drivers that shape the resistome, which is consistent with a previous study [15]. The *Enterobacteriaceae* family encompasses numerous pathogens commonly identified in clinical cultures and poses a major risk to human health as a result of their role as primary reservoirs for ARGs [16]. Our findings indicate that members of the *Enterobacteriaceae* are also the major ARG carriers in non-clinical settings, supporting the use of *Enterobacteriaceae* as indicators for AMR surveillance worldwide [17]. Additionally, we showed that *Enterobacteriaceae* carrying carbapenemase genes (*SHV-61*, *NDM-1*, *OXA-1*) were present in surface water, pork, and fly samples, highlighting the potential of regional *Enterobacteriaceae*-mediated dissemination of clinically important ARGs.

HGT mediated by MGEs is recognized as the major reason for AMR dissemination. ARGs found in MGEs are even more "global" than microbes, as illustrated by the ability of mobile ARGs to cross habitat boundaries [18]. The higher ratio of plasmids sequences in food samples we presented here suggests the higher risk of ARG transmission from food to humans (Fig. 3c). Besides plasmids, we showed that *Enterobacteriaceae* prophages/phages were also important ARG carriers in both our samples and the GPD database (Fig. 3e–f), supporting the significant role of phages in the transfer of ARGs [19]. Although HGT greatly contributes to ARG dissemination among microbial communities, strain transmission or clone spread can be regarded as a major route for regional epidemics of both AMR and infections. This was strongly supported by our finding that a single *E. coli* strain (SGB10068 Lineage I) was present in 235 samples from multiple sources, and an *E. coli* strain (cluster 1) carrying

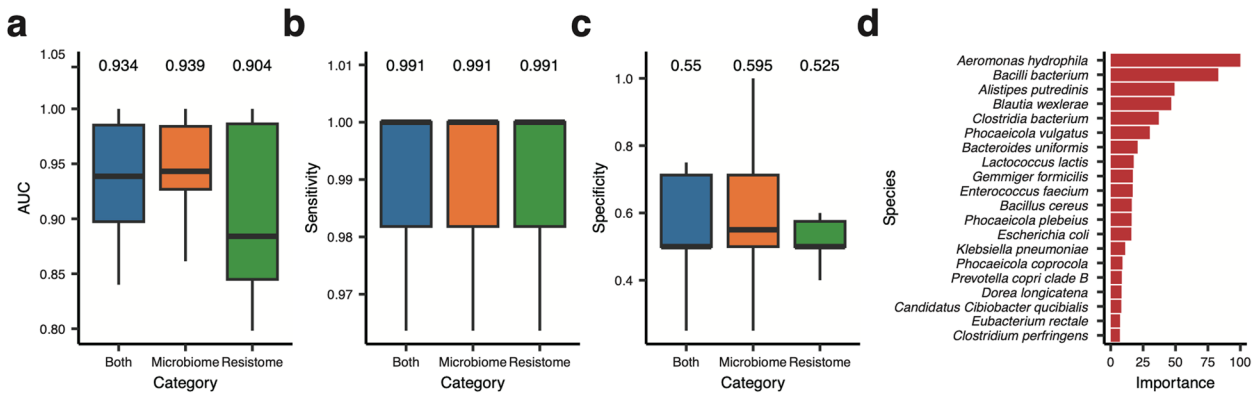

Fig. 6 Random forest model to predict the presence of carbapenem-resistant isolates. Performance of the random forest models used to predict the presence of carbapenem-resistant isolates, utilizing either microbiome profiles, resistome profiles, or a combination of both. The performance indicators, including the area under the curve (**a**), sensitivity (**b**), and specificity (**c**), were computed based on the results obtained from a tenfold cross-validation. The numbers displayed at the top represent the mean values of these three indicators. **d** The importance of the top 20 taxa was determined by the random forest model built on the microbiome profiles

carbapenemase genes was widely distributed among 141 samples. Notably, different carbapenemase genes were found to be carried by the same strain [i.e., *E. coli* strain (cluster 1)], suggesting HGT events (Fig. 4d). These findings imply that regional AMR dissemination involves strain transmission accompanied by HGT and ARG exchange on MGEs such as plasmids and prophages/phages. Additionally, the underestimation of carbapenemase genes by metagenomics may reflect the fact that many are harbored in microbes, especially ARB that are difficult to cultivate.

In this study, we observed a high prevalence of the carbapenemase gene *OXA-347* in human and animal feces, as well as in wastewater. *OXA-347* is associated with phenotypic resistance to penicillin, cephalosporins, and imipenem [20]. This gene has been found in species from the genera *Bacteroides*, *Myroides*, and *Capnocytophaga* from various sources [20–22]. In wastewater samples from the swine feedlots, *OXA-347* was found the most abundant beta-lactam resistance gene [23], and compared with swine fecal samples, the abundance of *OXA-347* was significantly higher in the human gut microbiome [24]. Notably, consistent with our results, recent genomic analyses suggest that *OXA-347* is likely located on MGEs, indicating its potential ability to move between microbes. Although *OXA-347* is currently not considered a major clinical concern, given the fact that the administration of antibiotics led to an increase in the abundance of *OXA-347*, possibly via HGT under selection pressure in the human microbiome [25], future research to evaluate its potential risk to human health should be a priority.

The dissemination of AMR among humans, animals, and environments includes both direct and indirect pathways involving various One Health interfaces [26]. Studies have suggested that food may act as a reservoir of ARB [27], indicating the critical role of diet in the transmission of ARGs from food to humans. This was evidenced by our finding that vegetarians shared more ARGs from vegetables or fruits than from pork. We also showed that vegetable/fruit microbes carried a relatively higher abundance of ARGs, probably because vegetables are accessible to microbial contamination via different routes such as manure fertilization and wastewater irrigation [28]. In addition to diet, we identified flies as a key environmental factor that facilitates ARG dissemination. Fly microbes were found to carry the highest relative abundance of ARGs among all the subgroup samples, and frequently shared *Enterobacteriaceae* strains as well as carbapenemase genes with other samples (Fig. 4d). As a result of their omnivorous diet and breeding habits [29], flies may play a significant but neglected role in the spread and transmission of pathogens and ARGs among the One Health sectors. Our findings highlight the importance of the vegetable- and fly-mediated ARG transmission routes and strongly suggest that these may be new points of intervention for the control of ARG spread. Additionally, we found potential person-to-person transmission of carbapenemase genes among vegan communes and showed the flow of the carbapenemase gene along the food supply chain (Additional file 1: Fig. S9), providing specific examples of the impact of lifestyle, dietary habits, and occupational exposure on the transmission of ARG to humans. Taken together, these results highlight the potential routes of the AMR spread, especially through the food-chain dissemination pathway and occupational exposure.

Conclusions

In summary, we give a landscape of regional ARG flow among humans, food, and the environment in a city in China. We highlight that the resistome profiles in samples of different origins showed habitat specificity, and *Pseudomonadota* members are the major contributors shaping the resistome through strain transmission and/or MGEs (plasmids and prophages or phages)-mediated HGT. We suggest that lifestyle, dietary habits, and occupational exposure are all risk factors that contribute to the spread of ARGs to humans. We also have evidence that the integration of metagenomic sequencing data and machine learning techniques could serve as a valuable approach for the surveillance of carbapenem-resistant strains. Collectively, the regional antimicrobial resistance gene flow we presented here provides real evidence of the AMR dynamics among the One Health sectors, and our findings highlight new points of focus for AMR surveillance and control in the future. We should stress that efforts are still needed to probe seasonal influences on the regional antimicrobial resistance gene flow, which was not taken into account in this study.

Methods

Sample collection

From October 2018 to April 2019, we collected a total of 592 samples from Dengfeng, Henan Province, China (Additional file 3), deriving from humans, food, and the environment. In detail, we collected fecal samples from individuals with various dietary patterns, including omnivores ($n=49$), vegetarians ($n=49$), pork abstainers ($n=50$), chicken abstainers ($n=46$), and aquatic product abstainers ($n=44$). We also collected fecal samples from humans with different lifestyles or occupational exposures, including vegan communes ($n=68$), boarding students ($n=48$), food processing workers ($n=50$), and livestock farmers ($n=50$). Additionally, 51 samples from various food sources were collected, including pork ($n=13$), chicken ($n=11$), and vegetables and

fruits ($n=27$); and 87 samples from different environments were collected, including soil ($n=10$), surface water ($n=9$), wastewater ($n=4$), flies ($n=20$), poultry feces ($n=24$), and swine feces ($n=20$). All samples were used for culturing carbapenem-resistant microbes and metagenomic sequencing.

DNA extraction and metagenomic sequencing

For metagenomic sequencing, DNA extraction was performed using the QIAamp Power Fecal DNA Kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. The DNA concentration was measured using a Qubit dsDNA Assay Kit and Qubit 2.0 fluorometer (Life Technologies, CA, USA), and the integrity was assessed using 0.8% agarose gel electrophoresis. Libraries were constructed using the MGIEasy FS DNA Library Prep Set (MGI Tech, Shenzhen, China). A paired-end library with an insert size of ~ 350 bp was constructed for each sample and sequenced on the MGISEQ-2000RS platform (MGI Tech). To avoid potential contamination indicated in a previous study [30], we also incorporated negative controls during the processes of library construction and metagenomic sequencing. The metagenomic sequencing of the 592 samples generated a total of 3492.7 Gb (6.0 Gb per sample after quality control, with a standard deviation of 1.9 Gb).

Microbiome profiling and resistome profiling

For the metagenomic sequencing data, we excluded low-quality bases and residual adapter contamination using Trim Galore (https://github.com/FelixKrueger/TrimGalore). To remove human DNA contamination, the sequencing data were mapped to the human genome (hg38) using Bowtie2 v2.3.5.1 [31]. Taxonomic profiling of the metagenomic sequencing samples was performed using MetaPhlAn v4.0.2 [32] with default parameters. The construction of the sample-specific strains of all species was performed on the samples using StrainPhlAn v4.0.2 [33]. We considered SGBs that met the following criterion: present in at least three samples in one subgroup of both human and food or environmental origin. To detect strain-sharing events, we first calculated the SGB-specific normalized phylogenetic distance (nGD) using pyphlan (https://github.com/SegataLab/pyphlan). Strain boundaries were set to below the threshold of 0.1 of the nGD. Resistome profiling of the samples was performed using deepARG v2 [34] with the pipeline short-read pipeline to predict ARGs directly from short reads. The diversity of the microbiome and the resistome was calculated using the R package vegan v2.6–4.

The presence of the crAssphage genome was regarded as an indicator of human or animal fecal contamination [35]. We used Bowtie v2.3.5.1 [31] to align the clean metagenomic sequencing reads to the reference genomes of *Carjivirus communis* (JQ9955537.1) and *Bacteroides phage B40-8* (NC_011222.1), as recommended previously [14]. The average coverage of each phage genome was calculated using SAMtools v1.9 [36]. Subsequently, the phage coverage was normalized by the data size of each sample (copies/Gb) to compare the profiles among the samples from various sources.

ExtrARG [37], a machine-learning approach using the extremely randomized tree algorithm, was utilized to identify discriminatory ARGs. Values with an importance greater than 0.004 were considered biomarkers from various sources. To identify the taxa that were associated with the resistome profiles, we applied feature selection by sorting the mean decrease in Gini values generated by the R package randomForest v4.7–1.1.

Metagenomic assembly and genome binning

The clean reads from each set of metagenomic sequencing data were independently assembled using MEGAHIT v1.1.3 [38]. For metagenomic binning, three methods, namely MetaBAT2 v2.12.1 [39], Maxbin2 v2.2.6 [40], and Concoct v1.0.0 [41], were used. A superior bin set from multiple original binning predictions was produced using the bin_refinement [42] module of metaWRAP v1.3.2 [43]. This module combines the three original binning predictions. The completeness and contamination of each bin from the superior bin set were evaluated using CheckM v1.0.12 [44]. Then, bins with a completeness of $\geq 70.0\%$ and a contamination rate of $\leq 10.0\%$ were retained. All MAGs were dereplicated at 99.0% ANI and 95.0% ANI using dRep v2.6.2 [45]. The bin annotation pipeline of CAT v5.2.3 [46] was used to assign taxonomy to MAGs. A genome was classified as a novel species if the ANI output was less than 95.0%. We also compared the MAGs with the isolate genomes by clustering at 99.0% ANI. The phylogenetic trees were inferred using PhyloPhlAn v3.0.60 [47].

Correlations between the MGEs and the microbiome

To determine the correlations between the resistome and both the microbiome (at the phylum level) and the MGEs in Additional file 1: Fig. S3b, the abundance matrices were analyzed through Procrustes analysis and a "protest" test. Here, the abundance of MGEs and the 16S rRNA sequence for each sample was calculated by mapping to the MobileGeneticElement [48] and SILVA [49] databases, respectively. The MGE profiles were normalized by the copies of MGEs per 16S rRNA gene.

ARG annotation and taxonomic assignment of ACCs

The gene contents in the assembled contigs were predicted using GeneMark-HM v2.07 [50], and ARGs were

further predicted using the "DeepARG-LS" mode of deepARG. To maintain consistency with the taxonomic assignment of the MAGs, we also utilized the contig annotation tool in CAT to predict the taxonomic assignment of the ACCs. To reveal the risk ranks of the discovered ARGs, the ARGs identified by deepARG were re-annotated against CARD [51] to ensure consistency with the risk rank assignments made in a previous study [13]. The ARG load index at different taxonomic levels was calculated using the following formula: the proportion of ARGs assigned to each taxon divided by the average relative abundance of the corresponding taxon.

Analyses of plasmid and prophage/phage sequences

We used geNomad v1.5.2 [52], a classification and annotation framework, to find plasmid and prophage/phage sequences from the ACCs. Taxonomic assignment of the phage fragments generated from ACCs was performed using PhaGCN2 [53], followed by host prediction using HostG [54]. Because some segments of the ACCs contained both prophage fragments and microbial genomes, we extracted the ARGs located within the prophage fragments using local scripts. The predicted plasmid sequences were typed using the mob_typer method implemented in MOB-suite v3.1.0 based on the replicon, relaxase, and cluster assignment [55]. This method provides in silico predictions of the replicon family, relaxase type, mate-pair formation type, and predicted transferability of the plasmid. The ARG annotation of the predicted prophage/phage-related and plasmid-related ACCs followed the aforementioned guidelines.

To reveal the prophage/phage-associated ARGs on a global scale, we downloaded the GPD [56], a collection of 142,809 non-redundant viral genomes obtained by mining a dataset of 28,060 globally distributed human gut microbiomes. The ARG annotation followed the same pipeline as mentioned above. We established the potential connections between the plasmid replicon types and the ARGs through further analysis of the PLSDB database [57], a plasmid database obtained from the NCBI nucleotide database, which comprises 34,513 sequences.

Determination of ARG-sharing networks

Over a shorter time span, the substitution rate of bacteria typically falls within the range of ~ 1 single nucleotide polymorphism per genome per year [58]. Thus, as suggested by a previous study [59], the recently transmitted ARGs should have identical sequences (100.0% identity) between two sources, considering the length of the ARGs. To identify potential ARG transmission among various sources, we conducted the following steps: (1) clustering of the DNA sequences of the ARGs discovered in the metagenomic assemblies from each source, with 100.0% identity and 85.0% coverage, using the scripts provided by checkV v1.0.1 [60] to generate the non-redundant set of ARGs from various sources; and (2) comparing the non-redundant sets of ARGs from various sources and constructing the ARG-sharing network between human and other samples.

Isolation and antimicrobial susceptibility testing of carbapenem-resistant strains

SuperCARBA (CHROMagar, Paris, France) was used to isolate carbapenem-resistant bacteria, mainly *Enterobacteriaceae*. Colonies with various morphologies were selected from each Petri dish. A matrix-assisted laser desorption/ionization time-of-flight mass spectrometry system (Zybio Inc., Chongqing, China) was used for the early identification of isolates.

Antimicrobial susceptibility testing of 28 antimicrobial agents (amikacin, amoxicillin-clavulanate, ampicillin-sulbactam, aztreonam, cefazolin, cefepime, cefoperazone-sulbactam, cefoxitin, ceftazidime, ceftriaxone, cefuroxime, chloramphenicol, ciprofloxacin, colistin, ertapenem, gentamicin, imipenem, levofloxacin, meropenem, minocycline, moxifloxacin, nitrofurantoin, norfloxacin, piperacillin-tazobactam, tetracycline, tigecycline, tobramycin, and trimethoprim-sulfamethoxazole) was performed using the agar dilution method following the standards and guidelines of the Clinical and Laboratory Standards Institute M100-2022 (V32) [61]. *E. coli* ATCC 25922 and *Enterococcus faecalis* ATCC 29212 were used as quality control strains in each run.

To obtain a more comprehensive view of carbapenem resistance gene carriage in bacteria, we analyzed the carbapenemase genes in complete genomes (41,185 genomes) downloaded from the National Center for Biotechnology Information (NCBI, July 29th, 2023) by CARD. We then used ISEScan v1.7.2.3 [62] to identify the insertion sequence elements surrounding the carbapenemase gene *OXA-347*. To determine the association between insertion sequences and ARGs on plasmids, the adjacent sequences, i.e. 5 kb upstream and downstream of ARGs, were extracted as previously suggested [63].

Genome sequencing and bioinformatic analyses of the isolates

For whole-genome sequencing, the genomic DNA of the 49 carbapenem-resistant isolates was extracted using a Wizard Genomic DNA Extraction Kit (Promega, Madison, WI, USA). The DNA concentration was measured using a Qubit dsDNA Assay Kit and a Qubit 2.0 fluorometer (Life Technologies), and the integrity was assessed using 1.0% agarose gel electrophoresis. Libraries were constructed using the MGIEasy FS DNA Library Prep Set (MGI Tech, Shenzhen, China). A paired-end library with

an insert size of 350 bp was constructed for each sample and sequenced on the MGISEQ-200RS platform (MGI Tech, Shenzhen, China).

Before assembling the microbial genomes, we performed quality control using Trim Galore (https://github.com/FelixKrueger/TrimGalore). The sequence reads were assembled using Unicycler v0.4.8 [64] with default parameters, followed by annotation using Prokka v1.13.3 [65]. To remove duplicate genomes from the same sample, we estimated genome similarity by calculating the pairwise Mash distance using Mash v2.3 [66] and removing the duplicated genomes with MASH distances less than 0.001. Phylogenetic trees of all isolates were constructed using PhyloPhlAn with the default parameters. We estimated maximum likelihood trees for microbial genomes of the same species from isolates and MAGs using IQ-TREE 2 v2.1.4-beta [67]. The best-fitting substitution model was automatically selected using the ModelFinder [68] program implemented in IQ-TREE 2 and performed with 1000 bootstrap replicates. The ARGs from the microbial genomes were annotated against CARD. All draft genomes of the isolates and MAGs belonging to the same strain were clustered at 99.0% ANI using dRep.

Predictive random forest models

Random forest models were utilized to forecast the presence of carbapenem-resistant isolates based on microbiome profiles, resistome profiles, or a combination of both, across 592 samples. To assess the predictive performance of these random forest classification models, we used three indicators (AUC, sensitivity, and specificity), derived from a tenfold cross-validation in this study. In this process, the dataset was randomly divided into 10 equal subsets. During each iteration, nine subsets were used for model training, while the remaining subset was used for prediction. The values of each iteration were recorded. These analyses were performed using the R packages randomForest v4.7–1.1 and caret v6.0–94.

Statistical analysis and visualization

Differential analyses were performed by Wilcoxon rank sum tests or Kruskal–Wallis tests. When analyzing more than two groups, multiple comparisons were conducted using the R package agricolae v1.3–7. For beta diversity statistical analyses, the "adonis" function from the R package vegan v2.6–4 and the "pairwise.adonis" function from the R package pairwiseAdonis v0.4.1 were utilized. We employed Procrustes analysis to establish correlations among various profiles using the "procrustes" function of the R package vegan v2.6–4. The correlations were determined by the "protest" function of the R package vegan v2.6–4. The correlations between two variables in this study were calculated using the R function "cor.test".

The intersecting sets were analyzed using the R package UpSetR v1.4.0 [69]. The co-occurrence networks were visualized using Cytoscape v3.10.1 [70]. The heatmaps were visualized using the R package ComplexHeatmap v2.16.0 [71]. The phylogenetic trees were visualized using the R package ggtree v3.8.2 [72] or iTOL v5 [73]. Figures were visualized primarily using the R packages ggplot2 v3.4.4 and patchwork v1.1.3.

Abbreviations

ACC	ARG-carrying contig
ANI	Average nucleotide identity
AUC	Area under the curve
AMR	Antimicrobial resistance
ARB	Antimicrobial-resistant bacteria
ARG	Antimicrobial resistance gene
GPD	Gut Phage Database
HGT	Horizontal gene transfer
MAG	Metagenome assembled genome
MGE	Mobile genetic element
MLS	Macrolide-lincosamide-streptogramin
NCBI	National Center for Biotechnology Information
nGD	Normalized phylogenetic distance
PCoA	Principal coordinate analysis

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-024-01983-x.

Additional file 1: Fig. S1: The abundance of the ARG types. (a) Average abundance of the main ARG types in various subgroups. (b) The abundance of the top 10 ARG types in the three groups. (c) The abundance of the 40 ARG types in all subgroups. Left panels: the ARG types (n = 20) with relative higher abundance; right panels: the ARG types (n = 20) with relative lower abundance. Fig. S2: Diversity analysis of the resistome profiles. (a) The Shannon indices of the resistome profiles. (b) Axis 1 values of PCoA of the resistome profiles from various subgroups. (c) Bray–Curtis distances between the resistome profiles of various human fecal samples and those from food and environmental samples. Figure S3: Metagenome-assembled genomes and ACC taxonomic assignment. (a) Phylogenetic tree of the 1,302 MAGs at the species level. A total of 14,787 MAGs (completeness ≥ 70.0% and contamination ≤ 10.0%) were constructed in this study. The novel species are labeled with light blue bars in the outer ring. The color of the inner ring represents the class level of the MAG. (b) Procrustes analysis of the correlations between the resistome and both MGEs and the microbiome in the three groups. The correlations were calculated by the function "protest" in R package vegan. (c) Proportion of taxonomic assignment of the ACCs by the software CAT. CAT exploits homology searches of individual open reading frames to classify ACCs directly. (d) The taxonomic assignment performance of the two methods based on the 54,831 ACCs, which were binned into MAGs out of 162,001 ACCs. (e) Proportion of the taxonomic assignment of the ACCs with the integration of taxonomic information of MAGs. Figure S4: Plasmid-associated ARGs. (a) The proportion of ARG types carried by plasmids in the different subgroups. (b) Co-occurrence network between plasmid replicon types and the ARGs revealed by the ACCs. The brown rounded rectangles represent the plasmid replicons, the colored circle nodes represent the ARG types, and the gray lines represent the associations between the ARGs and the plasmid replicons. (c) An expanded association between the ARGs identified in (b) and the plasmid types identified by integrating the results from the PLSDB database. The brown squares represent the plasmid replicons; the colored circles represent the ARG types; and the gray lines represent the associations between the ARGs and the plasmid replicons. Figure S5: Analysis of the prophages/phages identified from ACCs. The ratio of

different taxa of prophage/phage identified in samples from human (a), food (b), and environmental (c) sources. (d) Proportion of ARGs carried by phage/prophage fragments. Host prediction of prophage fragments at the phylum (e) and family (f) levels. Figure S6: Analysis of the ARG-sharing events. (a) ARG-sharing network. A cluster of shapes represents an ARG subtype shared among certain subgroups. The color of each node indicates the subgroup, while the different shapes represent human, food, and environmental samples. The size of the circle corresponds to the frequency of shared events. (b) Frequency of ARG-sharing events in the different risk ranks. Unknown represents the ARGs without a classification. (c) The ratio of ARGs found in food and environmental subgroups shared with human fecal groups. Figure S7: Strain-level phylogenies of species present in all 592 samples. Each node represents the main strain of a sample. Different colors represent various subgroups. Light blue rectangles indicate phylogenetic trees of the same strain (nGD < 0.1). The annotations are the number of samples harboring the same strain. Figure S8: Distribution of carbapenemase genes found in the metagenomic sequencing data. (a) The prevalence rates of the top 20 carbapenemase genes in all 592 samples. The ratio of positive samples for carbapenemase genes in the three groups (b) and the 17 subgroups (c). The numbers labeled on cells represent the ratio of positive samples for each carbapenemase gene. Figure S9: Phylogenetic trees of the top 20 carbapenemase genes. In the phylogenetic tree, each branch represents a sequence type of the carbapenemase gene. Orange circle nodes in the phylogenetic trees represent the reference carbapenemase gene sequences from CARD. Heatmap illustrates the prevalence of each sequence type in the different subgroups, and the color of the cells reflects the number of positive samples. The red border of each cell represents the sharing events of the carbapenemase gene across the different subgroups. Figure S10: Co-occurrence network of carbapenemase genes and their associated hosts identified in complete genomes. The carbapenemase genes includes those from both metagenomic sequencing data and the strain genomic data in our study. The blue nodes represent microbes; the orange nodes represent plasmid replicons, the green nodes represent the ARGs, and the gray lines represent the locations of the ARGs.

Additional file 2: Table S1: Sample information and sequencing statistics, related to Fig. 1a. Table S2: Statistics of the top 10 ARG types, related to Fig. 1b. Table S3: Average abundance of the Top 10 ARG types in different subgroups (copies per copy of 16S rRNA gene), related to Fig. S1a. Table S4: Statistical tests for the abundance of ARG types among different groups, related to Fig. S1b. Table S5: Statistical tests for the abundance of the 40 ARG subtypes (Kruskal–Wallis test), related to Fig. S1c. Table S6: Pairwise permutational multivariate ANOVA (PERMANOVA) for resistome profiles in the three groups, related to Fig. 1f. Table S7: Abundance of bacteriophages B40-8 and crAssphage, related to Fig. 1g. Table S8: Taxonomic assignment of the MAGs, related to Fig. S3a. Table S9: Pairwise permutational multivariate ANOVA (PERMANOVA) for the microbial compositions, related to Fig. 2c. Table S10: Variable importance, related to Fig. 2e. Table S10 Variable importance, related to Fig. 2e. Table S11: Statistical tests for the abundance of the six taxa, related to Fig. 2f. Table S12: Correlations between the abundance of the six taxa and the abundance/alpha-diversity of the resistome profiles, related to Fig. 2g. Table S13: Raw and improved taxonomic assignment of ACCs, related to Fig. S3e. Table S14: ARG load index of the ACCs at different taxonomic levels. Table S15: Plasmid replicons and their associated ARGs shared between human and food/environment, related to Fig. 3d. Table S16: Main ARGs carried by the phages/prophages, related to Fig. S5d. Table S17: Statistics of ARG sequence types in different subgroups. Table S18: Sharing events in the ARG-sharing network, related to Fig. S6a. Table S19: Statistics of the ARGs in the ARG-sharing network, related to Fig. S6a. Table 20: Number of ARGs from different risk ranks. Table S21: Statistics of ARGs shared between human and both food and environmental samples. Table S22: Number of sharing events between human and three potential sources, related to Fig. 4c. Table S23: Prevalence of SGBs detected in the present study, related to Fig. 4d and Fig. S7. Table S24: Number of carbapenemase genes found in the ACCs, related to Fig. S8a. Table S25: Antimicrobial susceptibility testing of carbapenem-resistant strains. Table S26: Clusters of the isolate genomes and the MAGs, related to Fig. 5b. Table S27: Host and genetic environment of OXA-347 revealed in the 14,875 complete genomes, related to Fig. 5c

Additional file 3: Introduction to Dengfeng

Acknowledgements
Not applicable.

Authors' contributions
X.L., Y.Y.Z., D.W., Y.H., and B.K. designed and supervised the study; Y.F., X.L., and Y.H. planned the methodology; Y.F. wrote the draft; Y.W., X.L., D.W., Y.H., and B.K., edited and reviewed the draft and provided critical comments; J.Z., H.L., J.X., Z.L., M.W., Y.P., T.T., G.Y., Y.Z., J.L., G.Q., Y.M., and W.G. carried out the experiments and collected the samples; Y.F., M.Z., and A.L.T.Z.L. performed the data analysis and visualization of the analyzed data, and X.L., and B.K. acquired the funding. All of the authors read and approved the final manuscript.

Funding
This work was supported by the National Key Research and Development Program of China (2022YFC2303900) and the major projects of the National Natural Science Foundation of China (22193064).

Data availability
All of the sequencing data have been deposited in the National Microbiology Data Center under accession numbers NMDC10018630 (https://nmdc.cn/resource/genomics/project/detail/NMDC10018630) and NMDC10018667 (https://nmdc.cn/resource/genomics/project/detail/NMDC10018667).

Declarations

Ethics approval and consent to participate
This study was conducted in compliance with the recommendations of the Declaration of Helsinki and the Ethics Committee of the Institute of Infectious Disease Prevention and Control, Henan Center for Disease Control and Prevention (ethical approval number: 2015-YM-006–02).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]State Key Laboratory of Animal Nutrition and Feeding, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China. [2]National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China. [3]Institute of Infectious Disease Prevention and Control, Henan Center for Disease Control and Prevention, Zhengzhou 450016, China. [4]Dengfeng Center for Disease Control and Prevention, Dengfeng, Zhengzhou 452470, China. [5]School of Light Industry Science and Engineering, Beijing Technology and Business University, Beijing 100048, China. [6]School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan 250012, Shandong, China. [7]NHC Key Laboratory of Food Safety Risk Assessment, China National Center for Food Safety Risk Assessment, Beijing 100022, China.

Received: 11 June 2024 Accepted: 19 November 2024
Published online: 07 January 2025

References

1. Hernando-Amado S, Coque TM, Baquero F, Martínez JL. Defining and combating antibiotic resistance from One Health and Global Health perspectives. Nat Microbiol. 2019;4:1432–42.
2. Hu Y, Liu D, Jin X, Feng Y, Guo Y. Synthetic microbiome for a sustainable poultry industry. Innovation. 2023;4:100357.

3. Liu J-H, Liu Y-Y, Shen Y-B, Yang J, Walsh TR, Wang Y, et al. Plasmid-mediated colistin-resistance genes: *mcr*. Trends Microbiol. 2024;32:365–78.
4. Walsh TR, Weeks J, Livermore DM, Toleman MA. Dissemination of *NDM-1* positive bacteria in the New Delhi environment and its implications for human health: An environmental point prevalence study. Lancet Infect Dis. 2011;11:355–62.
5. Liu YY, Wang Y, Walsh TR, Yi LX, Zhang R, Spencer J, et al. Emergence of plasmid-mediated colistin resistance mechanism *MCR-1* in animals and human beings in China: A microbiological and molecular biological study. Lancet Infect Dis. 2016;16:161–8.
6. He T, Wang R, Liu D, Walsh TR, Zhang R, Lv Y, et al. Emergence of plasmid-mediated high-level tigecycline resistance genes in animals and humans. Nat Microbiol. 2019;4:1450–6.
7. Sun J, Chen C, Cui C-Y, Zhang Y, Liu X, Cui Z-H, et al. Plasmid-encoded *tet*(X) genes that confer high-level tigecycline resistance in *Escherichia coli*. Nat Microbiol. 2019;4:1457–64.
8. Ellabaan MMH, Munck C, Porse A, Imamovic L, Sommer MOA. Forecasting the dissemination of antibiotic resistance genes across bacterial genomes. Nat Commun. 2021;12:2435.
9. Djordjevic SP, Jarocki VM, Seemann T, Cummins ML, Watt AE, Drigo B, et al. Genomic surveillance for antimicrobial resistance-a One Health perspective. Nat Rev Genet. 2024;25:142–57.
10. Hu Y, Yang X, Qin J, Lu N, Cheng G, Wu N, et al. Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. Nat Commun. 2013;4:2151.
11 Wyrsch ER, Nesporova K, Tarabai H, Jamborova I, Bitar I, Literak I, et al. Urban wildlife crisis: Australian silver gull Is a bystander host to widespread clinical antibiotic resistance. eSystems. 2022;7:e0015822.
12. Valiatti TB, Bessa-Neto FO, Santos FF, Silva RGB, Veiga R, Cassu-Corsi D, et al. Clonal dissemination of highly virulent Serratia marcescens strains producing *KPC-2* in food-producing animals. One Health. 2023;17:100591.
13. Zhang Z, Zhang Q, Wang T, Xu N, Lu T, Hong W, et al. Assessment of global health risk of antibiotic resistance genes. Nat Commun. 2022;13:1553.
14. Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GGZ, Boling L, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun. 2014;5:4498.
15. Munk P, Knudsen BE, Lukjacenko O, Duarte ASR, Van Gompel L, Luiken REC, et al. Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries. Nat Microbiol. 2018;3:898–908.
16 Pan Y, Zeng J, Li L, Yang J, Tang Z, Xiong W, et al. Coexistence of antibiotic resistance genes and virulence factors deciphered by large-scale complete genome analysis. mSystems. 2020;5:e00821-19.
17. Baker M, Zhang X, Maciel-Guerra A, Dong Y, Wang W, Hu Y, et al. Machine learning and metagenomics reveal shared antimicrobial resistance profiles across multiple chicken farms and abattoirs in China. Nat Food. 2023;4:707–20.
18. Pehrsson EC, Tsukayama P, Patel S, Mejía-Bautista M, Sosa-Soto G, Navarrete KM, et al. Interconnected microbiomes and resistomes in low-income human habitats. Nature. 2016;533:212–6.
19. Debroas D, Siguret C. Viruses as key reservoirs of antibiotic resistance genes in the environment. ISME J. 2019;13:2856–67.
20. Zangenah S, Andersson AF, Özenci V, Bergman P. Genomic analysis reveals the presence of a class D beta-lactamase with broad substrate specificity in animal bite associated *Capnocytophaga* species. Eur J Clin Microbiol Infect Dis. 2017;36:657–62.
21. Ming D-S, Chen Q-Q, Chen X-T. Analysis of resistance genes in pan-resistant *Myroides odoratimimus* clinical strain PR63039 using whole genome sequencing. Microb Pathog. 2017;112:164–70.
22. Schlesinger DJ, Shoemaker NB, Salyers AA. Possible origins of CTnBST, a conjugative transposon found recently in a human colonic Bacteroides strain. Appl Environ Microbiol. 2007;73:4226–33.
23. Wang M, Xiong W, Liu P, Xie X, Zeng J, Sun Y, et al. Metagenomic insights into the contribution of phages to antibiotic resistance in water samples related to swine feedlot wastewater treatment. Front Microbiol. 2018;9:2474.
24. Wang C, Song Y, Tang N, Zhang G, Leclercq SO, Feng J. The shared resistome of human and pig microbiota is mobilized by distinct genetic elements. Appl Environ Microbiol. 2021;87:e01910-e1920.
25. Willmann M, El-Hadidi M, Huson DH, Schütz M, Weidenmaier C, Autenrieth IB, et al. Antibiotic selection pressure determination through sequence-based metagenomics. Antimicrob Agents Chemother. 2015;59:7335–45.
26. Hu Y, Gao GF, Zhu B. The antibiotic resistome: gene flow in environments, animals and human beings. Front Med. 2017;11:161–8.
27. Losasso C, Di Cesare A, Mastrorilli E, Patuzzi I, Cibin V, Eckert EM, et al. Assessing antimicrobial resistance gene load in vegan, vegetarian and omnivore human gut microbiota. Int J Antimicrob Agents. 2018;52:702–5.
28. Heaton JC, Jones K. Microbial contamination of fruit and vegetables and the behaviour of enteropathogens in the phyllosphere: A review. J Appl Microbiol. 2008;104:613–26.
29. Stelder JJ, Kjær LJ, Jensen LB, Boklund AE, Denwood M, Carlsen M, et al. Livestock-associated MRSA survival on house flies (*Musca domestica*) and stable flies (*Stomoxys calcitrans*) after removal from a Danish pig farm. Sci Rep. 2021;11:3527.
30. Olomu IN, Pena-Cortes LC, Long RA, Vyas A, Krichevskiy O, Luellwitz R, et al. Elimination of "kitome" and "splashome" contamination results in lack of detection of a unique placental microbiome. BMC Microbiol. 2020;20:157.
31. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
32. Blanco-Míguez A, Beghini F, Cumbo F, McIver LJ, Thompson KN, Zolfo M, et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. Nat Biotechnol. 2023;41:1633–44.
33. Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. Microbial strain-level population structure and genetic diversity from metagenomes. Genome Res. 2017;27:626–38.
34. Arango-Argoty G, Garner E, Pruden A, Heath LS, Vikesland P, Zhang L. DeepARG: A deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome. 2018;6:23.
35. Salazar C, Giménez M, Riera N, Parada A, Puig J, Galiana A, et al. Human microbiota drives hospital-associated antimicrobial resistance dissemination in the urban environment and mirrors patient case rates. Microbiome. 2022;10:208.
36. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27:2987–93.
37. 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. 2019;7:123.
38. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31:1674–6.
39. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7: e7359.
40 Wu YW, Simmons BA, Singer SW. MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2016;32:605–7.
41. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014;11:1144–6.
42. Song WZ, Thomas T. Binning-refiner: improving genome bins through the combination of different binning programs. Bioinformatics. 2017;33:1873–5.
43. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome. 2018;6:158.
44. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.
45. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017;11:2864–8.
46. von Meijenfeldt FAB, Arkhipova K, Cambuy DD, Coutinho FH, Dutilh BE. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. Genome Biol. 2019;20:217.

47 Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P, et al. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nat Commun. 2020;11:2500.
48. Pärnänen K, Karkman A, Hultman J, Lyra C, Bengtsson-Palme J, Larsson DGJ, et al. Maternal gut and breast milk microbiota affect infant gut antibiotic resistome and mobile genetic elements. Nat Commun. 2018;9:3891.
49. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2012;41:590–6.
50 Lomsadze A, Bonny C, Strozzi F, Borodovsky M. GeneMark-HM: Improving gene prediction in DNA sequences of human microbiome. NAR Genom Bioinform. 2021;3:lqab047.
51. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, et al. CARD 2023: Expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res. 2023;51:D690–9.
52. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, et al. Identification of mobile genetic elements with geNomad. Nat Biotechnol. 2023. https://doi.org/10.1038/s41587-023-01953-y.
53. Shang J, Jiang J, Sun Y. Bacteriophage classification for assembled contigs using graph convolutional network. Bioinformatics. 2021;37:i25-33.
54. Shang J, Sun Y. Predicting the hosts of prokaryotic viruses using GCN-based semi-supervised learning. BMC Biol. 2021;19:250.
55 Robertson J, Bessonov K, Schonfeld J, Nash JHE. Universal whole-sequence-based plasmid typing and its utility to prediction of host range and epidemiological surveillance. Microb Genom. 2020;6:mgen000435.
56. Camarillo-Guerrero LF, Almeida A, Rangel-Pineros G, Finn RD, Lawley TD. Massive expansion of human gut bacteriophage diversity. Cell. 2021;184:1098-1109.e9.
57. Schmartz GP, Hartung A, Hirsch P, Kern F, Fehlmann T, Müller R, et al. PLSDB: Advancing a comprehensive database of bacterial plasmids. Nucleic Acids Res. 2022;50:D273–8.
58. Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. Within-host evolution of bacterial pathogens. Nat Rev Microbiol. 2016;14:150–62.
59. Groussin M, Poyet M, Sistiaga A, Kearney SM, Moniz K, Noel M, et al. Elevated rates of horizontal gene transfer in the industrialized human microbiome. Cell. 2021;184:2053-2067.e18.
60. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol. 2021;39:578–85.
61. CLSI. Performance standards for antimicrobial susceptibility testing, 33rd edition. Clinical and Laboratory Standards Institute. 2023.
62. Xie Z, Tang H. ISEScan: Automated identification of insertion sequence elements in prokaryotic genomes. Bioinformatics. 2017;33:3340–7.
63 Wang X, Zhang H, Yu S, Li D, Gillings MR, Ren H, et al. Inter-plasmid transfer of antibiotic resistance genes accelerates antibiotic resistance in bacterial pathogens. ISME J. 2024;18:wrad032.
64. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13:e1005595.
65. Prokka ST. Rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.
66. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: Fast genome and metagenome distance estimation using MinHash. Genome Biol. 2016;17:132.
67. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.
68. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.
69. Conway JR, Lex A, Gehlenborg N. UpSetR: An R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017;33:2938–40.
70. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:6.
71. Zuguang Gu. Complex heatmap visualization iMeta. 2022;1: e43.
72 Xu S, Li L, Luo X, Chen M, Tang W, Zhan L, et al. Ggtree: a serialized data object for visualization of a phylogenetic tree and annotation data. iMeta. 2022;1:e56.
73 Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;1:gkab301.

Publisher's Note

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