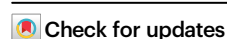


# Ecological connectivity of genomic markers of antimicrobial resistance in *Escherichia coli* in Hong Kong

Received: 27 November 2024

Accepted: 18 July 2025

Published online: 08 August 2025



Xiaoqing Xu<sup>1</sup>, Yunqi Lin<sup>1</sup>, Yu Deng<sup>1</sup>, Lei Liu<sup>1</sup>, Dou Wang<sup>1</sup>, Qinling Tang<sup>1</sup>, Chunxiao Wang<sup>1</sup>, Xi Chen<sup>1</sup>, You Che<sup>1</sup>, Ethan R. Wyrsh<sup>2</sup>, Veronica M. Jarocki<sup>2,3</sup>, Steven P. Djordjevic<sup>2</sup> & Tong Zhang<sup>1,4</sup>✉

Antibiotic-resistant *Escherichia coli* (*E. coli*) is a major contributor to the global burden of antimicrobial resistance (AMR). While the One Health concept emphasizes the connection of human, animal, and environmental health, genome-resolved and quantitatively integrated analyses of microbial exchange across ecological compartments remain limited. Here we show that *E. coli* populations from urban aquatic ecosystems in Hong Kong, representing human, animal, and environmental sources, exhibit close genetic relatedness. Using Nanopore long-read sequencing, we generated near-complete genomes for 1016 *E. coli* isolates collected over one year. These isolates encompassed all main phylogroups, 223 sequence types, 141 antibiotic resistance gene subtypes, and 2647 circular plasmids. 142 clonal strain-sharing events were detected between human-associated and environmental water samples. Additionally, 195 plasmids were shared across all three source-attributed sectors. Conjugation assays confirmed that several plasmids were functionally transmissible across ecological boundaries. To quantify these patterns, we established a genomic framework integrating sequence type similarity, genetic relatedness, and clonal sharing to assess ecological connectivity. Our results indicate that ecological connectivity may facilitate AMR dissemination, highlighting the importance of integrated strategies to monitor and manage resistance risks across sectors within the One Health framework.

Antimicrobial resistance (AMR) is one of the most significant threats to global public health, undermining the efficacy of existing antibiotics and complicating the treatment of bacterial infection<sup>1,2</sup>. Forecasts indicate that AMR could cause 1.91 million attributable and 8.22 million associated deaths annually by 2050<sup>3,4</sup>. Tackling these escalating threats requires the One Health approach that acknowledges the deep interconnections between human, animal, and environmental health<sup>5–8</sup>. However, major knowledge gaps remain regarding the extent and

mechanisms of genetic connectivity across these sectors and how they facilitate the dissemination of resistance. Without a clear understanding of these cross-sectoral linkages, efforts to mitigate AMR are likely to remain incomplete and less effective.

Aquatic environments, particularly wastewater systems, are increasingly recognized as key reservoirs for antibiotic resistance genes (ARGs) and resistant bacteria<sup>9,10</sup>. Wastewater-based epidemiology (WBE), also generally known as wastewater and environmental

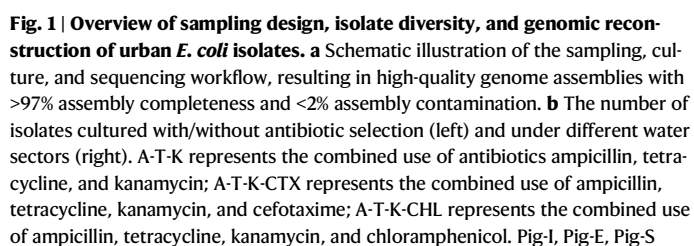
<sup>1</sup>Environmental Microbiome Engineering and Biotechnology Laboratory, Department of Civil Engineering, Center for Environmental Engineering Research, The University of Hong Kong, Hong Kong SAR, China. <sup>2</sup>Australian Institute for Microbiology and Infection, University of Technology Sydney, Sydney, NSW, Australia. <sup>3</sup>Cooperative Research Centre for Solving Antimicrobial Resistance in Agribusiness, Food and Environments, Adelaide, SA, Australia. <sup>4</sup>School of Public Health, The University of Hong Kong, Hong Kong SAR, China. ✉e-mail: [zhangt@hku.hk](mailto:zhangt@hku.hk)

In this study, we utilized the Nanopore R10.4.1 platform to generate 1016 high quality, near-complete genomes of *E. coli* isolates collected over one year from diverse aquatic environments across Hong Kong. Hong Kong, with its high population density and well-

Leveraging this city-scale genomic dataset, we systematically characterized genetic diversity, resistance profiles, and mobile genetic elements, particularly plasmids, across human-associated, animal-associated, and environmental sectors. Phylogenomic and pangenomic analyses revealed extensive strain sharing and resistome overlap across compartments. To quantify cross-sectoral ecological connectivity, we developed a multi-dimensional framework incorporating strain-sharing ratios, genetic similarity, and normalized core-genome distances. We further identified and experimentally validated the cross-sectoral transfer of high-risk ARGs and plasmids. Together, these findings underscore ecological connectivity as a potential contributor to AMR dissemination and provide actionable genomic insights to support integrated One Health surveillance and intervention strategies.

### *E. coli* genotypes and clinically important lineages are widely distributed across urban aquatic ecosystems

We analyzed 1016 high-quality *E. coli* genomes collected from 63 aquatic sampling sites across Hong Kong, covering human-associated ( $n=440$ ), animal-associated ( $n=194$ ), and environmental ( $n=382$ ) sources (Fig. 1a). Among these, 601 isolates were recovered under antibiotic selection and 415 were obtained without antibiotic exposure (Fig. 1b). All genomes exhibited >97% completeness and <2% contamination (Supplementary Data 1). Phylogenomic analysis revealed



represent influent, effluent, and sludge samples from the pig farm, respectively; WWTP-I, WWTP-E, WWTP-S represent influent, effluent, and activated sludge samples from wastewater treatment plants (WWTPs). **c** Maximum likelihood phylogeny based on core-genome alignment of 1016 *E. coli* genomes. Each tip in the tree represents a genome, and its phylogroup is indicated by the color of the dot. Inner ring: Sequence Types (STs, only STs with a minimum of 10 isolates are shown); Second ring: sampling source; Third ring: Antibiotic types; Dot: Sampling sites in WWTPs or pig farms; Outer ring: Ecological sectors. Source data are provided as a Source Data file.

broad genetic diversity across all major *E. coli* phylogroups. Phylogroups A and B1 were dominant across all three sectors. In human-associated and environmental samples, Phylogroup A was slightly more prevalent than B1 (42% vs. 29% and 38% vs. 35%, respectively), whereas in the animal-associated sector, B1 was more common (48% vs. 40%). Phylogroups B2, D, and F, which are often associated with extraintestinal infections, were more frequently identified in human-associated and environmental samples but were rare in animal-associated isolates. Only one isolate from Phylogroup E or clade I was detected among the animal-associated samples (Fig. 1c).

Sequence typing identified 899 isolates with known sequence types (STs). The most prevalent STs included ST457 (6%), ST10 (4%), and ST216 (4%). Specifically, ST457 and ST216 were predominant in human-associated samples, and ST457 was also frequently detected in environmental waters. ST5229 (15%) and ST101 (13%) were most common in animal-associated sources. The environmental sector showed the greatest ST diversity, comprising 122 distinct STs among 382 isolates, followed by the human-associated sector with 126 STs among 440 isolates and the animal-associated sector with 58 STs among 194 isolates. In total, 223 unique STs were identified across the dataset, with 117 isolates remaining untyped. Clinically significant extra-intestinal pathogenic *E. coli* (ExPEC) lineages, including ST69, ST73, ST95, and ST131, were widely detected. ST69, ST73, and ST95 were mainly found in human-associated and environmental water sources. Notably, ST131, a globally disseminated and multidrug-resistant lineage, was recovered from all three sectors, underscoring its ecological adaptability and potential for cross-sectoral dissemination (Supplementary Data 1).

### Widespread and source-specific resistance profiles shape the urban *E. coli* resistome landscape

We assessed antibiotic resistance phenotypes in 1016 *E. coli* isolates against 11 antibiotics across six drug classes (Supplementary Data 2). Among isolates recovered without antibiotic selection, high proportions of resistance were observed to amoxicillin (58%), ampicillin (56%), tetracycline (56%), and ciprofloxacin (52%) (Supplementary Fig. 1a). As expected, isolates recovered under antibiotic selection exhibited enrichment of multidrug resistance profiles, with each isolate resistant to an average of about 9 antibiotics compared to 4–6 in the non-selective group (Supplementary Fig. 1b–d). Notably, among the 1016 *E. coli* isolates, 0.9% exhibited resistance to meropenem.

Isolates from the animal-associated sector, particularly from the pig farm, exhibited the highest resistance levels, potentially reflecting localized selection pressures. Fishery water isolates, despite being collected without antibiotic selection, also showed substantial resistance to antibiotics commonly used in clinical and veterinary settings, including ampicillin, amoxicillin, and cefazolin. In contrast, isolates from marine water samples exhibited the lowest overall resistance levels. Bathing beach isolates showed elevated resistance to trimethoprim-sulfamethoxazole (SXT), norfloxacin, and ciprofloxacin (Fig. 2a). These findings highlight the risk of human exposure to resistant bacteria through recreational contact with contaminated water or aquatic food consumption. Consistent with expectations, isolates recovered under antibiotic selection exhibited higher overall resistance levels and a broader spectrum of resistance (Fig. 2b), indicating selective enrichment and the potential for environmental persistence of multidrug-resistant traits.

Genotypic profile mirrored these patterns. Isolates collected under antibiotic selection pressure harbored more ARGs per genome (mean: 23–26) than isolates collected without selection (mean: 10–17) (Fig. 2c, d). Animal-associated isolates, particularly from wastewater influent of the pig farm (Pig-I), carried the highest ARG burdens and exhibited distinct ARG enrichment patterns, although this reflects a single-site ecological setting. In contrast, isolates from activated

sludge of WWTPs (WWTP-S), bathing beaches, and marine sources recovered without selection carried fewer ARGs, reflecting lower baseline resistome content in these sources. Overall, we identified 141 ARG subtypes across 15 ARG types (classes), including widespread beta-lactamases (e.g., *bla*<sub>ampC</sub>, *bla*<sub>TEM-1</sub>), tetracycline resistance genes (*tet*(A)), aminoglycoside resistance genes (*aph*(3')-Ia), and *floR*, many of which were plasmid-encoded (Fig. 3).

Clinically important resistance genes were frequently detected, including 46 beta-lactamase ARG subtypes (covering ESBLs and carbapenemases), 13 quinolone ARG subtypes, 6 colistin ARG subtypes, 6 tetracycline ARG subtypes (including *tet*(X4)), and 12 SXT ARG subtypes. Notably, three *bla*<sub>NDM</sub>-producing isolates, conferring resistance to carbapenems, were found to co-harbor either *tet*(X4), associated with tigecycline resistance, or *mcr* genes, associated with colistin resistance. This co-occurrence suggests convergence of resistance to multiple last-resort antibiotics within individual strains. While resistance phenotypes generally correlated with corresponding ARGs, we also observed resistance to meropenem in isolates lacking known resistance genes (Supplementary Fig. 2), such as *bla*<sub>NDM</sub>, indicating the possible involvement of uncharacterized or novel mechanisms.

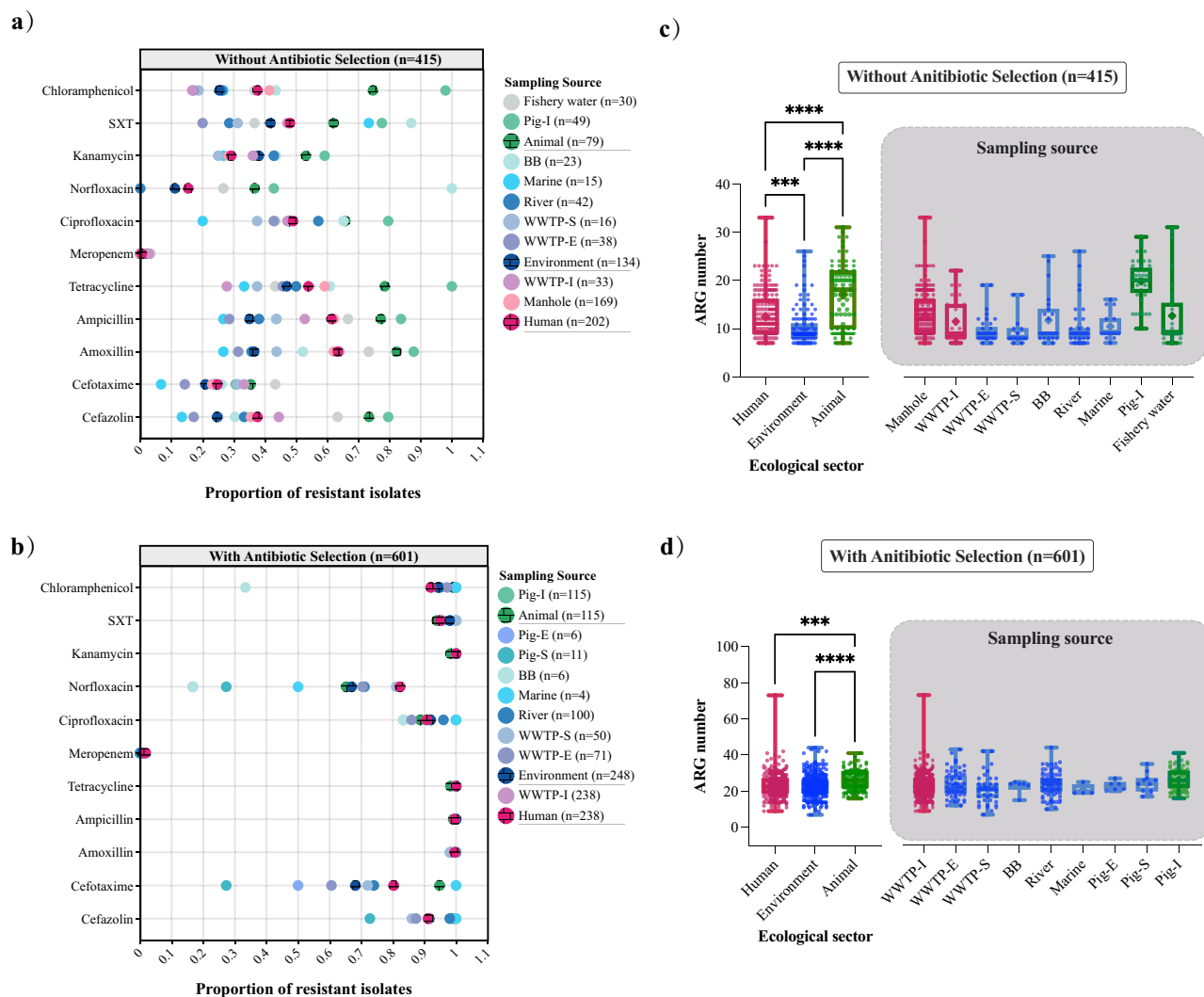
Notably, resistance profiles from human-associated and environmental isolates were more similar overall, with major differences driven by the distinct ARG composition observed in isolates from Pig-I (Fig. 3). Under non-selective conditions, Pig-I isolates exhibited distinct resistome profiles, including enrichment of *bla*<sub>TEM-1</sub>, *floR*, *tet*(A), and *tet*(M). Under antibiotic selection, resistome compositions became more homogeneous across sources. For example, isolates from wastewater effluent of the pig farm (Pig-E) and wastewater sludge of the pig farm (Pig-S) clustered more closely, indicating that antibiotic exposure promotes convergence of resistance profiles across sectors.

### Plasmid diversity and replicon-driven structuring of resistance and virulence across ecological sectors

Following the criteria described in Methods (Supplementary Fig. 3), we identified 3522 plasmid sequences from 1016 *E. coli* isolates, including 2647 circular plasmids, which were obtained from 957 isolates. Circular plasmid sizes ranged from 1531 to 382,796 bp. Based on mobility classification, 43% were conjugative, 33% mobilizable, and 24% non-mobilizable (Supplementary Data 3). Notably, 1126 plasmids (including 556 circular plasmids) lacked identifiable replicon types, which may reflect either previously uncharacterized plasmid lineages or limitations of current replicon reference databases in capturing environmental plasmid diversity.

Among 2396 plasmids with known replicons, 59 replicon types were detected, and 45% of these plasmids contained multiple replicons. As illustrated in Fig. 4, the five most prevalent replicons across all sectors were IncFIB(AP001918), IncX1, IncFII, IncFIC(FII), and IncII. These are primarily circular, conjugative, and low-copy plasmids and were frequently associated with aminoglycoside resistance genes. Most were identified in isolates collected under antibiotic selection pressure. Virulence factors (VFs), totaling 88,027 genes, were detected across 1016 isolates. Approximately 91% of these were chromosomally encoded and showed minimal variation across sectors (Supplementary Fig. 4). Notably, 29 plasmids carried VF genes, particularly those containing IncFIB(AP001918), IncFII, and IncX1 replicons, indicating a potential dual role in resistance and virulence dissemination.

Isolates from the animal-associated sector carried more circular plasmids per isolate genome, with a median of four, compared to a median of two in both the human-associated and environmental sectors. Similarly, the number of ARG-carrying plasmids per isolate was also higher in animal-associated isolates (Supplementary Fig. 5). Plasmid length distributions were bimodal, with peaks around 10 kb and 100 kb, and plasmids from the same source tended to show more similar size profiles than those from different sources (Supplementary Fig. 6).



**Fig. 2 | Antibiotic resistance phenotypes and ARG burdens in *E. coli* across ecological sectors and sampling sources.** Proportion of resistant isolates to 11 antibiotics across sampling sectors and sources without (a) and with (b) antibiotic selection, as determined by minimum inhibitory concentration (MIC) assays. Antibiotics tested include ampicillin, amoxicillin, cefazolin, cefotaxime, meropenem, tetracycline, ciprofloxacin, norfloxacin, kanamycin, chloramphenicol, and trimethoprim-sulfamethoxazole (SXT). BB represents bathing beach samples. The numbers in parentheses next to each sampling source represent the number of isolates recovered from that source. ARG number per *E. coli* isolate across ecological sectors and sampling sources without (c) and with (d) antibiotic selection. The

box plots display the minimum, 25th percentile, median (indicated by the line inside the box), 75th percentile, and maximum values for each group. The mean is represented by the “+” sign. Outliers are plotted as individual points. Statistical significance was assessed using two-way ANOVA with post-hoc comparisons, with *p*-values reported for pairwise comparisons: Environment vs. Animal ( $<1 \times 10^{-10}$ ), Human vs. Animal ( $<1 \times 10^{-10}$ ), and Human vs. Environment (0.00012) for (c); Environment vs. Animal ( $7.5 \times 10^{-5}$ ), Human vs. Animal ( $9.8 \times 10^{-4}$ ), and Human vs. Environment (0.73) for (d). Significance is indicated as follows: \*\*\* $p < 0.001$ ; \*\*\*\* $p < 0.0001$ . No significant differences were observed for Human vs. Environment in (d). Source data are provided as a Source Data file.

Replicon-based clustering showed that isolates from human-associated and environmental sectors were more similar to each other, forming a distinct cluster separate from the animal-associated sector. Within-sector clustering was also observed, for example, isolates from Pig-E and Pig-S, as well as from influent of WWTPs (WWTP-I) and effluent of WWTPs (WWTP-E), were highly similar. However, isolates from Pig-I displayed greater replicon diversity. Certain replicons showed higher prevalence in specific sources, for instance, the IncQ1 replicon was particularly prevalent in isolates from the Pig-E source (Fig. 4).

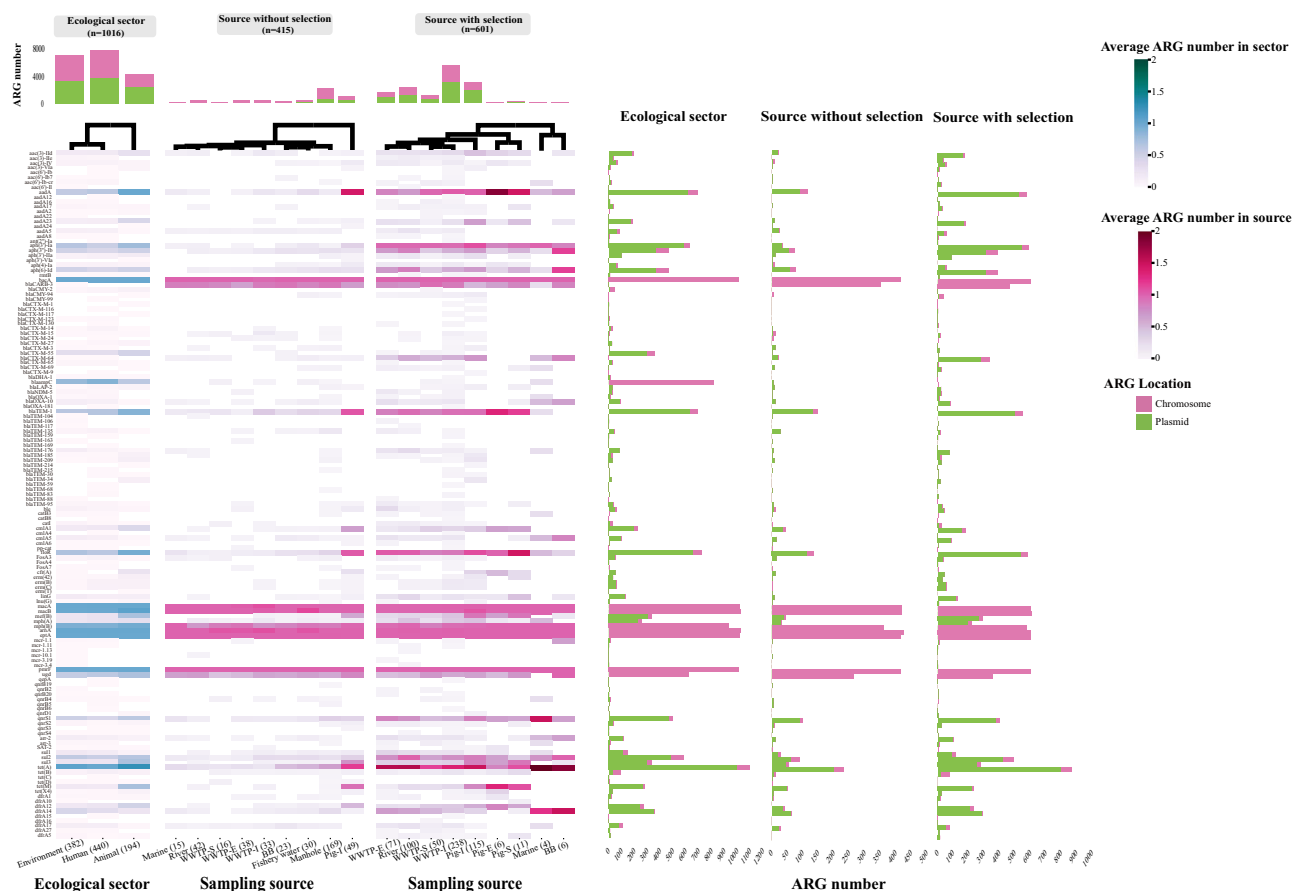
### Strain-level evidence reveals strong connectivity between human-associated and environmental *E. coli* populations

To investigate strain-level connectivity across ecological sectors, we first assessed the distribution of phylogroups and STs among 1016 *E. coli* genomes. All major phylogroups were detected in human-

associated, animal-associated, and environmental sectors, though relative abundance varied (Fig. 5a). Among the 223 identified STs, 27 were shared between human-associated and environmental samples, while 20 were found across all three sectors. Shared STs between human-associated and environmental sectors were not limited to different sampling sites within WWTPs but also occurred in other natural water environments and upstream human-associated inputs (Supplementary Fig. 7). In contrast, fewer STs were shared between human-associated and animal-associated water sectors (5 STs), and between environmental and animal-associated water sectors (6 STs) (Fig. 5b).

We further applied core-genome multi-locus sequence typing (cgMLST) to identify clonal relationships among isolates, finding that isolates from the same sector were generally more closely related than those between sectors (Fig. 5c), a pattern also supported by cgSNP analysis (Supplementary Fig. 8). Using a cgMLST allele difference threshold of fewer than 10, we identified 142 high-confidence strain-





**Fig. 3 | Distribution and genomic localization of ARG subtypes across ecological sectors and sampling sources.** This figure illustrates the distribution patterns of antibiotic resistance gene (ARG) subtypes across three dimensions: ecological sectors (left), sampling sources without antibiotic selection (middle), and sources with antibiotic selection (right). Each heatmap shows the average number of ARG subtypes detected per isolate within the respective group. Rows represent individual ARG subtypes, and color intensity reflects their average abundance. The hierarchical clustering was applied to both ARG subtypes and

sampling groups based on Bray–Curtis dissimilarity to highlight co-occurrence patterns and compositional similarities. Vertical bar plots above each heatmap summarize the total number of ARGs detected per ecological sector or sampling source. Horizontal bar plots to the right of each heatmap indicate the total number of detections for each ARG subtype across all three panels (ecological sectors, sources without selection, and sources with selection). Bar colors indicate the genomic location of the ARGs (green: plasmid-borne; pink: chromosomal). Source data are provided as a Source Data file.

sharing pairs involving 48 unique isolates, all of which were shared between human-associated and environmental sectors (Fig. 5d). These strain-sharing pairs were predominantly observed between WWTP influents and river water (48%), as well as between influent and sludge samples (39%), suggesting frequent transmission between human-associated wastewater and downstream environments. The 48 involved isolates corresponded to four STs: ST457 (50%), ST2003 (31%), ST127 (13%), and ST1638 (6%), corresponding to phylogroups F, D, B2, and A, respectively (Supplementary Data 4).

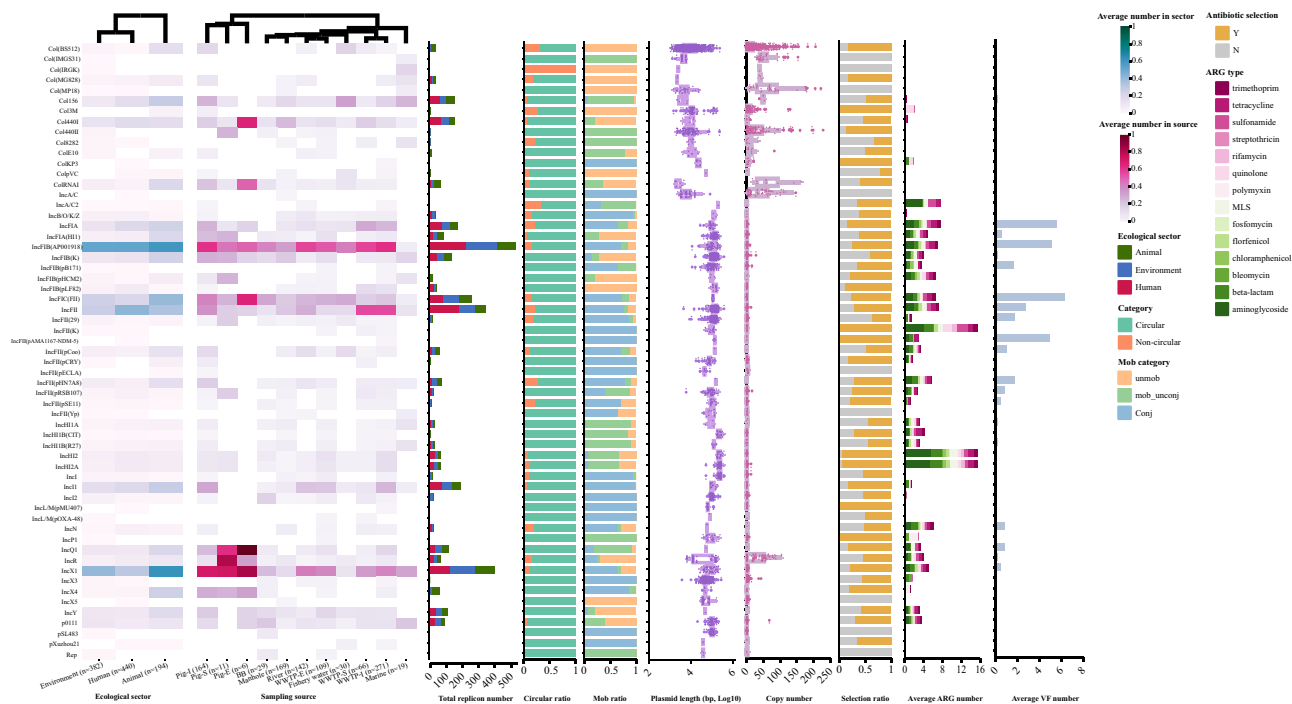
To evaluate the extent of functional relatedness, we calculated the Jaccard Index (JI) for both pangenomes and resistomes among sharing pairs. Pangenome similarity ranged from 0.89 to 0.98, with 6 pairs (4%) achieving nearly identical gene content (JI = 0.98; Fig. 5e). Resistome similarity ranged from 0.64 to 1, with 56% of sharing pairs carrying identical ARG sets (Fig. 5f). The results provide strong genomic evidence for intersectoral *E. coli* transmission and shared resistance determinants between human and environmental compartments.

### Cross-sectoral sharing and horizontal gene transfer of epidemiologically important ARGs among *E. coli* isolates

Horizontal gene transfer plays a critical role in the dissemination of ARGs, enabling the spread of high-risk resistance determinants across ecological compartments. In particular, third-generation

cephalosporin-resistant and fluoroquinolone-resistant *E. coli* strains are major contributors to both hospital- and community-acquired infections worldwide. Here, we further investigated the prevalence, mobility, and cross-sectoral sharing of both *bla*<sub>CTX-M</sub> genes, which confer resistance to third-generation cephalosporins, and fluoroquinolone resistance genes among urban waterborne *E. coli* populations.

We identified a total of 482 *bla*<sub>CTX-M</sub> genes spanning 18 subtypes, the majority of which were plasmid-encoded (Fig. 6a). Specifically, 13 subtypes were identified in isolates from human-associated sources, 10 in animal-associated sources, and 10 in environmental sources. The predominant subtype was *bla*<sub>CTX-M-55</sub>, which accounted for 353 genes (73%). Six subtypes, including *bla*<sub>CTX-M-55</sub> (73%), *bla*<sub>CTX-M-65</sub> (7%), *bla*<sub>CTX-M-14</sub> (7%), *bla*<sub>CTX-M-27</sub> (5%), *bla*<sub>CTX-M-15</sub> (2%), and *bla*<sub>CTX-M-226</sub> (1%), were shared across all three sectors. Additionally, three subtypes, including *bla*<sub>CTX-M-3</sub> (1%), *bla*<sub>CTX-M-123</sub> (1%), and *bla*<sub>CTX-M-64</sub> (0.4%), were shared between human-associated and environmental sectors (Fig. 6b, Supplementary Data 5). Among the 142 clonal sharing pairs previously identified, 135 pairs involved *bla*<sub>CTX-M-55</sub>, indicating a significant role in the cross-sectoral spread of third-generation cephalosporin resistance. The absence of other shared genes in these pairs highlights the prominence of *bla*<sub>CTX-M-55</sub> in the dataset and suggests its increased potential for horizontal gene transfer, emphasizing the urgent need for



**Fig. 4 | Plasmid replicon distribution and associated genomic features across 1016 *E. coli* isolates from urban aquatic environments.** The heatmap depicts the average number of replicons per isolate in different ecological sectors (left) and different sources (right). The hierarchical clustering was applied to replicons and sampling groups to reveal patterns of co-distribution. The following information is sequentially displayed for each detected replicon: total number of detected replicons, plasmid circular ratio, plasmid mobility, plasmid length, copy number, presence or absence of antibiotic selection, and the average number of carried ARGs

and VFs. The box plots display the interquartile range (IQR), with the 25th and 75th percentiles marking the lower and upper bounds of the box, respectively. The median is represented by the line inside the box. Whiskers extend to the minimum and maximum values within 1.5 times the IQR. Outliers are shown as individual points beyond the whiskers. In the ARG type legend, MLS refers to the macrolide-lincosamide-streptogramin class of antibiotics. Source data are provided as a Source Data file.

targeted surveillance and interventions to curb its spread in both clinical and environmental settings.

In the 470 shared genes, 388 (83%) were plasmid-borne, including 259 (67%) on conjugative plasmids, 54 (14%) on mobilizable plasmids, and 75 (19%) on non-mobilizable plasmids. Additionally, 82 genes (17%) were located on chromosomes. Interestingly, not all genes were on mobile plasmids. Therefore, we further examined the adjacent sequences of these genes for the presence of insertion sequences (IS), which have been shown to play a crucial role in facilitating gene transfer. We identified 30 IS elements across 18 ARG subtypes, with *ISEcp1* and *IS26* being associated with the highest number of ARG subtypes, carrying 15 and 8 subtypes, respectively (Fig. 6c).

We also identified 599 fluoroquinolone resistance genes belonging to 13 subtypes across 448 isolates, with the majority located on plasmids (Fig. 6d). The most frequently detected subtype was *qnrS1*, with 501 genes identified (84%). Four subtypes (including 501 genes of *qnrS1*, 34 of *qnrS2*, 11 of *qnrB4*, and 3 of *qnrS4*) were shared across all three sectors, while two subtypes (including 25 genes of *aac(6)-Ib-cr* and 3 of *qnrB20*) were shared between environmental and human-associated sectors. Only one subtype, *qnrS3*, was shared between animal- and human-associated sectors (Fig. 6e).

Notably, there is no pair in the 142 clonal sharing pairs having shared fluoroquinolone resistance genes, although these genes were widely shared. The majority of these genes (564, 94%) were plasmid-encoded, with 306 (51%) located on conjugative plasmids. Further analysis of IS elements in the flanking sequences of non-conjugative plasmids and chromosomal carriers revealed 31 distinct IS types. *IS26* was the most common, being associated with 6 subtypes, followed by *IS15DI*, an allele of *IS26*<sup>39</sup>, which was linked to 5 subtypes. Certain individual IS elements exhibited high prevalence in some specific

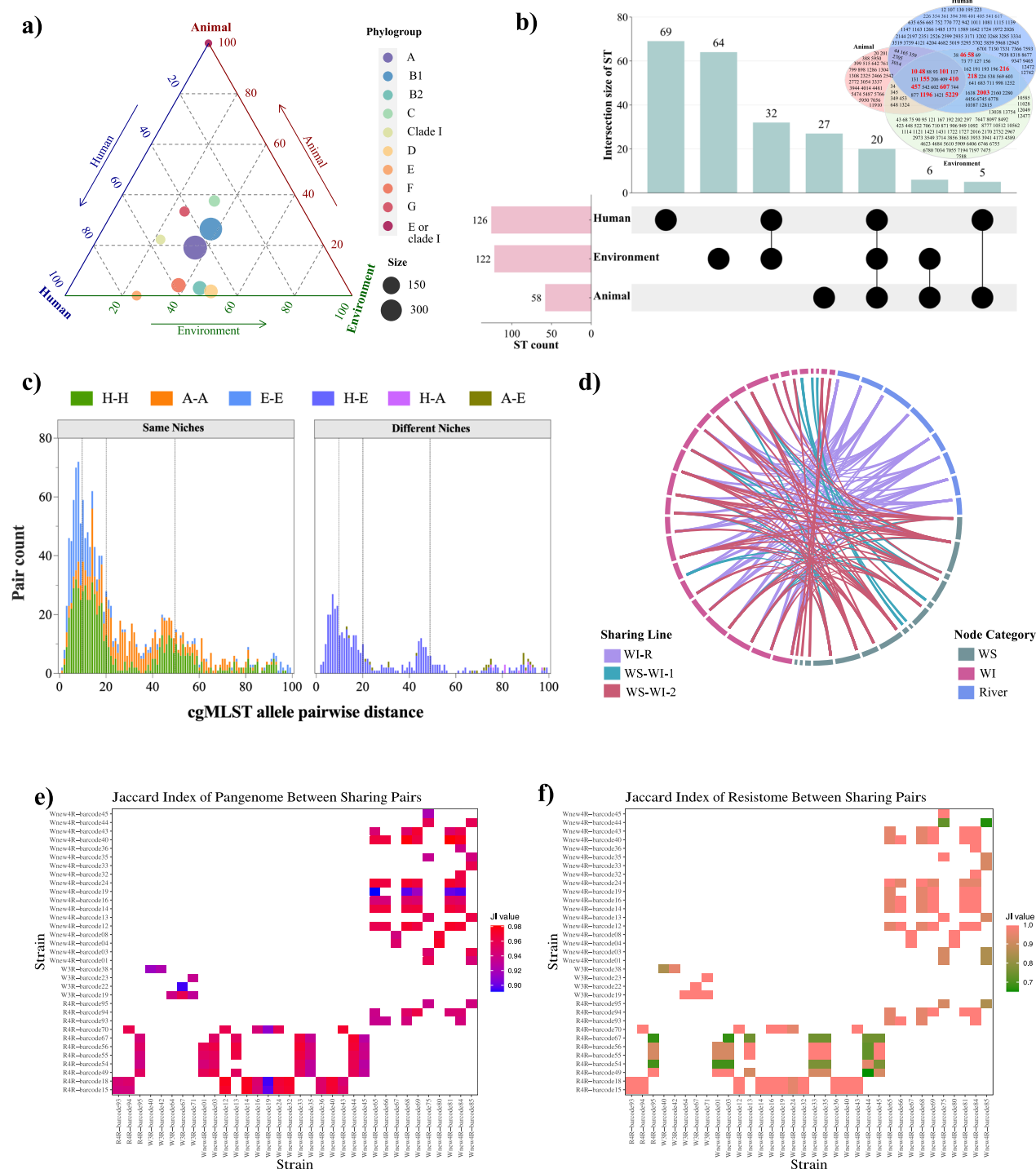
subtypes, for instance, *ISKpn19* was almost ubiquitously detected in the flanking sequences of the *qnrS3* gene (Fig. 6f).

### Global distribution and experimental validation of plasmid-mediated ARG transfer across aquatic ecosystems

To further investigate plasmid distribution and sharing across different sectors, we analyzed 2647 circular plasmids, categorizing them into 1516 groups based on pairwise similarity. Notably, 402 groups (27%) contained multiple plasmids, demonstrating the diversity of plasmids across sectors. Among these, 23 groups consisting of 195 plasmids (7% of 2647) were shared across all three sectors (Fig. 7a). The majority of these shared plasmids contained *IncX1* or *IncQ1* replicons, and were distributed across all sectors: 30% in human-associated, 32% in animal-associated, and 38% in environmental sector (Fig. 7b). The hosts of these plasmids were distributed across various phylogroups, primarily phylogroups A and B1. These plasmids were predominantly conjugative (67.7%) or mobilizable (27.7%), with a small fraction being non-mobilizable (4.6%).

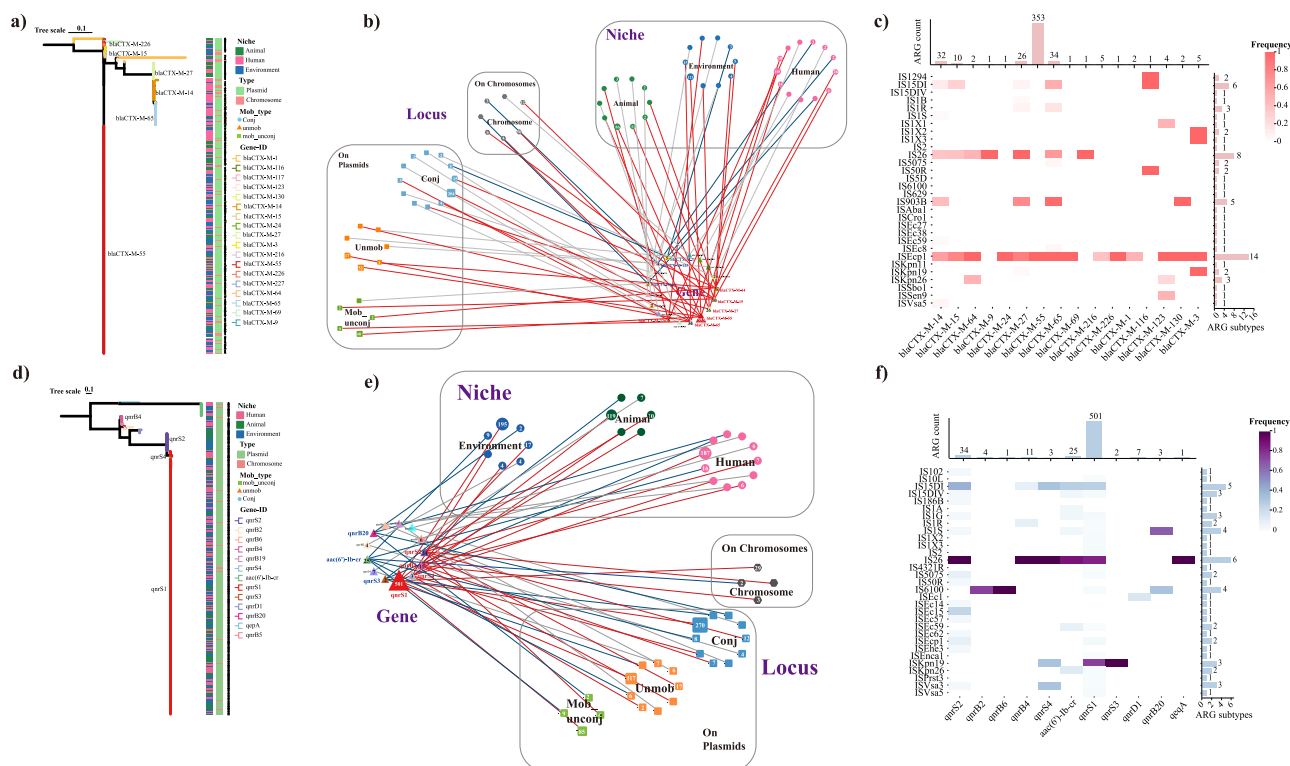
Of particular concern, 9 groups encompassing 115 low-copy plasmids carried ARGs, of which 101 plasmids carrying multiple ARGs, totaling 776 ARGs. In total, 464 plasmids were shared between two sectors, with 56% being conjugative plasmids (Supplementary Fig. 9). These included 19 groups with 85 plasmids shared between human-associated and animal-associated water sectors, 38 groups with 257 plasmids shared between human-associated and environmental water sectors, and 17 groups with 122 plasmids shared between environmental and animal-associated water sectors (Supplementary Data 3). These findings underscore the significant role of plasmids in facilitating the spread of ARGs across ecological boundaries.

To experimentally validate the cross-sector transmission potential, we conducted 108 groups of conjugation transfer experiments



**Fig. 5 | Cross-sectoral sharing and genomic similarity of *E. coli* isolates across aquatic environments.** **a** Ternary plot illustrating the distribution of *E. coli* phylogroups across human-associated, animal-associated, and environmental sectors. Each dot represents a phylogroup detected in a specific ecological sector combination, with dot color indicating the phylogroup and dot size proportional to the number of isolates. **b** Shared STs across ecological sectors. The UpSet and Venn diagrams represent the number and diversity of STs identified in different ecological sectors. **c** The frequency distribution of pairwise cgMLST distances among isolates from the same sector (left) and from different sectors (right). Colors indicate the sector combination for each pair. Only pairs that differ by fewer than 100 cgMLST loci are shown. The vertical dashed black line indicates the sharing

threshold (10, 20, and 50 cgMLST alleles). H, human; A, animal; E, environment. **d** Chord diagram representing the origin of 142 high-confidence strain-sharing pairs ( $\leq 10$  cgMLST allele differences). Connections indicate the source of shared pairs between influents and rivers (WI-R), within the same WWTP (WS-WI-1), or across different WWTPs (WS-WI-2). **e**, **f** Functional similarity among strain-sharing pairs. The Jaccard Index (JI) of the pangenome (**e**) and resistome (**f**) was calculated for each pair to assess genome-wide and ARG-level similarity, respectively. Matrix cells are colored by Jaccard similarity score, and the shared ST of each pair is indicated along the axes. Labels on the axes represent the names of isolates (also referred to as strains). Source data are provided as a Source Data file.



**Fig. 6 | Cross-sectoral dissemination and genomic context of high-risk ARGs.** **a, d** Phylogenetic trees of third-generation cephalosporin resistance *bla*<sub>CTX-M</sub> genes and quinolone resistance genes reconstructed from protein sequences. Genes transferred across all three sectors are annotated on the tree with their subtype name. The sector of origin, genomic locus (chromosome or plasmid), and plasmid mobility classification are shown alongside each tip. The branches of the tree are colored according to ARG subtype. **b, e** Bipartite networks linking each ARG subtype to its ecological sector (human-associated, animal-associated, environmental)

and genomic location. Red edges represent genes shared across all three sectors; blue edges indicate genes shared across two sectors. Node size reflects the number of observations, and node labels are shown when counts exceed one. **c, f** The frequency of insertion sequence (IS) detected adjacent to subtypes. The frequency is calculated as the ratio of genes with detected IS upstream or downstream to the total number of genes. The ARG count bar graph displays the number of genes for each subtype. The ARG subtype bar graph shows the number of IS-associated subtypes. Source data are provided as a Source Data file.

with donor and recipient strains from different sectors. The results confirmed that plasmids could transfer across sectors, with transfer efficiencies varying by the plasmid source and the recipient. Plasmids of the isolates from the human-associated sector transferred to the isolates from environmental and animal-associated sectors with efficiencies ranging from  $5.71 \times 10^{-9}$  to  $9.84 \times 10^{-3}$  and  $1.79 \times 10^{-8}$  to  $8.33 \times 10^{-3}$ , respectively; plasmids of the isolates from the environmental sector transferred to the isolates from the animal-associated and human-associated sectors with efficiencies ranging from  $1.82 \times 10^{-9}$  to  $8.50 \times 10^{-3}$  and  $8.90 \times 10^{-9}$  to  $9.07 \times 10^{-3}$ , respectively; and plasmids of the isolates from the animal-associated sector transferred to the isolates from the human-associated and environmental sectors with efficiencies ranging from  $7.07 \times 10^{-10}$  to  $2.66 \times 10^{-5}$  and  $2.00 \times 10^{-8}$  to  $8.37 \times 10^{-3}$ , respectively (Fig. 7c, Supplementary Data 6).

Furthermore, we investigated the global distribution of these plasmids by comparing them to complete plasmids in the IMG/PR database (Supplementary Data 7). Notably, Group P7, which comprises 14 IncII-type conjugative plasmids carrying *bla*<sub>CTX-M-55</sub>, exhibited high similarity to plasmids found in isolates from seven different countries (Fig. 7d, Supplementary Data 8). This underscores the global significance of these plasmids and their potential for widespread dissemination in both clinical and environmental contexts.

## Discussion

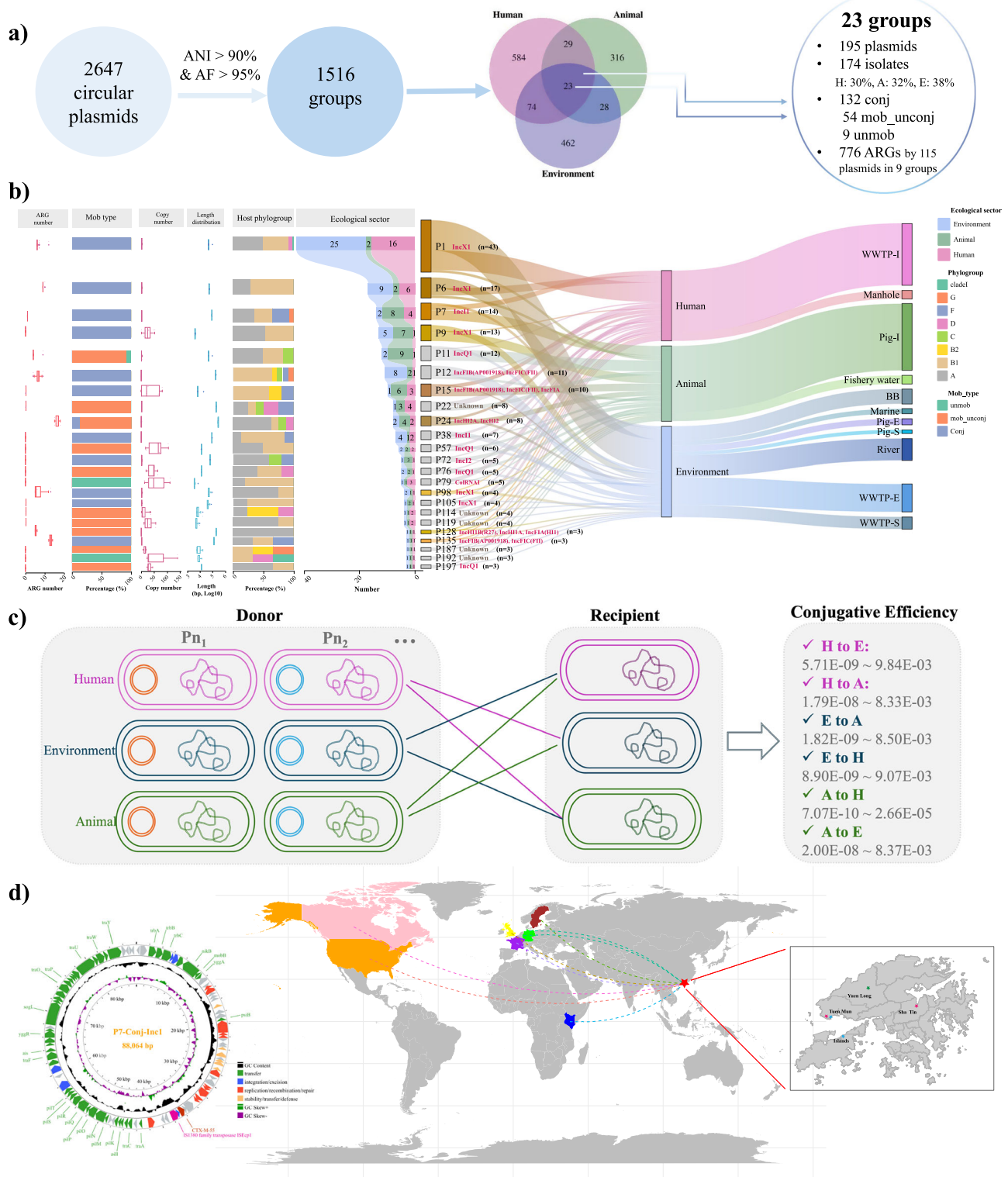
In this study, we provide multi-dimensional evidence that the isolates from human-associated, animal-associated, and environmental sectors are genetically interconnected, highlighting the ecological

connectivity in AMR dissemination. By leveraging city-scale *E. coli* genomic investigation in Hong Kong, a densely populated setting, we provide genome-scale evidence of measurable genetic connectivity of AMR across human-, animal-, and environmental-associated sectors, rather than confinement to one of the sectors.

Using the Nanopore long-read sequencing, we generated high-accuracy assemblies that enabled robust genomic comparisons across ecological compartments. Compared to the ONT R9.4.1 platform, R10.4.1 showed a substantial improvement in raw data accuracy, achieving nearly 99% modal read accuracy (Supplementary Data 9, Supplementary Fig. 10). Critically, Nanopore-only assemblies exhibited low indel errors and high consistency with hybrid assemblies generated using both Nanopore and Illumina data, indicating that long-read-only sequencing is reliable for comparative analyses of microbial populations. Importantly, the ability to reconstruct complete plasmids and resolve the chromosomal or extrachromosomal context of ARGs contributed to higher-resolution insights into the ecological connectivity. Furthermore, reliable assemblies were obtained at  $\sim 30\times$  coverage, illustrating the potential utility of this approach for high-throughput and rapid AMR surveillance (Supplementary Fig. 11).

Our dataset, comprising 1016 high-quality *E. coli* genomes across all major *E. coli* phylogroups, significantly expands the genomic landscape of *E. coli*. To contextualize these genomes globally, we compared them to 3361 complete genomes from the NCBI RefSeq database (Supplementary Data 10). Only 125 of the isolates in the present study shared  $>99.9\%$  average nucleotide identity (ANI) with any RefSeq genome, underscoring the novelty of the lineages obtained in our





study (Supplementary Fig. 12a). Phylogenetic analysis further revealed that several Hong Kong isolates formed distinct clades, not clustering with known global references (Supplementary Fig. 12b). These findings suggest that urban water systems in Hong Kong harbor genetically unique *E. coli* populations, supporting the value of regional genomic surveillance in complementing and expanding global AMR monitoring frameworks.

Unlike previous studies that focused on specific sublineages<sup>40,41</sup>, our sampling provides a city-wide, high-resolution snapshot of *E. coli* population structure across interconnected sectors. Among the

isolates, 117 remained unclassified, likely representing novel lineages or those that could not be identified within the current *E. coli* ST database. The detection of these unclassified strains highlights the broader genetic diversity within the urban microbial landscape, underscoring the need for continued genomic exploration to better understand these unclassified lineages and their potential public health implications. ST457 emerged as the most prevalent lineage among the isolates, consistent with its known association with human extraintestinal infections, such as urinary tract infections<sup>27,42</sup>. The widespread occurrence of ST457 across different sectors suggests a

**Fig. 7 | Sectoral sharing, mobility, and global dissemination of plasmids in urban *E. coli* isolates.** **a** Schematic summary of 2647 circular plasmids grouped into 1516 clusters based on pairwise sequence similarity (ANI > 90%, alignment fraction (AF) > 95%). Among them, 23 groups ( $n = 195$  plasmids) were shared across human-associated, animal-associated, and environmental water sectors. **b** Detailed characterization of the 23 cross-sectoral plasmid groups. Information shown includes the quantity of carried ARGs, plasmid category, copy number, length, host phylogroup, and ecological sector. Each group contains “ $n$ ” plasmids, as indicated in the figure, where “ $n$ ” represents the number of individual plasmids in each group. The box plots display the interquartile range (IQR), with the 25th and 75th percentiles marking the lower and upper bounds of the box, respectively. The median is represented by the line inside the box. Whiskers extend to the minimum and

maximum values within 1.5 times the IQR. Outliers are shown as individual points beyond the whiskers. The Sankey diagram on the right illustrates the sectoral origins of each group and their associated host sources. **c** An illustrative representation of the experimental validation of cross-sector transfer, along with the range of transfer efficiency of plasmids across different sectors. **d** Map of plasmid group P7, which harbors *bla*<sub>CTX-M-55</sub> and demonstrates global dissemination. The map displays its local distribution across sampling sites in Hong Kong and matched plasmids from international locations based on IMG/PR database comparisons. Circles represent recipient regions; symbols indicate local source types (pink: human-associated, blue: environmental, green: animal-associated). Source data are provided as a Source Data file.

high degree of ecological adaptability and highlights its potential role as a vehicle for the dissemination of clinically important resistance genes. This finding underscores the urgent need for continued One Health surveillance targeting ST457.

We also detected 13 isolates belonging to the globally dominant ST131 lineage, widely recognized for its multidrug resistance and pandemic spread. Although numerically limited, the detection of ST131 across multiple sectors and over an extended sampling period suggests its prevalence across the urban ecosystem. We observed diverse serotypes among ST131 isolates, potentially reflecting variations in virulence or environmental fitness<sup>43</sup>. Notably, all ST131 isolates harbored ESBL genes irrespective of antibiotic selection during isolation (Supplementary Data 11), aligning with other studies that highlight the propensity of ST131 for accumulating ESBL genes<sup>44</sup>.

By analyzing antibiotic resistance profiles, we identified a high prevalence of resistance to multiple antibiotics in isolates from both bathing beaches and animal-associated sources such as fishery water, which are directly or indirectly linked to human exposure. Isolates from bathing beaches showed elevated resistance to fluoroquinolones and SXT, both of which are first-line agents in clinical use. This raises concerns about possible transmission of resistant strains through recreational contact. Similarly, the detection of multidrug-resistant *E. coli* in fishery water suggests a risk of foodborne exposure through aquaculture products, particularly in regions where raw or undercooked seafood is consumed. These findings underscore ecological connectivity as a conduit for AMR threats, necessitating integrated surveillance. We propose leveraging WES, used during COVID-19 for real-time pathogen tracking<sup>45–47</sup>, as a scalable AMR sentinel system. For low- and middle-income countries, where clinical surveillance is resource-limited, WES offers a population-scale, cost-effective tool to map resistance genes and prioritize interventions in hotspots like hospitals and farms.

Discrepancies between phenotypic resistance and known genetic determinants were also observed, consistent with previous studies<sup>48,49</sup>, particularly for last-resort antibiotics such as meropenem. Specifically, some isolates exhibited resistance without detectable *bla*<sub>NDM</sub> or *bla*<sub>KPC</sub> genes, suggesting the presence of uncharacterized resistance mechanisms, epigenetic modifications, regulatory changes, or environmentally induced gene expression changes. These results emphasize the complexity of AMR phenotypes and the necessity for integrated genomic, transcriptomic, and functional studies to fully elucidate resistance mechanisms in environmental reservoirs.

Our analysis revealed that ARGs were predominantly plasmid-borne, reinforcing the importance of plasmids as vectors for ARG dissemination. Although recent studies have highlighted bacteriophages as potential carriers of ARGs<sup>50–52</sup>, our integrated analysis using geNomad, PlasFlow, and PlasX indicated that several initially classified “phage-associated” ARG carriers were in fact plasmids. This observation points to the technical challenges in distinguishing plasmids from phages, emphasizing the need for careful interpretation and further investigation. Importantly, we identified a total of 3522 plasmids, of which 1126 lacked identifiable replicons, reflecting both the

immense plasmid diversity and the current limitations of plasmid reference databases. The prevalence of unclassified plasmids highlights the urgent need to expand plasmid databases for accurate plasmid classification, which is crucial for a more complete understanding of mobile resistance dissemination across ecological sectors.

We systematically examined the ecological connectivity across sectors by analyzing the genome overlap of *E. coli* lineages, identifying clonal strain-sharing pairs, and investigating the cross-sectoral transfer of ARGs and plasmids. Our results revealed a significant overlap between isolates from human-associated samples and isolates from environmental samples, with clonal strain-sharing observed not only within different sampling sites of WWTPs but also between river water and human-associated samples. Notably, the extent of sharing appeared sensitive to the analytical thresholds applied, suggesting that the true magnitude of cross-sectoral connectivity may be underestimated.

To quantify these interconnections, we applied a hierarchical framework integrating JI, normalized distance (ND), and sharing pair ratio (SPR). All three metrics consistently indicated that the isolates from human-associated sectors and isolates from environmental sectors were more genetically connected to each other than connected to isolates from the animal-associated sector (JI: 0.36 vs. 0.19–0.20; ND: 1.00 vs. 1.04–1.05; SPR: 0.001 vs. 0). Importantly, each metric provides complementary ecological insights: the JI reflects broader population overlap, the ND captures fine-scale genetic relatedness, and the SPR highlights recent clonal transmission events.

Furthermore, the cross-sector dissemination of epidemiologically significant ARGs, notably *bla*<sub>CTX-M</sub> variants, was largely restricted to strain-sharing pairs, highlighting the role of clonal expansion in ARG spread. Plasmid mobility is often understudied in One Health research<sup>33</sup>. In the present study, among 2647 circular plasmids, 195 were found to be shared across the three sectors, including many instances of plasmid sharing independent of clonal relationships. Experimentally, conjugation assays further confirmed that representative ARG-carrying plasmids could successfully transfer between strains from different ecological sectors under laboratory conditions. However, transfer efficiencies varied considerably across donor–recipient pairs, reflecting differences in plasmid stability, host compatibility, and maintenance dynamics. Additionally, the differing growth rates of the donor, recipient, and transconjugant also quantitatively affect transfer efficiency<sup>54</sup>, necessitating further investigations.

While our study provides high-resolution insights into inter-sectoral microbial connectivity under the One Health framework, several limitations should be acknowledged. First, manhole and WWTP influent were used as proxies for human-associated samples, but we did not directly collect human samples. These wastewater samples predominantly originate from human feces, offering a collective, population-scale representation<sup>55,56</sup>, especially in urbanized areas like Hong Kong, where the separated sewer system minimizes contamination from non-human sources. Therefore, despite the absence of direct human sampling, wastewater remains an invaluable tool for population-level microbiome studies. Similar patterns likely exist in

other high-density cities such as Singapore and New York, where intricate human-animal-environment interfaces and extensive water networks contribute to AMR dissemination. Our results thus provide a scalable framework that may be adapted to assess ecological connectivity in other urban environments, supporting broader efforts toward integrated AMR surveillance. Second, while our study included isolates from multiple distinct sampling sites within each sector, the sampling design was tailored to capture cross-sectoral interactions rather than to enable robust within-sector comparisons. In particular, some subgroups (e.g., pig farm) were represented by a single site, and observations related to such locations should be interpreted as site-specific. Third, although over 1000 isolates in this study represent a large collection of *E. coli* genomes from different sectors, the number of isolates from one of the different sources is still limited, and may not fully capture the actual diversity of resistance profiles. Future studies with broader temporal and spatial resolution, especially across multiple environments, will be necessary to generalize sector-specific traits.

In summary, this study demonstrates that ecological connectivity between *E. coli* isolates from human-associated, animal-associated, and environmental sectors is not incidental but widespread, structured, and epidemiologically significant. Through high-resolution *E. coli* genomes, we show that strain- and plasmid-mediated exchanges actively link microbial communities across ecological boundaries, facilitating the circulation of clinically important ARGs. These findings identify ecological connectivity as a driver of AMR dissemination, with implications for public and environmental health under the One Health framework. By enabling the movement of resistant bacteria and mobile genetic elements across sectors, connectivity undermines traditional containment strategies, accelerating the bidirectional flow of resistance between environmental reservoirs and clinical settings.

Recognizing, monitoring, and managing ecological connectivity must therefore be placed at the forefront of global AMR control efforts. Our study underscores the urgent need for integrated genomic surveillance systems that transcend sectoral boundaries and unify human, animal, and environmental health responses. Moving forward, future interventions should not only focus on high-risk pathogens or hotspots but also target the ecological processes that underpin cross-sectoral transmission, contributing to the development of more sustainable strategies to address AMR, with the potential relevance at the global scale.

## Methods

### Isolates collection and DNA extraction

The samples used in this study were collected with the support of different departments of the Hong Kong Special Administrative Region Government, as part of the Hong Kong Theme-based Research Scheme project, which was funded by the Research Grant Council of Hong Kong SAR. This study focuses on a curated set of 1016 *E. coli* isolates collected during six major sampling rounds between November 2022 and February 2024. Sampling was conducted across 63 aquatic sites representing human-associated, environmental, and animal-associated compartments (sectors) (Supplementary Data 1). Specifically, human-associated samples were obtained from 7 manhole sites and influent of 14 WWTPs (WWTP-I), both regarded as major collective human-associated sources<sup>57</sup>. Environmental samples were collected from activated sludge of 6 WWTPs (WWTP-S), effluent of 8 WWTPs (WWTP-E), 12 river sites, 4 marine water sites, 5 bathing beach sites, 1 pig farm sludge (Pig-S) site, and 1 pig farm effluent (Pig-E) site. Animal-associated samples were derived from 1 pig farm influent (Pig-I) site and 4 fishery water sites. These samples were geographically distributed to capture diverse settings, and the isolates obtained from these samples were analyzed to characterize intersectoral microbial connectivity across different sectors.

For sample pretreatment, 1 mL of influent and sludge samples was centrifuged at  $5251 \times g$  for 10 min. River and bathing beach water

samples of 50 mL, and marine water samples of 150 mL were centrifuged for 30 min under the same conditions. Resulting pellets were resuspended using 100  $\mu$ L phosphate-buffered saline and plated on *E. coli* ChromoSelect Agar B (Sigma-Aldrich) for isolation.

To capture a broad spectrum of resistance phenotypes, isolates were cultured either on antibiotic-free media (“without antibiotic selection”) or on media supplemented with various antibiotic combinations (“under antibiotic selection”). Antibiotic selection involved media containing ampicillin (AMP, 100  $\mu$ g/mL), kanamycin (KAN, 50  $\mu$ g/mL), tetracycline (TET, 25  $\mu$ g/mL), chloramphenicol (CHL, 25  $\mu$ g/mL), or third-generation cephalosporins (CTX, 4  $\mu$ g/mL) in different combinations (Supplementary Data 1). These antibiotics were selected to isolate *E. coli* with diverse resistance profiles relevant to both clinical and environmental settings, representing major antibiotic classes commonly detected in urban aquatic ecosystems<sup>33</sup>.

All isolates were stored in 20% glycerol at  $-80^\circ\text{C}$  and purified through two rounds of streaking to ensure clonal colonies. From this pool, 1016 isolates were randomly selected based on stratification by sampling site, time, and antibiotic exposure to maximize ecological diversity. The final set included 415 isolates recovered without antibiotic selection and 601 obtained under selective pressure. The species identity of each isolate was first screened via *E. coli* selective agar and further confirmed through genome-based taxonomic classification. Genomic DNA was extracted from 3 mL overnight *E. coli* cultures using Qiagen DNeasy PowerSoil DNA Kit. DNA purity and concentration were assessed using Nanodrop and Qubit fluorometric quantification. Detailed metadata, including sampling information, are provided in Supplementary Data 1.

### Antimicrobial susceptibility testing

In this study, we tested 11 antibiotics from six broad categories, including (1) aminoglycoside (kanamycin), (2) beta-lactams (penicillin: amoxicillin and ampicillin; first-generation cephalosporin: cefazolin; third-generation cephalosporin: cefotaxime; carbapenem: meropenem), (3) chloramphenicol (chloramphenicol), (4) quinolones (ciprofloxacin and norfloxacin), (5) sulfonamide (sulfamethoxazole/trimethoprim), and (6) tetracycline (tetracycline). For each antibiotic, we set up concentration gradients. Fresh cultures of the test strains were adjusted to an OD<sub>600</sub> of 0.4–0.6 and inoculated at 10% into Mueller-Hinton Broth containing the different antibiotic concentrations. Initial OD<sub>600</sub> values (OD<sub>600</sub><sub>initial</sub>) were recorded using a microplate reader, and cultures were incubated at  $37^\circ\text{C}$  for 24 h with shaking at 180 rpm, after which the final OD<sub>600</sub> values (OD<sub>600</sub><sub>final</sub>) were measured. Strains were considered growing only when both OD<sub>600</sub><sub>final</sub> > 0.1 and [OD<sub>600</sub><sub>final</sub> – OD<sub>600</sub><sub>initial</sub>] > 0.03<sup>58</sup>. The results were interpreted according to the minimum inhibitory concentration (MIC) breakpoints following Clinical & Laboratory Standards Institute (United States) and European Committee on Antimicrobial Susceptibility Testing guidelines, categorizing the isolates as susceptible, intermediate, or resistant. Specific concentration gradients and breakpoint values for each antibiotic are detailed in Supplementary Data 2. Each antibiotic test was performed in duplicate, and only the consistent results in two tests were accepted. *E. coli* DH10B (NEB) was used as the negative control for quality control in each test batch.

### Benchmarking of sequencing and assembly strategies

To assess the feasibility of using long-read-only sequencing for large-scale bacterial genome reconstruction, we benchmarked the performance of the Nanopore R10.4.1 platform against PacBio HiFi and hybrid assembly strategies using three representative *E. coli* isolates collected from WWTP influent samples. Genomic DNA from these isolates was sequenced using multiple platforms. PacBio HiFi sequencing was performed on the PacBio Revio platform (Nextomics Biosciences Institute, China), and Illumina short-read sequencing ( $2 \times 150$  bp) was conducted on the HiSeq platform (Novogene, China).



Nanopore sequencing was performed using both R9.4.1 and R10.4.1 flow cells on GridION and PromethION devices, respectively. Library preparation followed the manufacturer's Rapid Barcoding Kit protocols (SQK-RBK110.96 and SQK-RBK114.96).

Illumina reads were adapter-trimmed using fastp (v0.23.2). Nanopore reads were first basecalled using super-accurate mode and then demultiplexed by Guppy (v6.3.9). For both PacBio and Nanopore reads, sequences  $\geq 1$  kb in length and with a quality score  $\geq 10$  were retained using Filtlong (v0.2.1). (<https://github.com/rrwick/Filtlong>). To evaluate the impact of sequencing depth on genome assembly quality, subsampling of the Nanopore and Illumina reads was performed using Seqkit (v2.6.1)<sup>59</sup>.

Assemblies were performed using Flye (v2.9.2)<sup>60</sup> and our in-house assembler NanoPhase (v0.2.2)<sup>61</sup>, both with and without Illumina short-read polishing. Specifically, PacBio HiFi data were assembled using Flye with the "--pacbio-hifi -g 5m" option. Nanopore reads were assembled using Flye with the "--nano-hq -g 5m" option and Nanopore in "isolate" mode. Hybrid assemblies were generated using NanoPhase with the "--hybrid" option.

Assembly quality was evaluated using CheckM (v1.1.10)<sup>62</sup> for genome completeness and contamination, and GTDB-Tk (v2.1.1)<sup>63</sup> for taxonomic classification. Open reading frames were predicted using Prodigal (v2.6.3)<sup>64</sup>, and genome annotation was performed using Prokka (v1.14.6)<sup>65</sup>. IDEEL scores<sup>66</sup> were calculated to assess protein-coding gene integrity, defined as the proportion of predicted proteins  $\geq 95\%$  the length of their best BLASTP hit in a curated protein database. QUAST (v5.0.2)<sup>67</sup> was used to assess indel rates, and FastANI (v1.34)<sup>68</sup> was applied for genome-wide similarity estimation relative to hybrid assemblies.

Benchmarking results showed that the Nanopore R10.4.1 platform, when combined with our in-house assembler NanoPhase, yielded high-quality genomes that were comparable to those generated by PacBio HiFi and hybrid assemblies. These assemblies showed elevated IDEEL scores, low indel error rates, and strong concordance in core-genome comparisons, including small cgSNP differences and high ANI values (Supplementary Data 12, Supplementary Fig. 13). Plasmid structures were also reliably reconstructed (Supplementary Data 13), indicating that long-read-only Nanopore sequencing is suitable for accurate genome and plasmid reconstruction of bacterial isolates.

### Genome reconstruction of 1016 *E. coli* isolates

Building on the benchmarking results, we applied the same sequencing and data processing pipeline to 1016 *E. coli* isolates. Genomic DNA of these isolates was sequenced using the Nanopore R10.4.1 platform with SQK-RBK114.96 barcoding kits on PromethION devices, enabling simultaneous sequencing of up to 96 bacterial isolates per flow cell. Raw Nanopore reads were processed as described above. Genome assemblies were generated using NanoPhase in the "isolate" mode, without additional Illumina polishing. Assemblies were quality filtered based on CheckM estimates: genomes with  $>97\%$  completeness and  $<2\%$  contamination were retained as high quality for downstream analysis.

### Pangenome, phylogenetic, and MLST analyses

The pangenome was estimated using Roary (v3.13.0)<sup>69</sup> with the following options "--e --mafft --group\_limit 60000." Sequence types (STs) were assigned to assembled genomes using MLST (v2.23.0) (<https://github.com/tseemann/mlst>). Phylogroups of the collected isolates were determined using ClermonTyping<sup>70</sup>. Phylogenetic analysis of the core-genome alignment obtained by Roary was performed using IQTREE2 (v2.2.6)<sup>71</sup> with the options "--m MFP -bb 1000." The resulting phylogenetic tree and associated metadata were visualized using Chiplot (<https://www.chiplot.online/>).

### Identification of bacterial sharing pairs

Pairwise core-genome MLST (cgMLST) and core-genome single-nucleotide polymorphism (cgSNP) analyses were conducted on isolate genomes, respectively, using cgmlstfinder (<https://github.com/kcric-tz/cgmlstfinder>) and snp-dists (v0.8.3) (<https://github.com/tseemann/snp-dists>). The genetic distance matrix of all pairwise-allelic-profile comparisons was calculated using a Python script. Isolate pairs with both pairwise cgSNP and cgMLST distances less than 10 were considered sharing pairs, consistent with previously published thresholds used to identify close genetic relatedness<sup>16</sup>. The similarity of the pangenome and resistome between isolates was assessed using the Jaccard Index (JI), defined as the size of the intersection divided by the size of the union of the gene sets. The JI value ranges from 0 to 1, with the value closer to 1 indicating higher similarity.

### VF, ARGs, and ISs identification and analyses

VFs were annotated through a BLASTN homology search ( $E$ -value  $\leq 10^{-10}$ ) against the experimentally verified VFs in the VFDB<sup>72</sup>, requiring a minimum of 80% identity and 70% query coverage. ARG profiles of the isolates were identified by BLASTP against the SARG database<sup>73</sup> ( $E$ -value  $\leq 10^{-5}$ ) with a minimum similarity of 90% over 90% query coverage. In this study, we excluded the ARG subtype categorized as "multidrug" in the SARG database, as this category primarily includes efflux pumps, such as those from the ABC transporter family, whose roles in AMR are often broad, indirect, and may not always correspond to specific antibiotic targets<sup>74</sup>. Recently transferred ARGs for resistance to third-generation cephalosporins and quinolones were defined as those with  $>99\%$  protein identity and coverage, identified using BLASTP ( $E$ -value  $\leq 10^{-7}$ ). Gene alignments were conducted using Clustal Omega v1.2.4<sup>75</sup>, and phylogenetic trees were constructed with IQTREE2 and visualized using Chiplot (<https://www.chiplot.online/>). Gene association networks were constructed with Cytoscape software<sup>76</sup>. IS elements within the 5 kb regions flanking the ARGs were identified based on a homology search against the ISfinder database with BLASTN ( $E$ -value  $\leq 10^{-10}$ ) with both identity and coverage higher than 80%.

### Identification, classification, clustering, and global distribution of plasmids

The plasmids in 1016 isolates were identified using three different tools: PlasX<sup>77</sup>, PlasFlow<sup>78</sup>, and geNomad<sup>79</sup>, with identification thresholds set at 0.7, 0.9, and 0.7, respectively. These tools respectively identified 3051, 3105, and 3197 sequences as plasmids, with an intersection of 2526 sequences (Supplementary Fig. 14). To address inconsistencies among the tools and ensure high-confidence identification, we developed a stringent workflow (Supplementary Fig. 3). This workflow involved initially identifying 3637 plasmid candidates by three tools, which we categorized into circular and non-circular sequences. For circular sequences, a sequence was confirmed as a plasmid if any of the following criteria were met:

- 1) It scored above the threshold in one tool and had a virus score below 0.7, as assigned by geNomad;
- 2) geNomad or PlasX scores were above the threshold, and PlasFlow also scored above the threshold;
- 3) Both PlasX and geNomad scored above the threshold;
- 4) High-confidence plasmid score (above 0.95) from geNomad or PlasX, and the virus score assigned by geNomad was below 0.9.

For non-circular sequences, in addition to ensuring they were not identified as viruses, they also had to be classified as non-chromosomal by PlasFlow. Plasmids were categorized into three types (conjugative, mobilizable, and non-mobilizable) using PlasmidC<sup>80</sup> based on the presence of the protein machinery associated with DNA transfer. In this pipeline, a plasmid was considered conjugative if it carried relaxase,



T4CP, and T4SSs; mobilizable if it encoded only relaxase; and non-mobilizable if it lacked all these elements. The replicon sequences of all plasmids were identified using Plasmidfinder<sup>81</sup>. Plasmid copy numbers were inferred by comparing the coverage of plasmid sequences to chromosomal coverage using CoverM v0.7.0 (<https://github.com/wood/CoverM>).

In this study, to investigate plasmid sharing across different sectors, we clustered plasmids with high confidence. Pairwise comparisons were conducted, grouping plasmids with an ANI higher than 90% and coverage above 95% for both query and target sequences. Additionally, the global distribution of the 195 shared plasmids was explored by comparing them with 13,209 circular plasmids isolated from bacterial isolates in the IMG/PR database<sup>82</sup>. Plasmids with pairwise ANI higher than 95% and coverage above 50% were considered to have a similar backbone, indicating global distribution. Plasmid maps were generated using Proksee (<https://proksee.ca/>).

### Plasmid conjugation experiments

To determine if plasmids from one sector could transfer to isolates from other sectors, we experimentally validated the cross-sector transfer of plasmids. For each group, one or more donor isolates carrying plasmid-borne ARGs were selected from their respective ecological sectors. Corresponding recipient isolates were chosen based on resistance to either tigecycline or colistin while remaining susceptible to the resistance marker carried by the donor plasmid. To test inter-sectoral transferability, each donor strain was conjugated with two recipient strains from each of the other two sectors. A total of 108 donor–recipient combinations were tested, involving 27 unique donor strains and 12 unique recipient strains spanning different ecological sectors (details are provided in Supplementary Data 6).

Specifically, 500 µL overnight cultures of each donor and recipient strain were mixed, and the bacterial mixture was spotted on 0.45 µM nitrocellulose membranes pre-plated on LB agar. Each plate was incubated at 37 °C for 12 h, after which the bacteria on the membrane were washed off using LB broth and plated on LB agar plates containing tigecycline (2.5 µg/mL) or colistin (2 µg/mL) and kanamycin (50 µg/mL) or third-generation cephalosporins (4 µg/mL). These plates were incubated for 24 h at 37 °C, followed by colony counting. All these experiments were performed in duplicate. The recipient strain was plated and counted at the start of the conjugation experiment, and conjugation efficiency was calculated using the formula: (number of transconjugants/number of recipients) × 100%.

### Quantification framework for the connectivity between different sectors

To assess the connectivity between different sectors, we developed a quantitative framework using a multi-resolution approach. This framework quantifies connectivity at various levels: ST, genetic distance, and clonal sharing, offering deeper insights into cross-sectoral overlap and genetic exchange within a One Health context.

1. *The JI* is used to measure the similarity between two or more habitats based on the shared STs. It is calculated as:

$$JI = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

Where:

- $A \cap B$  is the intersection of STs between habitats *A* and *B*.
- $A \cup B$  is the union of STs between habitats *A* and *B*.

2. *The normalized distance (ND)* accounts for internal genetic diversity within habitats, giving a more balanced view of genetic distinction or connectivity. It compares the genetic distance between habitats to the average internal distances within each habitat,

helping to reveal whether closer distances correlate with increased ecological connectivity. ND is calculated as follows:

$$ND = \frac{D_{AB}}{\frac{D_A + D_B}{2}} \quad (2)$$

Where:

- ND is the normalized distance between habitats *A* and *B*.
- $D_{AB}$  is the average genetic distance (e.g., cgMLST) between habitats *A* and *B*.
- $D_A$  is the average genetic distance within habitat *A*.
- $D_B$  is the average genetic distance within habitat *B*.

3. *The sharing pair ratio (SPR)* quantifies the proportion of highly similar isolate pairs across habitats, calculated as:

$$SPR = \frac{P_1}{P_2} \quad (3)$$

Where:

- SPR is the sharing pair ratio between habitats *A* and *B*.
- $P_1$  is the number of pairs of isolates between habitats *A* and *B* with a cgMLST distance of less than 10.
- $P_2$  is the total number of isolate pairs between habitats *A* and *B*.

### Statistical methods and analysis

All statistical comparisons of antibiotic resistance phenotypes, total ARG counts, VF counts, and plasmid counts across the three ecological sectors were performed using two-way ANOVA, with  $p < 0.05$  considered statistically significant. Hierarchical clustering of ARG presence/absence and plasmid replicon profiles was performed using Bray–Curtis distance and the single linkage method. Clustering was conducted in an unsupervised manner, and no prior group labels were used to guide clustering. These profile-based comparisons were used for exploratory visualization, and no hypothesis testing was applied to the clustering results.

### Reporting summary

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

### Data availability

All the assemblies and sequencing data from this study have been deposited in the NCBI GenBank under BioProject accession number PRJNA1185485. The global plasmid data and 3361 reference *E. coli* genomes used for comparison are available in the IMG/PR database and NCBI RefSeq database (Supplementary Data 7 and 10). Source data are provided with this paper.

### Code availability

All custom Python and R codes have been uploaded to GitHub: [https://github.com/xxqxq1994/EC\\_Ecological-connectivity/](https://github.com/xxqxq1994/EC_Ecological-connectivity/).

### References

1. Islam, M. S., Rahman, A. T., Hassan, J. & Rahman, M. T. Extended-spectrum beta-lactamase in *Escherichia coli* isolated from humans, animals, and environments in Bangladesh: a One Health perspective systematic review and meta-analysis. *One Health* **16**, 100526 (2023).
2. Bertagnolio, S. et al. WHO global research priorities for antimicrobial resistance in human health. *Lancet Microbe* **5**, 100902 (2024).
3. Naghavi, M. et al. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *Lancet* **404**, 1199–1226 (2024).

4. Kariuki, S. Global burden of antimicrobial resistance and forecasts to 2050. *Lancet* **404**, 1172–1173 (2024).
5. Larsson, D., Gaze, W., Laxminarayan, R. & Topp, E. AMR, One Health and the environment. *Nat. Microbiol.* **8**, 754–755 (2023).
6. Tripartite and UNEP support OHLEP's definition of "One Health". <https://www.who.int/news/item/01-12-2021-tripartite-and-unep-support-ohlep-s-definition-of-one-health> (World Health Organization, 2021).
7. Rabinowitz, P. & Conti, L. Links among human health, animal health, and ecosystem health. *Annu. Rev. Public Health* **34**, 189–204 (2013).
8. Li, X. et al. Population structure and antibiotic resistance of swine extraintestinal pathogenic *Escherichia coli* from China. *Nat. Commun.* **15**, 5811 (2024).
9. Yin, X. et al. Global environmental resistome: Distinction and connectivity across diverse habitats benchmarked by metagenomic analyses. *Water Res.* **235**, 119875 (2023).
10. Mao, X. et al. Longitudinal metagenomic analysis on antibiotic resistome, mobilome, and microbiome of river ecosystems in a subtropical metropolitan city. *Water Res.* **274**, 123102 (2025).
11. Mao, K. et al. The potential of wastewater-based epidemiology as surveillance and early warning of infectious disease outbreaks. *Curr. Opin. Environ. Sci. Health* **17**, 1–7 (2020).
12. Smit, C. C. et al. One Health determinants of *Escherichia coli* antimicrobial resistance in humans in the community: an umbrella review. *Int. J. Mol. Sci.* **24**, 17204 (2023).
13. Tenaillon, O., Skurnik, D., Picard, B. & Denamur, E. The population genetics of commensal *Escherichia coli*. *Nat. Rev. Microbiol.* **8**, 207–217 (2010).
14. Jang, J. et al. Environmental *Escherichia coli*: ecology and public health implications—a review. *J. Appl. Microbiol.* **123**, 570–581 (2017).
15. Mahfouz, N. et al. High genomic diversity of multi-drug resistant wastewater *Escherichia coli*. *Sci. Rep.* **8**, 8928 (2018).
16. Muloi, D. M. et al. Population genomics of *Escherichia coli* in livestock-keeping households across a rapidly developing urban landscape. *Nat. Microbiol.* **7**, 581–589 (2022).
17. Djordjevic, S. P. et al. Genomic surveillance for antimicrobial resistance—a One Health perspective. *Nat. Rev. Genet.* **25**, 142–157 (2024).
18. Delgado-Blas, J. F. et al. Population genomics and antimicrobial resistance dynamics of *Escherichia coli* in wastewater and river environments. *Commun. Biol.* **4**, <https://doi.org/10.1038/s42003-021-01949-x> (2021).
19. Cassini, A. et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect. Dis.* **19**, 56–66 (2019).
20. Islam, K. et al. Epidemiology of extended-spectrum  $\beta$ -lactamase and metallo- $\beta$ -lactamase-producing *Escherichia coli* in South Asia. *Future Microbiol.* **16**, 521–535 (2021).
21. Murray, C. J. et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* **399**, 629–655 (2022).
22. WHO Bacterial Priority Pathogens List, 2024: Bacterial Pathogens of Public Health Importance to Guide Research, Development and Strategies to Prevent and Control Antimicrobial Resistance <https://www.who.int/publications/i/item/9789240093461> (World Health Organization, 2024).
23. Denamur, E., Clermont, O., Bonacorsi, S. & Gordon, D. The population genetics of pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* **19**, 37–54 (2021).
24. Mageiros, L. et al. Genome evolution and the emergence of pathogenicity in avian *Escherichia coli*. *Nat. Commun.* **12**, 765 (2021).
25. Li, D. et al. Genomic comparisons of *Escherichia coli* ST131 from Australia. *Microb. Genom.* **7**, 000721 (2021).
26. Tarabai, H., Wyrsh, E. R., Bitar, I., Dolejska, M. & Djordjevic, S. P. Epidemic H12 plasmids mobilising the carbapenemase gene bla<sub>IMP-4</sub> in Australian clinical samples identified in multiple sub-lineages of *Escherichia coli* ST216 colonising silver gulls. *Microorganisms* **9**, 567 (2021).
27. Nesporova, K. et al. *Escherichia coli* sequence type 457 is an emerging extended-spectrum- $\beta$ -lactam-resistant lineage with reservoirs in wildlife and food-producing animals. *Antimicrob. Agents Chemother.* **65**, <https://doi.org/10.1128/aac.01118-01120> (2020).
28. Guenther, S., Ewers, C. & Wieler, L. H. Extended-spectrum beta-lactamases producing *E. coli* in wildlife, yet another form of environmental pollution? *Front. Microbiol.* **2**, 246 (2011).
29. Koren, S. & Phillippy, A. M. One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Curr. Opin. Microbiol.* **23**, 110–120 (2015).
30. Shaw, L. P. et al. Niche and local geography shape the pangenome of wastewater-and livestock-associated Enterobacteriaceae. *Sci. Adv.* **7**, eabe3868 (2021).
31. Castañeda-Barba, S., Top, E. M. & Stalder, T. Plasmids, a molecular cornerstone of antimicrobial resistance in the One Health era. *Nat. Rev. Microbiol.* **22**, 18–32 (2024).
32. Walker, A. Welcome to the plasmidome. *Nat. Rev. Microbiol.* **10**, 379–379 (2012).
33. Che, Y. et al. High-resolution genomic surveillance elucidates a multilayered hierarchical transfer of resistance between WWTP-and human/animal-associated bacteria. *Microbiome* **10**, 16 (2022).
34. Chin, C.-S. et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013).
35. Moss, E. L., Maghini, D. G. & Bhatt, A. S. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nat. Biotechnol.* **38**, 701–707 (2020).
36. Sereika, M. et al. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat. Methods* **19**, 823–826 (2022).
37. Loman, N. J., Quick, J. & Simpson, J. T. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods* **12**, 733–735 (2015).
38. Xu, X. et al. High-resolution and real-time wastewater viral surveillance by Nanopore sequencing. *Water Res.* **256**, 121623 (2024).
39. Pong, C. H., Harmer, C. J., Ataide, S. F. & Hall, R. M. An IS 26 variant with enhanced activity. *FEMS Microbiol. Lett.* **366**, fnz031 (2019).
40. Ho, P.-L. et al. Extensive dissemination of CTX-M-producing *Escherichia coli* with multidrug resistance to 'critically important' antibiotics among food animals in Hong Kong, 2008–10. *J. Antimicrob. Chemother.* **66**, 765–768 (2011).
41. Ho, P. et al. Dissemination of plasmid-mediated fosfomycin resistance fosA3 among multidrug-resistant *Escherichia coli* from livestock and other animals. *J. Appl. Microbiol.* **114**, 695–702 (2013).
42. Kocsis, B., Gulyás, D. & Szabó, D. Emergence and dissemination of extraintestinal pathogenic high-risk international clones of *Escherichia coli*. *Life* **12**, 2077 (2022).
43. Mora, A. et al. Virulence patterns in a murine sepsis model of ST131 *Escherichia coli* clinical isolates belonging to serotypes O25b: H4 and O16: H5 are associated to specific virotypes. *PLoS ONE* **9**, e87025 (2014).
44. Qureshi, Z. A. & Doi, Y. *Escherichia coli* sequence type 131: epidemiology and challenges in treatment. *Expert Rev. Anti-infect. Ther.* **12**, 597–609 (2014).
45. Aarestrup, F. M. & Woolhouse, M. E. Using sewage for surveillance of antimicrobial resistance. *Science* **367**, 630–632 (2020).
46. Deng, Y. et al. Use of sewage surveillance for COVID-19 to guide public health response: a case study in Hong Kong. *Sci. Total Environ.* **821**, <https://doi.org/10.1016/j.scitotenv.2022.153250> (2022).

47. Xu, X. et al. Real-time allelic assays of SARS-CoV-2 variants to enhance sewage surveillance. *Water Res.* **220**, 118686 (2022).
48. Yee, R., Dien Bard, J. & Simmer, P. J. The genotype-to-phenotype dilemma: How should laboratories approach discordant susceptibility results? *J. Clin. Microbiol.* **59**, <https://doi.org/10.1128/jcm.00138-00120> (2021).
49. Tofteland, S. L. et al. Effects of phenotype and genotype on methods for detection of extended-spectrum- $\beta$ -lactamase-producing clinical isolates of *Escherichia coli* and *Klebsiella pneumoniae* in Norway. *J. Clin. Microbiol.* **45**, 199–205 (2007).
50. Enault, F. et al. Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *ISME J.* **11**, 237–247 (2017).
51. Calero-Cáceres, W. & Muniesa, M. Persistence of naturally occurring antibiotic resistance genes in the bacteria and bacteriophage fractions of wastewater. *Water Res.* **95**, 11–18 (2016).
52. Moon, K. et al. Freshwater viral metagenome reveals novel and functional phage-borne antibiotic resistance genes. *Microbiome* **8**, 1–15 (2020).
53. Matlock, W. et al. Enterobacteriales plasmid sharing amongst human bloodstream infections, livestock, wastewater, and waterway niches in Oxfordshire, UK. *eLife* **12**, e85302 (2023).
54. Lopatkin, A. J. et al. Bacterial metabolic state more accurately predicts antibiotic lethality than growth rate. *Nat. Microbiol.* **4**, 2109–2117 (2019).
55. Becsei, Á. et al. Time-series sewage metagenomics distinguishes seasonal, human-derived and environmental microbial communities potentially allowing source-attributed surveillance. *Nat. Commun.* **15**, 7551 (2024).
56. Deng, Y. et al. Use of sewage surveillance for COVID-19: a large-scale evidence-based program in Hong Kong. *Environ. Health Perspect.* **130**, <https://doi.org/10.1289/ehp9966> (2022).
57. Mills, M., Mollenkopf, D., Wittum, T., Sullivan, M. P. & Lee, J. One Health threat of treated wastewater discharge in urban Ohio rivers: implications for surface water and fish gut microbiome and resistome. *Environ. Sci. Technol.* **58**, 13402–13414 (2024).
58. Ma, L., Yang, H., Guan, L., Liu, X. & Zhang, T. Risks of antibiotic resistance genes and antimicrobial resistance under chlorination disinfection with public health concerns. *Environ. Int.* **158**, 106978 (2022).
59. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **11**, e0163962 (2016).
60. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
61. Liu, L., Yang, Y., Deng, Y. & Zhang, T. Nanopore long-read-only metagenomics enables complete and high-quality genome reconstruction from mock and complex metagenomes. *Microbiome* **10**, 209 (2022).
62. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
63. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**, 5315–5316 (2022).
64. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 1–11 (2010).
65. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
66. Stewart, R. D. et al. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat. Commun.* **9**, 870 (2018).
67. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
68. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
69. Page, A. J. et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).
70. Beghain, J., Bridier-Nahmias, A., Le Nagard, H., Denamur, E. & Clermont, O. ClermonTyping: an easy-to-use and accurate in silico method for *Escherichia* genus strain phylotyping. *Microb. Genom.* **4**, e000192 (2018).
71. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
72. Liu, B., Zheng, D., Zhou, S., Chen, L. & Yang, J. VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res.* **50**, D912–D917 (2022).
73. Yin, X. et al. ARGs-OAP v3. O: antibiotic-resistance gene database curation and analysis pipeline optimization. *Engineering* **27**, 234–241 (2023).
74. El-Awady, R. et al. The role of eukaryotic and prokaryotic ABC transporter family in failure of chemotherapy. *Front. Pharmacol.* **7**, 535 (2017).
75. Sievers, F. & Higgins, D. G. Clustal Omega, accurate alignment of very large numbers of sequences. In *Multiple Sequence Alignment Methods*. 105–116 (Humana Press, Totowa, NJ, 2013).
76. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
77. Yu, M. K., Fogarty, E. C. & Eren, A. M. Diverse plasmid systems and their ecology across human gut metagenomes revealed by PlasX and MobMess. *Nat. Microbiol.* **9**, 830–847 (2024).
78. Krawczyk, P. S., Lipinski, L. & Dziembowski, A. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Res.* **46**, e35–e35 (2018).
79. Camargo, A. P. et al. Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* **42**, 1303–1312 (2024).
80. Che, Y. et al. Conjugative plasmids interact with insertion sequences to shape the horizontal transfer of antimicrobial resistance genes. *Proc. Natl. Acad. Sci. USA* **118**, e2008731118 (2021).
81. Carattoli, A. et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **58**, 3895–3903 (2014).
82. Camargo, A. P. et al. IMG/PR: a database of plasmids from genomes and metagenomes with rich annotations and metadata. *Nucleic Acids Res.* **52**, D164–D173 (2024).

## Acknowledgements

This study was financially supported by the Theme-based Research Scheme grant (T21-705/20-N to T.Z.) and General Research Fund (17209823 to Y.D. and T.Z.) from the Research Grants Council of the Hong Kong Special Administrative Region, China, as well as by the Shenzhen Science and Technology Innovation Bureau (no. SGD20230821091559021 to T.Z.). X.X. and D.W. would like to thank the University of Hong Kong for the postdoctoral fellowship. X.C. would like to thank the University of Hong Kong for the Postgraduate Studentship. Technical support from Ms. Vicky Fung is greatly appreciated.

## Author contributions

Conceptualization: X.X., T.Z.; methodology: X.X., Y.L., Y.D., L.L., D.W., Q.T., C.W., X.C., Y.C., E.W., V.J., S.D., T.Z.; investigation: X.X., Y.L.; visualization: X.X.; writing—original draft: X.X., T.Z.; writing—review and editing: X.X., Y.L., Y.D., L.L., D.W., Q.T., C.W., X.C., Y.C., E.W., V.J., S.D., T.Z.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-62455-w>.

**Correspondence** and requests for materials should be addressed to Tong Zhang.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

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

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

**Open Access** This article is licensed under a Creative Commons Attribution-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/>.

© The Author(s) 2025