

# From clinics to sewers: leveraging environmental surveillance and whole genome sequencing to inform transmission of ESBL-*Escherichia coli* in Switzerland

Sheena Conforti,<sup>1,2,3</sup> Louis du Plessis,<sup>2,3</sup> Claudia Bagutti,<sup>4</sup> Jens Becker,<sup>5</sup> Silvio D. Brugger,<sup>6</sup> Alexia Cusini,<sup>7</sup> Adrian Egli,<sup>8</sup> Valeria Gaia,<sup>9</sup> Gilbert Greub,<sup>10</sup> Cornelia Guler,<sup>11</sup> Jana S. Huisman,<sup>12</sup> Claudine Kocher,<sup>13</sup> Roger D. Kouyos,<sup>6</sup> Karoline Leuzinger,<sup>14</sup> Carola Maffioli,<sup>15</sup> Mara Neacșu,<sup>2</sup> Oliver Nolte,<sup>8</sup> Alban Ramette,<sup>13</sup> Salome N. Seiffert,<sup>16</sup> Sarah Tschudin-Sutter,<sup>17</sup> Jon Paulin Zumthor,<sup>18</sup> Tanja Stadler,<sup>2,3</sup> Timothy R. Julian<sup>1</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 15.

**ABSTRACT** Extended-spectrum  $\beta$ -lactamase-producing *Escherichia coli* (ESBL-*E. coli*) is a major antimicrobial resistance concern spreading across human, animal, and environmental domains. To assess between-source transitions, we analyzed 762 ESBL-*E. coli* genomes collected from wastewater (used as a community shedding indicator), clinical settings, cattle, and wildlife across Switzerland (2021–2023). ST131 was the most prevalent sequence type (ST), and 76% of isolates carried resistance to at least two antibiotic classes in addition to  $\beta$ -lactams. Phylogenetic analysis showed isolates were interspersed across sources, yet genetically similar strains were more common within compartments. Clonal isolates (0 SNPs) were rarely shared ( $n = 2$ ) between wastewater and corresponding clinics. Ancestral state reconstruction revealed compartmentalization of isolates between wastewater and clinics across the whole phylogeny. However, this pattern disappears within human-associated ST131, ST69, and ST1193, highlighting exchange between clinics and communities. These findings show that wastewater surveillance captures community circulation of ESBL-*E. coli*, which overlaps with circulating clinically relevant strains.

**IMPORTANCE** Extended-spectrum  $\beta$ -lactamase-producing *Escherichia coli* (ESBL-*E. coli*) is a major antimicrobial resistance concern globally, with ready transmission between people, animals, and the environment. Understanding transmission patterns among reservoirs can help inform risks and inform intervention strategies. To date, surveillance is heavily reliant on aggregation of clinical data, resulting in limited insights on ESBL-*E. coli* circulation in the population. Wastewater-based surveillance has emerged as a powerful tool to understand epidemiology of antimicrobial resistance circulating in communities, thereby complementing clinical surveillance approaches. Here, we demonstrate integrated surveillance of ESBL-*E. coli* isolates across sectors using whole genome sequencing, providing an opportunity to understand exchange of bacterial isolates between clinics, communities, livestock, and wildlife.

**KEYWORDS** antimicrobial resistance, One Health, ESBL-*E. coli*, wastewater-based surveillance, phylogenetics, sequencing

In 2019, infections caused by extended-spectrum  $\beta$ -lactamase-producing *Escherichia coli* (ESBL-*E. coli*) accounted for approximately 100,000 of the nearly 5 million deaths associated with bacterial antimicrobial resistance (AMR) (1). Initially predominant in hospitals, over the past 40 years, ESBL-*E. coli* has spread to communities, livestock, and wildlife (2, 3). Its widespread detection highlights the importance of not only

**Editor** Satoshi Ishii, University of Minnesota Twin Cities, St. Paul, Minnesota, USA

Address correspondence to Timothy R. Julian, tim.julian@eawag.ch, or Tanja Stadler, tanja.stadler@bsse.ethz.ch.

The authors declare no conflict of interest.

See the funding table on p. 15.

**Received** 18 September 2025

**Accepted** 27 February 2026

**Published** 6 April 2026

Copyright © 2026 Conforti et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

human-to-human transmission but also interactions within and between humans, livestock, and wildlife (3). Bacteria and antibiotic resistance genes (ARGs) exchange across these compartments, driving the spread of resistance through multiple transmission pathways (4–6). Accordingly, the World Health Organization (WHO) advocates a One Health approach for tackling AMR, emphasizing the interconnectedness of human, animal, and environmental health (6–8).

The environment (soil, water, wastewater, and hospital equipment) often acts as an intermediary in transmission, with humans and animals shedding ESBL-*E. coli* into it and being exposed through contact. The environment can also act as a reservoir where ESBL-*E. coli* can grow under suitable conditions, such as in soil (9, 10). Human activities that disturb natural habitats intensify inter-compartmental transmission, increasing AMR bacterial carriage in wildlife, while agricultural practices such as manure application contribute to ARGs' cross-sectoral transmission (11, 12). Direct transmission from livestock to farmers can further expose humans to animal-derived resistance factors (4). Unlike many infectious diseases, the spread of AMR is driven not only by infection and colonization but also by horizontal gene transfer and long-term, asymptomatic carriage, both of which facilitate the persistence and dissemination of resistant organisms (13, 14). While there is evidence of cross-compartmental transmission events (15), understanding their direction and frequency may help design effective interventions to reduce the spread of AMR.

Wastewater-based surveillance (WBS) is emerging as a useful tool for AMR monitoring within a One Health framework, tracking the presence of resistant bacteria in wastewater and providing insights into prevalence of both pathogenic and non-pathogenic resistant bacteria in communities (16–18). WBS offers a comprehensive view of AMR dissemination by capturing resistant bacteria from diverse sources, including sewer-connected entities like households, hospitals, and industries, as well as non-point sources like agricultural or urban runoff (19). The diversity of resistant bacteria from various sources can be further resolved using whole genome sequencing (WGS), which provides detailed insights into their distribution across human, animal, and environmental compartments, enhancing AMR surveillance (20). In the United Kingdom, the adoption of WGS for real-time surveillance of resistant pathogens resulted in enhanced public health outcomes by enabling quicker outbreak resolution, cost reduction, and increased confidence in linking sources to cases (21). Furthermore, WGS of wastewater isolates may help map transmission across compartments, contribute to epidemiological modeling, and guide public health strategies (16, 22).

In this study, we investigated the diversity and transmission of 762 ESBL-producing *E. coli* isolates across animal, clinical, and environmental compartments, including wastewater, in Switzerland to elucidate potential transmission pathways within a One Health framework. Specifically, we aimed to (i) characterize the genetic diversity of ESBL-producing *E. coli* across multiple compartments, (ii) assess whether ESBL-producing *E. coli* diversity shows clustering based on source, geographic regions, or both, and (iii) determine transition events between sources using phylogenomic approaches. Given that wastewater captures bacteria shed almost exclusively by human populations, we use wastewater isolates as an indicator for community-level ESBL-*E. coli*. Our findings provide insights into the diversity and transmission dynamics of ESBL-*E. coli* in Switzerland, contributing to a comprehensive AMR surveillance strategy at the national scale.

## RESULTS

### Sampling and sequencing

A total of 834 *Escherichia coli* isolates producing extended-spectrum  $\beta$ -lactamases (ESBL) were recovered from four sources across Switzerland between November 2021 and October 2023: wastewater ( $n = 438$ ), healthcare facilities ( $n = 328$ ), cattle ( $n = 26$ ), and wildlife ( $n = 42$ ) (Fig. 1; Fig. S1; Table S1). Clinical isolates were obtained through a convenience sampling strategy, whereby only confirmed ESBL-producing *E. coli* from inpatients and outpatients at healthcare facilities were included. Of the 834

isolates sequenced, 72 (9%) were excluded due to incorrect species assignment and/or contamination ( $n = 45$ ), low assembly quality ( $n = 21$ ), or low sequencing coverage ( $n = 6$ ) (Table S2). The remaining 762 (91%) passed quality control and were included in downstream analyses (Table S3 and S4). Exclusions were relatively consistent across sources, with 406 of 438 (93%) wastewater, 302 of 328 (92%) clinical, 25 of 26 (96%) cattle, and 29 of 42 (69%) wildlife isolates retained. Isolates were distributed across sources as follows: 406 (53%) from wastewater, 302 (40%) clinical isolates from in/out-patients at healthcare facilities, 29 (4%) from wildlife, and 25 (3%) from cattle (Table S2). To monitor contamination and sequencing performance, 12 negative extraction controls (AE blanks) were included across both sequencing batches, with no evidence of contamination detected. Reference strains of *E. coli* (strain 1 and 2;  $n = 8$  and 8 in batch 1,  $n = 2$  and 1 in batch 2) were processed alongside study samples and, together with the 15 study isolates sequenced in duplicate, were used to assess SNP variations among re-sequenced genomes. In total, 96 pairwise comparisons were evaluated, with SNP differences ranging from 0 to 13, with a median of 2 (Fig. S2A). Pairwise comparisons were performed between re-sequenced reference genomes and their corresponding original assemblies generated in an earlier study. For strain 1, SNP differences between the original genome and 10 re-sequenced genomes ranged from 6 to 14, with a median of 7.5 (Fig. S2B). For strain 2, SNP differences between the original genome and 9 re-sequenced genomes ranged from 1 to 13, with a median of 4 (Fig. S2C). *Klebsiella pneumoniae* controls (strain 1 and 2;  $n = 8$  and 5 in batch 1,  $n = 1$  and 2 in batch 2) were clearly identified as *K. pneumoniae* after assembly, confirming the accuracy of both sequencing and downstream species assignment.

### Diversity and phylogenetic distribution of ESBL-*E. coli*

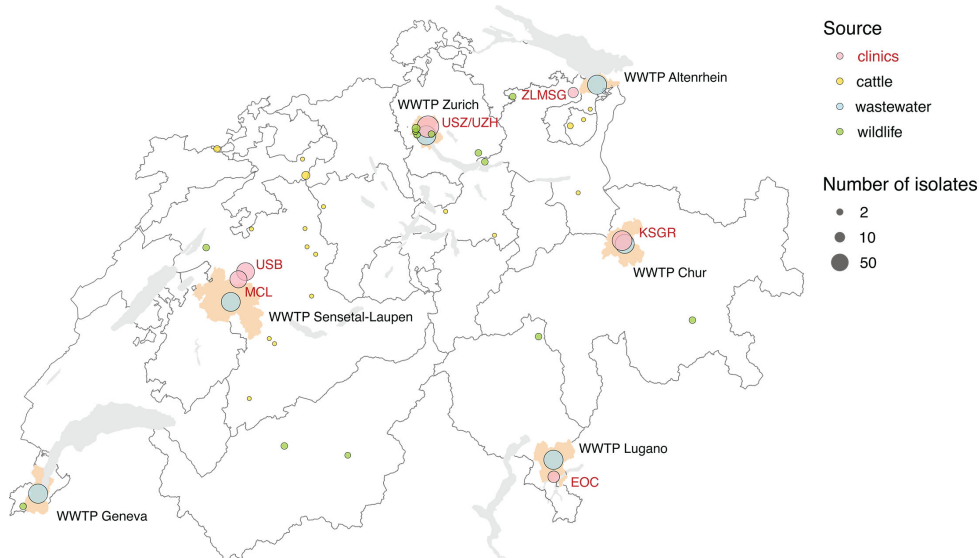
A maximum-likelihood phylogenetic tree based on the core genome of 762 ESBL-producing *E. coli* isolates reveals clustering into two main clades (Fig. 2). EzClermont typing identified eight phylogroups: B2 ( $n = 279$ , 37%), A ( $n = 162$ , 21%), D ( $n = 153$ , 20%), B1 ( $n = 92$ , 12%), F ( $n = 48$ , 6%), G ( $n = 13$ , 2%), E ( $n = 10$ , 1%), and cryptic ( $n = 4$ , 0.5%) (Table S5). Additionally, one isolate ( $n = 1$ , 0.1%) was categorized as unknown. Phylogroups A, B1, D, E, and F were present across all compartments, while B2 was predominantly detected in clinical samples ( $n = 169$ , 61%) and wastewater ( $n = 107$ , 38%), with three isolates (1%) from a single black-headed gull and none from cattle.

Among the isolates, 148 sequence types (STs) were identified (Table S4). The most prevalent was ST131 ( $n = 197$ , 26%), followed by ST10 ( $n = 66$ , 9%), ST38 ( $n = 55$ , 7%), ST69 ( $n = 53$ , 7%), and ST1193 ( $n = 44$ , 6%) (Fig. 2; Table S6). Clinical samples were dominated by phylogroup B2 and ST131. In contrast, cattle-derived isolates ( $n = 25$ ) largely belonged to less common STs, with only one identified as ST69. Notably, ST131 isolates ( $n = 197$ ) were detected in humans, wastewater, and wildlife but were absent in cattle.

The phylogenetic tree reveals extensive intermixing of ESBL-producing *E. coli* isolates across sources (clinics, wastewater, cattle, and wildlife) and regions (Fig. 2). Clade 1 is the most diverse and includes isolates primarily from phylogroups A, B1, and D, as well as less prevalent sequence types (ST10, ST38, and ST69). Clade 2 is dominated by ST131 and ST1193 isolates, with a marked concentration of clinical isolates. Notably, isolates from wastewater are widely dispersed in the tree. This holds for isolates from WWTPs with (e.g., Zurich) and without hospitals (e.g., Altenrhein) in their catchment areas, reflecting their genetic diversity independently from the hospital wastewater input. Wildlife and cattle isolates are similarly interspersed, with cattle isolates predominantly clustering within phylogroup B1. No exclusive clades were formed by isolates from a single source or region.

### Antibiotic-resistant genes associated with ESBL-*E. coli*

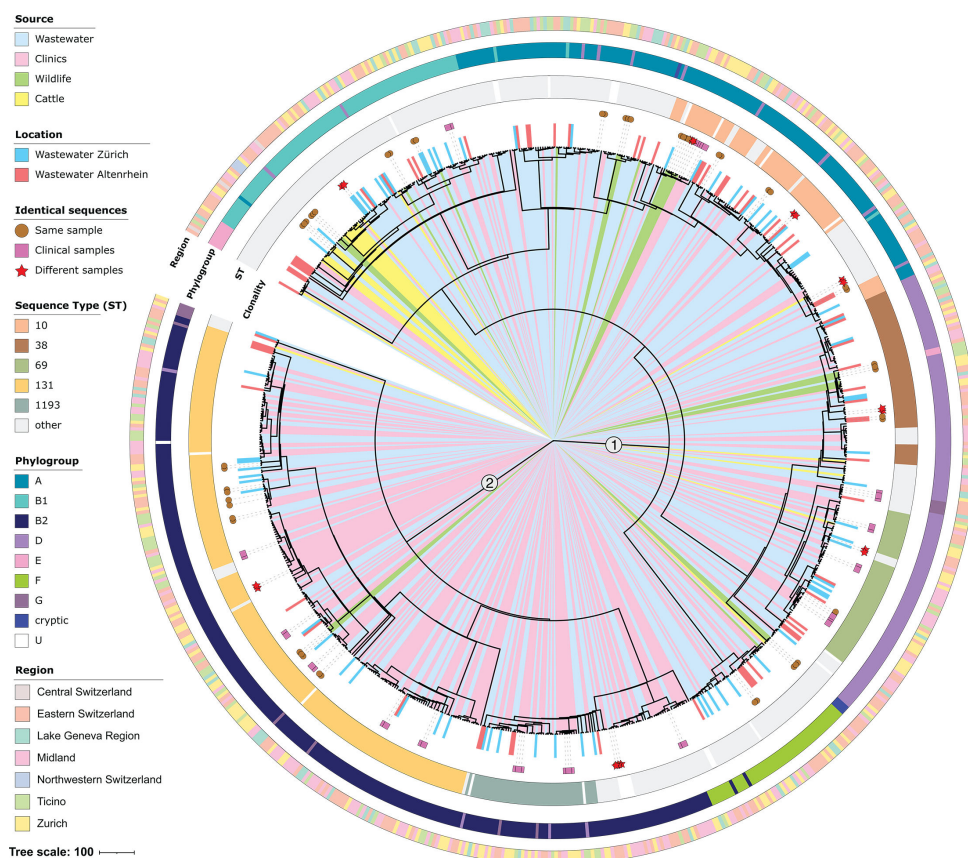
A total of different 46  $\beta$ -lactamase genes were detected among the ESBL-*E. coli* isolates (Fig. 3A; Table S7). In addition to ESBL-associated *bla*<sub>CTX-M</sub> genes, other  $\beta$ -lactamase gene



**FIG 1** Geographic distribution of sampled ESBL-producing *E. coli* isolates across Switzerland. Map of Switzerland showing the distribution of sampled ESBL-*E. coli* isolates, with the size of the circle indicating the number of isolates and the color representing the source. The sources include clinics (red), cattle (yellow), wastewater (blue), and wildlife (green). The locations of six wastewater treatment plants (WWTPs) from which ESBL-*E. coli* are isolated are marked by their corresponding catchments: WWTP Zurich, WWTP Altenrhein, WWTP Chur, WWTP Lugano, WWTP Sensetal-Laupen, and WWTP Geneva. Clinical settings from where the isolates were collected are labeled in red: Ente Ospedaliero Cantonale (EOC), Kantonsspital Graubünden (KSGR), Medizinische Laboratorien Niederwangen (MCL), Inselspital Bern (ISB), University Hospital Zurich (USZ/UZH), and Zentrum für Labormedizin (ZLMSG). The map was generated in R (v4.1.1) and modified in Inkscape (v1.1.1).

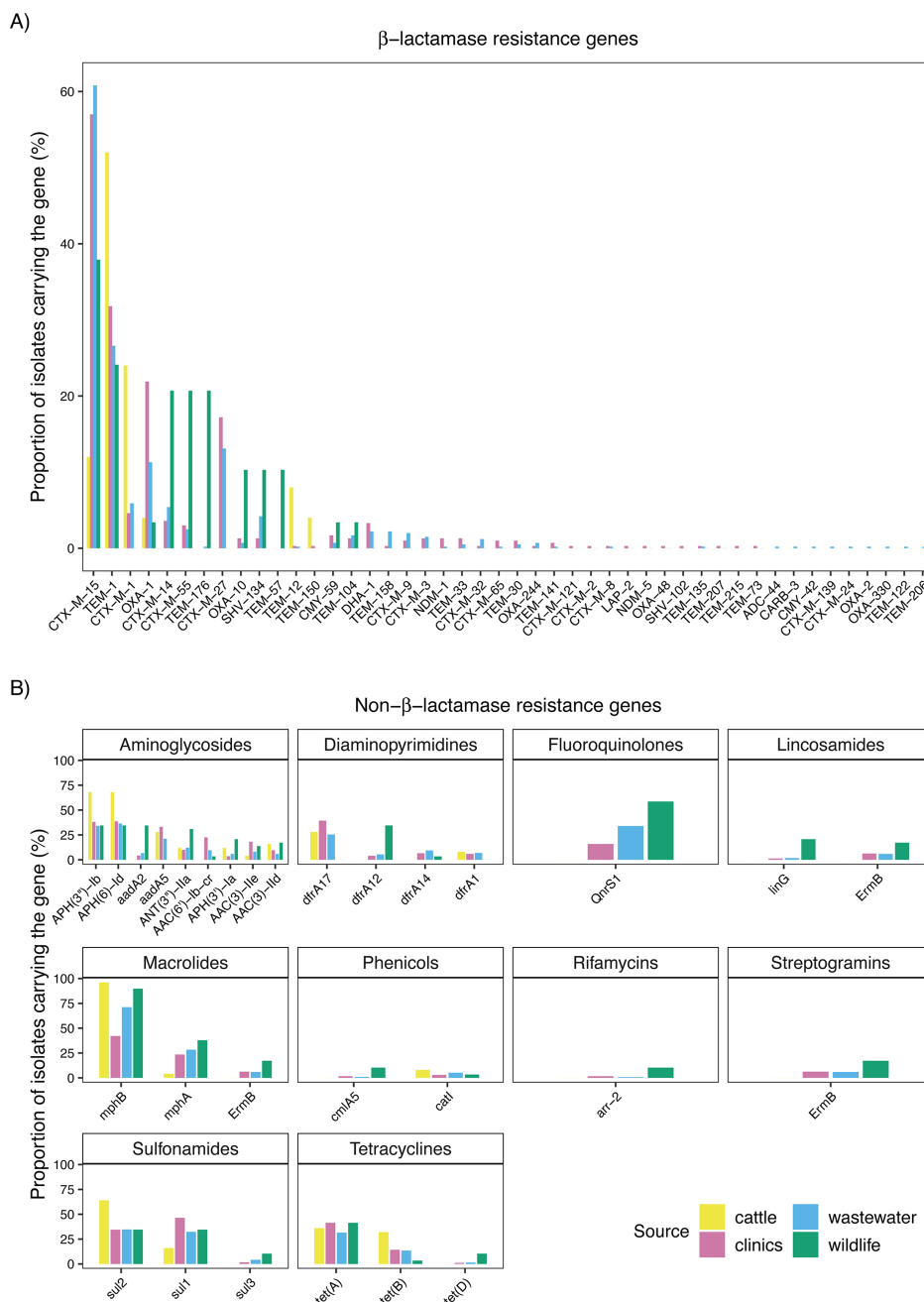
families such as *bla*<sub>TEM</sub>, *bla*<sub>OXA</sub>, and *bla*<sub>SHV</sub> were also identified, with some gene families showing frequent co-occurrence within individual isolates (Fig. S3). *bla*<sub>CTX-M-15</sub> was the most prevalent ESBL gene, detected in 61% of wastewater, 57% of clinical, 38% of wildlife, and 12% of cattle isolates, followed by *bla*<sub>CTX-M-1</sub>, which was common in cattle (24%), present in wastewater (6%) and clinics (5%) but absent in wildlife. Additionally, 14 other *bla*<sub>CTX-M</sub> type genes were identified in all sources except cattle. The most prevalent were *bla*<sub>CTX-M-14</sub>, found in 21% of wildlife, 6% of wastewater, and 4% of clinical isolates, and *bla*<sub>CTX-M-55</sub>, which was present across wildlife (21%), wastewater (3%), and clinical sources (3%). *bla*<sub>SHV</sub> type genes were rare, with only *bla*<sub>SHV-102</sub> found exclusively in clinical isolates and *bla*<sub>SHV-134</sub> detected in clinical, wastewater, and wildlife isolates. Among non-ESBL  $\beta$ -lactamases, *bla*<sub>TEM-1</sub> was the dominant *bla*<sub>TEM</sub> type gene, most frequently found in cattle (52%). *bla*<sub>OXA-1</sub>, also a non-ESBL gene, was the most prevalent *bla*<sub>OXA</sub> type gene, detected in all sources but mainly in clinical (22%) and wastewater (11%) isolates. *bla*<sub>OXA-10</sub> was predominantly found not only in wildlife (10%,  $n = 3$ ) but also in clinics (1%,  $n = 4$ ) and wastewater (1%,  $n = 3$ ). The carbapenem-hydrolyzing  $\beta$ -lactamase genes *bla*<sub>OXA-2</sub>, *bla*<sub>OXA-244</sub>, and *bla*<sub>OXA-48</sub> were found at low prevalence (<1%) in clinics and wastewater. Carbapenemase genes NDM-1 and NDM-5 were detected at low prevalence (<1%), while CARB-3 was identified in a single wastewater isolate.

Resistance genes were detected across 11 antibiotic classes beyond  $\beta$ -lactams based on annotations from the Comprehensive Antibiotic Resistance Database (CARD, accessed 4 November 2024). These included macrolides (73% of isolates), aminoglycosides (64%), sulfonamides (55%), diaminopyrimidines (52%), tetracyclines (50%), fluoroquinolones (30%), lincosamides (9%), phenicolis (8%), streptogramins (6%), rifamycins (2%), and phosphonic acids (0.7%), with notable variation across sources (Fig. 3B; Table 1). Macrolide resistance genes were present in 100% of wildlife isolates, 96% of cattle isolates, 79% of wastewater isolates, and 60% of clinical isolates. Similarly, aminoglycoside resistance genes were more frequent in cattle (84%) and wildlife (83%) than in clinics (70%) and wastewater (56%). Fluoroquinolones, lincosamides, phosphonic acids, streptogramins,



**FIG 2** Phylogenetic structure and source distribution of ESBL- *E. coli* isolates. Maximum-likelihood phylogenetic tree of the core genome of 762 ESBL-*E. coli* isolates, illustrating phylogroups, sequence types (STs), collection sources, geographic regions, and clonal relationships. Each leaf corresponds to a distinct isolate, with source indicated by the inner color band extending from the tree center to the leaf label: wastewater (light blue), clinical settings (pink), cattle (yellow), and wildlife (green). Wastewater isolates from Zurich and Altenrhein treatment plants are specifically annotated in blue and red, respectively. Numbers “1” and “2” on internal branches mark the ancestral nodes defining the two major clades discussed in the results. Identical isolates are indicated in the “Clonality” ring as: same sample (brown circles; applicable to wastewater, wildlife, and cattle, where up to three isolates per sample were selected; not applicable to clinical samples, for which one isolate per sample was selected), multiple clinical samples (pink squares; identical isolates recovered across distinct clinical samples), and different sources (red stars; identical isolates spanning source categories). Major STs ( $\geq 5\%$  prevalence: ST38, ST69, ST131, ST10, ST1193) are highlighted with specific colors, while less frequent STs are shown in grey and undefined STs in white. Phylogroups (A, B1, B2, D, E, F, G, cryptic, and unknown) are represented by distinct colored rings encircling the tree. The outer ring indicates the geographic region where each isolate was collected.

and rifamycin resistance genes were absent in cattle but detected at varying levels in other sources. Fluoroquinolone resistance genes were present in 59% of wildlife, 39% of wastewater, and 19% of clinical isolates. Lincosamides resistance genes were most common in wildlife (38%) but rare in clinics (8%) and wastewater (8%). Streptogramin genes followed a similar pattern, with 17% in wildlife and 6% in both clinics and wastewater, while rifamycin resistance genes were found only in a small proportion of wildlife (10%), clinical (3%), and wastewater (1%) isolates. Wildlife had the highest prevalence of phosphonic acid resistance genes (10%) though they were also detected at low levels in clinical (0.3%) and wastewater (0.2%) isolates. Seventy-six percent (76%) of isolates were multidrug-resistant, carrying resistance genes to at least three antibiotic classes. This analysis considered resistance genes across the 11 non- $\beta$ -lactam classes listed above (macrolides, aminoglycosides, sulfonamides, diaminopyrimidines, tetracyclines, fluoroquinolones, lincosamides, phenicols, streptogramins, rifamycins, and phosphonic



**FIG 3** Distribution of  $\beta$ -lactamase and other antibiotic resistance genes across sources. (A) Proportion of isolates from clinics (pink), cattle (gold), wildlife (green), and wastewater (blue) carrying specific  $\beta$ -lactamase genes, identified using the Comprehensive Antibiotic Resistance Database (CARD). Each bar represents the proportion of isolates within each source harboring the corresponding gene. (B) Proportion of isolates carrying resistance genes from other antibiotic classes, identified using CARD. Only genes present in  $\geq 5\%$  of isolates in at least one source are shown. Antibiotic classes include aminoglycosides, diaminopyrimidines, fluoroquinolones, lincosamides, macrolides, phenicols, phosphonic acids, rifamycins, sulfonamides, tetracyclines, and streptogramins. *ermB* is displayed in the lincosamides, streptogramins, and macrolides panels because it was annotated as conferring resistance to all three antibiotic classes. For an expanded CARD analysis, see Fig. S8.

acids), as well as  $\beta$ -lactamases. Notably, four isolates exhibited resistance genes to up to nine classes (Table S8). The number of ARGs per isolate was similar across clinics, cattle, wastewater, and wildlife in most Swiss regions, with no significant differences detected in Eastern Switzerland, Lake Geneva Region, or Ticino (Kruskal-Wallis,  $P > 0.3$ ). Marginally significant differences were observed in Zurich ( $P = 0.03$ ) and Midland ( $P = 0.04$ ), but no

**TABLE 1** Number and percentage of isolates carrying resistance genes for each antibiotic class, with percentages also stratified by source (clinics, cattle, wastewater, and wildlife)

Drug class	Total no. of isolates	Total % of isolates	% of clinical isolates (n = 302)	% of cattle isolates (n = 25)	% of wastewater isolates (n = 406)	% of wildlife isolates (n = 29)
Aminoglycosides	484	63.5	69.5	84	56.4	82.8
Diaminopyrimidines	396	52	57.9	40	49.3	37.9
Fluoroquinolones	232	30.4	18.9	0	38.9	58.6
Lincosamides	66	8.7	7.9	0	7.6	37.9
Macrolides	556	73	59.9	96	79.3	100
Phenicol	64	8.4	6.3	8	9.6	13.8
Phosphonic acids	5	0.7	0.3	0	0.2	10.3
Rifamycins	16	2.1	3	0	1	10.3
Streptogramins	48	6.3	6.3	0	5.9	17.2
Sulfonamides	422	55.4	59.9	68	51	58.6
Tetracyclines	382	50.1	54.6	68	45.3	55.2

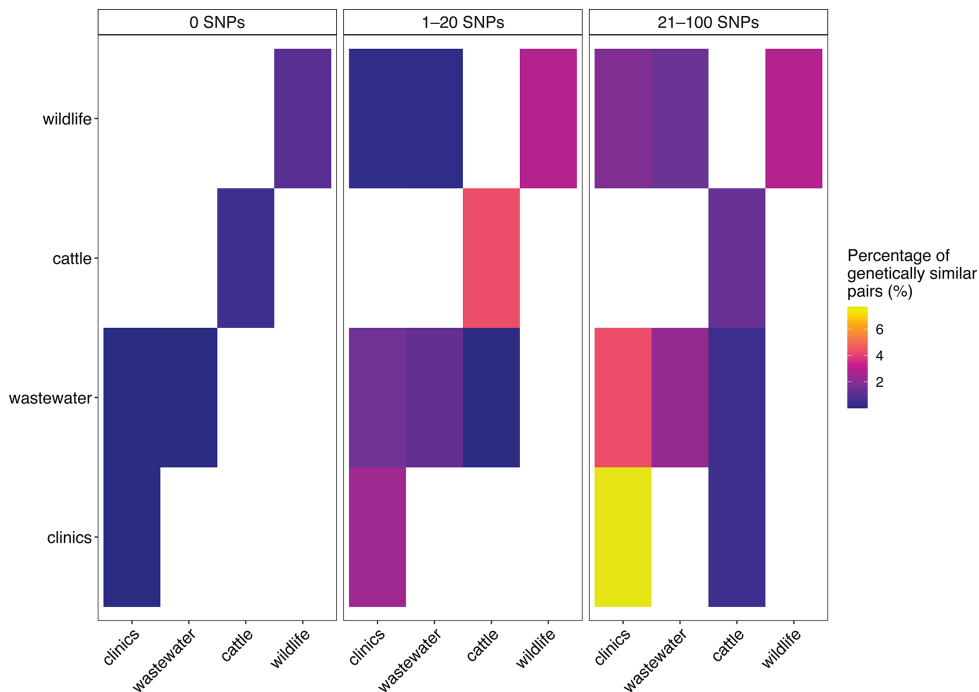
pairwise comparison remained significant after global Bonferroni correction (Table S9; Fig. S4).

### Genetic similarity

Across all SNP thresholds, we observed 5,272 within-compartment and 5,318 between-compartment genetically similar pairs at 21–100 SNPs, 2,050 within-compartment and 1,841 between-compartment pairs at 1–20 SNPs, and 39 within-compartment and 2 between-compartment pairs at 0 SNPs (Table S10). Within compartments, the proportion of genetically similar pairs calculated relative to all possible isolate pairs per compartment, to account for differences in sampling density, varied by source (Fig. 4), with 7.7% of clinical isolates forming genetically similar pairs at 21–100 SNPs, compared to 2.3% in wastewater, 1.3% in cattle, and 2.9% in wildlife (Table S10). The proportion of genetically similar pairs at stricter thresholds (1–20 and 0 SNPs) decreased across all compartments, with 5 zero-SNP clonal pairs in wastewater, 1 in wildlife, 1 in cattle, and 32 in clinics. Between compartments, the highest proportion of genetically similar pairs was observed between clinics and wastewater (4.3% at 21–100 SNPs), followed by clinics-wildlife (1.9%), wastewater-wildlife (1.4%), and cattle-wastewater (0.4%) (Fig. 4). Genetically identical isolate pairs (0 SNPs) were detected both within and between compartments (Table S11). The majority were observed within the same compartment, with multiple matches in clinics, as well as in wastewater, wildlife, and cattle. Clonal pairs between compartments were rare, with only two cases observed: one clinical isolate from St. Gallen that was detected in the wastewater of the nearest treatment plant (Altenrhein) and one from a Zurich hospital found in Zurich wastewater.

### Ancestral state reconstruction and transition analysis

Ancestral state reconstruction on 1,000 bootstrap maximum likelihood (ML) phylogenetic trees using the parsimony criterion demonstrated that transitions between sources in the full phylogeny of 792 ESBL-producing *E. coli* were not uniformly distributed (Fig. 5). Z-score analysis was applied to compare observed transitions to the null distribution based on the approximation that Z-scores follow a standard normal distribution. This approach was supported by the near-normal distribution of transition counts under randomization in the consensus phylogeny (Fig. S5). Using this framework, transitions between community (wastewater) and clinics, as well as from wildlife to clinics, occurred significantly less frequently than expected under the null model ( $Z < -1.96$ ). Transitions between other source pairs on the full phylogeny fell within the range expected under random permutations, with the exception of wildlife to cattle, which exhibited a slight excess of transitions relative to the null model, exceeding the central 95% confidence interval in 292 out of 1,000 bootstrap trees (Fig. 5; Table S12). Transitions involving



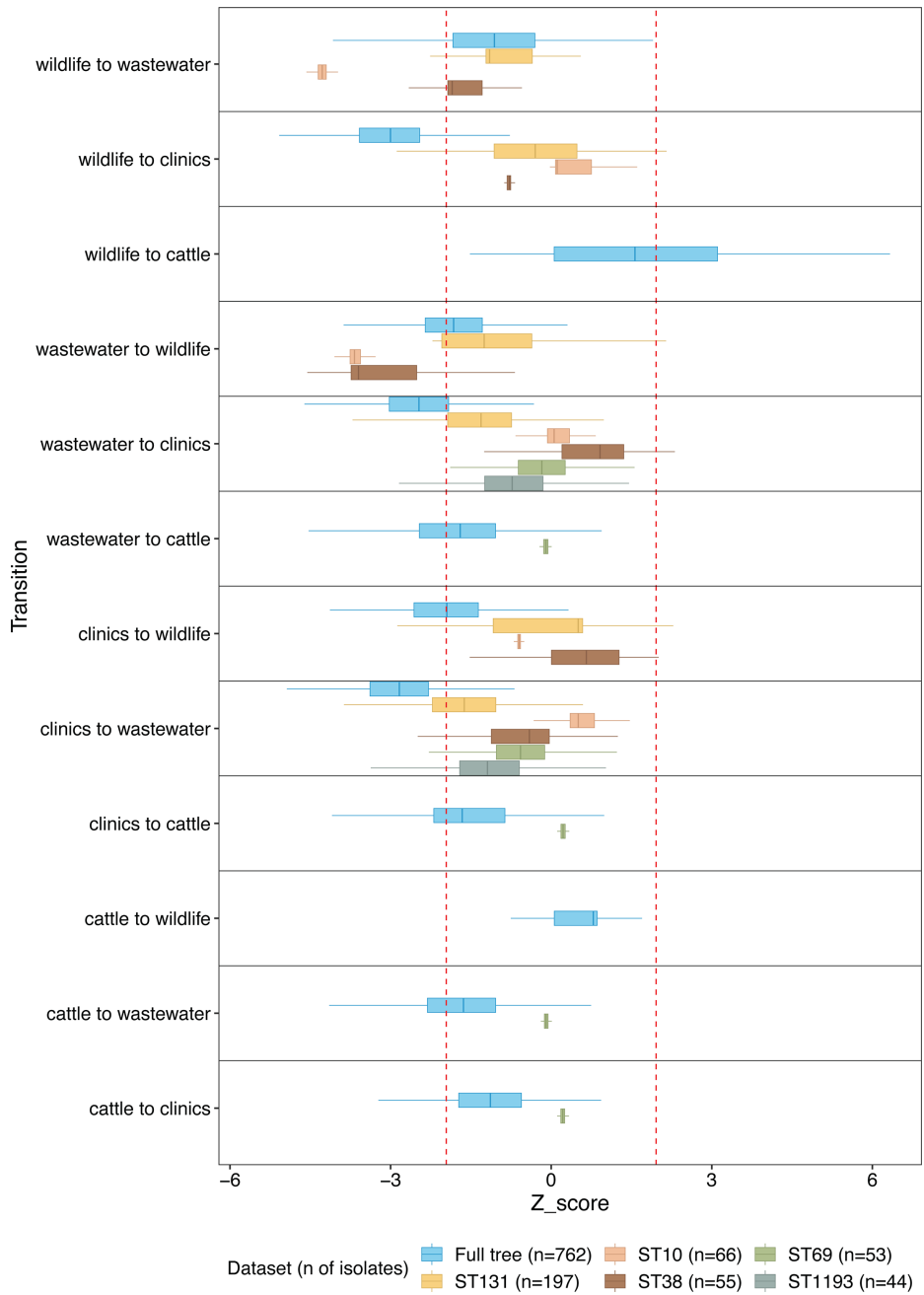
**FIG 4** Within- and between-compartment genetic similarity. Percentage of genetically similar ESBL-*Escherichia coli* isolate pairs within and between compartments across SNP thresholds. Diagonal cells represent within-compartment percentages; off-diagonal cells represent between-compartment percentages. White tiles indicate unobserved or zero values.

wildlife and cattle were based on fewer isolates due to the smaller representation of these compartments in the phylogeny (Table S1). For transitions involving wildlife and cattle, the null distributions were skewed and bounded at zero (Fig. S5), meaning that Z-scores in these cases must be interpreted cautiously, as decreases in transition frequency cannot be reliably detected.

When analyzed within the five most dominant sequence types (ST131, ST10, ST38, ST69, ST1193), the non-random patterns observed in the full data set were not retained. Transitions between community (wastewater) and clinics occurred at frequencies consistent with random expectations in all five STs, with no significant deviations observed. In ST10, however, transitions between wildlife and community were significantly underrepresented in both directions ( $Z < -1.96$ ). A similar pattern was observed in ST38, where transitions from community to wildlife occurred significantly less frequently than expected. Transitions among all other source pairs within individual STs, including between cattle, wildlife, clinics, and community, did not deviate from random expectations. The same pattern was observed when analyses were restricted to a subtree including only wastewater ( $n = 406$ ) and clinical ( $n = 302$ ) isolates, to control for potential sampling biases (Fig. S6). On the full subset phylogeny, transitions between wastewater and clinics were consistently less frequent than expected under the null model, across all 1,000 bootstrap trees (Table S13). Within sequence types, however, transitions between these sources conformed to random expectations.

## DISCUSSION

Our study highlights that within clinically relevant human-associated sequence types (STs), ESBL-producing *E. coli* circulate between clinics and the broader community, with wastewater surveillance capturing strains that overlap with those found in clinical settings. Although the phylogeny showed extensive intermixing across sources and no source- or region-specific clades, genetic similarity was higher within than between compartments. Clonal isolates (0 SNPs) were rare between compartments ( $n = 2$ , both between clinics and wastewater) in contrast to 32 within-clinic and a few within other



**FIG 5** Z-score distributions of collection source transitions across 1,000 bootstrap phylogenies for the full data set and dominant sequence types (STs). Transitions between collection sources (clinics, wastewater, wildlife, cattle) were inferred using maximum parsimony ancestral state reconstruction across 1,000 bootstrap maximum likelihood (ML) phylogenetic trees. For each tree, tip states were randomized 1,000 times preserving source frequencies to generate a null distribution of transition counts. Z-scores represent standardized deviations of observed transitions relative to the null. Each boxplot shows Z-score distributions per transition type, stratified by data set subset (full data set and the five dominant STs: ST131, ST10, ST38, ST69, ST1193). The legend shows the number of isolates retained in the analysis for each ST. Red dashed lines indicate  $\pm 1.96$ , corresponding to the central 95% confidence interval under the null model. Transitions were reconstructed using the *acctran* and *ancestral.pars* functions from the *ape* package (v5.7.1) in R (v4.1.2). Positive Z-scores indicate transitions occurring more frequently than expected by chance; negative Z-scores indicate less frequent transitions.

sources. Ancestral state reconstruction on the whole phylogeny revealed significantly fewer transition events than expected by chance between clinics and both wastewater

(used as a proxy for community shedding) and wildlife, suggesting partial compartmentalization and limited spillover, and indicating potential epidemiological separation. The limited connectivity between wildlife and clinical settings may reflect reduced exposure opportunities between wildlife and the human-built environment in Switzerland. Conversely, transitions between livestock and wildlife exceeded the null expectation in a small subset of bootstrap trees, which may indicate some degree of exchange or reflect reduced power to detect significant structure due to small sample sizes. Similarly, the observation that transitions between wildlife or livestock and wastewater fell within the range expected under randomization may reflect the fact that wastewater is not entirely distinct from these sources, as it can receive input from agricultural runoff and surrounding environments, such as slaughterhouses.

When analyzed by sequence type (ST), this structured signal diminished. ST131, the most prevalent lineage (26%), showed no deviation from random expectations in transitions between clinics and the community (as captured by wastewater), suggesting free circulation between symptomatic and asymptomatic hosts, and highlighting the potential of wastewater as an indicator of human gut colonization (23). Downstream of WWTPs in Switzerland, ST131 has been detected in rivers and other surface waters, indicating environmental dissemination (24). ST131 dominated clinical isolates, but it was absent from cattle, aligning with studies reporting genetic divergence between human- and cattle-derived *E. coli* strains (25). A slight, non-significant excess of transitions from clinics to wildlife was also observed, suggesting rare spillover events. Supporting this, one black-headed gull isolate belonging to ST131 clustered with wastewater and clinical isolates, suggesting cross-compartment interactions. Wildlife may act as a reservoir for clinically relevant AMR, particularly in species interacting with WWTPs, consistent with evidence of human-wildlife AMR exchange in urban birds in Kenya and of gulls accumulating AMR from WWTP sediment basins in Sweden (26–29). Nonetheless, transmission from wildlife to clinics or wastewater appeared rare (11).

ST1193 and ST69, two other human-associated lineages (30, 31), also showed transition frequencies between clinics and community consistent with random expectations, supporting their capacity to circulate across these compartments resulting in their epidemiological success (31–33). These results suggest that the separation observed in the full tree is not due to universal barriers but rather reflects lineage-specific ecological behavior.

Genetically identical strains were rarely shared between compartments. In both observed zero-SNP matches between wastewater and clinical isolates, the wastewater isolate was detected 2–3 weeks after the clinical counterparts, suggesting that clinical cases may contribute to subsequent detection in wastewater through shedding (34) or that these clones are more broadly circulating within the community, including via asymptomatic carriage. The rare detection of matched isolates between compartments may be influenced by the frequency and timing of sampling. Our sampling includes intermittent and repeated wastewater samples during a time frame when clinical submissions were ongoing. Timing wastewater sample collections to more closely align with the collection of clinical submissions may improve the likelihood of detecting connectivity between compartments. The highest proportion of genetically similar between-compartments pairs occurred between clinics and wastewater (4.3% at 21–100 SNPs), suggesting that exchange occurs mostly through closely related but non-identical strains, likely due to indirect transmission, rapid divergence following environmental entry, or limited sampling resolution. These findings emphasize how SNP thresholds shape interpretations of genetic relatedness and underscore the need for context-specific thresholds in genomic surveillance (35).

$\beta$ -Lactamase genes were widely distributed across all sources, with 46 genes detected, consistent with global reports of widespread  $\beta$ -lactam resistance (36, 37). *bla*<sub>CTX-M-15</sub> was the most prevalent, detected in 61% of wastewater and 57% of clinical isolates, reinforcing its dominance in human-associated environments (38). Its abundance in both compartments suggests wastewater may reflect the broader community

reservoir and detect resistance circulating before clinical detection. Supporting this, wastewater from the Altenrhein catchment (without hospital input) showed comparable ESBL-*E. coli* diversity to that in Zurich (with hospital input), highlighting the contribution of community sources and the value of wastewater monitoring for population-level AMR surveillance (39). In cattle, *bla*<sub>CTX-M-1</sub> was common, consistent with its prevalence in food-producing animals (40), while *bla*<sub>CTX-M-14</sub>, previously reported in cattle, was absent, indicating possible temporal shifts in gene prevalence. Notably, in Switzerland, ESBL-producing Enterobacterales are more prevalent in pigs and poultry than in cattle (41). In wildlife, *bla*<sub>CTX-M-15</sub>, *bla*<sub>TEM</sub>, *bla*<sub>SHV</sub>, and *bla*<sub>OXA-10</sub> were present, but *bla*<sub>CTX-M-1</sub> was not, mirroring regional differences observed in wild ruminants and boar in Germany (42, 43). In Switzerland, wild birds have been identified as reservoirs of ESBL-producing Enterobacterales (44, 45). These results point to wildlife as a reservoir for specific resistance genes and emphasize the need for cross-sectoral AMR monitoring in line with One Health principles (11, 46). Among the 76% of isolates in our data set carrying resistance to at least three antibiotic classes, consistent with global reports estimating a 66% prevalence of multidrug resistance among ESBL-*E. coli* (32, 47–49), one duck-derived isolate exhibited genotypic resistance to nine antibiotic classes. This extreme genotype reinforces the role of wildlife in AMR dissemination and underscores growing global concerns about the environmental spread of resistance (50, 51).

Although we systematically collected wastewater isolates, cattle, wildlife, and clinical isolates were obtained through convenience sampling, potentially limiting representativeness and introducing source-specific biases. A subset of clinical isolates was restricted to invasive infections and did not include colonization, which may underestimate within-clinic diversity. Furthermore, our clinic-wastewater comparisons rely on the assumption that isolates from a given hospital primarily enter the WWTP serving the same city. This assumption may introduce bias due to patients commuting between catchments or due to misalignment between the hospital catchment and the home catchment of patients. We note we did not have access to home addresses to ensure patient anonymity. The small sample sizes for wildlife and cattle may have underestimated AMR diversity and reduced statistical power to detect transitions involving these compartments. Furthermore, livestock sampling focused exclusively on cattle, omitting pigs and poultry, which may have closer epidemiological links to humans. Wildlife samples were primarily collected outside WWTP catchments, potentially overlooking interactions with peri-urban or slaughterhouse environments that may serve as transmission interfaces. Moreover, while we interpret ESBL-*E. coli* in wastewater as reflecting community shedding, wastewater may also include hospital, industrial, and animal inputs, and strains may persist in sewage networks independent of human input via mechanisms such as biofilm formation (52–54). Finally, this study focused on bacterial transmission rather than plasmid-mediated transfer of ESBL genes. Future work investigating mechanisms by which resistance determinants move across compartments, including via plasmids, would benefit from further incorporation of plasmid-typing or long-read sequencing.

Our results indicate that while transitions between sources occur, compartments remain partially structured on the full phylogeny, particularly between clinics and wastewater. This separation disappears within clinically relevant human-associated sequence types (ST131, ST69, ST1193), where the absence of structure suggests dynamic circulation between clinical settings and the broader community. Notably, wastewater isolates captured lineages also detected in clinical settings, supporting its utility as a proxy for community-level AMR carriage. Although zero-SNP matches between compartments were rare, the predominance of genetically similar isolates within clinics and wastewater points to localized clonal expansion and potential indirect exchange. These findings highlight how the interpretation of compartmental connectivity depends on the analytical resolution used and underscore the importance of context-specific SNP thresholds in One Health genomic surveillance (35). While expanding surveillance in underrepresented reservoirs such as wildlife and cattle remains important, our study

reinforces the value of wastewater monitoring for capturing clinically relevant ESBL-*E. coli* circulating in the human population and supports its integration into national One Health surveillance frameworks.

## MATERIALS AND METHODS

### Sample collection

A total of 834 ESBL-*E. coli* isolates were collected from wastewater ( $n = 438$ ), healthcare facilities ( $n = 328$ ), cattle ( $n = 26$ ), and wildlife ( $n = 42$ ) across Switzerland from 2021 to 2023 (Fig. 1; Table S1). Cattle and wildlife isolates were obtained through two independent field studies conducted during the same period, which focused on antimicrobial resistance in Swiss dairy herds and free-roaming wild animals, including mammals and birds (55, 56), and correspond to the same isolates processed and identified in a previous study (57). Wastewater samples were obtained from six wastewater treatment plants (WWTPs) using 24-h flow-proportional composite sampling. Samples were transported to the laboratory on ice, stored at 4°C, and processed within 48 h. To isolate ESBL-*E. coli*, 100  $\mu$ L of undiluted wastewater was plated onto CHROMagar Orientation supplemented with ESBL selective supplement (CHROMagar, France). Presumptive ESBL-*E. coli* colonies were identified by their characteristic dark pink to reddish coloration. Every 4 weeks, three to eight colonies per WWTP were streaked on LB agar (Merck KGaA, Germany), incubated at 37°C for 20–24 h, and further cultured in Luria Broth (Merck KGaA, Germany). Single-colony cultures were preserved in 12% glycerol stocks at –70°C for downstream analyses. Clinical isolates were retrospectively collected from six healthcare facilities across Switzerland: Ente Ospedaliero Cantonale (Ticino), Kantonsspital Graubünden (Graubünden), MCL Medizinische Laboratorien Niederwangen (Bern), Inselspital (Bern), University Hospital Zurich (Zurich), and Zentrum für Labormedizin (St. Gallen). All clinical isolates were confirmed ESBL-producing *E. coli* as part of routine diagnostic workflows and were selected accordingly for this study. In facilities where the number of available ESBL-*E. coli* isolates exceeded sequencing capacity (e.g., Zurich, Bern, and Chur), a random subset was selected. Isolates originated from individuals screened for ESBL resistance were revived from 12% glycerol stocks or Microbank Cryovial Storage Systems (Pro-Lab Diagnostic, Canada), streaked onto LB agar, incubated at 37°C for 20–24 h, and cultured in Luria Broth for DNA extraction.

### DNA extraction and WGS

DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. WGS was performed using the LITE Library Prep method on an Illumina NextSeq 1000 P2 flow cell (300 bp paired-end) at the Earlham Institute, UK (58). Sequencing was conducted in two batches (June 2023 and January 2024) for 834 ESBL-producing *E. coli* isolates. Unique Dual Indexes were used to minimize index hopping (59). Quality control measures included duplicate samples ( $n = 15$  pairs), negative extraction controls (AE blanks), and reference isolates of *E. coli* (strain 1 and 2) and *Klebsiella pneumoniae* (strain 1 and 2). Negative controls ( $n = 12$ ; 8 in batch 1, 4 in batch 2) were processed alongside study samples to monitor for contamination or sequencing artifacts. Reference *E. coli* strains (strain 1:  $n = 8$  in batch 1,  $n = 2$  in batch 2; strain 2:  $n = 8$  in batch 1,  $n = 1$  in batch 2) had been previously sequenced using alternative platforms and were included to confirm sequencing consistency. *K. pneumoniae* strains (strain 1:  $n = 8$  in batch 1,  $n = 1$  in batch 2; strain 2:  $n = 5$  in batch 1,  $n = 2$  in batch 2) were used to verify that the sequencing and bioinformatics pipelines could accurately distinguish *E. coli* from other species.

## Bioinformatic analyses and isolates characterization

The bioinformatic workflow is summarized in Fig. S7. Illumina reads were trimmed with trimmomatic v 0.35 in paired-end mode, removing Illumina Nextera adapters with specific trimming parameters (2:30:10:1) and applying quality-based trimming with a 4 bp sliding window and an average quality threshold of 20 (60). Reads shorter than 36 bp were removed. Read quality before and after trimming was assessed using FastQC v0.11.4 (61). Screening for contamination was performed using FastQ Screen v0.15.3 (62), which maps reads against a panel of reference genomes to detect potential co-isolated or misidentified species. To this end, we downloaded seven reference genomes from NCBI corresponding to bacterial species known to grow on CHROMagar ESBL plates and potentially co-occur with *E. coli*: *Acinetobacter baumannii* (ASM863263v1), *Citrobacter freundii* (ASM381234v1), *Escherichia coli* (MG1655), *Enterobacter roggenkampii* (ASM172980v1), *Klebsiella pneumoniae* (ASM24018v2), *Pseudomonas aeruginosa* (ASM676v1), and *Proteus mirabilis* (ASM6996v1). Filtered reads were *de novo* assembled using SPAdes v3.15.9, without further error correction and set to “careful” (63). Assembly quality was assessed with QUAST v5.2.0 (Table S3) (64). Genome annotation was performed using Bakta v1.9.1, specifying *E. coli* as the genus and species (65). Sequence types (STs) were assigned using MLST v2.23.0 with the Achtman scheme (<https://github.com/tseemann/mlst>). Phylogenetic groups were determined using EzClermont v0.6.3 based on an *in-silico* PCR approach (66). Serotyping was performed using *ecoli*-serotyping ([https://github.com/phac-nml/ecoli\\_serotyping](https://github.com/phac-nml/ecoli_serotyping)).

Antibiotic-resistant genes (ARGs) were identified using Abricate v1.0.1 (67), blasting against the Comprehensive Antibiotic Resistance Database (CARD, accessed 4th November 2023) (68). ARGs were classified in antibiotic classes based on CARD annotations. Genes conferring resistance to at least four different classes and identified as efflux pumps were classified separately. Efflux pumps and genes intrinsic to *E. coli* were excluded from the main analyses but included in the supplementary material (Fig. S8). Statistical differences in the prevalence of ARGs among sources and regions were assessed using a Kruskal-Wallis and Dunn’s pairwise test with global Bonferroni adjustment. Regions are displayed in Fig. S1.

## Phylogenetic reconstruction and genetic similarity

A pangenome was defined using Roary v3.13.0, identifying a total of 29,477 gene clusters, of which 2,312 core genes (present in  $\geq 99\%$  of isolates) were retained for phylogenetic analysis. The core genome was aligned using MAFFT (69) and concatenated into a single alignment comprising 2,099,419 nucleotide positions. Recombinant regions were identified and masked using ClonalFrameML v1.13 (70). A maximum-likelihood phylogeny was constructed using IQ-TREE with automatic model selection and rooted with an *Escherichia albertii* reference genome (NCBI Accession: [ASM1690475v2](https://ncbi.nlm.nih.gov/nucl/ASM1690475v2)) included in the pangenome analysis (71). The resulting tree was annotated and visualized using iTOL (72). Pairwise SNP differences between isolates from the same or different compartments (clinics, cattle, wastewater, and wildlife) were computed using snp-dists v0.8.2 on the recombination-free core genome alignment (<https://github.com/tseemann/snp-dists>). Identical sequences from the same sample were collapsed post-SNP calculation, retaining one representative per sample to avoid inflation of pairwise comparisons in the analysis of genetically similar pairs (Table S14). The proportion of genetically similar pairs was calculated in R (v4.3.1), within and between compartments, using SNP thresholds of 0, 1–20, and 21–100 SNPs to assess genetic similarity and potential transmission links. These thresholds span both the typical outbreak-oriented relatedness thresholds used in clinical investigations ( $\leq 20$  SNPs) and the broader  $\leq 100$  SNP window used in One Health studies to identify cross-source linkages (35).

## Linkage between clinical isolates and wastewater catchments

Patient residential addresses were not available to ensure sample anonymity, and thus frequent inter-canton commuting in Switzerland may introduce bias in spatial matching of isolates. Clinical isolates were compared with wastewater from the WWTP catchment within which each hospital is located, under the assumption that isolates from a given hospital primarily enter its corresponding WWTP catchment area.

## Ancestral state reconstruction and transition analysis

Ancestral state reconstruction of isolate collection sources (clinics, wastewater, cattle, wildlife) was conducted using maximum parsimony to infer the most likely source transitions along phylogenetic tree branches (73). Wastewater was used as a proxy for community-level ESBL-*E. coli*, as it captures bacteria shed by the general population. Tip states were assigned based on isolate origin and mapped onto both the consensus maximum-likelihood (ML) phylogeny and 1,000 bootstrap ML trees. Ancestral states at internal nodes were inferred using the *acctran* and *ancestral.pars* functions in the *ape* package (v5.7.1) within R (v4.1.2) (74). Transitions between sources were quantified for each phylogeny by identifying changes in reconstructed states along branches, excluding self-transitions (i.e., source remaining unchanged). To evaluate whether observed transitions deviated from random expectations, a null model was implemented whereby tip states were permuted 1,000 times for each bootstrap tree, maintaining the original frequency distribution of sources. For each randomized replicate, ancestral states were re-estimated using maximum parsimony, and the number of transitions between each pair of sources was recorded, resulting in a null distribution of transition counts for every source combination. For each randomization, ancestral states were reconstructed and transitions were counted, generating a null distribution for each source pair. For each source transition, a Z-score was calculated according to the following formula:

$$Z = \frac{(T_{\text{obs}} - \mu_{\text{null}})}{\sigma_{\text{null}}},$$

where  $T_{\text{obs}}$  is the observed transition count, and  $\mu_{\text{null}}$  and  $\sigma_{\text{null}}$  are the mean and standard deviation of the null distribution, respectively. Positive Z-scores indicated transitions more frequent than expected by chance, and negative Z-scores indicated fewer transitions. Z-scores exceeding  $\pm 1.96$  were considered significant at the central 95% confidence level. Analyses were performed on the full data set and separately on the five dominant sequence types (ST131, ST10, ST38, ST69, ST1193), by pruning the phylogenies to retain only isolates from each ST. An additional analysis was conducted using only isolates from wastewater and clinical sources and to the five dominant STs within this subset. Ancestral reconstruction, randomization testing, and Z-score computation were repeated accordingly.

## ACKNOWLEDGMENTS

We thank Lea Caduff, Charlie Gan, Laura Brülisauer, Camille Hablützel, and Aurélie Holschneider for their assistance in processing wastewater samples, Gladys Martinetti Lucchini for providing clinical isolates, and Aitana Neves for providing feedback. We are grateful to the Wastewater-based Surveillance for Infectious Diseases and Surveillance (WISE) group for their valuable insights and contributions to the general discussion. We also extend our appreciation to Timothy G. Vaughan, Etthel M. Windels, and Cecilia Valenzuela Agüi for their support with the phylodynamic analyses. Finally, we thank the employees of the wastewater treatment plants IDA CDA Lugano (Ticino), ARA Werdhölzli (Zurich), ARA Chur (Graubünden), ARA Sensetal Laupen (Bern), and STEP d'Aire Genève (Geneva) for providing wastewater samples.

The work was funded through the Swiss National Science Foundation grants 192763 to T.R.J., 205933 to T.S., and 211422 to S.D.B. and through a Swiss Federal Office of Public Health grant to Christoph Ort and T.R.J.

## AUTHOR AFFILIATIONS

<sup>1</sup>Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

<sup>2</sup>Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland

<sup>3</sup>Swiss Institute of Bioinformatics, Lausanne, Switzerland

<sup>4</sup>State Laboratory Basel-Stadt, Basel, Switzerland

<sup>5</sup>Clinic for Ruminants, Vetsuisse-Faculty, University of Bern, Bern, Switzerland

<sup>6</sup>Department of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, University of Zurich, Zurich, Switzerland

<sup>7</sup>Division of Infectious Diseases, Cantonal Hospital of Grisons, Chur, Switzerland

<sup>8</sup>Institute of Medical Microbiology, University of Zurich, Zurich, Switzerland

<sup>9</sup>Ente Ospedaliero Cantonale Bellinzona, Bellinzona, Switzerland

<sup>10</sup>Institute of Microbiology, Lausanne University Hospital (CHUV) & University of Lausanne (UNIL), Lausanne, Switzerland

<sup>11</sup>Central Laboratory, Department of Microbiology, Cantonal Hospital of Grisons, Chur, Switzerland

<sup>12</sup>Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

<sup>13</sup>Institute for Infectious Diseases, University of Bern, Bern, Switzerland

<sup>14</sup>Clinical Virology, University Hospital Basel, Basel, Switzerland

<sup>15</sup>Department of Microbiology, MCL-Dr Risch AG, Niederwangen, Switzerland

<sup>16</sup>Division of Human Microbiology, Center for Laboratory Medicine, St. Gallen, Switzerland

<sup>17</sup>Division of Infectious Diseases, University Hospital Basel, University Basel, Basel, Switzerland

<sup>18</sup>Amt für Lebensmittelsicherheit und Tiergesundheit (ALT), Chur, Switzerland

## AUTHOR ORCID*s*

Sheena Conforti  <http://orcid.org/0000-0002-0173-6170>

Louis du Plessis  <http://orcid.org/0000-0003-0352-6289>

Silvio D. Brugger  <http://orcid.org/0000-0001-9492-9088>

Adrian Egli  <http://orcid.org/0000-0002-3564-8603>

Gilbert Greub  <http://orcid.org/0000-0001-9529-3317>

Jana S. Huisman  <http://orcid.org/0000-0002-1782-8109>

Roger D. Kouyos  <http://orcid.org/0000-0002-9220-8348>

Oliver Nolte  <http://orcid.org/0000-0002-1761-0812>

Sarah Tschudin-Sutter  <http://orcid.org/0000-0003-4875-5368>

Tanja Stadler  <http://orcid.org/0000-0001-6431-535X>

Timothy R. Julian  <http://orcid.org/0000-0003-1000-0306>

## FUNDING

Funder	Grant(s)	Author(s)
<a href="#">Swiss National Science Foundation</a>	192763	Timothy R. Julian
<a href="#">Swiss National Science Foundation</a>	205933	Tanja Stadler
<a href="#">Bundesamt für Gesundheit</a>		Timothy R. Julian

## AUTHOR CONTRIBUTIONS

Sheena Conforti, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review and editing | Louis du Plessis, Investigation, Methodology, Writing – review and editing | Claudia Bagutti, Resources, Writing – review and editing | Jens Becker, Resources, Writing – review and editing | Silvio D. Brugger, Resources, Writing – review and editing | Alexia Cusini, Resources, Writing – review and editing | Adrian Egli, Resources, Writing – review and editing | Valeria Gaia, Resources, Writing – review and editing | Gilbert Greub, Resources, Writing – review and editing | Cornelia Guler, Resources, Writing – review and editing | Jana S. Huisman, Methodology, Software, Writing – review and editing | Claudine Kocher, Resources, Writing – review and editing | Roger D. Kouyos, Resources, Writing – review and editing | Karoline Leuzinger, Resources, Writing – review and editing | Carola Maffioli, Resources, Writing – review and editing | Mara Neacșu, Resources, Writing – review and editing | Oliver Nolte, Resources, Writing – review and editing | Alban Ramette, Resources, Writing – review and editing | Salome N. Seiffert, Resources, Writing – review and editing | Sarah Tschudin-Sutter, Resources, Writing – review and editing | Jon Paulin Zumthor, Resources, Writing – review and editing | Tanja Stadler, Conceptualization, Project administration, Resources, Supervision, Writing – review and editing | Timothy R. Julian, Conceptualization, Project administration, Resources, Supervision, Writing – review and editing

## DATA AVAILABILITY

The sequence data are available at SRA under the BioProject ID [PRJNA1271740](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1271740). The bioinformatic pipeline, implemented using the Snakemake workflow management system, is available at <https://github.com/sheenaconforti/ecoli-wgs/>. All scripts used for data analysis and figure generation are available at [https://github.com/EawagPHH/wgs\\_esblecoli](https://github.com/EawagPHH/wgs_esblecoli).

## ADDITIONAL FILES

The following material is available [online](#).

### Supplemental Material

**Supplemental material (AEM01848-25-S0001.docx).** Fig. S1 to S8; supplemental table captions.

**Supplemental tables (AEM01848-25-S0002.xlsx).** Tables S1 to S14.

## REFERENCES

- Murray CJL, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, Han C, Bisignano C, Rao P, Wool E, et al. 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* 399:629–655. [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0)
- Levy SB, Marshall B. 2004. Antibacterial resistance worldwide: causes, challenges and responses. *Nat Med* 10:S122–9. <https://doi.org/10.1038/nm1145>
- Radhouani H, Silva N, Poeta P, Torres C, Correia S, Igrejas G. 2014. Potential impact of antimicrobial resistance in wildlife, environment and human health. *Front Microbiol* 5:23. <https://doi.org/10.3389/fmicb.2014.00023>
- Bengtsson-Palme J, Kristiansson E, Larsson DGJ. 2018. Environmental factors influencing the development and spread of antibiotic resistance. *FEMS Microbiol Rev* 42:68–80. <https://doi.org/10.1093/femsre/fux053>
- McEwen SA, Collignon PJ. 2018. Antimicrobial resistance: a one health perspective. *Microbiol Spectr* 6. <https://doi.org/10.1128/microbiolspec.rba-0009-2017>
- Djordjevic SP, Jarocki VM, Morgan B, Donner E. 2020. Genomic surveillance for one health antimicrobial resistance: understanding human, animal, and environmental reservoirs and transmission. *Handbook of Environmental Chemistry* 91:71–100. [https://doi.org/10.1007/698\\_2020\\_626](https://doi.org/10.1007/698_2020_626)
- Larsson DGJ, Flach CF. 2022. Antibiotic resistance in the environment. *Nat Rev Microbiol* 20:257–269. <https://doi.org/10.1038/s41579-021-00649-x>
- World Health Organization. 2021. WHO integrated global surveillance on ESBL producing *E. coli* using a “One Health” approach: implementation and opportunities. Available from: <http://apps.who.int/bookorders>
- Navab-Daneshmand T, Friedrich MND, Gächter M, Montealegre MC, Mlambo LS, Nhwitiwa T, Mosler HJ, Julian TR. 2018. *Escherichia coli* contamination across multiple environmental compartments (soil, hands, drinking water, and handwashing water) in urban harare: correlations and risk factors. *Am J Trop Med Hyg* 98:803–813. <https://doi.org/10.4269/ajtmh.17-0521>
- Julian TR. 2016. Environmental transmission of diarrheal pathogens in low and middle income countries. *Environ Sci Process Impacts* 18:944–955. <https://doi.org/10.1039/c6em00222f>
- Vittecoq M, Godreuil S, Prugnolle F, Durand P, Brazier L, Renaud N, Arnal A, Aberkane S, Jean - Pierre H, Gauthier - Clerc M, Thomas F, Renaud F. 2016. Antimicrobial resistance in wildlife. *Journal of Applied Ecology* 53:519–529. <https://doi.org/10.1111/1365-2664.12596>
- Heuer H, Schmitt H, Smalla K. 2011. Antibiotic resistance gene spread due to manure application on agricultural fields. *Curr Opin Microbiol* 14:236–243. <https://doi.org/10.1016/j.mib.2011.04.009>

13. von Wintersdorff CJH, Penders J, van Niekerk JM, Mills ND, Majumder S, van Alphen LB, Savelkoul PHM, Wolffs PFG. 2016. Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. *Front Microbiol* 7:174871. <https://doi.org/10.3389/fmicb.2016.00173>
14. Ling W, Peri AM, Furuya-Kanamori L, Harris PNA, Paterson DL. 2022. Carriage duration and household transmission of enterobacteriales producing extended-spectrum beta-lactamase in the community: a systematic review and meta-analysis. *Microb Drug Resist* 28:795–805. <https://doi.org/10.1089/mdr.2022.0035>
15. Woolhouse MEJ, Ward MJ. 2013. Sources of antimicrobial resistance. *Science* 341:1460–1461. <https://doi.org/10.1126/science.1243444>
16. Parkins MD, Lee BE, Acosta N, Bautista M, Hubert CRJ, Hrudehy SE, Frankowski K, Pang XL. 2024. Wastewater-based surveillance as a tool for public health action: SARS-CoV-2 and beyond. *Clin Microbiol Rev* 37:e0010322. <https://doi.org/10.1128/cmr.00103-22>
17. Philo SE, De León KB, Noble RT, Zhou NA, Alghafri R, Bar-Or I, Darling A, D'Souza N, Hachimi O, Kaya D, et al. 2024. Wastewater surveillance for bacterial targets: current challenges and future goals. *Appl Environ Microbiol* 90. <https://doi.org/10.1128/aem.01428-23>
18. Conforti S, Holschneider A, Sylvestre É, Julian TR. 2024. Monitoring ESBL-*Escherichia coli* in Swiss wastewater between November 2021 and November 2022: insights into population carriage 9:e0076023. <https://doi.org/10.1128/msphere.00760-23>
19. Conforti S, Pruden A, Acosta N, Anderson C, Buergermann H, Calabria De Araujo J, Cristobal JR, Drigo B, Ellison C, Francis Z, et al. 2025. Strengthening policy relevance of wastewater-based surveillance for antimicrobial resistance. *Environ Sci Technol* 59:2339–2343. <https://doi.org/10.1021/acs.est.4c09663>
20. Muloi DM, Jauneikaite E, Anjum MF, Essack SY, Singleton DA, Kasudi MR, Wade MJ, Egyir B, Nunn JG, Midega JT, Peacock SJ, Feasey NA, Baker KS, Zadoks RN. 2023. Exploiting genomics for antimicrobial resistance surveillance at One Health interfaces, p e1056–e1062. In *The Lancet Microbe*. Vol. 4. . Harnessing Genomics for Antimicrobial Resistance Surveillance
21. Grant K, Jenkins C, Arnold C, Green J, Zambon M, Blackburn L, Gaynor L, Kroese M. 2018. Implementing pathogen genomics Implementing pathogen genomics: a case study We acknowledge the contributions of Sobia Raza Implementing pathogen genomics: a case study
22. Gay N, Rabenandrasana MAN, Panandiniaina HP, Rakotoninidrina MF, Ramahatafandry IT, Enouf V, Roger F, Collard J-M, Cardinale E, Rieux A, Loire E. 2023. One Health compartment analysis of ESBL-producing *Escherichia coli* reveals multiple transmission events in a rural area of Madagascar. *J Antimicrob Chemother* 78:1848–1858. <https://doi.org/10.1093/jac/dkad125>
23. Zarfel G, Galler H, Feierl G, Haas D, Kittinger C, Leitner E, Grisold AJ, Mascher F, Posch J, Pertschy B, Marth E, Reinthaler FF. 2013. Comparison of extended-spectrum-β-lactamase (ESBL) carrying *Escherichia coli* from sewage sludge and human urinary tract infection. *Environ Pollut* 173:192–199. <https://doi.org/10.1016/j.envpol.2012.09.019>
24. Biggel M, Hoehn S, Frei A, Dassler K, Jans C, Stephan R. 2023. Dissemination of ESBL-producing *E. coli* ST131 through wastewater and environmental water in Switzerland. *Environ Pollut* 337:122476. <https://doi.org/10.1016/j.envpol.2023.122476>
25. Miltgen G, Martak D, Valot B, Kamus L, Garrigos T, Verchere G, Gbaguidi-Haore H, Ben Cimon C, Ramiandrisoa M, Picot S, Lignereux A, Masson G, Jaffar-Bandjee MC, Belmonte O, Cardinale E, Hocquet D, Mavingui P, Bertrand X. 2022. One Health compartmental analysis of ESBL-producing *Escherichia coli* on Reunion Island reveals partitioning between humans and livestock. *J Antimicrob Chemother* 77:1254–1262. <https://doi.org/10.1093/jac/dkac054>
26. Gwenzi W, Chaukura N, Muisa-Zikalani N, Teta C, Musvuugwa T, Rzymyski P, Abia ALK. 2021. Insects, rodents, and pets as reservoirs, vectors, and sentinels of antimicrobial resistance. *Antibiotics (Basel)* 10:68. <https://doi.org/10.3390/antibiotics10010068>
27. Hassell JM, Ward MJ, Muloi D, Bettridge JM, Robinson TP, Kariuki S, Ogendo A, Kiiru J, Imboma T, Kang'ethe EK, Öghren EM, Williams NJ, Begon M, Woolhouse MEJ, Fèvre EM. 2019. Clinically relevant antimicrobial resistance at the wildlife-livestock-human interface in Nairobi: an epidemiological study. *Lancet Planet Health* 3:e259–e269. [https://doi.org/10.1016/S2542-5196\(19\)30083-X](https://doi.org/10.1016/S2542-5196(19)30083-X)
28. Mamedova Y, Chaplygina A. 2024. Population size and nesting peculiarities of the black-headed gull *Chroicocephalus ridibundus* (Linnaeus, 1766) on the territory of water treatment facilities. *Studia Biologica* 18:201–218. <https://doi.org/10.30970/sbi.1802.770>
29. Woksepp H, Karlsson K, Börjesson S, Karlsson Lindsjö O, Söderlund R, Bonnedahl J. 2023. Dissemination of carbapenemase-producing Enterobacteriales through wastewater and gulls at a wastewater treatment plant in Sweden. *Sci Total Environ* 886:163997. <https://doi.org/10.1016/j.scitotenv.2023.163997>
30. Valenza G, Werner M, Eisenberger D, Nickel S, Lehner-Reindl V, Höller C, Bogdan C. 2019. First report of the new emerging global clone ST1193 among clinical isolates of extended-spectrum β-lactamase (ESBL)-producing *Escherichia coli* from Germany. *J Glob Antimicrob Resist* 17:305–308. <https://doi.org/10.1016/j.jgar.2019.01.014>
31. Nüesch-Inderbinen MT, Baschera M, Zurfluh K, Hächler H, Nüesch H, Stephan R. 2017. Clonal diversity, virulence potential and antimicrobial resistance of *Escherichia coli* causing community acquired urinary tract infection in Switzerland. *Front Microbiol* 8:2334. <https://doi.org/10.3389/fmicb.2017.02334>
32. Pitout JDD, DeVinney R. 2017. *Escherichia coli* ST131: a multidrug-resistant clone primed for global domination. *F1000Res* 6:F1000 Faculty Rev-195. <https://doi.org/10.12688/f1000research.10609.1>
33. Pitout JDD, Peirano G, Chen L, DeVinney R, Matsumura Y. 2022. *Escherichia coli* ST1193: following in the footsteps of *E. coli* ST131. *Antimicrob Agents Chemother* 66:e0051122. <https://doi.org/10.1128/aac.00511-22>
34. Yuan T, Pian Y. 2023. Hospital wastewater as hotspots for pathogenic microorganisms spread into aquatic environment: a review. *Front Environ Sci* 10:1091734. <https://doi.org/10.3389/fenvs.2022.1091734>
35. Watt AE, Cummins ML, Donato CM, Wirth W, Porter AF, Andersson P, Donner E, Jennison AV, Seemann T, Djordjevic SP, Howden BP, Australian Pathogen Genomics One Health Working Group. 2025. Parameters for one health genomic surveillance of *Escherichia coli* from Australia. *Nat Commun* 16:17. <https://doi.org/10.1038/s41467-024-5510-3-2>
36. Bush K, Bradford PA. 2020. Epidemiology of β-lactamase-producing pathogens. *Clin Microbiol Rev* 33:e00047-19. <https://doi.org/10.1128/CMR.00047-19>
37. ECDC/EMEA Joint Working Group. 2009. The bacterial challenge: time to react. European Centre for Disease Prevention and Control. EMEA doc. ECDC/EMEA Joint Technical Report. <https://doi.org/10.2900/2518>
38. Bevan ER, Jones AM, Hawkey PM. 2017. Global epidemiology of CTX-M β-lactamases: temporal and geographical shifts in genotype. *J Antimicrob Chemother* 72:2145–2155. <https://doi.org/10.1093/jac/dkx146>
39. Tiwari A, Kurittu P, Al-Mustapha AI, Heljanko V, Johansson V, Thakali O, Mishra SK, Lehto K-M, Lippinen A, Oikarinen S, Pitkänen T, WastPan Study Group, Heikinheimo A. 2022. Wastewater surveillance of antibiotic-resistant bacterial pathogens: a systematic review. *Front Microbiol* 13:977106. <https://doi.org/10.3389/fmicb.2022.977106>
40. Geser N, Stephan R, Korczak BM, Beutin L, Hächler H. 2012. Molecular identification of extended-spectrum-β-lactamase genes from Enterobacteriaceae isolated from healthy human carriers in Switzerland. *Antimicrob Agents Chemother* 56:1609–1612. <https://doi.org/10.1128/AAC.05539-11>
41. Geser N, Stephan R, Hächler H. 2012. Occurrence and characteristics of extended-spectrum β-lactamase (ESBL) producing Enterobacteriaceae in food producing animals, minced meat and raw milk. *BMC Vet Res* 8:21. <https://doi.org/10.1186/1746-6148-8-21>
42. Stephan R, Hächler H. 2012. Discovery of extended-spectrum beta-lactamase producing *Escherichia coli* among hunted deer, chamois and ibex. *Schweiz Arch Tierheilkd* 154:475–478. <https://doi.org/10.1024/0036-7281/a000390>
43. Homeier-Bachmann T, Schütz AK, Dreyer S, Glanz J, Schaufler K, Conraths FJ. 2022. Genomic analysis of ESBL-producing *E. coli* in wildlife from north-Eastern Germany. *Antibiotics (Basel)* 11:123. <https://doi.org/10.3390/antibiotics11020123>
44. Zurfluh K, Nüesch-Inderbinen M, Stephan R, Hächler H. 2013. Higher-generation cephalosporin-resistant *Escherichia coli* in feral birds in Switzerland. *Int J Antimicrob Agents* 41:296–297. <https://doi.org/10.1016/j.ijantimicag.2012.11.005>
45. Zurfluh Katrin, Albin S, Mattmann P, Kindle P, Nüesch-Inderbinen M, Stephan R, Vogler BR. 2019. Antimicrobial resistant and extended-spectrum β-lactamase producing *Escherichia coli* in common wild bird species in Switzerland. *Microbiologyopen* 8:e845. <https://doi.org/10.1002/mbo3.845>

46. Hill R, Stentiford GD, Walker DI, Baker-Austin C, Ward G, Maskrey BH, van Aerle R, Verner-Jeffreys D, Peeler E, Bass D. 2024. Realising a global One Health disease surveillance approach: insights from wastewater and beyond. *Nat Commun* 15:5324. <https://doi.org/10.1038/s41467-024-49417-4>
47. Pitout JDD, Peirano G, DeVinney R. 2023. The contributions of multidrug resistant clones to the success of pandemic extra-intestinal Pathogenic *Escherichia coli*. *Expert Rev Anti Infect Ther* 21:343–353. <https://doi.org/10.1080/14787210.2023.2184348>
48. Salleh MZ, Nik Zuraina NMN, Hajissa K, Ilias MI, Deris ZZ. 2022. Prevalence of multidrug-resistant diarrheagenic *Escherichia coli* in Asia: a systematic review and meta-analysis. *Antibiotics (Basel)* 11:1333. <https://doi.org/10.3390/antibiotics11101333>
49. Widodo A, Lamid M, Effendi MH, Khailrullah AR, Kurniawan SC, Silaen OSM, Riwu KHP, Yustinasari LR, Afrani DA, Dameanti F, Ramandinianto SC. 2023. Antimicrobial resistance characteristics of multidrug resistance and extended-spectrum beta-lactamase producing *Escherichia coli* from several dairy farms in Probolinggo, Indonesia. *Biodiversitas* 24:215–221. <https://doi.org/10.13057/biodiv/d240126>
50. World Health Organisation. 2015. Global Antimicrobial Resistance Surveillance System
51. European Centre for Disease Prevention and Control. 2014. Antimicrobial resistance surveillance in Europe annual report of the European Antimicrobial Resistance Surveillance Network (EARS-Net) 2013. Publications Office.
52. Neupane S, Pant ND, Khatiwada S, Chaudhary R, Banjara MR. 2016. Correlation between biofilm formation and resistance toward different commonly used antibiotics along with extended spectrum beta lactamase production in uropathogenic *Escherichia coli* isolated from the patients suspected of urinary tract infections visiting Shree Birendra Hospital, Chhauni, Kathmandu, Nepal. *Antimicrob Resist Infect Control* 5:1–5. <https://doi.org/10.1186/s13756-016-0104-9>
53. Surgers L, Boyd A, Girard PM, Arlet G, Decré D. 2019. Biofilm formation by ESBL-producing strains of *Escherichia coli* and *Klebsiella pneumoniae*. *Int J Med Microbiol* 309:13–18. <https://doi.org/10.1016/j.jimm.2018.10.008>
54. Carlsen L, Büttner H, Christner M, Cordts L, Franke G, Hoffmann A, Knobling B, Lütgehetmann M, Nakel J, Werner T, Knobloch JK. 2023. Long time persistence and evolution of carbapenemase-producing Enterobacterales in the wastewater of a tertiary care hospital in Germany. *J Infect Public Health* 16:1142–1148. <https://doi.org/10.1016/j.jiph.2023.05.029>
55. Köchle B, Bernier Gosselin V, Kaspar H, Becker J. 2024. Treated and untreated cows housed side by side in tie-stalls and their respective risk of harboring *E. coli* resistant to antimicrobials. *PLoS One* 19:e0310431. <https://doi.org/10.1371/journal.pone.0310431>
56. Höckstaller K, Marti IA, Bank C, Yilmaz B, Becker J. 2025. Antimicrobial resistance in wildlife across Switzerland and the principality of Liechtenstein: associations with environmental factors and taxonomic variation. *bioRxiv*. <https://doi.org/10.2139/ssrn.5202202>
57. Conforti S, Kapopoulou A, Bank C, Köchle B, Yilmaz B, Becker J. 2025. Extended-spectrum beta-lactamase (ESBL)-producing *E. coli* in livestock and free-roaming wildlife: a combined phenotyping – whole-genome sequencing One Health approach. *bioRxiv*. <https://doi.org/10.1101/2025.04.08.647752>
58. Perez-Sepulveda BM, Heavens D, Pulford CV, Predeus AV, Low R, Webster H, Dykes GF, Schudoma C, Rowe W, Lipscombe J, Watkins C, Kumwenda B, Shearer N, Costigan K, Baker KS, Feasey NA, Hinton JCD, Hall N, 10KSG consortium. 2021. An accessible, efficient and global approach for the large-scale sequencing of bacterial genomes. *Genome Biol* 22:349. <https://doi.org/10.1186/s13059-021-02536-3>
59. MacConaill LE, Burns RT, Nag A, Coleman HA, Slevin MK, Giorda K, Light M, Lai K, Jarosz M, McNeill MS, Ducar MD, Meyerson M, Thorner AR. 2018. Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genomics* 19:30. <https://doi.org/10.1186/s12864-017-4428-5>
60. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
61. Andrews S. 2010. Online. FastQC: a quality control tool for high throughput sequence data. Available from: <http://www.bioinformatics.braham.ac.uk/projects/fastqc>
62. Wingett SW, Andrews S. 2018. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res* 7:1338. <https://doi.org/10.12688/f1000research.15931.2>
63. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>
64. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
65. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. 2021. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb Genom* 7:685. <https://doi.org/10.1099/mgen.0.000685>
66. Waters NR, Abram F, Brennan F, Holmes A, Pritchard L. 2020. Easy phylotyping of *Escherichia coli* via the EzClermont web app and command-line tool. *Access Microbiol* 2:acmi000143. <https://doi.org/10.1099/acmi.0.000143>
67. Seemann T. 2018. ABRicate: mass screening of contigs for antimicrobial and virulence genes. Department of Microbiology and Immunology, The University of Melbourne, Melbourne, Australia.
68. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S, et al. 2020. CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 48:D517–D525. <https://doi.org/10.1093/nar/gkz935>
69. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31:3691–3693. <https://doi.org/10.1093/bioinformatics/btv421>
70. Didelot X, Wilson DJ. 2015. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput Biol* 11:e1004041. <https://doi.org/10.1371/journal.pcbi.1004041>
71. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>
72. Letunic I, Bork P. 2021. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 49:W293–W296. <https://doi.org/10.1093/nar/gkab301>
73. Joy JB, Liang RH, McCloskey RM, Nguyen T, Poon AFY. 2016. Ancestral reconstruction. *PLoS Comput Biol* 12:e1004763. <https://doi.org/10.1371/journal.pcbi.1004763>
74. Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290. <https://doi.org/10.1093/bioinformatics/btg412>