Communications Medicine

https://doi.org/10.1038/s43856-026-01402-2

Article in Press

Genomic characterization and sub-clustering of *Escherichia coli* clonal complex 38 reveal host associated genetic markers

Received: 18 August 2025

Accepted: 16 January 2026

Cite this article as: Roer, L., Rasmussen, A., Hansen, F. *et al.* Genomic characterization and sub-clustering of *Escherichia coli* clonal complex 38 reveal host associated genetic markers. *Commun Med* (2026). https://doi.org/10.1038/s43856-026-01402-2

Louise Roer, Astrid Rasmussen, Frank Hansen, Lars E. B. Christoffersen, Raphael Sieber, Flemming Scheutz, Rene S. Hendriksen, Brian D. Johnston, James R. Johnson, Barbara J. Holzknecht, Lillian Søes, Kristian Schønning, Dennis B. Holmgaard, Ulrik S. Justesen, Claus Østergaard, Turid S. Søndergaard, Marc T. K. Nielsen, Mikala Wang, Hans L. Nielsen, Sam Abraham, Daniel E. Park, Maliha Aziz, Lance B. Price, Anette M. Hammerum, Henrik Hasman & Marc Stegger

Title: Genomic Characterization and Sub-Clustering of *Escherichia coli* Clonal Complex 38 reveal Host Associated Genetic Markers

Louise Roer[a,u]*, Astrid Rasmussen[a,u], Frank Hansen[a], Lars E. B. Christoffersen[b], Raphael Sieber[b], Flemming Scheutz[a], Rene S. Hendriksen[c], Brian D. Johnston[d,e], James R. Johnson[d,e], Barbara J. Holzknecht[a,f,g], Lillian Søes[h], Kristian Schønning[i,j], Dennis B. Holmgaard[k], Ulrik S. Justesen[l], Claus Østergaard[m], Turid S. Søndergaard[n], Marc T.K. Nielsen[o], Mikala Wang[p], Hans L. Nielsen[q,r], Sam Abraham[s,u], Daniel E. Park[t], Maliha Aziz[t], Lance B. Price[b,t,u], Anette M. Hammerum[a,u], Henrik Hasman[a,u], Marc Stegger[b,s,u]

a Department of Bacteria, Parasites and Fungi, Statens Serum Institut, Copenhagen, Denmark

b Department of Sequencing and Bioinformatics, Statens Serum Institut, Copenhagen, Denmark

c Reseach group Global Capacity building, National Food Institute, Technical University of Denmark, Kgs. Lyngby, Denmark

d Minneapolis Veterans Affairs Health Care System, Minneapolis, MN, USA

e Department of Medicine, University of Minnesota, Minneapolis, MN, USA

f Department of Clinical Microbiology, Copenhagen University Hospital - Herlev and Gentofte, Herlev, Denmark

g Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

h Department of Clinical Microbiology, Copenhagen University Hospital - Amager and Hvidovre, Hvidovre, Denmark

i Department of Clinical Microbiology, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark

j Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

k Department of Clinical Microbiology, Slagelse Hospital, Slagelse, Denmark

l Department of Clinical Microbiology, Odense University Hospital, Odense, Denmark

m Department of Clinical Microbiology, Lillebaelt Hospital, Vejle, Denmark

n Department of Clinical Microbiology, Sønderjylland Hospital, Aabenraa, Denmark

o Department of Clinical Diagnostics, Esbjerg and Grindsted Hospital, Denmark

p Department of Clinical Microbiology, Aarhus University Hospital, Aarhus, Denmark

q Department of Clinical Microbiology, Aalborg University Hospital, Aalborg, Denmark

r Depart of Clinical Medicine, Aalborg University, Aalborg, Denmark

s Antimicrobial Resistance and Infectious Diseases Laboratory, Harry Butler Institute, Murdoch University, Perth, Australia

t Milken Institute School of Public Health, George Washington University, Washington, DC, USA

u European Society of Clinical Microbiology and Infectious Diseases (ESCMID) Study Group for Mobile Elements and Plasmids (ESGMAP), Basel, Switzerland

Corresponding author: Louise Roer, loro@ssi.dk

Abstract

Background. *Escherichia coli* clonal complex 38 (CC38) is a genetically diverse lineage increasingly linked to antimicrobial resistance and extraintestinal infections in humans. Despite its clinical and epidemiological relevance, its population structure, zoonotic potential, and ecological associations remain poorly understood.

Methods. We analyzed 242 human *E. coli* CC38 bloodstream isolates collected through Danish national surveillance, 83 isolates from food and production animals, and 2,313 international genomes to investigate host associations and transmission dynamics. Phylogenetic reconstruction, Bayesian host prediction based on mobile genetic elements, and statistical testing of plasmid–host associations were used to delineate population structure and identify potential host-associated markers.

Results. Here we show that Danish CC38 isolates belong to multiple sub-lineages, with no evidence of foodborne outbreaks and limited hospital transmission. Bayesian host prediction supports a poultry origin for several distinct human sub-lineages. Global analyses of 2,638 genomes reveal two major clusters: a poultry-associated Cluster I and a predominantly human-associated Cluster II, which subdivides into eight sub-lineages with distinct host, resistance, and virulence profiles. Two small plasmids, ColRNAI and Col(MG828), are strongly enriched in poultry and livestock isolates but largely absent from human-associated sub-clusters, indicating their value as host-associated genetic markers.

Conclusions. Our findings refine the phylogenetic structure of *E. coli* CC38 and identify plasmid markers that may enhance genomic surveillance of zoonotic transmission. These results highlight the importance of a One Health approach to monitor antimicrobial resistance across human, food, and animal reservoirs. Together, these insights support data-driven One Health surveillance and intervention strategies.

Plain language summary

Antibiotic-resistant bacteria are a growing concern for both human and animal health. This study examined a specific group of *E. coli* bacteria, which can cause serious infections. By comparing the genomes of bacteria, isolated from people, food, and animals in Denmark and around the world, the researchers mapped how different lineages are related and where they are most likely to come from. The results show that some strains found in humans may have originated in poultry. The study also identified two small DNA elements strongly linked to bacteria from poultry and livestock. These elements could help trace how resistant bacteria move between animals and humans, supporting more effective disease surveillance and antibiotic resistance prevention.

Introduction

Extraintestinal pathogenic *Escherichia coli* (ExPEC) are a major cause of urinary tract infections (UTIs) and bloodstream infections (BSIs) worldwide, contributing significantly to antimicrobial resistance (AMR)-related mortality (1). Among ExPEC, specific clonal lineages are recognized as high-risk global clones due to their ability to cause severe infections and spread multidrug resistance determinants. One such lineage is clonal complex (CC) 38, which predominantly consists of sequence type (ST) 38 and closely related sequence types (e.g., ST963 and ST3666). ST38 has been increasingly implicated in invasive infections (2,3). Despite its clinical relevance, ST38 has been underexplored compared to other dominant ExPEC lineages such as ST10, ST69, ST131, and ST167.

E. coli is a highly diverse species, classified into eight phylogroups (A, B1, B2, C, D, E, F, and G). While phylogroup B2 lineages, such as ST131, have dominated attention due to their adaption to human hosts (4) and high levels of antimicrobial resistance (5), phylogroup D, including CC38, has traditionally been overlooked despite emerging clinical significance.

Phylogroup D strains tend to harbor a diverse set of virulence factors, facilitating extraintestinal infections, and emerging clones like ST38 are showing increasing levels of multidrug resistance (6,7).

Several studies have identified *E. coli* ST38 as a uropathogenic-enteroaggregative (UPEC-EAEC) hybrid clone. Functionally, this lineage combines traits such as biofilm that promote intestinal persistence (from EAEC) with virulence factors enabling extraintestinal infections (from UPEC). This dual nature may enhance its capacity to colonize multiple host environments and cause both intestinal and extraintestinal infections (2,8,9). Third-generation cephalosporin- and carbapenem-resistant (3GC/C-R) *E. coli* ST38 has been implicated in both hospital- and community-associated outbreaks (10,11). Additionally, it has been recognized as a cross-border disseminated clone (12). In Denmark, national surveillance data (www.danmap.org) indicate that 3GC/C-R ST38 is the second most prevalent ESBL-producing *E. coli* lineage causing BSIs, surpassed only by ST131 (13).

Growing evidence suggests that ST38 thrives in multiple reservoirs beyond human clinical settings. Recent studies have identified ST38 in human patients (6,10,11,14), wild birds, food-producing animals, food (2,6,15,16) and environmental sources (2,17,18). Additionally, ST38 isolates have been shown to carry chromosomal mutations in quinolone resistance-determining regions (QRDRs), contributing to fluoroquinolone resistance in both clinical and animal-associated settings (19,20). The combination of antimicrobial resistance and ecological versatility raises concerns about potential zoonotic transmission and environmental persistence, aligning with ST38's global dissemination and adaptation to different ecological niches (3,12,17). While Fonseca *et al.* demonstrated ST38 as a globally distributed *E. coli* lineage (17), Mo *et al.* were among the first to clearly distinguish two major subgroups within ST38, one predominantly associated with humans and the other with broiler chickens (6). However, it remains unclear whether ST38 *E. coli* represents a broad or narrow host range sequence type,

with human-pathogenic ST38 populations primarily originating from animal and environmental reservoirs or representing distinct, human-adapted lineages.

To address these knowledge gaps, we performed a comprehensive characterization of *E. coli* isolates belonging to the broader CC38 from diverse sources in Denmark. By integrating genomic analysis with epidemiological data, we delineate the population structure of CC38 and identify its major sub-lineages. We show that human and poultry isolates form distinct clusters, with limited evidence of hospital transmission. Using a global collection of genomes, we expand current knowledge on the resistome and phylogenetic relationships of *E. coli* CC38 and identify potential genomic markers indicative of host association patterns and zoonotic transmission. Together, these findings provide critical insights into the emergence, dissemination, and persistence of *E. coli* CC38 within the global One Health framework, supporting surveillance and control strategies.

Methods

Genome collection

As part of the Danish surveillance program DANMAP (www.danmap.org), resistant bacteria from food animals, food and human patients are collected, analyzed and monitored. Regarding human isolates, the Danish Departments of Clinical Microbiology submitted expected ESBL/pAmpC-producing *E. coli* blood isolates as well as carbapenemase-producing organisms (CPO) from all human samples for verification and genome sequencing at the national reference laboratory at Statens Serum Institut. For CPO cases, travel history was routinely collected as part of the submission process. Only the first positive isolate, within a rolling 12-month period, was included for each patient, except if the second isolated showed a different ESBL, pAmpC or CPO resistance gene compared to the first isolate. The genome collection was screened for *E. coli* isolates obtained between 2014 through 2022, belonging to CC38.

Bacteria from food animals and food were routinely collected by the Danish Veterinary and Food Administration (DVFA) and further analyzed in collaboration with the Technical University of Denmark (DTU) as part of the DANMAP surveillance program. Isolates collected from 2012 through 2021 were screened for isolates belonging to CC38.

In addition to the genomes collected as part of the Danish surveillance program, sequencing data and associated metadata were included from isolates collected between 2009 and 2019 in Australia (wildlife and companion animals) and the USA (human, poultry, and companion animal). Finally, EnteroBase (https://enterobase.warwick.ac.uk/) was searched for *E. coli* CC38 genomes (accessed 16 March 2023), where duplicates from the above-mentioned raw reads collection were removed. A total of 18 samples with source metadata failing into categories represented by fewer than ten isolates were excluded from the analysis (aquatic animal (n = 6), laboratory (n = 6), plant (n = 4), and unknown (n = 2)).

In this study, livestock was defined as non-poultry food-producing animals, including goats and horses, to distinguish them analytically from poultry.

In total, 2,638 genomes were included in this study, comprising 242 from Danish human cases, 83 from Danish food and veterinary sources, and 2,313 publicly available genomes from international collections (EnteroBase and other datasets). The complete list of draft genomes is provided in Supplementary Data 1, together with relevant metadata and accession numbers.

Whole-genome sequencing and quality control

DNA extraction and quantification

For whole genome sequencing (WGS), DNA was extracted using either DNeasy Blood & Tissue Kit (Qiagen, Copenhagen, Denmark) or MagNA Pure 96 DNA Multi-Sample Kit (Roche Life Science, Penzberg, Germany). Extracted DNA was quantified using the Quant-iT dsDNA BR

Assay Kit (Thermo Fisher Scientific, Carlsbad, CA, USA) and fluorescence was measured on a FLUOstar Omega (BMG LabTech, Ortenberg, Germany).

Library preparation and sequencing

Extracted DNA was diluted to 0.2 ng/µL, and 2.5 µL was used as input for library preparation. All WGS libraries were quantified and pooled according to genome size only if the post-library concentration was above 0.5 ng/µL.

Sequencing libraries for short-read sequencing were prepared with Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA). Sequencing was performed on Illumina MiSeq or NextSeq 500 platforms, generating paired-end reads of either 2 x 150bp or 2 x 250bp, depending on the year of sequencing. The use of multiple platforms reflects routine technological updates at the sequencing facility.

Bioinformatics analysis and quality control

Raw sequencing data were analyzed using the in-house Bifrost pipeline v. 2.0.8 (https://github.com/ssi-dk/bifrost), which performs automated read trimming, quality control, assembly, contamination screening, and species verification prior to downstream analyses. An average coverage of ≥ 30x was applied. *De novo* assembly of WGS data was performed using SKESA v. 2.2 (21), as implemented in the Bifrost pipeline.

The quality of all draft genomes was assessed, and only genomes with a total assembly size > 4 Mb and < 800 contigs were retained for downstream analysis. These thresholds were chosen to ensure sufficient assembly quality while retaining representative diversity across datasets, as stricter cutoffs would have excluded a substantial number of valid genomes.

Genotypic characterization

The applied identity and coverage thresholds for each analysis followed published recommendations and tool-specific defaults to ensure reproducible and biologically reliable gene detection.

Typing and classification

To verify that the genomes belonged to CC38, MLST was identified by using the Wirth scheme for *E. coli* (22) with ABRicate v. 1.0.1 (Seemann T, ABRIcate, Github https://github.com/tseemann/abricate).

fimH subtypes were identified by using KMA v. 1.4.9 (23) by mapping draft genomes against the *fimH* database (https://bitbucket.org/genomicepidemiology/fimtyper_db/src/master/), applying thresholds of 100% identity and 100% coverage.

Phylotyping was performed on each strain using EzClermont v. 0.7.0 (24).

Pathotype classification into ExPEC, UPEC, ETEC and EAEC was performed based on virulence gene profiles, as described by Tetzschner *et al.* (25), using KMA v. 1.4.9 (23) to map the draft genomes against the VirulenceFinder database (https://bitbucket.org/genomicepidemiology/virulencefinder_db/src/master/), with thresholds of 90% identity and 95% coverage.

Antimicrobial resistance genes

Resistance genes and mutations conferring antimicrobial resistance were identified with abriTAMR v. 1.0.14 (26), using a threshold of 98% identity (ID) and a minimum coverage of 80%.

Plasmid replicons and ColV screening

Plasmid replicons were identified using PlasmidFinder database for Enterobacteriales (plasmidfinder_db-enterobacteriales:

https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/) (27), with a threshold of 95% identity and a minimum coverage of 80%, except for Col-plasmids where an 80% identity threshold was applied, according to the recommendation by Carattoli *et al.* Plasmid replicon prediction was performed using ABRicate v. 1.0.1 (Seemann T, ABRIcate, Github https://github.com/tseemann/abricate).

Screening for ColV-positive genomes was conducted by the methodology described by Liu *et al.* (28). Briefly, a strain was classified as ColV-positive if it carried at least one gene from four or more of the following six gene sets: (i) *cvaABC* and *cvi* (ColV operon), (ii) *iroBCDEN* (salmochelin operon), (iii) *iucABCD* and *iutA* (aerobactin operon), (iv) *etsABC*, (v) *ompT* and *hlyF*, and (vi) *sitABCD*.

Genomic epidemiology and Phylogenetic Analysis

Identification of SNPs

SNPs were identified using the Northern Arizona SNP Pipeline v. 1.2 (NASP) (29). Briefly, duplicate regions of the reference chromosome of *E. coli* strain 190693 (GenBank accession: CP048934.1) were first identified using NUCmer (30), followed by mapping the draft genomes against the reference using the Burrows-Wheeler Aligner (BWA) (31). GATK was used for SNP calling (32) within NASP, applying minimum thresholds of 10× coverage and a base-call proportion of 0.9; positions not meeting these criteria were excluded. Recombinant regions were detected and masked with Gubbins v. 2.2 (33), using default parameters that also exclude taxa with more than 25% gaps in the alignment.

Phylogenetic analysis

Phylogenetic analyses were performed by maximum-likelihood approximation on purged alignments using the generalized time-reversible model in FastTree, v. 2.1.8 (34), with local

support assessed by 1,000 resamples. Phylogenies were used for visualization of population structure, annotated with relevant metadata and visualized using iTOL (35).

Identification of possible outbreaks and patient links in the Danish collection

Pairwise SNP differences between isolates were assessed to predict possible clones and outbreaks. Possible clones (PCs) were identified by applying an predefined SNP distance criterion of ≤ 40 SNPs to the nearest neighboring isolate (PC40), whereas possible outbreaks (POs) were identified by applying a SNP distance criterion of ≤ 10 SNPs (PO10), as previously described (36,37). For investigation of possible outbreaks from nosocomial transmission between patients, hospitalization data were retrieved from the Danish National Patient Registry for all patients included in a PO10 cluster. The data included dates of admission and discharge on patient level, hospital and department codes (38). Epidemiological links between patients were defined as; direct link when an overlap where any two patients were admitted to the same hospital department on the same date; hospital link when an overlap where any two patients were admitted to the same hospital on the same date; and indirect link when an overlap of any two patient were admitted to the same hospital department, but with a time intervening period of maximum 14 days between them. In addition, potential departmental accumulation was investigated by identifying departments with repeated occurrences of epidemiologically linked cases over longer periods, to account for the possibility of persistent environmental contamination or undetected transmission chains within a specific hospital unit.

Prediction of most likely source

Host-associated mobile genetic elements

The Danish collection was screened for host-associated mobile genetic elements (MGEs), with a focus on model-defined host specificity. Samples annotated as “Unknown food” and the single livestock isolate were excluded from the analysis.

Samples were annotated with the host-associated MGEs described by Liu *et al.* (39), using K-mer alignment (23) with default settings to scan the draft genomes for genes that comprise predefined host MGEs. A gene was called as present if the hit had at least 90% coverage and 90% identity. A host element was called as present if at least one gene from the host element was present.

Bayesian latent class model

The Bayesian latent class model was run as previously described (39), with a modification to exclude host MGE marker 45 from the MGE panel. This marker was removed based on model fit and sensitivity analyses indicating minor improvements in determinacy of classifications following exclusion (1% reduction in indeterminates) without changes in posterior class assignments. Clades and host elements were used as input for the model, from which probabilities of belonging to one of ten clade-host combinations were assigned. The results from the model were summarized to estimate if a sample was of human or meat origin. Response probabilities were generated using independent logistic normal priors with mean 0 and standard deviations of 4/9 while class probabilities use a Beta (1,1) prior.

Statistical Analysis

Associations between plasmid replicons and source categories were assessed using multinomial logistic regression with poultry as the reference group. Predictor variables included ColRNAI, Col(MG828), and their interaction. Odds ratios and 95% confidence intervals were estimated using maximum likelihood in R v. 4.2.2 (40) with the nnet package (v. 7.3-19). A two-sided p-value < 0.05 was considered statistically significant.

Human Ethics

This study was conducted as part of the statutory tasks of Statens Serum Institut under the Danish Health Care Act (§222). According to national legislation (implementing decree no.

1338, 1 September 2020), ethics approval and informed consent are not required. Only aggregated, non-identifiable data were analyzed, and the study does not involve personal data covered by the European GDPR.

Results

Prevalence and Genetic Diversity of *E. coli* CC38 from Danish human cases

Among 242 Danish CC38 isolates from 240 patients, 219 belonged to ST38 (91%), followed by ST3666 (n = 5, 2%). All isolates were assigned to phylogroup D and encompassed 15 *fimH*-types, with *fimH5* (n = 136, 56%), *fimH0* (n = 45, 19%), *fimH65* (n = 21, 9%), *fimH54* (n = 16, 7%) and *fimH116* (n = 11, 5%) being most common.

Sixteen distinct ESBL-, pAmpC-, or carbapenemase-encoding genes were identified among the 242 isolates, based on genomic detection (Supplementary Data 2). The most frequent were $bla_{\text{CTX-M-14}}$ (n = 101, 42%) and $bla_{\text{CTX-M-15}}$ (n = 62, 26%), followed by $bla_{\text{OXA-244}}$ (n = 45, 19%), and $bla_{\text{OXA-48}}$ (n = 34, 14%) (Supplementary Data 2). Most genomes carried resistance genes belonging to a single β-lactamase class (n = 165, 68%): ESBL (n = 142, 59%), pAmpC (n = 7, 3%), and carbapenemase (n = 16, 7%). The remaining 77 genomes (32%) encoded combinations of two classes, most frequently ESBL and carbapenemase (n = 58, 24%), followed by ESBL and pAmpC (n = 11, 5%), and pAmpC and carbapenemase (n = 8, 3%). No genomes carried determinants from all three categories.

On February 18th, 2020, the European Centre for Disease Prevention and Control (ECDC) published a Rapid Risk Assessment on the increase in *E. coli* producing the carbapenemase OXA-244, which was updated on July 20th, 2021 (41). Between January 2016 and August 2019, 23 isolates of OXA-244-producing *E. coli* were identified in Denmark, including 13 ST38 and one ST3268 (11), all included in the Rapid Risk Assessment. An additional 31 OXA-244-producing *E. coli* CC38 isolates were detected between September 2019 and December 2022.

In the current study, 35 of the OXA-244-producing CC38 *E. coli* co-carried other ESBL-encoded gene(s) and six carried the pAmpC gene *bla*$_{CMY-2}$.

Clonal Relatedness and Outbreaks of *E. coli* CC38 from Danish human cases

To investigate clonal relatedness, potential outbreaks, and the sources of CC38 3GC/C-R *E. coli* in Denmark, we conducted a detailed analysis of the collection. Among the 242 human genomes, 48,050 SNPs were identified within ~69% of the reference chromosome of *E. coli* strain 190693, with 5,159 phylogenetically informative SNPs (piSNPs) after filtering recombinant regions. The analysis identified 14 possible clonal clusters and 10 potential outbreak clusters (see Supplementary Data 3). The carbapenemase *bla*$_{OXA-244}$ gene was detected in two PO10s, while *bla*$_{OXA-48}$ appeared in five PO10s.

Only one case showed a possible direct epidemiological link involving three patients (PC40-4, PO10-6). PO10-6 included five isolates (0-11 SNPs) collected between 2014 and 2019. These isolates belonged to ST38-*fimH5* and all carried *bla*$_{CTX-M-24}$ and *bla*$_{OXA-48}$. Three patients had overlapping hospitalizations and could be directly linked epidemiologically. Their isolates differed by ≤ 1 SNPs. The remaining two patients could not be linked through hospitalization data.

Additionally, a potential case of departmental accumulation was observed in PO10-2, where five patients could be linked to the same hospital units over time, although no formal epidemiological links could be established.

Among the seven PO10s involving carbapenemase-producing isolates, travel information was reported in three PO10s. For the remaining clusters, no direct epidemiological links could be established.

Zoonotic Transmission: The Foodborne Link

To further assess the route of infection and dissemination of CC38 in Denmark, the Danish human collection (n = 242) was compared to the veterinary collection (n = 83) from Danish Veterinary and Food Administration (DVFA), to identify possible animal sources. A total of 51,459 SNPs were identified within ~68% of reference chromosome of *E. coli* strain 190693. After pruning of recombinant regions, 6,751 piSNPs remained. Applying a threshold of ≤ 40 SNP differences, no direct genomic links were observed between the two collections. However, two distinct phylogenetic groups were represented: DK-Group I, comprising predominantly human isolates; and DK-Group II, comprising isolates from poultry, livestock, and other food sources (Figure 1A).

To estimate the zoonotic contribution within the collection, source-associated mobile genetic elements were predicted from isolates of either human or poultry origin. The livestock sample was excluded due to the limited number (n = 1), and unknown food sources were not supported by the model.

The model predicted that isolates in DK-Group II were of poultry origin (Figure 1B), consistent with their actual isolation source. Conversely, DK-Group I was predominately predicted as human-associated, with the majority of PO10s found within this group. Notably, three groups within Group I were predicted to have non-human origins: Group Ia and Group Ib showed predictions of mixed sources. Group Ic was mainly predicted as poultry-derived (Figure 1B).

Global Phylogenetic Reconstruction and Cluster Analysis of *E. coli* CC38

A total of 2,638 genomes met the criteria for inclusion in the characterization. The isolates were collected between 1979 and 2022 (Figure 2), from human (n = 1,682, 64%), poultry (n = 462, 18%), wildlife (n = 149, 6%), water/river (n = 107, 4%), companion animal (n = 101, 4%), unknown food (n = 63, 2%), livestock (non-poultry food-producing animals including goats and horses, n = 52, 2%), and environment (n = 22, <1%). Before 2011, only few isolates were

collected (n = 152, 6%), whereas the following years the number increased as illustrated in Figure 2. Most isolates were collected in Europe (n = 950, 36%), followed by North America (n = 667, 25%), Oceania (n = 467, 18%), Asia (n = 399, 13%), South America (n = 94, 4%), and Africa (n = 61, 2%).

Genetic and clonal relationships within the global collection were assessed, providing insights into evolutionary and epidemiological patterns. A total of 97,145 SNPs were identified within ~45% the reference chromosome as previously described, with 28,291 piSNPs after filtering of recombinant regions. Assessing the phylogenetic reconstruction, two distinct clusters were observed: Cluster I and Cluster II (Figure 3). Key results across clusters were summarized in Table 1, with complete aggregated data in for all 2,638 isolates available in Supplementary Data 4.

The poultry-associated Cluster I was characterized by moderate antimicrobial resistance, primarily mediated by ESBL genes, combined with a remarkably high proportion of ExPEC-UPEC hybrid strains. Cluster I uniquely harbored the *fimH*270 subtype. It exhibited a distinct plasmid profile, notably with ColV markers and enriched with ColRNAI and Col(MG828) replicons.

In contrast to Cluster I, Cluster II encompasses a diverse set of subgroups (Clusters II-A to II-H), predominantly associated with human clinical settings, but also reflecting a broader host range and evolutionary dynamics. The Danish isolates were distributed across several of these subgroups: DK-Group Ib corresponded to Cluster II-A, DK-Group Ia to Cluster II-B, DK-Group Ic was primarily found in Cluster II-E but also occurred in II-F, II-H, and II-C, while DK-Group II aligned with Cluster II-D.

Clusters II-A and II-B were predominantly human-derived, displaying high numbers of ESBL- and carbapenemase genes. These clusters were strongly associated with UPEC and ExPEC

virulence profiles, and the combination of UPEC-ExPEC-EAEC were found in 13% of the genomes of Cluster II-B. The clusters were primarily associated with *fimH*5, and a proportion of isolates in Cluster II-B lacked the *fimH* gene entirely (*fimH*0).

Cluster II-D carried a high number of pAmpC resistance genes, primarily bla_{CMY-2}. QRDR mutations were absent from Clusters II-D, II-E, and II-G. Cluster II-D showed high prevalence of ExPEC classification and ColV, as well as enrichment of ColRNAI and Col(MG828) replicons. The predominant *fimH* subtype in this cluster was *fimH65*.

Cluster II-E was predominantly human-associated, but the Danish poultry-associated isolates (DK-Group Ic) were also present within this cluster. It exhibited a high level of resistance genes but a low level of virulence. The *fimH* subtype *fimH*54 was found in 81% of the isolates in this cluster.

Cluster II-F and II-C carried low levels of antimicrobial resistance and virulence genes. Cluster II-C displayed mixed human-poultry origin, occasional presence of pAmpC genes, and the *fimH65* as the predominant allele. More than half of the isolates in Cluster II-C harbored QRDR mutations. Cluster II-F was associated with *fimH*65 and *fimH*27. Both clusters showed levels of ColV, ColRNAI and Col(MG828).

Cluster II-G included isolates from human, wildlife and companion animal, and was enriched in pAmpC genes. The Col156 replicon was particularly prevalent in Cluster II-G.

Cluster II-H was human associated and exhibited high levels of antimicrobial resistance determinants, including the highest proportion of QRDR mutations among all sublineages (82%), while displaying low virulence.

Host Association Markers in *E. coli* CC38

Cluster-level analysis indicated that ColRNAI and Col(MG828) were highly enriched in poultry-associated sublineages, particularly in Cluster I and Cluster II-D. In contrast, ColRNAI was nearly absent from human-associated clusters such as Cluster II-H. Although Cluster II-H was 89% human-associated, Col(MG828) was still detected in 23% of isolates, possibly reflecting historical acquisition from livestock reservoirs.

Multinomial logistic regression (Table 2) confirmed that both plasmid replicons were strongly associated with poultry origin. Isolates from human, wildlife, water, environmental, companion animal, and livestock sources had significantly lower odds of carrying ColRNAI (ORs 0.04–0.14, all $p < 0.001$) or Col(MG828) (ORs 0.02–0.09, all $p < 0.001$) compared to poultry. Notably, the co-carriage (interaction) between the two plasmids was significant in livestock isolates (OR = 40.95, $p = 0.004$), despite both replicons being individually underrepresented in this group. This reflects a specific rather than frequent signature, where nearly all plasmid-positive livestock isolates carried both replicons. A similar interaction effect was observed in food isolates of unknown origin (OR = 15.20, $p = 0.014$). However, neither ColRNAI ($p = 0.35$) nor Col(MG828) ($p = 0.059$) alone were significantly associated with this category. All 52 co-carrier strains carrying both ColRNAI and Col(MG828) from unknown food belonged exclusively to Cluster II-D, a sub-lineage also enriched in poultry and livestock isolates.

Group-level distribution further reinforced these patterns. Co-carriers from livestock were found in multiple groups, but most frequently in Cluster II-D and II-F, which were previously identified as plasmid-enriched lineages. Poultry co-carriers, although more common overall, were distributed more broadly across groups. These phylogenetic associations suggest that co-carriage of ColRNAI and Col(MG828) is not random but instead reflects distinct host association patterns within specific CC38 sub-lineages.

Discussion

In this study, we characterize the prevalence, genetic diversity, and host associations of *E. coli* CC38 in Denmark and globally. CC38 was the second most common ESBL-producing lineage in Danish BSIs, with limited hospital transmission but potential community and zoonotic links. Global phylogenetic analysis revealed distinct human- and poultry-associated clusters, shaped by resistance profiles and plasmid content. Notably, ColRNAI and Col(MG828) were identified as potential markers of livestock association. These findings underscore CC38's role in antimicrobial resistance dissemination across human and animal reservoirs.

During the survey years, ST38 was the second most prevalent lineage of ESBL-producers identified in *E. coli* from human BSI in Denmark (7% of 2,060 isolates), following the dominating ST131 high-risk clone (55%). Notably, $bla_{\text{CTX-M-14}}$ was predominating in CC38, in contrast to $bla_{\text{CTX-M-15}}$ in the globally dominant ST131 lineage (36). Previous studies identified travel and community sources as possible drivers of OXA-244 spread (11,41), while a Norwegian study identified an intraregional hospital outbreak with OXA-244-producing ST38 *E. coli* in 2020 (10).

At the clonal level, hospitalization data retrieved from the Danish National Personal Registry, revealed only a few potential links between patients within PC40. However, the recorded links occurred either years before or after the positive sample dates, making direct transmission unlikely. The genomic and epidemiological data suggest that PO10-6 involved nosocomial transmission.

Overall, the data indicate a limited level of nosocomial transmission of CC38 in Denmark. However, most of the isolates originated from BSI, representing only the most severe cases. Travel Information is routinely collected for reporting of CPOs, and among the seven PO10s involving carbapenemase-producing isolates, travel was reported in three, highlighting the putative role of international and community-associated settings in the prevalence of CC38 in Denmark. These findings underscore the complex transmission dynamics of CC38. Although

nosocomial transmission appeared rare, the detection of clonal clusters and epidemiological links in specific cases emphasizes the need for continued genomic surveillance.

To explore possible zoonotic transmission routes in Denmark, we compared the 242 human isolates with 83 veterinary isolates obtained through food surveillance. Phylogenetic analysis of these isolates revealed two distinct groups: DK-Group I, comprising predominantly human isolates, and DK-Group II, consisting mainly of isolates from poultry, livestock, and other food sources (Figure 1A). This division aligns with previous findings by Mo *et al.* (6).

A large proportion of the isolates in DK-Group II originated from food sources. However, based on the model predictions, it is likely that these food isolates, and the entire DK-Group II, are linked to poultry. Overall, these results suggested zoonotic contribution to the CC38 linage in Denmark, consistent with overlapping genetic signatures between human and poultry isolates, but without direct epidemiological confirmation. However, interpretation can be flawed due to the sparse and potentially biased sampling of non-human reservoirs, as the routine surveillance system is primarily human-focused and limited to resistant *E. coli*, often excluding susceptible strains. Additionally, some zoonotic isolates infecting Danish patients may originate from international food sources, which could explain the lack of phylogenetic links to domestic poultry isolates.

To further understand the emergence of CC38 in Denmark, the collection was investigated in a global context. This distribution, in both source origin and geographic location, may not reflect the true prevalence of CC38. The dataset is inherently biased due to disparities in surveillance and research infrastructure and a predominance of human-derived samples. Acknowledging these limitations is crucial for contextualizing the findings and should be kept in mind when assessing the results.

In the global phylogeny, the CC38 population separated into two main lineages, Cluster I and Cluster II. The analysis revealed substantial genetic and phenotypic diversity among *E. coli* CC38 clusters, clearly reflecting distinct patterns of host association. The distinct plasmid composition observed in the poultry-associated Cluster I suggests that these plasmids may represent poultry-associated genetic elements linked to CC38. Both replicons, ColRNAI and Col(MG828), have been previously identified in poultry-associated *Salmonella* (42–44), while ColRNAI has been shown to facilitate transfer of bla_{KPC-2} resistance genes across species (45). This suggests a potential role in the dissemination of antimicrobial resistance, reinforcing concerns about its contribution to the persistence and spread of CC38 in poultry reservoirs. These observations have implications for zoonotic transmission and resistance gene dissemination in *E. coli.*

Clusters II-A and II-B showed strong associations with UPEC and ExPEC virulence profiles. This observation aligns with previous studies indicating that ST38 isolates can carry multiple virulence factors across different pathotypes, contributing to their adaptability in various hosts (2). The clusters were primarily associated with *fimH*5, a subtype previously linked to UTIs and BSIs (46–49). Interestingly, Cluster II-B included a notable proportion lacking the *fimH* gene entirely (*fimH*0), indicating reliance on alternative adhesion mechanisms.

A clear contrast emerged with the poultry-associated Cluster II-D, which was characterized by high number of pAmpC resistance genes largely driven by bla_{CMY-2}, a gene associated with poultry, and particularly broiler chickens (50,51). QRDR mutations were notably absent from Clusters II-D, II-E, and II-G, suggesting that fluoroquinolone resistance is not a defining feature of these sublineages and may reflect limited antimicrobial pressure in their respective host environments. The almost universal ExPEC classification and high prevalence of ColV point to a livestock-associated genetic signature (28). The ColRNAI and Col(MG828) replicons mirrored the trend observed in Cluster I, further suggesting that these plasmids represent livestock

associated genetic elements linked to persistence and fitness in poultry environments. The predominant *fimH65* allele observed in Cluster II-D has not previously been associated with specific sources, and its functional role in adhesion and colonization remains unknown. Given its recurrent presence in poultry-associated clusters, further investigations are needed to elucidate its role in host adaption and pathogenicity.

Together, Cluster II-D and II-E exemplify contrasting patterns of host association within CC38; while II-D shows a distinct poultry-associated genetic signature, II-E reflects mixed-origin characteristics, encompassing 74% human isolates alongside representatives of all other source categories, indicating potential cross-species transmission dynamics. However, the predominance of *fimH*54 in Cluster II-E, typically linked to environmental and animal isolates (52–56), points to potential cross-species transmission events shaping this cluster. The presence of Danish poultry-associated isolates (DK-Group Ic) within Cluster II-E further supports this mixed-origin hypothesis. This suggests that some isolates in this cluster may have origins from poultry, but have since diversified into human-associated groups, demonstrating a potential historical cross-species transmission event with subsequent establishment in human-associated populations. While 74% of isolates in Cluster II-E are of human origin, the genetic traits observed highlight the complex evolutionary dynamics of CC38.

Clusters exhibiting lower antimicrobial resistance and reduced virulence, notably II-F and II-C, were distinguished by diverse host origins, *fimH* subtypes, and plasmid profiles, suggestive of cross-host association pattern and environmental persistence. Particularly, Cluster II-C displayed mixed human-poultry signatures with occasional presence of pAmpC genes, and the *fimH65* as in the poultry-associated Cluster II-D. In contrast, Cluster II-F was associated with both *fimH*65 and *fimH*27, the latter previously linked to human-associated extraintestinal infections (57), suggesting a potential cross-host association of this cluster. Notably, over half of the isolates in Cluster II-C harbored QRDR mutations, indicating a substantial level of

fluoroquinolone resistance within this group. This contrasts with Cluster II-F, where QRDR mutations were less common, further distinguishing the two clusters in terms of resistance potential. Despite their differences, Cluster II-C and II-F share common genetic features indicative of cross-host association, reflecting the dynamic exchange of resistance and virulence elements between human, poultry, and environmental reservoirs. Both clusters showed levels of ColV, ColRNAI and Col(MG828) consistent with poultry-associated signatures. This reinforces the hypothesis of frequent genetic exchange among ST38 strains across hosts and environments.

Clusters primarily shaped by antimicrobial selection pressures rather than classical virulence determinants, such as Cluster II-H (human-associated, strongly resistant, low virulence) and Cluster II-G (human, wildlife and companion animal, predominantly pAmpC resistant), highlight the potential clinical impact of these isolates and illustrate the complex interplay between ecological niches, resistance dissemination, and pathogenicity evolution. Cluster II-H exhibited the highest proportion of QRDR mutations among all sublineages (82%), reflecting strong fluoroquinolone selection pressure. This extensive resistance, coupled with low virulence, suggests a lineage adapted to survival and transmission in antimicrobial-rich environments. The Col156 replicon was particularly prevalent in Cluster II-G, and has previously been reported in *E. coli* from silver gulls and human isolates in Oceania (52). These findings suggest that Cluster II-G may represent a lineage with strong environmental and wildlife associations. However, the presence of isolates from wildlife and companion animals could also reflect anthropozoonotic spillover of human-adapted strains, as previously observed for *E. coli* ST131 and ST1193 (58,59).

The observed patterns within Cluster II reflect a complex mosaic of host association patterns and resistance evolution, markedly different from the more poultry-associated Cluster I. This underscores the multifaceted adaptation strategies of *E. coli* CC38, highlighting distinct pattern

of host association, antimicrobial resistance, virulence determinants, and plasmid composition, which reflects this lineage's divergent evolutionary paths and clinical relevance.

Together, these findings indicate that carriage of ColRNAI and Col(MG828) individually predicts poultry origin, while their combined presence serves as a more specific marker of livestock-associated CC38. These associations are correlative and reflect statistical enrichment rather than demonstrated functional adaptation. The clear clustering of co-carriers in Cluster II-D, across multiple sources, suggests a consistent host associated distribution for this lineage, potentially linked to zoonotic transmission and antimicrobial resistance dissemination. This host association pattern may also have practical implications for foodborne transmission, as these plasmid combinations are enriched in poultry- and livestock-associated lineages that can enter the food chain. Incorporating such replicon markers into routine genomic surveillance could enhance early detection of zoonotic spillover and help prioritize monitoring efforts across animal and human reservoirs.

Conclusion

This study provides a comprehensive genomic characterization of *Escherichia coli* CC38, highlighting its global distribution, antimicrobial resistance patterns, and host association. Phylogenetic, resistome, and plasmid analyses revealed distinct host-associated lineages, with clear divergence between human and poultry clusters. As the study primarily includes resistant isolates collected through surveillance, the results should be interpreted within this context. The data may also reflect biases in representativeness across hosts and regions, and potential zoonotic links can only be inferred indirectly from genomic associations.

A key finding is the consistent association of ColRNAI and Col(MG828) co-carriage with poultry and livestock sub-lineages, particularly Cluster I and Cluster II-D. Multinomial regression confirmed their predictive value for poultry origin, while co-carriage was a specific marker of

livestock-associated strains. The phylogenetic clustering of co-carriers, most notably the exclusive presence of all unknown food co-carriers in Cluster II-D, supports the role of these replicons as host associated markers. However, functional studies are needed to confirm their ecological and evolutionary significance.

Additional subtyping confirmed strong host associations of specific *fimH* alleles, further underscoring ecological structuring within CC38. The analysis of Danish bloodstream and other human isolates corroborated these findings and showed limited evidence of nosocomial spread, suggesting a larger role for environmental and community transmission in CC38 dissemination.

Altogether, these findings emphasize the importance of a One Health perspective, integrating genomic surveillance across human, food, and animal sectors to monitor CC38 and related lineages. Tracking host associated plasmid markers such as ColRNAI and Col(MG828) in routine monitoring could provide early warning of emerging zoonotic lineages and inform risk-based interventions at the human–animal interface.

Data Availability

All *E. coli* CC38 sequence data generated as part of the Danish National Surveillance that have not previously been published are deposited in the European Nucleotide Archive (ENA) as part of the BioProjects *National Surveillance of ESBL-producing Escherichia coli* (BioProject PRJEB75816) and *National Surveillance of Carbapenemase-Producing Organisms* (BioProject PRJEB75178). Accession numbers for the raw sequencing read data included in this study are provided in Supplementary Data 1.

In addition, global *Escherichia coli* CC38 genome sequences analyzed in this study were obtained from public repositories, including the ENA Sequence Read Archive (SRA), NCBI SRA, and EnteroBase. Accession numbers and assembly barcodes for all publicly available

genomes are listed in Supplementary Data 1. The EnteroBase *E. coli* collection is accessible at https://enterobase.warwick.ac.uk/species/index/ecoli.

The source data for Figure 1 are provided in Supplementary Data 5, and the corresponding phylogenetic tree is proved in Supplementary Data 6. The source data for Figure 2 are provided in Supplementary Data 7. The phylogenetic tree for Figure 3 is provided in Supplementary Data 8, and source file for Figure 3 and Table 1 are provided in Supplementary Data 1 and 4.

Acknowledgement

Pia Thurø Hansen, Line Toft Madsen and Søren Iversen at Statens Serum Institut are thanked for their excellent laboratory assistance. We also thank Jette Sejer Kjeldgaard, Pimlapas Shinny Leekitcharoenphon, and Natasia Rebekka Thornval at the National Food Institute, Technical University of Denmark, for assistance with sequencing data.

Part of this work was supported by the Danish Ministry of Health.

Author contributions

L.R. and M.S. contributed conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, validation, visualization, writing—original draft, and writing—review and editing. A.M.H. and H.H. contributed conceptualization, original draft, and writing—review and editing. A.R., L.E.B.C., R.S., D.E.P and M.A. contributed bioinformatic analyses, figure preparation, and writing—review and editing. F.H. contributed laboratory analyses, sequencing, and registration of Danish surveillance isolates, as well as writing—review and editing. F.S. contributed virulence profiling and writing—review and editing. S.A. and L.B.P. contributed resources, supervision, and writing—review and editing, with additional input on conceptualization and interpretation. R.S.H., B.D.J., J.R.J., B.J.H., L.S., K.S., D.B.H., U.S.J., C.Ø., T.S.S., M.T.K.N., M.W., and H.L.N. contributed resources (isolate collection), data curation, and writing—review and editing.

Competing Interests

The authors declare no competing interests.

References

1. Murray CJ, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, *et al*. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet. 2022;399(10325):629–55.

2. Roy Chowdhury P, Hastak P, demaere M, Wyrsch E, Li D, Elankumaran P, *et al.* Phylogenomic analysis of a global collection of *Escherichia coli* ST38: evidence of interspecies and environmental transmission? mSystems. 2023;8(5).

3. Ahlstrom CA, Woksepp H, Sandegren L, Ramey AM, Bonnedahl J. Exchange of carbapenem-resistant *Escherichia coli* Sequence Type 38 intercontinentally and among wild bird, human, and environmental niches. Elkins CA, editor. Appl Environ Microbiol [Internet]. 2023 Jun 28;89(6):1–8. Available from: https://journals.asm.org/doi/10.1128/aem.00319-23

4. Clark JR, Maresso AM. Comparative pathogenomics of *Escherichia coli*: Polyvalent vaccine target identification through virulome analysis. Monack D, editor. Infect Immun [Internet]. 2021 Jul 15;89(8). Available from: https://journals.asm.org/doi/10.1128/IAI.00115-21

5. Moez NM, Mashouf RY, Sedighi I, Shokoohizadeh L, Taheri M. Phylogroup classification and investigation the relationships between phylogroups and antibiotic resistance patterns of uropathogenic *E. coli* isolated from pediatric urinary tract infection. Gene Reports [Internet]. 2020 Sep;20(May):100758. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2452014420301722

6. Mo SS, Fiskebeck EZ, Slettemeås JS, Lagesen K, Nilsson O, Naseer U, *et al.* *Escherichia coli* multilocus sequence type 38 from humans and broiler production represent distinct monophyletic groups. Front Microbiol [Internet]. 2023 May

2;14(May):1173287. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1173287/full

7. Massella E, Giacometti F, Bonilauri P, Reid CJ, Djordjevic SP, Merialdi G, *et al.* Antimicrobial resistance rrofile and EXPEC virulence potential in commensal *Escherichia coli* of multiple sources. Antibiotics [Internet]. 2021 Mar 26;10(4):351. Available from: https://www.mdpi.com/2079-6382/10/4/351

8. Chattaway MA, Jenkins C, Ciesielczuk H, Day M, Donascimento V, Day M, *et al.* Evidence of evolving extraintestinal enteroaggregative *Escherichia coli* ST38 clone. Emerg Infect Dis. 2014;20(11):1935–7.

9. Mostafa HH, Cameron A, Taffner SM, Wang J, Malek A, Dumyati G, *et al*. Genomic surveillance of ceftriaxone-resistant *Escherichia coli* in Western New York suggests the extended-spectrum β-lactamase *bla*CTX-M-27 is emerging on distinct plasmids in ST38. Front Microbiol. 2020;11(July).

10. Lindemann PC, Pedersen T, Oma DH, Janice J, Grøvan F, Chedid GM, *et al.* Intraregional hospital outbreak of OXA-244-producing *Escherichia coli* ST38 in Norway, 2020. Eurosurveillance. 2023;28(27):1–8.

11. Hammerum AM, Porsbo LJ, Hansen F, Roer L, Kaya H, Henius A, *et al.* Surveillance of OXA-244-producing *Escherichia coli* and epidemiologic investigation of cases, Denmark, January 2016 to August 2019. Eurosurveillance [Internet]. 2020;25(18):1–9. Available from: http://dx.doi.org/10.2807/1560-7917.ES.2020.25.18.1900742

12. Falgenhauer L, Nordmann P, Imirzalioglu C, Yao Y, Falgenhauer J, Hauri AM, *et al.* Cross-border emergence of clonal lineages of ST38 *Escherichia coli* producing the OXA-48-like carbapenemase OXA-244 in Germany and Switzerland. Int J Antimicrob Agents [Internet]. 2020;56(6):106157. Available from:

https://doi.org/10.1016/j.ijantimicag.2020.106157

13. DANMAP 2022 - Use of antimicrobial agents and occurrence of antimicrobial resistance in bacteria from food animals, food and humans in Denmark. [Internet]. 2023. Available from: www.danmap.org

14. Alghoribi MF, Gibreel TM, Farnham G, Al Johani SM, Balkhy HH, Upton M. Antibiotic-resistant ST38, ST131 and ST405 strains are the leading uropathogenic *Escherichia coli* clones in Riyadh, Saudi Arabia. J Antimicrob Chemother. 2015;70(10):2757–62.

15. Guenther S, Semmler T, Stubbe A, Stubbe M, Wieler LH, Schaufler K. Chromosomally encoded ESBL genes in *Escherichia coli* of ST38 from Mongolian wild birds. J Antimicrob Chemother. 2017;72(5):1310–3.

16. Vogt D, Overesch G, Endimiani A, Collaud A, Thomann A, Perreten V. Occurrence and genetic characteristics of third-generation cephalosporin-resistant *Escherichia coli* in Swiss retail meat. Microb Drug Resist. 2014;20(5):485–94.

17. Fonseca EL, Morgado SM, Caldart R V., Vicente AC. Global Genomic epidemiology of *Escherichia coli* (EXPEC) ST38 lineage revealed a virulome associated with human infections. Microorganisms. 2022;10(12).

18. Xu C, Li Y, Li Y, Huang L, Zhou M, Xu Y, *et al.* Characterization of an ST38 carbapenem-resistant and highly virulent *Escherichia coli* carrying conjugatively transferable ColV virulence–resistance and *bla*NDM-5-positive resistance plasmids. J Antimicrob Chemother. 2024;79(2):447–52.

19. Grevskott DH, Ghavidel FZ, Svanevik CS, Marathe NP. Resistance profiles and diversity of β-lactamases in *Escherichia coli* strains isolated from city-scale sewage surveillance in Bergen, Norway mimic clinical prevalence. Ecotoxicol Environ Saf [Internet]. 2021

Dec;226(July):112788. Available from: https://doi.org/10.1016/j.ecoenv.2021.112788

20. Belotindos L, Villanueva M, Miguel J, Bwalya P, Harada T, Kawahara R, et al. Prevalence and characterization of quinolone-resistance determinants in *Escherichia coli* isolated from food-producing animals and animal-derived food in the Philippines. Antibiotics [Internet]. 2021 Apr 9;10(4):413. Available from: https://www.mdpi.com/2079-6382/10/4/413

21. Souvorov A, Agarwala R, Lipman DJ. SKESA: Strategic k-mer extension for scrupulous assemblies. Genome Biol. 2018;19(1):1–13.

22. Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, *et al.* Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol [Internet]. 2006 Jun;60(5):1136–51. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16689791

23. Clausen PTLC, Aarestrup FM, Lund O. Rapid and precise alignment of raw reads against redundant databases with KMA. BMC Bioinformatics. 2018;19(1):1–8.

24. Waters NR, Abram F, Brennan F, Holmes A, Pritchard L. Easy phylotyping of *Escherichia coli* via the ezclermont web app and command-line tool. Access Microbiol. 2020;2(9).

25. Malberg Tetzschner AM, Johnson JR, Johnston BD, Lund O, Scheutz F. *In silico* genotyping of *Escherichia coli* isolates for extraintestinal virulence genes by use of whole-genome sequencing data. Dekker JP, editor. J Clin Microbiol [Internet]. 2020 Sep 22;58(10):1–13. Available from: https://journals.asm.org/doi/10.1128/JCM.01269-20

26. Sherry NL, Horan KA, Ballard SA, Gonçalves da Silva A, Gorrie CL, Schultz MB, et al. An ISO-certified genomics workflow for identification and surveillance of antimicrobial resistance. Nat Commun. 2023;14(1).

27. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, *et al.* In

Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014 Jul;58(7):3895–903.

28. Liu CM, Stegger M, Aziz M, Johnson TJ, Waits K, Nordstrom L, *et al. Escherichia coli* ST131-H22 as a foodborne uropathogen. Mobley HLT, Ravel J, editors. Mbio [Internet]. 2018 Aug 28;9(4). Available from: https://journals.asm.org/doi/10.1128/mbio.00470-18

29. Sahl JW, Lemmer D, Travis J, Schupp JM, Gillece JD, Aziz M, *et al.* NASP: an accurate, rapid method for the identification of snps in WGS datasets that supports flexible input and output formats. Microb Genomics [Internet]. 2016 Aug 25;2(8):1–12. Available from: http://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000074

30. Delcher AL. Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res [Internet]. 2002 Jun 1;30(11):2478–83. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/30.11.2478

31. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics [Internet]. 2009 Jul 15;25(14):1754–60. Available from: https://academic.oup.com/bioinformatics/article/25/14/1754/225615

32. Mckenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, *et al.* The genome analysis toolkit: A mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res [Internet]. 2010 Sep;20(9):1297–303. Available from: http://genome.cshlp.org/lookup/doi/10.1101/gr.107524.110

33. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res [Internet]. 2015 Feb 18;43(3):e15. Available from: http://academic.oup.com/nar/article/43/3/e15/2410982/Rapid-phylogenetic-analysis-of-large-samples-of

34. Price MN, Dehal PS, Arkin AP. Fasttree 2--approximately maximum-likelihood trees for large alignments. Poon AFY, editor. Plos One [Internet]. 2010 Mar 10;5(3):e9490. Available from: http://dx.plos.org/10.1371/journal.pone.0009490

35. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res [Internet]. 2016 Jul 8;44(W1):W242–5. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw290

36. Roer L, Hansen F, Frølund Thomsen MC, Knudsen JD, Hansen DS, Wang M, *et al.* WGS-based surveillance of third-generation cephalosporin-resistant *Escherichia coli* from bloodstream infections in Denmark. J Antimicrob Chemother [Internet]. 2017 Jul 22;72(7):1922–9. Available from: https://academic.oup.com/jac/article-lookup/doi/10.1093/jac/dkx092

37. Roer L, Overballe-Petersen S, Hansen F, Schønning K, Wang M, Røder BL, *et al.* *Escherichia coli* sequence type 410 is causing new international high-risk clones . Castanheira M, editor. Msphere [Internet]. 2018 Jul 18;3(4):e00337-18. Available from: http://msphere.asm.org/lookup/doi/10.1128/msphere.00337-18

38. Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National patient registry: A review of content, data quality, and research potential. Clin Epidemiol. 2015;7:449–90.

39. Liu CM, Aziz M, Park DE, Wu Z, Stegger M, Li M, *et al*. Using source-associated mobile genetic elements to identify zoonotic extraintestinal *E. coli* infections. One Heal [Internet]. 2023;16(February):100518. Available from: https://doi.org/10.1016/j.onehlt.2023.100518

40. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2022. Available from: https://www.r-

project.org/

41. European Centre for Disease Prevention and Control. Increase in OXA-244-producing *Escherichia coli* in the European Union / European Economic Area and the UK since 2013, first update - 20 July 2021. Stockholm; 2021.

42. Kipper D, Carroll LM, Mascitti AK, Streck AF, Fonseca ASK, Ikuta N, *et al.* Genomic characterization of *Salmonella* Minnesota clonal lineages associated with poultry production in Brazil. Animals [Internet]. 2020 Nov 5;10(11):2043. Available from: https://www.mdpi.com/2076-2615/10/11/2043

43. Wang W, Cui J, Liu F, Hu Y, Li F, Zhou Z, *et al.* Genomic characterization of *Salmonella* isolated from retail chicken and humans with diarrhea in Qingdao, China. Front Microbiol [Internet]. 2023 Dec 18;14(December):1–13. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1295769/full

44. Flatgard BM, Williams AD, Amin MB, Hobman JL, Stekel DJ, Rousham EK, *et al*. Tracking antimicrobial resistance transmission in urban and rural communities in Bangladesh: a One Health study of genomic diversity of ESBL-producing and carbapenem-resistant *Escherichia coli*. Babiker A, editor. Microbiol Spectr [Internet]. 2024 Jun 4;12(6). Available from: https://journals.asm.org/doi/10.1128/spectrum.03956-23

45. Sugita K, Aoki K, Komori K, Nagasawa T, Ishii Y, Iwata S, *et al.* Molecular analysis of *bla*KPC-2-Harboring plasmids: Tn4401a interplasmid transposition and Tn4401a-carrying ColRNAI plasmid mobilization from *Klebsiella pneumoniae* to *Citrobacter europaeus* and *Morganella morganii* in a single patient. Msphere [Internet]. 2021 Dec 22;6(6):e0085021. Available from: http://www.ncbi.nlm.nih.gov/pubmed/34730375

46. Roer L, Hansen F, Hasman H, Hammerum AM, Cavaco LM. Characterisation of extended-spectrum β-lactamase/plasmid ampc-β-lactamase-producing *Escherichia coli*

isolates from long-term recurrent bloodstream infections. Int J Antimicrob Agents. 2020;56(1):10–4.

47. Emery A, Hocquet D, Bonnet R, Bertrand X. Genotypic characteristics and antimicrobial resistance of *Escherichia coli* ST141 clonal group. Antibiotics. 2023;12(2):1–10.

48. Roer L, Tchesnokova V, Allesoe R, Muradova M, Chattopadhyay S, Ahrenfeldt J, *et al.* Development of a web tool for *Escherichia coli* subtyping based on fimh alleles. Diekema DJ, editor. J Clin Microbiol [Internet]. 2017 Aug;55(8):2538–43. Available from: http://jcm.asm.org/lookup/doi/10.1128/JCM.00737-17

49. Bunduki GK, Heinz E, Phiri VS, Noah P, Feasey N, Musaya J. Virulence factors and antimicrobial resistance of uropathogenic *Escherichia coli* (UPEC) isolated from urinary tract infections: a systematic review and meta-analysis. BMC Infect Dis. 2021;21(1):1–13.

50. Koga VL, Maluta RP, da Silveira WD, Ribeiro RA, Hungria M, Vespero EC, *et al.* Characterization of CMY-2-type beta-lactamase-producing *Escherichia coli* isolated from chicken carcasses and human infection in a city of South Brazil. BMC Microbiol [Internet]. 2019 Dec 30;19(1):174. Available from: https://bmcmicrobiol.biomedcentral.com/articles/10.1186/s12866-019-1550-3

51. Pietsch M, Irrgang A, Roschanski N, Brenner Michael G, Hamprecht A, Rieber H, *et al.* Whole genome analyses of CMY-2-producing *Escherichia coli* isolates from humans, animals and food in Germany. BMC Genomics [Internet]. 2018 Dec 9;19(1):601. Available from: https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-018-4976-3

52. Díaz-Jiménez D, García-Meniño I, Herrera A, García V, López-Beceiro AM, Alonso MP, *et al*. Genomic characterization of *Escherichia coli* isolates belonging to a new hybrid aEPEC/EXPEC pathotype O153:H10-A-ST10 *eae*-beta1 occurred in meat, poultry,

wildlife and human diarrheagenic samples. Antibiotics [Internet]. 2020 Apr 17;9(4):192. Available from: https://www.mdpi.com/2079-6382/9/4/192

53. Manageiro V, Jones-Dias D, Ferreira E, Caniça M. Plasmid-mediated colistin resistance (*mcr*-1) in *Escherichia coli* from non-imported fresh vegetables for human consumption in Portugal. Microorganisms [Internet]. 2020 Mar 18;8(3):429. Available from: https://www.mdpi.com/2076-2607/8/3/429

54. Fu Y, Nawrocki EM, m'ikanatha NM, Dudley EG. Host species shapes genotype, antimicrobial resistance, and virulence profiles of enterotoxigenic *Escherichia coli* (ETEC) from livestock in the United States. Dozois CM, editor. Appl Environ Microbiol [Internet]. 2024 Aug 21;90(8). Available from: https://journals.asm.org/doi/10.1128/aem.00749-24

55. De Lagarde M, Vanier G, Arsenault J, Fairbrother JM. High risk clone: A proposal of criteria adapted to the One Health context with application to enterotoxigenic *Escherichia coli* in the pig population. Antibiotics [Internet]. 2021 Feb 28;10(3):244. Available from: https://www.mdpi.com/2079-6382/10/3/244

56. Hayashi W, Tanaka H, Taniguchi Y, Iimura M, Soga E, Kubo R, *et al.* Acquisition of *mcr*-1 and cocarriage of virulence genes in avian pathogenic *Escherichia coli* isolates from Municipal wastewater influents in Japan. Elkins CA, editor. Appl Environ Microbiol [Internet]. 2019 Nov 15;85(22):1–11. Available from: https://www.mdpi.com/2079-6382/9/4/192

57. Mandomando I, Vubil D, Boisen N, Quintó L, Ruiz J, Sigaúque B, *et al*. *Escherichia coli* ST131 clones harbouring Aggr and AAF/V fimbriae causing bacteremia in Mozambican children: Emergence of new variant of *fimH*27 subclone. Blanco J, editor. Plos Negl Trop Dis [Internet]. 2020 May 1;14(5):e0008274. Available from: https://dx.plos.org/10.1371/journal.pntd.0008274

58. Kidsley AK, White RT, Beatson SA, Saputra S, Schembri MA, Gordon D, *et al.* Companion animals are spillover hosts of the multidrug-resistant human extraintestinal *Escherichia coli* pandemic clones ST131 and ST1193. Front Microbiol [Internet]. 2020 Sep 2;11(September):1–10. Available from: https://www.frontiersin.org/article/10.3389/fmicb.2020.01968/full

59. Mukerji S, Stegger M, Truswell AV, Laird T, Jordan D, Abraham RJ, *et al.* Resistance to critically important antimicrobials in Australian silver gulls (*Chroicocephalus novaehollandiae*) and evidence of anthropogenic origins. J Antimicrob Chemother [Internet]. 2019 Sep 1;74(9):2566–74. Available from: https://academic.oup.com/jac/article/74/9/2566/5523291

Table 1. Summary of predominant features across global CC38 *E. coli* clusters

Cluster	Source	Resistance Profile	Virulence Profile	*fimH* subtype	Key Plasmids
I	Poultry (83%)	ESBL (29%)	ExPEC (96%), UPEC hybrid (50%)	*fimH*270 (82%)	IncFIB, ColV ColRNAI Col(MG828)
II-A	Human (85%)	ESBL (88%) Carb (35%)	UPEC (100%), ExPEC hybrid (61%)	*fimH*5 (90%)	Low, IncF diversity
II-B	Human (85%)	ESBL (85%)	UPEC (92%) ExPEC (63%) EAEC (23%)	*fimH*5 (64%) *fimH*0 (29%)	IncFIB/FII
II-C	Human (63%) Poultry (24%)	QRDR (56%) ESBL (48%)	UPEC/ExPEC (58%)	*fimH*65 (80%)	IncFIB, ColV ColRNAI Col(MG828)
II-D	Poultry (79%)	pAmpC (78%)	ExPEC (97%)	*fimH*65 (95%)	IncFIB p0111 ColV ColRNAI Col(MG828)
II-E	Human (74%)	ESBL (85%) Carb (39%)	None (93%)	*fimH*54 (81%)	Diverse plasmids
II-F	Mixed host	None (68%)	None (61%)	*fimH*65 (59%) *fimH*27 (30%)	IncF diversity, ColV ColRNAI Col(MG828)
II-G	Human (54%) Wildlife (29%)	pAmpC (86%)	UPEC (98%)	*fimH*26 (92%)	IncFIB/FII Col156
II-H	Human (89%)	QRDR (82%) ESBL (61%)	None (82%)	*fimH*5 (80%)	IncFIA/FIB, Col(MG828)

Table 2. Distribution and predictive value of ColRNAI and Col(MG828) carriage across isolate sources in comparison to poultry

		No. (%) with plasmid			Multinomial logistic regression [OR (95% CI, p) [b]]		
Source	**n[a]**	**ColRNAI**	**Col(MG828)**	**Co-carriers**	**ColRNAI**	**Col(MG828)**	**Interaction**
Poultry	462	326 (71)	368 (80)	290 (63)			
Companion Animal	101	7 (7)	15 (15)	5 (5)	0.04 (0.01–0.17) p<0.0001	0.09 (0.04–0.19) p<0.0001	3.50 (0.56–21.87) p=0.1798
Environment	22	2 (9)	3 (17)	2 (9)	0.00 (0.00–4e+158) p=0.9463	0.04 (0.01–0.30) p=0.0018	2.4e+5 (0.00–4.3e+169) p=0.9489
Human	1682	133 (8)	216 (13)	63 (4)	0.08 (0.05–0.13) p<0.0001	0.08 (0.06–0.12) p<0.0001	1.37 (0.74–2.54) p=0.3153
Livestock	52	16 (31)	15 (29)	14 (27)	0.09 (0.02–0.41) p=0.0016	0.02 (0.00–0.16) p=0.0002	40.95 (3.27–512.65) p=0.0040
Unknown food	63	54 (86)	54 (86)	52 (83)	0.46 (0.09–2.34) p=0.3494	0.21 (0.04–1.06) p=0.0590	15.20 (1.74–132.92) p=0.0139
Water/River	107	13 (12)	12 (11)	6 (6)	0.13 (0.05–0.31) p<0.0001	0.05 (0.02–0.12) p<0.0001	2.10 (0.49–8.97) p=0.3168
Wildlife	149	7 (5)	12 (8)	4 (3)	0.04 (0.01–0.12) p<0.0001	0.04 (0.02–0.10) p<0.0001	3.72 (0.66–20.96) p=0.1359

[a]The total number of isolates (n) for each source category.

[B]Odds ratios (OR), 95% confidence intervals (CI), and p-values were derived from a multinomial logistic regression model using poultry as the reference group. Significant values ($p < 0.05$) are interpreted as evidence of differential plasmid association by source. Livestock was defined as non-poultry food-producing animals, including goats and horses, to distinguish them analytically from poultry

Figure 1. Phylogenetic structure and host prediction of Danish *E. coli* CC38 isolates.

(A) Midpoint rooted maximum-likelihood phylogeny of 325 Danish *E. coli* CC38 isolates based on core genome SNPs. The tree reveals two major groups (Group I and II) colored by sample source, sequence type (ST), and possible outbreak groups. **(B)** Host prediction results using a Bayesian latent class model based on accessory genome markers. Arcs indicate predicted source (>80% certainty) and most likely host (chicken, turkey, or human). Group II was predominantly poultry-associated, while Group I showed more diverse predictions.

Figure 2. Temporal and geographic distribution of global *E. coli* CC38 isolates

From left, yearly distribution of 2,638 CC38 isolates grouped by source, with isolates prior to year 2000 grouped together. Each dot represents the number of isolates per year per source, with red dots summarizing total counts per year. Right part of figure demonstrates the geographic distribution of isolates by continent. Livestock was defined as non-poultry food-producing animals, including goats and horses, to distinguish them analytically from poultry

Figure 3. Phylogenetic structure and genomic features of global *E. coli* CC38 clusters.

Midpoint rooted maximum-likelihood phylogeny of global CC38 isolates, highlighting two major clusters: Cluster I (poultry-associated) and Cluster II (human-associated, divided into sub-clusters II-A to II-H). Colored dots indicate aggregated presence/absence of selected features observed ≥50% of isolates within each cluster: Source (●Human, ●Poultry), Sequence Type (●ST115, ●ST38, ●ST963,,●ST2659), *fimH*-subtype (●*fimH*270, ●*fimH*5, ●*fimH*65, ●*fimH*54, ●*fimH*26), Resistance Profile (●ESBL, ●QRDR mutations, ●pAmpC), Virulence Profiles (●UPEC, ●ExPEC), and Plasmid Replicons (●incFIA, ●IncFIB, ●IncFII, ●p0111). ColV and host

association features were included if present ≥20% of the isolates within each cluster: ●ColV+ ●ColRNAI+ and ●Col(MG828)+.

ED Summary:

Roer et al. analyze genomes of Escherichia coli clonal complex 38 from humans, animals, and food to investigate its spread and host associations. They find distinct human and poultry lineages and identify plasmid markers linked to animal adaptation and zoonotic transmission.

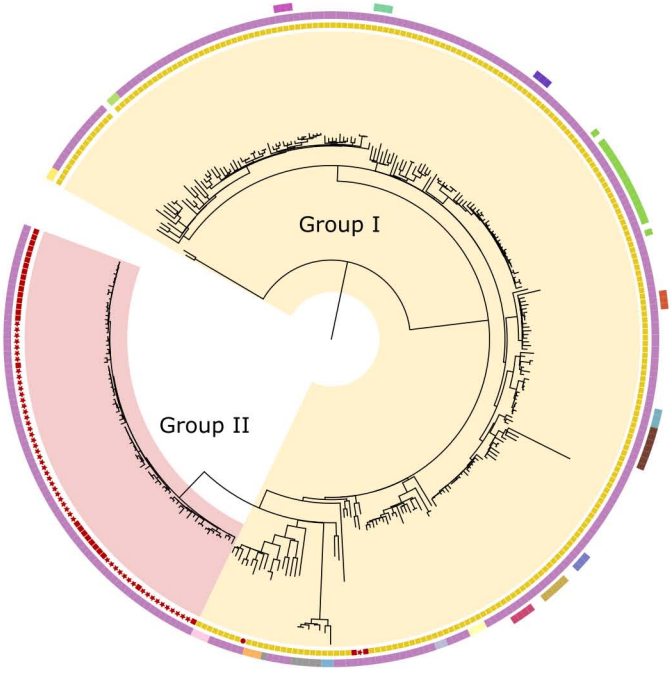
Peer review information: *Communications Medicine* thanks Reuben Maghembe and Ivan Sserwadda for their contribution to the peer review of this work. A peer review file is available.

A

Sample source (inner ring)
- Human
- Poultry
- Livestock
- Unknown food

Sequence type (middle ring)
- ST38
- ST315
- ST963
- ST1177
- ST2003
- ST2659
- ST3052
- ST3268
- ST3666
- ST15137

Possible outbreaks (outer ring)
- PO-1
- PO-2
- PO-3
- PO-4
- PO-5
- PO-6
- PO-7
- PO-8
- PO-9
- PO-10



Tree scale: 0.01

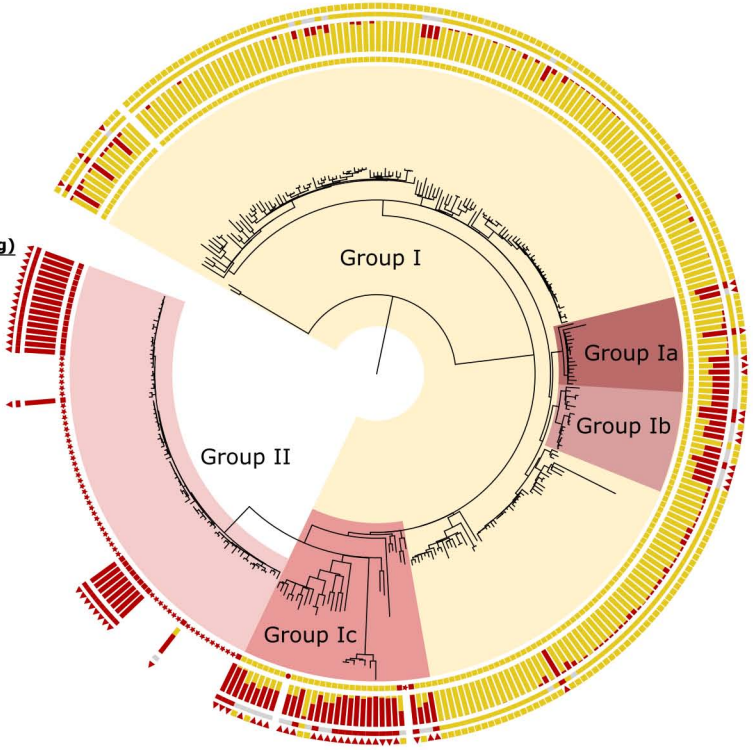
B

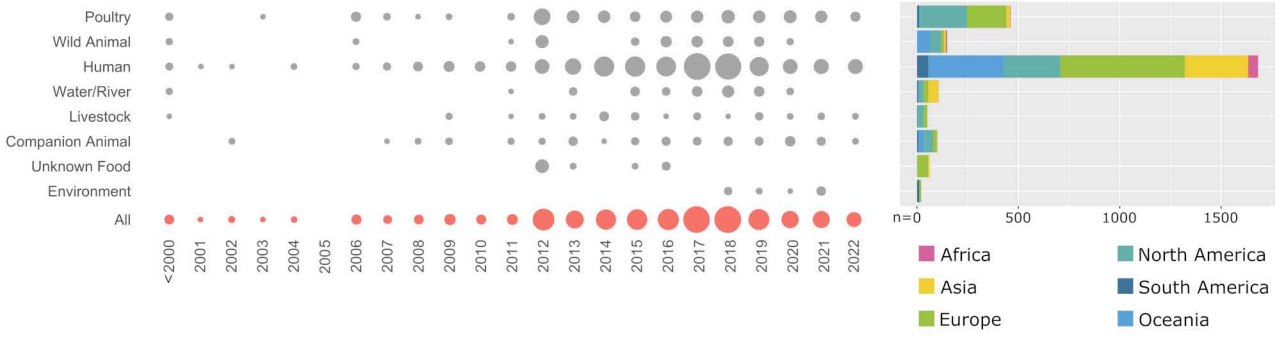
Sample source (inner ring)
- Human
- Poultry
- Livestock
- Unknown food

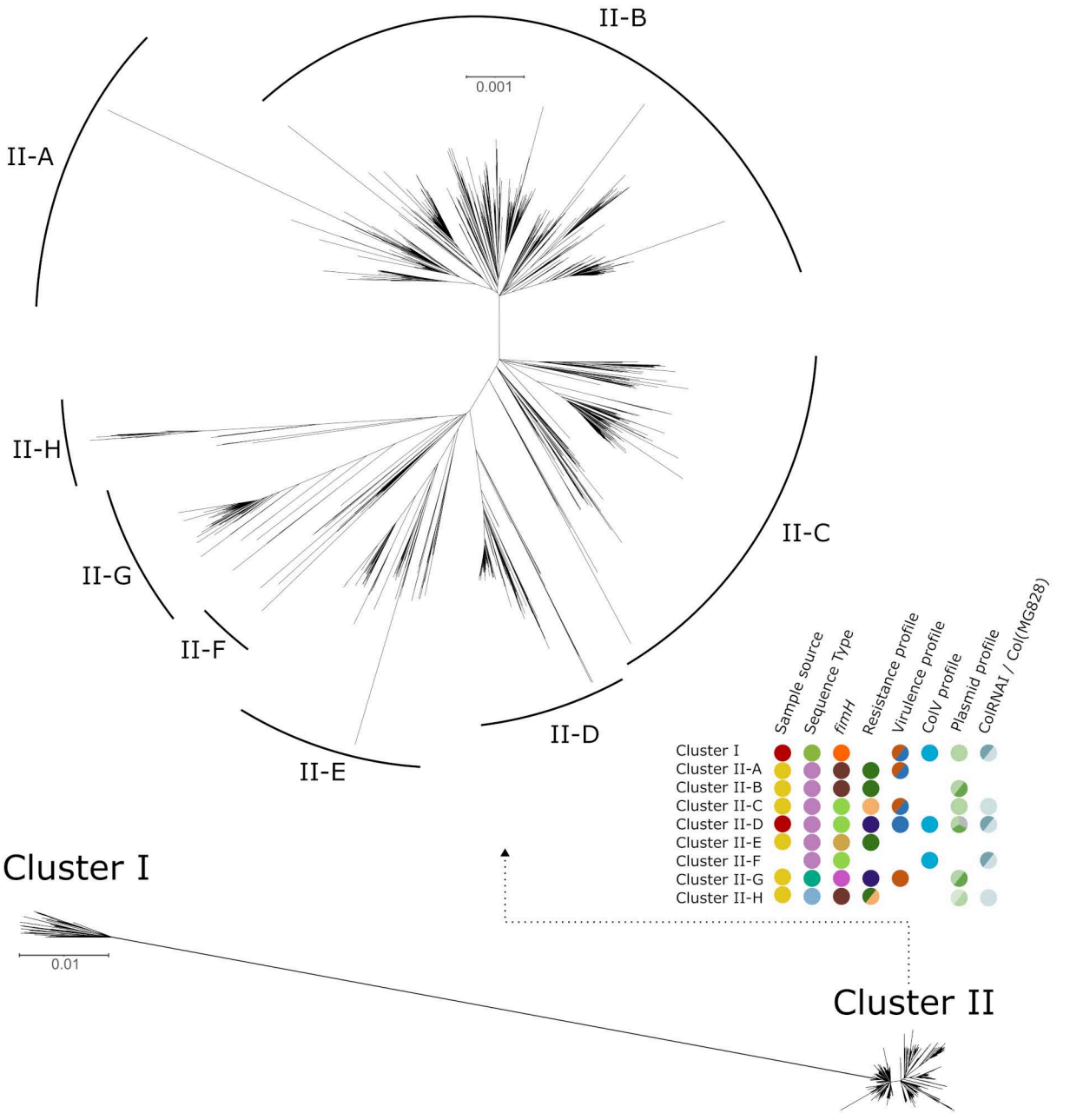
% Predicted (second ring)
- % Predicted Human
- % Predicted Meat

Predicted host > 80% (third ring)
- Meat
- Human
- Indeterminate

Host assignment (outer ring)
- Chicken
- Turkey
- Human






II-B
0.001
II-A
II-H
II-G
II-F
II-E
II-D
II-C
Sample source
Sequence Type
fimH
Resistance profile
Virulence profile
ColV profile
Plasmid profile
ColRNAI / Col(MG828)
Cluster I
Cluster II-A
Cluster II-B
Cluster II-C
Cluster II-D
Cluster II-E
Cluster II-F
Cluster II-G
Cluster II-H
Cluster I
0.01
Cluster II