Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Global Genomic surveillance of Enterobacter cloacae complex reveals species-specific, income-stratified, and One Health heterogeneity in antimicrobial resistance during 2000- 2026

Sudip Dey
ICMR - National Institute for Research in Bacterial Infections
Santasabuj Das
santasabujdas@yahoo.com

ICMR - National Institute for Research in Bacterial Infections

Research Article

Keywords:

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11723092/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Global Genomic surveillance of Enterobacter *cloacae* complex reveals species-specific, income-stratified, and One Health heterogeneity in antimicrobial resistance during 2000-2026

Sudip Dey[1,2], Dr. Santasabuj Das[1,#]

[1]*ICMR - National Institute for Research in Bacterial Infections*

P-33, CIT Rd, Subhas Sarobar Park, Phoolbagan, Beleghata, Kolkata, West Bengal 700010

[2]*Jadavpur University*

188, Raja Subodh Chandra Mallick Rd, Jadavpur, Kolkata, West Bengal 700032

[#]*Corresponding author: Dr. Santasabuj Das: santasabujdas@yahoo.com*

Abstract:

The Enterobacter cloacae complex(ECC) represents an increasingly important group of opportunistic pathogens, yet its global antimicrobial resistance landscape remains fragmented across taxonomic, geographical and ecological boundaries. Global comparisons of antimicrobial resistance using publicly available bacterial genomes may be distorted by unequal sampling across countries and ecological reservoirs. We analyzed 9,157 non-redundant ECC genomes collected during 2000 - 2026 from 97 countries to investigate species resolved resistance diversity across socio-economic settings and One Health reservoirs. We have revealed extensive heterogeneity in the distribution of antimicrobial resistance determinants with distinct resistance repertoires across ECC species and human, animal and environmental sources. Notably, environmental populations of selected species exhibited greater annotated ARG diversity than corresponding human associated isolates, challenging the assumption that clinically derived collections adequately capture ECC resistome. Clinically important determinants, including *blaCTX-M-15, tet(X4), mcr10.1* showed heterogeneous distributions. Crucial socio-economic differences in species composition observed in pooled genome level analysis were not consistently supported after accounting for unequal country representation, revealing how surveillance imbalance can influence interpretations of global bacterial population patterns. Together, these findings demonstrate that ECC antimicrobial resistance cannot be adequately understood through aggregated genomic surveillance alone and establish the importance of integrating species identity, ecological origin and geographical sampling structure when interpreting the global resistome.

Introduction:

Antimicrobial resistance (AMR) threatens the effective treatment of bacterial infections worldwide. In 2021, an estimated 4.71 million deaths were associated with bacterial AMR , including 1.14 million directly attributable deaths[1]. The World Health Organization has prioritized resistant Enterobacterales because resistance to third generation cephalosporins and

carbapenems limits therapeutic options[2]. Within this group, the Enterobacter cloacae complex (ECC) comprises closely related, opportunistic pathogens associated with bloodstream, respiratory, urinary tract and healthcare associated infections. Genome based classification distinguishes clinically important species, including E. hormaechei ,E. cloacae, E. asburiae, E. kobei , E. ludwigii ,E. roggenkampii and E. bugandensis , whose ecological distributions and resistance repertoires are not interchangeable[3, 4] [5]

The ECC resistome includes both intrinsic and acquired resistance determinants. Chromosomally encoded, inducible AmpC beta-lactamases contribute to the intrinsic resistance properties of several ECC members, while regulatory alterations can further modify susceptibility to beta-lactams[6]. The acquisition of extended spectrum Beta-lactamases, including CTX-M variants and carbapenemases belonging to the NDM, KPC, IMP, VIM and selected OXA families further expands the potential resistance repertoire. In addition, plasmid mediated and other resistance associated determinants can affect aminoglycosides, quinolones, tetracyclines and polymyxin producing complex combinations of resistance mechanisms within individual bacterial genomes[7, 8]. The clinical significance of these genomic findings nevertheless depends on the specific determinant and its expression, genetic background and functional consequences. For example, genomic detection of *mcr*, associated genes does not invariably establish phenotypic colistin resistance, particularly in ECC, where intrinsic and heterogeneous resistance mechanisms are also important[9, 10].

Although ECC is commonly investigated as a hospital associated pathogen, its members occupy diverse ecological niches. Extending beyond clinical environment Enterobacter species have been recovered from human and animal microbiota, wastewater, surface water, agricultural settings and plants, supporting their relevance to the broader ecology of antimicrobial resistance. Prospective One Health investigations have demonstrated that clinically important Enterobacterales and resistance determinants occur across human, animal and environmental components, while studies of less frequently investigated enterobacterales have highlighted environmental reserves that may be overlooked by predominantly clinical surveillance systems[11, 12]. However, identifying related resistance determinants across reservoirs does not in itself establish transmission between them. One health perspective therefore requires both comparative analysis of ecological sources and careful distinction between shared resistance gene carriage and demonstrated epidemiological connectivity[13].

A major challenge in interpreting the global distribution of resistance bacteria is the unequal representation of countries, ecological sources and bacterial populations in genomic repertoires. Publicly available genomes are increasingly valuable for comparative surveillance but they arise from heterogeneous research programs, clinical investigations, outbreak responses and sequencing initiatives with differing objectives and sampling intensities. The amr.watch (https://amr.watch/) Genomic surveillance platform, encompassing more than 620,000 genomes documented, pronounced geographical inequalities with high income countries accounting for the overwhelming majority of deposited genomes represented in that resource[14]. Consequently,

differences in the number or proportion of resistance genomes among geographical regions cannot automatically be interpreted as differences in population level resistance burden. The potential dependence of genomic observations on country representation, isolation, selection and sequencing effort is particularly important when comparing populations grouped by national economic classification.

Socioeconomic context introduces an additional dimension to global AMR surveillance. Differences in healthcare infrastructure, infection prevention capacity, diagnostic access, sanitation, antimicrobial availability, and surveillance resources may contribute to the unequal detection and reporting of resistance pathogens. Large scale genomic investigations have begun integrating national socioeconomic, environmental and health system indicators to examine geographical variations in resistance, demonstrating the potential value of linking genomic information with broader public health contexts[15,16]. Nevertheless, country income classifications represent broad economic groupings rather than direct measurements of antimicrobial use. Genomic comparisons across income categories must therefore account for unequal sampling and avoid attributing observe differences to socio-economic factors without supporting exposure data and appropriate analytical adjustments. Global and regional studies have identified extensive species and sequence type diversity, geographically distributed carbapenemase determinants and substantial heterogeneity in resistance characteristics of E. hormaechei and other ECC species [9, 17, 18]. Nevertheless, the complementary question of how taxonomically resolved ECC genomes and their annotated rhythms represented across global geographical, country, income and human animal environmental categories, and how sensitive such comparisons are to unequal genomic sampling remains important for interpreting publicly available surveillance data.

Here, we collected the publicly available NCBI assemblies, get them filtered., non-redundant ECC, genomes with geographical, temporal and isolation source metadata, spanning collection years from 2000 up to the available portion of July,2026 we investigated species identification, multi local sequence typing (MLST), world bank income classification and genome-based resistance determinant profiling to characterize the global distribution and diversity of ECC. Specially, we examined geographical and temporal patterns of genome representation; evaluated species composition and diversity across isolation sources and income categories, including country level rarefaction and equal country weighting; and compared species resolved resistance gene reporters and clinically relevant acquired determinants across human., animal and environmental collections. By distinguishing genome-weighted observations from country level patterns, this study aims to clarify the opportunities and interpretative limitations of public genome based ECC surveillance and provide a more geographically informed One health perspective on its antimicrobial resistome.

Results:

1.Global geographical distribution, temporal presentation and sequence type diversity of Enterobacter cloacae complex:

To characterize the global genomic presentation and population structure of Enterobacter cloacae complex (ECC), we analyzed 9,157 high quality filtered non redundant genome assemblies collected between 2000 and July, 2026(fig. 1a-c). The collection showed substantial geographical heterogeneity, with largest contributions originated from China (2,322 genomes;25.4%) followed by the United States (963,10.5%), France (830;9.1%), Germany (674;7.4%), and United Kingdom (556;6.1%). Stratification by World Bank income classification revealed an imbalance in genomic representation; 5258 genomes (57.4%) originated from 42 high income locations, 3117(34%) from 24 upper middle-income locations and 767(8.4%) from 27 low- and lower middle-income countries. Additional 15 (0.2%) could not be assigned. The genome level accession identifiers, geographical metadata, collection years and income classifications underlying the distributions are provided in **Supplementary Data 1(Supplementary data 1_ECC_FINAL_2000_2026_MASTER).**

The temporal distribution revealed marked changes in genome collection over the study period (fig. 1b). Only 454 genomes (5%) collected between 2000-2009 whereas 7,619(83.2%) originated from 2015 onwards. Source stratified analysis demonstrated a predominance of human-associated genomes (7283;79.5%) followed by environmental (853;9.3%), animal (686;7.5%) plant associated (247;2.7%) and food associated (88;1.0%) isolates. Taxonomic normalization identified Enterobacter hormaechei as the most extensively represented species (5,876;64.2%) followed by E. cloacae (757;8.3%), E. roggenkampii (650;7.1%), smaller contributions were observed for E. ludwigii (322;3.5%) and E. bugandensis(242;2.6%), while 275(3.0%) genomes were unresolved at species level. MLST further revealed extensive lineage diversity within the ECC collection (fig. 1c). Among 8,882 genomes assigned to seven resolved species, 8218(92.5%) got sequence type (ST) assignments, encompassing 1,252 distinct species -ST combinations. E. hormaechei accounted for the largest proportion of typed genomes (5,628;68.5%) and exhibited 521 distinct STs, followed by E. cloacae (662 typed genomes ,212 STs), and E. roggenkampii(580 typed genomes ,147 STs). Within E. hormaechei the most frequent Sts were ST78(441 genomes), ST171 (348), ST66 (278), ST90(238) and ST114(223). Collectively these findings establish a heterogenous global ECC genomic collection characterized by pronounced by sampling inequalities, extensive taxonomic and sequence type diversity, and predominance of human associated E. hormaechei genomes.The species resolved sequence type assignments and corresponding lineage distributions are provided in **Supplementary 2.**

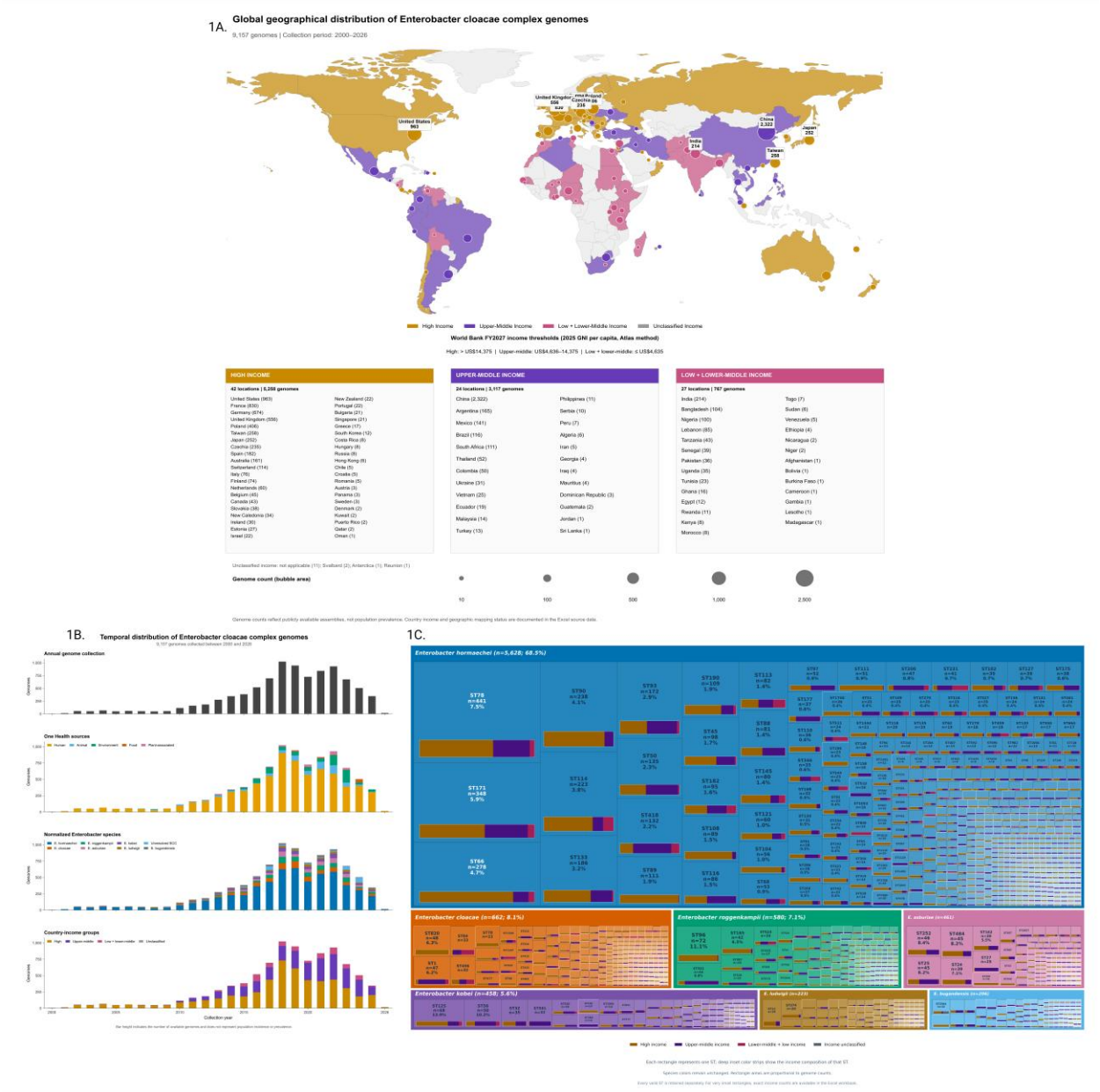


Fig. 1 | Global geographical distribution, temporal representation and multi-locus sequence type (MLST) diversity of Enterobacter cloacae complex: *Created in BioRender. https://BioRender.com/iwz87o2*

a. Global geographical distribution of 9,157 quality filtered non redundant E. cloacae complex genome collected between 2000-2026.Countries and territories are colored according to World Bank FY 2027 income classification based on 2025 gross national income per capita (Atlas method): high income (gold), upper middle income (purple) and low and lower middle income (pink). Locations without applicable income colored in grey. Proportional circle area corresponding to genome count. Country-level contributions are presented separately for high-income (42 locations, 5,258 genomes), upper middle income

(24 locations, 3,117 genomes) and low and lower middle income (27 locations, 767 genomes) groups. Fifteen genomes belong to unclassified income categories.

b. Temporal distribution of 9157 ECC genomes according to reported collection year. From top to bottom, histograms show annual genome counts, genome counts stratified by One Health isolation source, taxonomically normalized ECC species, country -income classification Isolation sources comprise human (gold), animal (light blue), environment (green) food associated (orange) and plant associated (pink) categories. Taxonomic groups are represented by E. hormaechei, E. cloacae, E. roggenkampii, E. asburiae, E. kobei, E. ludwigii, E bugandensis and unresolved ECC genomes. Income groups are represented using same colours in fig 1a. stacked bar heights represents absolute numbers of genomes collected in each year.

c. Hierarchical tree-map illustrating the distribution of MLSTs among the seven species resolved ECC groups. The analysis included 8,218 STs assigned genomes and 1,252 distinct species St combinations. Large colored regions distinguish ECC species, with their areas proportional to number of ST assigned genomes belonging to each species. Within each species, individual rectangles represent distinct STs. The relative contribution of high income (brown gold), upper middle income(purple), low and lower middle income (pink) and income unclassified (Grey) within each STs indicated by inset proportional color strips. Genomes lacking an assigned St and genomes without resolved species identification are excluded from tree-map **(Supplementary 2).**

2. Isolation source associated variation and geographical sampling inequalities in ECC species composition:

To investigate the ecological distribution of the ECC, we examined the species composition of 9,157 genomes across five isolation source categories (fog. 2a). Human associated isolates accounted for the majority of the collection (7,283 genomes ;79.5%), followed by environmental (853; 9.3%), animal (686;7.5%); plant associated (247; 2.7%) and food-associated (88; 1.0%) genomes. Enterobacter hormaechei predominated among human associated (67.9%), animal associated (67.2%) and food associated isolates (70.5%) but accounted for substantially smaller proportion of environmental (40.7%) and plant associated (23.9%) genomes. Environmental isolates exhibited a more heterogenous taxonomic composition, including increased representation of E. roggenkampii (18.8%), E. asburiae (11.4%), E. kobei (10.9%) and E. cloacae (10.4%) . Notably, E. ludwigii represented 27.9% of plant associated genomes, compared with 2.0% human associated isolates. Pearson's chi-square test identified an association between isolation source and species composition ($\chi 2$ =1,139.18 degree of freedom=28, $P< 0.001$), although the effect size was modest (Cramer's V = 0.176). These findings indicate that the taxonomic representation of ECC

varies across samples ecological compartments, with greater representation of non hormaechei species in environmental and plant- associated collections.

We next assessed whether ECC species composition differed among geographical groups stratified according to World Bank income classification (fig. 2b). The collection comprised 5,258 genomes from high income, 3,117 genomes from upper middle-income countries and 767 from low and lower middle-income countries; additional 15 genomes were assigned to unclassified. E. hormaechei remained the dominant species across all three classified groups, representing 65.7%, 61.4% and 65.7% genomes respectively. However, the relative representation of other ECC species differed between groups E. cloacae accounted for 17.7% of genomes from low- and lower middle-income countries, compared with 7.1% and 7.9% in high and upper middle-income countries, respectively. Conversely E. kobei was more frequent among upper middle-income genomes (8.2%). Species composition was associated with income classification in the genome level contingency analysis ($\chi 2$ =355.15, df=14, $P < 0.001$, Cramer's V= 0.139). Nevertheless, geographical sampling was highly uneven with China contributing 74.5% of genomes assigned to upper- middle income category. Thus, the observed differences may partly reflect the disproportionate representation of individual countries rather than consistent income associated patterns.

To account for differences in geographical sampling depth, we evaluated country level taxonomic diversity after restricting the analysis to countries represented by at least 20 species-resolved genomes (fig. 2c). This creation retained 42 countries, comprising 24 high income, 9 upper middle income and 9 lower middle-income countries. Species richness was standardized to common sampling depth of 20 genomes per country using expected rarefied richness. Median rarefied richness was 4.09, 4.49 and 3.90 species in the respective income groups, with no significant difference between groups (Kruskal-Wallis, P =0.675). Observed Shannon diversity similarly showed comparable median values of 1.06,1.07,0.99 respectively (Kruskal-Wallis test, P= 0.719). Although country level diversity varied considerably within income categories, this analysis did not identify systematic differences in ECC species diversity associated with country- income classification. We further examined the potential influence of unequal country contribution by comparing genome weighted and equal country weighted estimates of species composition across the same 42 eligible countries (fig.2d). In the genome – weighted analysis, E. hormaechei represented 65.9%, 61.3% and 67.5% of genomes from high-, upper middle-, and low- and lower-income groups respectively. When each country contributed equally to the estimated species composition, these proportions were 65.8%, 66.1%,67.9%. The largest change occurred in upper middle-income group where estimated contribution of E. hormaechei increased by 4.8 percentage points following equal-country weighting. A country level multivariate permutation test did not identify a significant association between income classification and species composition (pseudo-F= 0.372, P=0.827, 4,999 permutations). Together, these analyses demonstrate pronounced isolation -source- associated variation in samples ECC population, while indicating that apparent

income-stratified taxonomic differences are sensitive to geographical sampling structure and are not consistently supported by country level comparisons**. (Supplementary 3-6)**

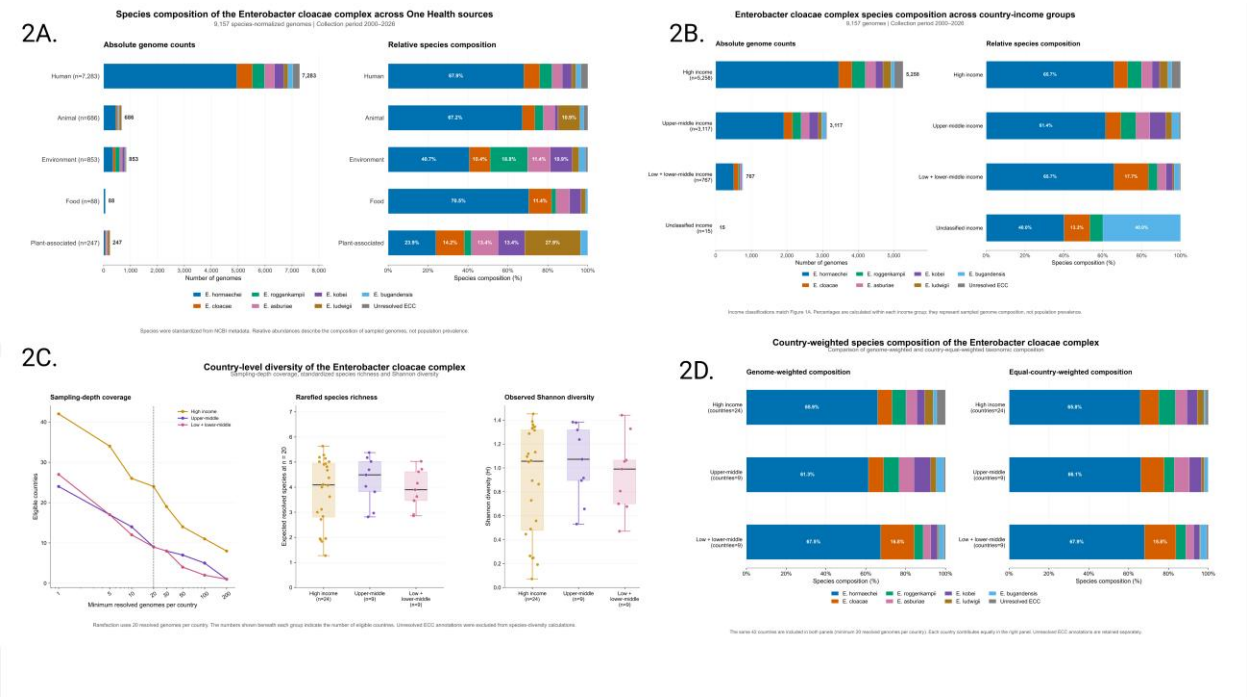


Fig. 2 | Species composition and country level diversity of the ECC across One health sources and country income groups. ***Created in BioRender. (2026) https://BioRender.com/666v2wm***

a. Absolute and relative species composition of 9,157 species-normalized Enterobacter cloacae complex (ECC) genomes stratified by reported isolation source. Horizontal stacked bars represent genome counts (left) and within-source species percentages (right) across human (n=7283), animal(n=686), environmental(n=853), food associated (n=88) and plant-associated (n=247) isolates. Species assignments include E. hormaechei, E. cloacae, E. roggenkampii, E. asburiae, E. kobei, E. ludwigii, E. bugandensis and unresolved ECC genomes. The association between isolation source and species composition was evaluated using Pearson's chi-square test ($\chi2$ =1,139.18 degree of freedom=28, P< 0.001, Cramer's V= 0.176). **(Supplementary 3)**

b. Absolute genome counts (left) and relative genome counts (right) stratified by World Bank country income classification. Income categories comprise high income (n= 5,258), upper-middle income (n= 3,117), low and lower middle income (n= 767) and income unclassified (n=15) genomes. Species percentages were calculated relative to the total number of genomes within each income category. Differences in species composition among the three classified income groups were evaluated using Pearson's chi-square test ($\chi2$ =355.15, d.f. =14, P< 0.001, Cramer's V= 0.139). (**Supplementary 4)**

c. Country -level species diversity across income groups. The left panel shows the number of eligible countries at increasing minimum sampling depths, with the dashed vertical line

indicating the primary inclusion threshold 20 species- resolved genomes per country, calculated under sampling without replacement. The right panel represents observed Shannon diversity (H), calculated from the original species composition of each eligible countries without rarefaction. Each point represents one country or territory; boxplots represent the distribution within income groups. Country level differences assessed using Kruskal Wallis tests (rarefied richness: P= 0.675; Shannon diversity P =0.719) (**Supplementary 5)**.

d. Composition of genome -weighted and unequal country weighted ECC species composition among 42 countries meeting the minimum sampling depth criteria. Left panel represents species composition calculated by pooling genomes across eligible countries within each income group. The right panel represents the arithmetic mean of within country species proportions such that each eligible country contributes equally regardless of it's genome count. Species categories are distinguished using consistent colors across panels. Country level differences in resolved species composition were assessed using a squared-Euclidean-distance-based pseudo-F statistic with 4,999 permutations of income- group labels (pseudo-F =0.372, P = 0.827) **(Supplementary 6)**

All panels describe the composition of publicly available ECC genome assemblies. The genome-level chi-squared tests should be interpreted as exploratory associations because individual genome may be clustered by country, sequencing project or lineage. Taxonomic proportions represent the composition of the samples genomic collection rather than population-level species prevalence.

3.Temporal heterogeneity and income-stratified resistome architecture across the Enterobacter cloacae complex

To investigate temporal changes in antimicrobial resistance gene burden across the Enterobacter cloacae complex (ECC) we quantified the number of distinct annotated resistant genes per genome across six collection periods spanning 2000-2026. Analysis was stratified by seven ECC species and three country income categories, encompassing 8,867 genomes with resolved species and income classifications (fig. 3a-g). Significant temporal heterogeneity was detected in seven of the 20 testable species-income combinations after Benjamini-Hochberg correction, with the most pronounced differences observed in Enterobacter hormaechei.

Within E. hormaechei, temporal variation was significant across high income (Kruskal-Wallis H= 124.98, $q=5.52 \times 10^{-24}$), upper middle income (H= 78.07, $q= 2.13 \times 10^{-14}$) and low and lower middle income (H=50.48, $q= 4.22 \times 10^{-10}$) groups (Fig. 3a). Among high income genomes, mean resistance-gene burden increases from 9.88 per genome in 2005-2009 to 13.67 in 2010-2014

(Holm-adjusted P= 4.71 x 10^{-11}), remained relatively stable through 2020-2024(13.63), and subsequently declined to 10.46 in 2025-26 (adjusted P =1.02 x 10^{-6}). Upper middle-income E. hormaechei genomes displayed a different trajectory, with mean gene burden decreasing from 14.32 in 2015-2019 to 12.31 in 2020-2024 and 7.69 in 2025-2026(adj P= 1.83 x 10^{-7}) and 4.10 x 10^{-7} respectively). In low- and lower middle-income settings, burden increased from 8.46 in 2010-2014 to 15.12 in 2015-2019 before declining to 11.26 in 2020-2024 and 7.17 in 2025-26. These contrasting trajectories indicate substantial temporal heterogeneity within sampled E. hormaechei population rather than a consistent directional increase in resistance – gene burden.

Other ECC species exhibited more restricted patterns of temporal variation. In E. roggenkampii , significant differences were detected among genomes from high- income (q= 1.81 x 10^{-6}) and upper middle income (q=0.0013) settings (fig. 3c).High income isolates showed an increase in mean resistance gene burden from 5.72 in 2015-2019 to 8.83 in 2020-24 (Holm adj P= 6.74 x 10^{-9}), whereas upper middle income isolates increased from 5.38 to 6.63 over the same periods (adj P= 0.00295), followed by a higher mean of 14.80 among ten genomes sampled in 2025-26.Temporal variation was also detected in high income E. ludwigii (q= 1.69 x 10^{-5} , fig. 3d) and upper middle income E. kobei (q = 0.00862; Fig . 3g). In the latter, mean gene burden decreased from 20.40 in 2000-2004 to 6.4 in 2025-26, although the earliest period comprised only five genomes. By contrast E. cloacae, E. bugandensis and E. asburiae showed no significant omnibus temporal differences after correction within any income category (fig. 3b, e, f). **(Supplementary 7)**

We next examined the income stratified distribution of resistance determinants using a genome-weighted network connecting country-income groups, resistance classes and individual genes (fig 3h). Across the 8,867 eligible genomes the network presented 601 distinct annotated gene identities, including 362 assigned to the beta-lactam class. Beta-lactam associated genes were detected in virtually all genomes across income categories (99.98-100%). While quinolone associated determinants (95.5-96.5%) and Fosfomycin associated determinants (89.1%-90.6%) were also highly represented. Aminoglycoside-associated genes occurred in 58.8%, 51.7% and 54.6% of genomes from high-, upper middle- low and lower middle-income categories respectively. Sulfonamide-associated genes exhibited corresponding prevalences of 57.1% ,49.6% and 50.9% whereas tetracycline associated determinants were detected in 30.3%,33.6% and 35.7%.

Additional heterogeneity was apparent for less frequently represented resistance classes. Colistin associated mcr determinants occurred in 2.2% of high income, 5.5% of upper middle income and 3.9% of low- and lower income genomes. Bleomycin associated annotations were proportionally more frequent in upper middle income (28.8%) and low- and lower middle-income categories (32.3%) than higher income genomes (9.96%). At the individual gene level, the network highlighted recurrent determinants including oqxA, oqxB, fosA, sul1, sul2, tet(A), mph(A) and aminoglycoside modifying enzyme genes, alongside a diverse collection of bet-lactamase alleles. These findings reveal widespread carriage of multiple resistance associated gene classes and

differences in their representation across country -income categories. However, the network summarizes genomic carriage rather than physical interactions, horizontal transmission or phenotypically conformed resistance.

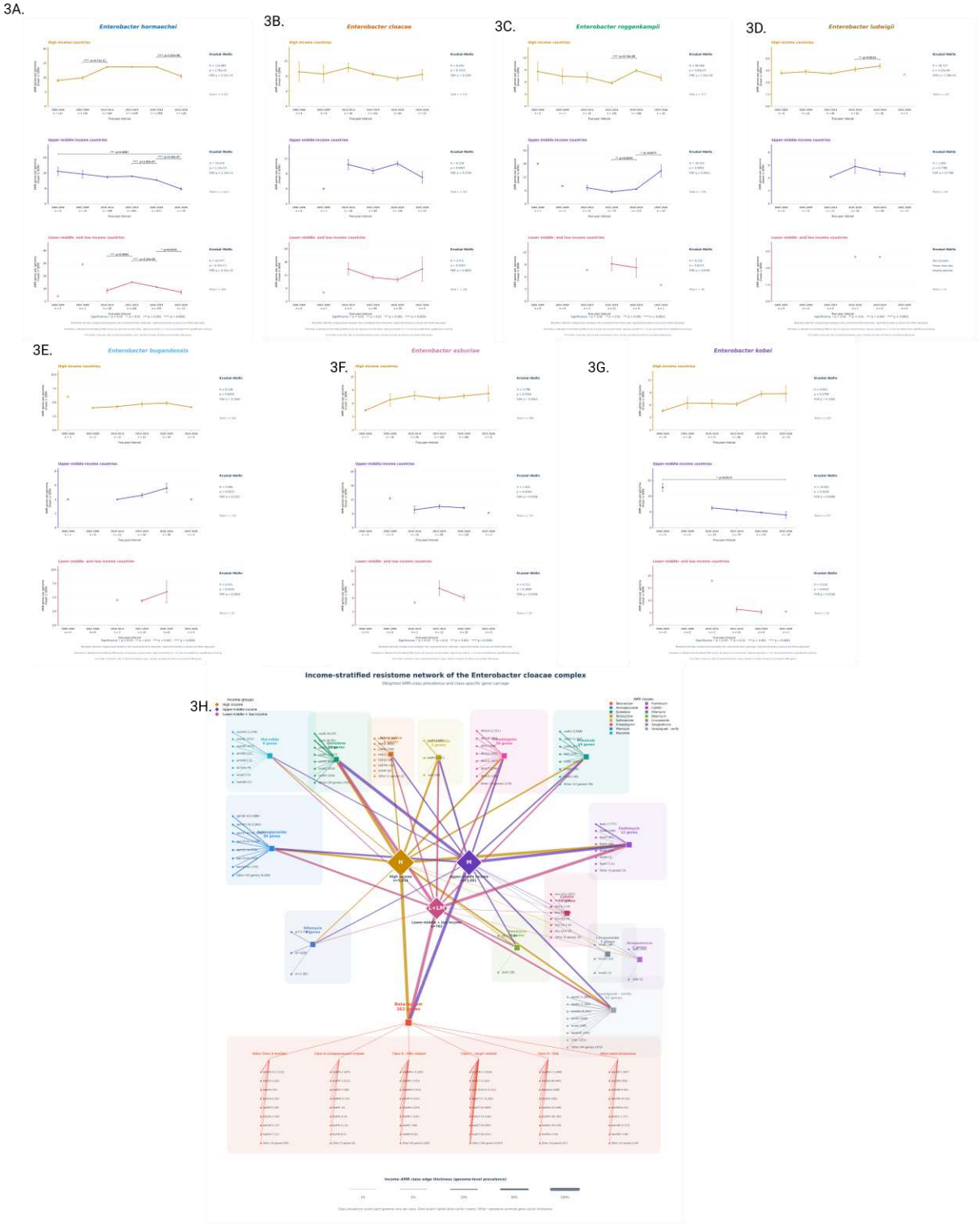


Fig. 3 | Temporal variation in antimicrobial resistance gene burden and income-stratified resistome architecture across the Enterobacter cloacae complex. ***Created in BioRender. (2026) https://BioRender.com/s8q7mgv***

a-g: Mean number of distinct annotated antimicrobial resistance genes per genome cross six collection periods (2000-2004,2005-2009,2010-2014,2015-2019,2020-2024 and 2025-2026) , stratified by country income group for E. hormaechei (a) , E. cloacae(b), E. roggenkampii (c), E. ludwigii (d), E. bugandensis(e), E. asburiae (f), E. kobei (g).Connected points show group mean ±standard errors of the mean insets report within stratum Kruskal -Wallis results. Omnibus P values were adjusted across species income comparisons using two sided Mann-Whitney U tests with Holm adjusted with strata. Significance is shown when both the adjusted omnibus and pairwise tests met $P < 0.05$. Tests required at least 5 genomes per period. **(Supplementary 7)**

h. Descriptive resistome network connecting the three income groups to antimicrobial classes and annotated genes across 8,867 genomes representing seven resolved ECC species (H, n=5014; M, n=3091; L+LM, n=762). Income to class edge widths indicate the within income percentage of genomes carrying at least one gene assigned to the class, class to gene edge widths reflect pooled gene carriage frequencies. Selected gene labels are displayed individually, with remaining genes grouped into other branches. Class level percentages count each genome once per class. Intrinsic genes are included. Network connections present annotation-based carriage, not physical genomic linkage, transmission or phenotypically verified resistance.

4. Species- and income- dependent heterogeneity in antimicrobial resistance gene burden across One health compartments

To examine variation in antimicrobial resistance gene repertoire across One Health compartments, we compared the number of distinct annotated AMR genes per genome among human, animal and environmental isolates of seven ECC species, stratified by country income classification (fig. 4a-g). The analysis included 8,532 genomes comprising 7027 human- associated, 667 animal-associated and 838 environmental isolates. Among 63 pairwise comparisons, ten remained statistically significant following Benjamini-Hochberg FDR correction, revealing species specific and geographically heterogenous associations between isolation source and AMR gene burden.

Environmental enrichment of annotated AMR gene repertoire was particularly evident in E. roggenkampii and E. kobei. Among high-income countries, environmental E. roggenkampii genomes carried a mean of 9.11 resistance genes per genome, compared with 6.83 in human associated genomes (adjusted $P = 6.63 \times 10^{-4}$, Fig. 4a). A similar difference was observed in upper middle-income countries (7.79 verses 6.27 genes per genome, adjusted $P = 8.36 \times 10^{-5}$), indicating consistent differences in distributions of annotated AMR gene burden between environmental and human associated isolates across these two

income categories. The largest observed environmental-human contrast occurred in E. kobei from high income countries, where environmental genes carried a mean of 13.26 resistance genes compared with 6.42 in human associated genomes (Cliff's $\delta = 0.612$, adjusted $P = 1.31 \times 10^{-6}$, Fig. 4c). The corresponding mean difference genes per genome was 6.83 genes per genome (95 % bootstrap confidence level ,4.35-9.21).

By contrast, E. hormaechei exhibited source associated differences that varied in direction across income groups (Fig. 4b). In high income countries animal associated genomes carried a higher mean AMR gene burden than Human – associated genomes (14.68 vs 13.09 ARGs per genome, adj $P = 3.83 \times 10^{-5}$) whereas environmental isolates exhibited lower burdens than animal associated genomes (12.20 vs 14.68, adjusted $P = 1.67 \times 10^{-6}$). However, in low- and lower- middle income countries, human associated genomes showed higher AMR gene burdens than animal associated genomes (14.09 vs 9.72 genes per genome; adj $P = 0.00491$). No corresponding source associated differences remained sufficient after correction in Upper middle-income countries highlighting the absence of uniform ecological pattern across geographical strata.

Similar geographical heterogeneity was observed in E. cloacae (Fig. 4e). Among upper-middle- income countries, animal associated genomes carried a mean of 14.38 ARGs, exceeds the values observed in human associated 9.63 ARGs per genome (adjusted P= 0.0347) and environmental genomes with 8.91 ARGs per genome (adj P =0.0427). Conversely, in low- and lower middle-income countries, human associated genomes exhibited a greater mean burden than animal associated genomes (13.01 vs 5.91; adjusted $P = 0.0197$). These contrasting associations suggest that the composition of sampled ECC populations rather than isolation source alone, influences the observed distribution of resistance associated determinants.

In E. ludwigii, high income human and animal associated genomes differed significantly in the rank-based distribution of AMR gene burden adjusted P= 0.00388, Cliff's $\delta = -0.283$, Fig. 4d), despite identical median burdens of four genes per genome. The mean difference was modest (animal vs human, -0.45 genes), with a 95% bootstrap confidence interval spanning zero (-1.26 to 0.41), indicating that significant distributional differences should not be interpreted as a statistically established in means. No source associated comparisons remained significant after multiple testing correction for E. asburiae or E. bugandensis (Fig. 4f, g). **(Supplementary 8)**

Collectively these findings demonstrate that the genomic burden of annotated AMR determinants differ among One health compartments in selected ECC species, with the direction and magnitude of these associations varying across country income groups. The greater gene repertoires observed in environmental E. roggenkampii and E. kobei underscore the importance of including non-clinical isolates in genomic resistance surveillance. Nevertheless, unequal sampling across sources and countries, together with

potential clustering by sequencing project and bacterial lineage, precludes direct attribution of these associations to ecological origin or socioeconomic conditions.

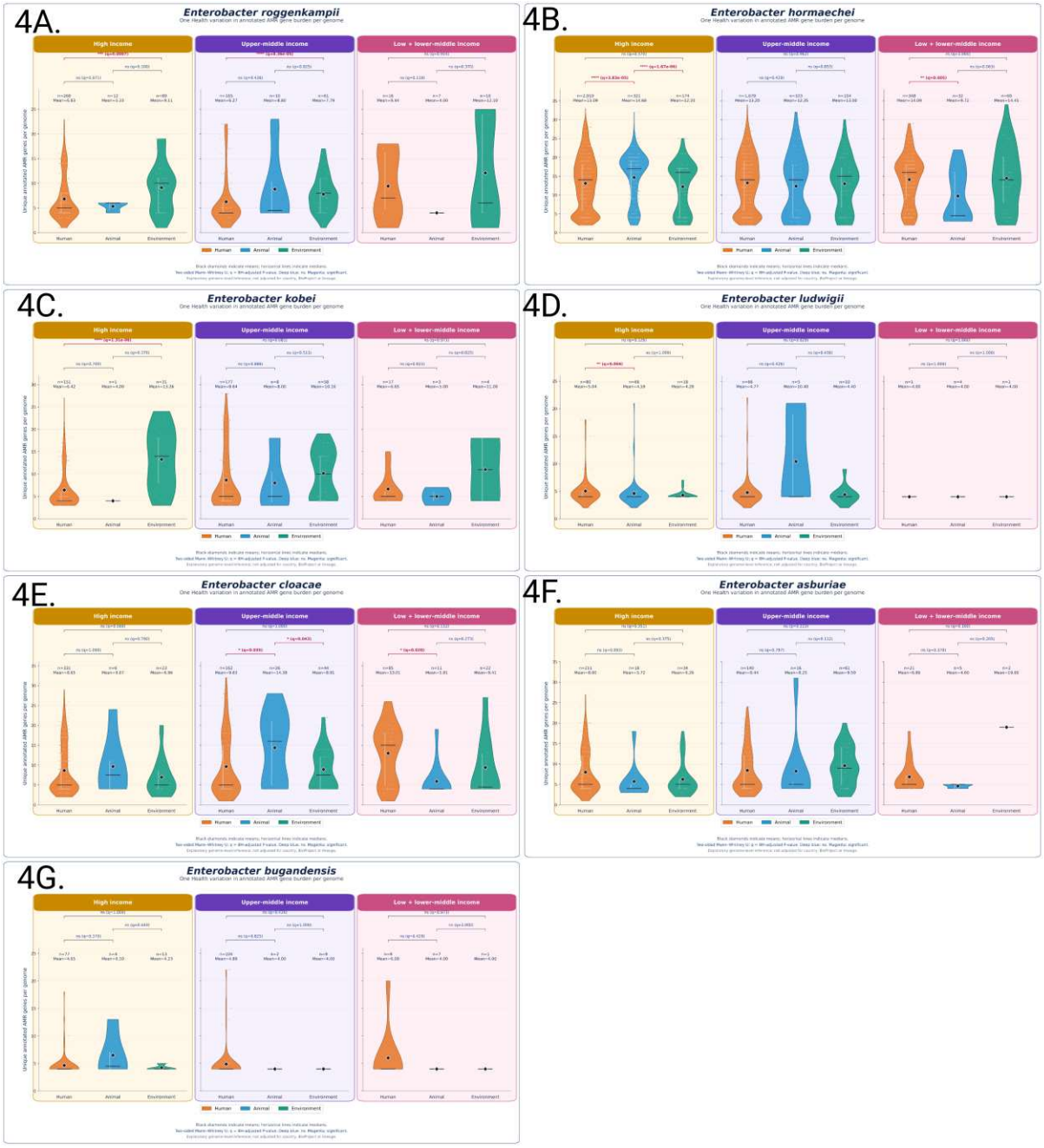


Fig 4 | Species and income -stratified variation in antimicrobial resistance gene burden across One Health sources within the Enterobacter cloacae complex. ***Created in BioRender (2026) https://BioRender.com dbd0anp***

a-g: Genome level distributions of distinct annotated antimicrobial resistance (AMR) gene counts among human, animal and environmental isolates of E. roggenkampii(a), E.

hormaechei (b), E. kobei (c), E. ludwigii (d), E. cloacae (e), E. asburiae (f), E. bugandensis(g). Analysis comprises 8,532 non-redundant genomes (human n= 7,027; animal n= 667 and environmental n= 838 genomes) subdivided into high income, upper middle income, low and lower middle-income countries. Violin widths indicate the distributions of AMR gene counts, horizontal lines denote medians and black diamonds indicate arithmetic means. Sample sizes and group means are displayed above each distribution. Pairwise comparisons were performed using two-sided Mann -Whitney U tests, with Benjamini Hochberg FDR correction across 63 comparisons. Significance brackets indicate adjusted P (q). Effect sizes were calculated using Cliff's delta and 95% bootstrap confidence intervals were calculated for differences in group means. **(Supplementary 8)**

5. Income stratified One Health analysis reveals species specific heterogeneity in the distribution of antimicrobial resistance determinants

To investigate how socioeconomic context and ecological origin shape the distribution of antimicrobial resistance determinants across the Enterobacter cloacae complex (ECC), we examined individual resistance gene prevalence across human, animal and environmental compartments stratified by World Bank country income classification (Fig. 5a-g) .The analysis included 8,532 genomes representing seven ECC species and revealed marked heterogeneity in the distribution of annotated resistance determinants . E. hormaechei exhibited the greatest annotated gene diversity (408 distinct gene identities), followed by E. cloacae (248), E. kobei (189), E. asburiae (171), E. roggenkampii (168), E. ludwigii (91) and E. bugandensis (84). Across these species, the resistome was characterized by widely distributed resistance associated determinants superimposed on heterogenous repertoire of less prevalent genes, whose representation varies among species, isolation sources and income categories. Gene prevalence calculated as:

Gene prevalence (%) = (Unique genomes carrying the gene in that group/ All eligible genomes in that group) x 100

A prominent finding was wide spread distribution of multidrug efflux associated genes oqxA and oqxB and the Fosfomycin resistance associated fosA across epidemiologically distinct ECC populations. Among human associated genomes from high income countries, oqxA and oqxB were detected in 87.5% and 87.6% of E. hormechei genomes respectively, compared with 97.0% for both genes in E. cloacae and 91.4% in E. roggenkampii. Similarly high prevalence was observed across several other species-source-income combinations, including complete detection of oqxA, oqxB and fosA among upper middle income human associated E. bugandensis (n=104). The wide spread occurrence of these determinants across multiple ecological compartments indicates that substantial compartment of

annotated ECC resistome is broadly distributed rather than restricted to a particular isolation source or national income category.

In contrast, clinically relevant acquired beta lactam is determinants displayed pronounced associated differences within selected species. Among human associated E hormaechei Genomes, blaCTX- M15 was detected in 56% of isolates from low and lower middle-income countries, Compared with 22% and 17.4% From high end upper middle-income countries respectively (Fig. 5e). A similar pattern was observed for blaOXA-1(42.8, 28.9% and 23.1% respectively), whereas blaTEM-1 was more evenly distributed (48% ,45.1% and 40.3%). This pattern was also apparent in E. cloacae, where blaCTX-M-15 occurred in 49.4% of human associated genomes from low and lower middle-income countries, respectively (Fig. 5f). The corresponding prevalence of blaTEM-1 Was 48.2%, 18.4% and 25.9%. Thus, the greater representation of selected extended spectrum beta lactamase associated genes in lower income strata was more than one major ECC species, Suggesting a potentially Important socio-economic pattern in the sampled clinical register. Nevertheless, these descriptive differences require adjustment for country level sampling, collection. And population structure before can be interpreted as independent income associated effects.

Further species level comparisons revealed that the income associated distribution of acquired resistance determinant was not uniform throughout the ECC. In human associated E. roggenkampii, blaCTX-M-15 prevalence increased from 0.4% in high income countries to 4.8% in upper middle-income countries and 31.2% in low and lower middle-income countries (Fig 5a). However, the latter category contained only 16 genomes, limiting the precision and generalizability of this observation. By contrast, blaCTX-M15 remained comparatively uncommon in human associated E. kobei, With prevalences of 1.3% and 1.1% in high and upper middle-income countries respectively (Fig. 5b). This contrasting indicates that the Apparent relations between national income classification and resistance gene carriage is strongly dependent on bacterial species and cannot be generalized uniformly across the ECC. The eight maps additionally revealed heterogeneous accessory resistance gene profiles among human animal and environmental genomes within the same species and income categories. Numerous determinants exhibited highly restricted distributions, whereas others were represented across multiple ecological compartments. This pattern is consistent with a combination of widely distributed resistant associated gene repertoires and more variable accessory determinants among sampled ECC populations. Importantly, the detection of shared across ecological compartments does not establish recent horizontal transfer or transmission between reservoirs.

Additional differences involved determinants across with polymyxin and tetracycline resistance. The mobile colistin resistance associated allele mcr-10.1 Was detected 38.6% of environmental E. cloacae genomes from upper middle-income countries, compared with 3.1% of corresponding human associated genomes (fig. 5f). The tigecycline resistance

Associated gene tet(X4) was observed in 8.7% of animal associated E. hormaechei genomes from upper middle-income countries, compared with 0.06% of human associated genomes. All nine tet(X4) Positive animal associated genomes in this stratum originated from China. Importantly, mcr-10.1 Carriage cannot be equated directly with phenotypic colistin resistance, and the detection of tet(X4) Requires corresponding functional or susceptibility evidence to establish its phenotypic consequences.

At the genome level, 66/8532 Genomes simultaneously carried at least one major carbapenemase family annotation and an mcr determinant. Such co-carriage was most frequently observed among upper middle income environmental genomes (33/375;8.8%), compared with human associated genomes from the same income group (23/2493;0.9%). However ,32 of 33 Environmental Co carriage records originated from China, with 29 represented by two bioprojects. This connection that the observed combination may reflect locally sampled genomic populations rather than a generalized income associated phenomenon Furthermore genome level co-carriage does not establish physical linkage between taste and determinants.

Collectively, these findings identify clinically differences in the distribution of carbapenemases and last line antimicrobial resistance Associated genes across ECC species, one health compartments and income strata. In particular, the repeated occurrence of blaNDM-1 across Several species in upper middle income human associated collections, together with concentrated environmental blaKPC-2 and mcr Associated determinants, highlights the need for lineage resort and geographically balanced genomic comparisons. These findings provide candidate patterns for further investigation but do not independently demonstrate resistance phenotypes, inter reservoir transmission or causal effect of national income.

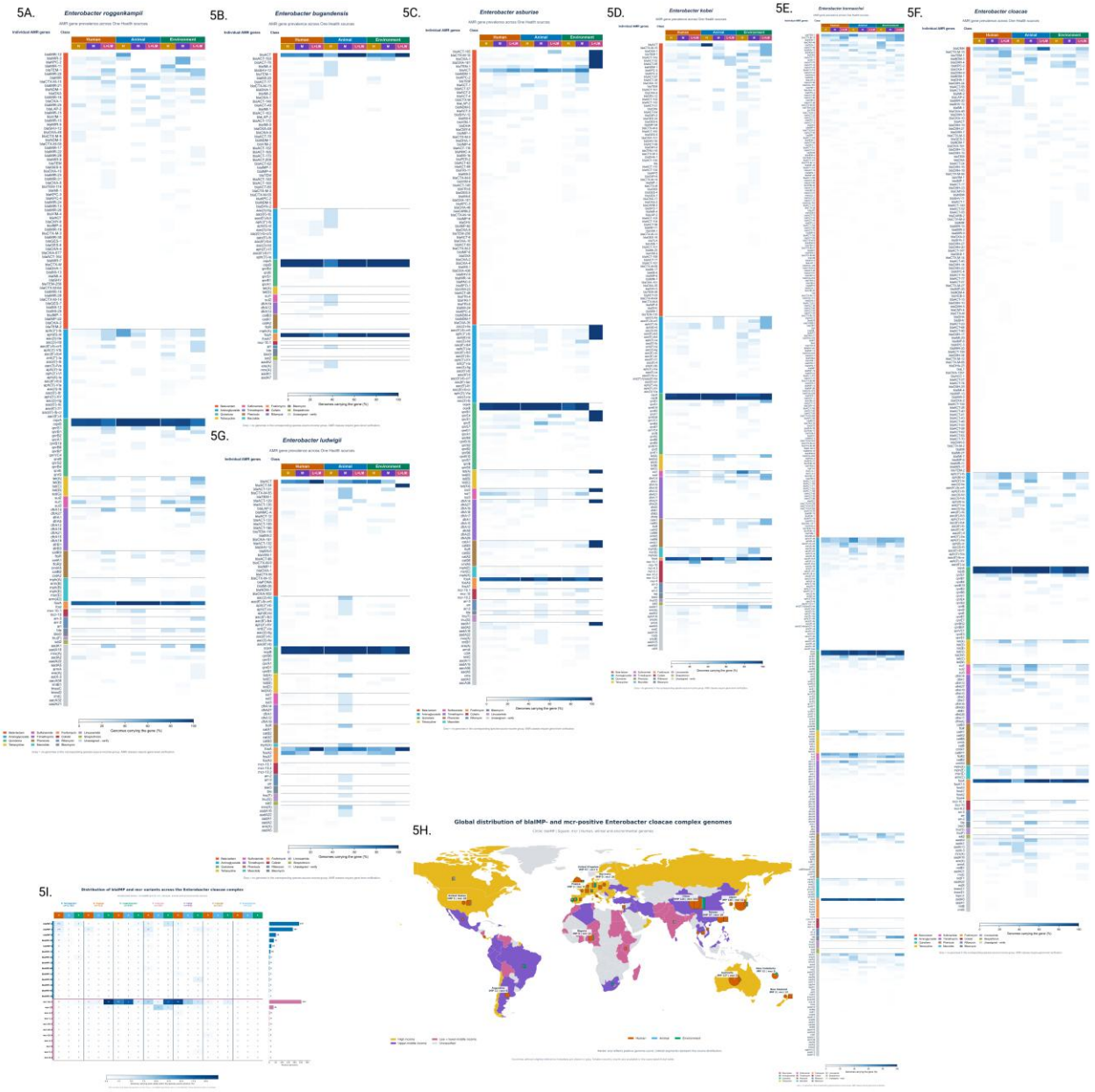


Fig. 5 | One Health and income stratified distributions of antimicrobial resistance determinants affecting critical treatment options in Enterobacter cloacae complex. ***Created in BioRender. https://BioRender.com/y8zynev***

a-i, Heatmaps showing genome level prevalence of individual antimicrobial resistance associated genes in E. roggenkampii (n=638), E. bugandensis(n=226), E. asburiae (n=508), E. kobei (n=450), E. hormaechei(n=5,749), E. cloacae(n=710) and E. ludwigii (n= 251). The analysis includes 8,532 genomes stratified by human animal and environmental origins and by high(H) income, upper middle income (M)and lower and lower income (L+ LM) countries**. (Supplementary 9)**

Rows represent individual annotated resistance, antimicrobial class and columns represent source income. Cell color intensity indicates the percentage of genomes carrying each determinant within the corresponding species source income stratum (white, 0%; dark blue, 100 %). Gray represents strata without eligible. Colored row annotations indicate antimicrobial classes. Particular emphasis is placed on carbapenemase determinants (*blaNDM, blaKPC, blaIMP, blaVIM and blaOXA-48*- like), Mobile colistin resistance associated *mcr* and tigecycline resistance associated *tet(X)* genes. Prevalence was calculated using unique gene positive genomes divided by the number of eligible genomes in each stratum. All detected genes, including low frequency variants, are retained in the heatmaps. Further special attention is given on 14 *blaIMP* variants and 11 *mcr* variants prevalence (Fig. 5I) across One health compartments and distribution in global geography across different countries (Fig. 5H). **(Supplementary 11)**

These heatmaps describe annotation-based gene carriage without inferential testing. They do not independently establish phenotypic resistance, physical linkage of resistance loci or transmission between One Health reservoirs.

6. Temporal and socioeconomic heterogeneity in plasmid replicon diversity across the Enterobacter cloacae complex

To investigate temporal changes in the plasmid replicon repertoire across the Enterobacter cloacae complex (ECC), we examined genome-level replicon diversity across seven ECC species collected between 2000-2026, stratifying isolates according to World Bank income classification of their country of origin (Fig. 6). Of 9,157 genomes included in the analysis, 7,125 fulfilled the eligibility criteria for replicon-based analysis. Replicon burden defined as the number of distinct plasmid replicon types per genome, allowing comparisons across successive collection periods and socioeconomic strata. The analysis revealed substantial species-specific and income dependent heterogeneity in temporal replicon burden, indicating that changes in plasmid replicon diversity were not uniform across the ECC.

The most pronounced temporal differences were observed in E. hormaechei, particularly among genomes originating from high-income countries (fig. 6A,6B). Mean replicon burden increased from 2.61 replicon types per genomes during 2000-2004 to 3.02 during 2010-2014, reaching 3.33 during 2015-2019 before declining to 3.14 during 2020-2024 and 2.61 during 2025-2026. These differences were statistically significant (Kruskal Wallis $H = 87.99$, BH adjusted $q = 3.37 \times 10^{-16}$). Pairwise comparisons identified a significant increase between 2010-2014 and 2015-2019(Holm-adjusted $P = 9.91 \times 10^{-6}$), followed by a significant decline between 2020-2024 and 2025-2026(Holm adj $P = 4.14 \times 10^{-5}$). Temporal variation was also significant in low- and lower-middle-income group did not exhibit a significant overall temporal difference after multiple testing correction ($q= 0.333$; $n=1,624$). These observations indicate that the temporal trajectory of replicon burden in E. hormaechei differed substantially across income groups.

Significant temporal variation was also identified in E.cloacae and E.roggenkampii originated from high income countries (Fig. 6C, D, M, N). In E. cloacae, mean, replicon burden increased from 2.86 during 2000-2004 to 3.24 during 2015-2019, subsequently declining to 1.86 during 2025-2026. The overall temporal difference was significant (H= 18.75, q= 0.00812; n=217). Similarly, E. roggenkampii showed significant temporal heterogeneity (H= 23.69, q = 8.77 x 10^{-4}, n=312), with mean replicon burden increasing from 2.65 during 2015-2019 to 3.49 during 2020-2024 and 3.79 during 2025-2026. By contrast, upper -middle income country comparisons for both species did not remain significant flowing correction for multiple testing (E. cloacae, q= 0.115, E. roggenkampii, q= 0.0757).

The remaining ECC species exhibited more heterogenous temporal patterns (Fig. 6E-6L) . Significant differences were detected among upper middle income country genomes of E. bugandensis (H=9.99, q= 0.0214, n=32) and E. asburiae (H= 12.84, q=0.00774, n=153)> In contrast, no significant temporal differences were detected for E. kobei or E. ludwigii in the evaluated income strata after correction for multiple comparisons. The absence of significant findings in some strata should be interpretated in the context of limited sample availability, particularly among low and lower middle-income countries less frequently represented ECC species.

Complementary replicon-specific prevalence heatmaps revealed heterogenous distributions of individual plasmid replicon types across successive collection periods (Fig. 6B,D,F,H,J,L,N) . These analyses demonstrate that changes in overall replicon burden are accompanied by variation in the prevalence of individual replicon types, rather than representing a uniform increase or decrease across the entire replicon repertoire. Collectively, the findings reveal distinct temporal and socioeconomic patterns of plasmid replicon diversity among ECC species, while emphasizing the importance of accounting for unequal geographical representation and temporal sampling when interpreting global plasmid epidemiology.

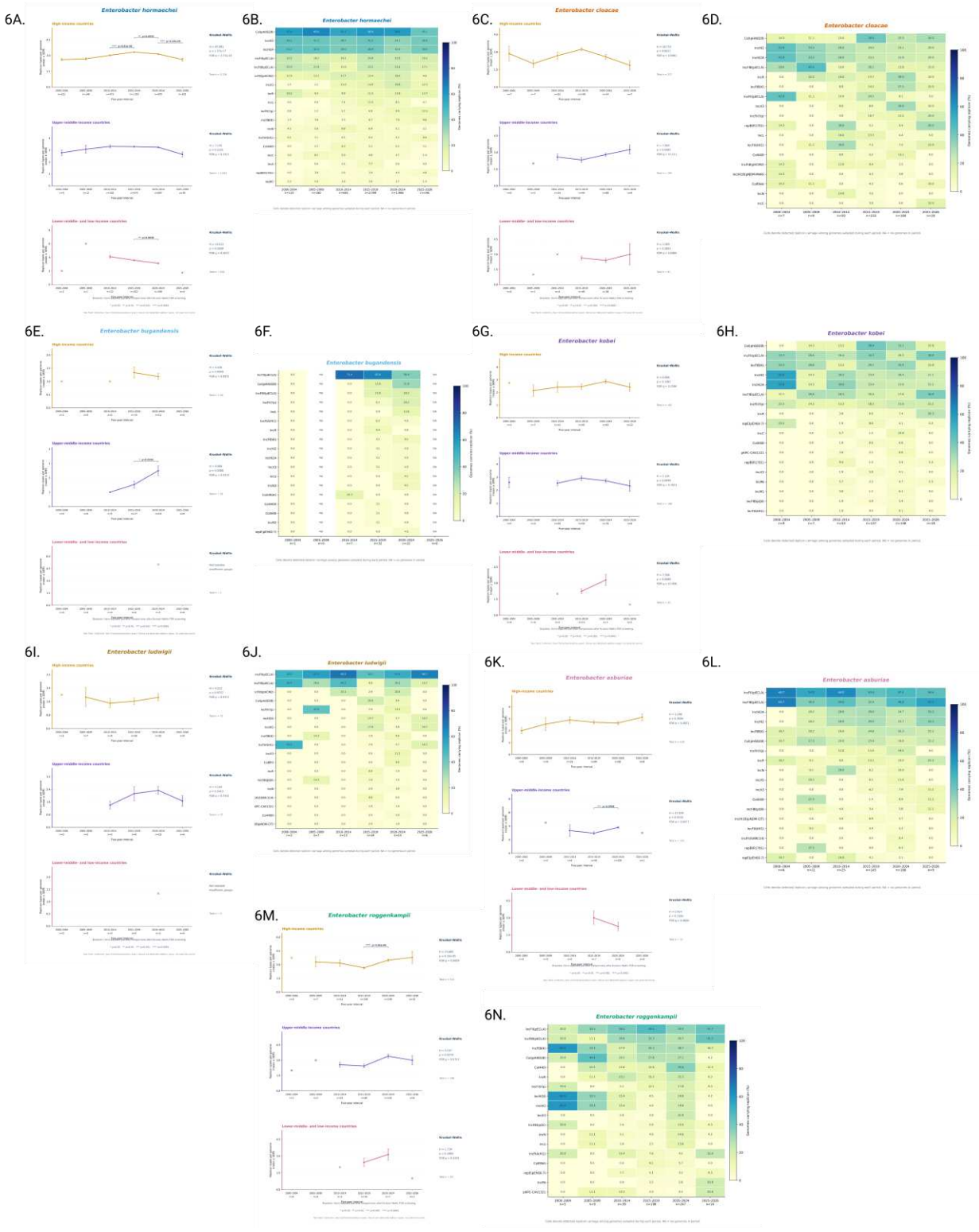


Fig. 6 | Temporal and income stratified heterogeneity in plasmid replicon burden and replicon type prevalence across the Enterobacter cloacae complex*Created in BioRender. (2026) https://BioRender.com/142hzxg*

Temporal variation in the number of distinct detected plasmid replicon types per genomes and prevalence of individual replicon types across seven Enterobacter cloacae complex species from 2000-2026.Analyses were conducted using 7,125 eligible genomes with available replicon annotation data.(A, B) Enterobacter hormaechei;(6C, D) E. cloacae;(6E, F) E. bugandensis ;(6G, H) E. kobei;(6I, J) E. ludwigii (6K, L), E. asburiae; and (6M, N) E. roggenkampii.(6A, C, E, G, I, K, M) Temporal trends in plasmid replicon burden, stratified according to the World Bank income classification of the isolate's country of origin: high-income (golden amber), upper -middle-income (violet), and combined low and lower middle income (rose pink). Replicon burden was defined as the number of distinct plasmid replicon types detected per genome rather than complete plasmids. Collection years grouped in six periods. Temporal heterogeneity within each species-income stratum was evaluated using Kruskal Wallis test with Benjamini Hochberg FDR correction across 21 species-income comparisons. Pairwise differences between collection periods were assessed using two-sided Mann Whitney U tests with Holm adjustment within each species income stratum. (6B, D, F, H, J, L, N) Replicon specific prevalence Heatmaps illustrating temporal differences in the detection frequencies of individual plasmid replicon types within each ECC species. Cell values represent the percentage of eligible genomes carrying the indicated replicon type within the corresponding species, income category and collection period grouping.

7. Species-specific plasmid replicon-AMR gene co-occurrence networks across the Enterobacter cloacae complex

Species-resolved spherical networks showing associations between plasmid replicon types and AMR genes across seven major Enterobacter cloacae complex species.The central node in each network represents the focal species, circular nodes denote plasmid replicon types, and square nodes denote AMR genes. Replicon nodes and their outgoing edges share the same colour enabling vidualization of replicon specific AMR association patterns. Outer ring segments around replicon nodes indicate the source composition classified as human , animal environment. AMR gene nodes are coloured by resistance class. Edge indicate observed genome level co-occurance between a plasmid replicon and AMR gene within the same genome.(fig. 7)

The networks highlight marked interspecies differences in plasmid repertoire complexity and replicon-AMR association structurewith particularly extensive association landscapes observed in E. hormaechei and E. cloacae . Several AMR genes are shared across multiple species associated plasmid backgrounds, whereas other associations are more species restricted.Overall, these data indicate that plasmid diversity is a major contributor to resistome organization across the Enterobacter cloacae complex. A total 8,532 genoems, representing 7 ECC species, 101 plasmid replicon types, 598 AMR genes and 18,861 observed co-occurrence associations were included.These associations represent genome

638 level co-detection and do not on their own establish physical localization of AMR genes
639 on that specific plasmids.

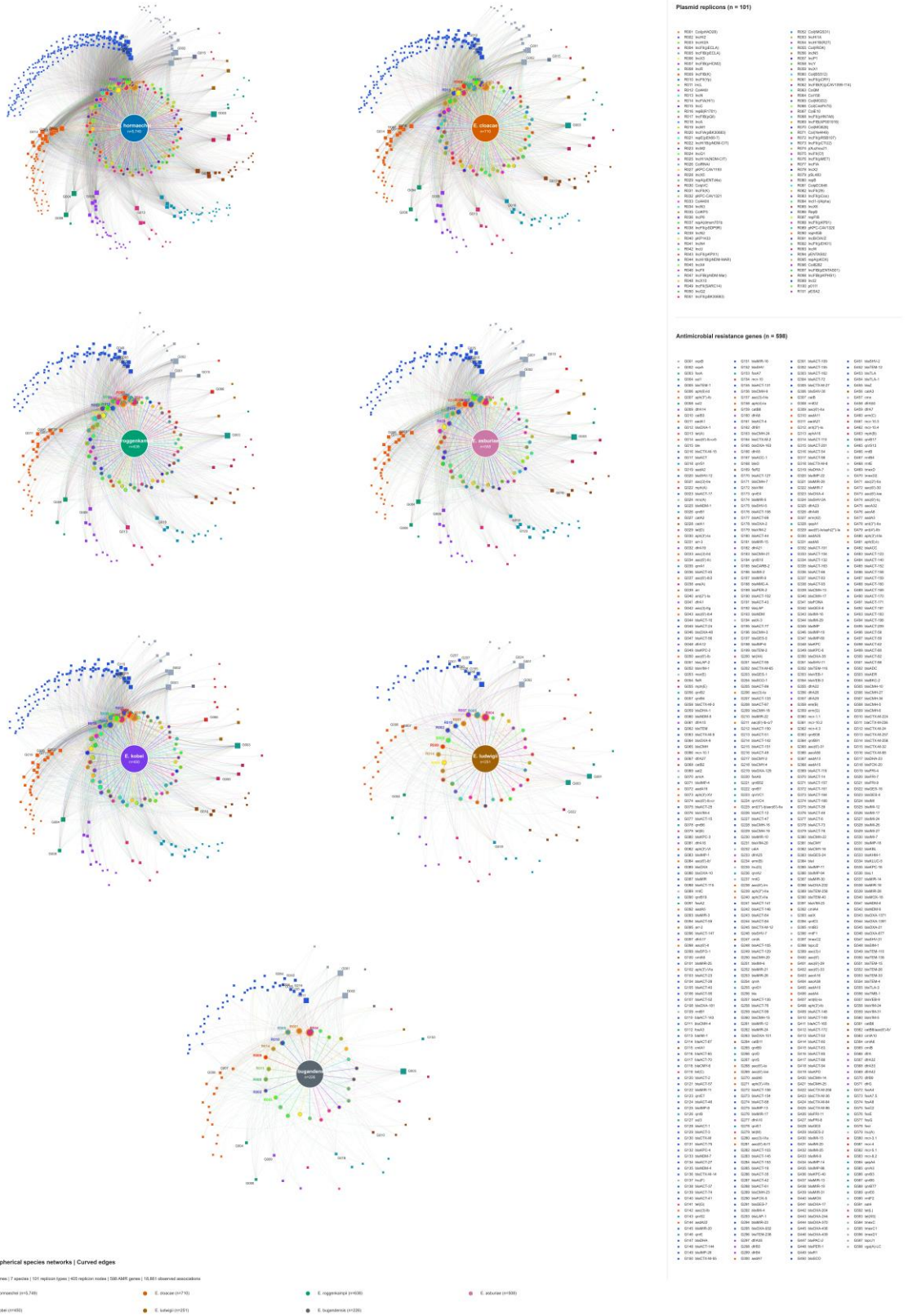

Plasmid replicons (n = 101)
Antimicrobial resistance genes (n = 598)
Seven spherical species networks | Curved edges
8,532 genomes | 7 species | 101 replicon types | 403 replicon nodes | 598 AMR genes | 16,661 observed associations
E. hormaechei (n=5,748)
E. cloacae (n=710)
E. roggenkampii (n=636)
E. asburiae (n=598)
E. kobei (n=450)
E. ludwigii (n=251)
E. bugandensis (n=226)

Fig. 7| Species-specific plasmid replicon-AMR gene co-occurance networks across the Enterobacter cloacae complex(ECC)

Spherical networks showing associations between plasmid replicon types and AMR genes across seven major ECC species . Central nodes denote species , circular nodes denote plasmid replicons, the square nodes denote AMR genes. Replicon nodes and their connecting edges share same colour and replicon ring segment indicate the source composition of positive genomes(Human, Animal,Environment). AMR genes are coloured by resistance class. Across 8,532genomes, the analysis identified 101 replicon types,598 AMR genes and 18,861 observed co-occurance associations **(Supplymentary 10)**. The networks reveal pronounced interspecies differences in plasmid repertoire complexity and in replicon-AMR association structure , with essentially extensive networks in E. hrmaechei and E. cloacae .Associations represent genome-level co-occurance and do not by confirm physical localization of AMR genes.

Discussion:

This global analysis of the Enterobacter cloacae complex (ECC) Identifies an important limitation of large-scale AMR surveillance and the apparent strength of socio-economic associations depend on how unevenly sampled genomes aggregated. Across 9,157 curated assemblies from 97 countries, we found substantial differences in the species and annotated response gene repertoire represented across sources and income groups. Yet a global-level association between national income category and ECC species composition was not reproduced by country level analysis that reduced the influence of heavily sampled countries. This contrast provides the central insight of the study: more genomes do not necessarily yield more representative cross-country comparisons. Geographical coverage was broad but unbalanced. High income countries contributed most assemblies, while low and lower middle-income countries contributed less than $1/10^{th}$ of the collection. China alone accounted for nearly three quarters of upper middle-income genomes. These imbalances reflect the organization of public genome generation and deposition, not necessarily differences in the occurrence of ECC or AMR in the underlying population. Public repositories integrate outbreak investigations, targeted resistance surveys and diverse research projects; their observations cannot be interpreted as incidence estimates. The geographical inequalities reported by amr.watch **(https://amr.watch/)** provide context for this limitation[14].

Taxonomic resolution proved essential. The 8,882 species-assigned assemblies encompassed 1,252 species-sequence type combinations, with E. hormaechei predominating. Nevertheless, E roggenkampii, E. asburiae,E. kobei, and E. ludwigii contributed more prominently in selected environmental collections. The association between isolation source and species composition was statistically detectable but modest in effect size (Cramer's V = 0.176). Because sampling protocols and isolation selection

criteria differed between source categories, these proportions describe deposited collections rather than ecological prevalence or niche adaptation.

Garcia-Martin and colleagues demonstrated extended spectrum beta-lactams producing Enterobacterales across clinical, stated and food associated compartments, identifying resistance, gene diversity, and genomic overlap between selected sources[11]. Similarly, longitudinal household surveillance in Malawi revealed extensive environmental contamination and resistance associated exposure in human and animal settings[12]. Unlike those prospective investigations, our analysis uses retrospective assembled public genomes; consequently, it cannot establish direct environmental exposure pathways or movement of strains between reservoirs. Its contribution is instead to reveal which ECC species and resistance repertoires are represented across these components at a broader geographical scale.

One of the most consequential findings of this study is the divergence between genome level and country level interpretations of ECC species composition. At the genome level, the relative distribution of ECC species differed significantly among World Bank income categories. However, the corresponding effect size was modest, and the upper middle-income group was particularly influenced by the disproportionate contribution of Chinese genomes. Countries, are rarefied species richness and observed Shannon diversity did not differ significantly among income categories. Furthermore, equal country weighting altered the estimated composition of upper middle-income collection, and a country level multivariate permutation analysis did not identify a significant. Income associated difference in species composition (pseudo-F =0.372, P= 0.827).

This distinction has broader implications for studies linking bacterial genomic characteristics with socioeconomic conditions. Large scale analysis of Salmonella has investigated geographical AMR variation in relation to national economic, environmental, healthcare and agricultural indicators, illustrating the scientific value of country level contextual data[15]. However, the reliability of such analysis depends on the geographical coverage of genomic sampling the compatibility of exposure measurements and the treatment of non-independence within countries and bacterial populations. The World Bank income categories used in our study describe broad differences in national economic context but do not directly capture antibiotic consumption, infection prevention practices, healthcare, accessibility or population awareness of AMR. Accordingly, the income stratified findings presented here should be regarded as contextual genomic comparisons rather than evidence that economic status independently determines ECC, species composition or resistance gene carriage.

The temporal analysis revealed substantial variation in annotated resistance, gene burden among ECC, species and income groups but did not demonstrate any foreign longitudinal increase across the complex. Seven of the twenty testable species income combinations

exhibited significant temporal heterogeneity after FDR correction for multiple comparisons, with the clearest patterns occurring in E. hormeichei. Notably, the direction and magnitude of changes differed between income categories. High income. E. hormaechei Genomes exhibited an increase in mean annotated gene burden between selected early sampling periods, whereas later collections showed stability or decreases. Different trajectories were evident in upper, middle- and lower-income strata, while other ECC species exhibited fewer significant temporal contrasts.

These findings differ from simple expectation that resistance associated gene burden should increase continuously over calendar time. However, temporal trends derived from public genome collections must be interpreted cautiously. The sampling frequency, source composition, species representation, sequencing project structured and clinical indications. Earlier periods were represented by relatively few genomes, and some species income groups contained only a small number of isolates. The final collection was also incomplete at the time of analysis. Consequently, changes in mean annotated gene counts may reflect both genuine differences among the sample bacterial populations and changes in the types of isolates selected for the sequencing. Clinical surveillance studies have documented changing antimicrobial resistance profiles over time for including differences among healthcare settings but these standardized surveillance populations are not directly equivalent to retrospective public genome collections[20].

A central finding of our one health analysis was that resistance gene burden differed among isolation sources within selected ECC species, with substantial variation across income strata. Pairwise comparisons remained significant after FDR correction. Among high income genomes, environmental E. roggenkampii exhibited a higher mean number of annotated resistance genes than human associated isolates, with a similar environmental - human difference observed in upper middle-income collections. The most pronounced contrast occurred in E. Kobei from higher income countries, where environmental genomes carried a mean of 13.26 Annotated resistance genes compared with 6.42 in human associated genomes, corresponding to a mean difference of 6.83 genes and a substantial rank-based effect size (Cliff's $\delta = 0.612$).These observations suggest that non-clinical ECC collections can contain extensive resistance associated genomic repertoires and that surveillance focused exclusively on human clinical isolates may provide an incomplete picture of the diversity represented within the complex genomes. Studies examining extended spectrum beta-lactamase producing Enterobacterales across Human, animal and environmental sources have similarly identified clinically relevant resistance genes outside healthcare settings[21, 22].

Nevertheless, the greater number of annotated resistance genes in an environmental genome does not necessarily imply higher pathogenicity, greater clinical resistance or stronger antimicrobial selection in that environment. Some differences may arise from intrinsic species specific. Gene content, variation in types of environments of the over

representation of particular strains and sequencing studies. Phenotypic resistance measurements and more controlled sampling would be required to establish the clinical and ecological significance of these differences.

Importantly, source associated resistance gene burden did not follow a common pattern across ECC species or country income groups. In high-income segment, E. hormaechei collections, animal associated genomes carried a greater AMR gene burden than human associated genomes, whereas the opposite relationship was observed in low and lower middle-income collections. E. cloacae displayed a compatible reversal, where animal associated genomes had higher AMR burdens in the upper middle-income groups, but human associated genomes had higher burdens in the lower and lower middle-income group. In addition, some statistically significant differences in rank distributions, including the higher income E. ludwigii comparison, when not accompanied by confidence intervals supporting a clear difference in group means.

The species reserved gene level analysis revealed marked. Variation in the distribution of acquired beta lactamase determinants. Particularly notable was the higher observed carriage of blaCTX-M-15 among human associated E. hormaechei genomes from low- and lower-income countries (56.0%) than among corresponding high income (22.0%) and upper middle income (17.4%) collections. Elevated representation of this determinant was also observed in selected lower income E. cloacae and E. roggenkampii groups, although the latter contained relatively few genomes. These observations identify potentially important disparities in the sampled distribution of an established extended spectrum beta-lactamase allele across ECC species. CTX-M type extended spectrum beta-lactamases are prominent among resistance. Studies in low- and middle-income countries have demonstrated substantial carriage of blaCTX-M-15 and. Clinically important beta lactamase genes among. Gram native organisms associated with neonatal infections and intestinal colonization[23, 24]. The geographical and ecological distribution of ESBL producing enterobacterales It has also been investigated in community and environmental settings, providing evidence that these genes are not restricted to hospital associated populations[11]. Our study complements this literature by comparing individual acquired determinants across several ECC species within a common genomic surveillance framework. Nevertheless, the greater representation of blaCTX-M-15 in selected lower income strata does not establish an independent income effect because gene level prevalence was not adjusted for country, sequence type, Bio project or clinical sampling intensity. These findings should be therefore interpreted as surveillance signals.

The distribution of carbapenemase associated determinants provides a further clinically relevant dimension to the global ECC resistome. Genes belonging to the NDM, KPC, IMP, VIM and OXA-48 like families represent major acquired mechanisms of carbapenem resistance in Enterobacterales, although they are relative representation considerably across species and geographical regions. Nationwide genomic surveillance studies have

demonstrated that carbapenem is producing Enterobacterales may carry distinct resistance genes on different plasmid backgrounds, contributing to geographically heterogeneous resistance populations[7, 8]. In our collection, the occurrence of carbapenem has associated determinants across several ECC species and ecological components reinforce their relevance to genomic surveillance. However, the presence of carbapenemase genes, does not establish the location of the gene on a plasmid, the mobility of its surrounding genetic element or the epidemiological direction of its spread. Such questions require genomic context analysis and where appropriate, complete assemblies or experimental evidence.

The detection of mcr associated tet(X4) Determinants also highlights the importance of distinguishing potentially important resistance gene carriage from phenotypically confirmed resistance. The mcr-10.1 alleles were particularly represented among environmental E. cloacae genomes from upper middle-income countries, where it's reported frequency exceeded that in corresponding human associated collections. Previous investigations have identified mcr-9 and mcr-10 among interactive species and documented their occurrance across diverse geographical settings and plasmid backgrounds[10, 25]. Nevertheless, susceptibility in ECC is complicated by species independent intrinsic mechanisms and hetero-resistance, and the presence of an mcr annotation cannot be assumed to confer a uniform phenotypic effect[9]. Similarly, the occurrence of tet(X4) among animal associated E. hormaechei genomes is noteworthy because the determinant is associated with an enzymatic inactivation of Tetracycline class antimicrobials, but its detection in public assemblies does not substitute for antimicrobial susceptibility measurements. The geographical concentration of these findings is particularly informative. All nine tet(X4) positive animal associated E. hormaechei genomes in the upper middle class income stratum originated from China. This concentration illustrates how one apparently source and income associated resistance pattern may arise from a geographically localization of deposited genome. Without information on independent sampling events, host populations and genomic relatedness, the observed frequency cannot be generalized to animal associated ECC populations across all upper middle-income countries. At the same time, the finding identifies specific genomic surveillance signal that could be investigated using geographically distributed, phenotype-linked isolates.

In addition, finding was the Co-occurrence of a major carbapenemase family annotation and an *mcr* determinant within genomes included in the three-compartment analysis. These combinations were particularly represented among upper middle income, environmental genomes, with 33 positive genomes among 375 eligible assemblies 8.8%, Compared with 23/2493 corresponding human associated assemblies or 0.9%. However. Thirty two out of the 33 environmental co-carriage records originated from China, and 29 belong to only two bio projects. This highly concentrated distribution substantially limits any interface that searches combinations are generally more prevalent in environmental populations or in upper middle-income settings. The co-occurrence of determinants associated with

resistance to critically important antimicrobial classes is a legitimate genomic surveillance concern, But the interpretation requires caution. Genome level co-carriage does not show whether genes are located on the same plasmid on separate replicas or on the chromosome. Nor does it demonstrate co- transfer, co-selection or resistance to both antimicrobial classes in the same isolate. Investigations combining complete plasmid reconstruction, longitudinal surveillance and laboratory characterization has shown how additional genomic and epidemiological evidence can distinguish the dissemination of mobile resistance elements from the spread of bacterial lineages[8, 26]. Our findings therefore identify candidate resistance gene combinations for prioritized investigation without establishing their physical linkage or transmission dynamics. Though we have identified the replicon abundance on the ECC genomes and also have a network of co-occurrence of the plasmid replicon and the AMR genes but only co-occurrence does not establish the concerned plasmid mediated transmission.

Nevertheless, several limitations should be considered. First, the analysis is based on publicly available genomes assembled from heterogeneous sampling programs rather than prospectively designed global surveillance network. Countries, ecological sources and lineages unevenly represented and genomes originating from the same bio project or closely related epidemiological populations may not constitute independent observations. Although the country level species composition analysis at least one aspect of this problem, the corresponding resistance gene comparisons were not comprehensively adjusted for country, sequencing project and bacterial population structure. Their statistical significance should therefore be interpreted as exploratory evidence of heterogeneity within the sampled collection. Second, Source metadata were harmonized into broad categories that do not fully resolve host species, clinical sampling identifications, environmental. Microhabitats and the epidemiological relationships between isolates. Third, AMRfinderplus based annotation identifies genomic resistance associated determinants but does not directly establish antimicrobial susceptibility, gene expression, resistance, gene mobility, or the contribution of regulatory and chromosomal mechanisms. This limitation is particularly important for intrinsic ECC, resistance mechanisms, polymerics in associated genes and the distinction between gene carriage and clinically defined multi-drug resistance. Fourth, the temporal distribution is affected by unequal sampling across years, limited representation of early collection periods. Finally, national income categories provide broad contextual classification rather than direct measurements of socioeconomic exposures, anti-microbial consumption or public awareness. The use of a single income classification or snapshot across historically collected genomes further limit longitudinal socioeconomic interpretation. These limitations do not negate the descriptive and competitive value of the findings, but define the level of inference that the data can support. Future investigations would benefit from geographically balanced sampling of ECC across human animal and environmental settings combined with harmonized isolate metadata,

antimicrobial susceptibility measurements and information on relevant health care and environmental exposures.

In conclusion, this study reveals substantial geographical, taxonomy, temporal and ecological heterogeneity in the sampled global ECC resistome. Species composition and resistance associated gene repertoires differed across isolation sources, while several clinically important, acquired resistance determinants showed pronounced but geographically uneven distributions across country income categories. Importantly, genome level differences in income stratified species composition were not consistently reproduced by country level diversity and composition analysis, demonstrating the influence of surveillance structure on conclusions. Drawn from public genomic repositories, these findings emphasize that ECC antimicrobial resistance surveillance requires species resolve analysis, geographically balanced sampling and careful interpretation of genomic carriage in relation to ecological and socioeconomic context. By identifying both resistances associated patterns and the limits imposed by uneven genome representation, this study provides a framework for more reliable on health interpretation of publicly available ECC genomic data and highlights priorities for future surveillance of this clinically important bacterial complex.

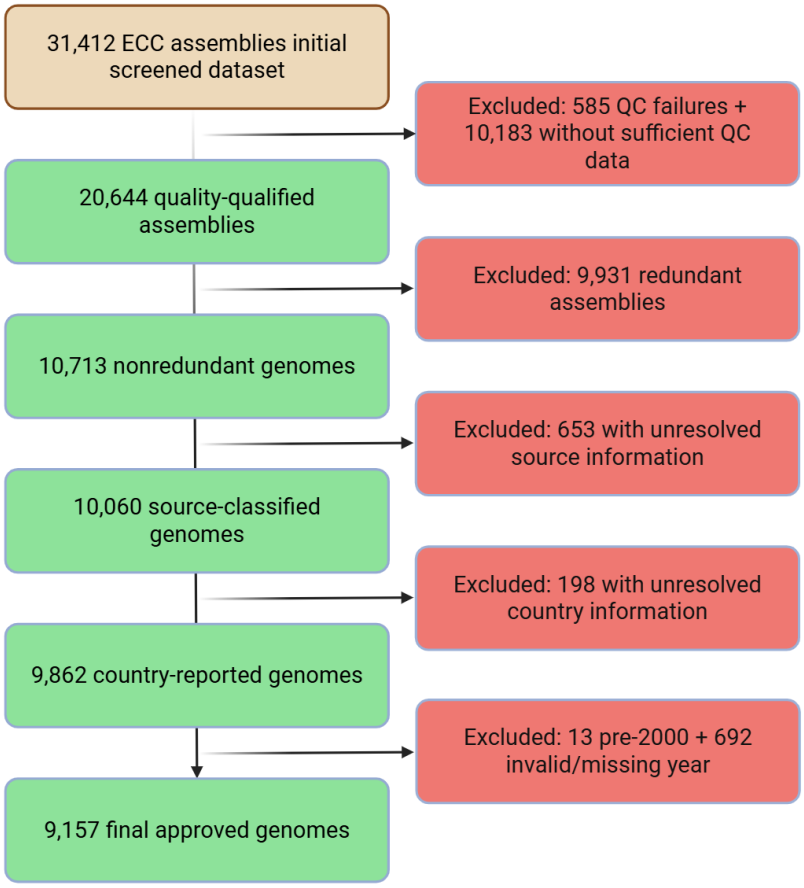
Method:

Genomic inclusion, exclusion and quality control criteria:

A systematic genomic selection framework was implemented to construct a high quality, non-redundant and metadata complete collection of Enterobacter cloacae complex genomes for global comparative genomics. Followed by taxonomic screening 31,412 ECC genome assemblies were included in initial quality assessment data. Genome quality assessed using CheckM derived completeness and contamination estimates. Assemblies were excluded when completeness of the genomes were ≤ 95 % and contamination having ≥ 5%. A total of 585 assemblies failed quality criteria and 10,183 assemblies has data unavailable. To minimize redundancy genomes associated with same bio sample were evaluated and a single representative of the assemblies were retained. This produce excluded 9,931 redundant assemblies, resulting in 10,713 non-redundant quality qualified genome records. Subsequently genomes were screened for availability of One health source metadata. Isolates classified into Human, Animal, Environmental, Food and plant associated categorized harmonized host and isolation source annotations. A total of 653 genomes with missing and ambiguous source information were excluded, retaining 10,060 genomes. Geographical eligibility was established using standardized country level metadata. 198 Genomes lacking country assignment were excluded., yielding 9,862 country reported genomes. Temporal eligibility was restricted to genomes collected between 1st January 2000 and 18th July, 2026

(**Supplementary Data 1**) based on verified collection year. Thirteen genomes were collected before 2000 and 692 genomes with missing collection years were excluded.

Following these sequential filtering procedures; 9,157 high quality genomes were retained in the final curated ECC dataset for global geographical, taxonomic, One health, temporal and genomic analysis. The final collection included 8,882 genomes assigned to seven resolved ECC species and 275 genomes retained as unresolved ECC. All inclusions and exclusion decisions were implemented sequentially and genomes association identifiers were maintained to preserved traceability between the original genomes collection and the final analytical.



(generated at :https://BioRender.com/sx2ol38)

Multi-locus typing and lineage distribution:

Multi-locus sequence typing was performed using mlst v2.35.0 and the Enterobacter cloacae seven locus scheme comprising dnaA, fusA, gyrB, leuS,pyrG, rplB and rpoB. Allelic profiles and sequence type assignments were generated from genome assemblies using corresponding reference allele database. Genome level MLST were integrated with taxonomic assignments and harmonized epidemiological metadata by assembly accession.

Antimicrobial resistance gene identification and resistome profiling:

Antimicrobial resistance determinants were identified using AMRfinderPlus v4.2.7 with NCBI AMRFinderPlus reference database released 2026-08-07.1. Genome assemblies were analyses individually using nucleotide sequence search functionality. Gene prevalence was calculated as the number of unique genomes carrying the respective determinant by total number of genomes within the specified analytical stratum. Multiple detections of the same gene within a genome were distinguished from the number of gene positive genomes to avoid inflation of genome level prevalence estimates.

Plasmid replicon identification:

Plasmid replicons were identified from assembled Enterobacter cloacae complex genomes using Plasmidfinder. Analysis was done using BLASTn based sequence alignment, applying minimum thresholds of 95% nucleotide sequence identity and 60% reference sequence coverage.

Data processing and Visualization:

Data integration and numerical summarization and visualization were performed using Python 3.14.7. Tabular data were processed using pandas and NumPy 2.5.3, with spreadsheet input and output handled using openpyxl 3.1.5. Statistical calculations using relevant python analytical libraries including SciPy 1.18.1. Publication figures were generated using Matplotlib 3.11.2, Geographical visualizations were prepared using GeoPandas 1.2.0.

Statistical analysis:

Descriptive analysis was used to characterize geographical, temporal taxonomic and epidemiological distributions of ECC genomes. Categorical variables were summarized as absolute counts and within group percentages. Species composition was assessed across One health source categories and country income groups, with percentages calculated using total number of genomes within the corresponding category as denominator. Association between species composition and country income classification were examined by Pearson's Chi-squared test of independence. Country level taxonomic diversity was characterized using observed using species richness and Shannon diversity index. Species richness was additionally evaluated after rarefaction to a common sampling depth of 20 genomes per country. Antimicrobial resistance burden was quantified using means medians and number of genomes. Within species and income stratified groups, differences in genome level ARG burden between one health compartments were examined using Mann- Whitney U tests. Resulting P values were adjusted for multiple comparisons using Benjamini Hochberg false discovery rate (FDR). Adjusted P values were reported as q values. Temporal variation in ARGs evaluated across six collection periods: 2000-2004,2005-2009,2010-2014,2015-2019,2020-2024 and 2025-2026.Perisod specific ARG burdens were summarized with standard errors of mean. Kruskal-Wallis tests were performed P values were adjusted with FDR similarly. Pairwise comparisons between adjacent temporal periods and between the earliest and latest periods were performed using two-sided Mann-Whitney U tests

with Holm adjustment for multiple comparisons within each species. Pairwise differences were considered significant only when the corresponding adjusted omnibus and pairwise results satisfied the predefined significance threshold.

Declarations

Abbreviations:

Antimicrobial Resistance: AMR

Antibiotic Resistance Gene: ARG

Benjamini-Hochberg: BH

Basic Local Alignment Search Tool: BLAST

Confidence interval: CI

Enterobacter cloacae complex: ECC

Extended spectrum β-lactamase: ESBL

False discovery rate: FDR

Low-and Middle-income countries: LMICs

Multi-locus sequence Typing: MLST

National Center for Biotechnology Information: NCBI

Standard deviation: SD

Standard error of the mean: SEM

Sequence type: ST

World Health Organization: WHO

Ethics approval: Not applicable

Consent for publication: Not applicable

Funding: No financial support was received for this study.

Competing Interests: The authors declare no conflict of interest.

Usage of Generative AI: ChatGPT 6 used during the writing process to improve the readability and language of the manuscript. The authors have reviewed and edited all AI-assisted custom python scripts as needed and take full responsibility for the final content of the article.

Availability of data and materials: All the genomes in this study are published in NCBI and curated metadata added in supplementary data 1

Code availability: Custom python scripts used for analysis is available at https://github.com/deysandy007/Enterobacter-cloacae-complex-

Author's Contribution:

SD: Conceptualization, Manuscript writing, Data analysis, Figure preparation

SSD: Conceptualization, Manuscript writing, Data validation

All authors reviewed the Manuscript.

Acknowledgements:

The authors acknowledge Indian Council of Medical Research (ICMR) and Academy of Scientific and Innovative Research, Training and Development Complex, Ghaziabad, Uttar Pradesh for their constant support and also Figures containing graphical elements created using Bio Render were prepared from BioRender.com and also Debjani Bhattacharjee(ICMR-JRF at ICMR NIRBI , Kolkata) for her support in figure exporting and Dr. Rahul Shubhra Mandal (Research Associate at Perelman School of Medicine, University of Pennsylvania) for his immense support .

References:

1. Naghavi, M., et al., *Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050.* The Lancet, 2024. **404**(10459): p. 1199–1226.
2. Sati, H., et al., *The WHO Bacterial Priority Pathogens List 2024: a prioritisation study to guide research, development, and public health strategies against antimicrobial resistance.* The Lancet infectious diseases, 2025. **25**(9): p. 1033–1043.
3. Davin-Regli, A., J.-P. Lavigne, and J.-M. Pagès, *Enterobacter spp.: update on taxonomy, clinical aspects, and emerging antimicrobial resistance.* Clinical microbiology reviews, 2019. **32**(4): p. 10.1128/cmr. 00002–19.
4. Annavajhala, M.K., A. Gomez-Simmonds, and A.-C. Uhlemann, *Multidrug-resistant Enterobacter cloacae complex emerging as a global, diversifying threat.* Frontiers in microbiology, 2019. **10**: p. 44.
5. Subedi, D., et al., *Rational design of a hospital-specific phage cocktail to treat Enterobacter cloacae complex infections.* Nature Microbiology, 2025. **10**(11): p. 2702–2719.
6. Barceló Munar, I.M., et al., *Filling knowledge gaps related to AmpC-dependent β-lactam resistance in Enterobacter cloacae.* Scientific Reports, 2024, vol. 14, num. 189, p. 1-11, 2024.
7. Kayama, S., et al., *National genomic surveillance integrating standardized quantitative susceptibility testing clarifies antimicrobial resistance in Enterobacterales.* Nature Communications, 2023. **14**(1): p. 8046.

8. Koh, V., et al., *Plasmid dynamics driving carbapenemase gene dissemination in healthcare environments: a nationwide analysis of closed Enterobacterales genomes.* Nature communications, 2025. **16**(1): p. 9522.
9. Doijad, S.P., et al., *Resolving colistin resistance and heteroresistance in Enterobacter species.* Nature communications, 2023. **14**(1): p. 140.
10. Ikhimiukor, O.O., et al., *Dissemination dynamics of colistin resistance genes mcr-9 and mcr-10 across diverse Inc plasmid backbones.* Communications medicine, 2025. **5**(1): p. 399.
11. García-Martín, A.B., et al., *Prospective One-Health investigation into low-abundant extended-spectrum β-lactamase producing Enterobacterales enables detection of potential dissemination events and persistence.* Science of The Total Environment, 2024. **950**: p. 175078.
12. Cocker, D., et al., *Investigating One Health risks for human colonisation with extended spectrum β-lactamase-producing Escherichia coli and Klebsiella pneumoniae in Malawian households: a longitudinal cohort study.* The Lancet Microbe, 2023. **4**(7): p. e534–e543.
13. Djordjevic, S.P., et al., *Genomic surveillance for antimicrobial resistance—a One Health perspective.* Nature Reviews Genetics, 2024. **25**(2): p. 142–157.
14. David, S., et al., *Monitoring antimicrobial resistance trends from global genomics data: amr. watch.* PLOS global public health, 2025. **5**(11): p. e0005256.
15. Wang, Y., et al., *A global atlas and drivers of antimicrobial resistance in Salmonella during 1900-2023.* Nature communications, 2025. **16**(1): p. 4611.
16. Bottini, I., et al., *Environmental, socioeconomic, and livestock-related factors associated with antibiotic resistance in Escherichia coli, Shigella spp., and Salmonella enterica: A global One Health analysis.* One Health, 2026: p. 101600.
17. Yaikhan, T., et al., *Genomic analysis of Enterobacter cloacae complex from Southern Thailand reveals insights into multidrug resistance genotypes and genetic diversity.* Scientific Reports, 2025. **15**(1): p. 4670.
18. Long, X., et al., *Global genomic and molecular epidemiology of carbapenem-resistant Enterobacter hormaechei.* Communications Biology, 2026.
19. Mustafa, A., et al., *Genome-wide Analysis of Four Enterobacter cloacae complex type strains: Insights into Virulence and Niche Adaptation.* Scientific Reports, 2020. **10**(1): p. 8150.
20. Yan, S., et al., *Changing antimicrobial resistance profile of Enterobacter spp. isolates in hospitals across China: a seven-year analysis from the CHINET antimicrobial resistance surveillance program (2015–2021).* One Health Advances, 2024. **2**(1): p. 11.
21. Craddock, H.A., et al. *Tracking extended-spectrum beta lactamase-producing Enterobacterales across one Health among Bedouin communities*. in *Open Forum Infectious Diseases*. 2025. Oxford University Press US.
22. Abi, O., et al., *A One Health view of antimicrobial resistance in North Africa: Meta-analysis of ESBL-producing Enterobacteriaceae across sectors.* One Health, 2026: p. 101402.
23. Sands, K., et al., *Characterization of antimicrobial-resistant Gram-negative bacteria that cause neonatal sepsis in seven low-and middle-income countries.* Nature microbiology, 2021. **6**(4): p. 512–523.

24. Carvalho, M., et al., *Antibiotic resistance genes in the gut microbiota of mothers and linked neonates with or without sepsis from low-and middle-income countries.* Nature microbiology, 2022. **7**(9): p. 1337–1347.
25. Portal, E., et al., *Characterisation of colistin resistance in Gram-negative microbiota of pregnant women and neonates in Nigeria.* Nature Communications, 2024. **15**(1): p. 2302.
26. Marimuthu, K., et al., *Whole genome sequencing reveals hidden transmission of carbapenemase-producing Enterobacterales.* Nature communications, 2022. **13**(1): p. 3052.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary6CountryWeightedData.xlsx
- Supplementary5CountryDiversityData.xlsx
- Supplementary3SpeciesOneHealthData.xlsx
- Supplymentary11CORRECTEDIMPMCRDataPhylogeny.xlsx
- Supplementary8FINALALLGENOMESSTATISTICS.xlsx
- Supplementary4SpeciesIncomeData.xlsx
- Supplementary9ALLGENESPREVALENCEMASTER.xlsx
- Supplementary2CombinedAllSTDeepIncomeCompleteData.xlsx
- Supplementary7RefinedTemporalStatistics.xlsx
- Supplymentary10ECC7SPHERESMAXIMUMSOURCEDATA.xlsx
- Supplementarydata1ECCFINAL20002026MASTER.xlsx