Title

Genomic analysis of plasmid content in food isolates of E. coli strongly supports its role as a reservoir for the horizontal transfer of virulence and antibiotic resistance genes.

Authors

Del Carmen Rocha-Gracia, Rosa

Balbuena-Alonso, María

Cortés-Cortés, Gerardo

et al.

Publication Date

2022

DOI

10.1016/j.plasmid.2022.102650

Peer reviewed

HHS Public Access

Author manuscript

Plasmid. Author manuscript; available in PMC 2024 February 26.

Published in final edited form as:

Plasmid. 2022 ; 123-124: 102650. doi:10.1016/j.plasmid.2022.102650.

Genomic analysis of plasmid content in food isolates of *E. coli*, strongly supports its role as a reservoir for the horizontal transfer of virulence and antibiotic resistance genes

María G. Balbuena-Alonso[a], **Gerardo Cortés-Cortés**[a,b], **Jay Kim**[b], **Patricia Lozano-Zarain**[a], **Manel Camps**[b,*], **Rosa del Carmen Rocha-Gracia**[a,*]

[a]Posgrado en Microbiología, Centro de Investigaciones en Ciencias Microbiológicas, Instituto de Ciencias de la Benemérita Universidad Autónoma de Puebla. Instituto de Ciencias, Ciudad Universitaria, San Manuel, Puebla, 72570, Mexico.

[b]Department of Microbiology and Environmental Toxicology, University of California at Santa Cruz, Santa Cruz, CA, 95064, USA.

Abstract

The link between *E. coli* strains contaminating foods and human disease is unclear, with some reports supporting a direct transmission of pathogenic strains via food and others highlighting their role as reservoirs for resistance and virulence genes. Here we take a genomics approach, analyzing a large set of fully-assembled genomic sequences from *E. coli* available in *GenBank*. Most of the strains isolated in food are more closely related to each other than to clinical strains, arguing against a frequent direct transmission of pathogenic strains from food to the clinic. We also provide strong evidence of genetic exchanges between food and clinical strains that are facilitated by plasmids. This is based on an overlapped representation of virulence and resistance genes in plasmids isolated from these two sources. We identify clusters of phylogenetically-related plasmids that are largely responsible for the observed overlap and see evidence of specialization, with some food plasmid clusters preferentially transferring virulence factors over resistance genes. Consistent with these observations, food plasmids have a high mobilization potential based on their plasmid taxonomic unit classification and on an analysis of mobilization gene content. We report antibiotic resistance genes of high clinical relevance and their specific incompatibility group associations. Finally, we report a striking enrichment for adhesins in food plasmids and their association with specific IncF replicon subtypes. The identification of food plasmids with

*Corresponding authors: rochagra@yahoo.com, mcamps@ucsc.edu.

Authors and contributors

Maria Balbuena-Alonso: Conceptualization, Data curation, Formal analysis, Investigation, Visualization and Writing – original draft.
Gerardo Cortés-Cortés: Conceptualization, Resources, Supervision, Visualization and Writing – review and response to reviewers & extensive editing.
Jay Kim: Conceptualization, Visualization.
Patricia Lozano-Zarain: Conceptualization, Resources, Visualization.
Manel Camps: Conceptualization, Funding acquisition, Methodology, Resource, Supervision, Visualization, and Writing – review & editing.
Rosa del Carmen Rocha-Gracia: Conceptualization, Funding acquisition, Project administration, Resources, Visualization, Writing – review & editing.

Conflicts of interest

The authors declare that there are no conflicts of interest.

specialized PTU-Inc group markers as mediators of horizontal transfer between food and clinical strains opens new research avenues and should assist with the design of surveillance strategies.

Keywords

Escherichia coli ; food plasmids; antimicrobial resistance; plasmid origin of replication; plasmid taxonomic units; comparative genomics

1. Introduction

Escherichia coli (*E. coli*) is a bacterial species found in the environment that colonizes the gastrointestinal tract of humans and animals. Most strains of *E. coli* are harmless commensals, but there are pathogenic lineages capable of producing severe intestinal and extraintestinal disease (sepsis, urinary tract infections, meningitis) (1).

E. coli is characterized by frequent genetic exchanges across different strains and even with other species, a process known as horizontal gene transfer (HGT). The resulting high genomic plasticity has allowed *E. coli* to adapt to different selective pressures exerted in various environments, including the rapid evolution of resistance to antibiotics frequently used in the clinic, notably β-lactam antibiotics, fluoroquinolones, and aminoglycosides (2–5).

E. coli strains can be classified into large groups by phylotype (according to phylogenetic relationship) and by pathotype (according to pathogenic profile). Serotyping and Multi-Locus Sequence Typing (MLST) are two methods used to provide a higher level of resolution. Serotyping is based on an agglutination test performed with a panel of reference antisera with the ability to discriminate variations in the genes responsible for the biosynthesis of structures such as the flagellum, the lipopolysaccharide (LPS), and the capsule (6). The MLST classification uses a genomics approach based on the presence of unique combinations of alleles for seven housekeeping genes (7). A few sequence types (STs) have emerged as epidemic, being present at high frequencies all over the world. Prominent among these is ST131, a clone whose expansion has been accompanied by the acquisition of fluoroquinolone resistance, of an adhesion virulence factor, and of a set of plasmid-borne antibiotic resistance genes (ARGs)(8,9).

HGT is greatly facilitated by plasmids, which are autonomously replicating pieces of DNA. In addition to the genes needed for plasmid homeostasis, plasmids frequently contain a cargo of genes involved in adaptation to environmental challenges (10–12). Establishing phylogenetic relationships between plasmids is a major challenge for the field given the instability of their sequence, which is constantly shuffled by recombination. Cargo content and individual genes involved in replication (*ori*) (13) or conjugation (relaxase) (14) have been used to define genetic relatedness between plasmids with only partial success. Recently, a method based on the fraction of sequence that perfectly aligns between two plasmids (the alignment fraction) has been used to define Plasmid Taxonomic Units (PTUs), which cluster plasmids by phylogenetic relatedness more reliably than other methods (15).

This analysis in turn has allowed the systematic tracking of plasmid exchanges across different taxonomic boundaries (16).

Plasmid-mediated horizontal gene transfer is mediated by conjugation (transfer of ssDNA by a process involving nicking of the donor DNA and cell-cell contact), transformation (uptake of DNA from the environment) or transduction (genetic exchange mediated by a phage). The conjugation ability of a given plasmid can be inferred based on the presence of a complete complement of the machinery that is required for self-transmission, which consists of four MOB (mobilization) modules: an origin of transfer site (*oriT*); a relaxase gene that cleaves the *oriT*, generating a ssDNA; and two mating channel components: a gene encoding type IV coupling protein (T4CP) and a gene cluster for bacterial type IV secretion system (T4SS) (11,17). Some plasmids lack some MOB genes but can still be transferred if these factors are produced in *trans* by helper plasmids. These are called mobilizable plasmids (18). Within plasmids, mobilization is further facilitated by sequence elements that facilitate recombination such as transposons, which are generally known as mobile genetic elements (MBEs) (19). In this article, we focus on two types of cargo, namely antibiotic resistance genes (ARGs) and virulence genes. ARGs involve a variety of mechanisms, but the most frequent mechanism found in plasmid-borne ARGs is antibiotic inactivation. Inactivation can be dependent on degradation (*e.g.*, β-lactamases enzymes), on chemical modification (*e.g.*, aminoglycoside- and fluoroquinolone- modifying enzymes) or on the modification of the target (*e.g., qnr* for fluoroquinolone resistance, *mcr-1* for colistin resistance, and *RmtA* and *armA* 16S rRNA methylases for aminoglycoside resistance) (4,5,20). Virulence genes encode for (among others) adhesins, invasins, toxins, hemolysins, iron homeostasis genes, antioxidants, and other immune evasion genes, as well as secretion systems (21,22). Given that the exchange of ARGs and virulence genes is mediated by similar MGEs, these two categories of genes are frequently linked, which is one of the reasons why multidrug resistant strains (MDR) also tend to be more virulent and why MDR is becoming such a serious public health problem (23,24).

The transmission of virulent *E. coli* strains to humans involves different routes of infection. In this work, we investigate the risk posed by contaminated food. The role of food as a reservoir for pathogenic *E. coli* infection is still under discussion. Some studies argue against this idea based on differences in the clonal lineage profile of food-borne and clinical strains (25,26); however, other studies comparing *E. coli* isolates from humans with isolates from food-producing animals identified genetically related plasmids that are shared between the two pools, suggesting that plasmid-mediated HGT from food strains may contribute to the pathogenicity and drug resistance exhibited by clinical strains (27–29). Here we compared a dataset of fully assembled genomes from food *E. coli* isolates, with the genomes of clinical isolates available from *GenBank*. We found that strains isolated in food are consistently more closely related to each other than to clinical strains, arguing against a frequent transmission of food strains to the clinic. We also show that food plasmids have a high mobilization potential and provide strong evidence of genetic exchange between food and clinical plasmid pools. This exchange appears to depend on a small subset of genetically-related plasmids. We identify the PTU and Inc group markers associated with these plasmids and reveal a degree of specialization, with some PTU-Inc group clusters preferentially transferring virulence factors over ARGs. We also identify genes of high

clinical relevance and their incompatibility group associations. The identification plasmid associated with HGT between food and clinic opens new avenues of research and should assist with the design of surveillance strategies.

2. Methods

2.1 Selection of sequences of *E. coli* genomes

Completely assembled *E. coli* genomes deposited until December 31, 2020 in the GenBank database were selected and checked for the presence of chimeric, contaminated, duplicated, or misassembled sequences. The final database contains genomes that passed these quality controls. Ninety-three sequences were excluded because they were defined as anomalous, corresponded to partial chromosomal or plasmid DNA sequences, not available or duplicated due to updating their assembly number or were internal control sequences of sequencing platforms. Manual curation using NCBI Pathogens detection, and the Pathosystems Resource Integration Center (PATRIC) databases, information deposited in each Bioproject, Biosample or the reported associated was performed to confirm the type of sample and the source of isolation to select those of food and human clinical origin; for each of the strains associated with the genomes was considered uniquely from the source reported by the authors; thus, no section showed a strain collected from both origins at the same time. The files downloaded from each genome were fna (fasta nucleic acids), faa (fasta protein sequence), and gbk; these files were used depending on the input needed by each bioinformatics tool. Also, all plasmids were collected from the NCBI nucleotide database, from the resources INSDC (which includes DDBJ, EMBL-EBI and GenBank) and RefSeq; preferring RefSeq records over the INSDC records. The record was required to have the assembly tag “Complete genome”. When the plasmid record was linked to multiple assembly IDs only the assembly with the label “latest” was assigned to this record.

2.2 Genomic analysis

Genomes were assigned to an *E. coli* phylogroup using ClermonTyping 1.4 (30) (http://clermontyping.iame-research.center). In addition, *E. coli* MLST profiles were predicted using the Achtman seven gene MLST scheme at PubMLST (https://pubmlst.org/organisms/escherichia-spp (accessed on February 26, 2021), the serotypes (O:H) were predict with Serotype finder 2.0 (31) (https://cge.cbs.dtu.dk/services/SerotypeFinder/), and pathotypes were assigned according to the presence of specific marker virulence genes (32) using the BLASTn tool (85% identity and 60% minimum length, and E-value ≤ 0.000001) (www.ncbi.nlm.nih.gov/BLAST); those containing markers of different pathotypes were described as hybrids and named according to the literature. Furthermore, the source of isolation for the designation of UPEC pathotype was included. Where available, the pathotype description was described as the original publication. To analyze phylogenetic relationship, the chromosomes of food and clinical *E. coli* were used using the default parameters of CSI Phylogeny 1.4 (https://cge.cbs.dtu.dk/services/CSIPhylogeny) default parameters and *E. coli* K-12 was used as reference (NC_000913.3) (33). Subsequently, the dendrogram was performed by the maximum likelihood method with 1000 bootstrap replicates using Fastree 2.1 (34). Finally, it was visualized using iTOL 6.0 (https://itol.embl.de/) (35).

2.3 Plasmid analysis

To identify plasmid genomes in genomic assemblies of *E. coli*, a manual review of the plasmids registry in each genome in *GenBank* was performed and the PlasmidFinder 2.1 (https://cge.cbs.dtu.dk/services/PlasmidFinder/) (85% identity and 60% minimum length) was used to identify the replicon present for each plasmid (36), and pMLST (http://pubmlst.org/plasmid/) was used to determine the subtypes in those defined as IncF plasmids (37). Plasmid Taxonomic Units (PTU) were identified by COPLA (a taxonomic classifier of plasmids) using the recommended parameters (16) (https://castillo.dicom.unican.es/copla/).

The presence of conjugation elements in plasmids was performed by using the tool OriTfinder 1.1 (38) (https://bioinfo-mml.sjtu.edu.cn/oriTfinder/) with modificated parameters (Blast E-value 0.00001) and the identity of relaxases was confirmed with MOBScan (39) (https://castillo.dicom.unican.es/mobscan) with default settings. The classification of plasmids was established according to their MOB and MPF content: Conjugative (presence of *oriT*, relaxase, TC4P and TSS4); non-mobilizable (absence of *oriT* or relaxase) and mobilizable (presence of *oriT* and relaxase, with variable TC4P) (40).

To analyze the phylogenetic relationship of representative plasmids, the complete nucleotide sequences of the outstanding overrepresentation and overlapping plasmids in both data subsets were compared using the neighbor-joining method. This method was used to construct a phylogenetic tree with 1000 bootstrap replicates (41). Finally, it was visualized with iTOL 6.0 (https://itol.embl.de/) (35).

2.4 Identification of antimicrobial resistant and virulence genes in plasmids

ResFinder 4.1 (https://cge.cbs.dtu.dk/services/ResFinder/) and VirulenceFinder 2.0 (https://cge.cbs.dtu.dk/services/VirulenceFinder/) were used to identify acquired antimicrobial resistance genes as well as virulence gene content, respectively; with search parameters at a 90% threshold for identity and 60% minimum length (42,43). Genes were considered to overlap when the same label was identified in food and in clinical plasmids. To visualize the genetic structure of the plasmids carrying two genes of special clinical interest (NDM-5 and *mcr-1.1*); sequences were initially annotated with Prokka 1.14.6(44) and, mobile genetic elements related to antibiotic resistance were identified using Mobile Element Finder 1.0.3 from the Center of Genomic Epidemiology (45) using the default parameters. Relevant plasmids were aligned with BRIG 0.95(46) and visualized using Proksee.ca 1.0.

2.5 Statistical analysis

The analysis was mainly descriptive, showing proportions as percentages and continuous variables as mean and SD. The Mann-Whitney test was used to compare distributions. We estimated the probability of overlap of virulence genes and antibiotic resistance genes using Monte Carlo simulation with 10^4 iterations; the genes identified in both sets (175 resistance genes and 72 virulence genes) were used as input data; 161 data were randomly sampled for each group, and 10^4 iterations were performed, indicating the number of plasmids each time, calculating the overlap in each iteration. Rstudio statistical software 1.4.1103 was used to perform all analyses.

2.6 Jaccard Index calculations

The Jaccard index (JI) measures the co-occurrence of two binary features, and ranges from 0 (mutual exclusion) to 1 (total co-occurrence). To calculate this index, we used the following formula:

$$J(m_i, m_j) = \frac{1}{|M|} \sum_{m \in M} \frac{\#\ strains\ with\ in\ ST_m\ with\ m_a\ \mathbf{and}\ m_b}{\#\ strains\ in\ ST_m\ with\ m_a\ \mathbf{or}\ m_b}$$

where M is the set of plasmids belonging to a given Inc or PTU, ma is the number having resistance (JI for resistance) or virulence (JI for virulence) in food plasmids and mb the number having resistance (JI for resistance) or virulence (JI for virulence) in clinical plasmids.

3. Results

3.1 *E. coli* genomic database

We generated a database of 1794 complete *E. coli* genomes deposited in *GenBank* at the National Center for Biotechnology Information (NCBI) by the end of 2020. The quality controls are described in Methods. We focused on genomes from food (n=77) and from clinical sources (n=649); their corresponding accession numbers are available in Table S1A and location and phylogenetic characteristics in Table S1B of the supplementary materials.

These genomes came from six different sources (Figure S1A). Human clinical samples were the most abundant (36%). A total of 45% of these samples were recovered from feces, urine, and blood; the rest came from samples of other body fluids, secretions, expectorations, swabs and generically described by the authors as human clinical specimens. Strains obtained from food sources, by contrast, represented only 4% of the isolates. The majority of these food-associated samples such as meat (cattle, poultry, and swine) or dairy products (cheese and milk) were linked to food-producing animals. Other food-associated samples came from fresh produce and flours (Figure S1B). Note that the geographic distribution of the sampling was skewed for genomes from countries that have clinical environment- and foodborne outbreak-surveillance systems in place. Therefore, isolates from Africa, Oceania, Latin America, and Asian countries other than China were infrequent (see Table S1B).

3.2 Lineage classification

Population studies of *E. coli* have mainly described eight phylogroups (A, B1, B2, C, D, E, F, and G) (47,48). Seven of these eight phylogroups (A, B1, B2, C, D, E, F) were represented in both populations, demonstrating that *E. coli* exhibits an extremely wide genetic diversity in both populations. The relative abundance of these phylogroups was also similar in both populations (Figure 1), though phylogroup B2 was much more abundant in clinical isolates (21%, compared to 1%). This striking difference is consistent with several reports (49–51) and is likely due to the presence of UPEC isolates, which represent a significant percentage of the identified phylogroup B2.

When we go down to the level of STs, by contrast, we see a highly discordant ST profile: of the 218 STs present in the combined dataset, only 16 STs overlapped. This observation is in agreement with previous reports noting substantial lineage differences between food and clinical isolates (25,26). Of the overlapping STs, ST11 is the most prevalent (23% in food and 11% in clinical), suggesting that ST11 strains may exhibit a high level of fitness in both ecological niches, and is thus potentially highly transmissible. The high transmissibility of ST11 via food would be congruent with studies that report the classification ST11 with clone EHEC O157:H7, which is a strain linked to outbreaks of human infections caused by the consumption of contaminated meat (52,53). Clone EHEC O157:H7 has been reported as a pathotype adapted to cattle and disseminated to humans through direct contact with animals or through the environment or food (81–83).

Figure 2 shows the phylogenetic relatedness between food and clinical strains in the form of a cladogram, as well as the phylogroup and sequence type classification. An almost complete segregation between clinical and food strains is apparent, with food strains all clustered in two sectors (see Methods). Taken together, the discordance in STs and the segregation seen in the cladogram between food and clinical strains, strongly argue against a frequent transmission of food strains to the clinic, with the possible exception of some ST11 strains.

We identified four pathotypes in food genomes (EHEC, EPEC, STEC, DAEC) and three additional ones in clinical genomes (UPEC, EIEC, and EAEC). EHEC was the most frequent pathotype in both populations (Figure 1). The lack of EAEC strains or EAEC/EHEC hybrid pathotypes in food strains is surprising, based on reports of strains found in food matrices that are genetically similar to strains isolated from urinary tract infections (54,55) and the report of EAEC/EHEC hybrid pathotypes in a large foodborne outbreak in Europe (56). This observation is likely due to a combination of a small sample size of food strains and the biases inherent to epidemiological surveillance programs.

3.3 Profiling of plasmids based on general structural and functional features.

A total of 161 plasmids were identified in genomes from food strains based on circular assemblies of less than 500 kb in size linked to at least one plasmid replicon. Using the same approach, we found 1624 plasmids in the genomes of clinical strains (see Methods).

Both plasmid populations present a comparable, very broad range of sizes (1.1 kb - 404 kb) and a trimodal distribution (Figure S2). Food and clinical strains have a moderately different average number of plasmids per host (2.0 and 2.5 plasmids per genome, respectively, Figures 2, S2).

Plasmids are classified into incompatibility (Inc) groups based on the inability of two plasmids of the same group to be stably maintained in the same cell. This functional classification was initially proposed as a rough phylogenetic classification by identifying redundancy in plasmid copy number regulatory elements within a replicon (57,58) although it is known to be an imperfect marker of phylogenetic relatedness (16). When no known replicon was assigned by PlasmidFinder, we defined plasmid as Non-Typeable (NT).

PlasmidFinder identified 27 Inc categories in *E. coli*-plasmids isolated from food, compared to 40 for clinical strains. The frequency distribution of the 23 Rep categories that overlap between the two groups is similar, shown in Figure S3. IncF plasmids were further typed using the subtype classification described by Villa et al (37). Table 1A shows the relative representation of replicon categories as a ratio of frequency in the clinic plasmid pool versus frequency in the food plasmid pool. The following five Inc groups stand out for their overrepresentation in food plasmids: IncR (10x overrepresentation), IncHI1B (6x), and Inc X1 (3x) and two IncF subtypes: IncF(F^-:A^-:B1) (2.9x) and IncF23:A^-:B3 (2.75x).

3.4 Plasmid profiling according to mobilization and promiscuity features.

Food and clinical plasmids were sorted into three categories corresponding to their predicted transfer capacity: conjugative, mobilizable (*i.e.*, can be conjugated in the presence of a helper plasmid), and non-mobilizable (See Methods). For the food plasmid population, 27.3% of the plasmids were classified as conjugative, and 9.3% as mobilizable, compared to 30.3% and 6.3%, respectively, for clinical plasmids. Almost all conjugative plasmids were larger than 25 kb, while mobilizable plasmids were mostly small, suggesting that in most cases small plasmids are mobilized with the help of large conjugative plasmids (Figure 3). This size distribution is consistent with previous reports (59). Overall, we found that over one third of all food plasmids had the elements necessary for HGT, and that this fraction was similar in food and clinical plasmids.

We next grouped the plasmids by phylogenetic relatedness through COPLA, a recently developed method that classifies plasmids into Plasmid Taxonomic Units (PTUs) based on nucleotide identity alignment fractions (see Methods and 16). COPLA was only able to assign a PTU to 124 of the 161 food plasmids (distributed in 39 different PTUs); in the case of clinic plasmids, 1236 out of a total of 1624 plasmids were assigned to 60 different PTU. Three of the 39 food PTUs (PTU-E46, PTU-E54, and PTU-E79) were absent in the clinic (Table S2). Figure 4A shows the PTU classification of all classifiable food plasmids.

Compared to Inc group representation, PTU representation across the two compartments showed a wider range of representation bias, with a 57% increase in the standard deviation of the representation ratios (Tables 1A–B). Three PTUs stood out for their overrepresentation in food plasmids: PTU-HI1A, -E50, and -E52 (5–8.7-fold enrichment) and two PTUs were overrepresented in clinical strains: PTU-E9 and -I1 (2–2.9-fold) (Table 1B).

Based on promiscuity, PTUs fall into six categories or “grades”, denoted by Roman numerals. Grade I represents promiscuity restricted to the level of species, II to the level of genus, III to family, IV to order, and V to class. Figure 4A shows the host range for the PTUs found in food plasmids. Note that out of the 79 plasmids, 25.3% have predicted promiscuity up to the level of family, 7.6% to the level of order, and 2.5% going all the way to the level class.

A seminal study published in 2020 (16), mapped the presence of PTUs across *Enterobacteriaceae*. Figure 4B shows other bacterial genera identified as primary hosts for the PTUs found in *E. coli* food plasmids. These include eight genera of Enterobacteriaceae,

with *Salmonella* and *Klebsiell*a being the best represented ones (13 and 10 PTUs, respectively). Overall, the PTU analysis confirms the very high potential for horizontal transfer of these food-borne plasmids.

3.5 Plasmid ARG cargo

Known ARGs were identified in our plasmids using ResFinder (42). We found 175 ARGs belonging to at least 14 different antibiotic resistance families. The results are summarized in Table 2 and listed in Table S3. All families were similarly represented in both plasmid populations, with β-lactamases, aminoglycosides, fluoroquinolones, and trimethoprim ranking top, in this order. Only two families were found exclusively in clinical plasmids, namely fosfomycin resistance genes, and efflux pump encoding variants (Table 2).

The average number of ARGs per plasmid was higher in clinical than in food plasmids, but the difference corresponded almost completely to a smaller fraction of ARG-free plasmids (46.2%, compared to 65.9 % for food plasmids) (Figure S4). Of the 175 ARGs identified, only 51 genes overlapped between the two plasmid populations. Eleven ARG genes were unique to food plasmids (listed in Table 2) and 113 were exclusive to clinical plasmids. The observed gene overlap highly deviates from what would be expected from a random distribution between the two groups of plasmids, even taking into account the difference in sample size between the two groups (Table S4) ($p<0.5$). Relative to other antibiotic resistance families, β-lactamases and plasmid-mediated fluoroquinolone resistance genes exhibit an even lower degree of overlap (15–18%, compared to an average overlap of 27.2% for the four other families exhibiting an n>10).

3.6 Linkage between ARG cargo, Inc groups, and PTUs

Next, we examined the linkage between ARGs and specific Inc groups comparing food and clinic plasmids (Supplementary figures S5 and S6). Only three Inc groups have a resistance cargo that is overrepresented in food plasmids than expected based on their total of plasmids: IncR, IncH1B, and IncF(F1:A$^-$:B$^-$). IncR and IncH1B are also overrepresented in food plasmids, which explains this higher contribution, but not IncF(F1:A$^-$:B$^-$). The latter observation implies that this IncF plasmid subtype exhibits a differentially high cargo of ARGs relative to clinical plasmids.

We also looked at the overlap between individual ARGs found in food and clinical plasmids for each Inc group by calculating the Jaccard index (JI) for resistance (Table 1A, see Methods). The results for Inc types that have at least three plasmids in food and three plasmids in the clinic are shown in Table 1A. The JI values for resistance range between 0 and 0.18, with an average of 0.07. The highest values corresponded to IncR, IncF (F$^-$:A$^-$:B1), and p0111 (all JI=0.18), IncX1 (JI=0.17), and IncI1–1 (JI=0.14). This seems to suggest that the exchange of genes between the food and clinical compartments may be channeled through a specific Inc groups.

The Inc group classification groups plasmids by distinct modes of replication that do not necessarily reflect phylogenetic relatedness. To establish the association between ARG cargo and phylogenetically-related clusters of plasmids, we also calculated the JI for individual PTU types Table 1B. In this case, while the average JI stayed the same, the highest values

went up considerably (up to 0.38). The three PTUs with the highest values were PTU-E50 (JI=0.38), PTU-HI2 (0.24), and PTU-FE (JI=0.23); note that this increase in maximal values does not correspond to a reduction in sample size. The increase in maximal JI values for resistance in PTUs confirms the idea that the ARG exchange between food and clinical strains is dominated by a specific cluster of phylogenetically-related plasmids.

Some PTUs involve multiple Inc groups. To obtain additional granularity in our analysis, we looked at all the PTU-Inc group combinations for plasmids that are represented in both compartments; the results are shown in Figure 5. Given the large number of PTU-replicon combinations, the average sample size went down substantially. The plasmid representation in each compartment is shown in **panel c** of Figure 5, as it needs to be taken into consideration because a very low sample size in one of the compartments reduces the significance of the JI and inflates its value. Based on these considerations, the following PTU-replicon combinations appear to have genuine overlap in ARG gene content: PTU-B/O/K/Z Inc B/O/K/Z (JI=0.09, with 4 plasmids for food and 43 for the clinic), PTU-HIA IncHI1B (JI=0.12, with three plasmids in each compartment) and PTU-HI2 IncHIA (JI=0.2, with 4 and 21 plasmids in each compartment, respectively). Within IncF plasmids, the PTU-FE IncF(F^-:A^-:B1) combination stands out (JI=0.21, with 7 food plasmids and 11 clinic plasmids) (Figure S6). Supplemental figure S7 shows the dendrograms for the plasmids included in these PTU-Inc group combination, and in all cases, we see that plasmids found in food are closely related to plasmids found in the clinic, supporting the idea of a direct exchange between the two compartments. The overlapping genes found in plasmids belonging to these PTU-Inc categories can be seen in (Table 3). The genes *aph(6)-Id, aac(3)-IId, aadA2, bla*CTX-M-55, *bla*TEM-1, *sul1, sul2, sul3* and *tet(A)* stands out because it is found in two of these categories.

3.7 Presence of ARGs of particular interest in food plasmids

We also looked for the presence of ARG of particular clinical or epidemiological interest in our food plasmids. Among β-lactamases, we detected several ESBL genes conferring resistance to synthetic cephalosporins in food plasmids, notably $bla_{\text{CTX-M-15,55,65}}$, $bla_{\text{TEM-135,214,215}}$, and bla_{CMY2}, all of which were also found in clinical isolates. The ESBL $bla_{\text{SHV-12}}$ and the β-lactamases $bla_{\text{LAP-2}}$ and $bla_{\text{HERA-3}}$ however, were found exclusively in food plasmids (Tables 2, S3), in agreement with recent reports of their presence in the food chain (60,61). Strikingly, in food plasmids, the widest variety of ESBL resistance genes was mainly found in IncHI (A, 1B, 2A) and IncF plasmid groups (Table S5). While plasmid-borne ESBL resistance has already been widely reported in *E. coli* in livestock and food derived from it, it was previously associated with a broader range of Inc types (29). This exclusive association with two Inc groups that we see in our food plasmids is unlike clinical plasmids, where ESBL β-lactamases are associated found with almost all plasmid types and most prominent in ColRNA, IncF, IncI, IncN, IncX, IncH, NT, and p0111 (Table S5). The Inc group promiscuity for β-lactamases observed in clinical plasmids may have facilitated the β-lactamase diversification mentioned above.

We would like to point out that aminoglycoside resistance genes were widely prevalent in both food and clinical isolates. Approximately 70% of the genes were shared in

both populations. In food plasmids, we found three classes of aminoglycoside-modifying enzymes: acetyltransferases (aac), phosphoryltransferases (aph) and adenyltransferases (aad), with *aph(3")-lb* (21.8%), *aadA1* (20.0%), *aadA2* (18.1%), *aph(6)-lb* (18.1%) and *aac (3)-lld* (16.3%) being the mainly detected genes. No nucleotidyltransferases (ant) nor the *rmtB, rmtC* and *armA* genes were found. The following five genes: *acc(3)-IIa*, *aac(3)-IIg*, *aac(6')-IIc*, *aadA17*, and *aadA24* were present only in food (Table 2). Aminoglycoside resistance gene cargo was carried mainly in IncF plasmids in both groups.

Carbapenems are used exclusively in the human clinic; however, there are increasing reports of carbapenem-resistant *E. coli* isolates outside hospital environments. Isolates producing the NDM-5 cabapenemase have previously been reported in food animals such as cows, poultry, and swine but these were mainly carried by IncX3 plasmids; (62–64). Here we report a $bla_{\text{NDM-5}}$ gene on an IncF ($F1:A^{-}:B^{-}$) plasmid (pF070, AP023237. 1) isolated from a pork sample; this plasmid was already reported in an isolate from a Japanese resident (65). Figure S8 compares the genetic layout of the 26 plasmids (food and clinical) that carry NDM-5 in our study. This includes six IncX3 plasmids of clinical origin and a variety of IncF subtypes associated with the same PTU (PTU-FE). In terms of their genetic environment, all our NDM-5-bearing plasmids fall into four different categories (a to d) based on the analysis of gene structures, mainly the presence of specific ISs in the vicinity, but also the presence or absence of the following key genes: ble_{MBL} (bleomycin resistance gene), *trpF* (phosphoribosylanthranilate isomerase), *dsbD* (disulfide interchange protein) and of hypothetical proteins of different lengths (see Methods).

Colistin is a polymyxin antibiotic widely used in veterinary medicine and for which plasmid-mediated-resistance has recently emerged (66). This development is of particular concern, since colistin is the antibiotic currently used as a last resort (67). In our food plasmids, we found two plasmid-mediated colistin resistance genes: *mcr-1.1*, and *mcr-9.1* carried in IncHIA (both) and IncI plasmids (*mcr-1.1* only) (Table S5 and Figure S9). While a greater variety of types were found associated with *mcr-1.1* in clinical plasmids, IncHIA and IncI plasmids still represented 79% of all the plasmids bearing one of these genes, confirming the strong association between *mcr-1.1*, and *mcr-9.1* and these two replicons. This observation is consistent with previous reports from around the world (68–70). One of the PTU Inc-group combinations with high resistance JI (PTU-HI2 IncHIA) was associated with *mcr-1.1* in food and clinic plasmids, suggesting that group of phylogenetically-related plasmids likely acts as facilitator of horizontal gene transfer of colistin resistance between food and the clinic.

3.8 Plasmid virulence cargo

We identified 72 known virulence genes in our complete plasmid dataset. The average number of virulence genes per plasmid was higher in clinical plasmids (4 virulence factors on average, compared to 3 for food plasmids), but the overall distribution of number virulence genes/plasmid was similar (no statistical difference, $p>0.5$). Of these 72 virulence genes, 33 overlapped between the two plasmid pools and only five of them were unique to food plasmids. Interestingly, these five genes (*afaA, afaB, afaC, afaE8*, and *fim41a*) are all involved in adhesion; adhesins have been reported to facilitate the colonization of the small

intestine during infection in DAEC (Diffuse-Adherent *Escherichia coli*) and uroepithelium in UPEC infections (Uropathogenic *Escherichia coli* strains) (71,72). While most proteins that are present on the bacterial cell surface are encoded chromosomally (typically in genomic islands), they have also been reported on plasmid DNA (73). In our study, strain 18SC04V04-EC (CP063739.1) showed characteristics of a STEC/ExPEC heterohybrid, as it harbors two Stx1 variants, and a plasmid (pVPS18EC0467–2; CP063740.1) carrying virulence genes (*afaABCDE, cdtB, espP, iha, iucC, iutA, ompT and traT*); this heterotype has already been reported in several clinical patients and production animals (74,75) but not in food. These findings increase the concerns of HGT of virulence factors in the food chain and highlight the need for a more extensive analysis of plasmids in food strains with ST and non-classical serotypes or pathotypes of human infections.

Table S6 shows the virulence genes found in food and clinical plasmids, ranked by frequency. The most frequent virulence genes shared between the two populations are *traT, sitA, iucC, ompT,* and *hlyF. SitA* and *iucC* have been reported to play a role in iron sequestration and regulation of metabolism in ExPEC strains (73). The gene *ompT* encodes a well-characterized integral membrane endopeptidase described as participating in the adhesion to host epithelial cells in ExPEC, APEC, and DAEC strains (76–78); *hlyF* is a hemolysin contributing to ExPEC virulence by regulating toxin release during the establishment of infection (79,80). Finally, *traT* is a gene encoding for an outer membrane lipoprotein whose expression has been linked to enhanced serum resistance, although the mechanism has not been clearly described (81). Two genes appear to be enriched in food plasmids relative to clinical plasmids *espP*, and *ehxA*. The gene *espP* (along with *etpD* and *katP*) has been previously described in plasmid pO157 associated mainly with EHEC and ETEC isolates (82). The enterohemolysin *ehxA* has been cataloged of importance in STEC strains (83).

3.9 Linkage between virulence cargo, Inc groups, and PTUs

We looked at the linkage of virulence cargo to specific Inc groups and their relative presence in food plasmids (Supplementary figures S5 and S6). Virulence genes are overrepresented in IncR, IncH1B, both of which also had an overrepresentation of ARG genes in food. To a lesser degree, virulence factors are also overrepresented in food plasmids classified in the following Inc groups: ColRNA, IncB/O/K/Z, IncHI2, IncY and IncF(F23:A^{-}:B3).

We also looked at the association of virulence factors with Inc groups, grouped by functional categories (Figure 6). We noted a stronger association of virulence factors with specific replicons in food plasmids, compared with clinical plasmids. The majority of adhesins, which as we mentioned above are exclusive to the food plasmids in our database, exhibit a striking association with IncF (subtypes F-:A-:B- and F74:A-:B-) replicons. Virulence factors involved in iron metabolism, by contrast, are associated with a wide range of replicons.

To determine the overlap between virulence factors found in food and clinical plasmids for individual Inc groups, we calculated a JI for virulence (see Methods). The results for Inc types that have at least three plasmids in food and three plasmids in the clinic are shown in Table 1A. JI values range between 0 and 0.45, with an average of 0.133. Three Inc types

stand out: IncF(F23:A⁻:B3)(JI=0.45), IncF(F⁻:A⁻:B⁻) (JI=0.40), and IncB/O/K/Z (JI=0.33). Note that, compared to ARGs, virulence factors produce much higher JI values. This is at least partially due to the reduced genetic pool involved (72 genes, compared to 175).

We also calculated the JI for virulence cargo for individual PTU types Table 1B and the highest values went up even further (up to 0.5). The PTUs that stand out for their high values were PTU-HI2 (JI=0.5), PTU-E5 (0.45), PTU-FE (JI=0.41) and PTU-B/O/K/Z (JI=0.36). As in the case of ARGs, this result suggests that certain PTUs dominate the genetic exchange of virulence factors between food and clinic strains although the fact that only two of the PTUs overlap between the ARG and virulence (PTU-FE and PTU-HI2) suggests some specialization for the HGT of virulence factors.

To increase granularity in our analysis, we determined the virulence JI for Inc groups within individual PTUs, the results are shown in Figure 5. The plasmid representation in each compartment is shown in **panel c** of Figure 5 to give a sense of the significance of the JI. Based on these considerations, the following PTU-replicon combinations appear to have genuine overlap in gene content between food and clinical plasmids: PTU-HI2 IncHIA (JI=0.5, with 4 plasmids for food and 21 for the clinic), PTU-E69 Inc B/O/K/Z (JI=0.44, with 2 plasmids for food and 22 for the clinic), PTU-B/O/K/Z Inc B/O/K/Z (JI=0.36, with 4 plasmids for food and 43 for the clinic), and non-PTU ColRNAI (JI=0.33, with 3 and 12 plasmids in each compartment, respectively). Within IncF plasmids, the following PTU-Inc combinations stand out: PTU-E41 IncF(F-:A-:B15) (JI=0.5, with 2 food plasmids and 9 clinical plasmid), PTU-E5 IncF(F23: A⁻:B3) (JI=0.45, with 15 food plasmids and 56 clinical plasmids), and PTU-FE IncF(F⁻:A⁻:B1) (JI=0.44, with 7 food plasmids and 11 clinic plasmids). Supplemental figure S7 shows the dendrograms for plasmids belonging to five of these PTU-Inc group combinations, and in all cases we see that plasmids found in food are more closely related to plasmids found in the clinic than to each other, strongly supporting a direct plasmid exchange between the two compartments.

The specific virulence factors that are found to be overlapping in these PTU-Inc categories exhibiting high JI indexes are listed in (Table 3). The enterohemolysin gene *ehxA* stands out because it is found in five out of the 8 categories, and it is one of the virulence genes that is overrepresented in food plasmids. This observation suggests that the enrichment for this virulence factor in food is likely facilitated by horizontal transfer via plasmids.

Note that if there is plasmid-mediated HGT between food and clinical strains, some of the vectors (particularly large ones) would be expected to transfer resistance and virulence genes in the same HGT event. Indeed, if we look at our PTU-Inc group analysis, three of the eight PTU-Inc combinations of interest overlap, notably: PTU-B/O/K/Z Inc B/O/K/Z (JI=0.09 for resistance, 0.36 for virulence), PTU-HI2 IncHIA (JI=0.2 for resistance, 0.5 for virulence), and PTU-FE IncF(F⁻:A⁻:B1) (JI=0.21 for resistance and 0.44 for virulence). Other PTU-Inc combinations with substantial JI for both resistance and virulence include: PTU-HI1A IncHI1B (JI=0.11 for resistance, 0.25 for virulence, 3 plasmids in each compartment) and PTU-I1 IncC I1-I (JI=0.11 and 0.125, with 4 food plasmids and 100 clinical plasmids) (Figure 5, Table 4).

4. Discussion

This article analyzes the genomic sequence of 77 food *E. coli* isolates (with 161 plasmids); for comparison, we also included 649 clinical isolates (with 1624 plasmids) in our analysis. Admittedly, the two groups of genomes that are being were not matched for geographic location or timing so we cannot rule out that the differences origin of the samples. The ten-fold difference in sample size between food and clinical strains is not ideal either. It means that the degree to which features from food samples do not overlap with clinical samples is likely an underestimate, whereas the features from clinical samples that do not overlap with food samples are likely overestimated. Having said that, observed overlaps are still informative and so are the genetic associations between genes and Inc groups and/or PTUs because these represent intrinsic, genetically-based properties of the plasmids. Most of our observations fit well with and extend previous reports from the literature.

The main goal of this study was to shed light on the role that *E. coli* strains contaminating foods play in human disease. This is a topic that is actively being debated, with some reports supporting the idea of a direct transmission of pathogenic strains via food (84,85)and other reports highlighting the role of food strains as reservoirs for resistance and virulence genes (86,87) and demonstrating the presence of plasmids harboring resistance and virulence genes in food strains worldwide (69,88).

This article takes a rigorous genomic approach to address this question. We found that the vast majority of the strains isolated in food are more closely related to each other than to clinical strains, and that there is very little overlap in sequence types between the two compartments (exception of ST11). We also provide very strong evidence of gene exchange between plasmids found in the food and the clinic. Taken together, these two observations strongly suggest that plasmids found in strains of *E. coli* contaminating food represent reservoirs of resistance and virulence genes and facilitate their genetic exchange with clinical strains via HGT. These genetic information findings complement studies comparing antibiotic-resistant clinical strains, their mobile genetic elements isolated in the clinic versus strains from food(89,90) and support literature highlighting niche specificity (91) and the critical role of horizontal transfer for the spread and maintenance of antibiotic resistance and virulence more generally (92,93).

We report a large number of resistance and virulence genes in food plasmids. Selective pressures to retain resistance genes in food likely reflects the use of antibiotics as prophylactics, metaphylactics and growth enhancers for food production animals and the subsequent use of manure as a fertilizer for crops (94) or contamination during meat processing (95). We found the two resistance genes of high clinical relevance in food plasmids: *mcr 1.1* and *mcr 9.1* (colistin resistance) and *blaNDM5* (carbapenem resistance). Pressures to retain virulence genes are likely linked to survival in the host animal (in case of meat and dairy products) or to remaining attached to food surfaces, which could be involved in the adaptation to the food environment. The high number of virulence genes found in food plasmids and the great enrichment for adhesion factors in food plasmids that we found, are remarkable.

Here we also report a general landscape for overlaps in the representation of virulence and resistance genes across food and clinical plasmids. The observed overlap suggests that some of the selective pressures operating are shared in both food and the clinic. The overlap is larger for virulence factors, possibly because of more rapid acquisition of resistance genes in the clinic, which increases the number of resistance genes found in clinic plasmids and not in food plasmids. This is likely the result from the more consistent selective pressures that these strains are undergoing in the clinic, where antibiotics are routinely administered for prophylaxis and therapy.

The observed overlap in resistance and virulence genes is enriched in specific Inc groups, further enriched when we look at PTUs, which represent *bona fide* phylogenetically-related units, and even further enriched when we look at specific PTU-Inc group combinations (Figures 5, S5 and S6). In addition, food plasmids cluster with clinical plasmids within the PTU-Inc group combinations exhibiting overlaps in resistance and virulence genes (Figure S7). These observations strongly suggest plasmid-based genetic exchange of resistance and/or virulence cargo across the two ecological compartments (food and clinic). Given that we showed a clear segregation of host strains by source, we have to conclude that the apparent genetic exchange between food and the clinic is the result of HGT.

The observed overlap in resistance and virulence genes is enriched in specific Inc groups, further enriched when we look at PTUs, which represent *bona fide* phylogenetically-related units, and even further enriched when we look at specific PTU-Inc group combinations (Figures 5, S5 and S6). Consistent with these conclusions, we also showed that food plasmids have a high mobilization potential (Figure S7). This conclusion is based on the presence of complete sets of conjugation genes (over 1/3 of food plasmids are predicted to be competent for conjugation or mobilization) and on their PTU classification, with 25% of food plasmids having a predicted promiscuity at least at the level of family. Most conjugative plasmids are large (25–200 kb) in both plasmid populations. The upper size limit could be related to the energetic cost of megaplasmid transfer (96). By contrast, a substantial fraction (26.6%) of mobilizable plasmids is made up of small very plasmids (<20 kb). Note that small plasmids tend to have a high copy number. The trade-off would involve two different types of access to genetic diversity, namely mutagenesis (in multicopy plasmids), which allow genetic drift by retaining copies with a wild-type sequence (97) or access to the larger gene pool provided by other strains/genus/family/order/class/phyla (conjugation), producing repositories of genes captured from diverse clinical and environmental sources. Indeed, the pool of ARGs in clinical *E. coli* is known to derive from the environmental resistome (89,90).

The specific associations with Inc groups and PTUs are summarized in Table 4. Four Inc groups stand out for their overrepresentation in plasmids, namely IncR, IncHI1B, IncF(F^-:A^-:B1), and IncF(F23:A^-:B3). Plasmids associated with these Inc groups carry a high cargo of resistance and virulence genes (IncR, IncHI1B, IncF(F^-:A^-:B1)) or only virulence genes (IncF(F23:A^-:B3)). IncR and IncF(F^-:A^-:B1) plasmids also carry a high proportion of resistance genes that overlap between food and clinic plasmids (Table 4). An IncF (F^-:A^-:B^-) plasmid has previously been described as one of the most frequent BLEE-producing *E. coli* from community-acquired urinary tract infections and sewage

(98,99), while, IncR plasmid is increasingly described as a multidrug resistance reservoir in different locations of the world (100,101). In the case of IncHI1B and IncF(F23:A⁻:B3), the overlap is restricted to virulence. In this case, IncHIB, not only has been reported in *E. coli*, but also in other enterobacteria such as *Klebsiella pneumoniae* and *Salmonella enterica* from patients and animals, mainly carrying resistance genes against beta-lactams (92,102,103). Other Inc groups with high levels of overlap include Col156 (virulence), IncB/O/K/Z (resistance and virulence), IncC (resistance and virulence) Inc I1-I (resistance and virulence), IncX1 (resistance), and p0111 (resistance). These associations are congruent with previous analyses performed showing associations between Inc groups and replicons harboring clinically important resistance genes (104–106).

The present work, through the use of COPLA, goes much further in establishing the phylogenetic relatedness between the plasmids involved and their potential for mobilization beyond *E. coli*.

We find three PTUs that are exclusively found in food plasmids (PTU-E46, PTU-E54, PTU-E79). This is remarkable given the almost 10-fold difference in plasmid size going against food plasmids. Three additional PTUs are also substantially overrepresented in food plasmids: PTU-HI1A, PTU-E50, and PTU-E52. Note that PTU-E50 is associated with a broad host range (Order), which could confer an adaptive advantage.

IncHI1B is associated with PTU HIA and with virulence and resistance cargo, IncF(F⁻:A⁻:B1) with PTU-FE (resistance and virulence), IncF(F23:A⁻:B3) with PTU-E5 (virulence), Col156 with PTU-E7 (virulence), IncB/O/K/Z with PTU- B/O/K/Z (resistance and virulence) and with PTU-E69 (virulence), Inc I1-I with PTU-I1 (resistance) (Table 4). Other PTU-combinations with significant JI indexes and therefore candidates to mediate horizontal transfer between food and clinical strains are: PTU-FE with IncF(F2:A⁻:B⁻) (resistance and virulence) and with IncF(F⁻:A⁻:B⁻) (virulence), PTU-HI2 Inc HIA (resistance and virulence), PTU-E41 IncF(F⁻:A⁻:B15) (virulence) (Table 4). There are also a few Inc groups associated with Non-PTU plasmids (Figure S10). While there is a substantial overlap between the PTU-Inc group combinations showing a substantial overlap for both resistance and virulence genes, often we see only overlap for one type of cargo. This is particularly true for virulence factors, with five PTU-Inc group combinations showing a strong preference for virulence genes (Figure S10). Note that two of these combinations (the ones with PTU-E7 Col156 and Non-PTU ColRNA) represent very small plasmids of generally less than 10,000 bp. In this case, size limitations may have driven their specialization for virulence factors.

Our observations highlight the importance of collecting genomic sequences of food isolates from all over the world for epidemiological surveillance and have both basic and translational relevance. The biological significance of Inc groups is still unclear. Given the clear phylogenetic segregation of food and clinical strains, the preferential association of specific Inc groups with food plasmids suggests some niche-dependent host preference. The same is true for the even stronger preference seen for PTUs, which suggests that PTU evolution is linked to the evolution of the host strain.

The high density of virulence genes found in food plasmids (75% that of clinical plasmids), and the high representation of adhesion factors are also of biological relevance and begs the question of what their role is for the maintenance of the host strains in the food chain. Also, one would like to know whether these observations are exclusive to *E. coli* or whether they also apply to other Gram-negatives that frequently contaminate foods such as *Salmonella*. On the translational side, the identification of specific PTU-Inc group combinations that concentrate the majority of the cargo overlap between food and clinic plasmids should facilitate the design of strategies for epidemiological surveillance of *E. coli* contaminating foods. It is important to acknowledge, though, that the movement of genes facilitated by MGE and recombination will diminish the specificity of any association between resistance and virulence genes and specific replicons or PTUs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank Miguel Castañeda Lucio, Margarita María de la Paz Hernández Arenas and Candelario Vázquez Cruz from “Centro de Investigaciones en Ciencias Microbiológicas, Instituto de Ciencias” of Benemérita Universidad Autónoma de Puebla, for their invaluable contribution, guidance and support to the development of this work. We would also like to thank María Pilar Garcillán-Barcia (University of Cantabria, Spain) for pointing us to COPLA.

Funding information

This work was supported by the Consejo Nacional de Ciencia y Tecnología (CONACyT) [CB-2017-2018/A1-S-22136]. The CONACyT also provided a fellowship [589143 to MGBA]. We are also grateful for the UC MEXUS-CONACYT Postdoctoral Fellowship 2017-2019 and the CONACYT National Postdoctoral Fellowship across “Convocatoria de Estancias Posdoctorales por México 2020” during the development of the research to GCC [335026]. JK was supported by CITRIS Seed Funding proposal 2015-324 and by NIAID award 1R41AI122740-01A1 in partnership with Maverix Genomics.

Abbreviations

MLST	Multilocus Sequence Typing
STs	sequence types
MGEs	Mobile Genetic Elements
TA	Toxin-antitoxin module
MDR	multidrug resistance
ssDNA	single-stranded DNA
MOB	mobilization module
PTU	Plasmid Taxonomic Unit
PSK	Post-Segregational Killing
HUS	Hemorrhagic Uremic Syndrome

Inc	Incompatibility Group
Par	partition gene
OriT	transfer origin of replication
JI	Jaccard index

References

1. Santos AC de M, Santos FF, Silva RM, Gomes TAT. Diversity of Hybrid- and Hetero-Pathogenic Escherichia coli and Their Potential Implication in More Severe Diseases. Front Cell Infect Microbiol. 2020 Jul;0:339. DOI: 10.3389/fcimb.2020.00339.
2. Baquero F, Tedim AP, Coque TM. Antibiotic resistance shaping multi-level population biology of bacteria. Front Microbiol. 2013;4(MAR). DOI: 10.3389/fmicb.2013.00015.
3. Palmer AC, Kishony R. nderstanding, predicting and manipulating the genotypic evolution of antibiotic resistance. Nat Rev Genet. 2013 Apr;14(4):243. DOI: 10.1038/nrg3351 [PubMed: 23419278]
4. Hughes D, Andersson DI. Evolutionary trajectories to antibiotic resistance. Annu Rev Microbiol. 2017;71:579–96. DOI: 10.1146/annurev-micro-090816-093813. [PubMed: 28697667]
5. Baquero F Martínez JL F. Lanza V Rodríguez-Beltrán J, Galán JC San Millán A, Cantón R Coque TM Evolutionary pathways and trajectories in antibiotic resistance. Clin Microbiol Rev. 2021 Oct, 34(4):1–72. DOI: 10.1128/cmr.00050-19.
6. Fratamico PM, DebRoy C, Liu Y, Needleman DS, Baranzoni GM, Feng P. Advances in molecular serotyping and subtyping of Escherichia coli. Frontiers in Microbiology. Frontiers Media S.A.; 2016 May. 7:644. DOI: 10.3389/fmicb.2016.00644. [PubMed: 27199968]
7. Clermont O, Gordon D, Denamur E. Guide to the various phylogenetic classification schemes for *Escherichia coli* and the correspondence among schemes. Microbiology. 2015;161(5):980–8. DOI: 10.1099/mic.0.000063. [PubMed: 25714816]
8. Nicolas-Chanoine M-H, Bertrand X, Madec J-Y. *Escherichia coli* ST131, an intriguing clonal group. Clin Microbiol Rev. 2014;27(3):543. DOI: 10.1128/cmr.00125-13. [PubMed: 24982321]
9. Mathers AJ, Peirano G, Pitout JDD. *Escherichia coli* ST131: The Quintessential Example of an International Multiresistant High-Risk Clone. Adv Appl Microbiol. 2015 Jan 1;90:109–54. DOI: 10.1016/Bs.aambs.2014.09.002. [PubMed: 25596031]
10. Carroll AC, Wong A. Plasmid persistence: costs, benefits, and the plasmid paradox. Canadian Journal of Microbiology. 2018. 64:293–304. DOI: 10.1016/BS.aambs.2014.09.002 [PubMed: 29562144]
11. Virolle C, Goldlust K, Djermoun S, Bigot S, Lesterlin C. Plasmid transfer by conjugation in Gram-negative bacteria: From the cellular to the community level. Genes (Basel). 2020 Nov 1;11(11):1–33. DOI: 10.3390/genes11111239.
12. Lehtinen S, Huisman JS, Bonhoeffer S. Evolutionary mechanisms that determine which bacterial genes are carried on plasmids. Evol Lett. 2021 Jun 1;5(3):290. DOI: 10.1002/evl3.226. [PubMed: 34136276]
13. Carloni E, Andreoni F, Omiccioli E, Villa L, Magnani M, Carattoli A. Comparative analysis of the standard PCR-Based Replicon Typing (PBRT) with the commercial PBRT-KIT. Plasmid. 2017 Mar 1;90:10–4. DOI: 10.1016/j.plasmid.2017.01.005. [PubMed: 28137396]
14. Garcillán-Barcia MP, Francia MV, de La Cruz F. The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiol Rev. 2009;33(3):657–87. DOI: 10.1111/j.1574-6976.2009.00168.x. [PubMed: 19396961]
15. Redondo-Salvo S, Bartomeus-Peñalver R, Vielva L, Tagg KA, Webb HE, Fernández-López R, de la Cruz F. COPLA, a taxonomic classifier of plasmids. BMC Bioinforma 2021 221. 2021 Jul 31;22(1):1–9. DOI: 10.1186/S12859-021-04299-X.
16. Redondo-Salvo S, Fernández-López R, Ruiz R, Vielva L, de Toro M, Rocha EPC, Garcillán-Barcia MP, de la Cruz F. Pathways for horizontal gene transfer in bacteria revealed by a global

map of their plasmids. Nat Commun. 2020 Dec;11(1):1–13. DOI: 10.1038/s41467-020-17278-2. [PubMed: 31911652]

17. Cabezón E, Ripoll-Rozada J, Peña A, de la Cruz F, Arechaga I. Towards an integrated model of bacterial conjugation. FEMS Microbiol Rev. 2015 Jan;39(1):81–95. DOI: 10.1111/1574-6976.12085. [PubMed: 25154632]
18. Werisch M, Berger U, Berendonk TU. Conjugative plasmids enable the maintenance of low cost non-transmissible plasmids. 2017.91:96–104. DOI: 10.1016/j.plasmid.2017.04.004.
19. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile genetic elements associated with antimicrobial resistance. Clinical Microbiology Reviews. American Society for Microbiology; 2018 Oct.31:1–63. DOI: 10.1128/cmr.00088-17.
20. Poirel L, Madec J-Y, Lupo A, Schink A-K, Kieffer N, Nordmann P, Schwarz F. Antimicrobial resistance in Escherichia coli. Microbiol Spectr.2018 Jul.6(4):1–32. DOI: 10.1128/microbiolspec.arba-0026-2017.
21. Bakour S, Sankar SA, Rathored J, Biagini P, Raoult D, Fournier P-E. Identification of virulence factors and antibiotic resistance markers using bacterial genomics. 2016 Mar;11(3):455–66. DOI:10.2217/fmb.15.149.
22. Munson GP. Virulence regulons of enterotoxigenic *Escherichia coli*. Immunol Res. 2013 571. 2013 Nov;57(1):229–36. DOI: 10.1007/s12026-013-8453-4. [PubMed: 24203442]
23. Ramirez MS, Traglia GM, Lin DL, Tran T, Tolmasky ME. Plasmid-mediated antibiotic resistance and virulence in Gram-negatives: the Klebsiella pneumoniae paradigm. Microbiol Spectr. 2014 Sep;2(5):PLAS-0016–2013. DOI: 10.1128/microbiolspec.plas-0016-2013.
24. Guillard T, Pons S, Roux D, Pier GB, Skurnik D. Antibiotic resistance and virulence: Understanding the link and its consequences for prophylaxis and therapy. BioEssays. 2016 Jul;38(7):682–93. DOI: 10.1002/bies.201500180. [PubMed: 27248008]
25. Ludden C, Raven KE, Jamrozy D, Gouliouris T, Blane B, Coll F, de Gorffaus M, Naydenova P, Horner C, Hernandez-Garcia J, Wood P, Hadjiri N, Radacovic M, Brown N, Holmes M, Parkhill J, Peacock S. One health genomic surveillance of Escherichia coli demonstrates distinct lineages and mobile genetic elements in isolates from humans versus livestock. MBio. 2019 Jan 1;10(1). DOI: 10.1128/mBio.02693-18.
26. Nguyen MN, Hoang HTT, Xavier BB, Lammens C, Le HT, Hoang NTB, Nguyen S, Pham N, Goossens H, Dang A, Malhotra-Kumar S. Prospective One Health genetic surveillance in Vietnam identifies distinct blaCTX-M-harbouring *Escherichia coli* in food-chain and human-derived samples. Clin Microbiol Infect. 2021; 27(10). P1515.E1–1515.E8. DOI: 10.1016/j.cmi.2021.01.006.
27. Day MJ, Rodríguez I, van Essen-Zandbergen A, Dierikx C, Kadlec K, Schink A-K, Wu G, Chattaway M, DoNascimento V, Wain J. Helmuth R, Guerra B, Schwarz S, Threfall J, Woodward J, Coldham N, Mevius D, Woodford N. Diversity of STs, plasmids and ESBL genes among *Escherichia coli* from humans, animals and food in Germany, the Netherlands and the UK. J Antimicrob Chemother. 2016 May,;71(5):1178–82. DOI: 10.1093/jac/dkv485. [PubMed: 26803720]
28. Reid CJ, McKinnon J, Djordjevic SP. Clonal ST131-H22 *Escherichia coli* strains from a healthy pig and a human urinary tract infection carry highly similar resistance and virulence plasmids. Microb Genomics. 2019;5(9). DOI: 10.1099/mgen.0.000295.
29. Madec JY, Haenni M. Antimicrobial resistance plasmid reservoir in food and food-producing animals. Plasmid. 2018 Sep 1;99:72–81. DOI: 10.1016/j.plasmid.2018.09.001. [PubMed: 30194944]
30. Beghain J, Bridier-Nahmias A, Nagard H Le, Denamur E, Clermont O. ClermonTyping: An easy-to-use and accurate in silico method for Escherichia genus strain phylotyping. Microb Genomics. 2018 Jul 1;4(7):e000192. DOI: 10.1099/mgen.0.000192.
31. Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheutz F. Rapid and easy in silico serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. J Clin Microbiol. 2015 Aug 1;53(8):2410–26. DOI: 10.1128/jcm.00008-15. [PubMed: 25972421]
32. Robins-Browne RM, Holt KE, Ingle DJ, Hocking DM, Yang J, Tauschek M. Are Escherichia coli pathotypes still relevant in the era of whole-genome sequencing?. Frontiers in Cellular and

Infection Microbiology. Frontiers Media S.A.; 2016. 6;141. DOI: 10.3389/fcimb.2016.00141. [PubMed: 27917373]

33. Kaas RS, Leekitcharoenphon P, Aarestrup FM, Lund O. Solving the Problem of Comparing Whole Bacterial Genomes across Different Sequencing Platforms. PLoS One. 2014;9(8):104984. DOI: 10.1371/journal.pone.0104984.
34. Price MN, Dehal PS, Arkin AP. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix.2009 Jul; 26(7):1641–50 DOI: 10.1093/molbev/msp077.
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. 2016 Jul.;44 (W1): W242–5. DOI: 10.1093/nar/gkw290. [PubMed: 27095192]
36. Carattoli A, Zankari E, Garciá-Fernández A, Larsen MV, Lund O, Villa L, Aarestrup F, Hasman H. In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014;58(7):3895–903. DOI: 10.1128/aac.02412-14. [PubMed: 24777092]
37. Villa L, García-Ferná Ndez A, Fortini D, Carattoli A. Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. 2010 Dec.65(12):2518–29. DOI: 10.1128/aac.02412-14.
38. Li X, Xie Y, Liu M, Tai C, Sun J, Deng Z, Ou H. OriTfinder: A web-based tool for the identification of origin of transfers in DNA sequences of bacterial mobile genetic elements. Nucleic Acids Res. 2018 Jul 2;46(W1):W229–34. DOI: 10.1093/nar/gky352. [PubMed: 29733379]
39. Garcillán-Barcia MP, Redondo-Salvo S, Vielva L, de la Cruz F. MOBscan: Automated annotation of MOB relaxases. In: Methods in Molecular Biology. Humana Press Inc.; 2020. 295–308. DOI: 0.1007/978-1-4939-9877-7_21.
40. Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EPC, de la Cruz F. Mobility of Plasmids. Microbiol Mol Biol Rev. 2010 Sep;74(3):434. DOI: 10.1128/mmbr.00020-10. [PubMed: 20805406]
41. Hong Y, Guo M, Wang J. ENJ algorithm can construct triple phylogenetic trees. Mol Ther Nucleic Acid. 2021;23:286–93. DOI: 10.1016/j.omtn.2020.11.004.
42. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe, , Rebelo A, Florensa A, Fagelhauer L, Chakraborty T, Neumann B. ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother. 2020 Dec 1;75(12):3491–500. DOI: 10.1093/jac/dkaa345. [PubMed: 32780112]
43. Tetzschner AMM, 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. J Clin Microbiol. 2020 Oct 1;58(10). DOI: 10.1128/jcm.01269-20.
44. Seemann T Genome analysis Prokka: rapid prokaryotic genome annotation. 2014;30(14):2068–9. DOI: 10.1093/bioinformatics/btu153.
45. Johansson MHK, Bortolaia V, Tansirichaiya S, Aarestrup FM, Roberts AP, Petersen TN. Detection of mobile genetic elements associated with antibiotic resistance in Salmonella enterica using a newly developed web tool: MobileElementFinder. 2021 Jan. 76(1);101–109. DOI: 10.1093/jac/dkaa390.
46. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. 2011 Aug.12:408. DOI: 10.1186/1471-2164-12-402.
47. Clermont O, C Christenson JK, Denamur E, Gordon DM. The Clermont *Escherichia coli* phylo-typing method revisited: Improvement of specificity and detection of new phylo-groups. Environ Microbiol Rep. 2013 Feb 1;5(1):58–65. DOI: 10.1111/1758-2229.12019. [PubMed: 23757131]
48. Clermont O, Dixit OVA, Vangchhia B, Condamine B, Dion S, Bridier-Nahmias A, Denamur E, Gordon D. Characterization and rapid identification of phylogroup G in *Escherichia coli*, a lineage with high virulence and antibiotic resistance potential. Environ Microbiol. 2019 Aug 1;21(8):3107–17. DOI: 10.1111/1462-2920.14713. [PubMed: 31188527]
49. Clermont O, Bonacorsi S, Bingen E. Rapid and simple determination of the *Escherichia coli* phylogenetic group. Appl Environ Microbiol. 2000;66(10):4555. DOI: 10.1128/AEM.66.10.4555-4558.2000. [PubMed: 11010916]

50. Nojoomi F, Ghasemian A. The relation of phylogroups, serogroups, virulence factors and resistance pattern of *Escherichia coli* isolated from children with septicemia. New Microbes New Infect. 2019 May 1;29:100517. DOI: 10.1016/j.nmni.2019.100517. [PubMed: 31080621]
51. Oteo J, Cercenado E, Cuevas Ó, Bautista V, Delgado-Iribarren A, Orden B, Pérez-Vázquez M, Carcía-Cobos S, Campos J. AmpC β-lactamases in Escherichia coli: emergence of CMY-2–producing virulent phylogroup D isolates belonging mainly to STs 57, 115, 354, 393, and 420, and phylogroup B2 isolates belonging to the international clone O25b–ST131. Diagn Microbiol Infect Dis. 2010 Jul 1;67(3):270–6. DOI: 10.1016/j.diagmicrobio.2010.02.008. [PubMed: 20462723]
52. Page AV, Liles WC Enterohemorrhagic Escherichia coli Infections and the Hemolytic-Uremic Syndrome. Medical Clinics of North America. Elsevier; 2013 Jul. 97(4):681–95 DOI: 10.1016/j.mcna.2013.04.001. [PubMed: 23809720]
53. Bolukaoto JY, Kock MM, Strydom KA, Mbelle NM, Ehlers MM. Molecular characteristics and genotypic diversity of enterohaemorrhagic *Escherichia coli* O157:H7 isolates in Gauteng region, South Africa. Sci Total Environ. 2019 Nov 20;692:297–304. DOI: 10.1016/j.scitotenv.2019.07.119. [PubMed: 31351277]
54. Singer RS. Urinary tract infections attributed to diverse ExPEC strains in food animals: evidence and data gaps. Front Microbiol. 2015 Feb:4;6:28. DOI: 10.3389/fmicb.2015.00028. [PubMed: 25699025]
55. Yamaji R, Friedman CR, Rubin J, Suh J, Thys E, McDermott P, Hung F, Riley L. A Population-based surveillance study of shared genotypes of *Escherichia coli* isolates from retail meat and suspected cases of urinary tract infections. mSphere. 2018 Aug 29;3(4):e00179–18. DOI: 10.1128/msphere.00179-18. [PubMed: 30111626]
56. Muniesa M, Hammerl JA, Hertwig S, Appel B, Brüssow H. Shiga toxin-producing *Escherichia coli* O104:H4: A new challenge for microbiology. Applied and Environmental Microbiology. 2012 Jun. 78(12):4065–73. DOI: 10.1128/aem.00217-12. [PubMed: 22504816]
57. Bouet J-Y, Funnell BE. Plasmid localization and partition in Enterobacteriaceae. EcoSal Plus. 2019 Feb 6;8(2). DOI: 10.1128/ecosalplus.esp-0003-2019.
58. Novick RP. Plasmid incompatibility. Microbiol Rev. 1987;51(4):381. DOI: 10.1128/mr.51.4.381-395.1987. [PubMed: 3325793]
59. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, Mevius D, Hordjik. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. J Antimicrob Chemother. 2018;73(5):1121–37. DOI: 10.1093/jac/dkx488. [PubMed: 29370371]
60. Irrgang A, Zhao G, Juraschek K, Kaesbohrer A, Hammerl JA. Characterization of *E. coli* isolates producing extended spectrum beta-Lactamase SHV-Variants from the food chain in Germany. Microorg. 2021 Sep 10;9(9):1926. DOI: 10.3390/microorganisms9091926.
61. Tadesse DA, Li C, Mukherjee S, Hsu C-H, Jones SB, Gaines SA, Kabera C, Lonegaran G, Torrence M, Harhay D, McDermott P, Zhao S. Whole-Genome sequence analysis of CTX-M Containing *Escherichia coli* isolates from retail meats and cattle in the United States. Microb drug Resist. 2018 Sep 1;24(7):939–48. DOI: 10.1089/mdr.2018.0206. [PubMed: 30148698]
62. Tang B, Chang J, Cao L, Luo Q, Xu H, Lyu W, Qiang M, Ji X, Zhang Q, Xia X, Yang H. Characterization of an NDM-5 carbapenemase-producing Escherichia coli ST156 isolate from a poultry farm in Zhejiang, China; 2019 Apr. 19(1):82. DOI: 10.1186/s12866-019-1454-2.
63. He T, Wei R, Zhang L, Sun L, Pang M, Wang R, Wang Y. Characterization of NDM-5-positive extensively resistant *Escherichia coli* isolates from dairy cows. Vet Microbiol. 2017 Aug 1;207:153–8. DOI: 10.1016/j.vetmic.2017.06.010. [PubMed: 28757017]
64. Wang S, Cheng G, Yu Y, R-s Y, Z-y Q, X-p L, Kuang X, Zhang Y, Liu J, Yang R, Qiu Z, Sun J, Liao X, Liu Y. Molecular Epidemiology of New Delhi Metallo-β-Lactamase-Producing Escherichia coli in Food-Producing Animals in China. Front Microbiol. 2022;13:912260. DOI: 10.3389/fmicb.2022.912260. [PubMed: 35847092]
65. Sugawara Y, Hagiya H, Akeda Y, Takeuchi D, Sakamoto N, Matsumoto Y, Mootoka D, Nichi I, Tomono K, Hamada S. Community spread and acquisition of clinically relevant *Escherichia coli* harbouring blaNDM among healthy Japanese residents of Yangon, Myanmar. J Antimicrob Chemother. 2021 May 12;76(6):1448–54. DOI: 10.1093/jac/dkab070. [PubMed: 33758942]

66. Skov RL, Monnet DL. Plasmid-mediated colistin resistance (mcr-1 gene): three months later, the story unfolds. 2016;21(9):30155. DOI: 10.2807/1560-7917.es.2016.21.9.30155.

67. Ahmed MAE-GE-S, Zhong L-L, Shen C, Yang Y, Doi Y, Tian G-B. Colistin and its role in the Era of antibiotic resistance: an extended review (2000–2019). Emerg Microbes Infect. 2020 Jan 1;9(1):868–85. DOI: 10.1080/22221751.2020.1754133. [PubMed: 32284036]

68. Kim Y-J, Seo K-H, Kim S, Bae S. Phylogenetic Comparison and Characterization of an mcr-1-Harboring Complete Plasmid Genome Isolated from Enterobacteriaceae. 2022 Apr; 28(4):492–497. DOI: 10.1089/mdr.2021.0164.

69. Darphorn TS, Bel K, Koenders-van Sint Anneland BB, Brul S, Ter Kuile BH. Antibiotic resistance plasmid composition and architecture in *Escherichia coli* isolates from meat. Sci Rep. 2021;11(1):1–13. DOI: 10.1038/s41598-021-81683-w. [PubMed: 33414495]

70. Matamoros S, Van Hattem JM, Arcilla MS, Willemse N, Melles DC, Penders J, Nguyen V, Hoa N, Bootsma M, Van Genderen P, Goorhuis A, Grobusch M, Molhoek N, Lashof A, Stobberingh E, Verbrug H, De Jong M, Schultsz C. Global phylogenetic analysis of Escherichia coli and plasmids carrying the mcr-1 gene indicates bacterial diversity but plasmid restriction. 2017 Nov:7(1):15364 DOI: 10.1038/s41598-017-15539-7.

71. Ageorges V, Monteiro R, Leroy S, Burgess CM, Pizza M, Chaucheyras-Durand F, Desvaux M. Molecular determinants of surface colonisation in diarrhoeagenic Escherichia coli (DEC): from bacterial adhesion to biofilm formation. FEMS Microbiol Rev. 2020 May 1;44(3):314–50. DOI: 10.1093/femsre/fuaa008. [PubMed: 32239203]

72. Servin AL. Pathogenesis of Afa/Dr Diffusely Adhering *Escherichia coli*. Clin Microbiol Rev. 2005;18(2):264–92. DOI: 10.1128/cmr.18.2.264-292.2005. [PubMed: 15831825]

73. Sarowska J, Futoma-Koloch B, Jama-Kmiecik A, Frej-Madrzak M, Ksiazczyk M, Bugla-Ploskonska G, Choroszy-Krol I. Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: recent reports. Gut Pathog. 2019 Feb 21;11(1). DOI: 10.1186/s13099-019-0290-0.

74. De Rauw K, Thiry D, Caljon B, Saulmont M, Mainil J, Piérard D. Characteristics of Shiga toxin producing- and enteropathogenic *Escherichia coli* of the emerging serotype O80:H2 isolated from humans and diarrhoeic calves in Belgium. Clin Microbiol Infect. 2019 Jan 1;25(1):111.e5–111.e8. DOI: 10.1016/j.cmi.2018.07.023.

75. Cointe A, Birgy A, Mariani-Kurkdjian P, Liguori S, Courroux C, Blanco J, Delannoy S, Fach P, Loukiadis E, Bidet P, Bonacorsi S.Emerging multidrug-resistant hybrid pathotype shiga toxin–producing *Escherichia coli* O80 and related strains of clonal complex 165, Europe. Emerg Infect Dis. 2018;24(12):2262–9. DOI: 10.3201/eid2412.180272. [PubMed: 30457551]

76. Kukkonen M, Korhonen TK. The omptin family of enterobacterial surface proteases/adhesins: From housekeeping in *Escherichia coli* to systemic spread of Yersinia pestis. Int J Med Microbiol. 2004;294(1):7–14. DOI: 10.1016/j.ijmm.2004.01.003. [PubMed: 15293449]

77. Torres AN, Chamorro-Veloso N, Costa P, Cádiz L, Canto F Del, Venegas SA, Nitsche M, Coloma-Rivero R, Montero D, Vidal R. Deciphering Additional Roles for the EF-Tu, l-Asparaginase II and OmpT Proteins of Shiga Toxin-Producing Escherichia coli. 2020 Aug; 8(8):1184. DOI: 10.3390/microorganisms8081184.

78. Wolfgang MC, Anderson G, Hassan HM, Nolan LK, Barbieri NL, Vande Vorde JA. FNR Regulates the Expression of Important Virulence Factors Contributing to the Pathogenicity of Avian Pathogenic Escherichia coli. Front Cell Infect Microbiol | 2017;1:265. DOI: 10.3389/fcimb.2017.00265.

79. Gigliucci F; van Hoek Angela; Chiani Paola; Knijin Arnoldo;Minelli Fabio; Scavia Gaia; Franz Eelco; Morabito Stefano; Michelacci V. Genomic Characterization of hlyF -positive Shiga toxin – Producing Escherichia coli, Italy and the Netherlands, 2000 – 2019. 2019;27(3):1–6. DOI: 10.3201/eid2703.203110.

80. Murase K, Martin P, Porcheron G, Houle S, Helloin E, Pénary M, Nougayrède J, Philippe D, Hayashi T, Oswald E.HlyF Produced by Extraintestinal Pathogenic *Escherichia coli* Is a Virulence Factor That Regulates Outer Membrane Vesicle Biogenesis. J Infect Dis. 2016 Mar 1;213(5):856–65. DOI: 10.1093/infdis/jiv506. [PubMed: 26494774]

81. Koraimann G. Spread and Persistence of Virulence and Antibiotic Resistance Genes: A Ride on the F Plasmid Conjugation Module. EcoSal Plus. 2018 Feb 8;8(1). DOI: 10.1128/ecosalplus.ESP-0003-2018.
82. Yang SC, Lin CH, Aljuffali IA, Fang JY. Current pathogenic *Escherichia coli* foodborne outbreak cases and therapy development. Arch Microbiol. 2017;199(6):811–25. DOI: 10.1007/s00203-017-1393-y. [PubMed: 28597303]
83. Lorenz SC, Monday SR, Hoffmann M, Fischer M, Kase JA. Plasmids from shiga toxin-producing Escherichia coli strains with rare enterohemolysin gene (ehxA) subtypes reveal pathogenicity potential and display a novel evolutionary path. Appl Environ Microbiol. 2016;82(21):6367. DOI: 10.1128/aem.01839-16. [PubMed: 27542930]
84. Kluytmans JAJW, Overdevest ITMA, Willemsen I, Kluytmans-Van Den Bergh MFQ, Van Der Zwaluw K, Heck M, Rijnsburger M, Vandenbroucke-Grauls C, Savelkoul P, Johnston B, Gordon D, Johnson J.,Extended-Spectrum β-Lactamase–Producing *Escherichia coli* From Retail Chicken Meat and Humans: Comparison of Strains, Plasmids, Resistance Genes, and Virulence Factors. Clin Infect Dis. 2013 Feb 15;56(4):478–87. DOI: 10.1093/cid/cis929. [PubMed: 23243181]
85. Leverstein-van Hall MA, Dierikx CM, Cohen Stuart J, Voets GM, van den Munckhof MP, van Essen-Zandbergen A, Platteel T, Fluit A, van de Sande-Bruinsma, Scharinga J, Bonten M, Mevius D. Dutch patients, retail chicken meat and poultry share the same ESBL genes, plasmids and strains. Clin Microbiol Infect. 2011 Jun 1;17(6):873–80. DOI: 10.1111/j.1469-0691.2011.03497.x. [PubMed: 21463397]
86. De Been M, Lanza V*F, De Toro M, Scharringa J, Dohmen W, Du y, Hu J, Lei Y, Li N, Tooming-Klunderud, Heederik D, Fluit A, Bonten M, Willems R, de la Cruz F, Van Schaik W.Dissemination of Cephalosporin Resistance Genes between *Escherichia coli* Strains from Farm Animals and Humans by Specific Plasmid Lineages. PLoS Genet. 2014;10(12):1004776. DOI: 10.1371/journal.pgen.1004776.
87. Sunde M, Skov Simonsen G, Slettemeås JS, Böckerman I, Norström M. Integron, Plasmid and Host Strain Characteristics of Escherichia coli from Humans and Food Included in the Norwegian Antimicrobial Resistance Monitoring Programs. 2015 Jun 5;10(6):e0128797. DOI: 10.1371/journal.pone.0128797.
88. Chuanchuen R, Puangseree J, Prathan R, Srisanga S, Angkittitrakul S. Epidemiology and Infection Plasmid profile analysis of Escherichia coli and Salmonella enterica isolated from pigs, pork and humans. 2022 May 10;150:e110. DOI: 10.1017/S0950268822000814.
89. Day MJ, Hopkins KL, Wareham DW, Toleman MA, Elviss N, Randall L, Teale C, Cleary P, Wiuff C.Extended-spectrum β-lactamase-producing *Escherichia coli* in human-derived and foodchain-derived samples from England, Wales, and Scotland: an epidemiological surveillance and typing study. Lancet Infect Dis. 2019 Dec 1;19(12):1325–35. DOI: 10.1016/S1473-3099(19)30273-7. [PubMed: 31653524]
90. Dorado-García A, Smid JH, Van Pelt W, Bonten MJM, Fluit AC, Van Den Bunt G, Wagennar J, Jordijk J, Dierikx C, Veldman K, De Koeijer A, Dohmen W, Schmitt H, Liakopoulos A, Pacholewicz E, Lam T, Velthius A.Molecular relatedness of ESBL/AmpC-producing Escherichia coli from humans, animals, food and the environment: a pooled analysis. 2018 Feb 1;73(2):339–347. DOI: 10.1093/jac/dkx397.
91. Yu D, Banting G, Neumann NF. A review of the taxonomy, genetics, and biology of the genus Escherichia and the type species Escherichia coli. 2021 Aug;67(8):553–571. DOI: 10.1139/cjm-2020-0508.
92. Zurfluh K, Jakobi G, Stephan R, Hächler H, Nüesch-Inderbinen M. Replicon typing of plasmids carrying 1 blaCTX-M-1 in Enterobacteriaceae of animal, environmental and human origin. Front Microbiol. 2014 Oct 30;5:555. DOI: 10.3389/fmicb.2014.00555. [PubMed: 25400623]
93. Schultsz C, Geerlings S. Plasmid-mediated resistance in enterobacteriaceae: Changing landscape and implications for therapy. Drugs. 2012;72(1):1–16. DOI: 10.2165/11597960-000000000-00000.
94. Madec JY, Haenni M. Antimicrobial resistance plasmid reservoir in food and food-producing animals. Plasmid. 2018 Sep 1;99:72–81. DOI: 10.1016/j.plasmid.2018.09.001. [PubMed: 30194944]
95. Projahn M, Von Tippelskirch P, Semmler T, Guenther S, Alter T, Roesler U. Contamination of chicken meat with extended-spectrum beta-lactamase producing-*Klebsiella pneumoniae* and

Escherichia coli during scalding and defeathering of broiler carcasses. 2019 Feb;77:185–191. DOI: 10.1016/j.fm.2018.09.010.

96. Baltrus DA. Exploring the costs of horizontal gene transfer. Trends Ecol Evol. 2013 Aug 1;28(8):489–95. DOI: doi: 10.1016/j.tree.2013.04.002. [PubMed: 23706556]
97. Rodriguez-Beltran J, Hernandez-Beltran JCR, Delafuente J. Multicopy plasmids allow bacteria to escape from fitness trade- offs during evolutionary innovation. 2018;2018(5):873–81. DOI: doi: 10.1038/s41559-018-0529-z.
98. Soncini JGM, Cerdeira L, Sano E, Koga VL, Tizura AT, Tano ZN, Nakazato G, Kobayashi R, Aires A, Lincopan N, Vespero E. Genomic insights of high-risk clones of ESBL-producing *Escherichia coli* isolated from community infections and commercial meat in southern Brazil. Sci Rep. 2022;12(1):1–12. DOI: 10.1038/s41598-022-13197-y. [PubMed: 34992227]
99. Jarocki VM, Heß S, Anantanawat K, Berendonk TU, Djordjevic SP. Multidrug-Resistant Lineage of Enterotoxigenic Escherichia coli ST182 With Serotype O169:H41 in Airline Waste. Front Microbiol. 2021;12(September):1–18. DOI: 10.3389/fmicb.2021.731050.
100. Zhou D, Nawaz M, Xiong Y, Haenni M, Bai baili L, Zengqi Yang cfsanetcn. Characterization of Five Escherichia coli Isolates Co-expressing ESBL and mcr-1 Resistance Mechanisms From Different Origins in Characterization of Five *Escherichia coli* Isolates Co-expressing ESBL and mcr-1 Resistance Mechanisms From Different Origins in China. China Front Microbiol. 2019 Aug 29;10:1994. DOI: 10.3389/fmicb.2019.01994. [PubMed: 31555232]
101. Kocsis E, Gužvinec M, Butić I, Krešić S, Crnek SŠ, Tambić A. *bla*NDM-1 Carriage on IncR Plasmid in Enterobacteriaceae Strains. Microb Drug Resist. 2016;22(2):123–8. DOI: 10.1089/mdr.2015.0083. [PubMed: 26484384]
102. Bloomfield S, Duong VT, Tuyen HT, Campbell JI, Thomson NR, Parkhill J. Mobility of antimicrobial resistance across serovars and disease presentations in non-typhoidal Salmonella from animals and humans in Vietnam. Microb Genomics. 2022;8:798. DOI: 10.1099/mgen.0.000798.
103. Wang Y, Zhu B, Liu M, Dong X, Ma J, Li X. Characterization of IncHI1B Plasmids Encoding Efflux Pump TmexCD2-ToprJ2 in Carbapenem-Resistant *Klebsiella variicola, Klebsiella quasipneumoniae, and Klebsiella michiganensis* Strains. Front Microbiol. 2021 Oct 6;12:759208. DOI: 10.3389/fmicb.2021.759208. [PubMed: 34691010]
104. Zhang Q, Lv L, Huang X, Huang Y, Zhuang Z, Lu J. Rapid Increase in Carbapenemase-Producing Enterobacteriaceae in Retail Meat Driven by the Spread of the bla NDM-5-Carrying IncX3 Plasmid in China from 2016 to 2018. 2019 Jul 25;63(8):e00573–19. DOI: 10.1128/AAC.00573-19.
105. Wang J, Zeng Z-L, Huang X-Y, Ma Z-B, Guo Z-W, Lv L-C. Evolution and Comparative Genomics of F33:A:B Plasmids Carrying bla CTX-M-55 or bla CTX-M-65 in Escherichia coli and Klebsiella pneumoniae Isolated from Animals, Food Products, and Humans in China. 2018 Jul 18;3(4):e00137–18. DOI: 10.1128/mSphere.00137-18.
106. Carattoli A, Villa L, Fortini D, García-Fernández A. Contemporary IncI1 plasmids involved in the transmission and spread of antimicrobial resistance in Enterobacteriaceae. 2021 Nov;118:102392. doi: 10.1016/j.plasmid.2018.12.001.

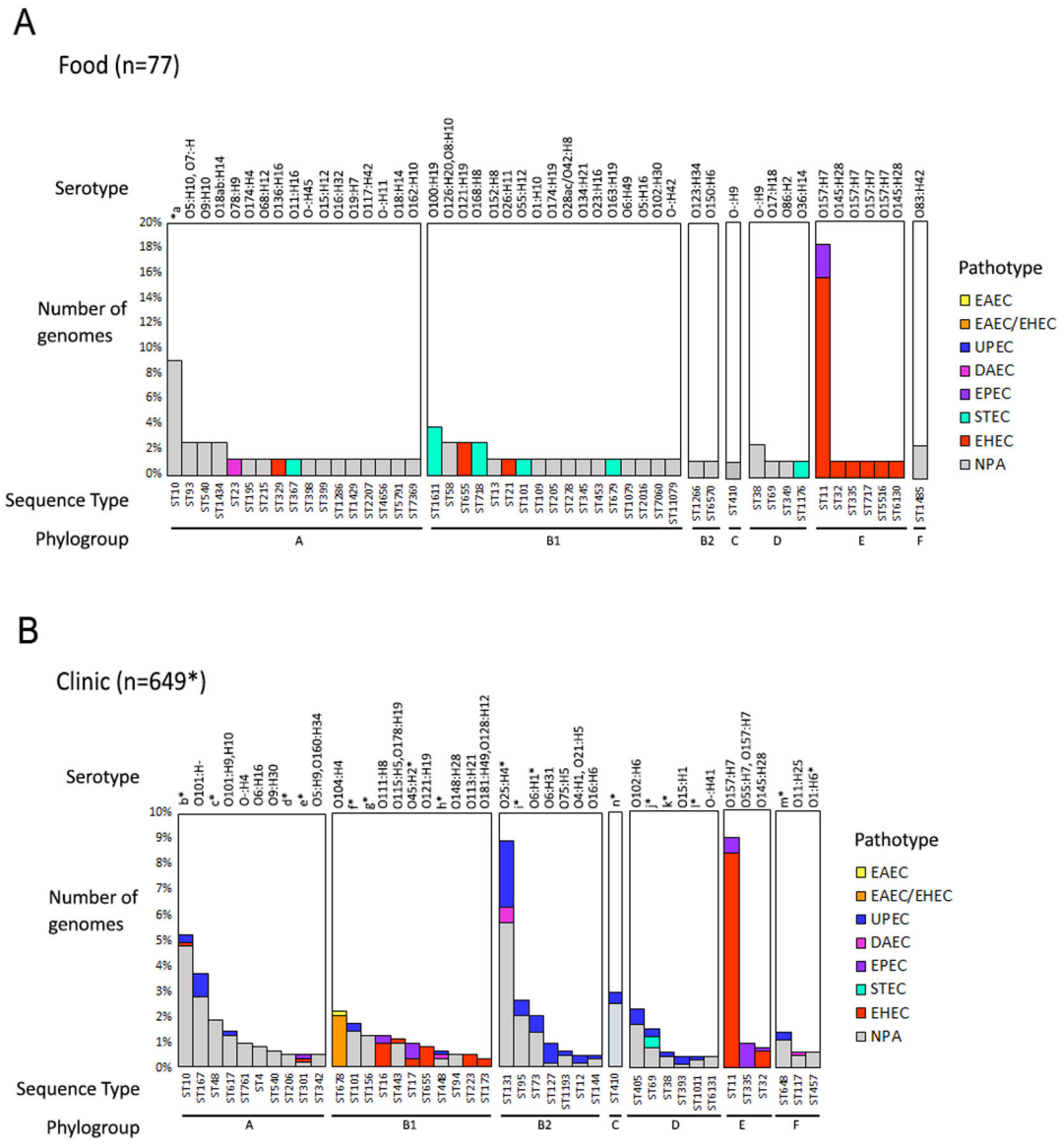

Figure 1. Phylogroup, serotype and pathotype classification.

A Food strains**. B** Clinical strains. Different pathotypes are highlighted using the following color code; yellow (EAEC), orange (EAED/EHEC), blue (UPEC), pink (DAEC), purple (EPEC), cyan (STEC), red (EHEC) and gray (NPA, non-pathotype associated). Only STs with at least three genomes from clinical source are shown. The asterisks indicate the STs in which more than three different serotypes are found. The complete information can be found in Table S1A and Table S1B of supplementary materials.

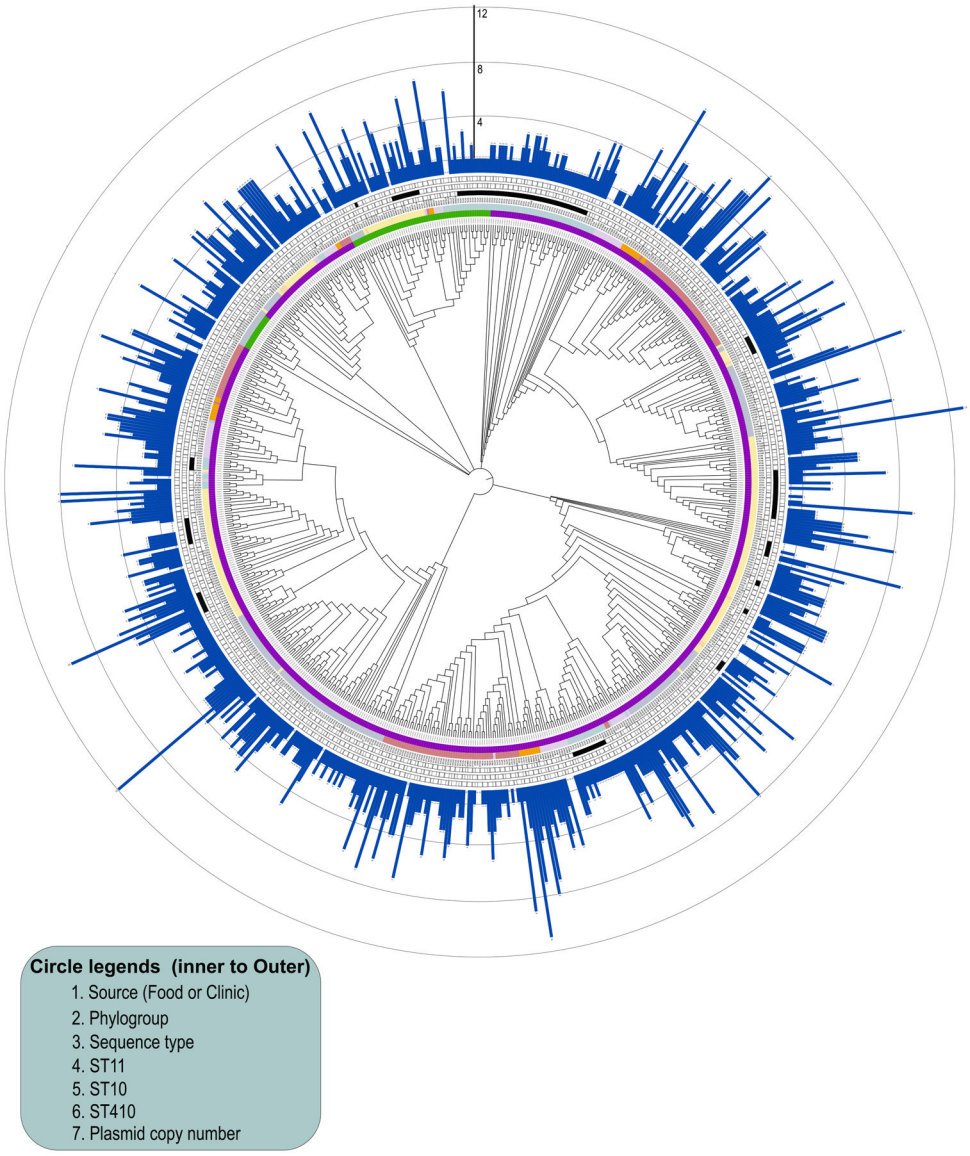

Figure 2. Phylogenetic analysis.

The phylogenetic relationships between the strains included in this study (726 sequenced strains, 77 from food and 649 from clinical sources genomes) are shown as a cladogram. This graphic representation was generated using the CSI Phylogeny platform, with *Escherichia coli* K-12 (NC_000913.3) as a reference genome. From inner to outer, the circles represent the following: Circle 1: Source. Food (green) or Clinic (purple). Circle 2: Phylogroup. A (light yellow), B1(light blue), B2 (pink), C (light orange), D (violet), E (sky blue) and F (orange). Circle 3: Sequence type assigned using MLST platform (see Methods). Circle 4–6 (Top 3 ST in clinic and food): ST11 (Circle 4), ST10 (Circle 5), ST410 (Circle 6). Circle 7: Plasmid copy number, *i.e.* the number of plasmids per host strain (dark blue) with lines as references for: 1, 4, 8 and 12 plasmids.

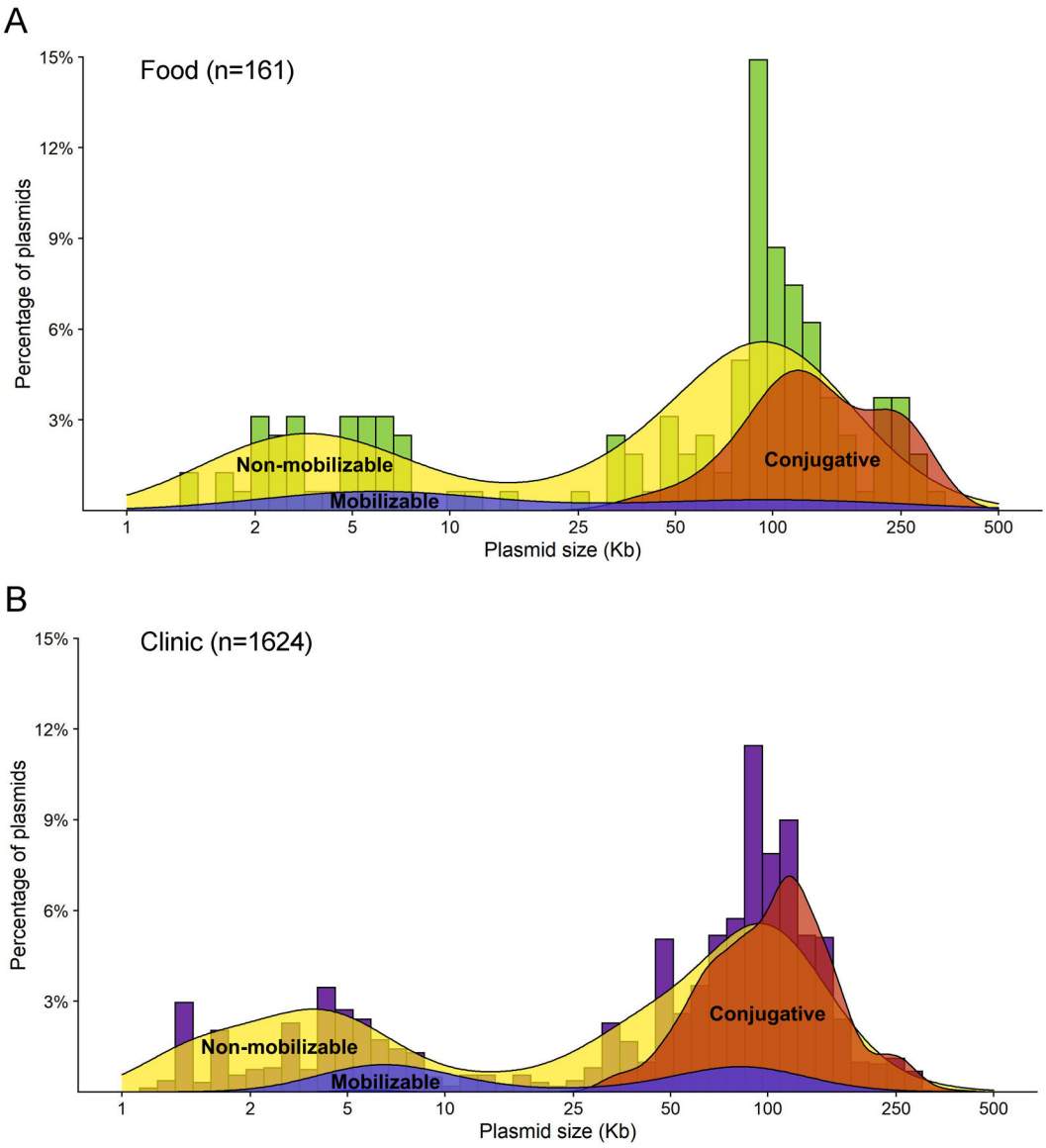

Figure 3. Plasmid size distribution and mobilization profile.

The plasmid size is shown in bins of 1kb. The fraction of conjugative (brown), mobilizable (blue) and non-mobilizable (yellow) plasmids is also shown as area under the curve along the X axis, *i.e.* across increasing plasmid size. **A. Food plasmids.** Plasmid size shows a trimodal distribution with the following median peak sizes: 3.2, 7.3 and 93.2 kb. **B. Clinical plasmids**. Plasmid size shows a trimodal distribution with the following median peak sizes: 1.5, 6.6 and 109.5 kb.

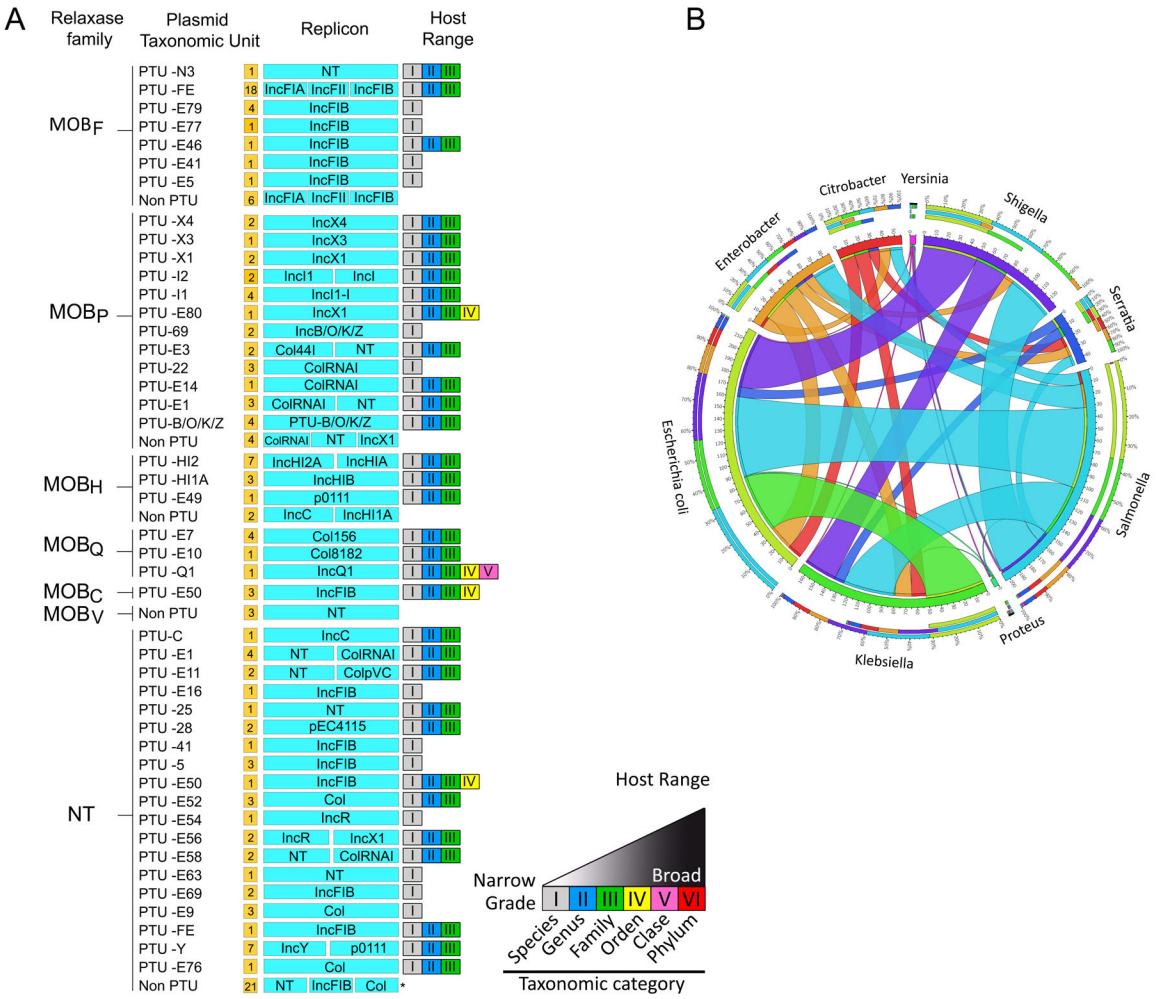

Figure 4.

A. PTU classification of food plasmids. The PTU classification of our 161 plasmids as determined by COPLA (16) is shown. The program was able to ascribe a PTU to 92 of them, for a total of 39 different PTUs. **A. PTU classification.** The 39 different PTUs found are listed, and their correspondence to relaxase families (MOB) is shown to the left of the column listing PTU classification. The number of plasmids ascribed to each of the PTUs is listed to the right, inside the orange square. Their general Rep (replicon identified by PlasmidFinder) correspondence and host range are indicated next to the corresponding boxes with roman numerals. The predicted promiscuity (*i.e.*, host range) is indicated as one of six categories or "grades". Grade I represents a predicted promiscuity restricted to the level of species, II to the level of genus, III to family, IV to order, and V to the level of class (see inset at the bottom). B. **Chord diagram showing potential plasmid exchanges between genera.** This diagram shows shared PTU presence across different genera of enterobacteria. These are listed at the periphery of the outer circle using the following color code: *E. coli* (orange), *Enterobacter* (yellow), *Klebsiella* (green), *Proteus* (aquamarine), *Salmonella* (light blue), *Serratia* (dark blue), *Shigella* (purple), *Yersinia* (pink) and *Citrobacter* (red). The genera sharing a given PTU are shown joined by a curved edge, the thickness of which is

proportional to the number of shared plasmids. The color indicative of the genus where the plasmid is thought to have originated.

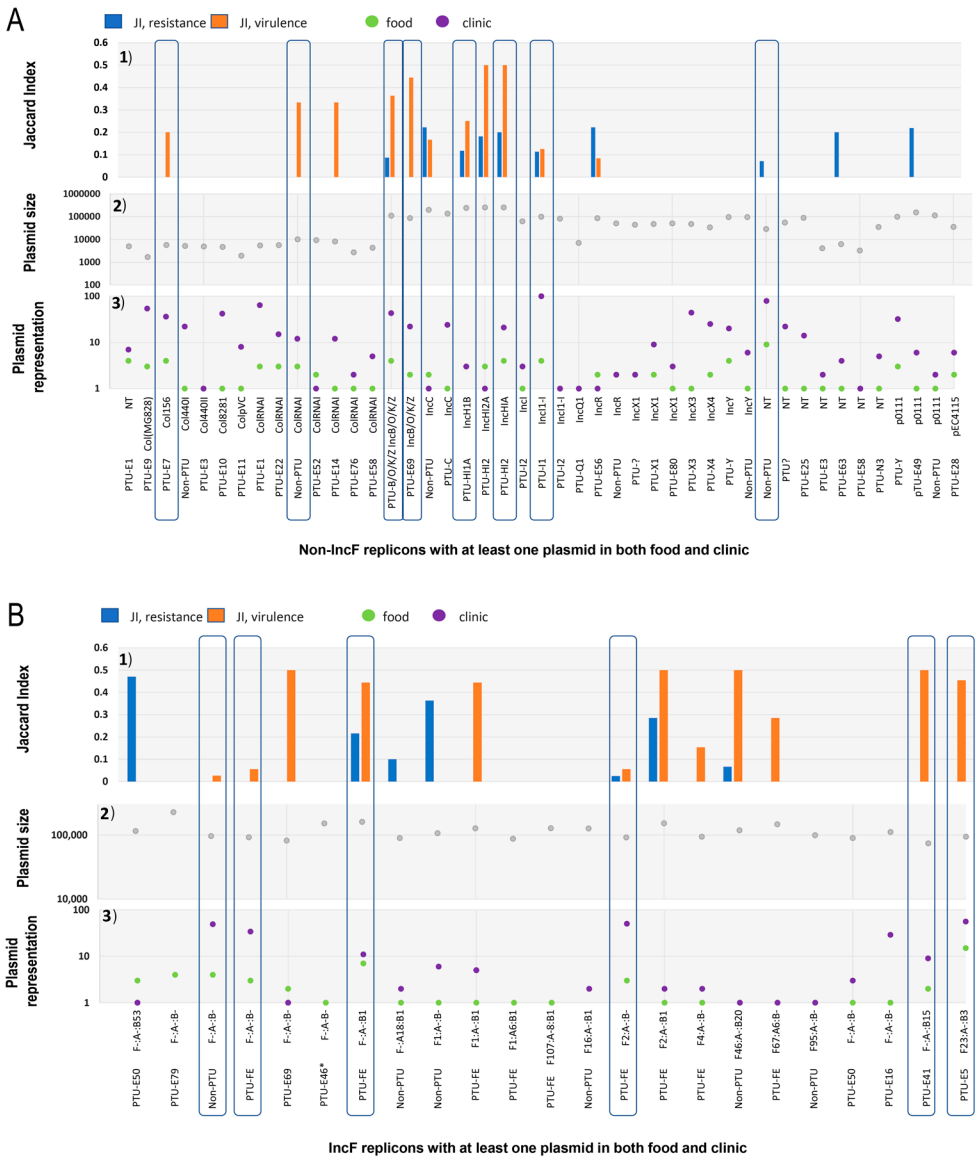

Figure 5. Resistance and virulence factor overlap, broken down by specific PTU-Inc group combinations.

The X axis lists PTU-Inc group combinations with representation in both food and clinical sources. **Panel a.** Jaccard Index (JI) of resistance (blue bars) and of virulence (orange bars) as an indicator of genetic overlap between food and clinical plasmids. Values range between 0 (no overlap) and 1 (complete overlap). Note that the large difference in sample size between the two groups of plasmids, biases the index against overlap. **Panel b**. Shows the average size (bp) of plasmids assigned to each of the PTU-Inc group combination listed on the X axis (light gray dots). **Panel c.** The number of plasmids identified in for each PTU-Inc group combination broken down by source, food (green dots), and clinical (purple dots).

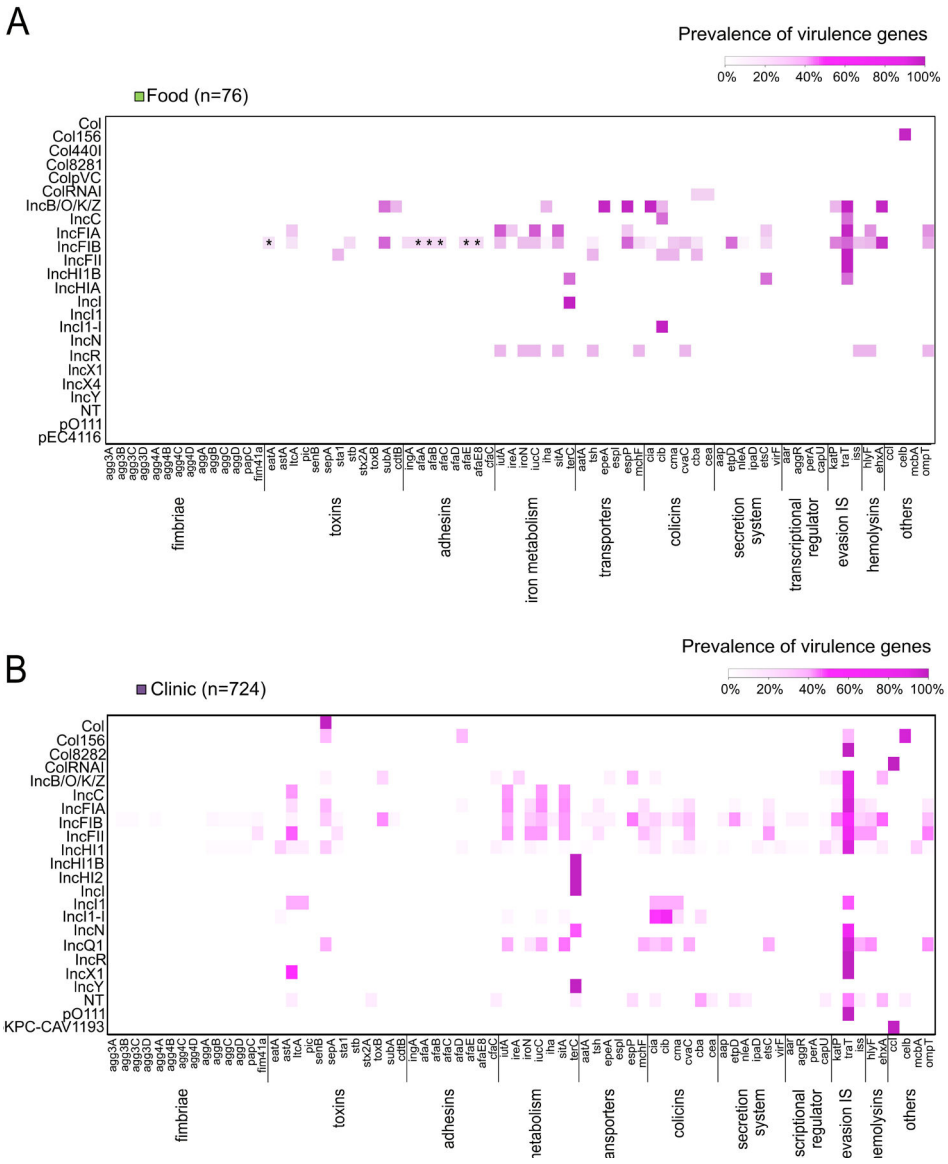

Figure 6. Association between virulence gene content and Inc type.

A. Food plasmids. **B. Clinical plasmids.** The association between virulence genes listed on the X-axis (grouped by categories) and plasmids carrying specific replicons as defined by PlasmidFinder listed on the Y-axis is shown as a heatmap, with the intensity of the pink color is proportional to the prevalence of the gene-Inc type association. **A. Food plasmids** (n=76). **B. Clinical plasmids** (n=724).

Plasmid. Author manuscript; available in PMC 2024 February 26.

Table 1A.

Relative representation of replicon categories as a ratio in the clinic plasmid vs food plasmid and JI indexes for plasmids with n=>3 for both food and clinical plasmids

Inc Group	Total number of plasmids	Number of food plasmids	Number of clinic plasmids	Ratio of clinic vs food representation	JI resistance	JI virulence	JI total
Col	102	9	93	1.02	0.00	0.00	0.00
Col156	53	4	49	1.21	0.00	0.20	0.20
ColRNAI	118	9	109	1.20	0.00	0.00	0.00
IncB/O/K/Z	70	6	64	1.06	0.09	0.33	0.21
IncC	27	3	24	0.79	0.10	0.11	0.10
IncF (F-:A-:B-)	146	16	130	0.81	0.02	0.11	0.06
IncF (F-:A-:B1)	31	7	24	0.34	0.18	0.40	0.25
IncF (F2:A-:B-)	58	3	55	1.82	0.02	0.06	0.02
IncF (F23:A-:B3)	70	15	55	0.36	0.00	0.45	0.29
IncHI1B	8	3	5	0.17	0.03	0.25	0.05
IncI1-I	90	4	86	2.13	0.14	0.11	0.13
IncR	12	6	6	0.10	0.18	0.08	0.15
IncX1	22	5	17	0.34	0.17	0.00	0.15
IncY	36	5	31	0.61	0.03	0.00	0.03
NT	221	20	201	1.00	0.03	0.00	0.02
p0111	41	5	36	0.71	0.18	0.00	0.18
Average	69.1	7.5	61.6	0.85	0.07	0.13	0.12

Inc group: lists replicons with at least three plasmids represented in both the food and clinical plasmid datasets **Total number of plasmids**: indicates the total of plasmids corresponding to the Inc group listed on the left, adding both plasmid datasets. **Number of food plasmids**: lists the plasmid content in each Inc group for the food dataset. **Number of clinic plasmids**: lists the plasmid content in each Inc group for the clinic dataset **Ratio of clinic vs. food representation:** indicates the representation of clinic plasmids in relationship to food plasmids corresponding to each replicon (% clinical plasmids for a given Inc group divided by the % food plasmids for a given Inc group). **JI resistance** indicates the Jaccard Index value for the resistance genes cargo, **JI virulence** indicates the Jaccard Index value for the virulence genes cargo, and **JI total** indicates the Jaccard Index value for the charge of the sum of virulence and resistance genes of the corresponding replicon (**See** Methods).

Table 1B.

The relative representation of PTU categories as a ratio of in the clinic plasmid vs food plasmid

PTU	Total number of plasmids	Frequency in food plasmids	Frequency in clinic plasmids	Ratio of clinic vs food representation	JI resistance	JI virulence	JI total
Non-PTU	88	9	79	1.00	0.07	0.00	0.05
Non-PTU	53	4	49	1.40	0.00	0.03	0.02
Non-PTU	15	3	12	0.46	0.00	0.33	0.25
PTU-?	127	3	124	4.73	0.00	0.00	0.00
PTU-B/O/K/Z	47	4	43	1.23	0.09	0.36	0.22
PTU-E1	84	7	77	1.26	0.00	0.00	0.00
PTU-E22	18	3	15	0.57	0.00	0.00	0.00
PTU-E5	80	15	65	0.50	0.00	0.45	0.31
PTU-E50	10	4	6	0.17	0.38	0.00	0.38
PTU-E52	8	3	5	0.19	0.00	0.00	0.00
PTU-E69	35	4	31	0.89	0.00	0.44	0.44
PTU-E7	40	4	36	1.03	0.00	0.20	0.20
PTU-E9	57	3	54	2.06	0.00	0.00	0.00
PTU-FE	277	19	258	1.55	0.23	0.41	0.28
PTU-HI1A	6	3	3	0.11	0.12	0.25	0.14
PTU-HI2	31	7	24	0.39	0.24	0.50	0.25
PTU-I1	104	4	100	2.86	0.11	0.13	0.11
PTU-Y	61	7	54	0.88	0.00	0.00	0.00
Average	63.39	5.89	57.50	1.18	0.07	0.17	0.15

PTU: lists replicons with at least three plasmids represented in both the food and clinical plasmid datasets **Total number of plasmids**: indicates the total of plasmids corresponding to the PTU listed on the left, adding both plasmid datasets. **Number of food plasmids**: lists the plasmid content in each PTU for the food dataset. **Number of clinic plasmids**: lists the plasmid content in each PTU for the clinic dataset **Ratio of clinic vs. food representation:** indicates the representation of clinic plasmids in relationship to food plasmids corresponding to each replicon (% clinical plasmids for a given Inc group divided by the % food plasmids for a given PTU). **JI resistance** indicates the Jaccard Index value for the resistance genes cargo, **JI virulence** indicates the Jaccard Index value for the virulence genes cargo, and **JI total** indicates the Jaccard Index value for the charge of the sum of virulence and resistance genes of the corresponding replicon (**See** Methods).

Table 2.

Representation of ARGs, grouped by families and overlap between food and clinical plasmid datasets.

Plasmid ARG Families	Food	Clinic	Total	Unique to food plasmids				
Beta-lactamases	14	58	72	*bla*$_{\text{SHV-12}}$	*bla*$_{\text{HERA-3}}$	*bla*$_{\text{LAP-2}}$		
Aminoglycoside acetyl-, phospho-, adenylyl-transferases	17	33	50	*aac(3)-IIa*	*aac(3)-IIg*	*aac(6')-IIc*	*aadA17*	*aadA24*
Fluoroquinolone resistance	6	16	22	*qnrB19*	*qnrS11*			
Trimethoprim resistance	5	14	19					
Phenol resistance	3	9	12					
Tetracycline resistance	4	8	12					
Macrolide resistance	1	5	6					
Sulfonamide resistance	3	3	6					
Lincosamide resistance	3	3	6					
Colistin resistance	2	3	5	*mcr-9.1*				
Rifamycin resistance	1	2	3					
Fosfomycin resistance	0	3	3					
RNA methylases	1	2	3					
Eflux pumps	0	3	3					
Glycopeptide resistance	1	1	2					
Nucleoside resistance	1	1	2					

Plasmid ARG Families: lists all ARGs represented in our plasmid database, grouped by families**. Food/Clinic**: number of individual ARGs found in food plasmids/clinical plasmids, including multiple hits of the same ARG. **Total**: adds up the combined number of ARGs. **Overlap:** tallies the genes overlapping between the two plasmid sources, in absolute number.

*Expresses overlap as % of combined ARGs.

**lists the ARGs that are unique to food plasmids. Fosfomycin resistance genes and efflux pumps were exclusively found in clinical plasmids.

Table 3.

Gene Content for the PTU-Inc Combinations with high Jaccard Index.

PTU-Inc Combinations	Resistance	Virulence
PTU-B/O/K/Z IncB/O/K/Z	***aph(6)-Id,*** *aph(3")-Ib*	***cia,*** *cib,* ***ehxA,*** *epeA,* ***espP,*** *iha, subA,* ***traT***
PTU-HI2 IncHIA	***aac(3)-IId***, ***aadA2***, *aadA22, mcr-1.1,* ***bla*** $_{\textbf{TEM-1}}$, *arr-2,* ***bla*** $_{\textbf{CTX-M-55}}$, *dfr14,linG, qnrs1, mphA,* ***sul1, sul2***, ***sul3***	*terC*
Non-PTU ColRNAI		*cea*
PTU-E69 IncB/O/K/Z		***ehxA, espP***, ***toxB, katP***
PTU-HI2 IncHI2A	*aadA1,* ***aadA2,*** *aph(3')-Ia, mcr-1.1,* ***sul3***, ***tet(A)***	-
PTU-FE F:-:A-:B1	***aac(3)-IId,*** *aph(3')-IIa,* ***aph(6)-Id, bla*** $_{\textbf{CTX-M-55}}$, ***sul1, sul2***, ***bla*** $_{\textbf{TEM-1}}$, ***tet(A)***	*cvaC,* ***ehxA,*** *etsC, hlyF, iroN, iss, iucC, iutA, mchF, ompT, sitA,* ***traT,*** *tsh*
PTU-E41 F-:A-:B15	-	***ehxA, etpD, traT***, *etpP, nleA,* ***cia***
PTU-E5 F23:A-:B3	-	***ehxA, espP, etpD, toxB, katP***

PTU-Inc Combinations: PTU-Replicon combinations with highe Jaccard Index are listed. **Resistance:** indicates the resistance genes contained in the plasmids assigned to the corresponding PTU-Replicon combination. **Virulence**: lists the virulence genes carried by the plasmids assigned to the corresponding PTU-Replicon combination. **Bold**, genes that are found in more than one PTU-Inc group category.

Plasmid. Author manuscript; available in PMC 2024 February 26.

Table 4.

Summary of the specific associations Inc group, PTU or PTU- Inc group combinations for food plasmids and associations with high resistance and/or virulence cargo overlap between food and clinic plasmids.

Characteristics	Food	Food and clinic: high overlap	Food and clinic: other PTU-Inc overlap;
General Inc type overrepresentation	IncR, IncHI1B, IncX1, IncF(F-:A-:B1), IncF(F23:A-:B3)		
General PTU overrepresentation	PTU-HI1A, PTU-E50, PTU-E52. Unique to food plasmids PTU-E46, PTU-E54, PTU-E79		
ARG: Inc type overrepresentation	IncR, IncHI1B IncF(F1:A-:B-)		
ARG: high Inc type JI		IncR, p0111, IncX1, IncI1-I IncF(F-:A-:B1)	
ARG: high PTU JI		PTU-E50, PTU-HI2, PTU-FE	
ARG: high JI Inc type-PTU		PTU-B/O/K/Z IncB/O/K/Z PTU-HIA IncHI1B PTU-HI2 IncHIA PTU-FE IncF(F-:A-:B1)	PTU-I1 IncI1-I Non-PTU NT PTU-FE IncF (F2:A-:B-)
Virulence: Inc type overrepresentation	IncR, IncHI1B, ColRNAI, IncB/O/K/Z, IncHI2, IncY and IncF(F23:A-:B3)		
Virulence: PTU overrepresentation			
Virulence: high Inc type JI		Inc-B/O/K/Z IncF (F23:A-:B3) IncF (F-:A-:B-)	
Virulence: high PTU JI		PTU-HI2, PTU-E5, PTU-FE, PTU-B/O/K/Z	
Virulence: high JI Inc type-PTU		PTU-B/O/K/Z IncB/O/K/Z PTU-E69 IncB/O/K/Z PTU-HI2 IncHIA Non-PTU ColRNAI PTU-FE IncF(F-:A-:B1) PTU-FE IncF (F23:A-:B3) PTU-E41 IncF(F-:A-:B15)	PTU-E7 Col156 PTU-HIA IncHI1B Non-PTU IncF(F-:A-:B-) PTU-FE IncF(F-:A-:B-)

Column A Lists the different categories. **Column B**. indicates the assignment of Inc groups and PTUs to food-borne plasmids according to the corresponding category. **Column C** Lists the Inc groups, PTU, and Inc group-PTU combinations associated with high overlap between food and clinical plasmids for the categories indicated in **Column A**. C**olumn D** indicates other PTU-Inc combinations with overlap between food and clinical plasmids that are of interest for the categories listed in **Column A**. PTU-Inc group combinations exhibiting overlap for both virulence and resistance genes are underlined.