



Article

Genomic Insights into Drug Resistance and Virulence Platforms, CRISPR-Cas Systems and Phylogeny of Commensal *E. coli* from Wildlife

Carla Andrea Alonso ^{1,2,†} , María de Toro ^{3,†}, Fernando de la Cruz ⁴ and Carmen Torres ^{1,*}

¹ Área de Bioquímica y Biología Molecular, Universidad de La Rioja, 26006 Logroño, Spain; caalonso@riojasalud.es

² Servicio de Microbiología, Hospital San Pedro, 26006 Logroño, Spain

³ Plataforma de Genómica y Bioinformática, Centro de Investigación Biomédica de La Rioja (CIBIR), 26006 Logroño, Spain; mthernando@riojasalud.es

⁴ Departamento de Biología Molecular, Universidad de Cantabria and Instituto de Biomedicina y Biotecnología de Cantabria (Universidad de Cantabria-CSIC), 39011 Santander, Spain; fernando.cruz@unican.es

* Correspondence: carmen.torres@unirioja.es; Tel.: +34-941299750; Fax: +34-941299721

† These authors contributed equally to this work.

Abstract: Commensal bacteria act as important reservoirs of virulence and resistance genes. However, existing data are generally only focused on the analysis of human or human-related bacterial populations. There is a lack of genomic studies regarding commensal bacteria from hosts less exposed to antibiotics and other selective forces due to human activities, such as wildlife. In the present study, the genomes of thirty-eight *E. coli* strains from the gut of various wild animals were sequenced. The analysis of their *accessory* genome yielded a better understanding of the role of the mobilome on inter-bacterial dissemination of mosaic virulence and resistance plasmids. The study of the presence and composition of the CRISPR/Cas systems in *E. coli* from wild animals showed some viral and plasmid sequences among the spacers, as well as the relationship between CRISPR/Cas and *E. coli* phylogeny. Further, we constructed a single nucleotide polymorphisms-based *core* tree with *E. coli* strains from different sources (humans, livestock, food and extraintestinal environments). Bacteria from humans or highly human-influenced settings exhibit similar genetic patterns in CRISPR-Cas systems, plasmids or virulence/resistance genes-carrying modules. These observations, together with the absence of significant genetic changes in their *core* genome, suggest an ongoing flow of both mobile elements and *E. coli* lineages between human and natural ecosystems.

Keywords: *E. coli*; CRISPR-Cas; antimicrobial resistance; wild animals; WGS; PLACNETw



Citation: Alonso, C.A.; de Toro, M.; de la Cruz, F.; Torres, C. Genomic Insights into Drug Resistance and Virulence Platforms, CRISPR-Cas Systems and Phylogeny of Commensal *E. coli* from Wildlife. *Microorganisms* **2021**, *9*, 999. <https://doi.org/10.3390/microorganisms9050999>

Academic Editor: Vincenzo Scarlato

Received: 22 March 2021

Accepted: 30 April 2021

Published: 5 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Escherichia coli is a well-known commensal of the gut of humans and a wide-range of other animals, but it can also reproduce and persist for long periods of time in extraintestinal natural environments. Beside these habitats, some strains of *E. coli* have the potential to cause severe intestinal and extraintestinal illness, such as meningitis, septicemia, pneumonia or urinary tract infections. This diversity in terms of niche distribution and host–pathogen interactions is due to the high plasticity of the *E. coli* genome, which allows the bacteria to adapt to the varying selective pressures exerted by the different environments. All this, plus the easy and fast growth characteristics of *E. coli*, makes this species an excellent model to study the evolution and epidemiology of antimicrobial resistance or bacterial virulence [1].

The dynamics of bacterial genomes involve both single mutations in the *core* genome (microevolution) as well as horizontal gene transfer events (macroevolution) that shape a species pangenome [2]. Both types of evolutionary phenomena have influenced the acquisition of new phenotypes and host/niche adaptation. Horizontal gene transfer (HGT)

is largely mediated by mobile genetic elements (MGEs) such as insertion sequences (ISs), transposons, integrons, prophages, integrative conjugative elements, genomic islands and plasmids. Human activities promote the use of selective agents (antibiotics, disinfectants, heavy metals) that induce a bacterial SOS response and, thus, increase HGT rates [3]. This leads to the assembly of mosaic genetic elements and the development of antimicrobial and/or virulence platforms of increasing complexity. Besides the selective pressure exerted by chemicals, the genetic background of bacteria also seems to play a role in the antimicrobial resistance phenotype. As an example, the strong correlation observed between *E. coli* phylogroup B2 and antimicrobial susceptibility patterns [4].

Classically, PCR-based genotypic tests and other traditional typing methods (serotyping, multilocus enzyme electrophoresis, pulsed-field gel electrophoresis, multilocus sequence typing and phylogrouping multiplex PCR) were used to ascertain the genetic relatedness and transmission dynamics of bacteria [5]. The advent of whole genome sequencing (WGS) technology provides a finer-scale analysis tool that enables the high resolution detection of antimicrobial resistance and virulence determinants, the analysis of genome structural variations and genetic linkages or the identification of potential adaptations of the bacterial genomes to the host conditions, among others. Since most genome sequencing projects focus on the analysis of multi-drug resistant clones, clinically relevant pathogens or epidemiological outbreaks, there is a lack of genomic data on commensal *E. coli* populations, even more so in those non-directly associated to humans. The study of randomly selected microbial communities, less exposed to antibiotics and other selective forces associated with the human environment, could provide new insights into the evolution of *E. coli* genomes and the flow of clinically relevant genes among different ecosystems.

In the present work, a collection of 38 *E. coli* genomes from the gut of various wild animals were sequenced. Analysis of their *accessory* genome yielded a better understanding of the role of the mobilome on inter-bacterial dissemination of mosaic virulence and resistance platforms. The study of the presence and composition of the CRISPR/Cas systems in *E. coli* from wild animals showed some phage- and plasmid-related sequences among the spacers, as well as the relationship between CRISPR/Cas and *E. coli* phylogeny. Further, we constructed a single nucleotide polymorphism (SNP)-based *core* tree with *E. coli* strains from different origins (humans, livestock, food and extraintestinal environments) in order to elucidate whether the commensal population of wild animals is subjected to a parallel independent microevolution or, on the contrary, an on-going inter-host transmission is more likely occurring.

2. Materials and Methods

2.1. Bacterial Collection, Antibiotic Susceptibility Testing and Phylotyping

In this work, 38 *E. coli* strains randomly selected among those previously isolated and identified as part of other studies [6–9] were subjected to sequencing. The basic genomic features of these 38 genomes are included in Supplementary Materials Table S1. The strains were isolated from fecal specimens (or intestine segments with fecal content in the case of wild boars) collected between 2013 and 2015 in various geographic locations of Spain (Castilla-La Mancha, Cádiz, Aragón) from healthy wild animals as previously described [6–9]. Eleven strains were isolated from red deer (*Cervus elaphus*), 11 from wild boars (*Sus scrofa*), 9 from small rodents ((mice, $n = 8$ (*Apodemus sylvaticus* and *Mus musculus*); rats, $n = 1$ (*Rattus rattus*)), and 7 from birds of prey (eagles, $n = 2$ (*Aquila chrysaetos* and *Circus gallicus*); vultures, $n = 4$ (*Gyps fulvus*); osprey, $n = 1$ (*Gypaetus barbatus*)). Intestine segments with fecal content were collected from wild boars legally hunted in their own habitat during the regular hunting season (October to February 2014–2015) approved by the Consejería de Agricultura of Castilla-La Mancha [8]; no approval was needed from an ethical committee since the sacrifice of animals was not performed for research purposes, following the Spanish Policy for Animal Protection RD53/2013 and the European Union Directive 2010/63.

All isolates were tested for antimicrobial susceptibility with ampicillin, amoxicillin/clavulanate, ceftazidime, ceftriaxone, cefoxitin, imipenem, nalidixic acid, ciprofloxacin, gentamicin, amikacin, tobramycin, streptomycin, chloramphenicol, sulfonamides, trimethoprim/sulfamethoxazole and tetracycline, using the CLSI disk-diffusion method. Identification of the *E. coli* phylogroups was analyzed by a multiplex PCR-based assay [10], followed by *in silico* validation.

For comparative phylogenomic analysis, another 242 *E. coli* full genomes were taken from the NCBI database. These belonged to bacteria from different sources: humans, livestock, wild animals and the environment. Information about all these genomes is included in Supplementary Materials Table S2.

2.2. Sequencing of the *E. coli* Genomes

Genomic DNA from *E. coli* strains was extracted using the QIAmp DNA Mini Kit (Qiagen, Hilden, Germany). Subsequently, libraries were prepared with the TruSeq DNA PCR-free Sample Preparation Kit (Illumina, San Diego, CA, USA) and 100 bp paired-end reads were sequenced in a HiSeq 1500 System (Health in Code Facility). Illumina reads were quality checked with FastQC software (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and trimmed to optimize their quality with TrimGalore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore).

2.3. Genome Analysis

Clean reads were uploaded to the PLACNETw plasmid reconstruction tool [11] in order to assemble and to elucidate genome components. After manual pruning based on scaffold link and coverage analysis as well as reference sequences comparisons, chromosomal and plasmid sequences were identified, clustered and downloaded in separated files. Besides genome reconstruction, PLACNETw enabled the classification of the plasmids by identification of relaxases, plasmid replication initiator proteins and incompatibility groups (Inc).

E. coli MLST profiles were predicted *in silico* using the Achtman 7 gene scheme at Enterobase (<http://enterobase.warwick.ac.uk> (accessed on 18 December 2017)). The genomes of six strains showing novel MLST alleles or allele combinations were uploaded to the database for sequence type (ST) assignment. ResFinder 3.1 and VirulenceFinder 2.0 (90% identity and 60% minimum length) were used to estimate the antimicrobial resistance and virulence gene content, respectively [12,13]. To determine the presence of CRISPR/Cas systems and the number and sequences of spacers, the CRISPRfinder program was employed, with manual validation [14]. Spacer homologues were identified by BLASTn search against the nucleotide collection (nr/nt) database (query coverage 100%, identity $\geq 88\%$), excluding matches to other clustered, regularly interspaced, short palindromic repeats (CRISPR) sequences.

Whenever necessary, contigs were manually assembled, curated and annotated using additional bioinformatic tools as: Artemis and SnapGene softwares, BLAST (www.ncbi.nlm.nih.gov/BLAST), ORFfinder (www.ncbi.nlm.nih.gov/projects/gorf), ISfinder (www-is.biotoul.fr) and UniProt database (<http://www.uniprot.org> (accessed on 22 December 2017)). To visualize resistance and virulence gene comparisons, EasyFig 2.1 was used [15].

Phylogenetic analysis of the *cas* set of genes or specific virulence determinants were carried out using MEGA7 software [16] from alignments generated with CLUSTALW. Concatenated amino acid or nucleotide sequences trees were constructed using the UPGMA method, with distances calculated by the Poisson correction or maximum composite likelihood, respectively, on a pairwise-deletion comparison.

For the phylogenomic study, the *core* genome was defined as the assembly of genes present in all the genomes analyzed, with more than 80% similarity and 60% coverage. Genes were clustered with CD-HIT-EST [17], selecting those that were present at least in one copy per genome. This subset was concatenated and aligned with Muscle (v3.8.31) (<http://www.drive5.com/muscle>). SNPs were extracted and visualized in a SNP-matrix

with HarvestTools (v1.0) (<https://harvest.readthedocs.io/en/latest/content/harvest-tools.html>). Finally, RAxML [18] was used to build the core genome phylogenetic tree, by using 100 replicates for bootstrap determination.

2.4. Nucleotide Sequence Accession Numbers

Raw sequence reads reported in this paper were deposited under NCBI BioProject PRJNA699864.

3. Results and Discussion

Our analysis of the 38 genomes of *E. coli* isolated from wild animals focused on their CRISPR/Cas systems (Section 3.1), plasmidome and resistome (Section 3.2), virulome (Section 3.3) and phylogenomics (Section 3.4).

3.1. CRISPR/Cas Systems

The discovery of CRISPRs was reported in the late 1980s in *Bacteria* and, a few years later, in *Archaea* [19,20]. CRISPRs are direct repeats of around 21–47 bp length separated by other short sequences named *spacers*, which share homology with phage- and plasmid-related segments. Frequently, CRISPR loci are found associated with genes encoding nucleases, the *cas* genes. Although different biological roles have been proposed, various functional assays demonstrated that CRISPR/Cas acts as an adaptive and heritable immunity system able to target and neutralize exogenous DNA through a pathway similar to RNA interference (RNAi) in eukaryotes [21]. Two monophyletic CRISPR/Cas systems have been reported in *E. coli*: I-E and I-F1. Subtype I-E consists of two CRISPR arrays (CRISPR-1, CRISPR-2) and a set of eight *cas* genes, while subtype I-F1 has six *cas* genes flanked by CRISPR-3 and CRISPR-4 arrays [22]. In contrast to the I-E system, which seems to be inactive due to H-NS regulation under laboratory conditions [23], the I-F system was demonstrated to be constitutively expressed, interfering targeted foreign elements [24,25].

We examined the CRISPR/Cas and spacer repertoire in our collection of 38 *E. coli* from wildlife. CRISPR/Cas I-E and I-F1 systems were detected in 28 (73.4%) and four genomes (10.5%), respectively. As commonly occurs, none carried both I-E and I-F1 subtypes simultaneously.

Supplementary Materials Figure S1 shows the different genetic organizations for I-E and I-F1 systems regarding the module located between the two CRISPR arrays, comprising the *cas* set of genes and adjacent ORFs. The *cas* genes nomenclature follows the recommendations of Makarova et al., 2020 [26]. As other authors described, we found a higher diversity of structures in I-E systems, probably due to deletions and rearrangements promoted by IS elements [27]. In fact, a truncated *IS186*, which has been suggested to guide the CRISPR/Cas evolution owing to its affinity for GC-rich regions (very frequent in distinct points of the system), was found inserted within the *cas3* gene in C7973 and C7974 strains. Among the remaining *cas* genes, only *cas2* was present (variant B in Supplementary Materials Figure S1). Further, in the strain C7347 (phylogroup B2) carrying the CRISPR/Cas I-E a partial deletion of *cas7* and *cse1* genes and the absence of the complete copy of *cse2* was observed (variant A * in Supplementary Materials Figure S1). However, CRISPR 1 and 2 arrays were present, bearing low number of spacers (four and one, respectively), which probably reflect the non-functionality of the system and the progressive loss of spacers (Figure 1).

In addition, the evolutionary relationship of *cas* genes inferred from the phylogenetic tree of the concatenated amino acid sequences of *cas*-E and *cas*-F1 genes showed, as previously reported [27], two clearly differentiated clusters in CRISPR/Cas I-E systems (E1 and E2) and a minor heterogeneity among CRISPR/Cas I-F1 (Supplementary Materials Figure S2). As it was described among ECOR collection [27], the majority of the *E. coli* strains from wildlife, belonging to a variety of phylogroups (A, B1, D, B2), were included in cluster E1, whereas cluster E2 comprised few strains of the A group. Although there was a limited number of strains, the *cas*-F1 set of genes exhibited lower genetic variability

and, although the concatenated amino acid sequence differed by 8.6% between clade V and B2 strains (C7969 vs. C7570/C7975/C7963), they shared a common origin, in contrast to what it is proposed for E1 and E2 variants. Notably, the genetic sequence of *cas-F1* cluster and adjacent genes in clade V strain suggests a close relation with *Escherichia marmotae*, a previously described species isolated from the feces of wild marmots in China [28].

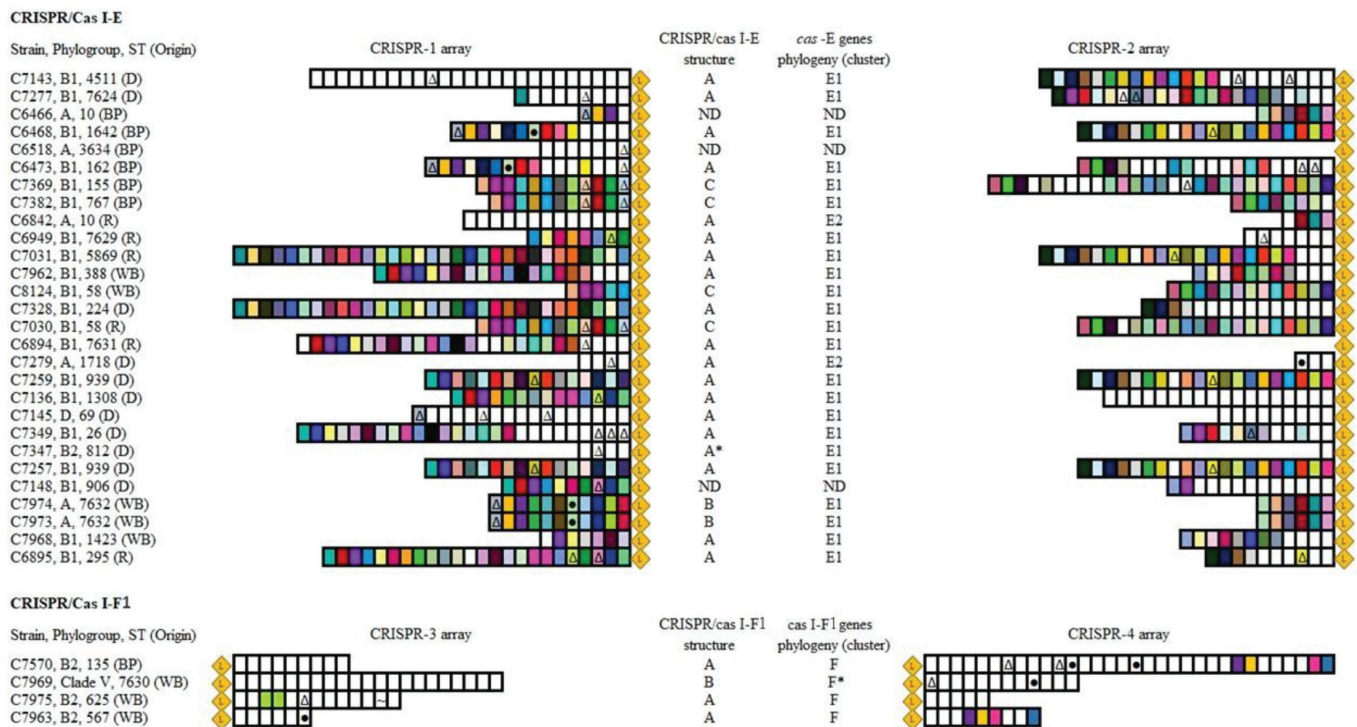


Figure 1. Scheme of the CRISPR arrays located in CRISPR/Cas I-E and I-F1 systems found in the genome of our *E. coli* collection. Spacers are represented as vertical boxes and are oriented with respect to the leader (◊). Boxes with the same color represent spacers that appear in ≥ 2 strains. There were no overlapping spacer sequences between CRISPR 1, 2, 3 and 4 arrays, thus colors must be independently interpreted for each array. The different symbols indicate spacers matching at least 88% with phages (Δ), plasmids (\bullet) or IS (\sim)-like sequences in databases. Origin: deer (D); bird of prey (BP); rodent (R); wild boar (WB). CRISPR/cas I-E (A, A*, B, C) and I-F1 structures (A, B) as well as *cas* genes clusters (E1, E2, F, F*) correspond to those represented in Supplementary Materials Figures S1 and S2, respectively. ND: non-determined.

Concerning the CRISPR arrays, there were a total of 404, 331, 48 and 58 spacers in CRISPR-1 to -4, respectively; 108, 63, 1 and 4 appeared in more than one isolate, regardless of the clonal relatedness of bacteria (Figure 1). The analyses of the spacers of CRISPR loci have been used in several bacterial species for strain typing. As shown in Figure 1 for the studied *E. coli* collection, a good correlation between CRISPR-1/-2 arrays and ST was observed, with some exceptions. We found strains sharing a large CRISPR-1 array that were ascribed to different ST, and even phylogroup (7031 and C7328). The coincidence in spacer segments between non-clonally related strains can be explained by recombination between homologous CRISPR arrays [27], vertical inheritance and differential spacer deletion [29] or a combination of both. Moreover, in agreement with Díez-Villaseñor et al., 2010 [30], we observed that CRISPR-3/-4 arrays contained a larger proportion of unique spacers than CRISPR-1/-2 (90.6% vs. 30.7%), which seem to reflect the higher activity of CRISPR/Cas I-F1 systems demonstrated in the aforementioned experimental studies.

Regarding the homology of spacer sequences with known genes (proto-spacers), we found significant differences among CRISPR arrays. Globally, among the 841 spacers present in CRISPR 1 to 4, 59 (7.0%) showed high homology ($\geq 88\%$) to distinct mobile elements available in public databases. Specifically, in decreasing order of prevalence, 49 (5.8%), 9 (1.1%) and 1 (0.1%) matched viral, plasmid and IS proto-spacers. How-

ever, according to previous studies, the distribution of these homologues varies among CRISPR/Cas I-E and I-F1 [22,30]. Supporting the findings of Díez-Villaseñor et al., 2010 [30], and contrary to Aydin et al., 2017 [22], a significant number of spacers in CRISPR-1 and -2 showed homology with different regions of *E. coli* prophage genomes (percentage among the total viral proto-spacers in I-E/reference sequence): E24377A (46.7%/CP000800.1), P1 (15.5%/AF234172), D6 (2.2%/MF356679) and uncharacterized bacteriophages (35.6%). In CRISPR/Cas I-F1 systems, apart from phages, various spacers matched at least 97% with conserved regions of plasmids. We also identified one spacer 100% identical to a specific region of the ISEc66 element (IS110 family, IS1111 group). The complete nucleotide sequences of these plasmid- or IS- proto-spacers are shown in Figure 2a. Previous research has established a positive correlation between the presence of CRISPR/Cas I-F1 systems and antimicrobial susceptible phenotypes in *E. coli* [22]. Authors concluded that the presence of type I-F1 spacers matching conserved regions within IncI1-, IncFII- and IncFIB-plasmids interfered with the acquisition of resistance plasmids. It is worth noting that despite the different geographical and host origins of the strains, two out of the five spacers they reported in *E. coli* from clinical human specimens were also present in our collection of wildlife commensal strains. This is most likely due to the existence of common ancestral CRISPR loci which evolve, retaining certain spacers in response to the environmental selective pressures. In this regard, the detection of two novel plasmid proto-spacers in CRISPR/Cas I-F1 systems belonging to *E. coli* from wildlife may reflect an improvement in the fitness of bacteria lacking resistance plasmids in environments less exposed to antibiotics.

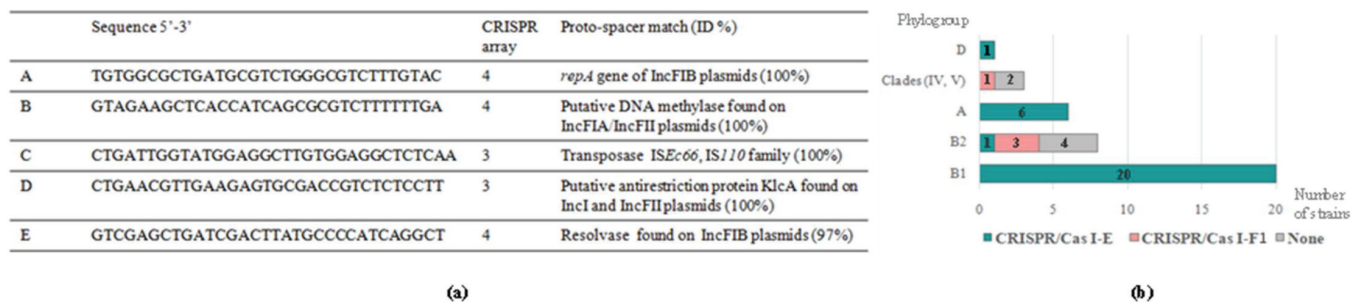


Figure 2. (a) Complete spacer sequences matching at least 97% with known plasmids or IS regions. (b) Distribution of CRISPR/Cas I-E and I-F1 systems among *E. coli* phylogroups.

As previously reported [22,30,31], we observed a clear dependence of the CRISPR/Cas subtype on phylogeny (Figure 2b). On the one hand, concerning CRISPR/Cas I-F1, we mainly found it in B2 group strains (three out of four I-F1-harboring strains). As mentioned before, the presence of this I-F1 system has been associated to low levels of antimicrobial resistance, a frequent pattern observed in B2 isolates [4]. However, some members of the B2 phylogroup, such as those belonging to ST131, have contributed in the last decades to the global dissemination of multi-drug resistance [32]. Notably, in accordance with Aydin et al., 2017 [22], the two strains of the studied collection ascribed to the ST131 complex did not carry CRISPR/Cas systems, which may favor them in the uptake of resistance and virulence determinants under selective environmental conditions. In addition, this suggests a divergent evolution of B2 group members prior to the acquisition of *cas* genes. The remaining strain of the collection carrying a CRISPR/Cas I-F1 system belonged to *Escherichia* cryptic clade V (new ST7630). On the other hand, considering CRISPR/Cas I-E systems, its presence was confirmed in strains ascribed to phylogroups B1, A, D and, more relevantly, B2. Previous analyses on large collections reported the absence of CRISPR/Cas I-E in isolates belonging to B2 group [22,30] or, less frequently, an incomplete I-E system lacking CRISPR 1 array as a consequence of the deletion/truncation of one or more *cas* genes [31].

3.2. Plasmidome and Resistome

PLACNETw reconstructions revealed the presence of plasmids in 29/38 *E. coli* genomes (76.3%), showing ample diversity in size, replicon families and number/isolate, which ranged from one to up to five.

As discussed by Smilie et al., 2010 [33], plasmid size distributions follow a bimodal curve, with average sizes of 5 kb for small plasmids and 200 kb for larger ones. Table 1 summarizes the sizes and genetic features of small plasmids that were present in 28.9% of the isolates. In most cases, they only contained genes involved in plasmid replication and/or mobility. However, they showed a high phylogenetic diversity, encoding replication and relaxase proteins belonging to four different families. Among mobilizable small plasmids, we observed associations between the MOB_{P51} relaxase subfamily and RNAI-encoding loci, which characterized ColE1 plasmids (such as C7382-1), but MOB_{P51} was also present in plasmids with a non-typeable replicon sequence. The same was shown for small MOB_Q plasmids, often but not always carrying a Rep similar to that involved in pColE2 replication [col(156)]. The relation of MOB_{P5} and MOB_Q small plasmids with diverse replication systems has been already pointed out [2]. Further, due to its scarcity among *Enterobacteriaceae*, it is worth noting the detection of a 3348 bp plasmid (C7382-2) showing the conserved domains of MOB_{V2} and a RepL-like replication protein. Besides small mobilizable plasmids, we found several mini-plasmids carrying barely the information for self-perpetuation (plasmids C6466-2, C7369-1, C7347). They show the smallest length of the entire collection (1.5–2.2 kb) and are homologous to the pKST21 family [col(MG828)], a mosaic group of plasmids only transferable by vertical inheritance that need two *rep* genes (*repA-repB*) for autonomous replication [34].

Regarding adaptive and cargo genes of small plasmids, most appeared to be cryptic, since their detectable ORFs (one to three) encoded unknown hypothetical proteins. They could represent adaptive genes to the less studied natural environment. Only three of the small plasmids were involved in the production of colicins, carrying specifically the colE1, colN and colE8 operons. Interestingly, comparing our set of small plasmids with those deposited in public databases, five of them were identical or highly similar ($\geq 99\%$) to others previously reported among *Enterobacteriaceae* of human and livestock origin (Table 1). Hence, according to previous observations, these small replicons, a priori imposing a metabolic cost but scarce benefit to the cell, are common and circulate among *E. coli* from both clinical and environmental settings [35]. The question about whether these structures are simply selfish genetic elements or play a role as moldable vectors in the acquisition of adaptive genes, remains unclear. However, latest evidence points towards the second scenario [36].

Before focusing on larger plasmids, it is necessary to remark the identification of extrachromosomal prophages or phage-like elements in two strains. A 5386 bp sequence, assembled in a unique contig, gave a perfect BLASTn match with the genome of the coliphage phi-X174 (C7347 strain). Further, the reconstruction of the C7136 genome evidenced a large cluster of 13 contigs (~90 kb), well-differentiated from the rest of the genome elements but showing a poor coverage. It contains features and large sequence segments highly homologous to P1 bacteriophages, known to lysogenize bacteria and replicate as independent low-copy number plasmid-like elements [37]. The *repA* protein of P1 and P7 phages belonged to IncY, the incompatibility group associated to C7136 strain. These plasmid-like elements have been involved in the spread of relevant antimicrobial resistance determinants (*bla*_{SHV-12}, *mcr-1* . . .) and are distributed both in human as well as natural environments [37,38].

As detailed in Table 2, among the large conjugative plasmids residing in *E. coli* from wildlife, members of the MOB_{F12} subfamily (IncF complex) were predominant, followed by MOB_{P12} (mainly, IncI1). This was not surprising, since these conjugative replicons are well-known to have higher prevalence in *E. coli*, in part due to the lower fitness cost they impose based on their capability for self-transfer repression [39]. The most remarkable was the heterogeneity of plasmids found among our antimicrobial resistance (AMR) strains.

Although the resistance rate was low among this collection of *E. coli* from wildlife (21.0%), AMR gene-carrying plasmids belonged to a variety of MOB/Inc families (MOB_{F12}/IncF, MOB_{P12}/IncI1, MOB_{F11}/IncN; MOB_{P3}/IncX1; -/IncR). In contrast to what happens in wild natural environments, the overrepresentation of particular AMR-plasmids or clones seems to be more frequent in human-influenced settings. The high antimicrobial pressures tend to select well-adapted replicons (or even clones) with inherent or newly acquired resistance genes to proliferate and rapidly spread.

Table 1. Diversity and genetic features of small plasmids found among studied *E. coli* collection. Plasmids have been named according to the strain in which they are harbored (if ≥ 1 per isolate, they appear enumerated). Those replicons unable to be typed by using PlasmidFinder database are indicated as “nd”.

Small Plasmid	Size (bp)	Relaxasa	Replicon Type	Accessory Genes (Proteins)	$\geq 99\%$ Identical Plasmids (Strain Host)
C7369-2	3371	MOB _{P51}	nd	<i>orf1</i> (hypothetical poetin)	CP012927 (human)
C8124-2	4633	MOB _{P51}	col(RNAI)	<i>orf1</i> (hypothetical poetin)	-
C7382-1	6493	MOB _{P51}	col(RNAI)	<i>cea</i> (colicin E1), <i>cei</i> (colicin E1 immunity), <i>cel</i> (colicin E1 lysis), <i>exc1</i> (entry exclusion 1), <i>exc2</i> (entry exclusion 2)	-
C7962	4203	MOB _{P51}	nd	<i>orf1</i> , <i>orf2</i> , <i>orf3</i> (hypothetical poeins)	-
C7971	4182	MOB _{P51}	nd	<i>orf1</i> , <i>orf2</i> , <i>orf3</i> (hypothetical poeins)	-
C7969	5428	-	col(RNAI)	<i>cna</i> (colicin N), <i>cni</i> (colicin N immunity), <i>cnl</i> (colicin N lysis), <i>exc1</i> (entry exclusion 1), <i>exc2</i> (entry exclusion 2)	-
C6466-1	3609	-	col(RNAI)	<i>orf1</i> (hypothetical poetin)	KU166868 (human), CP006052 (chicken)
C6466-2	1506	-	col(MG828)	<i>orf1</i> (hypothetical)	CP010877 (human)
C7369-1	2255	-	col(MG828)	<i>orf1</i> , <i>orf2</i> , <i>orf3</i> (hypothetical)	-
C7347	1736	-	col(MG828)	<i>orf1</i> (hypothetical)	-
C6466-3	5132	MOB _{Q12}	col(156)	<i>orf1</i> (hypothetical)	-
C7369-3	3904	MOB _{Q12}	col(156)	<i>orf1</i> (hypothetical)	CP012638 (human), CP019896 (beef)
C8124-3	8138	MOB _{Qu}	col(156)	<i>col</i> (colicin E8), <i>cei</i> (3 genes) (colicin immunity), <i>cel</i> (2 genes) (colicin lysis); <i>orf1</i> , <i>orf2</i> , <i>orf3</i> (hypothetical)	AP010964 (human)
C8124-1	4110	MOB _{Qu}	nd	-	-
C7382-2	3348	MOB _{V2}	nd	<i>orf1</i> (hypothetical)	-
C7328	3087	-	Col440I	<i>orf1</i> , <i>orf2</i> , <i>orf3</i> (hypothetical)	-
C7974	2717	-	nd	<i>orf1</i> , <i>orf2</i> , <i>orf3</i> (hypothetical)	-
C7973	2717	-	nd	<i>orf1</i> , <i>orf2</i> , <i>orf3</i> (hypothetical)	-

Figure 3 summarizes the genetic surrounding of the main resistance determinants found in the present study. Most of them were included in larger resistance complexes involving different hybrid transposons and insertion sequences, but the resolution of the complete structure was not possible in all cases due to the limitations of short-read sequence data. We therefore showed those well-characterized gene associations owing to a high coverage and previous experimental assays. It should be noted that resistance gene patterns predicted by in silico analyses were consistent with the phenotypic resistance patterns obtained by the disk-diffusion method.

Table 2. Genetic features of commensal *E. coli* isolates from wildlife based on in silico WGS analysis.

Strain, Origin ^a	No. Plasm.	Inc/Rep Types, Phages	Relaxase Protein ^c	Resistance Genes		Virulence Genes		CRISPR/Cas Systems		ST/PG ^e
				Chromosome	Plasmids	Chromosome	Plasmids	Subtype	No. Spacers	
C6466, BP	5	IncFII, IncX1, col(156), col(MG828) (2)	MOB _{P3} , MOB _{Q12}		-	<i>gad</i>	-	I-E	4, 6	10/A
C6468, BP	1	IncFIA-IncFIB	MOB _{F12}		<i>tet(B)</i> , <i>bla</i> _{TEM-1b} , <i>catA1</i> , <i>dfrA17</i> , <i>aadA5</i> , <i>aph(3')-Ia</i>	<i>gad</i> , <i>lpfA</i> , <i>iss</i>	-	I-E	14, 20	1642/B1
C6473, BP	1	IncFIB-IncFIC(FIIA)	MOB _{F12}		<i>tet(B)</i> , <i>bla</i> _{TEM-1b} , <i>catA1</i> , <i>dfrA17</i> , <i>aadA5</i>	<i>gad</i> , <i>lpfA</i>	<i>iss</i> , <i>cva</i>	I-E	16, 20	162/B1
C6518, BP	1	IncX1	MOB _{P3}		<i>bla</i> _{TEM-1b}	<i>gad</i> , <i>lpfA</i>	-	I-E	7, ND	3634/A
C6842, R	1	IncI1	MOB _{P12}		<i>bla</i> _{TEM-1a} , <i>strA</i> , <i>strB</i> , <i>aph(4)-Ia</i> , <i>aac(3)-IVa</i>	<i>gad</i> , <i>iss</i>	-	I-E	13, 4	10/A
C6847, R	2	IncFIA-IncFII, IncX1-X4	MOB _{F12}		-	<i>gad</i> , <i>ast</i>	<i>iss</i> , <i>iroN</i> , <i>cva</i>	-	-	322/Cl. IV
C6894, R	-	-	-		-	<i>gad</i> , <i>lpfA</i>	-	I-E	26, ND	7631/B1
C6895, R	1	IncFIA-IncFIB-IncFIC(FII)	MOB _{F12}		-	<i>gad</i> , <i>lpfA</i> , <i>iss</i>	<i>astA</i>	I-E	24, 10	295/B1
C6949, R	nd	IncFII, IncY	nd		-	<i>gad</i> , <i>lpfA</i>	-	I-E	8, 7	7629/B1
C6950, R	2	IncFIA-IncFII, IncX1-X4	MOB _{F12}		-	<i>gad</i> , <i>ast</i>	<i>iss</i> , <i>iroN</i> , <i>cva</i>	-	-	322/Cl. IV
C7030, R	-	-	-		-	<i>gad</i> , <i>lpfA</i>	-	I-E	12, 20	58/B1
C7031, R	-	-	-		-	<i>gad</i> , <i>lpfA</i> , <i>iss</i>	-	I-E	31, 23	5869/B1
C7032, R	1	IncFIB-IncFII	MOB _{F12}		-	<i>gad</i> , <i>iss</i> , <i>vat</i> , <i>pic</i> , <i>mcmA</i> , <i>mchB</i> , <i>mchC</i> , <i>mchE</i> , <i>iroN</i>	<i>cma</i>	-	-	104/B2
C7136, D	1	IncFII, IncY (phage-like plasmid)	-		-	<i>gad</i> , <i>lpfA</i> , <i>iss</i>	-	I-E	14, 18	1308/B1
C7143, D	-	-	-		-	<i>gad</i> , <i>lpfA</i>	-	I-E	25, 23	4511/B1
C7145, D	1	IncB/O/K/Z	MOB _{P12}		-	<i>gad</i> , <i>lpfA</i> , <i>iss</i> , <i>air</i> , <i>eilA</i>	<i>aaiC</i> , <i>astA</i>	I-E	17, 9	69/D
C7148, D	1	IncFIA-IncFIB-IncFII	MOB _{F12}		-	<i>gad</i> , <i>lpfA</i>	-	I-E	10, 13	906/B1

Table 2. Cont.

Strain, Origin ^a	No. Plasm.	Inc/Rep Types, Phages	Relaxase Protein ^c	Resistance Genes		Virulence Genes		CRISPR/Cas Systems		ST/PG ^e
				Chromosome	Plasmids	Chromosome	Plasmids	Subtype	No. Spacers	
C7257, D	-	-	-		-	<i>gad, lpfA</i>	-	I-E	16, 20	939/B1
C7259, D	-	-	-		-	<i>gad, lpfA</i>	-	I-E	16, 20	939/B1
C7277, D	1	IncI1	MOB _{P12}		-	<i>gad, lpfA</i>	-	I-E	9, 22	7624/B1
C7279, D	-	-	-	<i>tet(B)</i>		<i>gad, iss</i>	-	I-E	4, 3	1718/A
C7328, D	3	IncN, IncR-IncFIA (HI1), SCP ^b	MOB _{F11}		<i>tet(A), bla_{PSE-1}, bla_{CTX-M-1}, dfrA16, aadA2, mph(A), sul3</i>	<i>gad, lpfA</i>	-	I-E	31, 15	224/B1
C7347, D	1	col(MG828), phage φX174	-		-	<i>gad, iha, ireA, subA, espI</i>	-	I-E	4, 1	812/B2
C7349, D	1	IncFIB-IncFII	MOB _{F12}		-	<i>gad, lpfA, iss, iha, ireA, astA, mchC, mchB, mchF, subA^d</i>	<i>ehxA (hlyA)</i>	I-E	26, 12	26/B1
C7369, BP	4	IncFIB-IncFIC(FIIA), col(MG828), col(156), SCP ^b	MOB _{F12} , MOB _{P51} , MOB _{Q12}		-	<i>gad, lpfA, iss</i>	<i>iss, iroN, cva, cma, cba</i>	I-E	12, 27	155/B1
C7382, BP	3	IncFIB-IncC(FIIA), colRNAI, SCP ^b	MOB _{F12} , MOB _{P51} , MOB _{V2}		-	<i>gad, lpfA, iss, ast</i>	<i>iss, cva, cea</i>	I-E	11, 8	767/B1
C7570, BP	-	-	-		-	<i>gad, vat</i>	-	I-F1	9, 32	135/B2
C7962, WB	2	IncFII, SCP ^b	MOB _{F12} , MOB _{P51}		-	<i>gad, lpfA, iss, ast</i>	<i>ast, cma, cba</i>	I-E	20, 11	388/B1
C7963, WB	1	IncFIB-IncFII	MOB _{F12}		-	<i>gad, iss, iroN, vat, sfaS</i>	<i>iss, cva, cma, cba, tsh</i>	I-F1	6, 9	567/B2
C7968, WB	1	IncFIA-IncFIB	nd		-	<i>gad, lpfA</i>	-	I-E	7, 12	1423/B1
C7969, WB	2	IncFIB-FII, col(RNAI)	nd		-	<i>gad, astA</i>	<i>cna</i>	I-F1	21, 12	7630/Cl.V
C7970, WB	1	IncFIB-IncFIC (FII)	MOB _{F12}		-	<i>gad, iss</i>	<i>iroN</i>	-	-	131/B2
C7971, WB	1	SCP ^b	MOB _{P51}		-	<i>gad, astA, vat</i>	-	-	-	1170/B2
C7973, WB	2	IncI1, col(RNAI)	MOB _{P12}	<i>tet(B)</i> ^d	-	<i>gad, iss, capU</i>	-	I-E	11, 6	7632/A

Table 2. Cont.

Strain, Origin ^a	No. Plasm.	Inc/Rep Types, Phages	Relaxase Protein ^c	Resistance Genes		Virulence Genes		CRISPR/Cas Systems		ST/PG ^e
				Chromosome	Plasmids	Chromosome	Plasmids	Subtype	No. Spacers	
C7974, WB	nd	IncQ1, col(RNAI)	MOB _{P12}	<i>tet(B)</i> ^d	<i>tet(A), bla_{TEM-1b}, strA, strB, sul2</i>	<i>gad, iss, capU</i>	-	I-E	11, 6	7632/A
C7975, WB	-	-	-	-	-	<i>gad, iss, iroN, vat, mcmA, mchC, mchB, mchF</i>	-	I-F1	13, 5	625/B2
C7979, WB	1	IncFIC(FII)	MOB _{F12}	-	-	<i>gad, astA, vat</i>	-	-	-	1317/B2
C8124, WB	4	IncFIB-IncFII-IncQ1, col(RNAI), col(156), SCP ^b	MOB _{F12} , MOB _{P51} , MOB _{Qu}	-	<i>dfrA5, strA, strB, sul2</i>	<i>gad, lpfA, iss</i>	<i>iss, iroN, cva, cea</i>	I-E	5, 13	58/B1

^a BP: bird of prey; D: deer; WB: wild boar; R: rodent. ^b SCP: The strain harbors a small cryptic plasmid but its replicase shows not significant homology with known proteins. ^c nd: “non-determined” (plasmid number is unclear due to a complex contig assembly and genome reconstruction) or “not detected” (relaxase and REP domains are not present or could not be found). ^d Encoded in fragmented contigs of unclear origin (plasmid or chromosome). ^e Sequence type/phylogroup (Cl.: cryptic clade).

Antimicrobial family	Resistance gene	Genetic location	Plasmid (MOB/Inc)	Linkage to MGE	Genetic environment
β-lactams	<i>bla_{TEM-1b}</i>	Plasmid	MOB _{P23} /IncX1	Tn2	
	<i>bla_{TEM-1b}</i>	Plasmid	nd	IS26	
	<i>bla_{PSE-1}</i>	Plasmid	~IncR-IncFIA	class 1 integron (GC)	see MDR complex B
	<i>bla_{TEM-1a}</i>	Plasmid	MOB _{P12} /IncI1	ΔIS1133, IS15DII-ΔTn3	see MDR complex A
	<i>bla_{CTX-M-1}</i>	Plasmid	MOB _{F11} /IncN	IS26	
Tetracyclines	<i>tet(A)</i>	Plasmid	~IncR-IncFIA	Tn1721	see MDR complex B
	<i>tet(B)</i>	Plasmid	MOB _{F12} /IncFIA-IncFIB	Tn10	
	<i>tet(B)</i>	nd	-	Tn10	
Aminoglycosides	<i>aadA1/aadA2</i>	Plasmid	~IncR-IncFIA	class 1 integron (GC)	see MDR complex B
	<i>aadA5</i>	Plasmid	MOB _{F12} /IncFIA-IncFIB	class 1 integron (GC)	
	<i>strA/strB</i>	Plasmid	MOB _{P12} /IncI1	ΔTn5393	see MDR complex A
	<i>strA/strB</i>	Plasmid	MOB _{F12} /IncFIB-IncFII-IncQ1	IS26	
	<i>aac(3)-IVa</i>	Plasmid	MOB _{P12} /IncI1	ISEc59, IS15DII	see MDR complex A
	<i>aph(4)-Ia</i>	Plasmid	MOB _{P12} /IncI1	ISEc59, IS15DII	see MDR complex A
Macrolides	<i>Δmef(B)</i>	Plasmid	~IncR-IncFIA	IS26	see MDR complex B
	<i>mph(A)</i>	Plasmid	MOB _{F11} /IncN	IS26	
Chloramphenicols	<i>catA1</i>	Plasmid	MOB _{F12} /IncFIA-IncFIB	Tn21, ΔIS10	
	<i>cmlA</i>	Plasmid	~IncR-IncFIA	class 1 integron (GC)	see MDR complex B
Sulphonamides	<i>sul1</i>	Plasmid	MOB _{F12} /IncFIA-IncFIB	class 1 integron (3'CS)	
	<i>sul2</i>	Plasmid	MOB _{F12} /IncFIB-IncFII-IncQ1	IS26	
	<i>sul3</i>	Plasmid	~IncR-IncFIA	class 1 integron (3'CS)	see MDR complex B
Diaminopyrimidines	<i>dfr:A5</i>	Plasmid	MOB _{F12} /IncFIB-IncFII-IncQ1	class 1 integron (GC)	
	<i>dfi:A16</i>	Plasmid	~IncR-IncFIA	class 1 integron (GC)	see MDR complex B
	<i>dfi:A17</i>	Plasmid	MOB _{F12} /IncFIA-IncFIB	class 1 integron (GC)	

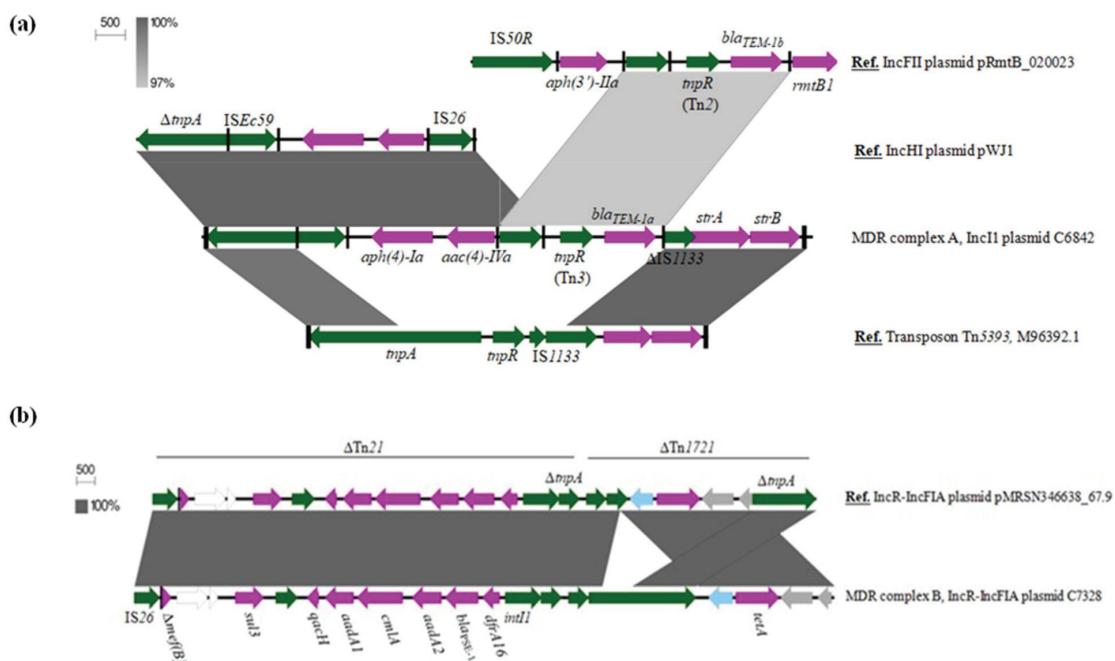


Figure 3. Genetic environment and location of antimicrobial resistance determinants identified in *E. coli* from wildlife. (a) BLAST comparison of the novel resistance complex Δ Tn5393-*bla_{TEM-1a}*-*strA*-*strB* (MDR complex A) with other reference sequences from public databases; (b) BLAST comparison of a hybrid Tn21-1721 transposon carrying an atypical class 1 integron and *tet(A)* in non-conjugative IncR-IncFIA plasmids (MDR complex B). Genes are colored according to their function: purple, resistance to drug or metal; green, recombination/transposition; blue, replication and regulation of gene expression; grey, known function; white, unknown function.

Among the five *bla*_{TEM-1} genes detected, three different genetic surroundings could be identified. The variant *bla*_{TEM-1b}, more prevalent, was located upstream of the Tn2 *tnpR* gene, as commonly occurs, but was also found flanked by two IS26 elements. The *bla*_{TEM-1a} variant, different in three nucleotides but encoding the same TEM-1 protein, was found in a truncated Tn3 transposon. An IS26 element was inserted downstream of the Tn3 *tnpR* gene, likely recruiting this IS26-*tnpR* (Tn3)-*bla*_{TEM-1a} gene pool and integrating it in a Δ Tn5393 resistance complex by homologous recombination, since no direct target site duplications were identified. Thus, the Δ Tn5393-*aph*(4)-*Ia-aac4-IVa-bla*_{TEM-1a}-*strA-strB* complex (MDR complex A in Figure 3a), not previously described, emerged from diverse insertion events within the original Tn5393 backbone. Probably, the *tnpA* transposase gene was disrupted by the inclusion of the genetic unit ISEc59-*aph*(4)-*Ia-aac4*-IS26, frequently found in many bacterial genomes, and was followed by the insertion of the IS26-*tnpR*(Tn3)-*bla*_{TEM-1a} segment due to a recombinatory event between two directly oriented IS26.

The IS26- Δ *repA-repC-sul2-strA-strB*-IS26 composite transposon, conferring resistance towards sulfonamides and streptomycin, was present in the multireplicon IncFIA-IncFIB-IncQ1 carried by C8124 strain. This cluster, composed by part of the IncQ1 prototype plasmid RSF1010 (M28829), also derived from the Tn5393 transposon. In this case, Tn5393 harboring *strA* and *strB* genes was likely transposed into the resistance gene region of RSF1010, containing the *sul2* gene and a mobile element named CR2 [40]. Secondary events removed parts of Tn5393 and CR2 regions, leading to the conserved Δ *repA-repC-sul2-strA-strB* configuration [40], which is widely spread as an IS26-bounded transposon. This MGE is present in many bacterial plasmids and integrated chromosomal structures from human, veterinary and environmental settings [40].

Regarding tetracycline efflux pumps encoding genes, *tet*(A) and *tet*(B), they were located at the well-characterized transposons Tn1721 and Tn10, respectively. In C6468 isolate, a copy of IS26 disrupted the IS10-L transposase gene. Furthermore, three class 1 integron structures were identified, containing the following gene cassette arrangements: (i) *dfrA5*; (ii) *dfrA17-aadA5* and (iii) *dfrA16-bla*_{PSE-1}-*aadA2-cmlA-aadA1*. The integron carrying the *dfrA17-aadA5* array showed the classical 3'-conserved segment (3'-CS), composed by a truncated *qac* Δ E1 and the sulfonamide resistance *sul1* genes. The results of a Norwegian survey suggested a linkage between this integron and a subpopulation of *E. coli* adapted to a human host [41]. We found this element in two different *E. coli* from wild birds (C6468, C6473), belonging to different STs and harboring distinct MOB_{F12} plasmids. However, although we could not achieve the complete assembly of these plasmids, they showed a very similar resistance gene content. Hence, the spread of *dfrA17-aadA5* array seems to be more associated to a stable plasmid-encoded resistance complex rather than to certain clonal lineages. The larger *dfrA16-bla*_{PSE-1}-*aadA2-cmlA-aadA1* array, which will be mentioned latter, was an atypical *sul3*-associated class 1 integron. Notably, *dfrA5* gene cassette carrying integron lacked the 3'-CS due to an IS26-mediated deletion. The *intI1-dfrA5*-IS26 configuration has been less reported in conventional PCR-based studies targeting integrons, owing to the absence of the hybridization site of the 3'-CS primer. However, as shown, they are present in both human and animal *E. coli* populations [42].

Figure 4 shows the PLACNETw reconstruction of the extended-spectrum β -lactamase producing C7328 strain. Plasmid 2 showed a MOB_{F11} relaxase (red node) and an IncN replicon type (yellow node) and harbored the *bla*_{CTX-M-1} in an IS26 composite transposon. An incomplete macrolide resistance cluster, with an entire *mphA* and a truncated *mrx* gene, were located downstream of *bla*_{CTX-M-1}, leading to the gene block IS26-*mph*(A)- Δ *mrx-orf477-bla*_{CTX-M-1}- Δ ISEc1-IS26. This module, first reported in IncN plasmids of human clinical isolates, has also been described in IncI and IncB/O replicons, which suggest the occurrence of IS26-mediated transposition events between different plasmid backbones [43,44]. In addition, C7328 strain carried a nontransferable IncR-FIA plasmid of around 70 kb, lacking the relaxase gene (plasmid 3 in Figure 4). Its backbone was highly homologous (99% coverage, 100% identity) to plasmid pMRSN346638_67.9, reported in a clinical *E. coli* isolate [45]. Further, both plasmids carried *aadA1/aadA2*, *cmlA1*, *bla*_{PSE-1}, *dfrA16*, *sul3* and

tet(A) genes, encoding for resistance against aminoglycosides (streptomycin), phenicols, β -lactams (penicillin, carbenicillin), trimethoprim, sulfonamides and tetracycline, respectively. Their resistance regions, almost identical, consisted of a hybrid Δ Tn21/Tn1721 transposon containing a large class 1 integron (MDR complex B in Figure 3b). The C7328 strain sequenced in this work corresponded to that included in a previous study carried out using conventional PCR-based approaches [7]. There, we demonstrated the conjugal transferability of the CTX-M-1 encoding gene, but we could not determine neither its complete genetic surrounding nor the genetic basis behind the impossibility of a parallel transference of the class 1 integron and the *tet(A)* determinant. WGS data enabled the genome reconstruction of the strain, which provides a better understanding of the genetic background and the mechanisms involved in the antimicrobial resistance spread.

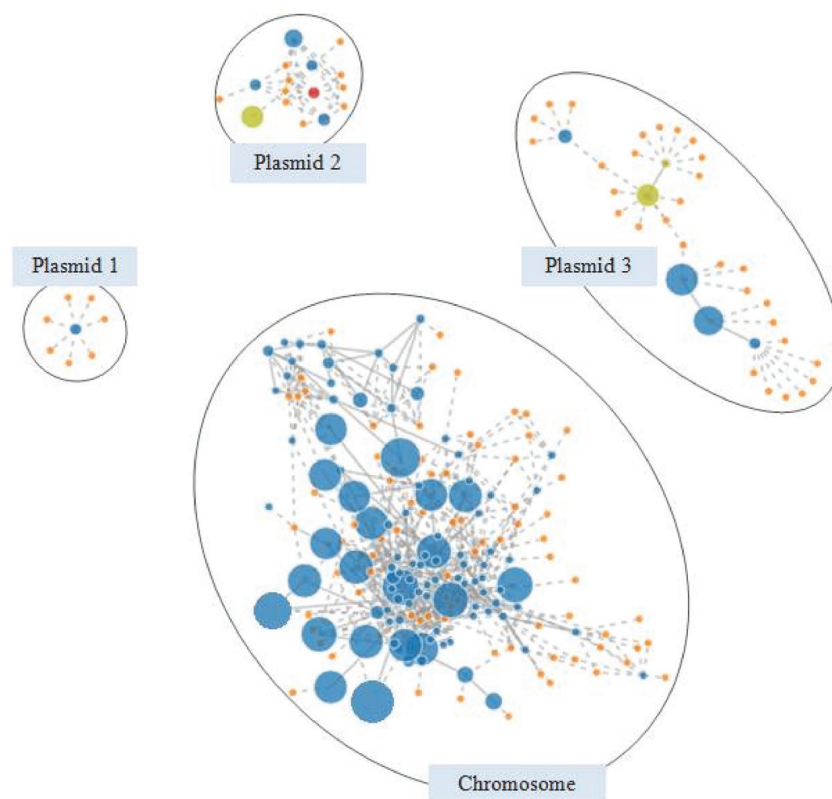


Figure 4. PLACNETw reconstruction of the *E. coli* C7328 genome. After a manual pruning of the original network, four components were differentiated: the chromosome, plasmid 1 (small cryptic plasmid), plasmid 2 (IncN replicon) and plasmid 3 (IncR-IncFIA multi-replicon). Contigs are represented as blue nodes of size proportional to their length. Orange nodes indicate reference genomes. Contigs encoding plasmid relaxases and replication proteins are colored in red and yellow, respectively.

3.3. Virulome

It is necessary to specify that, although a web-based program was used to search for the virulence traits, the high sequence homology between particular determinants (e.g., *iss*/*bor* or *cva*/*mchF*) required manual validation. Unlike acquired AMR, preferentially plasmid-encoded, virulence genes were found in a variety of MGE, including pathogenicity islands (PAIs), prophages and large virulent plasmids. Among the 25 detected virulence-related genes, the highest frequencies were observed for *gad*, *lpfA* and *iss*. Notably, at least one of the two isoenzymes of Gad (glutamate decarboxylase A and B), both in most cases, were found encoded in all genomes, invariably located in previously defined chromosomal loci [46]. These stress-related genes, involved in the decarboxylation of glutamate for the maintenance of neutral cytoplasmic pH, seem to be essential in the colonization of the host gut.

As shown in Figure 5, the occurrence of some virulence factors (VFs) varied among *E. coli* phylogroups. For instance, the fimbrial adhesin gene *lpfA* was present in all strains belonging to phylogenetic group B1, but it was absent in B2 isolates. On the contrary, vacuolin autotransporter toxin (*vat*) and, to a lesser extent, enterobactin siderophore receptor (*iroN*) were more prevalent among B2 phylogroup. These differences in VF distribution seem to be in relation with the propensity of certain *E. coli* phylogroups to cause particular infections. In this sense, *lpfA* carriage and B1 phylogroup have been associated to persistent mastitis in cattle [47], whilst B2 phylogroup is more related with avian pathogenic (APEC) and human extraintestinal pathogenic (ExPEC) *E. coli* strains.

		Adhesins			Toxins						Siderophores		Colicins/Microcins						Miscellaneous							
		<i>lpfA</i>	<i>sfaS</i>	<i>iha</i>	<i>ast</i>	<i>eltx</i>	<i>sub</i>	<i>vat</i>	<i>espI</i>	<i>aaiC</i>	<i>pic</i>	<i>iroN</i>	<i>ireA</i>	<i>cea</i> (E1)	<i>cea</i> (E8)	<i>ena</i> (N)	<i>cma</i> (M)	<i>cba</i> (B)	<i>cva</i> (V)	<i>mchF</i>	<i>gad</i>	<i>iss</i>	<i>eilA</i>	<i>capU</i>	<i>air</i>	<i>tsh</i>
A	C6466	■																			■					
	C6518	■																			■					
	C7279	■																			■					
	C7974																					■				
	C7973																					■				
B1	C6842																									
	C7369	■																				■				
	C7382	■																				■				
	C6468	■																				■				
	C6473	■																				■				
	C7143	■																				■				
	C7277	■																				■				
	C7328	■																				■				
	C7259	■																				■				
	C7136	■																				■				
	C7349	■																				■				
	C7257	■																				■				
	C7148	■																				■				
	C8124	■																				■				
	C7968	■																				■				
C6949	■																				■					
C7031	■																				■					
C7962	■																				■					
C7030	■																				■					
C6894	■																				■					
C6895	■																				■					
D	C7145																					■				
B2	C7570																					■				
	C7347																					■				
	C7975																					■				
	C7979																					■				
	C7963																					■				
	C7970																					■				
	C7971																					■				
	C7032																					■				
CLADES (IV and V)	C6847																					■				
	C6950																					■				
	C7969																					■				

Figure 5. Distribution of the virulence determinants among *E. coli* phylogroups.

Focusing on these latter ExPEC/APEC-associated VFs, *vat* and *iroNBCDE* cluster, our results suggest that their linkage to B2 phylogroup is in relation to their preferential location in chromosomal PAIs (Figure 6). As it will be described, *iroN* was also found in B1 and clade IV strains, but in these cases, encoded in large pColV-like MOB_{F12} plasmids (Figure 6a). Regarding *vat*, as previously described in human UPEC (CFT073) and avian APEC (Ec222) strains, it was located on a PAI inserted at the *thrW* tRNA locus, between the *proA* and *yagU* genes. This site seems to be a hotspot for recombination and has been found to carry different VFs such as *iroN*, *sfaS* and microcins in some *E. coli* isolates harboring PAI-III₅₃₆ [48]. However, six out of the eight strains belonging to the B2 phylogroup of our collection only carried *vat* at this insertion point (Figure 6b). Most of them also harbored *iroN*, but in a different genomic location. In fact, this siderophore gene was occupying two different PAIs inserted in the *serX* locus, the so-called PAI-CFT073-*serX* and the island V described in IHE3034 strain [49] (Figure 6a). Apart from the iron locus, PAI-CFT073-*serX* co-harbored the genetic system encoding microcin H47 (*mcmA*, *mchF*, *mchB*, *mchC*) and the *foc* operon, encoding F1C fimbriae. The genetically related SfaS adhesin, also known to play a role in uropathogenesis, was adjacent to *iroNBCDE* locus in PAI-V-IHE3034. However, as mentioned before, the plasmid-borne *iro* cluster was also frequent among *E. coli* from wildlife belonging to various phylogroups (B1, B2, clade IV). pColV-plasmids, coding for the production of colicin V (*cva* gene), were found in 18.4% of the strains. All belonged to MOB_{F12} family and carried different virulence traits such as *iroN*, *iss* (increased serum survival), *tsh* (temperature-sensitive hemagglutinin), *cma* (colicin M) and/or *cmb* (colicin B). These plasmids have been associated to the APEC pathotype and, although their

background and virulence region is variable, the so-called “conserved” portion (containing among others *iroN* cluster) shows high homology with certain chromosomal PAIs [50]. Indeed, most of the pColV plasmids identified in *E. coli* from wildlife contained *iss*, *iroN* and *cva* (C7369, C8124, C6847 and C6950 strains), a set of genes associated to the conserved virulence region of these plasmids. Taking into consideration the frequent occurrence of the mentioned ExPEC/APEC-related VFs (e.g., Vat serin-protease, *iroNBCDE* cluster) among the commensal *E. coli* population, these traits seem to primarily contribute to bacterial fitness, competitiveness and gut colonization, rather than being directly involved in infection.

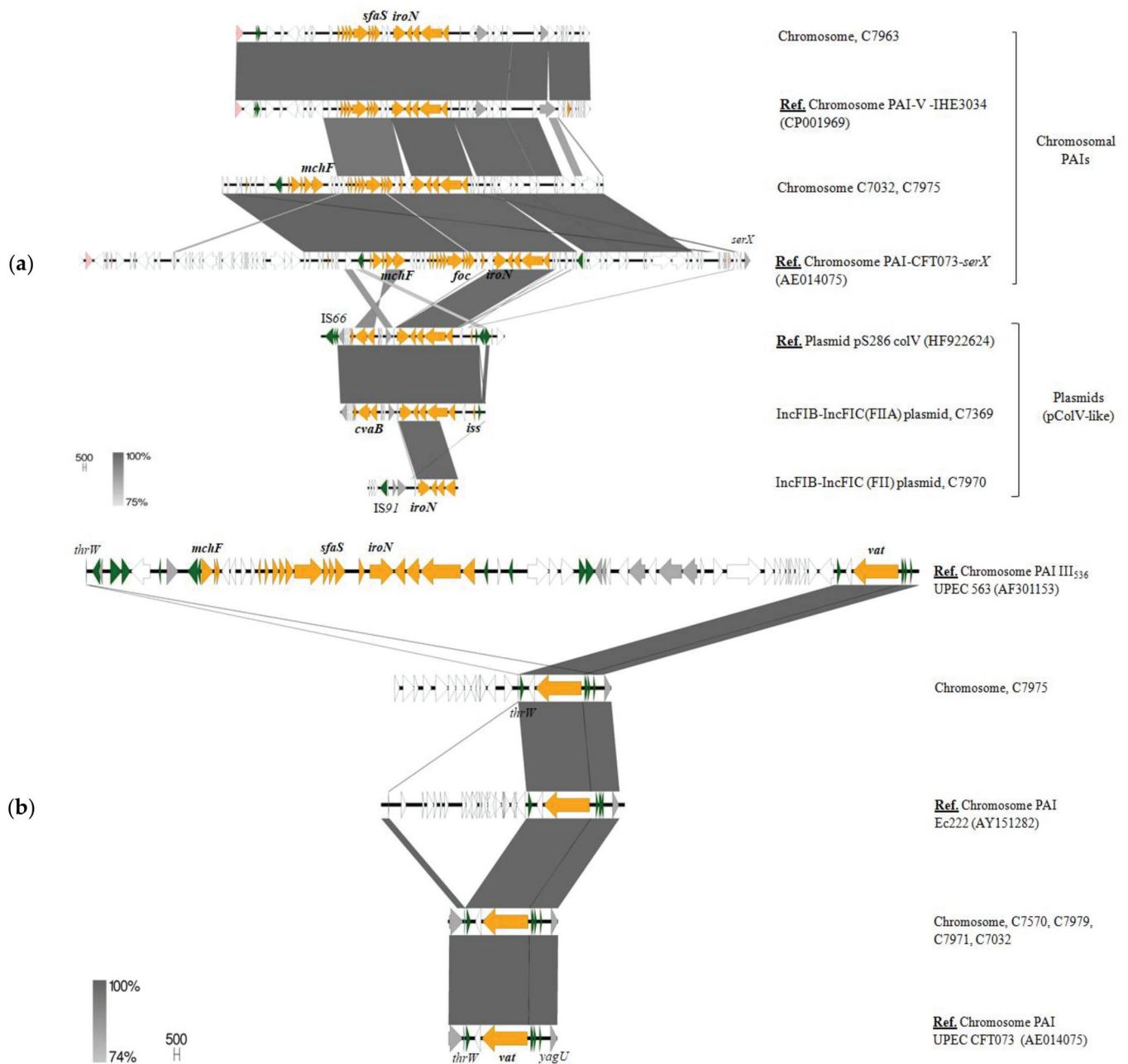


Figure 6. Genetic location and surrounding of *iroN* (a) and *vat* (b) virulence genes in some of the *E. coli* strains from our collection. The reference genomes used for BLASTn comparisons are indicated as “Ref. Genetic location-Strain name-(accession number)”. Genes are colored according to their function: orange, virulence genes; green, recombination/transposition; pink, prophage genes; grey, known function; white, unknown function.

In this regard, bacteriocin synthesis has also been linked to ExPEC strains due to the positive correlation observed between bacteriocin-encoding genes and other virulence determinants [51]. The present study on commensal *E. coli* corroborated this virulence gene association for particular bacteriocins, such as colicins V, B and M and microcin H47. As detailed in Table 3, colicins V, B and M were located on large MOB_{F12} plasmids, while microcin H47 was mainly found in chromosomal PAIs. Both platforms usually co-harbored *iroN* and other aforementioned virulence determinants. However, colE1, E8 and N encoding genes occupied very small plasmids, lacking additional traits. The integration of bacteriocins–siderophores apparently helps *E. coli* to adapt to competitive environments, such as the gut, thereby posing an advantage for commensal populations. However, in combination with additional VFs, these strains can also be more prone to invade other tissues causing extraintestinal infections [52].

Table 3. Distribution and genetic location of bacteriocins encoding genes among commensal *E. coli* from wildlife.

Colicin/Microcin	Gene Cluster	Genetic Location	Plasmid or PAI Features		Strain (Phylogroup)	Host
			MOB/Inc	Size (kb)		
E1	<i>cea, cei, cel</i>	Plasmid	MOB _{P51} /col(RNAI)	6.6	C7382 (B1)	Bird of prey
E8	<i>cea, cei, cel</i>	Plasmid	MOB _{Qu} /col(156)	8.1	C8124 (B1)	Wild boar
H47	<i>mchB, mchC, mchF, mcmA</i>	Chromosome	PAI-CFT073- <i>serX</i>	-	C7975 (B2)	Wild boar
H47	<i>mchB, mchC, mchF, mcmA</i>	Chromosome	PAI-CFT073- <i>serX</i>	-	C7032 (B2)	Rodent
H47-like	<i>mchB, mchC, mchF</i>	Chromosome		-	C7349 (B1)	Deer
M, B	<i>cma-cmi, cba-cbi</i>	Plasmid	MOB _{F12} /IncFIB-IncFIC(FIIA)	~120	C7369 (B1)	Bird of prey
M, B	<i>cma-cmi, cba-cbi</i>	Plasmid	MOB _{F12} /IncFII	~100	C7962 (B1)	Wild boar
M, B	<i>cma-cmi, cba-cbi</i>	Plasmid	MOB _{F12} /IncFIB-IncFII	~130	C7963 (B2)	Wild boar
M, B (truncated)	<i>cma-cmi, cba-cbi</i>	Plasmid	MOB _{F12} /IncFIB-IncFII	~120	C7032 (B2)	Rodent
N	<i>cna, cni, cnl</i>	Plasmid	-/col(RNAI)	5.4	C7969 (Clade V)	Wild boar
V	<i>cvaA, cvaB, cvaC</i>	Plasmid	MOB _{F12} /IncFIB-IncFIC(FIIA)	~150	C6473 (B1)	Bird of prey
V	<i>cvaA, cvaB, cvaC</i>	Plasmid	MOB _{F12} /IncFIA-IncFII	~150	C6847 (Clade IV)	Rodent
V	<i>cvaA, cvaB, cvaC</i>	Plasmid	MOB _{F12} /IncFIA-IncFII	~150	C6950 (Clade IV)	Rodent
V	<i>cvaA, cvaB, cvaC</i>	Plasmid	MOB _{F12} /IncFIB-IncFII	~130	C7963 (B2)	Wild boar
V	<i>cvaA, cvaB, cvaC</i>	Plasmid	MOB _{F12} /IncFIB-IncFII-IncQ1	~150	C8124 (B1)	Wild boar
V	<i>cvaA, cvaB, cvaC</i>	Plasmid	MOB _{F12} /IncFIB-IncFIC(FIIA)	~100	C7382 (B1)	Bird of prey
V	<i>cvaA, cvaB, cvaC</i>	Plasmid	MOB _{F12} /IncFIB-IncFIC(FIIA)	~120	C7369 (B1)	Bird of prey

The *Iss* factor, frequently associated with colibacillosis in poultry and found on pColV/BM plasmids, was detected in the genome of 55.3% *E. coli* strains. The high prevalence of this gene, located on both plasmids and/or prophage-like elements on the chromosome, has been previously reported [53]. However, three *iss* alleles have been differentiated according to their sequence [54]. The tree of Supplementary Materials Figure S3 was constructed based on most of the *iss* sequences from our collection as well as reference genes from Johnson et al., 2008 [54] and Xu et al., 2018 [53]. Johnson et al., 2008 [54] reported

a positive correlation between *iss* type 3 and human ExPEC as well as *iss* type 1 and avian APEC. Xu et al., 2018 [53] demonstrated that differences in the *iss* nucleotide sequence and mRNA copy affect its virulence and serum resistance. As shown in Supplementary Materials Figure S3, the *iss* sequence associated with higher serum livability clustered with *iss* type 1 group, while other representing a lower livability (“type 13”) clustered with *iss* type 2 group. Although all *iss* alleles were represented among *E. coli* from wildlife, most of them encoded the chromosomal, and apparently less virulent, *iss* type 2.

Considering the VFs involved, to a greater or lesser extent, in intestinal colonization and pathogenicity, we reported *astA*, *aaiC*, *air*, *eilA*, *pic*, *hlyA*, *subA*, *espI*, *ireA*, *iha* and *capU* (Figure 4). EAST-1 heat-stable toxin encoding gene (*astA*) was mainly found embedded within the putative IS1414 transposase, as previously described [55]. The C7145 strain harbored different VFs related to the enteroaggregative *E. coli* (EAEC) pathotype (Table 2). In addition to *astA*, it also carried *aaiC* (type VI secretion protein), *air* (enteroaggregative immunoglobulin-repeat protein) and *eilA* (salmonella Hil-A homolog) determinants. Although the type VI secretion system (*aaiA*-*aaiY* genes) was first described in a PAI inserted at *pheU* in the chromosome of the prototype EAEC 042 strain, in C7145, a partial cluster (*aacA*-*P*) was identified on a plasmid. This cluster shared 94% identity with that described in the pAA700-09 plasmid, which also comprised 16 genes [56]. The *air* and *eilA* determinants were found to be chromosomally co-located at *selC* locus. Regarding Pic serine protease autotransporter, implicated in intestinal colonization, it was identified in the chromosome of the C7032 strain showing a configuration highly homologous to that reported in the CFT073 strain. In C7349, genes encoding α -hemolysin (*hlyCABD*) were harbored in a MOB_{F12} plasmid, as widely occur in enteropathogenic strains [57]. Finally, the hexosyltransferase homolog CapU, which is frequently plasmid-borne in EAEC strains, was located in the chromosome of C7973 and C7974 strains, conserving the configuration reported at the same locus in the prototype EAEC 042 strain.

3.4. Phylogenomics

A phylogenetic tree based on single nucleotide polymorphisms (SNPs) in the *core* genome was constructed in order to examine the occurrence of potential adaptive mutations associated with the host. In a first approach, with the aim of defining the genetic diversity of our *E. coli* collection, we performed a phylogenetic analysis considering only the *core* SNPs of the 38 genomes from wild animals. The resulting tree was highly concordant with the population structure inferred by MLST. Thus, additional *E. coli* strains were selected from Enterobase, based on MLST criteria. Up to 31 *E. coli* genomes per sequence type, representing various sources and geographic locations, were downloaded from the NCBI BioProjects database. Information about these genomes is detailed in Supplementary Materials Table S2. The final dataset for the phylogenomic reconstruction, represented by the tree in Supplementary Materials Figure S4, comprised 280 genomes, including those belonging to our collection ($n = 38$). Of the total gene clusters (108,518), only 0.6% (674) were shared among all genomes, resulting in a *core* length of $575,173.5 \pm 3802.95$ bp. This small *core* genome evidenced the high genetic diversity of the studied collection, particularly highlighted by the remarkable divergence of cryptic *Escherichia* clades (C6847 and C6950 strains-Clade IV/ST322; C7969 strain-Clade V/ST7630). Among the five major clades defined by Walk et al., 2009 [58] (I to V), members of clades II, III, IV and V appeared overrepresented in samples from soils, rivers and aquatic sediments. Wild animals seemed to be spillover hosts of these environmentally well-adapted *Escherichia* lineages. Luo et al., 2009 [59] analyzed nine genomes belonging to these cryptic clades and, by comparing them with enteric *E. coli* strains (including clade I), demonstrated that both differed in the content of a significant number of genes. Further, they did not find the genetic exchange of *core* genes between enteric and environmental clades, contrary to what it was observed within members of each group. Thus, they suggested that the evolution of *Escherichia* genus appeared to be mainly driven by gene acquisition and/or deletion, rather than homologous recombination, which supports our observations. The other obvious reason that could affect

the *core* size is the high number of sequences taken from public databases. When using sequence data from different research groups, two questions affecting its quality should be taken into consideration: genome annotation and sequence quality [60]. Although we avoided the bias by assembling and predicting CDS from all reference genomes using Velvet [61] and Prodigal software [62], the quality of the sequences likely varies.

As previously observed [60], the *core* genome tree was highly consistent with MLST data as well as with the phylogroup distribution of our bacterial collection. Only ST7632, ST1642, ST4511, ST939, ST5869 and ST767 fit inside other ST clusters. ST767 and ST7632 are single locus variants of ST58 and ST10, respectively, which explains their close relation. The ST224 clade is split by ST939 and ST1642, which show two locus change, but also by ST5869 and ST4511. These latter ones shared only two and three alleles with ST224, respectively, a fact that remarks the higher resolution of WGS data (Figure 7). Although MLST is highly discriminatory, discordant relationships might be attributed to recombination events in MLST loci.

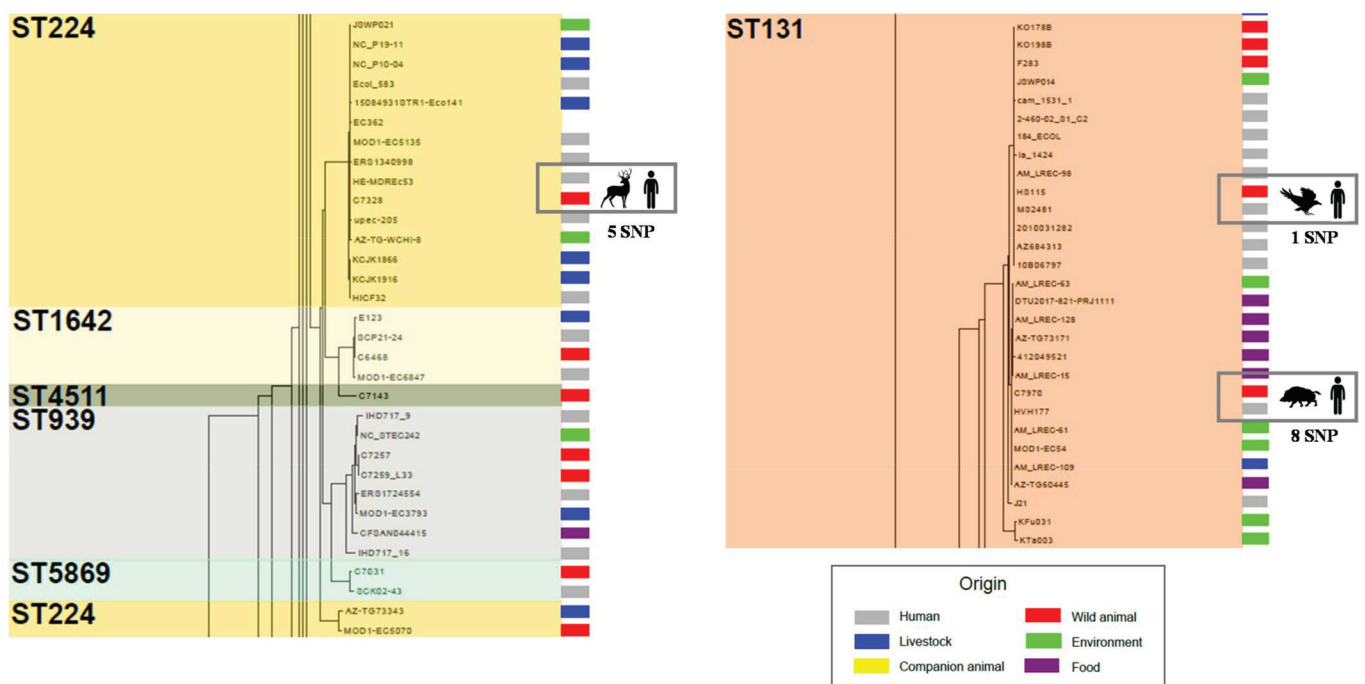


Figure 7. Overview of two of the most populated branches of the *core* genome tree. The very low number of SNP differences observed between strains from wild animals and humans are highlighted in squares. Moreover, ST224 cluster also shows the higher discriminatory power of WGS when compared to conventional MLST typing.

Regarding the distinct clusters of the tree, most of them contained strains recovered from different sources and distant locations. Further, in many cases, the SNP number (Figure 7, Supplementary Materials Table S3) was significantly lower between strains from different species of animals than within members of the same species. We can point out various examples observed among some of the most populated tree branches (ST10, ST58, ST224 and ST131 clusters). For instance, in the ST58 clade, with an average of 282.8 SNPs between its members, the C7030 strain belonging to a rodent showed the closest relationship with the bovine origin strain ERS1801995 (33 SNPs). A similar observation was reported among members of the widely distributed ST10 (SNP average: 670.6). The *core* genome of the C6842 isolate (rodent, Spain) shared more similarities with MOD1-EC6577 (bovine, Canada) than with other *E. coli* from wildlife hosts. More remarkable was the extremely low number of SNPs (≤ 10) that differentiated the C7328 strain (deer, Spain) from others belonging to humans (HE-MDREc53, USA; MOD1EC5135, USA; Ecol583, USA) or poultry (NC_P10-04, Uganda), even more so considering the average SNP within ST224 ($n = 625.8$) (Figure 7). That was also the case (≤ 10 SNP) between C7970 (wild boar,

Spain) and HVH 177 (human, Denmark), two members of the major ST131 cluster (SNP average = 259.6) (Figure 7). Thus, no clear evidence of adaptive mutations associated to the *core* genome were found in strains from different hosts. Although further studies with a more homogeneous dataset are required to generalize this assumption, our results reflect an on-going inter-host transmission of *E. coli* strains. Adaptation to the host environment does not seem to leave a genetic mark in the bacterial *core* genome.

4. Conclusions

This study contributes to better characterize the commensal gut niche from wildlife, which is crucial for understanding the adaptive pathways and genome diversification of *E. coli*, the role of different elements in bacterial fitness and the flow of genes and gene modules among interconnected ecosystems. Whole genome sequencing enables a more consistent characterization of the composition and diversity of the plasmidome, facilitating the study of small replicons, the differentiation of extra-chromosomal phages or phage-like elements and the detection of non-typeable plasmids. Further, it offers a general view of different genomic regions related to immunity, phylogeny, pathogenesis or resistance, which favors more robust assumptions than those made on the basis of a few set of genes screened by conventional methods.

Our analyses seem to rule out the occurrence of a parallel evolution of the *E. coli* genome in the gut of wild animals. Bacteria from humans or highly human-influenced settings exhibit similar genetic patterns in CRISPR-Cas systems, plasmids or virulence/resistance gene-carrying modules. These observations, together with the absence of significant genetic changes in their *core* genome, suggest an ongoing flow of both mobile elements and *E. coli* lineages between human and natural ecosystems.

Supplementary Materials: The following are available online at <http://www.mdpi.com/xxx/s1>. Figure S1: Structural diversity of the CRISPR/Cas I-E and I-F1 systems regarding the module located between the two CRISPR arrays, comprising the *cas* set of genes and adjacent ORFs. CRISPR arrays depicted in this figure illustrate only its position within the CRISPR/Cas systems (the spacer repertoire of each isolate is represented in Fig. 1). The *cas*-E (*cas2*, *cas1*, *cas6*, *cas5*, *cas7*, *cse2*, *cse1*, and *cas3*) and *cas*-F1 (*cas1-cas2/3-csy1-csy2-csy3-cas6f*) genes, in pink, are located between CRISPR-1/CRISPR-2 loci in I-E system and CRISPR-3/CRISPR 4 in I-F1 system., Figure S2: Phylogenetic tree constructed according to the concatenated amino acid sequences of the *cas* genes. Colour circles denote the different phylogenetic groups (dark blue: B1; red: A; green: D; pink: B2; light blue: Clade V). The concatenated *cas*-E genes of *E. coli* K-12 MG1655 are also included as reference., Figure S3: Phylogenetic tree constructed according to the nucleotide sequences of 21 putative *iss* genes from our *E. coli* collection. Some reference sequences used by Johnson *et al.* 2008 to define the different *iss* alleles are marked with a red circle. Two reference sequences, distinctly associated with high (AF042279) and low (*iss* type13) serum livability by Xu *et al.* 2018, appear indicated with a blue circle., Figure S4: Phylogenetic relationship of the 38 studied *E. coli* and 242 reference genomes (Table S2). The tree was based on a $575,173.5 \pm 3,802.95$ bp *core* genome including 674 orthologous common genes, from a total of 108,518 gene clusters, with at least 80% identity and 60% coverage and using 100 bootstrapping replicates. ST types are highlighted in different colors and origin has been indicated in the right part. The pairwise SNP distance matrix used to build the tree is shown in Table S3; Table S1: Genomic features of the 38 *E. coli* genomes from wildlife. Table S2: Information about the genome sequences downloaded from the NCBI for the construction of the phylogenomic *core* tree. Table S3: Pairwise SNP distance matrix calculated from the *core* genome in the set of 242 *E. coli* full genomes from the NCBI database plus 38 *E. coli* genomes from this study. Genomes are ordered according to their Sequence Type (ST).

Author Contributions: C.T. and F.d.l.C. conceived the study; C.A.A. and M.d.T. performed the experimental work and the genomic analysis. C.T., M.d.T. and F.d.l.C. supervised the study; C.A.A. wrote the first draft of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by project SAF2016-76571-R and PID2019-106158RB-I00 from the Agencia Estatal de Investigación (AEI) of Spain and the Fondo Europeo de Desarrollo

Regional (FEDER) of EU. During the experimental work of this study, Carla Andrea Alonso Alonso had a predoctoral fellowship FPI from MINECO.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Raw sequence reads reported in this paper were deposited under NCBI BioProject PRJNA699864.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Touchon, M.; Hoede, C.; Tenaillon, O.; Barbe, V.; Baeriswyl, S.; Bidet, P.; Bingen, E.; Bonacorsi, S.; Bouchier, C.; Bouvet, O.; et al. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.* **2009**, *5*, e1000344. [[CrossRef](#)] [[PubMed](#)]
2. De Toro, M.; Garcillán-Barcia, M.P.; de la Cruz, F. Plasmid diversity and adaptation analyzed by massive sequencing of *Escherichia coli* plasmids. *Microbiol. Spectr.* **2014**, *2*. [[CrossRef](#)] [[PubMed](#)]
3. Gillings, M.R.; Stokes, H.W. Are humans increasing bacterial evolvability? *Trends Ecol. Evol.* **2012**, *27*, 346–352. [[CrossRef](#)] [[PubMed](#)]
4. Johnson, J.R.; Kuskowski, M.A.; Gaiowski, A.; Sahm, D.F.; Karlowsky, J.A. Virulence characteristics and phylogenetic background of multidrug-resistant and antimicrobial-susceptible clinical isolates of *Escherichia coli* from across the United States 2000–2001. *J. Infect. Dis.* **2004**, *190*, 1739–1744. [[CrossRef](#)]
5. Tenaillon, O.; Skurnik, D.; Picard, B.; Denamur, E. The population genetics of commensal *E. coli*. *Nat. Rev. Microbiol.* **2010**, *8*, 207–217. [[CrossRef](#)]
6. Alcalá, L.; Alonso, C.A.; Simón, C.; González-Esteban, C.; Orós, J.; Rezusta, A.; Ortega, C.; Torres, C. Wild birds, frequent carriers of extended-spectrum β -lactamase (ESBL) producing *Escherichia coli* of CTX-M and SHV-12 types. *Microb. Ecol.* **2016**, *72*, 861–869. [[CrossRef](#)]
7. Alonso, C.A.; González-Barrio, D.; Tenorio, C.; Ruiz-Fons, F.; Torres, C. Antimicrobial resistance in faecal *Escherichia coli* isolates from farmed reed deer and wild small mammals. Detection of a multiresistant *E. coli* producing extended-spectrum beta-lactamase. *Comp. Immunol. Microbiol. Infect. Dis.* **2016**, *45*, 34–39. [[CrossRef](#)]
8. Alonso, C.A.; González-Barrio, D.; Ruiz-Fons, F.; Ruiz-Ripa, L.; Torres, C. High frequency of B2 phylogroup among non-clonally related fecal *Escherichia coli* isolates from wild boars, including the lineage ST131. *FEMS Microbiol. Ecol.* **2017**, *93*. [[CrossRef](#)]
9. Alonso, C.A.; Alcalá, L.; Simón, C.; Torres, C. Novel sequence types of extended-spectrum and acquired AmpC beta-lactamase producing *Escherichia coli* and *Escherichia* clade V isolated from wild mammals. *FEMS Microbiol. Ecol.* **2017**, *93*, fix097. [[CrossRef](#)]
10. Clermont, O.; Christenson, J.K.; Denamur, E.; Gordon, D.M. The Clermont *Escherichia coli* phylo-typing method revisited: Improvement of specificity and detection of new phylo-groups. *Environ. Microbiol. Rep.* **2013**, *5*, 58–65. [[CrossRef](#)]
11. Vielva, L.; de Toro, M.; Lanza, V.; de la Cruz, F. PLACNETw: A web-based tool for plasmid reconstruction from bacterial genomes. *Bioinformatics* **2017**, *33*, 3796–3798. [[CrossRef](#)]
12. Zankari, E.; Hasman, H.; Cosentino, S.; Vestegaard, M.; Rasmussen, S.; Lund, O.; Aarestrup, F.M.; Larsen, M.V. Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **2012**, *67*, 2640–2644. [[CrossRef](#)]
13. Joensen, K.G.; Scheutz, F.; Lund, O.; Hasman, H.; Kaas, R.S.; Nielsen, E.M.; Aarestrup, F.M. Real-time whole-genome sequencing for routine typing, surveillance and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* **2014**, *52*, 1501–1510. [[CrossRef](#)]
14. Grissa, I.; Vergnaud, G.; Pourcel, C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinform.* **2007**, *8*, 172. [[CrossRef](#)]
15. Sullivan, M.J.; Petty, N.K.; Beatson, S.A. Easyfig: A genome comparison visualizer. *Bioinformatics* **2011**, *27*, 1009–1010. [[CrossRef](#)]
16. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)]
17. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for clustering the next generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152. [[CrossRef](#)]
18. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **2006**, *22*, 2688–2690. [[CrossRef](#)]
19. Ishino, Y.; Shinagawa, H.; Makino, K.; Amemura, M.; Nakata, A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **1987**, *169*, 5429–5433. [[CrossRef](#)]
20. Mojica, F.J.; Juez, G.; Rodríguez-Valera, F. Transcription at different salinities of *Haloflex mediterranei* sequences adjacent to partially modified PstI sites. *Mol. Microbiol.* **1993**, *9*, 613–621. [[CrossRef](#)]
21. Makarova, K.S.; Grishin, N.V.; Shabalina, S.A.; Wolf, Y.I.; Koonin, E.V. A putative RNA interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct.* **2006**, *1*, 7. [[CrossRef](#)]

22. Aydin, S.; Personne, Y.; Newire, E.; Laverick, R.; Russell, O.; Roberts, A.P.; Erne, V.I. Presence of type I-F CRISPR-Cas systems is associated with antimicrobial susceptibility in *Escherichia coli*. *J. Antimicrob. Chemother.* **2017**, *72*, 2213–2218. [[CrossRef](#)]
23. Westra, E.R.; Pul, U.; Heidrich, N.; Heidrich, N.; Jore, M.M.; Lundgren, M.; Stratmann, T.; Wurm, R.; Raine, A.; Mescher, M.; et al. H-NS-mediated repression of CRISPR based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol. Microbiol.* **2010**, *77*, 1380–1393. [[CrossRef](#)]
24. Cady, K.C.; Bondy-Denomy, J.; Heussler, G.E.; Davidson, A.R.; O’Toole, G.A. The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered phages. *J. Bacteriol.* **2012**, *194*, 5728–5738. [[CrossRef](#)]
25. Almendros, C.; Guzmán, N.M.; Díez-Villaseñor, C.; García-Martínez, J.; Mojica, F.J.M. Target motifs affecting natural immunity by a constitutive CRISPR-Cas System in *Escherichia coli*. *PLoS ONE* **2012**, *7*, e50797. [[CrossRef](#)]
26. Makarova, K.S.; Wolf, Y.I.; Iranzo, J.; Shmakov, S.A.; Alkhnbashi, O.S.; Brouns, S.J.J.; Charpentier, E.; Cheng, D.; Haft, D.H.; Horvath, P.; et al. Evolutionary classification of CRISPR-Cas system: A burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **2020**, *18*, 67–83. [[CrossRef](#)]
27. Almendros, C.; Mojica, F.J.M.; Díez-Villaseñor, C.; Guzmán, N.M.; García-Martínez, J. CRISPR-Cas functional module exchange in *Escherichia coli*. *mBio* **2014**, *5*, e00767-13. [[CrossRef](#)]
28. Liu, S.; Jin, D.; Lan, R.; Wang, Y.; Meng, Q.; Dai, H.; Lu, S.; Hu, S.; Xu, J. *Escherichia marmotae* sp. nov., isolated from faeces of *Marmota himalayana*. *Int. J. Syst. Evol. Microbiol.* **2015**, *65*, 2130–2134. [[CrossRef](#)]
29. Kupczok, A.; Landan, G.; Dagan, T. The contribution of genetic recombination to CRISPR array evolution. *Genome Biol. Evol.* **2015**, *7*, 1925–1939. [[CrossRef](#)]
30. Díez-Villaseñor, C.; Almendros, C.; García-Martínez, J.; Mojica, F.J.M. Diversity of CRISPR loci in *Escherichia coli*. *Microbiology* **2010**, *156*, 1351–1361. [[CrossRef](#)]
31. Touchon, M.; Charpentier, S.; Clermont, O.; Rocha, E.P.; Denamur, E.; Branger, C. CRISPR distribution within the *Escherichia coli* species is not suggestive of immunity-associated diversifying selection. *J. Bacteriol.* **2011**, *193*, 2460–2467. [[CrossRef](#)] [[PubMed](#)]
32. Nicolas-Chanoine, M.H.; Bertrand, X.; Madec, J.Y. *Escherichia coli* ST131, an Intriguing Clonal Group. *Clin. Microbiol. Rev.* **2014**, *27*, 543–574. [[CrossRef](#)] [[PubMed](#)]
33. Smillie, C.; Garcillán-Barcia, M.P.; Francia, M.V.; Rocha, E.P.; de la Cruz, F. Mobility of plasmids. *Microbiol. Mol. Biol. Rev.* **2010**, *74*, 434–452. [[CrossRef](#)] [[PubMed](#)]
34. Spaková, T.; Fecskeová, L.K.M.; Javorský, P.; Pristas, P. Two *rep* genes in small cryptic plasmid pKSTt21 of *Escherichia coli*. *Curr. Microbiol.* **2013**, *67*, 437–441. [[CrossRef](#)]
35. Brolund, A.; Franze’n, O.; Melefors, O.; Tegmark-Wisell, K.; Sandegren, L. Plasmidome-analysis of ESBL-producing *E. coli* using conventional typing and high-throughput sequencing. *PLoS ONE* **2013**, *8*, e65793. [[CrossRef](#)]
36. Attéré, S.A.; Vincent, A.T.; Paccaud, M.; Frenette, M.; Charette, S.J. The role for the small cryptic plasmids as moldable vectors for genetic innovation in *Aeromonas salmonicida* subsp. *salmonicida*. *Front. Genet.* **2017**, *8*. [[CrossRef](#)]
37. Billard-Pomares, T.M.; Fouteau, S.; Jacquet, M.E.; Roche, D.; Barbe, V.; Castellanos, M.; Bouet, J.Y.; Cruveiller, S.; Médique, C.; Blanco, J.; et al. Characterization of a P1-like bacteriophage carrying an SHV-2 extended-spectrum β -lactamase from an *Escherichia coli* strain. *Antimicrob. Agents Chemother.* **2014**, *58*, 6550–6557. [[CrossRef](#)]
38. Zhang, C.; Feng, Y.; Liu, F.; Jiang, H.; Qu, Z.; Lei, M.; Wang, J.; Zhang, B.; Hu, Y.; Ding, J.; et al. A phage-like IncY plasmid carrying the *mcr-1* gene in *Escherichia coli* from a pig farm in China. *Antimicrob. Agents Chemother.* **2017**, *61*, e02035. [[CrossRef](#)]
39. Haft, R.J.; Mittler, J.E.; Traxler, B. Competition favours reduced cost of plasmids to host bacteria. *ISME J.* **2009**, *3*, 761–769. [[CrossRef](#)]
40. Yau, S.; Liu, X.; Djordjevic, S.P.; Hall, R.M. RSF1010-like plasmids in Australian *Salmonella enteric* serovar *Typhimurium* and origin of their *sul2-strA-strB* antibiotic resistance gene cluster. *Microb. Drug Resist.* **2010**, *16*, 249–252. [[CrossRef](#)]
41. Sunde, M.; Simonsen, G.S.; Slettemeas, J.S.; Hockerman, I.; Norstrom, M. Integron, plasmid and host strain characteristics of *Escherichia coli* from humans and food included in the Norwegian Antimicrobial Resistance Monitoring Programs. *PLoS ONE* **2015**, *10*, e0128797. [[CrossRef](#)]
42. Dawes, F.E.; Kuzevski, A.; Bettelheim, K.A.; Hornitzky, M.A.; Djordjevic, S.P.; Walker, M.J. Distribution of class 1 integrons with IS26-mediated deletions in their 3’-CS conserved segments in *Escherichia coli* of human and animal origin. *PLoS ONE* **2010**, *5*, e12754. [[CrossRef](#)]
43. Cullik, A.; Pfeifer, Y.; Prager, R.; von Baum, H.; Witte, W. A novel IS26 structure surrounds *bla*CTX-M genes in different plasmids from German clinical *Escherichia coli* isolates. *J. Med. Microbiol.* **2010**, *59*, 580–587. [[CrossRef](#)]
44. Wang, J.; Stephan, R.; Power, K.; Yan, Q.; Hächler, H.; Fanning, S. Nucleotide sequences of 16 transmissible plasmids identified in nine multidrug-resistant *Escherichia coli* isolates expressing an ESBL phenotype isolated from food-producing animals and healthy humans. *J. Antimicrob. Chemother.* **2014**, *69*, 2658–2668. [[CrossRef](#)]
45. Snesrud, E.; Ong, A.C.; Corey, B.; Kwak, Y.I.; Clifford, R.; Gleeson, T.; Wood, S.; Whitman, T.J.; Lesho, E.P.; Hinkle, M.; et al. Analysis of serial isolates of *mcr-1*-positive *Escherichia coli* reveals a highly active IS*Apl1* transposon. *Antimicrob. Agents Chemother.* **2017**, *61*, e00056-17. [[CrossRef](#)]
46. Smith, D.K.; Kassam, T.; Singh, B.; Elliott, J.F. *Escherichia coli* has two homologous glutamate decarboxylase genes that map to distinct loci. *J. Bacteriol.* **1992**, *174*, 5820–5826. [[CrossRef](#)]

47. Dogan, B.; Rishniw, M.; Bruant, G.; Harel, J.; Schukken, Y.H.; Simpson, K.W. Phylogroup and *lpfA* influence epithelial invasion by mastitis associated *Escherichia coli*. *Vet. Microbiol.* **2012**, *159*, 163–170. [[CrossRef](#)]
48. Parham, N.J.; Pollard, S.J.; Desvaux, M.; Scott-Tucker, A.; Liu, C.; Fivian, A.; Henderson, I.R. Distribution of the serine protease autotransporters of the *Enterobacteriaceae* among extraintestinal clinical isolates of *Escherichia coli*. *J. Clin. Microbiol.* **2005**, *43*, 4076–4082. [[CrossRef](#)]
49. Moriel, D.G.; Bertoldi, I.; Spagnuolo, A.; Marchi, S.; Rosini, R.; Nesta, B.; Pastorello, I.; Corea, V.A.; Torricelli, G.; Cartocci, E.; et al. Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 9072–9077. [[CrossRef](#)]
50. Johnson, T.J.; Siek, K.E.; Johnson, S.J.; Nolan, L.K. DNA sequence of a ColV plasmid and prevalence of selected plasmid-encoded virulence genes among avian *Escherichia coli* strains. *J. Bacteriol.* **2006**, *188*, 745–758. [[CrossRef](#)]
51. Micenková, L.; Štaudová, B.; Bosák, J.; Mikalová, L.; Littnerová, S.; Vrba, M.; Ševčíková, A.; Woznicová, V.; Šmajš, D. Bacteriocin-encoding genes and ExPEC virulence determinants are associated in human fecal *Escherichia coli* strains. *BMC Microbiol.* **2014**, *14*, 109. [[CrossRef](#)]
52. Martin, P.; Tronnet, S.; Garcia, C.; Oswald, E. Interplay between siderophores and colibactin genotoxin in *Escherichia coli*. *IUBMB Life.* **2017**, *69*, 435–441. [[CrossRef](#)]
53. Xu, W.Y.; Li, Y.J.; Fan, C. Different loci and mRNA copy number of the increased serum survival gene of *Escherichia coli*. *Can. J. Microbiol.* **2018**, *64*, 147–154. [[CrossRef](#)]
54. Johnson, T.J.; Wannemuehler, Y.M.; Nolan, L.K. Evolution of the *iss* gene in *Escherichia coli*. *Appl. Environ. Microbiol.* **2008**, *74*, 2360–2369. [[CrossRef](#)]
55. McVeigh, A.; Fasano, A.; Scott, D.A.; Jelacic, S.; Moseley, S.L.; Robertson, D.C.; Savarino, S.J. IS1414, an *Escherichia coli* insertion sequence with a heat-stable enterotoxin gene embedded in a transposase-like gene. *Infect. Immun.* **2000**, *68*, 5710–5715. [[CrossRef](#)]
56. Jønsson, R.; Struve, C.; Boll, E.J.; Boisen, N.; Joensen, K.G.; Sørensen, C.A.; Jensen, B.H.; Scheutz, F.; Jenssen, H.; Krogfelt, K.A. A Novel pAA Virulence Plasmid Encoding Toxins and Two Distinct Variants of the Fimbriae of Enteroaggregative *Escherichia coli*. *Front. Microbiol.* **2017**, *8*, 263. [[CrossRef](#)]
57. Burgos, Y.; Beutin, L. Common origin of plasmid encoded alpha-hemolysin genes in *Escherichia coli*. *BMC Microbiol.* **2010**, *10*, 193. [[CrossRef](#)] [[PubMed](#)]
58. Walk, S.T.; Alm, E.W.; Gordon, D.M.; Ram, J.L.; Toranzos, G.A.; Tiedje, J.M.; Whittam, T.S. Cryptic lineages of the genus *Escherichia*. *Appl. Environ. Microbiol.* **2009**, *75*, 6534–6544. [[CrossRef](#)] [[PubMed](#)]
59. Luo, C.; Walk, S.T.; Gordon, D.M.; Feldgarden, M.; Tiedje, J.M.; Konstantinidis, K.T. Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 7200–7205. [[CrossRef](#)]
60. Kaas, R.S.; Friis, C.; Ussery, D.W.; Aarestrup, F.M. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. *BMC Genom.* **2012**, *13*, 577. [[CrossRef](#)] [[PubMed](#)]
61. Zerbino, D.R. Using the Velvet de novo assembler for short-read sequencing technologies. *Curr. Protoc. Bioinform.* **2010**. [[CrossRef](#)]
62. Hyatt, D.; Chen, G.L.; LoCascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **2010**, *11*, 119. [[CrossRef](#)]