



Article

Global Genomic Epidemiology of *Escherichia coli* (ExPEC) ST38 Lineage Revealed a Virulome Associated with Human Infections

Erica L. Fonseca ^{1,*} , Sergio M. Morgado ¹ , Raquel V. Caldart ² and Ana Carolina Vicente ¹

¹ Laboratório de Genética Molecular de Microrganismos, Instituto Oswaldo Cruz—FIOCRUZ, Rio de Janeiro 21040-360, RJ, Brazil

² Centro de Ciências da Saúde, Universidade Federal de Roraima, Boa Vista 69300-000, RR, Brazil

* Correspondence: ericafon@ioc.fiocruz.br; Tel.: +55-21-3865-8176

Abstract: Background: Most of the extraintestinal human infections worldwide are caused by specific extraintestinal pathogenic *Escherichia coli* (ExPEC) lineages, which also present a zoonotic character. One of these lineages belongs to ST38, a high-risk globally disseminated ExPEC. To get insights on the aspects of the global ST38 epidemiology and evolution as a multidrug-resistant and pathogenic lineage concerning the three axes of the One Health concept (humans, animals, and natural environments), this study performed a global phylogenomic analysis on ST38 genomes. Methods: A phylogenetic reconstruction based on 376 ST38 genomes recovered from environments, humans, livestock, and wild and domestic animals in all continents throughout three decades was performed. The global information concerning the ST38 resistome and virulome was also approached by in silico analyses. Results: In general, the phylogenomic analyses corroborated the zoonotic character of the ExPEC ST38, since clonal strains were recovered from both animal and human sources distributed worldwide. Moreover, our findings revealed that, independent of host sources and geographic origin, the genomes were distributed in two major clades (Clades 1 and 2). However, the ST38 accessory genome was not strictly associated with clades and sub-clades, as found for the type 2 T3SS ETT2 that was evenly distributed throughout Clades 1 and 2. Of note was the presence of the *Yersinia pestis*-like high-pathogenicity island (HPI) exclusively in the major Clade 2, in which prevails most of the genomes from human origin recovered worldwide (2000 to 2020). Conclusions: This evidence corroborates the HPI association with successful *E. coli* ST38 establishment in human infections.

Keywords: phylogenomics; zoonosis; high-risk; genomic epidemiology; high-pathogenicity island; accessory genome; ETT2



Citation: Fonseca, E.L.; Morgado, S.M.; Caldart, R.V.; Vicente, A.C. Global Genomic Epidemiology of *Escherichia coli* (ExPEC) ST38 Lineage Revealed a Virulome Associated with Human Infections. *Microorganisms* **2022**, *10*, 2482. <https://doi.org/10.3390/microorganisms10122482>

Academic Editor: Flavia Indrio

Received: 11 November 2022

Accepted: 13 December 2022

Published: 15 December 2022

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



Copyright: © 2022 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 commensal bacterium and a versatile pathogen capable of causing intestinal and extraintestinal infections. Extraintestinal pathogenic *E. coli* (ExPEC) presents a complex phylogeny and comprises many lineages with enhanced metabolic repertoires that affect not only humans' but also animals' health, besides occurring in the environment [1–5]. Due to its remarkable genome plasticity, the ExPEC group is able to colonize a variety of host sites, and includes uropathogenic *E. coli* (UPEC), neonatal meningitis *E. coli* (NMEC), sepsis-associated *E. coli* (SEPEC), and avian pathogenic *E. coli* (APEC) [6].

The ST38, together with other ExPEC lineages (ST131, ST405, and ST648), emerged recently as the globally dominant strains isolated from extraintestinal infections. The treatment of infections caused by such ExPEC lineages has become a challenge to public health due to the increase in the prevalence of multidrug-resistant strains presenting resistance to the first-line and the last-resort antibiotics [7], raising concerns about how ExPEC, such as ST38, evolves and diversifies in terms of virulence and antibiotic resistance.

Interestingly, some ST38 strains may present a uropathogenic/enteroaggregative hybrid character, due to the presence of both intestinal and extraintestinal virulence traits, which may contribute to the enhancement of ST38 virulence [8,9]. In spite of this, there are few studies focusing on the characterization of clinical hybrid strains [10].

Moreover, the ST38 has been associated with the global dissemination of the extended spectrum β -lactamase (ESBL)-coding *bla*_{CTX-M} gene [1,2,4]. On the other hand, studies on the occurrence of carbapenemase genes, such as *bla*_{NDM} and *bla*_{OXA}, in ST38 are scarce. Particularly, a recent study on ST38 harboring the carbapenemase OXA-244 and several virulence genes relevant to the establishment of infections suggested the emergence of a supranational clonal outbreak, at least, in Europe [11].

Therefore, based on the recent concern about ST38 impact on the three axes of the One Health concept (humans, animals, and environment), mainly considering its role in spreading antibiotic resistance, and the gaps regarding genomic studies on ST38 lineage and its epidemiology in a global context, we performed a large-scale comparative genomic analysis on publicly available ST38 genomes in order to gain insights into the antibiotic resistance, virulence, adaptability, and colonization capabilities of this lineage.

2. Materials and Methods

2.1. EnteroBase ST38 Meta-Analysis

EnteroBase is integrated software that supports the identification of global population structures within several bacterial genera [12]. Here, a meta-analysis based on the geographical information of 3290 ST38 entries included in EnteroBase (<https://enterobase.warwick.ac.uk> accessed on 29 November 2022) was performed (Table S1). The ST38 epidemiological map was generated using R software v4.1.1 with different libraries (maptools and ggplot2).

2.2. Comparative Genomic Analyses and In Silico Characterization of the Resistome and Virulome

All complete and draft available *E. coli* genomes ($n = 24,102$) were retrieved from GenBank and EMBL databases. The seven MLST alleles that define the ST38 (*adk-4*, *fumC-26*, *gyrB-2*, *icd-25*, *mdh-5*, *purA-5*, and *recA-19*) were separately used as queries in BLASTn analyses to recover all currently available ST38 genomes. The completeness of the draft ST38 genomes was assessed by using the BUSCO v5.3.2 tool, presenting a minimum completeness >93% [13]. A total of 376 ST38 genomes (complete and draft) recovered on all continents from humans, wild and companion animals, livestock, vegetables, poultry, sewage, fresh water, and seawater throughout three decades (1983–2020) were used in comparative genomic analyses. The genome sequence of one *E. coli* strain (EC35) (GenBank assembly accession no. GCA_019492145) recovered in 2016 from a urinary tract infection in a clinical setting in the Amazonian city of Boa Vista, Roraima, Brazil, was obtained in this study and also included in the analyses. The genomic DNA extraction was obtained with NucleoSpin Microbial DNA kit (Macherey-Nagel), and the genome libraries were constructed using Nextera paired-end libraries. The sequencing was performed by Illumina HiSeq 2500, generating reads of 250 bp length. The raw reads were filtered and trimmed using the NGS QC Toolkit v.2.3.3 [14], with a Phred quality score ≥ 20 . The genomes were *de novo* assembled with a SPAdes assembler v3.14.1 [15] and improved using a Pilon v1.23 [16]. Gene prediction and annotation were performed with Prokka v1.14.5 [17].

ARG prediction was conducted using the Comprehensive Antibiotic Resistance Database (CARD) [18]. The Resistance Gene Identifier (RGI) tool was used for predicting the resistome based on homology and single nucleotide polymorphism (SNP) models with the default parameters. The in silico detection and plasmid typing based on replicon sequence analysis was performed using the PlasmidFinder web tool [19]. The virulome mining was conducted with ABRicate software package against the *E. coli* virulence database based on Virulence Factor Database (VFDB) (<https://github.com/tseemann/abricate> accessed on 30 July 2021) [20].

2.3. Phylogenomic Analysis

A global-scale phylogenomic reconstruction was performed with the 376 ST38 *E. coli* genomes. These genomes were annotated using Prokka v1.14.6 [17] and the resultant gff3 files were submitted to Roary v3.13.0 [21] to determine the core genome. The SNP sites of the concatenated core genes (104,645 bp) were extracted using snp-sites v2.5.1 [22] and submitted to IQTree v1.6.12 [23] to obtain a maximum likelihood SNP tree, which used the model of substitution GTR + F + ASC + R8 and 1000 ultrafast bootstrap replicates. The SNP tree was generated using iTOL v4 [24].

3. Results and Discussion

3.1. Global Epidemiology of ST38

In the context of a study conducted in a clinical setting in the Amazon region, we identified an ExPEC strain, which was recovered from sputum, belonging to ST38, and carrying the bla_{CTX-M-15} and bla_{NDM-1} resistance genes. Manges et al. [5], based on a systematic revision, had established the worldwide distribution of ExPEC lineages, in which ST38 was in the fourth position on the top 20 ExPEC STs reported in the world. In this revision, ST38 occurrence was lower in the Americas relative to other continents and absent in Africa; not being reported prior to 2000 in the world. Based on this scenario, our finding could represent the index case of this lineage emergence in Brazil. In order to raise the epidemiological scenario of this ExPEC ST, we searched for ST38 information in another database, the EnteroBase, which includes published and unpublished MLST data and geographical information (Table S1) (last updated in November 2022). This meta-analysis revealed that ST38 occurred in 72 countries from all continents, with higher prevalence in the USA, Australia, Germany, and the Netherlands, contrasting with the inferences previously based on reports [5]. Moreover, this ExPEC lineage occurred in at least nine African countries (Figure 1). Even though most of ST38 deposited in EnteroBase was from 2000 onwards, there are strains isolated since 1984 (Table S1). In this way, we corroborated the importance of ST38 as an ExPEC lineage worldwide.

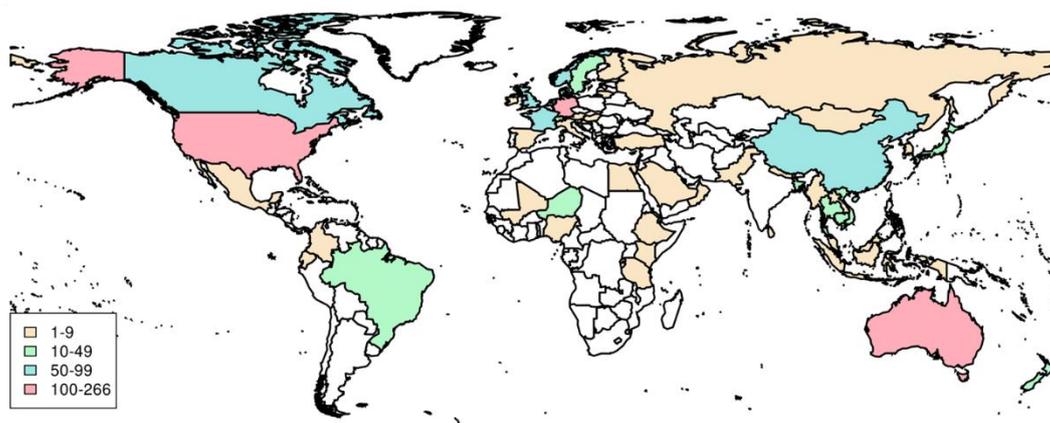


Figure 1. Map of epidemiological distribution of *E. coli* ST38 based on available metadata from EnteroBase.

3.2. ST38 Phylogenomics

To gain insights on the ST38 global epidemiology at genomic level, we considered 376 ST38 genomes, retrieved from GenBank and EMBL databases, recovered worldwide from humans, animals, and environmental sources. Interestingly, it was observed that ST38 occurs, at least, for more than three decades (1983–2020), being the oldest strains recovered from animals in the USA (Figure 2). The phylogenomic reconstruction revealed that ST38 genomes grouped in two clades including genomes from all sources (Figure 2). Clade 1 included 32 genomes isolated from 1983 to 2019, while Clade 2 presented several sub-clades containing the majority of genomes ($n = 344$) that were recovered from 2000 to

2020 (Figure 2). It was observed that animal and environmental genomes clustered together with human genomes; however, there are sub-clades represented only by genomes from human sources (Figure 2). Interestingly, identical genomes were observed recovered from humans, animals, and environmental sources in distinct spatio-temporal contexts, stressing the ability and plasticity of some lineages in disseminating among different backgrounds (Figure 2). For instance, there was a strain persisting for more than 30 years in distinct sources in the world: the genome of the Brazilian EC35 strain (GCA_019492145) obtained in 2016, which belonged to a sub-clade in Clade 1, was identical to genomes recovered from poultry in Australia (2008), and from animal and environmental sources in the USA (1986 and 1993, respectively). In the same way, there was a sub-clade in Clade 2 including 31 closely related genomes recovered worldwide from distinct sources between 2012 and 2019. The presence of closely related ST38 genomes globally distributed points to their epidemiological success in terms of persistence and niche adaptability. Recently, an ST38 genomic analysis revealed the presence of a clade comprising genomes from humans, wild and companion animals, and animal feed recovered in Asia, Europe, the Americas, and Australia between 2013 and 2019, stressing the spreading capacity and ubiquity of some ST38 clones [3]. Therefore, these findings altogether demonstrated that ST38 corresponds to a successful zoonotic or zooanthroponosis *E. coli* lineage.

3.3. ST38 Acquired Resistome

The analysis of the acquired resistome revealed the presence of a rich and diverse arsenal of antibiotic resistance genes (ARGs) with a quite heterogeneous distribution among ST38 genomes, independently of their source, genetic relatedness, and spatio-temporal context. Overall, genes involved with resistance to most of the clinically relevant antibiotics for treating *E. coli* infections, such as aminoglycosides, β -lactams, including fifth-generation cephalosporins and carbapenems, and fluoroquinolones were found within these genomes, except for those associated with tigecycline and fosfomycin resistance [25]. Although the most prevalent ARGs among ST38 were *aph*(3'')-Ib, *aph*(6'')-Id, *mphA*, *mphE*, *qacE* Δ 1, and *sul*, here we discuss ARGs with clinical impact, such as those coding for ESBLs (*bla*_{CTX-M} and *bla*_{CMY}), carbapenemases (*bla*_{KPC} and *bla*_{OXA}), metallo- β -lactamases (MBLs) (*bla*_{NDM}, *bla*_{IMP}, and *bla*_{VIM}), and pentapeptide-repeat proteins (*qnr* family).

It was found that *bla*_{CTX-M} alleles (*bla*_{CTX-M-1,-2,-3,-9,-14,-15,-16,-24,-27,-55}) were dispersed in the great majority of ST38 genomes (273/376 genomes), corroborating other studies demonstrating the occurrence of *bla*_{CTX-M} in the ST38 lineage (Figure S1) [26,27]. The *bla*_{CTX-M-14} and *bla*_{CTX-M-27} were the most prevalent among genomes and, in fact, *bla*_{CTX-M-14} is one of the most predominant genotypes among Enterobacteriaceae in the world [28]. A strong association was observed of *bla*_{CTX-M-27} with a particular sub-clade that included ST38 genomes recovered from Vietnam, the USA, France, the Netherlands, Germany, Qatar, Australia, and Brazil (the only genome in this sub-clade from an animal source), since 57/60 genomes from this sub-cluster harbored this ESBL allele. In the case of *bla*_{CTX-M-14}, it was the unique allele found in a sub-clade comprising genomes recovered between 2006 and 2019 from Asia, Europe, and Oceania from humans and wild animal (one genome from Mongolia). Similarly, *bla*_{CTX-M-14} was also found in a sub-clade characterized by genomes from the Americas, Asia, Europe, and Oceania recovered from all sources between 2012 and 2020. Sporadic genomes ($n = 4$) co-harbored two *bla*_{CTX-M} alleles: *bla*_{CTX-M-15} + *bla*_{CTX-M-27} in a genome from Vietnam; *bla*_{CTX-M-15} + *bla*_{CTX-M-9} from China; *bla*_{CTX-M-3} + *bla*_{CTX-M-27} and *bla*_{CTX-M-14} + *bla*_{CTX-M-15} from France. In general, clonal genomes harbored the same *bla*_{CTX-M} allele. However, eventually, distinct *bla*_{CTX-M} alleles were found among closely related genomes (GCA_016860125-*bla*_{CTX-M-14}; GCA_904863305-*bla*_{CTX-M-15}; GCA_002442325-*bla*_{CTX-M-27}), indicating events of microevolution or recombination. A diversity of *bla*_{CTX-M} alleles, mainly *bla*_{CTX-M-14} and *bla*_{CTX-M-15}, was identified among non-human genomes as previously observed elsewhere [28]. In comparison with *bla*_{CTX-M}, the ESBL-coding *bla*_{CMY-2} gene was less frequent among ST38, although it was present throughout the clades.

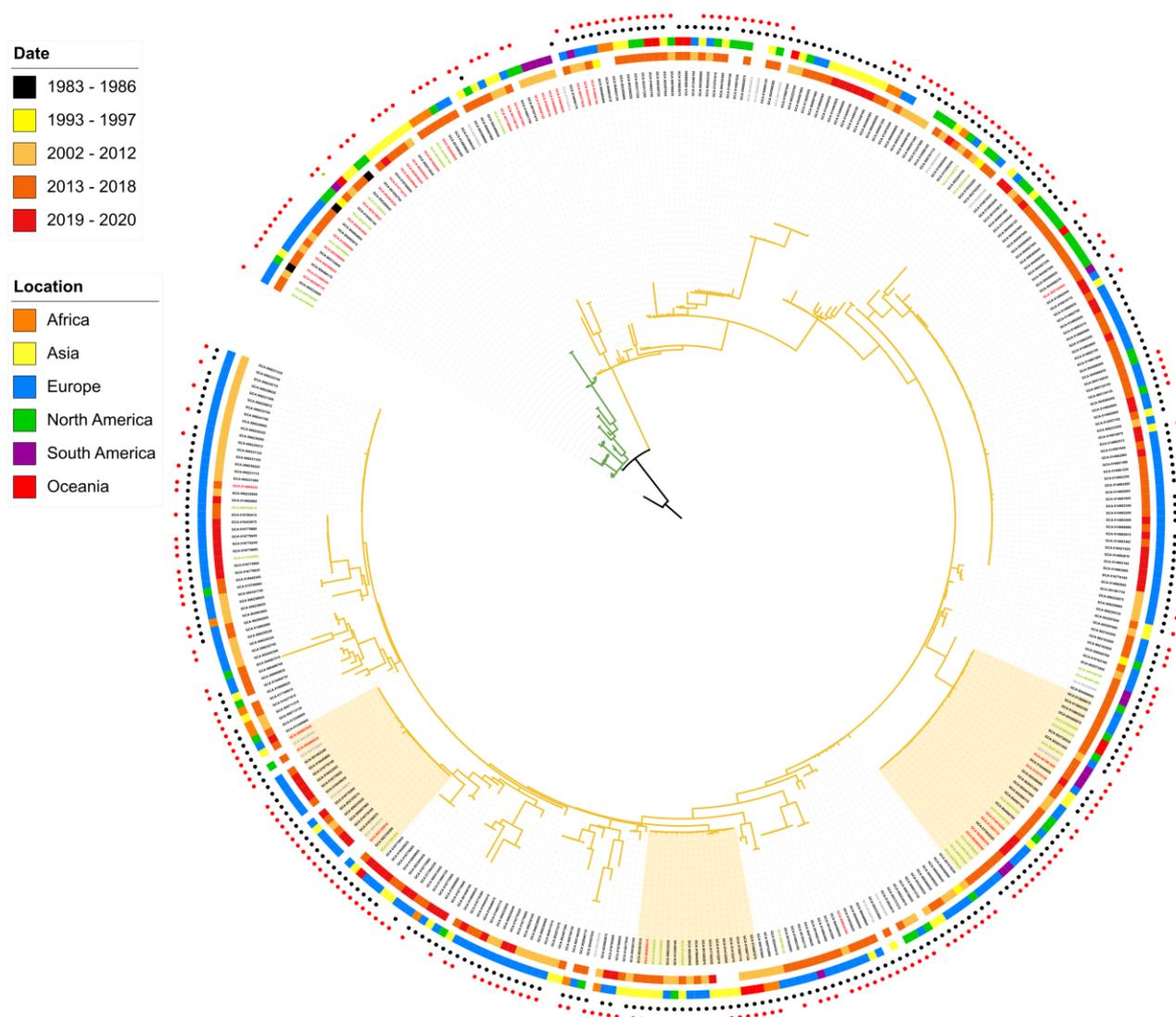


Figure 2. Core SNP tree based on *E. coli* ST38 genomes. Highlighted branches represent Clade 1 (green) and Clade 2 (yellow). The colored circles outside the tree, from innermost to outermost, represent metadata of isolation date and origin, respectively. Genomes carrying the high-pathogenicity island (HPI) and the second T3SS (ETT2) are marked with black and red circles, respectively. The EC35 genome from Brazil obtained in this study is featured by a green star. Highlighted labels represent genomes from human (black), animal (red), environmental (green), and unknown (grey) sources. The yellow hatched regions highlight the occurrence of identical genomes among different sources, demonstrating the zoonotic character of ST38. This figure with a better resolution was also provided in the Supplementary Material (Figure S3).

Carbapenems are drugs of choice to treat infections caused by ESBL-producing Enterobacteriaceae, such as CTX-M-producing ExPEC [29] and, therefore, the co-occurrence of ESBLs and MBLs, for example, bla_{NDM-1} , among these strains, is of great global concern due to its clinical outcomes. Here, in spite of the clinical origin of most of the ST38 genomes analyzed, we observed a low prevalence of carbapenem resistance genes (bla_{OXA-48} : 11%; $bla_{OXA-244}$: 14%; bla_{KPC-2} : 0.8%; bla_{NDM-1} : 0.5%; bla_{IMP-4} : 0.26%; bla_{VIM-4} : 0.26%) (Figure S1). The bla_{OXA-48} and $bla_{OXA-244}$ were found more frequently and spread among the carbapenemase genes analyzed, and *E. coli* ST38 has been associated with the spread of OXA-48 by clonal expansion [11,30,31]. In fact, we observed identical ST38 genomes carrying bla_{OXA-48} in both clinical strains from distinct countries (Netherlands, USA, and Germany) and also in animals (birds/Switzerland). On the other hand, our

large-scale resistome analysis revealed that bla_{OXA-48} distribution in ST38 was not only due to clonal expansion since this gene was found spread among unrelated genomes recovered from distinct countries in Asia (Qatar, Lebanon, and Turkey) distributed through several sub-clades (Figure S1). Concerning bla_{OXA-244}, a point-mutation derivative of bla_{OXA-48}, we verified that this allele was spread by clonal expansion of ST38 in Europe, as previously observed [11], and Asia, but also by horizontal transmission, since bla_{OXA-244} was found in clinical unrelated ST38 genomes from Qatar, Colombia, Switzerland, and Germany, and in an environmental genome from USA.

Among the 376 ST38 genomes MBL genes were quite rare: bla_{NDM-1} ($n = 2$; Brazil and Switzerland, 2016), bla_{IMP-4} ($n = 1$; China, 1983), and bla_{VIM-4} ($n = 1$; France, 2014) (Figure S1). So far, only bla_{NDM} gene had been reported in a few ST38 strains in Asia [32–35]. In fact, MBL genes are not highly prevalent in any *E. coli* ST when compared with other Enterobacteriaceae family members [36]. Interestingly, all of these MBL-positive genomes (except for the one from Switzerland) co-harbored a bla_{CTX-M} gene, all of them associated with human infection cases: bla_{CTX-M-14} + bla_{CTX-M-15} + bla_{VIM-4} in a genome recovered in France (GCA_009909835); bla_{CTX-M-14} + bla_{IMP-4}, in an isolate from China (GCA_002223705); and bla_{CTX-M-15} + bla_{NDM-1} in the Brazilian EC35 genome obtained in this study (GCA_019492145) (Figure S1). Although rare, such association has a relevant impact due to its clinical burden.

In contrast, the in vitro resistome analysis of the EC35 strain, recovered in the Amazon region and sequenced in the present study, had revealed an MDR phenotype, being resistant to most antibiotics tested, including the new drug combinations ceftolozane/tazobactam and ceftazidime/avibactam (Table 1). Its resistome was in agreement with this resistance phenotype since, besides bla_{CTX-M-15} and bla_{NDM-1}, it harbored genes conferring resistance to cephalosporins (bla_{CMY-2}), trimethoprim (dfrA14), sulfonamides (sul2), tetracycline (tet(A)), streptomycin (strBA) and quinolones (qnrS1 and a novel qnrB allele encoding a protein sharing 85% identity with QnrB1). Interestingly, the co-occurrence of these two qnr-type genes might synergistically contribute to the accumulation of activity, increasing quinolone resistance. Indeed, the observed MIC of ciprofloxacin for EC35 (0.5 mg/L) was slightly higher than that related to qnrB1 and qnrS1 alone (MICs 0.1–0.25 mg/L) [37]. This genome is closely related to two genomes from animals (Australia and USA) and one from the environment (water) USA. The resistome of these non-clinical genomes comprised only bla_{TEM-181}, dfrA5 and tet(A) (GCA_003318825/animal/Australia) and sul2 (GCA_002463095/animal/USA), while the environmental one (GCA_002231735) had no antibiotic resistance gene. Considering that these genomes are clonal, this finding strongly suggests that the huge resistome presented in the clinical EC35 genome is associated with an acquired mobilome. In fact, the in silico plasmid searches revealed the presence of different plasmid backbones from several incompatibility groups, such as IncHI2 and IncFIIb, and most of the ARGs were associated with these plasmid segments. However, due to the technical limitations imposed by the sequencing methodology applied, it was not possible to completely characterized these elements.

It is worth mentioning the occurrence of an expressive arsenal of ARGs among several non-human genomes involved with resistance emergence to a broad variety of antibiotics. These genomes were recovered from distinct spatio-temporal contexts and were distributed throughout Clade 1. It was the case of the genomes GCA_014712515 (aac(6'')-Ib9, ant(3'')-IIa, aph(3'')-Ib, aph(3')-Ia, aph(6)-Id, bla_{CMY-59}, bla_{OXA-10}, qnrVC4, bla_{TEM-181}, cmlA6, dfrA14, floR, qacL, sul2, tet(A)), recovered in China from poultry, 2016; GCA_011742905 (aph(3'')-Ib, aph(6'')-Id, aadA5, the carbapenemase bla_{OXA-48}, catI, bla_{TEM-181}, dfrA17, qacΔE1, sul1, sul2, tet(D)), recovered in Switzerland from the natural environment, 2019; GCA_002901285 (aacA6-Ib-cr6, bla_{CTX-M-15}, bla_{OXA-1}, aadA5, dfrA17, qacΔE1, sul1, tet(B)), recovered in Brazil from cabbage, 2016; and GCA_008642395 (aac(3'')-IId, aph(3'')-Ib, aph(6)-Id, bla_{CTX-M-14}, bla_{TEM-181}, aadA5, dfrA17, qacΔE1, sul1, sul2, tet(D)), recovered in Norway from mollusks, 2014.

Table 1. EC35 antimicrobial resistance profile.

Antibiotics	EC35 (MIC, mg/L)
Ampicillin	≥256
Ampicillin/sulbactam	≥256
Amoxicillin/clavulanate	≥256
Ticarcillin/clavulanic acid	≥256
Piperacillin/tazobactam	≥256
Cefoxitin	128
Ceftriaxone	128
Cefotaxime	64
Cefuroxime	128
Ceftazidime	≥256
Cefepime	32
Ceftolozane/tazobactam	≥256
Ceftazidime/avibactam	≥256
Aztreonam	16
Imipenem	32
Meropenem	16
Ertapenem	32
Doripenem	2 (I)
Tetracycline	64
Trimethoprim/sulfamethoxazole	>32
Ciprofloxacin	0.5 (I)
Streptomycin	96
Gentamicin	4
Tobramycin	4
Amikacin	8
Chloramphenicol	4
Fosfomycin	16
Tigecycline	0.125
Colistin	1.5

3.4. ST38 Virulome

The ST38 virulome revealed a robust arsenal of virulence-associated genes related to the successful establishment of both intestinal and extraintestinal infections. The ST38 virulome was characterized by genes involved with crucial roles in host–pathogen interactions and adaptability, competitiveness, and colonization capabilities (*aslA*, *hlyE*, *motAB*, *nadAB*, *malX*, *artJ*, *clpV*, *eaeh*, *hcp*, *hofBC*, *matF*, *ycfZ*, *ygdB*, *b2854/b2972*, and the operons *aec*, *ent*, *ppd*, *ecp*, *flg*, *flh*, *fli*, and *ycb*) [38,39]. A heterogeneous distribution of these genes was observed among ST38 regardless of source, phylogenetic relationship, and spatio-temporal context.

In addition to the aforementioned virulence factors, the virulome mining revealed interesting and particular features regarding the T3SS type 2 identified in most of ST38 genomes (Figure 2). The ETT2 secretion system corresponds to T3SS type 2 identified in most of *E. coli* that shares a significant relationship with the T3SS coded by the Salmonella pathogenicity island 1 (SPI-1). Although present in many *E. coli* strains, in most cases the ETT2 had undergone extensive mutational attrition, resulting in a functionless system [40,41]. The complete ETT2 system comprises the three operons *eprKJIH-epaSRQPO-eivJICAEGF*, involved with apparatus formation, and the *eip* locus, encoding putative effectors, as well as *eilA*, *eicA*, and *eaex*, encoding a transcriptional regulator, a chaperone, and an outer membrane invasion/intimin-like protein, respectively [41–43]. Here, in order to infer the putative functionality of ETT2 revealed in ST38, we performed a screening on the 376 ST38 genomes considering not only the presence of the genes associated with ETT2 but also their identity at the protein level (Figure S2). This analysis showed a remarkable prevalence of the ETT2 entire system, and the high protein identity (90–100%) shared with EAEC 042 ETT2 (reference of functional ETT2) indicating their putative functionality in most of ST38 genomes.

So far, only a small set of complete and canonical ETT2 has been reported [41–49], which includes the EAEC 042 and the ExPEC ST69 lineage. Concerning the distribution and functionality of ETT2 in the ST38 lineage, a recent study demonstrated its occurrence in ST38 with a high level of conservation relative to the canonical one found in EAEC 042 in only two ST38 strains [43].

In the virulome of ST38 we also identified the high-pathogenicity island (HPI) which was first described in pathogenic *Yersinia* species that encodes the siderophore yersiniabactin [50], and has been considered one of the most relevant virulence factors among ExPEC lineages [39,51–54]. Besides its pivotal role in iron uptake, the HPI orchestrates various virulence mechanisms and optimizes the overall fitness of ExPEC, such as motility, antibiotic tolerance, and exacerbating inflammatory processes [51–54].

Interestingly, we observed a strong association and high prevalence of HPI in genomes recovered from humans/clinical setting over space and time distributed throughout Clade 2. Conversely, this island was absent in Clade 1, which is composed mainly of animal and environmental ST38 genomes (Figure 2). In fact, previous studies have revealed that HPI is prevalent in strains associated with nosocomial infections [39,51–54]. Therefore, our results raise genomic evidence that corroborates the HPI association with successful *E. coli* ST38 establishment in human infections.

4. Conclusions

Here, we presented the large-scale genomic analyses of the high-risk ExPEC ST38, focusing on its resistome and virulome, providing novel insights into its global epidemiology in terms of the three axes of the One Health concept. It was demonstrated that ST38 is a host-broad pathogen since clones were found efficiently persisting in humans, animals, and natural environments, underlining its global dissemination. Due to the high prevalence of HPI in human/clinical genomes, we could conclude that HPI plays a major role in the establishment of ST38 as a successful pathogen in humans.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/microorganisms10122482/s1>, Figure S1: Heatmap of the ST38 resistome considering the carbapenemase and ESBL genes.; Figure S2: Heatmap of the ST38 virulome considering the ETT2 T3SS genes.; Figure S3: Core SNP tree based on *E. coli* ST38 genomes.; Table S1: Metadata of *E. coli* ST38 isolates/genomes from Enterobase.

Author Contributions: Conceptualization, E.L.F. and A.C.V.; resources, R.V.C.; methodology, E.L.F. and S.M.M.; writing—original draft preparation, E.L.F.; writing—review and editing, E.L.F., S.M.M. and A.C.V.; supervision, A.C.V.; funding acquisition, A.C.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Inova Fiocruz /VPPCB post-doctoral fellowship and Oswaldo Cruz Institute grant.

Data Availability Statement: Not applicable.

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

References

1. de Carvalho, M.P.N.; Fernandes, M.R.; Sellera, F.P.; Lopes, R.; Monte, D.F.; Hippólito, A.G.; Milanelo, L.; Raso, T.F.; Lincopan, N. International clones of extended-spectrum beta-lactamase (CTX-M)-producing *Escherichia coli* in peri-urban wild animals, Brazil. *Transbound. Emerg. Dis.* **2020**, *67*, 1804–1815. [CrossRef] [PubMed]
2. Fernandes, M.R.; Sellera, F.P.; Moura, Q.; Esposito, F.; Sabino, C.P.; Lincopan, N. Identification and genomic features of halotolerant extended-spectrum- β -lactamase (CTX-M)-producing *Escherichia coli* in urban-impacted coastal waters, Southeast Brazil. *Mar. Pollut. Bull.* **2020**, *150*, 110689. [CrossRef] [PubMed]
3. Lopes, R.; Fuentes-Castillo, D.; Fontana, H.; Rodrigues, L.; Dantas, K.; Cerdeira, L.; Henriques, I.; Lincopan, N. Endophytic Lifestyle of Global Clones of Extended-Spectrum β -Lactamase-Producing Priority Pathogens in Fresh vegetables: A Trojan Horse Strategy Favoring Human Colonization? *mSystems* **2021**, *6*, e01125-20. [CrossRef] [PubMed]

4. Salgado-Caxito, M.; Benavides, J.A.; Adell, A.D.; Paes, A.C.; Moreno-Switt, A.I. Global prevalence and molecular characterization of extended-spectrum β -lactamase producing-*Escherichia coli* in dogs and cats—A scoping review and meta-analysis. *One Health* **2021**, *12*, 100236. [[CrossRef](#)]
5. Manges, A.R.; Geum, H.M.; Guo, A.; Edens, T.J.; Fibke, C.D.; Pitout, J.D.D. Global Extraintestinal Pathogenic *Escherichia coli* (ExPEC) Lineages. *Clin. Microbiol. Rev.* **2019**, *32*, e00135-18. [[CrossRef](#)]
6. Kaper, J.B.; Nataro, J.P.; Mobley, H.L. Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* **2004**, *2*, 123–140. [[CrossRef](#)]
7. Dale, A.P.; Woodford, N. Extra-intestinal pathogenic *Escherichia coli* (ExPEC): Disease, carriage and clones. *J. Infect.* **2015**, *71*, 615–626. [[CrossRef](#)]
8. Chattaway, M.A.; Jenkins, C.; Ciesielczuk, H.; Day, M.; DoNascimento, V.; Day, M.; Rodríguez, I.; van Essen-Zandbergen, A.; Schink, A.-K.; Wu, G.; et al. Evidence of Evolving Extraintestinal Enteropathogenic *Escherichia coli* ST38 Clone. *Emerg. Infect. Dis.* **2014**, *20*, 1935–1937. [[CrossRef](#)]
9. Lindstedt, B.A.; Finton, M.D.; Porcellato, D.; Brandal, L.T. High frequency of hybrid *Escherichia coli* strains with combined Intestinal Pathogenic *Escherichia coli* (IPEC) and Extraintestinal Pathogenic *Escherichia coli* (ExPEC) virulence factors isolated from human faecal samples. *BMC Infect. Dis.* **2018**, *18*, 544. [[CrossRef](#)]
10. Nascimento, J.; Santos, F.; Valiatti, T.; Santos-Neto, J.; Santos, A.M.; Cayô, R.; Gales, A.; Gomes, T.A.T. Frequency and Diversity of Hybrid *Escherichia coli* Strains Isolated from Urinary Tract Infections. *Microorganisms* **2021**, *9*, 693. [[CrossRef](#)]
11. Falgenhauer, L.; Nordmann, P.; Imirzalioglu, C.; Yao, Y.; Falgenhauer, J.; Hauri, A.M.; Heinmüller, P.; Chakraborty, T. Cross-border emergence of clonal lineages of ST38 *Escherichia coli* producing the OXA-48-like carbapenemase OXA-244 in Germany and Switzerland. *Int. J. Antimicrob. Agents* **2020**, *56*, 106157. [[CrossRef](#)] [[PubMed](#)]
12. Zhou, Z.; Alikhan, N.-F.; Mohamed, K.; Fan, Y.; Achtman, M.; the Agama Study Group; Brown, D.; Chattaway, M.; Dallman, T.; Delahay, R.; et al. The Enterobase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia coli* core genomic diversity. *Genome Res.* **2020**, *30*, 138–152. [[CrossRef](#)] [[PubMed](#)]
13. Manni, M.; Berkeley, M.R.; Seppey, M.; Zdobnov, E.M. BUSCO: Assessing Genomic Data Quality and Beyond. *Curr. Protoc.* **2021**, *1*, e323. [[CrossRef](#)] [[PubMed](#)]
14. Patel, R.K.; Jain, M. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLoS ONE* **2012**, *7*, e30619. [[CrossRef](#)]
15. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Pribelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [[CrossRef](#)] [[PubMed](#)]
16. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **2014**, *9*, e112963. [[CrossRef](#)]
17. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [[CrossRef](#)]
18. Alcock, B.P.; Raphenya, A.R.; Lau, T.T.Y.; Tsang, K.K.; Bouchard, M.; Edalatmand, A.; Huynh, W.; Nguyen, A.-L.V.; Cheng, A.A.; Liu, S.; et al. CARD 2020: Antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **2020**, *48*, D517–D525. [[CrossRef](#)]
19. Carattoli, A.; Zankari, E.; García-Fernandez, A.; Larsen, M.; Lund, O.; Voldby Villa, L.; Møller Aarestrup, F.; Hasman, H. PlasmidFinder and pMLST: In silico detection and typing of plasmids. *Antimicrob. Agents Chemother.* **2014**, *58*, 3895–3903. [[CrossRef](#)]
20. Chen, L.; Yang, J.; Yu, J.; Yao, Z.; Sun, L.; Shen, Y.; Jin, Q. VFDB: A reference database for bacterial virulence factors. *Nucleic Acids Res.* **2005**, *33*, D325–D328. [[CrossRef](#)]
21. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.G.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**, *31*, 3691–3693. [[CrossRef](#)] [[PubMed](#)]
22. Page, A.J.; Taylor, B.; Delaney, A.J.; Soares, J.; Seemann, T.; Keane, J.A.; Harris, S.R. SNP-sites: Rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genom.* **2016**, *2*, e000056. [[CrossRef](#)] [[PubMed](#)]
23. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [[CrossRef](#)] [[PubMed](#)]
24. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259. [[CrossRef](#)]
25. Magiorakos, A.-P.; Srinivasan, A.; Carey, R.B.; Carmeli, Y.; Falagas, M.E.; Giske, C.G.; Harbarth, S.; Hindler, J.F.; Kahlmeter, G.; Olsson-Liljequist, B.; et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: An international expert proposal for interim standard definitions for acquired resistance. *Clin. Microbiol. Infect.* **2012**, *18*, 268–281. [[CrossRef](#)]
26. Peirano, G.; van der Bij, A.K.; Gregson, D.B.; Pitout, J.D. Molecular epidemiology over an 11-year period (2000 to 2010) of extended-spectrum beta-lactamase-producing *Escherichia coli* causing bacteremia in a centralized Canadian region. *J. Clin. Microbiol.* **2012**, *50*, 294–299. [[CrossRef](#)]
27. Pitout, J.D. Extraintestinal pathogenic *Escherichia coli*: A combination of virulence with antibiotic resistance. *Front. Microbiol.* **2012**, *3*, 9. [[CrossRef](#)]
28. Bevan, E.R.; Jones, A.M.; Hawkey, P.M. Global epidemiology of CTX-M β -lactamases: Temporal and geographical shifts in genotype. *J. Antimicrob. Chemother.* **2017**, *72*, 2145–2155. [[CrossRef](#)]

29. Gutiérrez-Gutiérrez, B.; Rodríguez-Baño, J. Current options for the treatment of infections due to extended spectrum beta-lactamase-producing *Enterobacteriaceae* in different groups of patients. *Clin. Microbiol. Infect.* **2019**, *25*, 932–942. [[CrossRef](#)]
30. Turton, J.F.; Doumith, M.; Hopkins, K.L.; Perry, C.; Meunier, D.; Woodford, N. Clonal expansion of *Escherichia coli* ST38 carrying a chromosomally integrated OXA-48 carbapenemase gene. *J. Med. Microbiol.* **2016**, *65*, 538–546. [[CrossRef](#)]
31. Pitout, J.D.D.; Peirano, G.; Kock, M.M.; Strydom, K.-A.; Matsumura, Y. The Global Ascendency of OXA-48-Type Carbapenemases. *Clin. Microbiol. Rev.* **2019**, *33*, e00102-19. [[CrossRef](#)] [[PubMed](#)]
32. Sekizuka, T.; Matsui, M.; Yamane, K.; Takeuchi, F.; Ohnishi, M.; Hishinuma, A.; Arakawa, Y.; Kuroda, M. Complete sequencing of the bla(NDM-1)-positive IncA/C plasmid from *Escherichia coli* ST38 isolate suggests a possible origin from plant pathogens. *PLoS ONE* **2011**, *6*, e25334. [[CrossRef](#)] [[PubMed](#)]
33. Yamamoto, T.; Takano, T.; Iwao, Y.; Hishinuma, A. Emergence of NDM-1-positive capsulated *Escherichia coli* with high resistance to serum killing in Japan. *J. Infect. Chemother.* **2011**, *17*, 435–439. [[CrossRef](#)] [[PubMed](#)]
34. Shrestha, B.; Tada, T.; Shimada, K.; Shrestha, S.; Ohara, H.; Pokhrel, B.M.; Sherchand, J.B.; Kirikae, T. Emergence of Various NDM-Type-Metallo-beta-Lactamase-Producing *Escherichia coli* Clinical Isolates in Nepal. *Antimicrob. Agents Chemother.* **2017**, *61*, e01425-17. [[CrossRef](#)] [[PubMed](#)]
35. Huang, Y.-S.; Tsai, W.-C.; Li, J.-J.; Chen, P.-Y.; Wang, J.-T.; Chen, Y.-T.; Chen, F.-J.; Lauderdale, T.-L.; Chang, S.-C. Increasing New Delhi metallo- β -lactamase-positive *Escherichia coli* among carbapenem non-susceptible *Enterobacteriaceae* in Taiwan during 2016 to 2018. *Sci. Rep.* **2021**, *11*, 2609. [[CrossRef](#)] [[PubMed](#)]
36. Behzadi, P.; García-Perdomo, H.A.; Karpiński, T.M.; Issakhanian, L. Metallo- β -lactamases: A review. *Mol. Biol. Rep.* **2020**, *47*, 6281–6294. [[CrossRef](#)] [[PubMed](#)]
37. Hu, F.-P.; Xu, X.-G.; Zhu, D.-M.; Wang, M.-G. Coexistence of *qnrB4* and *qnrS1* in a clinical strain of *Klebsiella pneumoniae*. *Acta Pharmacol. Sin.* **2008**, *29*, 320–324. [[CrossRef](#)]
38. Peirano, G.; Mulvey, G.L.; Armstrong, G.D.; Pitout, J.D.D. Virulence potential and adherence properties of *Escherichia coli* that produce CTX-M and NDM β -lactamases. *J. Med. Microbiol.* **2013**, *62*, 525–530. [[CrossRef](#)]
39. Shaik, S.; Ranjan, A.; Tiwari, S.K.; Hussain, A.; Nandanwar, N. Comparative Genomic Analysis of Globally Dominant ST131 Clone with Other Epidemiologically Successful Extraintestinal Pathogenic *Escherichia coli* (ExPEC) Lineages. *mBio* **2017**, *8*, e01596-17. [[CrossRef](#)]
40. Hayashi, W.; Ohsaki, Y.; Taniguchi, Y.; Koide, S.; Kawamura, K. High prevalence of blaCTX-M-14 among genetically diverse *Escherichia coli* recovered from retail raw chicken meat portions in Japan. *Int. J. Food Microbiol.* **2018**, *284*, 98–104. [[CrossRef](#)]
41. Ren, C.-P.; Chaudhuri, R.R.; Fivian, A.; Bailey, C.M.; Antonio, M.; Barnes, W.M.; Pallen, M.J. The ETT2 gene cluster, encoding a second type III secretion system from *Escherichia coli*, is present in the majority of strains but has undergone widespread mutational attrition. *J. Bacteriol.* **2004**, *186*, 3547–3560. [[CrossRef](#)] [[PubMed](#)]
42. Sheikh, J.; Dudley, E.G.; Sui, B.; Tamboura, B.; Suleman, A.; Nataro, J.P. EilA, a HilA-like regulator in enteroaggregative *Escherichia coli*. *Mol. Microbiol.* **2006**, *61*, 338–350. [[CrossRef](#)] [[PubMed](#)]
43. Fox, S.; Goswami, C.; Holden, M.; Connolly, J.P.R.; Mordue, J. A highly conserved complete accessory *Escherichia coli* type III secretion system 2 is widespread in bloodstream isolates of the ST69 lineage. *Sci. Rep.* **2020**, *10*, 4135. [[CrossRef](#)] [[PubMed](#)]
44. Hartleib, S.; Prager, R.; Hedenström, I.; Löfdahl, S.; Tschäpe, H. Prevalence of the new, SPI1-like, pathogenicity island ETT2 among *Escherichia coli*. *Int. J. Med. Microbiol.* **2003**, *292*, 487–493. [[CrossRef](#)] [[PubMed](#)]
45. Prager, R.; Bauerfeind, R.; Tietze, E.; Behrend, J.; Fruth, A.; Tschäpe, H. Prevalence and deletion types of the pathogenicity island ETT2 among *Escherichia coli* strains from edema disease and colibacillosis in pigs. *Vet. Microbiol.* **2004**, *99*, 287–294. [[CrossRef](#)] [[PubMed](#)]
46. Ideses, D.; Gophna, U.; Paitan, Y.; Chaudhuri, R.R.; Pallen, M.J.; Ron, E.Z. A degenerate type III secretion system from septicemic *Escherichia coli* contributes to pathogenesis. *J. Bacteriol.* **2005**, *187*, 8164–8171. [[CrossRef](#)]
47. Cheng, D.; Zhu, S.; Su, Z.; Zuo, W.; Lu, H. Prevalence and isoforms of the pathogenicity island ETT2 among *Escherichia coli* isolates from colibacillosis in pigs and mastitis in cows. *Curr. Microbiol.* **2012**, *64*, 43–49. [[CrossRef](#)]
48. Wang, S.; Liu, X.; Xu, X.; Zhao, Y.; Yang, D.; Han, X.; Tian, M.; Ding, C.; Peng, D.; Yu, S. *Escherichia coli* type III secretion system 2 (ETT2) is widely distributed in avian pathogenic *Escherichia coli* isolates from Eastern China. *Epidemiol. Infect.* **2016**, *144*, 2824–2830. [[CrossRef](#)]
49. Shulman, A.; Yair, Y.; Biran, D.; Sura, T.; Otto, A.; Gophna, U.; Becher, D.; Hecker, M.; Ron, E.Z. The *Escherichia coli* Type III Secretion System 2 Has a Global Effect on Cell Surface. *mBio* **2018**, *9*, e01070-18. [[CrossRef](#)]
50. Carniel, E.; Guilvout, I.; Prentice, M. Characterization of a large chromosomal “high-pathogenicity island” in biotype 1B *Yersinia enterocolitica*. *J. Bacteriol.* **1996**, *178*, 6743–6751. [[CrossRef](#)]
51. Magistro, G.; Magistro, C.; Stief, C.G.; Schubert, S. The high-pathogenicity island (HPI) promotes flagellum-mediated motility in extraintestinal pathogenic *Escherichia coli*. *PLoS ONE* **2017**, *12*, e0183950. [[CrossRef](#)] [[PubMed](#)]
52. Liu, C.; Shan, C.; Dong, Q.; Fu, G.; Zhao, R.; Yan, Y.; Gao, H. Pathogenic *E. coli* HPI upregulate the expression of inflammatory factors in porcine small intestinal epithelial cells by ubiquitin proteasome pathway. *Res. Vet. Sci* **2018**, *120*, 41–46. [[CrossRef](#)] [[PubMed](#)]

53. Galardini, M.; Clermont, O.; Baron, A.; Busby, B.; Dion, S.; Schubert, S.; Beltrao, P.; Denamur, E. Major role of iron uptake systems in the intrinsic extra-intestinal virulence of the genus *Escherichia* revealed by a genome-wide association study. *PLoS Genet.* **2020**, *16*, e1009065. [[CrossRef](#)] [[PubMed](#)]
54. Shan, C.; Miao, S.; Liu, C.; Zhang, B.; Zhao, W.; Wang, H.; Yang, W.; Cha, J.; Zhao, R.; Xiao, P.; et al. Induction of macrophage pyroptosis-related factors by pathogenic *E. coli* high pathogenicity island (HPI) in Yunnan Saba pigs. *BMC Vet. Res.* **2021**, *17*, 114. [[CrossRef](#)]