

Global genomic epidemiology, resistome, virulome and plasmidome of the Extraintestinal Pathogenic *Escherichia coli* (ExPEC) ST38 lineage

Érica L. Fonseca^{a,*}, Sérgio M. Morgado^a, Ana Carolina P. Vicente^a

^aLaboratório de Genética Molecular de Microrganismos, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil

ELF: ericafon@ioc.fiocruz.br

SMM: sergio.morgado@ioc.fiocruz.br

ACPV: anapaulo@ioc.fiocruz.br

Corresponding Author. Érica L. Fonseca. Laboratório de Genética Molecular de Microrganismos, Instituto Oswaldo Cruz, FIOCRUZ. Avenida Brasil 4365, Manguinhos, Rio de Janeiro. Brazil. E-mail address: ericafon@ioc.fiocruz.br. Telephone number: +55-21-3865-8176.

ABSTRACT

Most of the extraintestinal human infections worldwide are determined by some specific Extraintestinal Pathogenic *Escherichia coli* (ExPEC) lineages, which also present a zoonotic character. One of these lineages belongs to ST38, which is considered a uropathogenic/enteroaggregative *E. coli* hybrid and a high-risk globally disseminated ExPEC. To get insights on the aspects of ST38 epidemiology and evolution as a multidrug resistant and pathogenic lineage, this study performed a global comparative genomic analysis on ST38 genomes. A total of 376 genomes recovered from environments, humans, livestock, wild and domestic animals in all continents throughout three decades were analyzed considering the ST38 phylogenomics, resistome, plasmidome, and virulome. In general, our analyses revealed that, independently of host sources and geographic origin, these genomes were distributed in two clusters comprising clonal and non-clonal genomes. Moreover, the ST38 accessory genome was not strictly associated with clusters and sub-clusters. Interestingly, the High Pathogenicity Island (HPI) prevailed in the major cluster (cluster 2), which comprised most of the genomes from human origin recovered worldwide (2000 to 2020). In addition, the ExPEC ST38 harbors a huge and diverse plasmidome that might contribute to the evolution of this lineage, reflecting in ST38 ability to live in a diversity of niches as a commensal or pathogenic organism.

Keywords: Phylogenomics; zoonosis; high-risk; genomic epidemiology; High Pathogenicity Island; accessory genome

1. Introduction

Escherichia coli is a commensal bacterium and a versatile pathogen capable of causing intestinal and extraintestinal infections (Vila et al., 2016). Extraintestinal pathogenic *E. coli* (ExPEC) determines most of the human extraintestinal infections (Manges et al., 2019), and includes uropathogenic *E. coli* (UPEC), neonatal meningitis *E. coli* (NMEC), sepsis-associated *E. coli* (SEPEC), and avian pathogenic *E. coli* (APEC) (Kaper et al., 2004). ExPEC comprises many lineages that affect not only humans but also animal health, besides occurring in the environment. However, a subset of lineages/sequence types (STs) is responsible for the majority of extraintestinal pathologies in humans and animals. The recent emergence of multidrug-resistant ExPEC lineages (ST131, ST38, ST405, and ST648) raised concerns on how *E. coli* evolves and diversifies in terms of virulence and antibiotic resistance (Dale and Woodford, 2015).

The ST38 is considered a uropathogenic/enteroaggregative *E. coli* hybrid (Chattaway et al., 2004), which also presents a zoonotic character since it is often identified in natural environments and animals (Hayashi et al., 2018; de Carvalho et al., 2020; Fernandes et al., 2020; Lopes et al., 2021; Salgado-Caxito et al., 2021).

The Extend Spectrum β -Lactamase (ESBL)-coding *bla*_{CTX-M} gene is prevalent among ExPEC lineages, such as ST131, which has been associated with the global dissemination of this gene (Bevan et al., 2017; Shaik et al., 2017; Mostafa et al., 2020). Although *bla*_{CTX-M} genes have been found in ST38 (Peirano et al., 2012; Pitout, 2012; Shaik et al., 2017), this lineage is far less characterized than ST131 (Cameron et al., 2021). Therefore, to get insights on ST38 and *bla*_{CTX-M} dissemination, as well as, other

features involved with its high-risk character in clinics, a global genomic analysis of ST38 should be performed taking into account the three axes of the One Health Concept (humans, animals, and natural environments), and the virulence, adaptability, and persistence traits.

Thus, this study performed a large-scale comparative genomic analysis on ST38 genomes from environments, humans, livestock, wild and domestic animals recovered from all continents throughout three decades, and determined the global ST38 phylogenomics, resistome, plasmidome, and virulome. In general, our analyses corroborated the zoonotic character of the ExPEC ST38 since it was identified clonal genomes from both animal and human sources recovered worldwide. Interestingly, it was verified that a virulence trait present in most of the ST38 genomes, the High Pathogenicity Island (HPI) from *Yersinia pestis*, prevailed in ST38 from human origin. In addition, the ExPEC ST38 was characterized by a huge and diverse plasmidome that might contribute to the evolution of this lineage, reflecting in ST38 ability to in a diversity of niches as a commensal or pathogenic organism.

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 (Zhou et al., 2020). Here, a meta-analysis based on the geographical information of ST38 entries included in EnteroBase (<https://enterobase.warwick.ac.uk>) was performed (last updated in September 2021). The ST38 epidemiological map was generated using the R software v4.1.1 with different libraries (maptools and ggplot2).

2.2. Comparative genomic analyses and *in silico* characterization of the resistome, plasmidome and virulome

All complete and draft *E. coli* genomes (n=24,102) (last updated in July 2021) 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. A total of 376 ST38 genomes recovered in all continents from humans, wild and companion animals, livestock, vegetables, poultry, sewage, fresh, and seawater throughout three decades (1983 – 2020) were used in comparative genomic analyses.

ARG prediction was conducted using the Comprehensive Antibiotic Resistance Database (CARD) (Alcock et al., 2020). The Resistance Gene Identifier (RGI) tool was used for predicting the resistome based on homology and 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 (Carattoli et al., 2014). The virulome mining was conducted with ABRicate software package against the *E. coli* virulence database based on VFDB (updated in July 2021) (<https://github.com/tseemann/abricate>) (Chen et al., 2005). The heatmaps of ST38 resistome, virulome, and plasmidome were drawn using iTOL v4 (Letunic and Bork, 2019).

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 (Seemann, 2014) and the resultant gff3 files were submitted to Roary v3.13.0 (Page et al., 2015) to

determine the core genome. The SNP sites of the concatenated core genes (104,645 bp) were extracted using snp-sites v2.5.1 (Page et al., 2016) and submitted to IQTree v1.6.12 (Minh et al., 2020) 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 (Letunic and Bork, 2019).

3. Results and Discussion

3.1. Global ExPEC ST38 distribution

A previous study (Manges et al., 2019) established the worldwide distribution of ExPEC lineages based on a systematic revision and meta-analysis considering reports that had performed multilocus sequence typing (MLST) or whole-genome sequencing to genotype *E. coli*. The scenario raised showed that ST38 was in the fourth position on the top 20 ExPEC STs reported in the world. In addition, ST38 occurrence appeared to be lower in North America relative to other continents and absent in Africa (Manges et al., 2019). Here, we analyzed the global occurrence of ExPEC ST38 using the EnteroBase records based on ST38 MLST entries with geographical information (last updated in September, 2021), providing pieces of evidence that enlarge the ST38 prevalence and its worldwide burden.

This meta-analysis revealed that ST38 occurred in countries from all continents, and its prevalence was higher in the USA, Australia, Germany, and the Netherlands, contrasting with the inferences previously established based on reports (Manges et al., 2019). Moreover, this ExPEC lineage occurred in at least nine African countries (Fig. 1).

3.2. ST38 Phylogenomics

This study considered 376 ST38 genomes recovered worldwide from humans, animals, and environmental sources persisting for more than three decades (1983-2020). The phylogenomic reconstruction revealed that ST38 genomes grouped in two clusters including genomes from all sources (Fig. 2). Cluster 1 included 32 genomes isolated from 1983 to 2019, while cluster 2 presented several sub-clusters containing the majority of genomes (n=344) that were recovered from 2000 to 2020 (Fig. 2). It was not observed any sub-cluster comprising only genomes from animal and/or environment but there were sub-clusters represented only by genomes from human sources (Fig. 2). Interestingly, despite the overall non-clonal aspect of the ST38, it was observed clonal genomes recovered from humans, animals, and environmental sources in distinct spatio-temporal contexts (Fig. 2). For instance, a Brazilian genome (GCA_019492145), recovered from a urinary tract infection in the Amazon region in 2016 and belonging to a sub-cluster in cluster 1, presented a clonal relationship with genomes from poultry in Australia (2008), animal, and environmental sources recovered in the USA almost 30 years before (1986 and 1993, respectively). In the same way, there was a sub-cluster in cluster 2 comprehending 31 closely related genomes recovered worldwide from distinct sources between 2012 and 2019. The presence of globally distributed ST38 clones points to their epidemiological success in terms of persistence and niche exploration. So far, there were few phylogenomic studies concerning ExPEC lineages. A previous study (Shaik et al., 2017) performed a comparative genomic analysis with the high-risk ExPEC lineages ST131, ST38, ST405, and ST648, in which twelve ST38 genomes recovered from humans were considered. Overall, their study revealed clonality as well as distinctness among strains at the genomic level. Recently, an ST38 genomic analysis revealed the presence of a cluster comprising genomes from humans, wild and companion animals, and animal feed recovered in Asia, Europe, Americas, and

Australia between 2013 and 2019, stressing the spreading capacity and ubiquity of some ST38 clones (Lopes et al., 2021). Therefore, these findings altogether demonstrated that ST38 corresponds to a successful zoonotic or zooanthroponosis *E. coli* lineage, probably due to the versatile genome that allows it to persist in animal reservoir hosts and in clinical and natural environments.

3.3. ST38 Acquired Resistome

The analysis of the acquired resistome mining 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 (Supplementary Fig. S1). Overall, genes involved with resistance to most of the clinically relevant antibiotics for treating *E. coli* infections were found within these genomes, except by those associated with tigecycline and fosfomycin resistance (Magiorakos et al., 2012). A previous study showed a similar acquired resistome profile among other ExPEC lineages such as ST131, ST648, and ST405 (Shaik et al., 2017).

Although the most prevalent ARGs among ST38 were *aph(3'')-Ib*, *aph(6'')-Id*, *mphA*, *mphE*, *qacEΔ1*, and *sul* (Supplementary Fig. S1), here we focused on ARGs with a remarkable clinical impact such as those coding for ESBLs (*bla*_{CTX-M} and *bla*_{CMY-2}), carbapenemases (*bla*_{KPC-2}, *bla*_{OXA-48}, *bla*_{OXA-244}) and metallo-β-lactamases (MBLs) (*bla*_{NDM-1}, *bla*_{IMP-4} and *bla*_{VIM-4}).

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, corroborating other studies demonstrating the occurrence of *bla*_{CTX-M} in ST38 lineage (Peirano et al., 2012; Pitout, 2012). The *bla*_{CTX-M-14} and *bla*_{CTX-M-27} were the most prevalent among ST38 and were

dispersed in several sub-clusters (Supplementary Fig. S1). In fact, *bla*_{CTX-M-14} is one of the most predominant genotype among *Enterobacteriaceae* in the world (Bevan et al., 2017). However, we observed a strong association of *bla*_{CTX-M-27} with a particular sub-cluster that included ST38 genomes recovered from Vietnam, the USA, France, the Netherlands, Germany, Qatar, Australia, and Brazil (the only genome in this sub-cluster from an animal source). In the case of *bla*_{CTX-M-14}, it was the unique allele found in a sub-cluster 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-cluster 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 (Supplementary Fig. S1). 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}) (Supplementary Fig. S1), corroborating the strong association of *bla*_{CTX-M} with mobile genetic platforms. Moreover, 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 (Bevan et al., 2017). In comparison with *bla*_{CTX-M}, the ESBL-coding *bla*_{CMY-2} gene was less frequently among ST38, although it was present throughout the clusters (Supplementary Fig. S1).

Carbapenems are drugs of choice to treat infections caused ESBL producing *Enterobacteriaceae* (Gutiérrez- Gutiérrez and Rodríguez-Baño, 2019). Here, genes associated with carbapenem resistance were found among ST38 (*bla*_{OXA-48}, *bla*_{OXA-244},

*bla*_{KPC-2}, *bla*_{NDM-1}, *bla*_{IMP-4}, *bla*_{VIM-4}), even though in low prevalence, except by the two *bla*_{OXA} genes.

Concerning the carbapenemase-coding genes, the *bla*_{OXA-48} and *bla*_{OXA-244} were found to be the most prevalent and widespread among ST38 (Supplementary Fig. S1). *E. coli* ST38 has been associated with the spread of OXA-48 by clonal expansion (Turton et al., 2016; Pitout et al., 2019). However, 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 non-clonal genomes from several sub-clusters (Supplementary Fig. S1). Conversely, the *bla*_{KPC-2} gene is rare in ExPEC lineages. Indeed, this gene was found only in two ST38 genomes from Germany (2014) and in a non-clonal genome from Vietnam (2011), all of them from human sources.

In the same way, MBL genes seem to be rare in ST38 since, so far, only *bla*_{NDM} gene had been reported in a few ST38 strains in Asia (Sekizuka et al., 2011; Yamamoto et al., 2011; Shrestha et al., 2017; Huang, et al., 2021). In fact, our resistome mining analysis revealed a sparse occurrence of *bla*_{NDM-1} (n=2; Brazil and Switzerland, 2016), *bla*_{IMP-4} (n=1; China, 1983), and *bla*_{VIM-4} (n=1; France, 2014) among the 376 ST38 genomes. Interestingly, all of these MBL-positive genomes (except for the one from Switzerland) co-harbored a *bla*_{CTX-M} gene: *bla*_{CTX-M-15} + *bla*_{NDM-1} in a Brazilian genome (GCA_019492145); *bla*_{CTX-M-14} + *bla*_{CTX-M-15} + *bla*_{VIM-4} in a genome recovered in France (GCA 009909835); and *bla*_{CTX-M-14} + *bla*_{IMP-4}, in an isolate from China (GCA 002223705), all of them associated with human infection cases (Supplementary Fig. S1). Although rare, such association has a relevant impact due to its clinical burden.

3.4. ST38 Virulome

The ST38 virulome mining 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 adherence/aggregation, motility, colonization/invasion, biofilm formation, necroptosis blockage, capsule formation, serum resistance, acid tolerance, iron uptake, intracellular survival, evasion from the host's immune response, activation of innate immune response and long-term persistence in the urinary tract. However, a heterogeneous distribution of these genes was observed among ST38 regardless of source, phylogenetic relationship, and spatio-temporal context (Supplementary Fig. S2).

Concerning, the virulence genes present in ExPEC pathotypes (UPEC, NMEC, SEPEC, APEC) (Pitout, 2012; Dale and Woodford 2015; Sarowska et al., 2019), ST38 genomes harbored the genes *matF*, *ibeBC*, *kpsM*, *iss*, *fyuiA*, *ecpA* and the operons *fim*, *csg*, *chu*, and *che*, while other ExPEC virulence genes were not ubiquitous (Supplementary Fig. S2). On the other hand, all ST38 genomes carried genes that play crucial roles in host-pathogen interactions and are associated with 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*) (Peirano et al., 2013; Shaik et al., 2017). These findings reinforce that ST38 is well adapted to survive and proliferate in diverse hosts and environments.

Surprisingly, it was observed a particular and remarkable virulome feature concerning the High Pathogenicity Island (HPI) distribution within ST38 genomes. This island, firstly described in pathogenic *Yersinia* species, encodes the siderophore yersiniabactin (Carniel et al., 1996), and has been considered one of the most relevant

virulence factors among ExPEC lineages (Magistro et al., 2017; Shaik et al., 2017; Liu et al., 2018; Galardini et al., 2020, Shan et al., 2021). Besides its pivotal role on iron uptake, the HPI orchestrates various virulence mechanisms and optimizes the overall fitness of ExPEC, such as motility, antibiotic tolerance, and exacerbate inflammatory processes (Magistro et al., 2017; Liu et al., 2018; Galardini et al., 2020, Shan et al., 2021). Considering HPI distribution among ST38, it was observed that this island was absent in cluster 1 and in a specific sub-cluster of cluster 2. Of note is that most of the ST38 genomes in these HPI-missing clusters belonged to animal and environmental sources, contrasting with the strong association and high prevalence of HPI in genomes recovered from humans distributed throughout clade 2 (Fig. 2). Our finding corroborates previous studies on ExPEC lineages, in which HPI was associated with strains causing human infections (Magistro et al., 2017; Shaik et al., 2017; Liu et al., 2018; Galardini et al., 2020, Shan et al., 2021).

3.5. *ST38 plasmidome*

The plasmidome analysis was based on the identification of *rep* genes and their incompatibility group classification (Supplementary Fig. S3). In this way, it was revealed the presence of several plasmids from distinct incompatibility groups among ST38 genomes. The IncFIIb group was the most prevalent and evenly distributed in the sub-clusters, and, in fact, it was demonstrated that this plasmid type is spread in several ExPEC lineage such as ST131 (Shaik et al., 2017). In general, there was an abundance of genomes harboring several plasmids (at most 8 per genome). However, two ST38 genomes, one recovered from an animal source (GCA_014712515) and another recovered from human (GCA_015701015), harbored 10 plasmids and nine plasmids, respectively. In addition, some sub-clusters comprised genomes from animal and human

sources sharing the same plasmid profile, which included three to five plasmids (Supplementary Fig. S3).

Curiously, the pKPC-CAV1321 plasmid was observed in some ST38 genomes distributed in non-related sub-clusters, recovered from distinct sources and spatio-temporal contexts: one genome from human/Brazil/2016 (GCA_019492145), one from animal/Paraguay/2012 (GCA_016858745), one from wastewater/Czech Republic/2016 (GCA_009790745), and 10 genomes from human/Russia/2012 (Supplementary Fig. S3). The pKPC-CAV1321 had been previously found in *Citrobacter freundii* and *Klebsiella pneumoniae* (Sheppard et al., 2016), and it was recently identified in *Salmonella Choleraesuis* spread in different Brazilian regions (Meneguzzi et al., 2021). However, so far, pKPC-CAV1321 has never been reported in any ExPEC lineage. Interestingly, all pKPC-CAV1321-positive ST38 genomes co-harbored the IncHI2 plasmid. The IncHI2 replicon has been associated with human and animal disease, representing a threat in the context of One Health (Hooton et al., 2021) and, although abundant among *Enterobacteriaceae* (Wong et al., 2016; Matamoros et al., 2017; Zhao et al., 2018), the IncHI2 had never been reported in ST38 and is rare among other ExPEC strains (Shaik et al., 2017).

4. Conclusions

Here, we presented the large-scale genomic analyses of the high-risk ExPEC ST38, providing novel insights into its evolutionary mechanisms. It was demonstrated that ST38 is a host-broad pathogen since clones were found efficiently persisting in humans, animals, and natural environments, underlining the global dimension and the zoonotic character of this ExPEC lineage. Probably, the extensive array of resistance

genes and the well-equipped arsenal of ExPEC- and IPEC-associated virulence determinants contributed to the establishment of ST38 as a versatile and successful *E.coli* lineage.

Funding sources

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Inova Fiocruz /VPPCB post-doctoral fellowship.

Declaration of Competing Interest

The authors declare no competing interests.

References

Alcock, B.P., Raphenya, A.R., Lau, T.T.Y., Tsang, K.K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.V., Cheng, A.A., Liu, S., Min, S.Y., Miroshnichenko, A., Tran, H.K., Werfalli, R.E., Nasir, J.A., Oloni, M., Speicher, D.J., Florescu, A., Singh, B., Faltyn, M., Hernandez-Koutoucheva, A., Sharma, A.N., Bordeleau, E., Pawlowski, A.C., Zubyk, H.L., Dooley, D., Griffiths, E., Maguire, F., Winsor, G.L., Beiko, R.G., Brinkman, F.S.L., Hsiao, W.W.L., Domselaar, G.V., McArthur, A.G., 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 48 (D1), D517–D525. <https://doi.org/10.1093/nar/gkz935>.

- Bevan, E.R., Jones, A.M., Hawkey, P.M., 2017. Global epidemiology of CTX-M β -lactamases: temporal and geographical shifts in genotype. *J. Antimicrob. Chemother.* 72 (8), 2145-2155. doi: 10.1093/jac/dkx146.
- Cameron, A., Mangat, R., Mostafa, H.H., Taffner, S., Wang, J., Dumyati, G., Stanton, R.A., Daniels, J.B., Campbell, D., Lutgring, J.D., Pecora, N.D., 2021. Detection of CTX-M-27 β -Lactamase Genes on Two Distinct Plasmid Types in ST38 *Escherichia coli* from Three U.S. States. *Antimicrob. Agents Chemother.* 65, e00825-21. <https://doi.org/10.1128/AAC.00825-21>.
- Carattoli, A., Zankari, E., Garcia-Fernandez, A., Voldby Larsen, M., Lund, O., Villa, L., Aarestrup, F.M., Hasman, H., 2014. PlasmidFinder and pMLST: in silico detection and typing of plasmids. *Antimicrob. Agents Chemother.* 58 (7), 3895-3903. doi: 10.1128/AAC.02412-14.
- Carniel, E., Guilvout, I., Prentice, M., 1996. Characterization of a large chromosomal "high-pathogenicity island" in biotype 1B *Yersinia enterocolitica*. *J. Bacteriol.* 178 (23), 6743-6751. doi: 10.1128/jb.178.23.6743-6751.1996.
- Chattaway, M.A., Jenkins, C., Ciesielczuk, H., Day, M., Do Nascimento, V., Day, M., Rodríguez, I., van Essen-Zandbergen, A., Schink, A.-K., Wu, G., Threlfall, J., Woodward, M.J., Coldham, N., Kadlec, K., Schwarz, S., Dierikx, C., Guerra, B., Helmuth, R., Mevius, D., Woodford, N., Wain, J., 2014. Evidence of Evolving Extraintestinal Enteroaggregative *Escherichia coli* ST38 Clone. *Emerg. Infect. Dis.* 20 (11), 1935-1937. doi: 10.3201/eid2011.131845.

- Chen, L., Yang, J., Yu, J., Yao, Z., Sun, L., Shen, Y., Jin, Q., 2005. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* 33 (Database issue), D325-D328. doi:10.1093/nar/gki008
- Dale, A.P., Woodford, N., 2015. Extra-intestinal pathogenic *Escherichia coli* (ExPEC): Disease, carriage and clones. *J. Infect.* 71 (6), 615-626. doi: 10.1016/j.jinf.2015.09.009.
- 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., 2020. International clones of extended-spectrum beta-lactamase (CTX-M)-producing *Escherichia coli* in peri-urban wild animals, Brazil. *Transbound. Emerg. Dis.* 67, 1804-1815. doi: 10.1111/tbed.13558.
- Fernandes, M.R., Sellera, F.P., Moura, Q., Esposito, F., Sabino, C.P., Lincopan, A., 2020. 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.* 150, 110689. doi: 10.1016/j.marpolbul.2019.110689.
- Galardini, M., Clermont, O., Baron, A., Busby, B., Dion, S., Schubert, S., Beltrao, P., Denamur, E., 2020. 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. 16 (10), e1009065. doi: 10.1371/journal.pgen.1009065.
- Gutiérrez-Gutiérrez, B., Rodríguez-Baño, J., 2019. Current options for the treatment of infections due to extended spectrum beta-lactamase-producing *Enterobacteriaceae* in different groups of patients. Clin. Microbiol. and Infect. 25 (8), 932-942. <https://doi.org/10.1016/j.cmi.2019.03.030>.
- Hayashi, W., Ohsaki, Y., Taniguchi, Y., Koide, S., Kawamura, K., Suzuki, M., Kimura, K., Wachino, J.-I., Nagano, Y., Arakawa, Y., Nagano, N., 2018. High prevalence of blaCTX-M-14 among genetically diverse *Escherichia coli* recovered from retail raw chicken meat portions in Japan. Int. J. Food Microbiol. 284, 98-104. doi: 10.1016/j.ijfoodmicro.2018.08.003.
- Hooton, S.P.T., Pritchard, A.C.W., Asiani, K., Gray-Hammerton, C.J., Stekel, D.J., Crossman, L.C., Millard, A.D., Hobman, J.L., 2021. Laboratory Stock Variants of the Archetype Silver Resistance Plasmid pMG101 Demonstrate Plasmid Fusion, Loss of Transmissibility, and Transposition of Tn7/*pco/sil* Into the Host Chromosome. Front. Microbiol. 12, 723322. doi: 10.3389/fmicb.2021.723322.
- 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., 2021. Increasing New Delhi metallo- β -lactamase-positive *Escherichia coli* among carbapenem non-susceptible *Enterobacteriaceae* in Taiwan during 2016 to 2018. Sci. Rep. 11 (1), 2609. doi: 10.1038/s41598-021-82166-8.

Kaper, J.B., Nataro, J.P., Mobley, H.L., 2004. Pathogenic *Escherichia coli*. Nat. Rev. Microbiol. 2 (2), 123-140. doi: 10.1038/nrmicro818.

Letunic, I., Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 47 (W1), W256-W259. doi:10.1093/nar/gkz239.

Liu, C., Shan, C., Dong, Q., Fu, G., Zhao, R., Yan, Y., Gao, H., 2018. Pathogenic *E. coli* HPI upregulate the expression of inflammatory factors in porcine small intestinal epithelial cells by ubiquitin proteasome pathway. Res. Vet. Sci. 120, 41-46. doi: 10.1016/j.rvsc.2018.08.009.

Lopes, R., Fuentes-Castillo, D., Fontana, H., Rodrigues, L., Dantas, K., Cerdeira, L., Henriques, I., Lincopan, N., 2021. Endophytic Lifestyle of Global Clones of Extended-Spectrum β -Lactamase-Producing Priority Pathogens in Fresh vegetables: a Trojan Horse Strategy Favoring Human Colonization? MSystems. 6 (1), e01125-20. doi: 10.1128/mSystems.01125-20.

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., Paterson, D.L., Rice, L.B., Stelling, J., Struelens, M.J., Vatopoulos, A., Weber, J.T., Monnet, D.L., 2012. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clin. Microbiol. Infect. 18 (3), 268-281. doi: 10.1111/j.1469-0691.2011.03570.x.

- Magistro, G., Magistro, C., Stief, C.G., Schubert, S., 2017. The high-pathogenicity island (HPI) promotes flagellum-mediated motility in extraintestinal pathogenic *Escherichia coli*. PLoS One. 12 (10), e0183950. doi: 0.1371/journal.pone.0183950.
- Manges, A.R., Geum, H.M., Guo, A., Edens, T.J., Fibke, C.D., Pitout, J.D.D, 2019. Global Extraintestinal Pathogenic *Escherichia coli* (ExPEC) Lineages. Clin. Microbiol. Rev. 32 (3), e00135-18. doi: 10.1128/CMR.00135-18.
- Matamoros, S., van Hattem, J.M., Arcilla, M.S., Willemse, N., Melles, D.C., Penders, J., Vinh, T.N., Thi Hoa, N., Bootsma, M.C.J., van Genderen, P.J., Goorhuis, A., Grobusch, M., Molhoek, N., Oude Lashof, A.M.L., Stobberingh, E.E., Verbrugh, H.A., de Jong, M.D., Schultsz, C., 2017. Global phylogenetic analysis of *Escherichia coli* and plasmids carrying the *mcr-I* gene indicates bacterial diversity but plasmid restriction. Sci. Rep. 7 (1), 15364. doi: 10.1038/s41598-017-15539-7.
- Meneguzzi, M., Pissetti, C., Rebelatto, R., Trachsel, J., Kuchiishi, S.S., Reis, A.T., Guedes, R.M.C., Leão, J.A., Reichen, C., Kich, J.D., 2021. Re-Emergence of *Salmonellosis* in Hog Farms: Outbreak and Bacteriological Characterization. Microorganisms. 9 (5), 947. doi: 10.3390/microorganisms9050947.
- Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New Models and Efficient Methods

- for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* .37 (5), 1530-1534. doi:10.1093/molbev/msaa015.
- Mostafa, H.H., Cameron, A., Taffner, S.M., Wang, J., Malek, A., Dumyati, G., Hardy, D.J., Pecora, N.D., 2020. Genomic Surveillance of Ceftriaxone-Resistant *Escherichia coli* in Western New York Suggests the Extended-Spectrum β -lactamase blaCTX-M-27 Is Emerging on Distinct Plasmids in ST38. *Front Microbiol.* 11, 1747. doi: 10.3389/fmicb.2020.01747.
- 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., 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics.* 31 (22), 3691-3693. doi:10.1093/bioinformatics/btv421.
- Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann T., Keane, J.A., Harris, S.R., 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genom.* 2 (4), e000056. doi:10.1099/mgen.0.000056.
- Peirano, G., Mulvey, G.L., Armstrong, G.D., Pitout, J.D.D., 2013. Virulence potential and adherence properties of *Escherichia coli* that produce CTX-M and NDM β -lactamases. *J. Med. Microbiol.* 62 (4), 525-530. doi: 10.1099/jmm.0.048983-0.
- Peirano, G., van der Bij, A.K., Gregson, D.B., Pitout, J.D., 2012. 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.* 50, 294–299. <https://doi.org/10.1128/JCM.06025-11>.
- Pitout, J.D., 2012. Extraintestinal pathogenic *Escherichia coli*: a combination of virulence with antibiotic resistance. *Front. Microbiol.* 3, 9. <https://doi.org/10.3389/fmicb.2012.00009>.
- Pitout, J.D.D., Peirano, G., Kock, M.M., Strydom, K.-A., Matsumura, Y., 2019. The Global Ascendency of OXA-48-Type Carbapenemases. *Clin. Microbiol. Rev.* 33 (1), e00102-19. doi: 10.1128/CMR.00102-19.
- Salgado-Caxito, M., Benavides, J.A., Adell, A.D., Paes, A.C., Moreno-Switt, A.I., 2021. 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* 12, 100236. doi: 10.1016/j.onehlt.2021.100236.
- Sarowska, J., Futoma-Koloch, B. Jama-Kmiecik, A., Frej-Madrzak, M., Ksiazczyk, M., Bugla-Ploskonska, G., Choroszy-Krol, I., 2019. Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: recent reports. *Gut Pathog.* 11, 10. doi: 10.1186/s13099-019-0290-0.
- Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 30 (14), 2068-2069. doi:10.1093/bioinformatics/btu153.

Sekizuka, T., Matsui, M., Yamane, K., Takeuchi, F., Ohnishi, M., Hishinuma, A., Arakawa, Y., Kuroda, M., 2011. 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. 6 (9), e25334. doi: 10.1371/journal.pone.0025334.

Shaik, S., Ranjan, A., Tiwari, S.K., Hussain, A., Nandanwar, N., Kumar, N., Jadhav, S., Semmler, T., Baddam, R., Islam, M.A., Alam, M., Wieler, L.H., Watanabe, H., Ahmed, N., 2017. Comparative Genomic Analysis of Globally Dominant ST131 Clone with Other Epidemiologically Successful Extraintestinal Pathogenic *Escherichia coli* (ExPEC) Lineages. mBio. 8 (5), e01596-17. doi: 10.1128/mBio.01596-17.

Shan, C., Miao, S., Liu, C., Zhang, B., Zhao, W., Wang, H., Yang, W., Cha, J., Zhao, R., Xiao, P., Gao, H., 2021. Induction of macrophage pyroptosis-related factors by pathogenic *E. coli* high pathogenicity island (HPI) in Yunnan Saba pigs. BMC Vet. Res. 17 (1), 114. doi: 10.1186/s12917-021-02824-x.

Sheppard, A.E., Stoesser, N., Wilson, D.J., Sebra, R., Kasarskis, A., Anson, L.W., Giess, A., Pankhurst, L.J., Vaughan, A., Grim, C.J., Cox, H.L., Yeh, A.J., Modernising Medical Microbiology (MMM) Informatics Group, Sifri, C.D., Walker, A.S., Peto, T.E., Crook, D.W., Mathers, A.J., 2016. Nested Russian Doll-Like Genetic Mobility Drives Rapid Dissemination of the Carbapenem Resistance Gene blaKPC. Antimicrob. Agents. Chemother. 60 (6), 3767-3778. doi: 10.1128/AAC.00464-16.

Shrestha, B., Tada, T., Shimada, K., Shrestha, S., Ohara, H., Pokhrel, B.M., Sherchand, J.B., Kirikae, T., 2017. Emergence of Various NDM-Type-Metallo-beta-Lactamase-Producing *Escherichia coli* Clinical Isolates in Nepal. *Antimicrob. Agents Chemother.* 61 (12), e01425-17. doi: 10.1128/AAC.01425-17.

Turton, J.F., Doumith, M., Hopkins, K.L., Perry, C., Meunier, D., Woodford, N., 2016. Clonal expansion of *Escherichia coli* ST38 carrying a chromosomally integrated OXA-48 carbapenemase gene. *J. Med. Microbiol.* 65 (6), 538-546. doi: 10.1099/jmm.0.000248.

Vila, J., Sáez-López, E., Johnson, J.R., Römling, U., Dobrindt, U., Cantón, R., Giske, C.G., Naas, T., Carattoli, A., Martínez-Medina, M., Bosch, J., Retamar, P., Rodríguez-Baño, J., Baquero, F., Soto, S.M., 2016. *Escherichia coli*: an old friend with new tidings. *FEMS Microbiol. Rev.* 40 (4), 437-463. doi: 10.1093/femsre/fuw005.

Wong, M.H.-Y., Chan, E.W.-C., Xie, X., Li, R., Chen, S., 2016. IncHI2 Plasmids Are the Key Vectors Responsible for oqxAB Transmission among *Salmonella* Species. *Antimicrob. Agents Chemother.* 60 (11), 6911-6915. doi: 10.1128/AAC.01555-16.

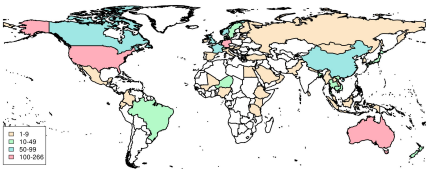
Yamamoto, T., Takano, T., Iwao, Y., Hishinuma, A., 2011. Emergence of NDM-1-positive capsulated *Escherichia coli* with high resistance to serum killing in Japan. *J. Infect. Chemother.* 17 (3), 435-439. doi: 10.1007/s10156-011-0232-3.

Zhao, H., Chen, W., Xu, X., Zhou, X., Shi, C., 2018. Transmissible ST3-IncHI2 Plasmids Are Predominant Carriers of Diverse Complex IS26-Class 1 Integron Arrangements in Multidrug-Resistant *Salmonella*. *Front. Microbiol.* 9, 2492. doi: 10.3389/fmicb.2018.02492.

Zhou, Z., Alikhan, N-F., Mohamed, K., Fan, Y., Agama Study Group; Achtman, M., 2020. The Enterobase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity. *Genome Res.* 30 (1), 138-152. doi: 10.1101/gr.251678.119.

Fig. 1. Map of epidemiological distribution of *E. coli* ST38 based on available metadata from EnteroBase (September, 2021).

Fig. 2. Core SNP tree based on *E. coli* ST38 genomes. Highlighted labels represent the clade 1 (green) and clade 2 (yellow). The circles outside the tree, from innermost to outermost, represent metadata of origin, isolation date, and isolation location, respectively. Genomes featuring the High Pathogenicity Island (HPI) are marked with a black circle. Red branches have bootstrap values greater than 70. An interactive version of the tree is available for specific observations: <https://itol.embl.de/tree/1578617753161201632757536>.



Colored background

- Clade 1 ST38
- Clade 2 ST38
- Other STs

Source

- Human
- Environment
- Animal

Date

- 1983 - 1986
- 1993 - 1997
- 2002 - 2012
- 2013 - 2018
- 2019 - 2020

Location

- Africa
- Asia
- Europe
- North America
- South America

