Revista Argentina de Microbiología 55 (2023) 111–119

asociación argentina de microbiología

REVISTA ARGENTINA DE

www.elsevier.es/ram

ORIGINAL ARTICLE

First molecular characterization of *Escherichia coli* O157:H7 isolates from clinical samples in Paraguay using whole-genome sequencing

Natalie Weiler[a,*], Lucia Jazmín Martínez[a], Josefina Campos[b], Tomas Poklepovich[b], Maria Veronica Orrego[a], Flavia Ortiz[a], Mercedes Alvarez[a], Karina Putzolu[b], Gisela Zolezzi[b], Elisabeth Miliwebsky[b], Isabel Chinen[b]

[a] *Central Public Health Laboratory, 1535 Asunción, Paraguay*
[b] *National Institute of Infectious Diseases – ANLIS ''Dr. Carlos G Malbrán'', C1282 AFF Buenos Aires, Argentina*

Received 6 October 2021; accepted 16 November 2022
Available online 2 January 2023

KEYWORDS
Escherichia coli;
O157:H7;
Foodborne pathogen;
Whole genome sequencing (WGS);
Clade 8

Abstract *Escherichia coli* O157:H7 is a foodborne pathogen implicated in numerous outbreaks worldwide that has the ability to cause extra-intestinal complications in humans. The Enteropathogens Division of the Central Public Health Laboratory (CPHL) in Paraguay is working to improve the genomic characterization of Shiga toxin-producing *E. coli* (STEC) to enhance laboratory-based surveillance and investigation of foodborne disease outbreaks. Whole genome sequencing (WGS) is proposed worldwide to be used in the routine laboratory as a high-resolution tool that allows to have all the results in a single workflow. This study aimed to carry out for the first time, the genomic characterization by WGS of nine STEC O157:H7 strains isolated from human samples in Paraguay. We were able to identify virulence and resistance mechanisms, MLST subtype, and even establish the phylogenetic relationships between isolates. Furthermore, we detected the presence of strains belonging to hypervirulent clade 8 in most of the isolates studied.

PALABRAS CLAVE
Escherichia coli;
O157:H7;
Patógeno transmitido por alimentos;

Primera caracterización molecular de aislamientos de *Escherichia coli* O157:H7 provenientes de muestras clínicas en Paraguay utilizando secuenciación de genoma completo

Resumen *Escherichia coli* O157:H7 es un patógeno transmitido por alimentos implicado en numerosos brotes en todo el mundo y es capaz de causar complicaciones extraintestinales en humanos. La sección de «Enteropatógenos» del Laboratorio Central de Salud Pública trabaja

* Corresponding author.
E-mail address: natalieweiler@gmail.com (N. Weiler).

https://doi.org/10.1016/j.ram.2022.11.002

Secuenciación de genoma completo; Clado 8

en mejorar la caracterización genómica de STEC, de modo de potenciar la vigilancia laboratorial y la investigación de brotes de enfermedades transmitidas por alimentos. La secuenciación de genoma completo (WGS, por sus siglas en inglés) se propone a nivel mundial como una herramienta de alta resolución para ser utilizada en el laboratorio de rutina, ya que permite obtener todos los resultados en un único proceso. El objetivo de este trabajo fue llevar a cabo, por primera vez, la caracterización genómica por WGS de nueve cepas STEC O157:H7 aisladas en Paraguay a partir de muestras de origen humano. Pudimos identificar los factores de virulencia, los mecanismos de resistencia, el subtipo MLST, e incluso pudimos establecer la relación filogenética entre los aislamientos. Además, detectamos que la mayoría de las cepas pertenecían al clado hipervirulento 8.

Introduction

Shiga toxin-producing *Escherichia coli* (STEC) strains constitute a subset of pathogenic *E. coli* that have been associated with sporadic cases and outbreaks of diarrhea (D), bloody diarrhea (BD) and hemolytic uremic syndrome (HUS) in humans[17–19]. STEC strains are characterized by the production of Shiga-toxins (Stx), which are AB5 cytotoxins encoded on the bacterial chromosome that inhibit protein synthesis and can cause cell death[3]. *E. coli* O157:H7 is the most well-known Stx-producing serotype within this group and most associated with severe disease and outbreaks[25,39,42]. There are two types of Stx (Stx_1 and Stx_2) and each one is associated with a different clinical outcome[4,21]. In fact, in mouse models, Stx_2 has shown to be 100 times more potent than Stx_1[43]. Moreover, in primate models the administration of Stx_2 alone can produce symptoms of HUS, which are not produced by the administration of Stx_1 alone[38,40].

Stx_1 and Stx_2 are further divided into subtypes associated with different levels of virulence[37]. For instance, Stx_{2a}, Stx_{2c}, Stx_{2d} are more often associated with HUS. In contrast, the other Stx_1 and Stx_2 subtypes are mainly associated with mild disease[4,21]. Besides Stx, additional virulence factors contribute to the pathogenicity of the STEC strains associated with severe diseases. The most common ones are located in the pathogenicity island called the locus of enterocyte effacement (LEE) and the O157 plasmid[17,30,31]. The LEE island encodes several proteins needed to produce attaching and effacing (A/E) lesions on the intestinal mucosa, such as intimin (encoded by the *eae* gene) that closely attaches the bacteria to the intestinal epithelium, the translocated intimin receptor (Tir), and other Type 3 secretion system (T3SS) effectors[17,29]. The O157 plasmid encodes for several virulence proteins, i.e., enterohemolysin (*ehxA*), a catalase-peroxidase (*KatP*), extracellular serine-protease (*espP*), adhesines (toxB) and proteins of T3SS[27].

STEC O157:H7 populations have shown to be genotypically diverse, and variations of disease severity in humans are associated with specific genotypes. Three main lineages (I, II, and I/II) differently distributed between animals, humans and geographical regions were described[22,47]. Furthermore, O157:H7 strains can be divided into nine clades based on outbreak studies associated with different clinical outcomes[28]. Strains belonging to clade 8 are known to be ''hypervirulent'' because they are associated with more severe disease and the potential capacity to cause severe outbreaks[28,36]. Although the biological basis behind the increased pathogenicity of these strains is not fully understood, it is suggested that the differential expression of some genes could play this role[1]. Moreover, Kulasekara et al.[23] found 7 coding sequences postulated as putative virulence determinants when comparing the genome of the TW14359 strain (clade 8) against two other sequenced strains from clade 1 and 3 (EDL-933 and Sakai).

Epidemiological surveillance of STEC is a fundamental public health issue to limit the spread, detect outbreaks, and trace back contamination sources, allowing the development of efficient and targeted intervention strategies. In Paraguay, the STEC surveillance program was implemented in 2002. Since then, 40 STEC strains have been isolated from clinical samples, 10 of which were attributed to the O157:H7 serotype (CPHL – Paraguay). This work aimed to achieve the genomic characterization and study the genetic diversity of STEC O157:H7 strains isolated from human samples in Paraguay to identify the subtypes associated with high pathogenicity in order to provide insights into the risk they pose to public health.

Materials and methods

Bacterial strains

From 2002 to 2018, 10 *E. coli* O157:H7 strains were detected in stool samples from BD or HUS cases in Paraguay through the STEC surveillance program. For this study, 9 out of 10 strains were retrieved and subjected to the traditional genotypic characterization and genomic analysis by WGS at the CPHL in Paraguay and INEI-ANLIS ''Dr. Carlos G. Malbrán'' in Argentina.

Characterization

Traditional characterization was performed by serotyping, biotyping and PCR genotyping of the main virulence

factors (*stx*, *eae*, *exhA*) following the protocol described in the Procedure Manual for the Diagnosis and Characterization of Shiga-toxin producing *E. coli* from human samples by molecular biology techniques[15] (INEI/ANLIS Malbrán).

S*tx* subtyping was performed as described by Scheutz et al.[37]. The strains were subtyped by PFGE using the *Xba*I enzyme according to the International PulseNet protocol[35]. The subsequent analysis was performed using the GelCompar II software version 5.1 (Applied Maths, Sint-Martens-Latem, Belgium) using the Dice coefficient of similarity and the UPGMA (Unweighted Pair Group Method with Arithmetic mean) grouping method to construct dendrograms (tolerance 1.5%) and determine the genetic relationship.

The resistance profile was determined by testing the following antibiotics: ampicillin (AMP), amoxicillin/clavulanic acid (AMC), azithromycin (AZI), cefotaxime (CTX), ceftazidime (CTZ), nalidixic acid (NAL), gentamicin (GEN), ciprofloxacin (CIP), tetracycline (TCY), trimethoprim/sulfamethoxazole (SXT), chloramphenicol (CHL), nitrofurantoin (NIT), by agar diffusion (Kirby-Bauer method) following the CLSI (Clinical and Laboratory Standards Institute) regulations[7] (2019 version).

Whole-genome sequencing

Genomic DNA (gDNA) was extracted and purified from the isolates using Magpurix equipment with the MagPurix Bacterial DNA Extraction Kit by Zinexts Life Science Corp. DNA concentrations were determined using the QuantiFluor® dsDNA System reagent kit with the Quantus™ Fluorometer (Promega Corporation).

Libraries were prepared using the Nextera XT DNA library prep kit (Illumina) according to the manufacturer's guidelines.

Sequencing was performed by the genomic platform of INEI/ANLIS Malbrán (Buenos Aires – Argentina) and by the CPHL (Asuncion/Paraguay) on the Illumina MiSeq platform using the MiSeq Reagent Kit V2.

Serotyping

The Short Read Sequence Typing performed the serotyping of bacterial pathogens using the (SRST2) (v.2.0) tool[14]. The reads in FASTQ format were mapped against the reference database (EcOH database available at https://github.com/katholt/srst2) in FASTA format using the following parameters: 85% threshold for sequence identity and 60% minimum length for a hit.

Virulome and AMR genes

The prediction of virulence genes was performed from the short-reads through the ARIBA software[13] (Ariba-2.14.4 version) using the virulence finder database[16] (2020-05-29 version) (available at https://bitbucket.org/genomicepidemiology/virulencefinder_db/src/master/) with a threshold of 95% identity and a minimum length of 80% for a hit.

The Stx subtype was determined from the assembled reads in FASTA format by an *in-silico* PCR performed using the Ipcress program developed by Guy St.C. Slater (version 0.8.3) using the primers described by Scheutz et al.[37]. Additional Stx subtyping was performed with the program *E. coli* Shiga-toxin-typing v.2.0 (Blast v2.9 against Shiga toxin-type databases from the Statens Serum Institut SSI and the Technical University of Denmark DTU available at https://github.com/aknijn/shigatoxin-galaxy) in the Galaxy platform (Galaxy version 2.0 https://aries.iss.it/).

Putative virulence factors described by Kulasekara et al.[23] were screened by *in-silico* PCR using the Ipcress program for the locus tags ECSP_3620 (encoding the anaerobic nitric oxide reductase NorV), ECSP_0242 (encoding a domain that facilitates protein–protein interactions), ECSP_3286 (encoding a product that binds with high affinity to heme), ECSP_2870/2872 (encoding proteins that are involved in the adaptation to plant hosts), ECSP_2687 (encoding proteins that reduce the host immune response) and ECSP_1773 (encoding a protein that interferes with innate immunity).

Antimicrobial resistance (AMR) was screened from both short-reads and assemblies in FASTA format. The short-reads were mapped against Res_Finder database[46] (v.4.1) (available at https://bitbucket.org/genomicepidemiology/resfinder_db/src/master/) using the ARIBA software with requirements of 95% identity threshold and a minimum length of 80%. Additional AMR screenings were performed using the AMRfinder program[9] from FASTA files.

Plasmid search was conducted by mapping FASTA files against Plasmid_Finder database[6] (available at https://bitbucket.org/genomicepidemiology/plasmidfinder_db/src/master/).

Epidemiological typing and phylogenetic analysis

In-silico multilocus sequence typing (MLST) was performed by comparing the raw reads in FASTQ format with the PubMLST database downloaded from the PubMLST website (https://pubmlst.org) using the ARIBA software. Seven house-keeping genes (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, *and recA*) were used for MLST analysis.

The lineage-specific polymorphism assay (LSPA-6) was performed by *in-silico* PCR using the Ipcress program according to the scheme described by Yang et al.[45] Isolates with genotype 111111 were classified as LSPA-6 LI, those with genotype 211111 as LSPA-6 LI/II and those with other derivations were classified as LSPA-6 LII.[8]

Determination of O157:H7 clades was performed by *in-silico* PCR using the Ipcress program according to the 4 SNPs algorithm developed by Riordan et al.[36].

In addition to the genomic characterization, we decided to study the genomic relatedness among our isolates and other *E. coli* O157:H7 isolated from humans around the world. To accomplish that goal, we submitted the short-read sequences to EnteroBase[48] (v1.1.2). We determined the closest relatives to our population of *E. coli* isolates, taking the HC100 cluster information obtained by using the hierarchical clustering (HierCC) algorithm in EnteroBase and chose 50 closest relatives worldwide whose FASTQ files were available at SRA GenBank from the same period that our strains were collected. Both phylogenetic analyses (one between our nine *E. coli* O157:H7 strains and the other between our strains

Table 1 Traditional characterization results, epidemiological information and PFGE profile of the isolates.

ID	Year	Origin	Patient age	Bio-type	Serotype	Stx	Stx subtype	Virulence profile	AMR profile	PFGE PUN
1049	2002	BD	1	C	O157:H7	Stx_2	stx_{2c}	*eae+/exhA+*	–	PYEXHX01.0001
1176	2002	BD	6	C	O157:H7	Stx_2	stx_{2c}	*eae+/exhA+*	AMP/SXT	PYEXHX01.0002
713	2004	HUS	11m	C	O157:H7	Stx_2	stx_{2c}	*eae+/exhA+*	–	PYEXHX01.0004
1985	2008	BD	5	C	O157:H7	Stx_2	stx_{2a+c}	*eae+/exhA+*	AMP	PYEXHX01.0005
1903	2008	HUS	2	C	O157:H7	Stx_2	stx_{2a+c}	*eae+/exhA+*	AMP	PYEXHX01.0005
13316	2014	BD	1	C	O157:H7	Stx_2	stx_{2c}	*eae+/exhA+*	–	PYEXHX01.0006
73755	2015	BD	33	C	O157:H7	Stx_2	stx_{2c}	*eae+/exhA+*	–	PYEXHX01.0007
203577	2017	BD	1	C	O157:H7	Stx_2	stx_{2a+c}	*eae+/exhA+*	AMP	PYEXHX01.0008
303191	2018	BD	2	C	O157:H7	Stx_2	stx_{2a}	*eae+/exhA+*	–	PYEXHX01.0009

and the 50 selected) were done from the paired end reads that were mapped against the *E. coli* O157:H7_str_EDL933 genome used as a reference. For this purpose, we used Lyve-SET, a pipeline that uses high-quality single-nucleotide polymorphism (hqSNPs) to create a phylogeny[20]. The output was then used as an input with the FigTree program developed by Andrew Rambaut (version 1.4.4.) to build the phylogenetic tree. Finally, trees obtained from FigTree were used in the iTOL v6 program[26] (https://itol.embl.de) for a friendlier visualization.

Publicly available genomes

All raw sequence data for the Paraguayan *E. coli* genomes analyzed in this publication are publicly available under Bioproject IDs PRJNA277984 and PRJNA764717. The accession numbers of the Paraguayan and the other strains used in the analyses are given in Supplementary Table 1.

Results

The genetic characterization showed that all the strains (9) carried the stx_2 gene. Furthermore, the subtypes detected were: stx_{2c} (5/9 strains), stx_{2a+c} (3/9) and stx_{2a} (1/9). All the strains harbored *eae*, *ehxA*, rfb_{O157}, and $fliC_{H7}$ genes. With regard to the antibiogram, 4 strains were ampicillin-resistant. At the same time, one ampicillin-resistant strain was also resistant to trimethoprim–sulfamethoxazole.

*Xba*I-PFGE, 7/9 strains generated different patterns except for two strains isolated in the same year and month (1985 and 1903) that showed one unique pattern (PYEXHX01.0005). These results are shown in Table 1, along with the epidemiological information about the isolates.

As expected, all the strains harbored the *wzx/wzy*-O157 gene and the $fliC_{H7}$ gene.

With regard to the virulence profile, all strains harbored stx_2, *eae* and *exhA* The virulence genotype analysis by WGS allowed the identification of a vast number of virulence factors that are crucial for the pathogenicity mechanism of the strains such as *tir*, the group of *esp* proteins (A, B, F, J), *katP*, *nle* (A, B), *tccP*, *toxB*, among others. Of all the virulence genes present in the database, Figure 1 shows only the results for the genes found for one strain.

In-silico PCR was performed from the FASTA files to determine the Shiga-toxin subtype. Of all nine samples, seven subtyping results agreed with those found by PCR. Two samples identified as stx_{2a+c} by PCR (1903 and 1985) were of the stx_{2a} subtype by *in-silico* PCR. These two sequences were then analyzed in a complementary manner through the program *E. coli* Shiga-toxin-typing in the Galaxy platform to verify and confirm results. The 1985 sample was found to be stx_{2a+c} subtype by this method (in agreement with PCR results), while the results for the 1903 sample was again the stx_{2a} subtype.

As for the putative virulence determinants, all the isolates carried ECSP_3620 and ECSP_0242, 8/9 ECSP_2687, 5/9 ECSP_3286, 4/9 ECSP_2870/2872 and 3/9 ECSP_1773 (Fig. 2).

AMR analysis identified 7 genes associated with resistance to β-lactamases (bla_{EC}, bla_{TEM}), aminoglycosides (*aph(6)-Id*, *aph(3')-Id*), sulfonamides (*sul2*) and trimethoprim (*dfrA8*) among the strains (Fig. 1). Three plasmid replicons sequences were found among the O157:H7 strains corresponding to IncFIB(AP001918), IncFII and IncI1-I (Gamma) incompatibility groups (Fig. 1)

All the strains were found to be ST 11 (allelic profile *adk*:12 *fumC*:12 *gyrB*:8 *icd*:12 *mdh*:15 *purA*:2 *recA*:2).

As for the LSPA-6 assay, we could not assign lineage to three isolates since the amplicon for the marker gene *Z5935* was not predictable by *in-silico* PCR. Five of the strains belonged to lineage I/II, and one to lineage II. Clade determination was performed according to the algorithm developed by Riordan et al.[36]. Six out of nine strains (6/9) belonged to the hypervirulent clade 8, and the three others belonged to clades 4–7 or 9 (Fig. 2). Clades 4–7 and 9 strains share the same SNP profile for the target genes in the clade typing method described by Riordan et al.[36]; therefore, they cannot be differentiated by this method.

The phylogenetic tree obtained by mapping our nine *E. coli* O157:H7 paired-end reads against a reference genome showed the genetic diversity of the strains, with strains 303191 and 13316 being the most phylogenetically distant (638 SNPs). However, strains 1903 and 1985 (collected in the same year) appeared to be closely related (1 SNP), thus confirming the results found by PFGE. Moreover, we could observe that strains belonging to clade 8 clustered together, supporting the results of the *in-silico* PCR for clade detection (Fig. 2).

We further investigated the genomic relatedness between our *E. coli* O157:H7 strains and other *E. coli* O157:H7 isolates from humans worldwide. Results showed that the Paraguayan strains are genotypically diverse, and

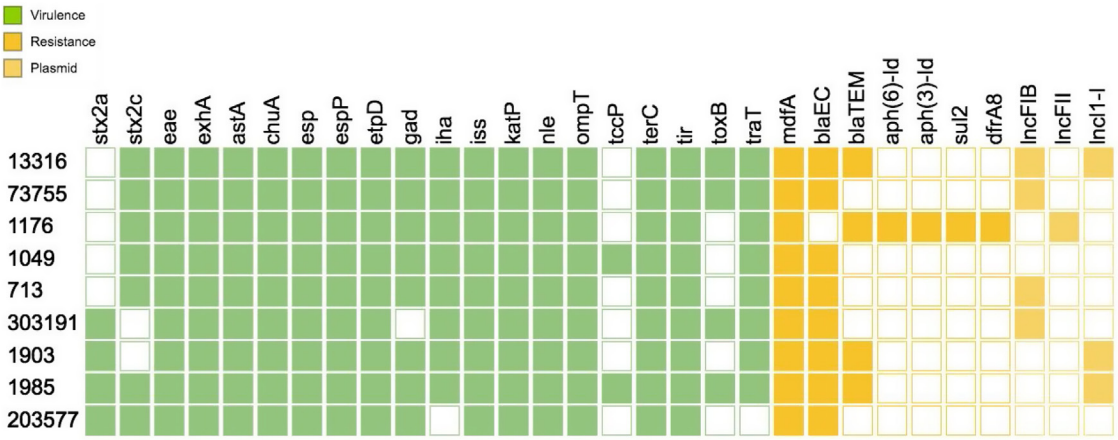

Figure 1 Summary of virulence, resistance and plasmid profiles found by the WGS analysis of the studied strains. The figure shows the genetic elements found in at least one strain among all those available in the databases.

Figure 2 hqSNPs phylogenetic tree and main characteristics of *E. coli* O157:H7 isolated from human samples in Paraguay.
[a] The amplicon for the marker gene *Z5935* was not predictable by *in-silico* PCR and hence no lineage was assigned to these isolates.
[b] The isolates belonged to one of clades 4, 5, 6, 7 or 9 since the subtyping scheme described by Riordan et al. cannot differentiate strains from clade 4, 5, 6, 7, and 9.

are not especially closely related to strains from other specific locations (Fig. 3).

Discussion

The objective of this study was to achieve a genomic characterization of isolated strains of *E. coli* O157:H7 in Paraguay by performing the WGS analysis. We decided to investigate STEC O157:H7 STEC strains from clinical cases because they are one of the most prevalent and virulent serotypes associated with severe human diseases such as HUS[17–19,42]. A study conducted in 2017 found that the predominant causative agent in bacterial diarrhea in Paraguay was diarrheagenic *E. coli*, which is responsible for 13.4% of cases[12,44]. Epidemiological data for the first quarter of 2021 showed that diarrheagenic *E. coli* continues to top the list of bacterial agents that cause diarrheal disease with a 14% prevalence (unpublished data-CPHL).

There are no extensive publications on HUS in Paraguay. The first documented case dates from 1987, with no description of the etiological agents. It was not until 2009 when a clinical case of HUS due to *E. coli* O157:H7 Stx_2 was documented for the first time[32]. Since 2002, when the STEC surveillance started, it was noted that STEC in diarrhea appears to be relatively low in Paraguay (0.6–0.8%). However, the monitoring and characterization of STEC are essential due to its high pathogenic power. In fact, all our O157:H7 strains were confirmed to have a high pathogenic potential, as they all harbor the *stx*$_2$, *eae*, and *exhA* genes. It has been shown that the strains encoding for *stx*$_2$ are more virulent and are associated with more severe disease than those encoding *stx*$_1$[10], and also that intimin (*eae*), as well as enterohaemolysin (*exhA*) play a significant role in STEC pathogenicity[29]. Moreover, WGS allowed to identify many other virulence genes implicated in adhesion (*toxB*, *iha*), survival (*chuA*, *katP*, *gad*, *iss*, *ompT*), transduction (*esp*), secretion (*etpD*), as well as important effectors (*nle*, *chuA*, *espP*, *tir*), and toxins (*astA*) from one single analysis. Besides *stx*$_2$, *eae* and *exhA*; all strains harbored *nle* genes. Another important virulence factor found in all strains was *espP*,

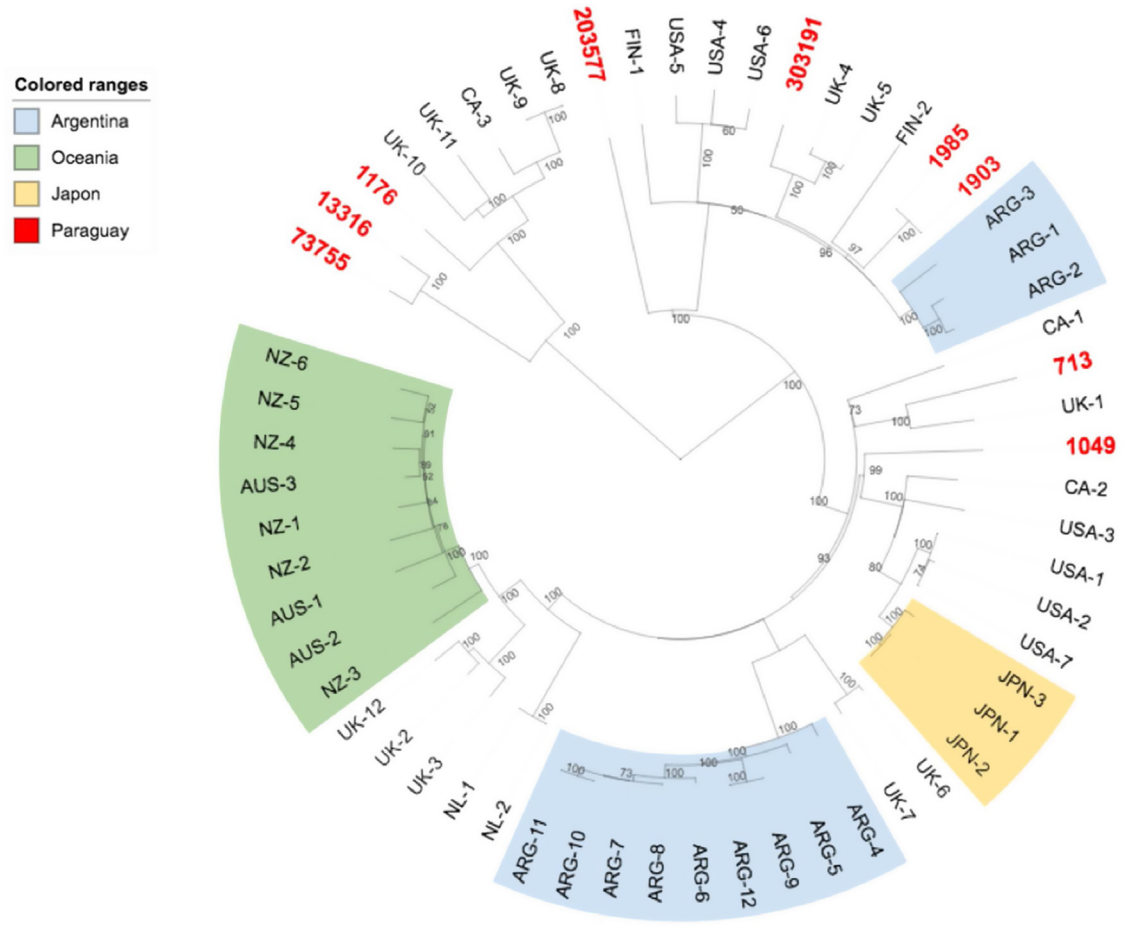

Figure 3 Phylogenetic tree obtained by mapping our *E. coli* O157:H7 strains and others collected worldwide against reference.

which might be implicated in the pathogenesis of bloody diarrhea, as EspPα can cleave human coagulation factor V. Further investigation regarding the subtype of the *espP* gene is necessary since EspPβ and EspPδ are impaired in their autotransporter or proteolytic activity[5,24,33]. Regarding the Stx subtype, stx_{2c} and stx_{2a+c} were the most prevalent. Subtype stx_{2a} was found exclusively among clade 8 strains. These subtypes are, in fact, more frequently associated with human disease[28]. It is important to point out that WGS analysis failed to identify stx_{2c} from one strain that harbored stx_{2a}. This situation was previously reported by others[2,11], and it may be due to the limitations in *de novo* assembly from short reads because the regions of variation between these subtypes (stx_{2a} and stx_{2c}) are concentrated at the 5′ and 3′ ends of the coding DNA sequence (CDS), with a largely homogenous region in the center[2]. With regard to AMR, 4 isolates carried the bla_{TEM} gene and 3 were resistant to ampicillin. One of the strains carrying the bla_{TEM} gene also carried genes that confer resistance to aminoglycosides, sulfonamides, and trimethoprim. Antibiotic susceptibility tests of this strain revealed resistance to ampicillin and trimethoprim–sulfamethoxazole. Interestingly, this strain is the only one with the presumption of carrying the IncFII plasmid incompatibility group.

Although the antibiotic treatment of STEC infections in humans is not recommended, AMR represents a severe threat, especially in highly pathogenic strains, since its ability to remain in the environment and/or animals is increased, thus facilitating its spread.

Regarding the epidemiological characteristics, all the strains studied were from ST 11. This ST was widely isolated worldwide from humans, food, and the environment (EnteroBase). The LSPA-6 assay revealed that none of the strains was from lineage I, which was initially thought to be more frequently associated with human isolates[24,45,47]. Concerning clade designation, a dominance of hypervirulent clade 8 (6/9 strains) was observed, coinciding with other researchers' findings in Argentina[34,41]. Furthermore, the two HUS cases were caused by strains belonging to clade 8. As for the putative virulence factors, ECS_3620, ECS_0242 and ECS_2687 were present in strains from clade 8 and non-8, while ECS_3286, ECS_1773 and ECS_2870/2872 were present in strains from clade 8 exclusively.

The phylogenetic tree obtained with the subset of *E. coli* O157:H7 strains from Paraguay and 50 closest isolates worldwide showed that while the Paraguayan O157:H7 strains are rather homogeneous regarding clade and lineage assignation, they are genetically diverse. We observed that they are mainly spread among different clusters contrary to Australia's, Argentine's, New Zealand's chosen subsets of samples that appeared to be more homogenous.

Conclusions

WGS has become an increasingly attractive tool for foodborne disease surveillance worldwide since its numerous advantages, such as enhanced resolution and time saving features, enable a broad characterization using a single process. Therefore, implementing WGS in Paraguay will allow us to identify, characterize and subtype foodborne pathogens and support an appropriate outbreak response by tracing the source of infection and elucidating transmission pathways, which is very important to identify potential intervention points. Through WGS we achieved an exhaustive molecular characterization of *E. coli* O157:H7 strains circulating in Paraguay for almost 20 years. For the first time in Paraguay, we describe the circulation of strains belonging to lineage I/II and clade 8. All strains harbored the *eae*, *exhA* and stx_2 genes. These characteristics describe the O157:H7 strains isolated from humans in Paraguay as a high risk for public health.

Funding

This work did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

This work was carried out with the close collaboration of the genomics department and physiopathogenesis service of the Malbrán Institute – Buenos Aires – Argentina.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ram.2022.11.002.

References

1. Abu-Ali GS, Ouellette LM, Henderson ST, Lacher DW, Riordan JT, Whittam TS, Manning SD. Increased adherence and expression of virulence genes in a lineage of *Escherichia coli* O157:H7 commonly associated with human infections. PLOS ONE. 2010;5:e10167, http://dx.doi.org/10.1371/journal.pone.0010167 [PMID: 20422047; PMCID: PMC2858043].
2. Ashton PM, Perry N, Ellis R, Petrovska L, Wain J, Grant KA, Jenkins C, Dallman TJ. Insight into Shiga toxin genes encoded by *Escherichia coli* O157 from whole genome sequencing. PeerJ. 2015;3:e739, http://dx.doi.org/10.7717/peerj.739 [PMID: 25737808; PMCID: PMC4338798].
3. Bai X, Fu S, Zhang J, Fan R, Xu Y, Sun H, He X, Xu J, Xiong Y. Identification and pathogenomic analysis of an *Escherichia coli* strain producing a novel Shiga toxin 2 subtype. Sci Rep. 2018;8:6756, http://dx.doi.org/10.1038/s41598-018-25233-x [PMID: 29712985; PMCID: PMC5928088].
4. Boerlin P, McEwen SA, Boerlin-Petzold F, Wilson JB, Johnson RP, Gyles CL. Associations between virulence factors of Shiga toxin-producing *Escherichia coli* and disease in humans. J Clin Microbiol. 1999;37:497–503, http://dx.doi.org/10.1128/JCM.37.3.497-503.1999 [PMID: 9986802; PMCID: PMC84443].
5. Brockmeyer J, Bielaszewska M, Fruth A, Bonn ML, Mellmann A, Humpf HU, Karch H. Subtypes of the plasmid-encoded serine protease EspP in Shiga toxin-producing *Escherichia coli*: distribution, secretion, and proteolytic activity. Appl Environ Microbiol. 2007;73:6351–9, http://dx.doi.org/10.1128/AEM.00920-07 [Epub 2007 17; PMID: 17704265; PMCID: PMC2075056].
6. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup F, Hasman H. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014;58:3895–903, http://dx.doi.org/10.1128/AAC.02412-14 [Epub 28.04.14; PMID: 24777092; PMCID: PMC4068535].
7. CLSI. Performance standards for antimicrobial susceptibility testing. CLSI supplement M100. 29th ed. Wayne, PA: Clinical and Laboratory Standards Institute; 2019.
8. Dallman TJ, Ashton PM, Byrne L, Perry NT, Petrovska L, Ellis R, Allison L, Hanson M, Holmes A, Gunn GJ, Chase-Topping ME, Woolhouse MEJ, Grant KA, Gally DL, Wain J, Jenkins C. Applying phylogenomics to understand the emergence of Shiga-toxin-producing *Escherichia coli* O157:H7 strains causing severe human disease in the UK. Microb Genom. 2015;1:e000029, http://dx.doi.org/10.1099/mgen.0.000029 [PMID: 28348814; PMCID: PMC5320567].
9. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, Tyson GH, Zhao S, Hsu CH, McDermott PF, Tadesse DA, Morales C, Simmons M, Tillman G, Wasilenko J, Folster JP, Klimke W. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype–phenotype correlations in a collection of isolates. Antimicrob Agents Chemother. 2019;63:e00483–519, http://dx.doi.org/10.1128/AAC.00483-19 [erratum in: Antimicrob Agents Chemother. 2020;64(4); PMID: 31427293; PMCID: PMC6811410].
10. Fuller CA, Pellino CA, Flagler MJ, Strasser JE, Weiss AA. Shiga toxin subtypes display dramatic differences in potency. Infect Immun. 2011;79:1329–37, http://dx.doi.org/10.1128/IAI.01182-10 [Epub 03.01.11; PMID: 21199911; PMCID: PMC3067513].
11. Holmes A, Allison L, Ward M, Dallman TJ, Clark R, Fawkes A, Murphy L, Hanson M. Utility of whole-genome sequencing of *Escherichia coli* O157 for outbreak detection and epidemiological surveillance. J Clin Microbiol. 2015;53:3565–73, http://dx.doi.org/10.1128/JCM.01066-15 [Epub 09.09.15; PMID: 26354815; PMCID: PMC4609728].
12. Huber C, Orrego MV, Ortiz F, Álvarez M, Weiler N. Prevalencia de patógenos causantes de enfermedad diarreica aguda en el área Metropolitana de Asunción y Central. Rev. Salud Pública Parag [online]. 2019;9:41–5, http://dx.doi.org/10.18004/rspp.2019.diciembre.41-45. ISSN 2307-3349.
13. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, Harris SR. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microb Genom. 2017;3:e000131, http://dx.doi.org/10.1099/mgen.0.000131 [PMID: 29177089; PMCID: PMC5695208].
14. Ingle DJ, Valcanis M, Kuzevski A, Tauschek M, Inouye M, Stinear T, Levine MM, Robins-Browne RM, Holt KE. In silico serotyping of *E. coli* from short read data identifies limited novel O-loci but extensive diversity of O:H serotype combinations within and

between pathogenic lineages. Microb Genom. 2016;2:e000064, http://dx.doi.org/10.1099/mgen.0.000064 [erratum in: Microb Genom 2017;3:e000109; PMID: 28348859; PMCID: PMC5343136].
15. Instituto Nacional de Enfermedades Infecciosas – A.N.L.I.S. ''Dr. Carlos G. Malbrán'' – OPS. Manual de Procedimientos. Detección de patógenos asociados a Enfermedad Diarreica Aguda, incluyendo *Vibrio cholerae*. Buenos Aires; 2011.
16. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, Aarestrup FM. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. J Clin Microbiol. 2014;52:1501–10, http://dx.doi.org/10.1128/JCM.03617-13 [Epub 26.02.14; PMID: 24574290; PMCID: PMC3993690].
17. Kaper JB, Nataro JP, Mobley HL. Pathogenic *Escherichia coli*. Nat Rev Microbiol. 2004;2:123–40, http://dx.doi.org/10.1038/nrmicro818 [PMID: 15040260].
18. Karmali MA, Petric M, Lim C, Fleming PC, Arbus GS, Lior H. The association between idiopathic hemolytic uremic syndrome and infection by verotoxin-producing *Escherichia coli*. J Infect Dis. 1985;151:775–82, http://dx.doi.org/10.1093/infdis/151.5.775 [PMID: 3886804].
19. Karmali MA. Infection by Shiga toxin-producing *Escherichia coli*: an overview. Mol Biotechnol. 2004;26:117–22, http://dx.doi.org/10.1385/MB:26:2:117 [PMID: 14764937].
20. Katz LS, Griswold T, Williams-Newkirk AJ, Wagner D, Petkau A, Sieffert C, Van Domselaar G, Deng X, Carleton HA. A comparative analysis of the Lyve-SET phylogenomics pipeline for genomic epidemiology of foodborne pathogens. Front Microbiol. 2017;8:375, http://dx.doi.org/10.3389/fmicb.2017.00375 [PMID: 28348549; PMCID: PMC5346554].
21. Kawano K, Okada M, Haga T, Maeda K, Goto Y. Relationship between pathogenicity for humans and stx genotype in Shiga toxin-producing *Escherichia coli* serotype O157. Eur J Clin Microbiol Infect Dis. 2008;27:227–32, http://dx.doi.org/10.1007/s10096-007-0420-3 [Epub 11.12.07; PMID: 18071766].
22. Kim J, Nietfeldt J, Benson AK. Octamer-based genome scanning distinguishes a unique subpopulation of *Escherichia coli* O157:H7 strains in cattle. Proc Natl Acad Sci U S A. 1999;96:13288–93, http://dx.doi.org/10.1073/pnas.96.23.13288 [PMID: 10557313; PMCID: PMC23940].
23. Kulasekara BR, Jacobs M, Zhou Y, Wu Z, Sims E, Saenphimmachak C, Rohmer L, Ritchie JM, Radey M, McKevitt M, Freeman TL, Hayden H, Haugen E, Gillett W, Fong C, Chang J, Beskhlebnaya V, Waldor MK, Samadpour M, Whittam TS, Kaul R, Brittnacher M, Miller SI. Analysis of the genome of the *Escherichia coli* O157:H7 2006 spinach-associated outbreak isolate indicates candidate genes that may enhance virulence. Infect Immun. 2009;77:3713–21, http://dx.doi.org/10.1128/IAI.00198-09 [Epub 29.06.19; PMID: 19564389; PMCID: PMC2738036].
24. Kuo KH, Khan S, Rand ML, Mian HS, Brnjac E, Sandercock LE, Akula I, Julien JP, Pai EF, Chesney AE. EspP, an extracellular serine protease from enterohemorrhagic *E. coli*, reduces coagulation factor activities, reduces clot strength, and promotes clot lysis. PLOS ONE. 2016;11:e0149830, http://dx.doi.org/10.1371/journal.pone.0149830 [PMID: 26934472; PMCID: PMC4775034].
25. Lang C, Hiller M, Konrad R, Fruth A, Flieger A. Whole-genome-based public health surveillance of less common Shiga toxin-producing *Escherichia coli* serovars and untypeable strains identifies four novel O genotypes. J Clin Microbiol. 2019;57:e00768–819, http://dx.doi.org/10.1128/JCM.00768-19 [PMID: 31366691; PMCID: PMC6760944].
26. Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49:W293–6, http://dx.doi.org/10.1093/nar/gkab301 [PMID: 33885785; PMCID: PMC8265157].
27. Lim JY, La HJ, Sheng H, Forney LJ, Hovde CJ. Influence of plasmid pO157 on *Escherichia coli* O157:H7 Sakai biofilm formation. Appl Environ Microbiol. 2010;76:963–6, http://dx.doi.org/10.1128/AEM.01068-09 [Epub 04.12.09; PMID: 19966025; PMCID: PMC2813019].
28. Manning SD, Motiwala AS, Springman AC, Qi W, Lacher DW, Ouellette LM, Mladonicky JM, Somsel P, Rudrik JT, Dietrich SE, Zhang W, Swaminathan B, Alland D, Whittam TS. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. Proc Natl Acad Sci U S A. 2008;105:4868–73, http://dx.doi.org/10.1073/pnas.0710834105 [Epub 2008 10; PMID: 18332430; PMCID: PMC2290780].
29. Mathusa EC, Chen Y, Enache E, Hontz L. Non-O157 Shiga toxin-producing *Escherichia coli* in foods. J Food Prot. 2010;73:1721–36, http://dx.doi.org/10.4315/0362-028x-73.9.1721 [PMID: 20828483].
30. McDaniel TK, Jarvis KG, Donnenberg MS, Kaper JB. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. Proc Natl Acad Sci U S A. 1995;92:1664–8, http://dx.doi.org/10.1073/pnas.92.5.1664 [PMID: 7878036; PMCID: PMC42580].
31. McDaniel TK, Kaper JB. A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. Mol Microbiol. 1997;23:399–407, http://dx.doi.org/10.1046/j.1365-2958.1997.2311591.x [PMID: 9044273].
32. Noceda LAC. Síndrome Urémico-Hemolítico por *E. coli* Entero-Hemorrágica 0157:H7 Stx2: Primer Caso Descrito en Paraguay. Pediatría Asunción. 2009;36:131–7.
33. Orth D, Ehrlenbach S, Brockmeyer J, Khan AB, Huber G, Karch H, Sarg B, Lindner H, Würzner R. EspP, a serine protease of enterohemorrhagic *Escherichia coli*, impairs complement activation by cleaving complement factors C3/C3b and C5. Infect Immun. 2010;78:4294–301, http://dx.doi.org/10.1128/IAI.00488-10 [Epub 19.07.10; PMID: 20643852; PMCID: PMC2950363].
34. Pianciola L, Chinen I, Mazzeo M, Miliwebsky E, González G, Müller C, Carbonari C, Navello M, Zitta E, Rivas M. Genotypic characterization of *Escherichia coli* O157:H7 strains that cause diarrhea and hemolytic uremic syndrome in Neuquén, Argentina. Int J Med Microbiol. 2014;304:499–504, http://dx.doi.org/10.1016/j.ijmm.2014.02.011 [Epub 18.03.14; PMID: 24702854].
35. Ribot EM, Fair MA, Gautom R, Cameron DN, Hunter SB, Swaminathan B, Barrett TJ. Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157:H7, *Salmonella*, and *Shigella* for PulseNet. Foodborne Pathog Dis. 2006;3:59–67, http://dx.doi.org/10.1089/fpd.2006.3.59 [PMID: 16602980].
36. Riordan JT, Viswanath SB, Manning SD, Whittam TS. Genetic differentiation of *Escherichia coli* O157:H7 clades associated with human disease by real-time PCR. J Clin Microbiol. 2008;46:2070–3, http://dx.doi.org/10.1128/JCM. 00203-08 [Epub 09.04.08; PMID: 18400915; PMCID: PMC2446876].
37. Scheutz F, Teel LD, Beutin L, Piérard D, Buvens G, Karch H, Mellmann A, Caprioli A, Tozzoli R, Morabito S, Strockbine NA, Melton-Celsa AR, Sanchez M, Persson S, O'Brien AD. Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. J Clin Microbiol. 2012;50:2951–63, http://dx.doi.org/10.1128/JCM.00860-12 [Epub 2012; PMID: 22760050; PMCID: PMC3421821].
38. Siegler RL, Obrig TG, Pysher TJ, Tesh VL, Denkers ND, Taylor FB. Response to Shiga toxin 1 and 2 in a baboon model of hemolytic uremic syndrome. Pediatr Nephrol. 2003;18:92–6,

http://dx.doi.org/10.1007/s00467-002-1035-7 [Epub 10.01.03; PMID: 12579394].
39. Snedeker KG, Shaw DJ, Locking ME, Prescott RJ. Primary and secondary cases in *Escherichia coli* O157 outbreaks: a statistical analysis. BMC Infect Dis. 2009;9:144, http://dx.doi.org/10.1186/1471-2334-9-144 [PMID: 19715594; PMCID: PMC2741466].
40. Stearns-Kurosawa DJ, Collins V, Freeman S, Tesh VL, Kurosawa S. Distinct physiologic and inflammatory responses elicited in baboons after challenge with Shiga toxin type 1 or 2 from enterohemorrhagic *Escherichia coli*. Infect Immun. 2010;78:2497–504, http://dx.doi.org/10.1128/IAI.01435-09 [Epub 22.03.10; PMID: 20308301; PMCID: PMC2876564].
41. Tanaro JD, Pianciola LA, D'Astek BA, Piaggio MC, Mazzeo ML, Zolezzi G, Rivas M. Virulence profile of *Escherichia coli* O157 strains isolated from surface water in cattle breeding areas. Lett Appl Microbiol. 2018;66:484–90, http://dx.doi.org/10.1111/lam.12873 [Epub 30.03.18; PMID: 29500840].
42. Tarr PI, Gordon CA, Chandler WL. Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. Lancet. 2005;365:1073–86, http://dx.doi.org/10.1016/S0140-6736(05)71144-2 [PMID: 15781103].
43. Tesh VL, Burris JA, Owens JW, Gordon VM, Wadolkowski EA, O'Brien AD, Samuel JE. Comparison of the relative toxicities of Shiga-like toxins type I and type II for mice. Infect Immun. 1993;61:3392–402, http://dx.doi.org/10.1128/iai.61.8.3392-3402.1993 [PMID: 8335369; PMCID: PMC281015].
44. Weiler N, Orrego M, Alvarez M, Huber C. Detección molecular de *Escherichia coli* diarreogénica en pacientes pediátricos con síndrome diarreico agudo en Paraguay. Mem Inst Investig Cienc Salud [Internet]. 2017;15:16–21, http://dx.doi.org/10.18004/mem.iics/1812-9528/2017.015(01)16-021. Available from: http://scielo.iics.una.py/scielo.php?script=sci_arttext&pid=S1812-95282017000100016&lng=es
45. Yang Z, Kovar J, Kim J, Nietfeldt J, Smith DR, Moxley RA, Olson ME, Fey PD, Benson AK. Identification of common subpopulations of non-sorbitol-fermenting, beta-glucuronidase-negative *Escherichia coli* O157:H7 from bovine production environments and human clinical samples. Appl Environ Microbiol. 2004;70:6846–54, http://dx.doi.org/10.1128/AEM.70.11.6846-6854.2004 [PMID: 15528552; PMCID: PMC525184].
46. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012;67:2640–4, http://dx.doi.org/10.1093/jac/dks261 [Epub 10.07.12; PMID: 22782487; PMCID: PMC3468078].
47. Zhang Y, Laing C, Steele M, Ziebell K, Johnson R, Benson AK, Taboada E, Gannon VP. Genome evolution in major *Escherichia coli* O157:H7 lineages. BMC Genomics. 2007;8:121, http://dx.doi.org/10.1186/1471-2164-8-121 [PMID: 17506902; PMCID: PMC1890555].
48. Zhou Z, Alikhan NF, Mohamed K, Fan Y, Achtman M, Agama Study Group. The EnteroBase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity. Genome Res. 2020;30:138–52, http://dx.doi.org/10.1101/gr.251678.119 [Epub 06.12.19; PMID: 31809257; PMCID: PMC6961584].