Characterisation of an emergent clone of Enteroinvasive *Escherichia coli* circulating in Europe

Running title: An emerging EIEC clone causing infections in Europe

Keywords: Enteroinvasive *Escherichia coli*, *Shigella,* outbreaks of infections, genomic characterisation, emergence of new pathogenic types

Category: Original Article – E-only

Michelacci V.[1#], Prosseda G.[2], Maugliani A.[1], Tozzoli R. [1], Sanchez S.[3], Herrera-León S.[3], Dallman T[4], Jenkins C.[4], Caprioli A.[1] and Morabito S.[1]

[1]*European Union Reference Laboratory for Escherichia coli, Department of Veterinary Public Health and Food Safety, Istituto Superiore di Sanità, Viale Regina Elena, 299, 00161, Rome Italy*

[2]*Istituto Pasteur-Fondazione Cenci Bolognetti, Department of Biology and Biotechnology "C. Darwin", Sapienza Università di Roma, Roma, Italy,*

[3]*Laboratory of Enterobacteriaceae, Service of Bacteriology, National Center of Microbiology, Instituto de Salud Carlos III, 28220, Majadahonda, Madrid, Spain*

[4]*Gastrointestinal Bacteria Reference Unit, Public Health England, 61 Colindale Ave, London NW9 5HT, UK*

[#]Corresponding author. Mailing address:

European Union Reference Laboratory for *Escherichia coli*

Department of Veterinary Public Health and Food Safety

Istituto Superiore di Sanità

Viale Regina Elena 299, 00161, Rome, Italy

Tel. +39-06-49902729; Fax +36-06-49387077

E-mail valeria.michelacci@iss.it

Abstract

Enteroinvasive *Escherichia coli* (EIEC) cause intestinal illness indistinguishable from that caused by *Shigella*, mainly in developing countries. Recently an upsurge of cases of EIEC infections has been observed in Europe, with two large outbreaks occurring in Italy and in the United Kingdom. We have characterised phenotypically and genotypically the strains responsible for these epidemics together with an additional isolate from a sporadic case isolated in Spain. The three isolates belonged to the same rare serotype O96:H19 and were of sequence type ST-99, never reported before in EIEC or *Shigella*. The EIEC strains investigated possessed all the virulence genes harboured on the large plasmid conferring the invasive phenotype to EIEC and *Shigella*, while showing only some of the known chromosomal virulence genes and none of the described pathoadaptative mutations. At the same time, they displayed motility abilities and biochemical requirements resembling more closely those of the non-pathogenic *E. coli* rather than the EIEC and *Shigella* strains used as reference.

Our observations suggested that O96:H19 strains belong to an emerging EIEC clone, which could be the result of a recent event of acquisition of the invasion plasmid by commensal *E. coli*.

Introduction

Enteroinvasive *Escherichia coli* (EIEC) are a group of pathogenic bacteria causing intestinal illness upon invasion of the human colonic mucosa [1]. The disease caused by EIEC is a bacillary dysentery with a clinical presentation indistinguishable from that caused by infection with strains of *Shigella* species, involving abdominal cramps, nausea, fever and bloody and mucus diarrhoea [2]. The pathogenesis of EIEC infection involves the colonic epithelial cell penetration preceded by the transcytosis through M cells, the lysis of

the endocytic vacuole, intracellular multiplication and extension into adjacent epithelial cells [1].

The main genes conferring the *Shigella* and EIEC invasive phenotype are harboured on a large plasmid and encode the components of a type three secretion system, including *mxi* (Membrane excretion of Ipa) and *spa* (Surface Presentation of invasion plasmid Antigens) and a number of translocated effectors, represented by the products of the genes *vir*, *ipa* (Invasion Plasmid Antigens) and *ipg* (Invasion Plasmid Genes) [3]. Several other virulence genes play accessory roles in the pathogenetic process and are differentially distributed in different *Shigella* and EIEC strains, and encode toxins, proteins interfering with the immune response of the host, factors facilitating the colonization process and iron-uptake systems favouring intracellular growth [3].

The morbidity and mortality of EIEC infections have not been fully assessed, but can be inferred from those ascribed to shigellosis. Mortality is especially high among children and it has been estimated that 99% of the appraised 165 million cases recorded annually worldwide occur in developing countries [4, 5]. The high circulation of these pathogens in low-income regions is plausibly linked to the mode of transmission of the infections, which involves the oral-faecal route. In the United States and Europe, where higher hygiene standards are in place, the subjects most often infected are travellers returning from high incidence countries, children in day care and migrant workers [6]. The information on the incidence of EIEC-associated disease is scanty, as the differentiation between these infections and those caused by *Shigella* is difficult and based on the use of multiple tests, such as the PCR targeting the *ipaH* gene, coupled with biochemical and serological typing [2]. In some cases the infections caused by *Shigella* and EIEC may be transmitted by contaminated food and water, but these appear not to be common sources of infections [7]. Historically, EIEC have not been associated with large outbreaks in industrialised countries. More commonly, EIEC causes sporadic cases afflicting specific risk groups.

However, recently an upsurge of cases of EIEC infections has been observed in Europe. In 2012, a large outbreak of bloody diarrhoea occurred among the employees of the Fire Brigade of the city of Milan, Italy [8]. The episode involved more than 100 cases of infections and the additional symptoms most commonly reported were fever, abdominal cramps and vomiting [8]. Laboratory investigations showed the presence of a positive PCR amplification of the *ipaH* gene in several stool samples and an *E. coli* strain positive for the presence of *ipaH* and belonging to serotype O96:H19 was isolated from six cases [8]. Cooked vegetables were suspected as the source for infection following a case-control study [8]. In 2013, an EIEC isolate of the same serotype was isolated from a sporadic case of traveller's diarrhoea in Spain (data not published). Finally, an outbreak of gastrointestinal disease occurred in the East Midlands in the United Kingdom (UK), involving 50 people and was suspected to be caused by the consumption of contaminated salad vegetables [9]. Again, an EIEC of serogroup O96 was isolated from some of the patients [9].

In this paper, we carried out the biochemical and phenotypic characterisation of the EIEC strains from the Italian and the UK outbreaks and from the sporadic case in Spain, as well as their whole genome sequencing. Our results show that these isolates belong to a same, emerging EIEC clone.

Materials and methods

Bacterial strains

The EIEC isolates involved in the study included the six strains EF432, EF433 and EF434 isolated in Italy in 2012 [8], H142690012 isolated in the United Kingdom 2014 [9] and CNM-2113/13, isolated from a case of severe diarrhoea occurred in Spain in 2014 (data not published). All the six EIEC strains possessed the *ipaH* gene, which is the hallmark for EIEC as well as for *Shigella* spp strains, as assessed by conventional PCR [10]. The strain

EF432 was chosen as representative of the clone that caused the Italian outbreak of infections and used in all the characterisation experiments, while the remaining two Italian isolates (EF433 and EF434) were only included in the PFGE cluster analysis.

The reference EIEC strains 6.81 and 4608 [11], the *Shigella flexneri* strain M90T [11], the *E. coli* K12 strain MG1655 [12] and the non-pathogenic human *E. coli* isolate ECOR1 part of the ECOR reference collection [13] were included in the study for comparative purposes.

Genomic characterisation of EIEC isolates

Whole genome sequencing

Whole genome sequencing of the EIEC strains was carried out to reach the coverage of at least 20X per isolate. The genomes of the EIEC strains EF432, CNM-2113/13 and 6.81 were sequenced with an Ion Torrent Personal Genome Machine (Life Technologies, Carlsbad, USA). The genome of the EF432 was sequenced in six different runs (three runs of a 200 bp library and three runs of a 400 bp library) while the genomes of CNM-2113/13 and 6.81 strains were sequenced in one and two 400 bp runs, respectively. Sequencing of the EIEC strain H142690012 was carried out by the PHE Genome Sequencing Unit using Nextera library preparation and the Illumina HiSeq 2500 in fast run mode according to manufacturers' instructions. The sequencing reads have been uploaded in the EMBL-EBI sequence database (EMBL European Nucleotide Archive accession no. ERP010762).

The reference sequences of EIEC 4608 (Acc. no. JTCO00000000), *Shigella sonnei* Ss046 (Acc. no. NC_007384, NC_007385, NC_009345, NC_009346, NC_009347), *S. boydii* CDC3083-94 (Acc. no. CP001063, CP001058, CP001059, CP001060, CP001061, CP001062), *S. dysenteriae* Sd197 (Acc. no. NC_007606, NC_007607, NC_009344) and *S. flexnerii* 2a str. 301 (Acc. no. NC_004337, NC_004851) were retrieved from GenBank database at NCBI.

Bioinformatics analysis

The sequencing reads of the EF432, CNM-2113/13, H142690012 and 6.81 isolates were assembled in contigs by using SPADES *de novo* assembling tool version 3.5.0 [14] and automatically annotated with PROKKA tool version 1.10 [15]. The two tools were operated through a local instance of the bioinformatics framework Galaxy [16].

The contigs obtained through the assembly process from the EIEC strains and the sequences of the EIEC reference strains retrieved from GenBank, were searched for the serotype-associated genes using the blastn tool present on the Galaxy against a precompiled database of reference sequences [17] provided by Dr. Flemming Scheutz at the Statens Serum Institut, Copenhagen, DK.

The Multi Locus Sequence Typing was performed according to the scheme developed by Wirth and colleagues [18] using the blastn tool to search the reference database of alleles downloaded from the University of Warwick website [19]. The combinations of alleles of the seven genes obtained through the blastn were translated into the corresponding sequence types (ST) using the online tool located at the University of Warwick website.

The phylo-group assignment was performed *in silico* by using the blastn tool to query the contigs of EIEC strains for the expected amplification products of the genes part of the scheme developed by Clermont and collegues [20].The reference database of the virulence genes of EIEC and *Shigella* strains has been compiled in house by merging the gene sequences retrieved from the VFDB website [21] with those of the genes *virB*, *virF*, *mxiE* and *ipgC* retrieved from GenBank (Gene ID: 1237991, 1238022, 876514 and 1238043, respectively), in a single multi fasta file and used with the blastn tool for the screening of the genomes for the presence of *Shigella*/EIEC virulence genes.

The presence of the described pathoadaptative mutations was investigated by progressive alignment of the annotated contigs of the three EIEC strains with the corresponding regions on the *E. coli* K12 chromosome using the software MAUVE [22] and by means of

157 phenotypic assays (API). In particular, the presence and the integrity of the genes *cadA*,
158 *cadB*, *cadC*, *speA*, *speB*, *speC*, *speD*, *speE*, *speF*, *speG*, *nadA* and *nadB* (Gene IDs:
159 85676884, 85676885, 85676886, 1789307, 1789306, 87082193, 1786311, 1786312,
160 1786909, 85675033, 16128718, 16130499, respectively) was evaluated.
161 The identification of the presence of known replicons in the whole genome sequences was
162 achieved using the PlasmidFinder tool available on the CGE webserver [23] and used as
163 an indicator of the plasmid content of the isolates.
164
165 *Pulsed Field Gel Electrophoresis*
166 The pulsed field gel electrophoresis (PFGE) was performed with XbaI enzyme as
167 previously described [24]. The similarity evaluation of PFGE profiles was performed with
168 the Bionumerics software version 7.1 (Applied Maths, Sint-Martens-Latem, Belgium) using
169 the algorithm UPGMA with tolerance and optimization set at 1.5%.
170
171 ***Phenotypic Characterization of EIEC in comparison with Shigella spp.***
172 *Growth curves*
173 Bacteria were grown statically O.N. in LB at 30°C, diluted 1:10 and evaluated
174 spectrophotometrically. Cultures were diluted to OD600 0.1 and plated on LB agar plates
175 in order to verify the equivalence between the readings and the number of viable cells in
176 the inocula. Each OD600 0.1 culture was diluted 1:100 in the same medium (LB) and five
177 0.1 ml aliquots were dispensed in 96-wells microplates and incubated at 37°C. Growth
178 was monitored by OD600 using the Wallac Victor counter (Perkin Elmer, Waltham, US-
179 MA). The data obtained were used to calculate the Gt (generation time) and the lag delay.
180 To perform the auxotrophy test for nicotinic acid, the strains were grown in M9 minimal
181 medium supplemented with 10 µg/ml thiamine, 0.2% glucose and w/o 40 µg/ml nicotinic
182 acid.

Motility assays

Motility assays (swimming) were carried out using Tryptone swim plates (1% tryptone, 0.5% NaCl, 0.3% agar). In detail single colonies of the test strains grown overnight on LB agar plates were inoculated at the surface by using a sterile needle and incubated for 16 h at 37°C. The diameters of the swimming motility zones were measured and the results recorded as mean values of three replicates.

Biochemical characterization

The EIEC strains were analysed for their biochemical profiles through the Analytical Profile Index (API) 20E test (BioMérieux SA, France) according to the manufacturer's instructions.

Results

Genomic characterisation of EIEC isolates

The draft genomes of the strains EF432, CNM-2113/13 and H142690012 and the reference EIEC strain 6.81, together with that of the EIEC strain 4608 obtained from GenBank (Acc. No. JTCO00000000) were analysed for the identification of the genes associated with the *E. coli* O and H antigens to determine or confirm the serotypes. The EIEC strains from the European outbreaks and the sporadic case showed the presence of the genes associated with the O96:H19 serotype, while the genomes of the reference strains 4608 and 6.81 revealed the presence of genes encoding the serotypes O143:H26 and O160:H26, respectively. It has to be noted that the latter strain was originally reported as belonging to serogroup O115 [11]. The three EIEC strains O96:H19 all belonged to the same sequence type ST-99, different from those of the reference EIEC strains 4608 and 6.91, respectively belonging to ST-280 and ST-279. Similarly, the three O96 strains were of B1 phylo-group, while the reference EIEC strains were all of phylo-group E.

Virulence genes asset of the EIEC strains

The presence of the virulence genes so far described in EIEC and *Shigella* was investigated in the EIEC O96:H19 strains and compared to the virulence genes asset of the two reference EIEC strains 6.81 and 4608 and with the whole genomes of the strains of *Shigella sonnei* Ss046, *S. boydii* CDC3083-94, *S. dysenteriae* Sd197, *S. flexneri* 2a str. 301. The complete results are reported in Table S1 in supplementary material.

The plasmid-borne virulence genes were present in all the EIEC strains analysed. These included those encoding the type three secretion system (TTSS), *mxi* and *spa*, as well as the *ipaA-H* and *ipgA-F* genes, encoding TTSS-secreted effectors, *icsA*, *icsB*, *icsP* and *virA*, altogether responsible for the cytoskeleton reorganisation and mobility of the bacteria inside the invaded host cells, *ospG*, whose product interferes with the innate immune responses, and the regulatory genes *virF* and *virB*, responsible for activating the transcription of all the other virulence genes (Table S1). All the plasmid genes were also present in all the *Shigella* strains assayed with the exception of *senB*, encoding a toxin involved in the early on-set of symptoms of gastrointestinal disease, which was absent from the *S. flexneri* and *S. dysenteriae* strains assayed (Table S1).

The whole set of the chromosomal *gsp* genes (*gspC-M*), encoding a type two secretion system typically present in *Shigella dysenteriae* and *S. boydii* was present in the genome of all the strains assayed but in the chromosomes of *S. sonnei* Ss046 and *S. flexneri* 2a str. 301 (Table S1). All the EIEC strains analysed were negative for the presence of the *set1A* and *set1B* genes encoding the *Shigella* enterotoxin 1 and *pic* gene, which encodes an autotransporter protease with mucinase and haemagglutinin activity involved in the colonisation process. The latter three genes occupy the same chromosomal region in the *S. flexnerii* 2a strain 301. The three EIEC O96:H19 strains, similarly to the EIEC strain 6.81, were negative for the presence of the *iucABC* and *iutA* genes encoding the

aerobactin system, playing an important role in the iron uptake especially during intracellular growth, which was instead present in EIEC strain 4608. The *sigA* gene, encoding a protease involved in intestinal fluid accumulation was absent in all the strains tested, with the exception of the reference strains EIEC 4608 and *S. sonnei* Ss46 (Table S1). Finally, the *gtr* genes associated with the lipopolysaccharide assembly could not be found in the sequences of all the EIEC and *Shigella* strains tested, with the exclusion of the *S. flexnerii* 2a strain 301 (Table S1).

Analysis of the pathoadaptative mutations

Shigella and EIEC are known to have accumulated mutations inactivating genes whose product is thought to be detrimental to the survival of the pathogen in the cellular environment. Such mutations have been hypothesized to favour the intracellular invasion and persistence. These events are defined pathoadaptative mutations and the genes involved, such as those of the *cadBA* operon, its *cadC* regulator and the *spe* genes intervening in the polyamine metabolism are termed antivirulence genes [25]. The integrity of the coding sequences of the mentioned genes has been evaluated in the EIEC strains in this study. In contrast to what is normally reported for EIEC, the three O96:H19 strains didn't show mutations in these loci. As expected, the presence of pathoadaptative mutations in these genes were confirmed for the EIEC strains 4608 and 6.81 used as references in this study. This result is in agreement with the positivity for the lysine decarboxylase activity detected through the API biochemical tests of the EIEC O96:H19. The latter strains were also negative for the presence of mutations in the *nadA*, *nadB* and *argT* genes, commonly observed in *Shigella* and EIEC.

Plasmid profiles

The plasmid profiles of the EIEC strains were determined by analysing their whole genome sequences with the PlasmidFinder tool available on the CGE webserver (https://cge.cbs.dtu.dk/services/PlasmidFinder/) [23]. The results are reported in Table 1.

All the isolates tested showed the presence of the FII replicon (Acc. No. AY458016) suggesting its correspondence with the invasion plasmid. As a matter of fact, the same replicon characterises the only EIEC virulence plasmid whose sequence is available in GenBank (Acc. No. NC_010719) as well as those of *Shigella* [23]. All the EIEC O96:H19 shared the presence of two additional replicons, ColRNAI and FIB (Acc. No. DQ298019 and AP001918, respectively) indicating the presence of as many plasmids.

The CNM-2113/13 and 6.81 strains showed the presence of two different B/O/K/Z replicons, described in plasmids harbouring beta-lactamases coding genes [23]. A replicon Q1 was identified in the sequence of the H142690012 strain only. Replicons corresponding to two different small plasmids could be identified in the genomes of CNM-2113/13, H142690012 and 6.81 strains.

Genomic correlation among the EIEC strains

Pulsed field gel electrophoresis (PFGE) was performed to estimate the correlation between the genomes of the EIEC strains included in the study. The PFGE profiles of three EIEC isolates obtained during the outbreak occurred in Italy in 2012 (EF432, EF433 and EF434), the EIEC strain isolated Spain in 2013 (CNM-2113/13) and the strain responsible for the outbreak occurred in UK in 2014 (H142690012) were produced and compared with those of the two reference EIEC strains 6.81 and 4608. The results are shown in Figure 1. As expected, the PFGE profiles of the three Italian isolates clustered with 100% identity and showed a high similarity level with the strains CNM-2113/13 and H142690012 strains, while being only distantly related to the EIEC reference strains 6.81 and 4608.

286

287 ***Phenotypic Characterization of EIEC in comparison with Shigella spp.***

288 Many EIEC share phenotypic features with strains of *Shigella*. Characteristically, they do
289 not ferment lactose [7] and lack lysine decarboxylase activity [25]. Additionally, in some
290 cases EIEC belong to O groups identical to the typical *Shigella* O-antigens [7].

291

292 *Fitness*

293 EIEC are considered an intermediate step in the transition of *E. coli* towards a full-blown
294 *Shigella* phenotype. This evolutionary process is characterised by an extensive and
295 progressive genetic decay through gene deletion and accumulation of pseudogenes. Such
296 a process involves, among the others, determinants coding for carbon sources utilization,
297 such as transporters or membrane proteins [26]. Therefore, the bacterial growth curve
298 analysis provides information about the magnitude of gene decay since the latter reduces
299 the bacterial metabolic activity.

300 Based on these considerations, we compared the growth curves of EF432, CNM-2113/13
301 and H142690012 strains, cultured in LB medium, at 37°C in microplate, with those of the
302 *E. coli* K12 laboratory strain MG1655, the EIEC strains 4608 and 6.81, the *S. flexneri*
303 strain M90T and the *E. coli* natural isolate ECOR1 used as reference isolates. In particular,
304 we analysed the Generation time (Gt), the stationary and the lag phase. The growth
305 analysis showed no significant differences in the Gt of the three EIEC investigated and the
306 EIEC 6.81 and ECOR1 strains used for comparison (from 23.8 to 23.9 minutes). The EIEC
307 strain 4608 and *E. coli* K12 MG1655 showed similar Gt (25.2 and 24.8 minutes) while the
308 highest value characterised the growth of the *Shigella* strain M90T (26.2 minutes) (Table
309 S2). Furthermore, the A_{600} of each strain after prolonged incubation (16 h), i.e. stationary
310 phase, were also considered and no significant differences were found. Finally, the lag
311 phase analysis evidenced that the ECOR1 natural isolate displayed the promptest

adaptation to fresh medium followed by the three EIEC subject of this study (4 minute delay) (Figure 2) and the EIEC reference strains 4608 and 6.81 (13 minute delay). Finally, the *E. coli* K12 MG1655 and *S. flexneri* M90T strains showed the longest lag phase before the cultures entered the exponential growth (20 and 43 minutes, respectively). These results suggest that strains EF432, CNM-2113/13 and H142690012 have a high metabolic versatility and could be more efficient at exploiting the available nutritional elements in the medium than the *E. coli* K12 strain and the reference EIEC and *S. flexneri* strains. This consideration was further confirmed by the analysis of the *nic* phenotype, assessed by the ability to grow in the absence of nicotinic acid and determined by the presence of functional *nad* genes [27]. The *nic*$^{-}$ phenotype is considered a final step characteristic in the evolutionary route of EIEC towards *Shigella*, since Shigella typically lacks a de novo pathway for the synthesis of nicotinamide adenine dinucleotide and requires nicotinic acid for growth in minimal medium while only some EIEC strains have this requirement [27]. Strains EF432, CNM-2113/13 and H142690012 grow on minimal medium without nicotinic acid indicating that *nad* genes are functional and correctly expressed in all the three EIEC strains analysed.

Biochemical characterisation

The three EIEC strain assayed EF432, CNM-2113/13 and H142690012 shared the same biochemical profile when tested through the Analytical Profile Index (API) 20E test. In particular, they were positive for the β-galactosidase and lysine decarboxylase (LDH) activity and were able to ferment glucose, mannose, sorbitol, rhamnose, sucrose, melibiose and arabinose. The observed phenotypes resulted in the API code 5004572, identical for all the three strains, which corresponded to a good identification of the *Escherichia coli* species (id=90.7%).

Motility test

The lack of motility constitutes important taxonomic and diagnostic criteria used to differentiate *Shigella* from other members of the *Enterobacteriaceae*, being the former non-motile, similarly to the majority of EIEC strains. In order to analyse the motility phenotype of the EIEC O96:H19, we performed a swim assay using the EF432, CNM-2113/13 and H142690012 strains and included the MG1655, 4608, 6.81, M90T and ECOR1 as reference strains. We inoculated the swim plates and, after 16 hours of incubation, no swimming was observed for M90T, 4608, 6.81 and H142690012, while MG1655 showed the strongest ability to swim (33 mm diameter) (Figure 3). ECOR1 showed an intermediate level of mobility (15 mm diameter) while EF432 and CNM-2113/13 showed residual mobility (5.5 and 4.1 mm) indicating a partial but significant impairment of flagella biosynthesis and functionality.

Discussion

EIEC infections share with those caused by bacteria belonging to genus *Shigella* the pathophysiological features and modes of transmission [7]. While the burden of EIEC infections is not known mainly due to mis-diagnosis, most of what we know derives from studies done on the prevalence of *Shigella* infections among the populations living in underdeveloped countries, where they cause millions of human cases of disease with more than one million deaths, yearly [4].

In addition to the heavy toll paid by low-income countries, *Shigella* infection is the third most common cause of bacterial gastroenteritis in the United States, after *Salmonella* and *Campylobacter* infections and followed by *E. coli* O157, Vibrio and Yersinia [28, 29].

No immunity to shigellosis has been described for specific groups, but certain individuals are at increased risk of getting infected. Children below five acquire *Shigella* infection at

363 the highest rate [30] and other immunocompromised groups, such as persons infected
364 with HIV, suffer from *Shigella* infections much more commonly than other individuals [31].
365 Insights into the phylogeny of EIEC and *Shigella* have been obtained from studies based
366 on the analysis of the variations accumulated in the DNA sequence of housekeeping
367 genes [18, 32]. Such studies provide evidences that these pathogens should be
368 considered as two different steps of the same evolutionary pathway rather than two
369 separated bacterial species [2, 18, 32]. It is likely that EIEC are precursors of *Shigella,*
370 which could have emerged following acquisition of the plasmid containing the invasion
371 genes [32], by a commensal *E. coli*, eventually transforming into the strains known as
372 *Shigellae* following the accumulation of pathoadaptative mutations that facilitated the
373 intracellular lifestyle [25]. In particular, according to the evolutionary relationships between
374 the two species, the strains of *Shigella* seem to be derived from *E. coli* belonging to the
375 ST-280 clonal complex (Cplx) [18].
376 In industrialised countries the information on EIEC disease are scanty. These infections
377 are believed to be mostly imported from developing countries and with limited possibility to
378 spread due to their inter-human transmission cycle [6]. Recently, a study was published
379 analysing the frequency of enteric pathogens in stool samples of patients affected with
380 acute gastroenteritis throughout Europe, reporting a very low prevalence for EIEC or
381 *Shigella* (10 positive out of 709 samples analysed) [33]. Nevertheless, recently an upsurge
382 in cases of EIEC infections in the European Union has been observed [8, 9]. Two large
383 outbreaks have been reported, in years 2012 and 2014, in Italy [8] and the UK [9],
384 respectively, both involving a high number of cases. It is noteworthy that the two episodes
385 have been linked to the consumption of contaminated food [8, 9]. In addition to these two
386 outbreaks, one sporadic case has been registered in Spain in 2013. We have
387 characterised a set of three EIEC strains isolated from the two outbreaks and the sporadic
388 case in Spain and found that all of them belonged to serotype O96:H19, a rare serotype

first attributed to the EIEC strain that caused the Italian outbreak in 2012 [8]. Additionally, we have found that the three EIEC all belonged to the ST-99 that in turn does not belong to any known ST Cplx and it is only distantly related to the ST-280 Cplx, having only one of the alleles of the scheme in common.

These observations, together with the observed difference in the phylo-group of EIEC O96, when compared to that of the EIEC reference strains, suggested that the EIEC O96:H19 could be the result of a recent event of acquisition of the invasion plasmid. This was confirmed by the analysis of the virulence genes content and pathoadaptative mutation pattern. The whole genomes of the three isolates was determined and compared with a database containing all the known virulence gene of *Shigella*/EIEC. The three EIEC O96:H19 showed the expected pattern of virulence genes displaying the complete plasmid-borne virulome of the EIEC and *Shigella* strains used as controls (Table S1). Nevertheless, the genome analysis showed that they had not acquired any of the described pathoadaptative mutations.

In order to support the hypothesis that the three EIEC O96:H19 were closer to *E. coli* than to *Shigella*, we have characterised them phenotypically with respect to their ability to move and diffuse into agar layers. As expected, some residual movement was observed in two out of the three EIEC O96:H19 in comparison to all the EIEC and *Shigella* control strains that were non-motile. Also, the observed prototrophic behaviour of the EIEC O96:H19 towards the use of nicotinic acid for growth seems to confirm this hypothesis. The auxotrophism for this compound is an evolutionary feature in *Shigella*, and is partially present in the EIEC strains described so far [27]. Even more strikingly, the three isolates showed a more rapid entry into the log phase than the *Shigella* and EIEC controls, but very similar to how observed with the *E. coli* natural isolate ECOR1 strain included in the experiments (Figure 2). This latter feature is interesting and may, at least partially, explain the described association of the two outbreaks with a food vehicle. The Italian episode

was epidemiologically linked to the consumption of cooked vegetables [8], while the outbreak occurred in the UK was reported as being caused by the consumption of salad [9]. In both cases the superior ability of the EIEC O96:H19 to grow with an "*E. coli* style" may have been driving the overgrowth in the food vehicle and could explain the size of the two episodes. In fact none of the known EIEC or *Shigella* would have had the possibility to reach a bacterial load able to cause such a high number of people affected.

In conclusion, our findings support the hypothesis that the EIEC O96:H19 that caused sporadic cases and outbreaks of infections in the EU belong to an emergent clone never described previously. The genomic characterisation of these isolates, carried out by PFGE, further confirms this hypothesis (Figure 1). Additionally, our results seem to indicate that EIEC may emerge as the result of the continuous acquisition, by commensal *E. coli*, of the plasmid containing the invasion determinants and that, under certain circumstances, as in the case of the *E. coli* strains of ST-280 Cplx, this process progresses with the accumulation of pathoadaptative mutations and the loss of some of the *E. coli* characteristics, such as the rapid entry into the exponential growth phase, more adapted to an out-of-the-cell lifestyle, eventually resulting in the emergence of *Shigellae* (Figure 4).

The hypothesis of a stepwise model for the emergence of *Shigella* from *E. coli* and the possibility that the process may reiterate from time to time could explain the emergence and the sudden appearance of the highly virulent EIEC O96:H19 clone and paves the way to the possible emergence of new EIEC clones causing outbreaks of infections in the future.

References

1 Nataro JP, Kaper JB. Diarrheagenic escherichia coli. *Clinical microbiology reviews*. 1998; **11**: 142-201.

2 van den Beld MJ, Reubsaet FA. Differentiation between shigella, enteroinvasive escherichia coli (eiec) and noninvasive escherichia coli. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology*. 2012; **31**: 899-904.

3 Sansonetti PJ. Rupture, invasion and inflammatory destruction of the intestinal barrier by shigella, making sense of prokaryote-eukaryote cross-talks. *FEMS microbiology reviews*. 2001; **25**: 3-14.

4 Kotloff KL, Winickoff JP, Ivanoff B, et al. Global burden of shigella infections: Implications for vaccine development and implementation of control strategies. *Bulletin of the World Health Organization*. 1999; **77**: 651-666.

5 Vieira N, Bates SJ, Solberg OD, et al. High prevalence of enteroinvasive escherichia coli isolated in a remote region of northern coastal ecuador. *The American journal of tropical medicine and hygiene*. 2007; **76**: 528-533.

6 Niyogi SK. Shigellosis. *Journal of microbiology*. 2005; **43**: 133-143.

7 Tozzoli R, Scheutz, F. Diarrheagenic escherichia coli infections in humans. In: Stefano M, ed. *Pathogenic escherichia coli: Molecular and cellular microbiology*: Caister Academic Press 2014.

8 Escher M, Scavia G, Morabito S, et al. A severe foodborne outbreak of diarrhoea linked to a canteen in italy caused by enteroinvasive escherichia coli, an uncommon agent. *Epidemiology and infection*. 2014; **142**: 2559-2566.

9 PHE PHE. Foodborne illness in humans: General outbreaks in england and wales in 2014. London, UK: Public Health England 2014.

10 Luscher D, Altwegg M. Detection of shigellae, enteroinvasive and enterotoxigenic escherichia coli using the polymerase chain reaction (pcr) in patients returning from tropical countries. *Molecular and cellular probes*. 1994; **8**: 285-290.

11 Sansonetti PJ, d'Hauteville H, Formal SB, Toucas M. Plasmid-mediated invasiveness of "shigella-like" escherichia coli. *Annales de microbiologie*. 1982; **133**: 351-355.

12 Guyer MS, Reed RR, Steitz JA, Low KB. Identification of a sex-factor-affinity site in e. Coli as gamma delta. *Cold Spring Harbor symposia on quantitative biology*. 1981; **45 Pt 1**: 135-140.

13 Ochman H, Selander RK. Standard reference strains of escherichia coli from natural populations. *Journal of bacteriology*. 1984; **157**: 690-693.

14 Bankevich A, Nurk S, Antipov D, et al. Spades: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology : a journal of computational molecular cell biology*. 2012; **19**: 455-477.

15 Seemann T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*. 2014; **30**: 2068-2069.

16 Goecks J, Nekrutenko A, Taylor J, Galaxy T. Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome biology*. 2010; **11**: R86.

17 Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F. Rapid and easy in silico serotyping of escherichia coli using whole genome sequencing (wgs) data. *Journal of clinical microbiology*. 2015.

18 Wirth T, Falush D, Lan R, et al. Sex and virulence in escherichia coli: An evolutionary perspective. *Mol Microbiol*. 2006; **60**: 1136-1151.

19 Achtman M, Velayudhan V., Zhou Z. *Mlst databases at university of warwick.* http://mlst.warwick.ac.uk/mlst/dbs/Ecoli

20 Clermont O, Christenson JK, Denamur E, Gordon DM. The clermont escherichia coli phylo-typing method revisited: Improvement of specificity and detection of new phylo-groups. *Environmental microbiology reports*. 2013; **5**: 58-65.

21 MOH Key Laboratory of Systems Biology of Pathogens IoPB, CAMS&PUMC, Bejing, China. http://www.mgc.ac.cn/cgi-bin/VFs/genus.cgi?Genus=Shigella
22 Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. Reordering contigs of draft genomes using the mauve aligner. *Bioinformatics*. 2009; **25**: 2071-2073.
23 Carattoli A, Zankari E, Garcia-Fernandez A, et al. In silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrobial agents and chemotherapy*. 2014; **58**: 3895-3903.
24 EFSA. Molecular typing of verocytotoxin-producing e. Coli (vtec) strains isolated from food, feed and animals: State of play and standard operating procedures for pulsed field gel electrophoresis (pfge)typing, profiles interpretation and curation. *wwwefsaeuropaeu/publications*. 2014; **European Food Safety Authority supporting publication 2014:EN-704**.
25 Prosseda G, Di Martino ML, Campilongo R, et al. Shedding of genes that interfere with the pathogenic lifestyle: The shigella model. *Research in microbiology*. 2012; **163**: 399-406.
26 Feng Y, Chen Z, Liu SL. Gene decay in shigella as an incipient stage of host-adaptation. *PloS one*. 2011; **6**: e27754.
27 Di Martino ML, Fioravanti R, Barbabella G, Prosseda G, Colonna B, Casalino M. Molecular evolution of the nicotinic acid requirement within the shigella/eiec pathotype. *Int J Med Microbiol*. 2013; **303**: 651-661.
28 Gupta A, Polyak CS, Bishop RD, Sobel J, Mintz ED. Laboratory-confirmed shigellosis in the united states, 1989-2002: Epidemiologic trends and patterns. *Clin Infect Dis*. 2004; **38**: 1372-1377.
29 Iwamoto M, Huang JY, Cronquist AB, et al. Bacterial enteric infections detected by culture-independent diagnostic tests--foodnet, united states, 2012-2014. *MMWR Morbidity and mortality weekly report*. 2015; **64**: 252-257.
30 Mohle-Boetani JC, Stapleton M, Finger R, et al. Communitywide shigellosis: Control of an outbreak and risk factors in child day-care centers. *Am J Public Health*. 1995; **85**: 812-816.
31 Baer JT, Vugia DJ, Reingold AL, Aragon T, Angulo FJ, Bradford WZ. Hiv infection as a risk factor for shigellosis. *Emerg Infect Dis*. 1999; **5**: 820-823.
32 Lan R, Alles MC, Donohoe K, Martinez MB, Reeves PR. Molecular evolutionary relationships of enteroinvasive escherichia coli and shigella spp. *Infect Immun*. 2004; **72**: 5080-5088.
33 Spina A, Kerr KG, Cormican M, et al. Spectrum of enteropathogens detected by the filmarray gi panel in a multicentre study of community-acquired gastroenteritis. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*. 2015; **21**: 719-728.