

# Acquisition of a large virulence plasmid (pINV) promoted temperature-dependent virulence and global dispersal of O96:H19 enteroinvasive *Escherichia coli*

Sydney L. Miles,<sup>1</sup> Vincenzo Torraca,<sup>1</sup> Zoe A. Dyson,<sup>1,2,3</sup> Ana Teresa López-Jiménez,<sup>1</sup> Ebenezer Foster-Nyarko,<sup>1</sup> Damián Lobato-Márquez,<sup>1</sup> Claire Jenkins,<sup>4</sup> Kathryn E. Holt,<sup>1,2</sup> Serge Mostowy<sup>1</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 8.

**ABSTRACT** Enteroinvasive *Escherichia coli* (EIEC) and *Shigella* are closely related agents of bacillary dysentery. It is widely viewed that EIEC and *Shigella* species evolved from *E. coli* via independent acquisitions of a large virulence plasmid (pINV) encoding a type 3 secretion system (T3SS). Sequence Type (ST)99 O96:H19 *E. coli* is a novel clone of EIEC responsible for recent outbreaks in Europe and South America. Here, we use 92 whole genome sequences to reconstruct a dated phylogeny of ST99 *E. coli*, revealing distinct phylogenomic clusters of pINV-positive and -negative isolates. To study the impact of pINV acquisition on the virulence of this clone, we developed an EIEC-zebrafish infection model showing that virulence of ST99 EIEC is thermoregulated. Strikingly, zebrafish infection using a T3SS-deficient ST99 EIEC strain and the oldest available pINV-negative isolate reveals a separate, temperature-independent mechanism of virulence, indicating that ST99 non-EIEC strains were virulent before pINV acquisition. Taken together, these results suggest that an already pathogenic *E. coli* acquired pINV and that virulence of ST99 isolates became thermoregulated once pINV was acquired.

**IMPORTANCE** Enteroinvasive *Escherichia coli* (EIEC) and *Shigella* are etiological agents of bacillary dysentery. Sequence Type (ST)99 is a clone of EIEC hypothesized to cause human disease by the recent acquisition of pINV, a large plasmid encoding a type 3 secretion system (T3SS) that confers the ability to invade human cells. Using Bayesian analysis and zebrafish larvae infection, we show that the virulence of ST99 EIEC isolates is highly dependent on temperature, while T3SS-deficient isolates encode a separate temperature-independent mechanism of virulence. These results indicate that ST99 non-EIEC isolates may have been virulent before pINV acquisition and highlight an important role of pINV acquisition in the dispersal of ST99 EIEC in humans, allowing wider dissemination across Europe and South America.

**KEYWORDS** EIEC, zebrafish, host-pathogen interactions, evolution, *Shigella*, *Enterobacteriaceae*, virulence determinants

Enteroinvasive *E. coli* (EIEC) and *Shigella* species are Gram-negative, human-adapted pathogens that cause bacillary dysentery. The greatest burden of bacillary dysentery is in low- and middle-income countries (LMICs) (1), although the true burden of EIEC infection is likely underestimated since it is difficult to distinguish from *Shigella*. Historically, *Shigella* was classified as its own genus, with four distinct species, but Multi-Locus Sequence Typing (MLST) and whole-genome sequencing data clearly show *Shigella* spp. are lineages of *E. coli*, as are EIEC (2, 3). Each *Shigella* and EIEC lineage evolved independently within the *E. coli* population, following the horizontal acquisition of a ~220 kbp virulence plasmid (also known as plasmid of invasion or pINV) from a

**Editor** Carmen Buchrieser, Institut Pasteur, Paris, France

Address correspondence to Kathryn E. Holt, kat.holt@shttm.ac.uk, or Serge Mostowy, serge.mostowy@shttm.ac.uk.

The authors declare no conflict of interest.

See the funding table on p. 9.

**Received** 10 April 2023

**Accepted** 17 April 2023

**Published** 31 May 2023

Copyright © 2023 Miles et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

currently unknown source (2). pINV encodes a type three secretion system (T3SS) that facilitates the invasion of human epithelial cells and is thermoregulated in both EIEC and *Shigella* (4).

A novel clone of EIEC, of serotype O96:H19 and Multi-Locus Sequence Type (ST) 99, was first described in 2012 in Italy and has since caused several foodborne outbreaks of moderate to severe diarrheal disease across Europe and South America (5–7). Before 2012, ST99 *E. coli* had not been reported in the literature as causing human disease but had been sporadically isolated from cattle and environmental sources (8). ST99 EIEC isolates have been characterized as possessing the virulence hallmarks of EIEC and *Shigella* (pINV and T3SS) (9), but its metabolic capacity closely resembles that of commensal *E. coli* and it has more recently been associated with *pga*-mediated biofilm formation (6, 9). It has therefore been proposed that ST99 EIEC diverged recently from ST99 *E. coli* due to the acquisition of pINV.

The zebrafish (*Danio rerio*) larvae model is widely used to study infection biology *in vivo* because of its rapid development and innate immune system that is highly homologous to that of humans (10, 11). Zebrafish have emerged as a valuable vertebrate model to study human enteropathogens like *Shigella* (12), highlighting the key roles of bacterial virulence factors (e.g., T3SS and O-antigen) (13, 14) and cell-autonomous immunity (e.g., autophagy and septin-mediated immunity) (12, 15) in host-pathogen interactions.

In this observation, we reconstruct a dated phylogeny of ST99 *E. coli* using publicly available whole genome sequences, to understand the role of pINV in its global dispersal. We develop a temperature-dependent zebrafish infection model to assess the virulence of EIEC and non-EIEC ST99 isolates, highlighting the power of zebrafish infection in studying the evolution of novel enteropathogens causing disease in humans.

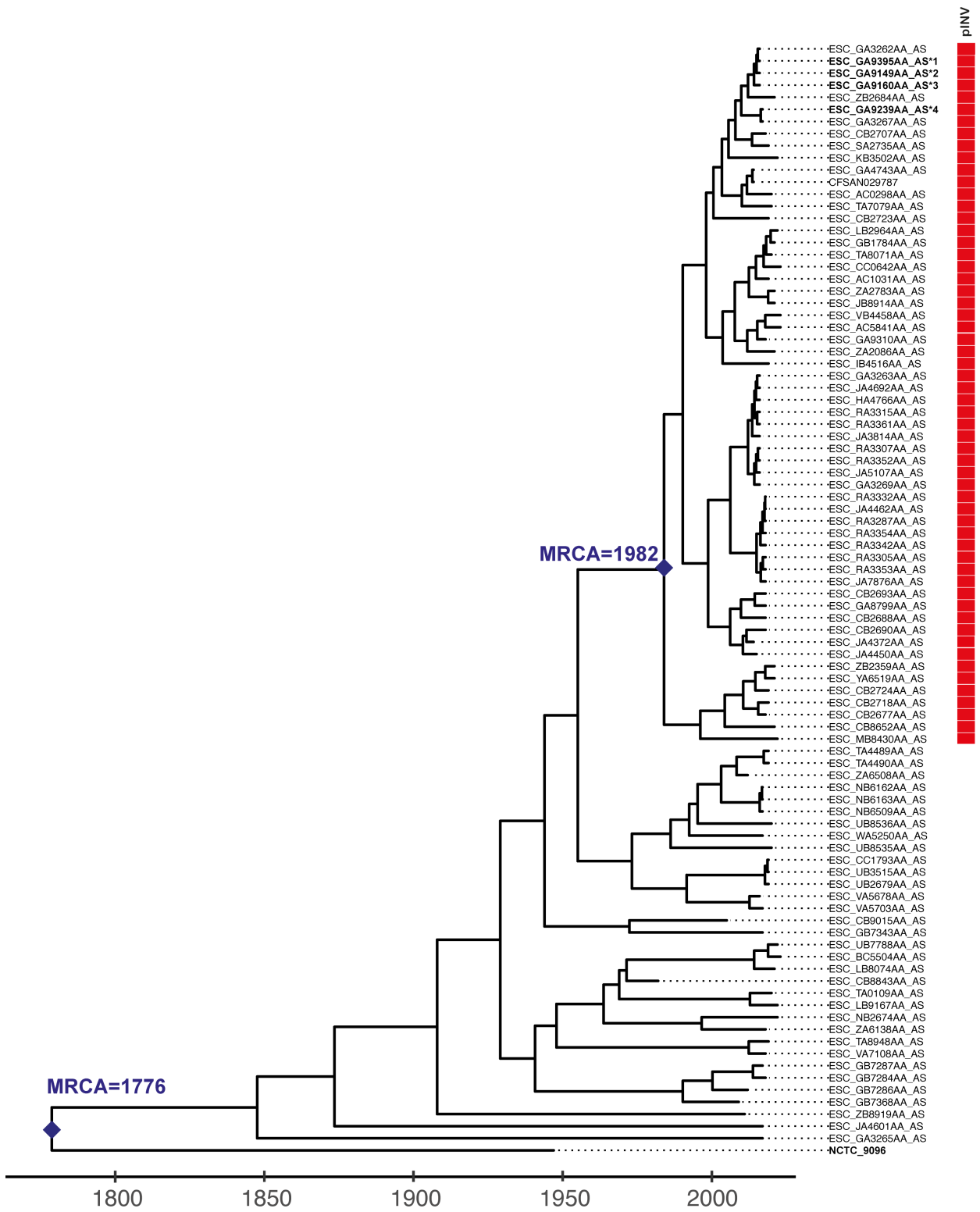
### ST99 EIEC diverged ~40 years ago

To dissect the evolution of the ST99 clone and its transition to EIEC, we analyzed all publicly available ST99 genomes ( $n = 92$ ), using the Enterobase integrated software environment (16). Enterobase routinely scans short-read archives and retrieves *E. coli* and *Shigella* sequences from the public domain or uses user-uploaded short reads. We used Gubbins v.3.2.1 (17) to filter recombinant sites, RaxML v.8.10 to infer a Maximum Likelihood phylogenetic tree and BactDating v.1.2 (18) to date the phylogeny (Fig. 1), as previously described by Didelot and Parkhill (19). Root-to-tip genetic distances were positively associated with the year of isolation ( $R^2 = 0.19$ ,  $P = 6 \times 10^{-3}$ ), and the date-randomization test showed no overlap between results of observed and date-randomized analyses (Fig. S1), indicating a moderate molecular clock signal to support dating analysis. From this analysis, we estimate that the most recent common ancestor (MRCA) of the whole ST99 group (pINV+ and pINV-) existed circa 1776 [95% highest posterior density (HPD), 1360–1927]. To test for the presence of pINV, we used ShigEiFinder, which scans the genomes for pINV-encoded genes and deems an isolate positive for pINV when 26 of 38 genes are present (20). The pINV+ isolates form a distinct cluster, with their MRCA existing circa 1982 (95% HPD, 1965–2011) (Fig. 1). This suggests that the ST99 EIEC may have been circulating undetected for ~30 years before being detected in the 2012 outbreak.

To test the role of pINV in the dispersal of ST99 EIEC, we selected: (i) four recent pINV+ isolates from moderate-to-severe diarrheal outbreaks in the United Kingdom in 2014 and 2015 (21, 22) to represent contemporary ST99 EIEC, (ii) a Congo red-negative colony to represent a T3SS-deficient strain isogenic strain, and (iii) the oldest available ST99 isolate (~1945, NCTC 9096, pINV-) to represent ancestral pINV- ST99 (see Fig. 1).

### ST99 EIEC virulence is temperature-dependent in zebrafish

The zebrafish infection model has generated fundamental advances in our understanding of *Shigella* and its ability to infect humans (23). To test the virulence of pINV+ ST99 EIEC strains, ~5,000 CFU was injected into the hindbrain ventricle (HBV) of zebrafish



**FIG 1** Time-calibrated phylogeny of 92 Sequence Type (ST)99 genomes. BactDating was used to infer a time-calibrated phylogeny, incorporating the output from the recombination detection software, Gubbins. Blue diamonds indicate the internal nodes representing the most recent common ancestors (MRCA) of interest. Tip labels represent assembly barcodes correlating to the isolate accession in Enterobase. Tip labels in bold represent isolates that we tested *in vivo*. As determined using ShigEiFinder (20), the presence of the invasion plasmid (pINV) is indicated by a red box in the pINV column. We estimate the MRCA of the whole group to be ~1776 and the MRCA for the pINV+ cluster to be ~1982.

larvae at 3 d post-fertilization (dpf) (Fig. S1A). Infected zebrafish larvae are typically incubated at 28.5°C for optimal development but we have shown they can also be maintained at 32.5°C (13), allowing the study of temperature-dependent virulence. For the pINV+ strains, we observed ~75% survival when larvae were incubated at 28.5°C but only ~30% survival when incubated at 32.5°C (Fig. 2A; Fig. S2B and C). In agreement with survival results, CFUs recovered at 6 h post-infection (hpi) were significantly lower at 28.5°C than CFUs recovered at 32.5°C (Fig. 2B; Fig. S2D and E), suggesting that larvae were more able to control infection at 28.5°C.

To test if the T3SS in ST99 EIEC is functional and thermoregulated, we compared the secretion of virulence factors by ST99 EIEC and *Shigella flexneri* *in vitro* (Fig. S3). The overall abundance of secreted proteins is lower for ST99 EIEC as compared to *S. flexneri*, but the relative abundance of major secreted effectors appears similar. One exception is SepA, a protein secreted independently of the T3SS, whose presence is known to be variable in other EIEC lineages (24). Although we do not observe significant differences in secretion between 28.5°C and 32.5°C under these *in vitro* conditions tested, the T3SS in ST99 EIEC is clearly thermoregulated (with optimal secretion *in vitro* at 37°C).

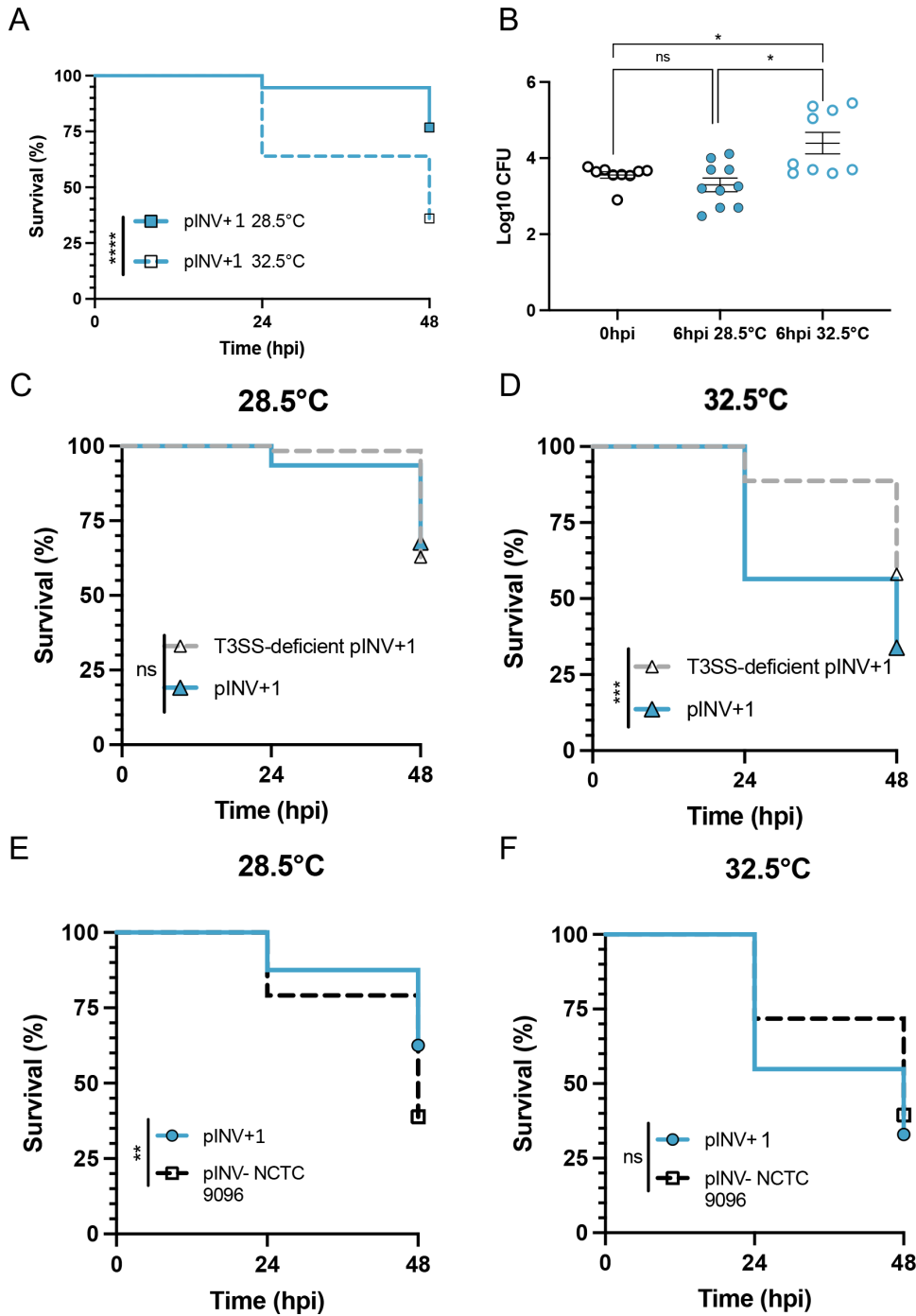
Having established a temperature-dependent EIEC-zebrafish infection model, it was next of great interest to test the virulence of an isogenic, T3SS-deficient ST99 EIEC strain and an ancestral pINV– ST99 isolate. Since we observed no significant differences in zebrafish survival or bacterial burden between the four pINV+ strains at either 28.5°C or 32.5°C (Fig. S2B to E), we chose one isolate (pINV+1) as a representative pINV+ isolate to compare with the T3SS-deficient and pINV– isolate (NCTC 9096).

### ST99 *E. coli* comprises temperature-dependent and -independent mechanisms of virulence

To test if the virulence of ST99 *E. coli* in zebrafish is dependent on the acquisition of pINV and the T3SS, we selected a naturally T3SS-deficient colony (Congo red negative) to compare against pINV+1. Colonies were screened for several pINV-encoded genes by colony PCR and found to be deficient in genes located in the T3SS-encoding region of pINV (*mxlG*, *mxlD*, and *icsB*), but positive for genes located outside (*ospF* and *ipaH*) (Fig. S4). In addition, we verified dysfunction of the T3SS, showing that T3SS effector proteins are not secreted by T3SS-deficient colonies at 37°C but are by the wild-type EIEC isolate (Fig. S3). These results suggest that the T3SS-encoding region has been lost in Congo red negative colonies, consistent with what has previously been reported for *S. flexneri* (25). Infection of zebrafish with these isolates shows that the thermoregulated virulence is lost in the T3SS-deficient strain, with no significant difference in zebrafish survival observed between 28.5°C and 32.5°C, whilst thermoregulated virulence is maintained in the wild-type strain (Fig. 2C and D; Fig. S5A and B). These results implicate acquisition of the T3SS (and pINV) in the temperature-dependent virulence of pINV+1.

Next, we compared the virulence of a non-EIEC (pINV–) ST99 isolate (NCTC 9096) and a pINV+ EIEC isolate (pINV+1) strains using our EIEC-zebrafish infection model. Strikingly, NCTC 9096 was significantly more virulent than pINV+1 at 28.5°C, with only ~35% of infected larvae surviving at 48 hpi (Fig. 2E). Although no change in survival of larvae infected with NCTC 9096 is observed at 32.5°C (as compared to that of 28.5°C), survival of pINV+1 infected larvae significantly decrease at 32.5°C, consistent with a role for temperature-dependent virulence. At 32.5°C, we found that both pINV+1 and NCTC 9096 isolates were equally virulent (Fig. 2F).

The trend in virulence was also reflected in the quantification of bacterial burden (Fig. S5C and D). When incubated at 32.5°C, we observed a ~2 log increase in pINV+1 CFUs enumerated from larvae at 6 hpi but not when incubated at 28.5°C. We observe a similar increase in NCTC 9096 CFUs quantified, irrespective of temperature. These results show temperature-dependent virulence of the pINV+1 strain and non-temperature-dependent virulence of the pINV– strain, consistent with our observations for the T3SS-deficient EIEC isolate.



**FIG 2** Temperature-dependent and -independent mechanisms of virulence in the ST99 group. Zebrafish larvae at 3 d post-fertilization were injected with 5,000 CFU of a representative pINV+ ST99 strain, a T3SS-deficient strain and an ancestral pINV- ST99 strain, before being separated for incubation at 28.5°C or 32.5°C. (A, B) pINV+ strain exhibits a temperature-dependent virulence with significantly more killing observed at 32.5°C. Enumeration of bacterial burden is also temperature dependent, with greater CFUs quantified at 6 h post-infection from larvae incubated at 32.5°C. Black circles indicate pINV+1 CFUs at 0 hpi, blue filled circles indicate pINV+1 CFUs at 6 hpi incubated at 28.5°C, and blue outlined circles indicate pINV+1 CFUs at 6 hpi incubated at 32.5°C. (C, D) Thermoregulated virulence is lost in a T3SS-deficient (Congo red negative) pINV+1 strain (gray dashed line). (E, F) pINV- strain NCTC 9096 (black dashed line) is virulent in the zebrafish model in a non-temperature-dependent manner. Significance was tested using Log-rank (Mantel-Cox) test for survival curves. For CFUs (panel B), significance was tested using a one-way ANOVA with Sidak's correction. \* $P < 0.0332$ ; \*\* $P < 0.0021$ ; \*\*\* $P < 0.0002$ ; and \*\*\*\* $P < 0.0001$ .

## DISCUSSION

It is widely recognized that the acquisition of pINV is a defining feature in the evolution of EIEC and *Shigella* (26). Here, we analyze the evolution of ST99 EIEC and propose that an MRCA for the pINV+ group existed in the early 1980s. This suggests that ST99 EIEC may have been circulating undetected for ~30 years until it was implicated in the 2012 outbreak in Italy, perhaps because EIEC infections are typically endemic in regions where surveillance and sequencing of enteropathogens are limited.

We prove that the virulence of ST99 EIEC strains is thermoregulated *in vitro* and *in vivo* (with zebrafish larvae less able to control infection), leading to increased killing and greater bacterial replication at 32.5°C. Some killing is still observed at 28.5°C, suggesting a low-level activation of the T3SS and/or non-T3SS mechanisms of virulence *in vivo*, which would be of interest to test in future studies. These data are consistent with previous reports for pINV-mediated virulence in both *S. flexneri* and *Shigella sonnei* (4, 13). Our zebrafish infection model highlights the importance of temperature in EIEC virulence and supports the hypothesis that pINV acquisition is the first key step in the evolutionary pathway toward becoming a human-adapted pathogen. Our data using zebrafish infection further show that non-EIEC ST99 isolates can also cause disease and that the ability of the ST99 clone to cause disease does not strictly rely on the acquisition of pINV and the transition to EIEC. Considering that pINV– ST99 strain NCTC 9096 is highly virulent *in vivo*, we conclude that it must encode separate, non-thermoregulated mechanism(s) of virulence that becomes less important for human infection once pINV is acquired.

Collectively, our findings illuminate the short history of ST99 EIEC and implicate pINV acquisition as a key factor in its epidemiological success. Our approach also reveals a separate, non-thermoregulated virulence mechanism in a pINV– ST99 isolate, suggesting that an already pathogenic *E. coli* may have acquired pINV. Further studies, including identifying the source of pINV and those isolates likely to acquire it, are important to fully understand and prevent the dispersal of novel EIEC and *Shigella* clones infecting humans.

## METHODS

### Bacterial strains

Four pINV+ EIEC strains isolated in diarrhoeal outbreaks from the United Kingdom in 2014/2015 were included in this study (Table 1), and strains were identified and sequenced through routine surveillance and kindly shared with us by the UK Health and Security Agency (UKHSA). A pINV– ST99 strain (Table 1) included in this study was obtained from the National Culture Type Collection (NCTC). *S. flexneri* M90T was used as a positive control for the *in vitro* secretion assay (27).

TABLE 1 Bacterial strains used in experimental work/as a reference genome<sup>a</sup>

Strain	Source	Serotype	pINV	Origin	Sequence accession no.	Enterobase assembly barcode
NCTC 9096	NCTC	O96:H19	–	Denmark, 1945	UGEL01000000	ESC_CC4859AA_AS
pINV+1	UKHSA	O96:H19	+	United Kingdom (Travel to Turkey), 2014	SRR3578973	ESC_GA9395AA_AS
pINV+2	UKHSA	O96:H19	+	Kingdom (Travel to Turkey), 2014	SRR3578582	ESC_GA9149AA_AS
pINV+3	UKHSA	O96:H19	+	Kingdom (Travel to Turkey), 2014	SRR3578593	ESC_GA9160AA_AS
pINV+4	UKHSA	O96:H19	+	United Kingdom, 2015	SRR3578770	ESC_GA9239AA_AS
CFSAN029787	NA	O96:H19	+	Italy, 2012	Chromosome: CP011416.1, pINV: CP011417.1	ESC_GA4743AA_AS
<i>S. flexneri</i> M90T	Institut Pasteur	5 a	+	Mexico, 1955	NA <sup>b</sup>	NA <sup>b</sup>

<sup>a</sup>Strains NCTC 9096 and the four pINV+ ST99 strains obtained from the UKHSA were used for the *in vivo* work. CFSAN029787 was used as a reference strain for the phylogenomic analyses. Enterobase assembly accessions correlate to tip labels in the dated phylogeny (Fig. 1).

<sup>b</sup>NA, not applicable.



To obtain a T3SS-deficient ST99 isolate, bacteria (pINV+1) were grown on trypticase soy agar (TSA) plates supplemented with 0.01% Congo red (Sigma-Aldrich) dye. A white colony was selected as a natural isogenic mutant, unable to secrete T3SS effector proteins, as previously described (28). We tested this isolate for the presence of five pINV-encoded genes (*mxIG*, *mxID*, *icsB*, *ipaH*, and *ospF*) by colony PCR, using primers described in Table 2.

## Genomic analysis

Enterobase was used to identify publicly available ST99 genomes, using the filter by ST function (16); all ST99 genomes with an associated assembly and isolation date were included in our study, sequence accessions and metadata can be found in Table S1. Complete genome sequences of strains NCTC 9096 and CFSAN029787 were downloaded from GenBank (accessions [UGEL01000000](#) and [CP011416.1](#), respectively). All genomic analyses were performed using the Cloud Infrastructure for Microbial Bioinformatics (CLIMB) (29). Snippy v.4.6 (<https://github.com/tseemann/snippy>) was used to generate a core genome alignment, using CFSAN029787 as the reference. Gubbins v.3.2.1 (17) was used to identify recombinant regions of the alignment, and RaxML v.8.10 (30) was used to build a maximum likelihood phylogenetic tree, using the General Time Reversible (GTR) GAMMA nucleotide substitution model. BactDating v.1.2 (31) was used to infer the dated phylogeny, using the “relaxedgamma” model, the option to incorporate Gubbins detected recombination was selected and  $10^5$  Markov chain Monte Carlo (MCMC) chain iterations were run. To confirm the temporal signal (association between genetic divergence and time) within our dataset, tip nodes were assigned random dates and the analysis was rerun (this was completed  $n = 10$  times). We saw no overlap between the substitution rates of our real data and the randomized datasets (Fig. S3) that shows that the data pass the stringent test CR2 for the presence of a temporal signal according to Duchene et al. (32). To screen assemblies for the presence of pINV, ShigEiFinder was used, which screens for 38 pINV-encoded genes and deems an isolate positive when at least 26 genes are present (20).

## Inoculate preparation

Single red colonies (pINV+ EIEC) or white colonies (pINV– NCTC 9096 and T3SS-deficient EIEC) were selected and inoculated into 5 mL trypticase soy broth (TSB) and incubated overnight at 37°C, shaking at 400 rpm. 400  $\mu$ L of overnight culture was subsequently diluted in 20 mL TSB and grown until an optical density of  $\sim 0.6$  (measured at 600 nm) was reached. For zebrafish larvae infections, inoculate preparation was carried out by resuspension of the bacteria at the desired concentration in phosphate buffer saline (PBS, Sigma-Aldrich) pH 7.4 containing 2% polyvinylpyrrolidone (Sigma-Aldrich) and 0.5% phenol red (Sigma-Aldrich) as previously described (13).

**TABLE 2** Primers used to detect for the presence of pINV-encoded genes

Primer name	Primer sequence (5'-3')
mxID_Fwd	CAGAATGTAAGTAATGCACTGGCTATGATAC
mxID_Rev	CTGTCTATAAAATCCTGATCTAGAGGAAGGTTATC
mxIG_Fwd	CTGATTGTTGGGATAAGGCTGG
mxIG_Rev	CCGAGATCCCCTGTTTACCTC
ospF_Fwd	AAAAGATGAAGGCCTGATGGGAGCATTAAAC
ospF_Rev	TGGTGGATAAAACCCGCCAGAATGAACA
icsB_Fwd	GGTTCCAAGATCTGGCGATTAAAGAGAATTGTAATAATC
icsB_Rev	GGGCCTATACGCGTTGAAGATACAGAG
ipaH1.4_Fwd	GGGCATGAAAAAGCTACATCC
ipaH1.4_Rev	CACCATTATTCGAGTATAGGGAGAG

## Zebrafish larvae infection

Wild-type-AB zebrafish embryos were used for *in vivo* studies. Embryos were kept in  $0.5 \times E2$  medium supplemented with  $0.3 \mu\text{g/mL}$  methylene blue and incubated at  $28.5^\circ\text{C}$  unless otherwise stated. Using a microinjector,  $\sim 1 \text{ nL}$  of bacterial suspension was injected into the HBV of 3 d post-fertilization (dpf) zebrafish larvae, following previously described procedures (13). The precise inoculum was determined retrospectively by homogenization of larvae at 0 h post-infection and plating on TSA plates supplemented with 0.01% Congo red.

For survival assays, zebrafish larvae were visualized using a light stereomicroscope at 24 and 48 hpi; the presence of a heartbeat was used to determine viability. For colony forming unit (CFU) counts, larvae were disrupted in PBS using a pestle pellet blender at 0 and 6 hpi. Serial dilutions in PBS and plating on TSA plates supplemented with 0.01% Congo red were then performed to estimate the bacterial load in each larva. Statistical analysis was performed in GraphPad Prism 9.

## *In vitro* secretion assay

Secretion of T3SS effectors was tested as previously described (33). Briefly, bacteria were grown overnight, subcultured and grown until exponential phase ( $\text{OD} = 0.4\text{--}0.5$ ) at either  $28.5^\circ\text{C}$ ,  $32.5^\circ\text{C}$ , or  $37^\circ\text{C}$ . Cultures were then incubated for 3 h in the presence or absence of Congo red to induce type 3 secretion. Secreted proteins were collected from culture supernatants, precipitated using trichloroacetic acid (Sigma-Aldrich), and then analyzed using SDS-PAGE and Coomassie Brilliant Blue R-250 (Bio-Rad) staining.

## ACKNOWLEDGMENTS

We thank members of the Mostowy and Holt labs for the helpful discussion. We thank the Biological Services Facility (BSF) at LSHTM for assistance with zebrafish care and breeding.

S.L.M. is supported by a Biotechnology and Biological Sciences Research Council LIDo Ph.D. studentship (BB/T008709/1). V.T. was supported by an LSHTM/Wellcome Trust Institutional Strategic Support Fund (ISSF) Fellowship (204928/Z/16/Z). A.T.L.J. is funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska – Curie Individual Fellowship (grant agreement No. H2020-MSCA-IF2020-895330). Work in the S.M. laboratory is supported by a European Research Council Consolidator Grant (grant agreement No. 772853-ENTRAPMENT) and Wellcome Trust Senior Research Fellowship (206444/Z/17/Z).

## AUTHOR AFFILIATIONS

<sup>1</sup>Department of Infection Biology, London School of Hygiene and Tropical Medicine, London, United Kingdom

<sup>2</sup>Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, Victoria, Australia

<sup>3</sup>Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom

<sup>4</sup>Gastrointestinal Pathogens and Food Safety (One Health), UK Health Security Agency, London, United Kingdom

## PRESENT ADDRESS

Vincenzo Torraca, School of Life Sciences, University of Westminster, London, United Kingdom

Damián Lobato-Márquez, National Center of Biotechnology CSIC, Madrid, Spain

## AUTHOR ORCIDs

Sydney L. Miles  <http://orcid.org/0000-0003-2291-4105>



Vincenzo Torraca [id](http://orcid.org/0000-0001-7340-0249) <http://orcid.org/0000-0001-7340-0249>

Zoe A. Dyson [id](http://orcid.org/0000-0002-8887-3492) <http://orcid.org/0000-0002-8887-3492>

Ana Teresa López-Jiménez [id](http://orcid.org/0000-0002-0289-738X) <http://orcid.org/0000-0002-0289-738X>

Ebenezer Foster-Nyarko [id](http://orcid.org/0000-0001-6620-9403) <http://orcid.org/0000-0001-6620-9403>

Damián Lobato-Márquez [id](http://orcid.org/0000-0002-0044-7588) <http://orcid.org/0000-0002-0044-7588>

Kathryn E. Holt [id](http://orcid.org/0000-0003-3949-2471) <http://orcid.org/0000-0003-3949-2471>

Serge Mostowy [id](http://orcid.org/0000-0002-7286-6503) <http://orcid.org/0000-0002-7286-6503>

## FUNDING

Funder	Grant(s)	Author(s)
<a href="#">UKRI   Biotechnology and Biological Sciences Research Council (BBSRC)</a>	BB/T008709/1	Sydney Leigh Miles
<a href="#">London School of Hygiene and Tropical Medicine (LSHTM)</a>	204928/Z/16/Z	Vincenzo Torraca
<a href="#">EC   H2020   PRIORITY 'Excellent science'   H2020 Marie Skłodowska-Curie Actions (MSCA)</a>	H2020-MSCA-IF2020-895330	Ana Teresa López-Jiménez
<a href="#">EC   H2020   PRIORITY 'Excellent science'   H2020 European Research Council (ERC)</a>	772853-ENTRAPMENT	Serge Mostowy
<a href="#">Wellcome Trust (WT)</a>	206444/Z/17/Z	Serge Mostowy

## AUTHOR CONTRIBUTIONS

Sydney L. Miles, Conceptualization, Formal analysis, Funding acquisition, Investigation, Writing – original draft, Writing – review and editing | Vincenzo Torraca, Supervision, Writing – review and editing | Zoe A. Dyson, Formal analysis, Writing – review and editing | Ana Teresa López-Jiménez, Investigation, Writing – review and editing | Ebenezer Foster-Nyarko, Formal analysis, Writing – review and editing | Damián Lobato-Márquez, Investigation, Writing – review and editing | Claire Jenkins, Resources, Writing – review and editing | Kathryn E. Holt, Conceptualization, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review and editing | Serge Mostowy, Conceptualization, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review and editing

## DATA AVAILABILITY STATEMENT

All genomes used in this study are publicly available in Enterobase, and assembly accessions are provided in Table S1.

## ETHICS APPROVAL

Animal experiments were performed according to the Animals (Scientific Procedures) Act 1986 and approved by the Home Office (Project license: P4E664E3C). All experiments were conducted up to 5 days post-fertilization (dpf).

## ADDITIONAL FILES

The following material is available [online](#).

### Supplemental Material

**Fig S1 (mBio00882-23-s0001.pdf)**. Results of date randomization test.

**Fig S2 (mBio00882-23-s0002.pdf).** Virulence of ST99 EIEC isolates is temperature dependent.

**Fig S3 (mBio00882-23-s0003.pdf).** Secretion of virulence factors *in vitro* by ST99 EIEC pINV+1 (Congo red+ colony), T3SS-deficient ST99 EIEC (Congo red- colony), and *S. flexneri*.

**Fig S4 (mBio00882-23-s0004.pdf).** Colony PCR to check for the presence of pINV-encoded genes.

**Fig S5 (mBio00882-23-s0005.pdf).** Bacterial burden of pINV+1 compared to its T3SS-deficient counterpart and the oldest available pINV- isolate, NCTC 9096.

**TABLE S1 (mBio00882-23-s0006.xlsx).** Details and metadata for all genomes downloaded from Enterobase.

## REFERENCES

- Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, Wu Y, Sow SO, Sur D, Breiman RF, Faruque AS, Zaidi AK, Saha D, Alonso PL, Tamboura B, Sanogo D, Onwuchekwa U, Manna B, Ramamurthy T, Kanungo S, Ochieng JB, Omere R, Oundo JO, Hossain A, Das SK, Ahmed S, Qureshi S, Quadri F, Adegbola RA, Antonio M, Hossain MJ, Akinsola A, Mandomando I, Nhampossa T, Acácio S, Biswas K, O'Reilly CE, Mintz ED, Berkeley LY, Muhsen K, Sommerfelt H, Robins-Browne RM, Levine MM. 2013. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the global enteric multicenter study, GEMS): a prospective, case-control study. *Lancet* 382:209–222. [https://doi.org/10.1016/S0140-6736\(13\)60844-2](https://doi.org/10.1016/S0140-6736(13)60844-2)
- Pupo GM, Lan R, Reeves PR. 2000. Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc Natl Acad Sci U S A* 97:10567–10572. <https://doi.org/10.1073/pnas.180094797>
- Sahl JW, Morris CR, Emberger J, Fraser CM, Ochieng JB, Juma J, Fields B, Breiman RF, Gilmour M, Nataro JP, Rasko DA. 2015. Defining the phylogenomics of *Shigella* species: a pathway to diagnostics. *J Clin Microbiol* 53:951–960. <https://doi.org/10.1128/JCM.03527-14>
- Falconi M, Colonna B, Prosseda G, Micheli G, Gualerzi CO. 1998. Thermoregulation of *Shigella* and *Escherichia coli* EIEC pathogenicity: a temperature-dependent structural transition of DNA modulates accessibility of virF promoter to transcriptional repressor H-NS. *EMBO J* 17:7033–7043. <https://doi.org/10.1093/emboj/17.23.7033>
- Escher M, Scavia G, Morabito S, Tozzoli R, Maugliani A, Cantoni S, Fracchia S, Bettati A, Casa R, Gesu GP, Torresani E, Caprioli A. 2014. A severe foodborne outbreak of diarrhoea linked to a canteen in Italy caused by enteroinvasive *Escherichia coli*, an uncommon agent. *Epidemiol Infect* 142:2559–2566. <https://doi.org/10.1017/S0950268814000181>
- Iqbal J, Malviya N, Gaddy JA, Zhang C, Seier AJ, Haley KP, Doster RS, Farfán-García AE, Gómez-Duarte OG. 2022. Enteroinvasive *Escherichia coli* O96:H19 is an emergent Biofilm-forming pathogen. *J Bacteriol Res* 204: e0056221. <https://doi.org/10.1128/jb.00562-21>
- Newitt S, MacGregor V, Robbins V, Bayliss L, Chattaway MA, Dallman T, Ready D, Aird H, Puleston R, Hawker J. 2016. Two linked enteroinvasive *Escherichia coli* outbreaks, Nottingham, UK, June 2014. *Emerg Infect Dis* 22:1178–1184. <https://doi.org/10.3201/eid2207.152080>
- Bai X, Scheutz F, Dahlgren HM, Hedenström I, Jernberg C. 2021. Characterization of clinical *Escherichia coli* strains producing a novel shiga toxin 2 subtype in Sweden and Denmark. *Microorganisms* 9:2374. <https://doi.org/10.3390/microorganisms9112374>
- Michelacci V, Prosseda G, Maugliani A, Tozzoli R, Sanchez S, Herrera-León S, Dallman T, Jenkins C, Caprioli A, Morabito S. 2016. Characterization of an emergent clone of enteroinvasive *Escherichia coli* circulating in Europe. *Clin Microbiol Infect* 22:287. <https://doi.org/10.1016/j.cmi.2015.10.025>
- Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch G-J, White S, Chow W, Kilian B, Quintais LT, Guerra-Assunção JA, Zhou Y, Gu Y, Yen J, Vogel J-H, Eyre T, Redmond S, Banerjee R, Chi J, Fu B, Langley E, Maguire SF, Laird GK, Lloyd D, Kenyon E, Donaldson S, Sehra H, Almeida-King J, Loveland J, Trevanion S, Jones M, Quail M, Willey D, Hunt A, Burton J, Sims S, McLay K, Plumb B, Davis J, Clee C, Oliver K, Clark R, Riddle C, Elliot D, Threadgold G, Harden G, Ware D, Begum S, Mortimore B, Kerry G, Heath P, Phillimore B, Tracey A, Corby N, Dunn M, Johnson C, Wood J, Clark S, Pelan S, Griffiths G, Smith M, Glithero R, Howden P, Barker N, Lloyd C, Stevens C, Harley J, Holt K, Panagiotidis G, Lovell J, Beasley H, Henderson C, Gordon D, Auger K, Wright D, Collins J, Raisen C, Dyer L, Leung K, Robertson L, Ambridge K, Leongamornlert D, McGuire S, Gilderthorpe R, Griffiths C, Manthavadi D, Nichol S, Barker G, Whitehead S, Kay M, Brown J, Murnane C, Gray E, Humphries M, Sycamore N, Barker D, Saunders D, Wallis J, Babbage A, Hammond S, Mashreghi-Mohammadi M, Barr L, Martin S, Wray P, Ellington A, Matthews N, Ellwood M, Woodmansey R, Clark G, Cooper JD, Tromans A, Grafham D, Skuce C, Pandian R, Andrews R, Harrison E, Kimberley A, Garnett J, Fosker N, Hall R, Garner P, Kelly D, Bird C, Palmer S, Gehring I, Berger A, Dooley CM, Ersan-Ürün Z, Eser C, Geiger H, Geisler M, Karotki L, Kirn A, Konantz J, Konantz M, Oberländer M, Rudolph-Geiger S, Teucke M, Lanz C, Raddatz G, Osoegawa K, Zhu B, Rapp A, Widaa S, Langford C, Yang F, Schuster SC, Carter NP, Harrow J, Ning Z, Herrero J, Searle SMJ, Enright A, Geisler R, Plasterk RHA, Lee C, Westerfield M, de Jong PJ, Zou LI, Postlethwait JH, Nüsslein-Volhard C, Hubbard TJP, Roest Crollius H, Rogers J, Stemple DL. 2013. The Zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496:498–503. <https://doi.org/10.1038/nature12111>
- Gomes MC, Mostowy S. 2020. The case for modeling human infection in Zebrafish. *Trends Microbiol* 28:10–18. <https://doi.org/10.1016/j.tim.2019.08.005>
- Mostowy S, Boucontet L, Mazon Moya MJ, Sirianni A, Boudinot P, Hollinshead M, Cossart P, Herbomel P, Levraud J-P, Colucci-Guyon E. 2013. The Zebrafish as a new model for the *in vivo* study of *Shigella flexneri* interaction with phagocytes and bacterial autophagy. *PLoS Pathog* 9:e1003588. <https://doi.org/10.1371/journal.ppat.1003588>
- Torraca V, Kafrou M, Watson J, Duggan GM, Guerrero-Gutiérrez H, Krokowski S, Hollinshead M, Clarke TB, Mostowy RJ, Tomlinson GS, Sancho-Shimizu V, Clements A, Mostowy S. 2019. *Shigella sonnei* infection of Zebrafish reveals that O-antigen mediates neutrophil tolerance and dysentery incidence. *PLoS Pathog* 15:e1008006. <https://doi.org/10.1371/journal.ppat.1008006>
- Willis AR, Torraca V, Gomes MC, Shelley J, Mazon-Moya M, Filloux A, Lo Celso C, Mostowy S. 2018. *Shigella*-induced emergency granulopoiesis protects Zebrafish larvae from secondary infection. *mBio* 9:e00933-18. <https://doi.org/10.1128/mBio.00933-18>
- Van Ngo H, Robertin S, Brokatzky D, Bielecka MK, Lobato-Márquez D, Torraca V, Mostowy S. 2022. Septins promote caspase activity and coordinate mitochondrial apoptosis. *Cytoskeleton*. <https://doi.org/10.1002/cm.21696>
- Zhou Z, Alikhan N-F, Mohamed K, Fan Y, Achtman M. 2020. The Enterobase user's guide, with case studies on salmonella transmissions, *Yersinia Pestis* Phylogeny, and *Escherichia coli* Genomic diversity. *Genome Res* 30:138–152. <https://doi.org/10.1101/gr.251678.119>
- Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 43:e15. <https://doi.org/10.1093/nar/gku1196>
- Didelot X, Croucher NJ, Bentley SD, Harris SR, Wilson DJ. 2018. Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Res* 46:e134. <https://doi.org/10.1093/nar/gky783>

19. Didelot X, Parkhill J. 2022. A scalable analytical approach from bacterial genomes to epidemiology. *Philos Trans R Soc Lond B Biol Sci* 377:20210246. <https://doi.org/10.1098/rstb.2021.0246>
20. Zhang X, Payne M, Nguyen T, Kaur S, Lan R. 2021. Cluster-specific gene markers enhance *Shigella* and enteroinvasive *Escherichia coli* in silico serotyping. *Microb Genom* 7:000704. <https://doi.org/10.1099/mgen.0.000704>
21. Michelacci V, Tozzoli R, Arancia S, D'Angelo A, Boni A, Knijn A, Prosseda G, Greig DR, Jenkins C, Camou T, Sirok A, Navarro A, Schelotto F, Varela G, Morabito S. 2020. Tracing back the evolutionary route of enteroinvasive *Escherichia coli* (EIEC) and *Shigella* through the example of the highly pathogenic o96: H19 eiec clone. *Front Cell Infect Microbiol* 10:260. <https://doi.org/10.3389/fcimb.2020.00260>
22. Cowley LA, Oresegun DR, Chattaway MA, Dallman TJ, Jenkins C. 2018. Phylogenetic comparison of Enteroinvasive *Escherichia Coli* isolated from cases of Diarrhoeal disease in England, 2005-2016. *J Med Microbiol* 67:884–888. <https://doi.org/10.1099/jmm.0.000739>
23. Duggan GM, Mostowy S. 2018. Use of Zebrafish to study *Shigella* infection. *Dis Model Mech* 11: <https://doi.org/10.1242/dmm.032151>.
24. Lan R, Alles MC, Donohoe K, Martinez MB, Reeves PR. 2004. Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and *Shigella* Spp. *IAI* 72:5080–5088. <https://doi.org/10.1128/IAI.72.9.5080-5088.2004>
25. Pilla G, McVicker G, Tang CM. 2017. Genetic plasticity of the *Shigella* virulence plasmid is mediated by Intra- and inter-molecular events between insertion sequences. *PLoS Genet* 13:e1007014. <https://doi.org/10.1371/journal.pgen.1007014>
26. Pasqua M, Michelacci V, Di Martino ML, Tozzoli R, Grossi M, Colonna B, Morabito S, Prosseda G. 2017. The intriguing evolutionary journey of Enteroinvasive *E. coli* (EIEC) toward pathogenicity. *Front Microbiol* 8:2390. <https://doi.org/10.3389/fmicb.2017.02390>
27. Sansonetti PJ, Kopecko DJ, Formal SB. 1982. Involvement of a plasmid in the invasive ability of *Shigella Flexneri*. *Infect Immun* 35:852–860. <https://doi.org/10.1128/iai.35.3.852-860.1982>
28. Qadri F, Hossain SA, Ciznár I, Haider K, Ljungh A, Wadstrom T, Sack DA. 1988. Congo red binding and salt aggregation as indicators of virulence in *Shigella species*. *J Clin Microbiol*. 26:1343–1348. <https://doi.org/10.1128/jcm.26.7.1343-1348.1988>
29. Connor TR, Loman NJ, Thompson S, Smith A, Southgate J, Poplawski R, Bull MJ, Richardson E, Ismail M, Thompson SE, Kitchen C, Guest M, Bakke M, Sheppard SK, Pallen MJ. 2016. CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. *Microb Genom* 2:e000086. <https://doi.org/10.1099/mgen.0.000086>
30. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
31. Didelot X, Croucher NJ, Bentley SD, Harris SR, Wilson DJ. 2018. Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Res* 46:e134. <https://doi.org/10.1093/nar/gky783>
32. Duchêne S, Duchêne D, Holmes EC, Ho SYW. 2015. The performance of the date-randomization test in phylogenetic analyses of time-structured virus data. *Mol Biol Evol* 32:1895–1906. <https://doi.org/10.1093/molbev/msv056>
33. Reinhardt J, Kolbe M. 2014. Secretion assay in *Shigella flexneri*. *Bio-protocol* 4. <https://doi.org/10.21769/BioProtoc.1302>