Identification and prevalence of in vivo-induced genes in enterohaemorrhagic Escherichia coli
Marion Gardette, Simon Le Hello, Patricia Mariani-Kurkdjian, Laetitia Fabre, François Gravey, Annie Garrivier, Estelle Loukiadis, Gregory Jubelin

To cite this version:
Marion Gardette, Simon Le Hello, Patricia Mariani-Kurkdjian, Laetitia Fabre, François Gravey, et al.. Identification and prevalence of in vivo-induced genes in enterohaemorrhagic Escherichia coli. Virulence, Landes Bioscience, 2019, 10 (1), pp.180-193. 10.1080/21505594.2019.1582976. hal-02154035

HAL Id: hal-02154035
https://hal-normandie-univ.archives-ouvertes.fr/hal-02154035
Submitted on 26 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Identification and prevalence of in vivo-induced genes in enterohaemorrhagic Escherichia coli

Marion Gardette, Simon Le Hello, Patricia Mariani-Kurkdjian, Laetitia Fabre, François Gravey, Annie Garrivier, Estelle Loukiadis & Grégory Jubelin

To cite this article: Marion Gardette, Simon Le Hello, Patricia Mariani-Kurkdjian, Laetitia Fabre, François Gravey, Annie Garrivier, Estelle Loukiadis & Grégory Jubelin (2019) Identification and prevalence of in vivo-induced genes in enterohaemorrhagic Escherichia coli, Virulence, 10:1, 180-193, DOI: 10.1080/21505594.2019.1582976

To link to this article: https://doi.org/10.1080/21505594.2019.1582976

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
Identification and prevalence of in vivo-induced genes in enterohaemorrhagic Escherichia coli

Marion Gardette\textsuperscript{a,b}, Simon Le Hello\textsuperscript{c,d}, Patricia Mariani-Kurkdjian\textsuperscript{e}, Laetitia Fabre\textsuperscript{e}, François Gravey\textsuperscript{c,d}, Annie Garrivier\textsuperscript{e}, Estelle Loukiadi\textsuperscript{a,b,f}, and Grégory Jubelin\textsuperscript{a,\textsuperscript{*}}

\textsuperscript{a}UCA, INRA, UMR454 MEDIS, Clermont-Ferrand, France; \textsuperscript{b}Laboratoire d’écologie microbienne de Lyon, Université de Lyon, CNRS, INRA, UCBL, VetAgro Sup, Marcy l’Étoile, France; \textsuperscript{c}Centre de Référence National des Escherichia coli, Shigella et Salmonella, Institut Pasteur, Paris, France; \textsuperscript{d}Université de Normandie, EA 2656 GRAM 2.0, UNICAEN, Caen, France; \textsuperscript{e}Service de Microbiologie, Centre National de Référence associé Escherichia coli, Hôpital Robert-Debré, AP-HP, Paris, France; \textsuperscript{f}Laboratoire national de référence des E. coli, Université de Lyon, VetAgro Sup, Marcy l’Étoile, France

\textsuperscript{*}Corresponding author. E-mail: gregory.jubelin@inra.fr

ABSTRACT

Enterohaemorrhagic Escherichia coli (EHEC) are food-borne pathogens responsible for bloody diarrhoea and renal failure in humans. While Shiga toxin (Stx) is the cardinal virulence factor of EHEC, its production by \textit{E. coli} is not sufficient to cause disease and many Shiga-toxin producing \textit{E. coli} (STEC) strains have never been implicated in human infection. So far, the pathophysiology of EHEC infection is not fully understood and more knowledge is needed to characterize the “auxiliary” factors that enable a STEC strain to cause disease in humans. In this study, we applied a recombinase-based \textit{in vivo} expression technology (RIVET) to the EHEC reference strain EDL933 in order to identify genes specifically induced during the infectious process, using mouse as an infection model. We identified 31 \textit{in vivo}-induced (\textit{ivi}) genes having functions related to metabolism, stress adaptive response and bacterial virulence or fitness. Eight of the 31 \textit{ivi} genes were found to be heterogeneously distributed in EHEC strains circulating in France these last years. In addition, they are more prevalent in strains from the TOP seven priority serotypes and particularly strains carrying significant virulence determinants such as Stx2 and intimin adhesin. This work sheds further light on bacterial determinants over-expressed \textit{in vivo} during infection that may contribute to the potential of STEC strains to cause disease in humans.

ARTICLE HISTORY

Received 10 December 2018
Revised 25 January 2019
Accepted 6 February 2019

KEYWORDS

EHEC; intestinal pathogen; virulence; metabolism; stress response; gene prevalence; RIVET

Introduction

Shiga toxin-producing \textit{Escherichia coli} (STEC) are recognized as an important cause of food-borne disease in humans and cause large outbreaks worldwide. Pathogenic STEC, designated as enterohaemorrhagic \textit{E. coli} (EHEC), are responsible for intestinal disorders, including watery or bloody diarrhoea, that may ultimately evolve to life-threatening diseases such as haemolytic uremic syndrome (HUS) or thrombotic thrombocytopenic purpura. The two cardinal virulence factors in EHEC are type III secretion system (T3SS) and Shiga toxins (Stx). T3SS, encoded by the Locus of Enterocyte Effacement (LEE), is involved in the formation of attaching and effacing (A/E) lesions on gut mucosa through the injection of specific effectors within enterocytes [1]. Interactions of bacterial effectors with eukaryotic signal transduction pathways lead to host cytoskeleton reorganization that is characterized by two key markers of A/E lesions: an effacement of microvilli and formation of pedestals beneath adherent EHEC. Among translocated bacterial proteins, Tir is crucial since it acts as a receptor for the adhesin intimin (encoded by gene \textit{eae}), upon its integration into the host cell plasma membrane [2]. Strong interaction between intimin and Tir leads to an intimate attachment of EHEC to infected epithelial cells. Thirty distinct intimin subtypes with variation in their C-terminal part (involved in Tir binding) have been described so far [3] and this represents a standard tool for typing EHEC strains as well as for epidemiological studies. The biological significance of \textit{eae} subtypes is not known but it has been suggested that intimin variants may be responsible for different host tissue cell tropisms [4]. The other cardinal virulence factor in EHEC is Stx, which is produced in the gut lumen and crosses the epithelial barrier to reach the bloodstream and the target organs including brain and kidneys. Stx binds specifically to globotriaosylceramide, a receptor present at the surface of glomerular...
endothelial cells, and blocks the translation in intoxicated cells resulting in cell death by apoptosis [5]. Two antigenically distinct forms of Stx can be produced by STEC, Stx1, and Stx2, and for each type of Stx up to seven subtypes have been described to date [6]. Stx2-producing *E. coli* strains, especially subtypes Stx2a and Stx2d, are more often associated with the most severe symptoms of the disease than Stx1-producing strains [7,8].

Because Stx induces bloody diarrhoea and renal dysfunction in infected patients, the toxin is considered as the main virulence factor of EHEC. However, many STEC strains isolated from the animal reservoir or from food products have never been associated with human infection and these strains have been classified as seropathotype E by Karmali and colleagues [9]. While several hypotheses could explain the occurrence of these putative non-pathogenic STEC strains (cryptic *stx* gene, undetected cases of human infection, weak EHEC survival during transmission to human, etc.), one hypothesis relies on the idea that Stx is necessary but not sufficient to cause disease in humans and that a minimal combination of virulence factors (including Stx) is required. In accordance with this hypothesis, epidemiological studies have emphasized that LEE is frequently associated with severe STEC infections [10,11]. However, about 10% of the confirmed STEC cases reported in the European Union between 2007 and 2010, including HUS patients, were caused by LEE-negative STEC strains [11]. These strains probably utilize other determinants to colonize the gut mucosa efficiently. Apart from intimin, EHEC produces many other potential adhesins or fimbriae that have been primarily described in vitro [12]. Similarly, many EHEC strains also carry large plasmids including the pO157 which encodes potential virulence factors such as the enterohaemolysin EhxA whose gene is used as a genetic marker for epidemiological studies [13]. However, the contribution of these determinants to the EHEC infectious process still needs to be evaluated. Over the past 15 y, the genome sequencing of many EHEC strains has provided new insights into their potential virulence determinants. In particular, the presence of additional DNA segments termed O-islands in the EHEC genome was linked to virulence potential since they contain several genes associated with virulence, including *stx* and T3SS-encoding genes [14,15]. However, these O-islands also carry hundreds of genes of unknown predictive function that might be involved in EHEC infectious process. The pathophysiology of STEC infection in humans is therefore still far from being fully understood and more knowledge is needed to characterize the “auxiliary” factors that enable a STEC strain to cause disease in humans.

In this study, we applied a recombinase-based in vivo expression technology (RIVET) to the EHEC reference strain EDL933 in order to identify genes specifically induced during the infectious process, using mouse as an infection model. These in vivo-induced (*ivi*) genes would give us a subclass of genes encoding products potentially important for STEC pathogenesis. Thirty-one *ivi* genes have been identified and most of them are linked to general metabolism or to stress adaptive responses, or have poorly characterized functions. Among them, eight genes were not homogeneously distributed in *E. coli* and their prevalence was analysed in EHEC strains circulating in France these last years. Finally, the potential role of selected *ivi* genes in EHEC fitness has been investigated by competition assays in mice.

**Materials and methods**

**Bacterial strains, plasmids, media and growth conditions**

The bacterial strains and plasmids used in this study are listed in Table S1. The well-characterized O157:H7 EDL933 strain, which is *stx1a*, *stx2a*, and LEE+, was used throughout this study. The EDL933 mutants Δ*mhpR*, Δ*yhbU*, Δ*yifE*, Δ*yifR*, Δ*Z4799*, Δ*Z4070* and Δ*Z3135* were constructed using the one-step PCR-based method [16]. Chromosomal deletion mutations were confirmed by PCR followed by DNA sequencing. The Δ*agaF* mutant was obtained using primers described in [17]. Bacteria were routinely grown in Luria-Bertani (LB) medium at 37°C unless otherwise indicated. For strain WM3064, 2,6-diaminopimelic acid (DAP) was added to a concentration of 100 µM. When required, antibiotics were used at the following concentrations: ampicillin (Amp), 50 µg.ml⁻¹; kanamycin (Kan), 25 µg.ml⁻¹; gentamicin (Gm), 15 µg.ml⁻¹; chloramphenicol (Cm), 25 µg.ml⁻¹; streptomycin (Sm), 50 µg.ml⁻¹.

**Construction of an EHEC RIVET library**

The EHEC RIVET library used in this study has been constructed using tools from the improved RIVET method described by Osorio et al. [18]. The strains used to construct the RIVET library, designated EDL-RES and EDL-RES1, are spontaneous streptomycin-resistant (SmR) strains, harbouring a chromosomally located cassette, respectively, *res-kan-sacB-res* or *res1-kan-sacB-res1*. The strains were constructed by amplifying a cassette composed of *kan* and *sacB* genes and
flanked by res sequences from plasmid pRES or pRES1 with primers RES-F (5′-AAGCTGTACCTCCACCCGC -3′) and RES-R (5′- GATCAGGCCCATATAGGGC GAATTCGTTACC-3′). The PCR products were digested with SacI and ApaI and inserted into the corresponding sites of pGP-Tn7-Gm. The corresponding constructs, pGP-Tn7-RES and pGP-Tn7-RES1, were then transferred to EDL933-SmR by mating experiments in order to integrate the kan-sacB cassette into the chromosome as described in [19]. Correct insertion of the cassette was validated by PCR and sequencing. Sucrose sensitivity of EDL-RES and EDL-RES1 strains was validated on LB plates without NaCl supplemented with 4% sucrose.

The EHEC RIVET library was constructed as follows. Genomic DNA from strain EDL933 was partially digested with Sau3AI. Fragments between 0.5 and 3 kb were gel-purified and cloned into the BglII site of vectors pGOA1193 or pGOA1195. The ligation mixtures were transformed into PIR1 strain. Transformants were selected on plates with Amp and pooled to perform a plasmid extraction en masse. The plasmid library was then transformed into WM3064, a strain used as the donor to mobilize plasmids into EDL-RES or EDL-RES1 by mating experiments, leading to plasmid integration into the EHEC genome due to its oriR6K origin of replication. Approximately 21,000 colonies, obtained on LB agar supplemented with Amp and Kan, were collectively resuspended and aliquots were frozen at –80°C in 20% glycerol.

**Screening of the RIVET library for in vivo-induced genes**

An aliquot of the RIVET library was grown 6 h at 37°C in LB with Amp and Kan. Bacteria were washed once in PBS and approximately 5.10^8 CFU was intragastrically administered to 6-week-old BALB/c mice. After 18 h of infection, mice were euthanized and faeces or large intestine contents were collected and homogenized in sterile PBS. Serial dilutions were next spotted on LB agar + Sm or LB agar without NaCl + Sm and 4% sucrose. The plates were incubated for 16 h at 30°C. Individual sucrose-resistant colonies were tested for their sensitivity to Kan and resistance to Amp. Correct clones were grown in LB + Amp + Gm in 96-well plates and stored at –80°C in 20% glycerol. DNA flanking the 5′ end of *tnpR* was amplified for each individual clone by SP-PCR [20] using the primer SP-PCR1 (5′-ATGACGTCACCTTCTCCAC-3′) and sequenced using the primer SP-PCR2 (5′-AGATGCGATTTGCTTTC ACG-3′). Obtained sequences were compared with the EHEC EDL933 genome using the MicroScope platform (https://www.genoscope.cns.fr/agc/microscope/home/index.php).

**Construction of p*chuA*-tnpR fusion strains**

EDL-RES and EDL-RES1 strains carrying *tnpR* gene under the control of *chuA* promoter region were constructed as follows. First, *chuA* promoter was amplified from genomic DNA of EDL933 using primers ChuA-F (5′-GAAGATCTTCGAGATAGGTTTCTACAACC-3′) and ChuA-R (5′-GAAGATCTTACGATTCTCCAT GAGA-3′), which contain BglII restriction sites in their 5′ ends. PCR product and vectors pGOA1193, pGOA1194 and pGOA1195 were digested with BglII and ligated separately to yield plasmids p1193-*chuA*, p1194-*chuA* and p1195-*chuA*. Correct orientation of *chuA* promoter was validated by PCR and sequencing. Plasmids were mobilized into EDL-RES and EDL-RES1 strains by mating experiments. Plasmid insertion into the *chuA* locus was verified by PCR and sequencing.

**Resolution assays**

Excision of the RES marker cassette was monitored in vitro after growth in LB or in vivo after mouse infection. The percentage of bacteria that have lost the RES marker cassette, which is termed resolution, is calculated for an individual sample by dividing the titre on LB without NaCl agar plates with Sm and 4% sucrose by the titre on LB agar plates with Sm.

**Gene prevalence analyses**

Distribution of *ivi* genes in 51 selected *E. coli* strains (listed in Table S2) was performed using the MicroScope platform (https://www.genoscope.cns.fr/agc/microscope/home/index.php) and completed when necessary using the BLAST program from the NCBI website. Then, distribution of eight *ivi* genes was evaluated on 228 selected strains obtained from the surveillance of EHEC infections from the French National Reference Centre for *Escherichia coli, Shigella* and *Salmonella*, Institut Pasteur, France (NRC). This collection represented all the 196 clinical EHEC strains received at NRC between 2015 and 2016 (n = 95 and 101, respectively, one clinical strain per patient) and was completed by food or environmental strains isolated in 2016 encompassing the serotype diversity. All the 32 non-human strains were isolated at the National Reference Laboratory, VetAgro Sup, France (LNR). Basic information as origin of isolates (human or food), year of isolation, clinical symptoms, and microbiological characteristics are shown in Table S3.
High-throughput whole-genome sequencing (WGS) was carried out on all *E. coli* strains at the “Plateforme de microbiologie mutualisée” (P2M) from the Pasteur International Bioresources network (Institut Pasteur, Paris, France). DNA extraction was carried out by using the MagNAPure 96 system (Roche). The libraries were prepared using the Nextera XT kit (Illumina) and the sequencing was done with the NextSeq 500 system (Illumina) generating 100 to 146 bp paired-end reads. Reads were trimmed and filtered using AlienTrimmer [21] with a quality Phred score threshold of 13 on a minimum length of 30 nucleotides. De novo assembly was performed with the SPAdes V.3.6.0 assembler [22]. Various genetic analyses were determined on SPAdes assembled sequences with web-based tools for WGS analysis developed by the Center for Genomic Epidemiology (CGE) (http://www.genomicepidemiology.org/). In particular, we searched O and H molecular serotyping, Multi-Locus-Sequence Typing (MLST) and main virulence gene content, i.e. *stx*1 and *stx*2 variants, *eae* variants, and *ehxA*. Furthermore, we extracted from each genome all the genes defining the phylogroup scheme [23]. A core-genome multi-alignment of all assembled genomes including an O157 *E. coli* reference genome (NC_002695) was done using the harvest v1.0.1 f ParSNP function [24]. The resulting single nucleotide polymorphisms (SNPs) were concatenated to generate a filtered multiple alignment that was used as input for the construction of a phylogenetic tree using MEGA6 [25], utilizing a maximum-likelihood (ML) approach. The final trees were visualized in the interactive Tree Of Life [26]. All genomes listed in supplementary Table S3 were uploaded on Enterobase (http://enterobase.warwick.ac.uk/species/ecoli) under the name indicated in “file NRC” item.

### Competition assays

Each of the *ivi* gene deletion strains was tested *in vivo* during competition assays in mice to evaluate if gene inactivation affects the bacterial fitness of EDL933 as previously described [27]. Briefly, individual EDL933 mutant and the wild-type (WT) strain were grown to mid-exponential phase in LB, washed once with PBS and mixed 1:1. Approximately $10^7$ CFU of both strains were administered intragastrically into 10 to 15 BALBc mice, 6 weeks of age, that were previously given drinking water containing Sm (5 g/L) for 24 h in order to selectively remove facultative anaerobic bacteria and open the niche for *E. coli* [27]. Eight days post-infection, faecal pellets were resuspended in PBS and plated on LB agar plates + Sm and LB agar plates + Sm + Kan to count WT + *ivi* mutant and *ivi* mutant alone, respectively. The WT population was obtained by subtracting CFU from LB + Sm + Kan plates to CFU from LB + Sm plates. Results are expressed as competitive indices obtained by dividing output ratio (WT/mutant day 8 post-infection) by the corresponding input ratio (WT/mutant from mouse inoculum).

### Statistical analyses

Clinical and microbiological data were entered anonymously. The Chi-square test or Fisher exact tests for proportions were performed. A *p*-value of <0.05 was considered significant. We used STATA version 12.0 (Stata Corporation, Texas, USA) for these analyses. Data in Figures 1 and 2 are presented as the mean ± the standard deviation of at least three independent experiments. Data in Figure 4 are presented as the median of two or three independent experiments depending on tested genes. ROUT method has been applied to exclude outliers.

**Figure 1. Efficiency of the RIVET system in EHEC.** Indicated strains carrying *tnpR* gene under the control of *chuA* promoter were grown in LB supplemented with various concentrations of 2,2’-dipyridyl. After 6 h of growth, cultures were enumerated on LB plates and on LB + sucrose plates to calculate the percentage of resolved CFU. Results are presented as mean and standard deviations from at least three independent experiments. Blue and red curves represent resolution obtained from the 3 *tnpR* variants inserted into EDL-RES and EDL-RES1 strains, respectively.
constructed by cloning random genomic DNA and RES cassettes may allow in vitro translation in vitro construct had a low resolution that did not...alleles have been conceived under the control of the promoter region...growth (white bars) or following mouse challenge (black bars). Results are presented as mean and standard deviations from at least three independent experiments. Differences between in vitro and in vivo conditions were analyzed by a two-tailed unpaired t-test (*P < 0.05; **P < 0.01).

**Results**

**Validation of the RIVET method in EHEC**

To evaluate the efficiency of the RIVET method described by Osorio et al. [18] in EHEC, we placed the reporter gene tnpR under the control of the promoter region of chuA, for which expression is well known to be induced in iron starvation conditions [28]. Transcriptional fusions were made with three different alleles of tnpR (tnpR, tnpR<sup>135</sup>, tnpR<sup>168</sup>) and transferred to EDL-RES or EDL-RES1 strains that carry the RES or RES1 cassette carrying sacB and kan genes (see M&M), respectively (Table S1). These RES cassette variants and tnpR alleles have been conceived in order to detect a wider range of ivi genes through alteration of cassette excision and tnpR translation efficiencies, respectively [18]. Resolution (i.e. percentage of bacterial cells that have lost the RES cassette) was quantified after 6 h of growth in LB supplemented with various concentrations of the iron chelator 2,2′-dipyridyl (Figure 1). In the presence of 2,2′-dipyridyl, the level of resolution increased for each strain except for strains carrying the P<sub>chuA</sub>-tnpR<sup>168</sup> fusion, which were therefore not used thereafter. In addition, patterns of resolution were different between the strains. The strain most sensitive to iron starvation was EDL-RES carrying the P<sub>chuA</sub>-tnpR construct since it reached almost 100% resolution. Strain EDL-RES1 carrying the same construct was less sensitive to iron deficiency, with a maximal resolution of 49%. This result was expected since the RES1 cassette is excised by TnpR with a lower efficiency in comparison with the original RES cassette [18]. Both strains harbouring the P<sub>chuA</sub>-tnpR<sup>135</sup> construct had a low resolution that did not exceed 15%. Taken together, these results indicate that the RIVET strategy based on RES cassette – tnpR combination can be readily used in EHEC and that the use of different alleles of tnpR and RES cassettes may allow the identification of promoter regions with a broad range of promoter strength.

**Identification of in vivo-induced genes in EHEC using mouse as a host model**

A genomic library of EHEC O157:H7 strain EDL933 was constructed by cloning random genomic DNA fragments upstream of tnpR genes (see M&M for details). More than 21,000 clones were obtained in EDL-RES or EDL-RES1 strains after selection on plates with kanamycin. Resistance to kanamycin implies that these clones still harbor the RES cassette and consequently do not carry an in vitro active promoter upstream of tnpR. The library was then inoculated intragastrically to conventional BALB/c mice. After 18 h of infection, bacteria were recovered from faeces either on LB with Sm to quantify the total number of EHEC or on LB with Sm and sucrose to select resolved strains. The percentage resolution after infection ranged from 0.01% to 6.75% between mice with a mean of 2.43%. We isolated a total of 1051 sucrose-resistant clones and 506 were confirmed to be sensitive to kanamycin. The DNA insert sequence was obtained from 335 isolates and 32 unique clones with integrated genes in the same orientation as tnpR were identified. Two clones correspond to the same gene (Z4799), thus leading to a list of 31 in vivo-induced (ivi) genes (Table 1). To validate our screening, we arbitrarily chose four genes and reconstructed transcriptional fusion with tnpR in EDL-RES strain. For each construct, resolution
Table 1. List of in vivo-induced genes identified in EHEC O157:H7 strain EDL933.

| COG category | Gene | Role or putative function |
|--------------|------|--------------------------|
| Metabolism   | Z0072 araB | L-Ribulokinase |
|              | Z4726 nirB | Large subunit of nitrite reductase |
|              | Z3960 gabT | 4-Aminobutyrate aminotransferase |
|              | Z3500 glpB | sn-Glycerol-3-phosphate dehydrogenase |
|              | Z2551 trpA | α-Subunit of tryptophan synthase |
|              | Z0607 ybaT | Putative nitrogen-containing metabolite transporter |
|              | Z2746 yjeH | Member of the APC superfamily of amino acid transporters |
|              | Z4486 aowW | Enzyme of the N-acetylgalactosamine phosphotransferase system |
|              | Z2540 btuR | Cobalamin adenoslytransferase |
|              | Z0061 pdaA | 4-Hydroxythreonine-4-phosphate dehydrogenase |
|              | Z2315 aozR | NADH-azoreductase |
|              | Z4799 yhiI | Transcriptional activator of nitric oxide reductase NorV |
|              | Z3960 gabT | Putative nitrogen-containing metabolite transporter |
|              | Z4070 norR | Putative HlyD family secretion protein |
|              | Z4524 norW | Multidrug efflux system protein |
|              | Z4199 ytfE | Putative transcription regulator |
|              | Z2551 trpA | Multidrug efflux system protein |
|              | Z5939 mdtM | Putative DNA processing protein |
|              | Z4799 - | Putative DNA processing protein |
| Information storage and processing | Z4409 cca | Multifunctional CCA protein |
|              | Z0444 mhpR | Transcriptional activator of 3-hydroxyphenylpropionic acid catabolism |
|              | Z20073 araC | Transcriptional regulator of L-arabinose transport and catabolism |
|              | Z4022 ascG | Transcriptional repressor of arbutin and salicin transport |
|              | Z4593 degQ | Putative transcriptional regulator |
|              | Z4017 norR | Transcriptional activator of nitric oxide reductase NorV |
|              | Z4799 - | Putative DNA processing protein |
| Cellular processes and signaling | Z5820 ythE | Iron–sulphur cluster repair protein |
|              | Z4886 yhiL | Putative HlyD family secretion protein |
|              | Z4926 mdeE | Multidrug efflux system protein |
|              | Z5393 mdmM(yijD) | Multidrug efflux system protein |
|              | Z4519 yhbU | Putative collagenase |
|              | Z4543 yghU | Putative glutathione S-transferase |
|              | Z4593 degQ | Serine endoprotease |
| Poorly characterized | Z4070 - | Uncharacterized protein |
|              | Z4357 ydlM | Uncharacterized protein; regulated by the SOS system regulator LexA |
|              | Z2064 - | Putative DNA processing protein of prophage CP-933K |
| No COG | Z3135 - | Putative adhesin/invasin |

COG category: [C] Energy production and conversion; [D] Cell cycle control, cell division, chromosome partitioning; [E] Amino acid transport and metabolism; [G] Carbohydrate transport and metabolism; [H] Coenzyme transport and metabolism; [I] Lipid transport and metabolism; [J] Translation, ribosomal structure and biogenesis; [K] Transcription; [L] Replication, recombination and repair; [M] Cell wall/membrane/envelope biogenesis; [O] Post-translational modification, protein turnover, and chaperones; [P] Inorganic ion transport and metabolism; [R] General function prediction only; [T] Signal transduction mechanisms; [U] Intracellular trafficking, secretion, and vesicular transport.

Functional classification of ivi genes

Functional categories of the 31 ivi genes were assigned using the COG database [29,30]. Thirteen genes (42%) had a COG associated with the general category metabolism, seven (22.5%) with the category information storage and processing, seven (22.5%) with the category cellular processes and signalling, three (10%) with the category poorly characterized and one (3%) had no COG assignment (Table 1). A large number of ivi genes are involved in the metabolism of i) carbohydrates: L-arabinose (araB, araC) and N-acetyl-galactosamine (agaW); ii) amino acids: tryptophan (trpA), glutamate (gabT) and others (yjeH); iii) lipids: sn-glycerol-3P (glpB) and fatty acids (fadA); iv) vitamins: B6 (pdbA) and B12 (btuR); v) aromatic compounds: 3-hydroxyphenylpropionic acid (3HPP) (mhpR) and arbutin (ascG) or vi) nitrogen-containing metabolites (nirB, ybaT). The second largest class of ivi genes is related to stress resistance. Genes mdIE and mdmM encode subunits of multidrug efflux pumps. yghU encodes a glutathione S-transferase, and yhiL encodes a putative protein belonging to the MFP family, which is involved in the export of a variety of compounds including drugs. Genes ythE and norR are related to nitrosative stress since it encodes, respectively, an iron-sulphur repair protein and a transcriptional activator of the norVW operon, whose products are involved in nitric oxide detoxification. There were also a number of ivi genes with unknown or poorly characterized function (ydlM, yhbU, Z0964, Z4070, Z4799, Z3135).
Distribution of ivi genes in E. coli

We next analyzed in silico the distribution of ivi genes among 51 sequenced E. coli strains (Table S2). Twenty-three of the 31 ivi genes are highly conserved since they were detected in at least 98% of all strains. The eight other ivi genes present a prevalence ranging from 38% to 90% and half are located in O-islands (Table 1 and S2). These genes encode transcriptional regulators (mhpR, ascG and yjiR), a protein from the Nle-encoding phage CP-933K (Z0964), a potential adhesin/invasin (Z3135), an efflux pump protein (mdtM), a putative DNA processing protein (Z4799) or an uncharacterized protein with a Cas3 domain (Z4070).

We next focused our attention on these eight ivi genes and analyzed their prevalence in the same E. coli collection after classification of strains as commensals, InPEC or ExPEC. As shown in Table 2, most of these ivi genes are not homogeneously distributed among the three categories. Interestingly, genes mhpR, ascG, mdtM, Z3135 and Z4070 present a higher prevalence in InPEC than in other groups (p < 0.03). Surprisingly, each ivi gene was almost exclusively present or absent in all strains from a given serotype (Table 3 and data not shown), suggesting a population-based distribution. For example, genes Z0964 and Z4799 were detected in all 27 O157:H7 strains but were not found in the genome of the 31 O26:H11 strains (Table 3). As the MLST types were concordant with the serotype, we found the same ivi gene distribution. Regarding their distribution by phylogroups (Table S4), some genes were more associated with a specific phylogroup than the others like genes Z0964 and Z4799 with phylogroup E and gene Z3135 with phylogroup A (p < 0.01). Distributions of the eight ivi genes according to serotype and phylogenetic population are shown in Figure 3.

Table 2. Distribution of eight ivi genes in 51 E. coli strains grouped as commensals, InPEC or ExPEC.

|        | mhpR | ascG | mdtM | yjiR | Z0964 | Z3135 | Z4070 | Z4799 |
|--------|------|------|------|------|-------|-------|-------|-------|
| commensal (n = 6) | 0.83 | 0.83 | 0.83 | 0.67 | 0.33 | 0 | 0.33 | 0.33 |
| InPEC (n = 31) | 0.97 | 0.97 | 0.97 | 0.74 | 0.71 | 0.52 | 0.68 | 0.35 |
| EAEc (n = 3) | 1 | 1 | 1 | 0.67 | 0.67 | 0 | 1 | 0 |
| EHEC (n = 18) | 1 | 1 | 1 | 0.78 | 0.72 | 0.72 | 0.72 | 0.72 |
| EPEC (n = 4) | 0.75 | 0.75 | 0.75 | 0.5 | 0.75 | 0.25 | 0.5 | 0.5 |
| ETEC (n = 6) | 0.5 | 0.71 | 0.43 | 0.43 | 0.5 | 0.21 | 0.14 | 0.5 |
| ExPEC (n = 14) | 0.5 | 0.71 | 0.43 | 0.43 | 0.5 | 0.21 | 0.14 | 0.5 |

Values indicate the proportion of strains carrying ivi genes (columns) for each category (lines).

To go further, we next evaluated the distribution of the eight ivi genes specifically in 228 EHEC strains isolated either from patients hospitalized in France between 2015 and 2016 or from contaminated foods (Table S3). The overall prevalence of the eight ivi genes ranges from 22% to 92%. First, we looked at the distribution of ivi genes in strains belonging to the 6 most commonly reported serotypes from severe cases of human STEC infections in the European Union [31] plus the emerging O80:H2 serotype [32,33] (defined here as the TOP seven priority serotype) versus other serotypes. We observed that all genes except Z3135 were more prevalent in strains from the TOP seven priority serotypes than in strains from other serotypes (p < 0.003). Surprisingly, each ivi gene was almost exclusively present or absent in all strains from a given serotype (Table 3 and data not shown), suggesting a population-based distribution. For example, genes Z0964 and Z4799 were detected in all 27 O157:H7 strains but were not found in the genome of the 31 O26:H11 strains (Table 3). As the MLST types were concordant with the serotype, we found the same ivi gene distribution. Regarding their distribution by phylogroups (Table S4), some genes were more associated with a specific phylogroup than the others like genes Z0964 and Z4799 with phylogroup E and gene Z3135 with phylogroup A (p < 0.01). Distributions of the eight ivi genes according to serotype and phylogenetic population are shown in Figure 3.

Table 3. Distribution of the eight ivi genes in 228 EHEC strains classified by serotypes.

|        | mhpR | ascG | mdtM | yjiR | Z0964 | Z3135 | Z4070 | Z4799 |
|--------|------|------|------|------|-------|-------|-------|-------|
| TOP 7 priority serotypes (n = 97) | 1 | 1 | 1 | 0.82 | 0.37 | 0.66 | 0.77 | 0.46 |
| O157:H7 (n = 37) | 1 | 1 | 1 | 1 | 0.96 | 0.96 | 1 | 1 |
| O26:H11 (n = 31) | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 |
| O103:H2 (n = 2) | 1 | 1 | 1 | 0.5 | 1 | 1 | 1 | 1 |
| O146:H21 (n = 12) | 1 | 1 | 1 | 0.08 | 1 | 1 | 1 | 1 |
| O91:H21 (n = 4) | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| O145:H28 (n = 4) | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| O80:H2 (n = 17) | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| Other serotypes (n = 131) | 0.86 | 0.87 | 0.86 | 0.77 | 0.11 | 0.58 | 0.72 | 0.24 |

Values indicate the proportion of strains carrying ivi genes (columns) for each category (lines).
Figure 3. Distribution of the eight ivi genes in the EHEC strain collection. A phylogenetic tree was constructed after a core-genome multi-alignment with all 228 STEC strains and the O157 E. coli reference genome (NC_002695). Presence of each ivi gene is indicated by different coloured forms. Phylogroups are shown by colorizing the numbering of each strain and serotypes are written. All this information is indicated at the end of each branch of the phylogenetic tree.

Table 4. Co-distribution of the eight ivi genes with known virulence factors in the EHEC strain collection.

|         | mhpR | ascG | mdtM | yjiR | Z0964 | Z3135 | Z4070 | Z4799 |
|---------|------|------|------|------|-------|-------|-------|-------|
| stx1 + stx2 (n = 52) | 1     | 1    | 1    | 1    | 0.37  | 0.71  | 0.92  | 0.67  |
| stx1a + stx2a (n = 13) | 1     | 1    | 1    | 1    | 0.38  | 0.62  | 1     | 0.38  |
| stx1a + stx2c (n = 7)  | 1     | 1    | 1    | 1    | 0.86  | 1     | 1     | 0.37  |
| stx1a + stx2d (n = 7)  | 1     | 1    | 1    | 1    | 0.43  | 1     | 0.71  | 0.43  |
| stx1c + stx2b (n = 23) | 1     | 1    | 1    | 1    | 0.17  | 0.61  | 0.91  | 0.87  |
| stx2 (n = 114)         | 0.99  | 1    | 0.99 | 0.76 | 0.23  | 0.75  | 0.75  | 0.3   |
| stx2a (n = 39)         | 1     | 1    | 1    | 0.92 | 0.31  | 0.51  | 0.9   | 0.28  |
| stx2b (n = 12)         | 0.92  | 1    | 0.92 | 1    | 0.17  | 0.17  | 0.5   | 0.92  | 0.83  |
| stx2c (n = 17)         | 1     | 1    | 1    | 0.88 | 0.41  | 1     | 0.82  | 0.47  |
| stx2d (n = 28)         | 1     | 1    | 1    | 0.46 | 0.07  | 0.96  | 0.32  | 0     |
| stx2e (n = 3)          | 1     | 1    | 1    | 0    | 0.33  | 0.33  | 0.67  | 0     |
| stx2g (n = 3)          | 1     | 1    | 1    | 0.67 | 0.33  | 1     | 1     | 0     |
| stx3 (n = 62)          | 0.73  | 0.73 | 0.73 | 0.68 | 0.1   | 0.27  | 0.56  | 0.11  |
| stx1a (n = 46)         | 0.63  | 0.63 | 0.63 | 0.61 | 0.07  | 0.17  | 0.59  | 0.07  |
| stx2c (n = 15)         | 1     | 1    | 1    | 0.93 | 0.2   | 0.53  | 0.47  | 0.27  |
| eae* (n = 102)         | 1     | 1    | 1    | 0.82 | 0.31  | 0.62  | 0.77  | 0.28  |
| eaeβ (n = 42)          | 1     | 1    | 1    | 1    | 0.17  | 1     | 0     | 0     |
| eaeγ (n = 31)          | 1     | 1    | 1    | 1    | 0.97  | 0.84  | 0.87  | 0.33  |
| eaeε (n = 6)           | 1     | 1    | 1    | 1    | 0.17  | 0.83  | 1     | 0     |
| eaeα (n = 17)          | 1     | 1    | 1    | 0.67 | 0    | 0.33  | 0.67  | 0     |
| eaeε (n = 126)         | 0.86  | 0.87 | 0.86 | 0.77 | 0.15  | 0.61  | 0.71  | 0.37  |

Values indicate the proportion of strains carrying ivi genes (columns) for each category (lines)
NB: only categories with a strain number ≥3 are shown.
We next evaluated the existence of a potential correlation between the presence of \textit{ivi} genes and genes encoding characterized virulence factors. As shown in Table 4, all eight \textit{ivi} genes were found to be more prevalent in strains harbouring \textit{stx2} or \textit{stx1} + \textit{stx2} than in strains carrying only \textit{stx1} gene (P < 0.001). In contrast, only genes \textit{mhpR}, \textit{ascG}, \textit{mdtM} and \textit{Z0964} were more prevalent in \textit{eae}-positive strains than in \textit{eae}-negative strains (P < 0.01). Interestingly, some \textit{ivi} genes were found to be highly associated with particular intimin variants. For example, \textit{Z0964} was detected in all 31 strains carrying \textit{eae} $\gamma$ whereas it was only found in 1 of the 71 strains carrying other intimin variants (Table 4; P < 0.001). We finally investigated the distribution of \textit{ivi} genes in strains for which we have clinical data in order to evaluate potential links between the presence of the 8 \textit{ivi} genes in EHEC and severity of the induced disease. From our collection, 34 strains have been isolated from patients who have developed a HUS, 33 strains from patients with bloody diarrhoea and 47 strains from patients with watery diarrhoea. As expected, HUS-associated strains carry \textit{stx2} and \textit{eae} genes more frequently than strains causing diarrhoea only (Table 5, p < 0.01). In contrast, \textit{stx1} is less prevalent in strains responsible for HUS. Regarding the eight \textit{ivi} genes, no significant association was determined with clinical symptoms except for \textit{Z0964}, which tends to be associated with HUS and/or bloody diarrhoea (P = 0.07).

\textbf{Involvement of \textit{ivi} genes in mouse gut colonization}

We next assessed the potential role of some \textit{ivi} genes in the gut colonization process. Seven \textit{ivi} genes have been arbitrarily selected based on their function potentially associated with bacterial fitness or virulence. This includes, for example, genes encoding putative adhesin or collagenase as well as genes encoding proteins involved in particular metabolic pathways. Of note, five of the seven selected genes belong to \textit{ivi} genes heterogeneously distributed in \textit{E. coli} and in EHEC strains as described above. We individually deleted the coding sequence of these seven \textit{ivi} genes in EDL933 and tested each mutant by a competition assay with the WT strain during mouse infection (Figure 4). As a control, we also performed competitive infection using a strain deleted for \textit{agaF}, which has been shown to be as efficient as the WT strain in colonizing mouse gut [34]. Deletion of genes \textit{ytfE}, \textit{yiiR}, \textit{Z3135}, \textit{Z4070} or \textit{Z4799} has no impact on EDL933 colonization efficiency since calculated competitive indices between the WT strain and each mutant were closed to 1. In contrast, inactivation of \textit{mhpR} or \textit{yhbu} altered the ability of EDL933 to colonize the gut since both mutants were significantly outcompeted by the WT strain (Figure 4).

\textbf{Discussion}

Since the first case of infection in the 1980s, the number of EHEC infections in humans has increased steadily and the pathogen is now recognized as a serious threat to human health. Stxs and T3SS are the main virulence factors produced by EHEC strains, but the
combinations of genetic traits that render a STEC strain highly virulent in humans are still poorly known. In this study, we performed RIVET screening using the EHEC O157:H7 reference strain EDL933 to identify genes that are induced in vivo, suggesting their potential role during the course of infection.

Among the 31 ivi genes identified, most are involved in adaptation of the pathogen to the gastrointestinal tract (GIT), as has been demonstrated in other studies [35,36]. First, genes involved in the metabolism of carbohydrates including arabinose (araB, araC), N-acetyl-galactosamine (agw) and β-glucosides (ascG) have been detected as ivi genes. Proteins from these three pathways were also shown to be specifically expressed by EHEC in infected patients [35]. Interestingly, C-source metabolic EHEC profiling has shown that arabinose and N-acetylglucosamine could be relevant for identification of recent diversification events among serogroups and could be linked with functional specialization for the most hazardous EHEC strains [37]. Additionally, the utilization of arabinose (but not N-acetyl-galactosamine) confers an advantage on EHEC in efficient colonization of the mouse gut [34]. ivi Genes also include genes required for the metabolism of amino acids (tryptophan, glutamate), vitamins (B6, B12), fatty acids and aromatic compounds. For the latter category, expression of gene mhpR is induced during mouse infection. We also showed that the WT EDL933 strain has a competitive advantage over the mhpR mutant in colonizing mouse intestine. MhpR is the transcriptional activator of the mhpABCDFE operon that encodes enzymes dedicated to the transformation of 3HPP into intermediates of the Krebs cycle [38,39]. 3HPP is found naturally in the GIT of animals including humans since it results from anaerobic bacterial metabolism of plant-derived flavonoids or aromatic amino acids. Taken together with a recent study demonstrating that mhp gene expression in EDL933 is higher after growth in the caecal content of human gut microbiota-associated rats than after growth in the caecal contents of axenic rats [36], these data strongly suggest that EDL933 utilizes gut microbiota-produced 3HPP to efficiently colonize the GIT. It is noteworthy that oxygen is required for 3HPP catabolism by Mhp enzymes, and it has been suggested that E. coli performs these transformation steps close to epithelial cells where oxygen diffuses from blood [38].

In addition to metabolic adjustment, EHEC adapts to the GIT through the expression of genes involved in stress responses. Indeed, genes encoding components of the efflux pumps MdtEF and MdtM as well as a putative HlyD family secretion protein (Yhil) were found to be induced in vivo. These membrane structures are specialized in the export of a variety of harmful molecules such as drugs [40]. Other stress response pathways activated in vivo correspond to those involved in defence against nitrosative stress, with overexpression of the genes mdtEF, ytfE, and norR. The efflux pump MdtEF has been shown to protect E. coli against nitrosative stress during anaerobic respiration [41] and YtfE is an iron-sulphur cluster repair protein highly produced in the presence of nitric oxide (NO) through a direct regulation via the NO-sensing transcriptional regulator, NsrR [42,43]. NorR is a second NO-sensing transcriptional regulator that controls the expression of norV which encodes a NO reductase [42,43]. NorV by itself was identified as one of the E. coli O157:H7 proteins expressed during human infection [35]. Moreover, it has been shown that NorV plays an important role for NO detoxification under anaerobic conditions as well as for EHEC survival and Stx2 production within macrophages [44,45]. Authors of these works proposed norV as a direct virulence determinant for the pathogenesis of EHEC. We and others have also demonstrated that NO strongly affects the expression of both stx and LEE genes in EHEC, and therefore modulates the expression of key virulence factors [46–48]. Altogether, these data argue that establishment of an adapted response to nitrosative stress is a key component of both the adaptation of EHEC to the GIT and the control of its virulence program.

Despite the use of several tnpR and RES cassette variants to cover a broad range of promoter strength, genes encoding key virulence genes such as stx and LEE genes were not revealed in our screening. This finding has been already observed in other studies identifying ivi genes from different pathogens [18,49] and demonstrates that RIVET screenings do not provide an exhaustive list of fitness or virulence genes for a given pathogen. In addition to possible bias during construction of the library, identified ivi genes rely on their expression level during both in vitro and in vivo growth and can therefore be affected by experimental parameters (culture medium, animal model, infection route, bacterial inoculum dose, spatial and temporal sampling, etc.). Among these possibilities, absence of stx and LEE genes in our list of ivi genes might be explained by the fact that these genes are known to be expressed during growth in laboratory conditions and were probably discarded during the selection of inactive promoters in vitro. Nonetheless, several identified ivi genes are linked to processes known to affect the production of these two cardinal virulence factors. Indeed, gene ydjM is known to be part of the LexA regulon and its expression is highly induced following activation of the SOS system [50]. Strikingly, proteins encoded by...
the SOS-induced genes *uvrB* and *uvrY* have been detected as immunogenic O157:H7 proteins in HUS patients [35]. This converges towards an activation of the SOS response during infection and, as a consequence, induction of Stx phage lytic cycle and Stx2 production and release [51]. [46,47,52] Our list of *ivi* genes also includes *btuR* encoding a cobalamin adenyltransferase involved in the assimilation of vitamin B<sub>12</sub> [53] and *yjbB* encoding a transporter of inorganic phosphate [54]. Vitamin B<sub>12</sub> and inorganic phosphate were both identified as molecules affecting the expression of Stx2 synthesis in EHEC [55,56].

Among the 31 identified *ivi* genes, 8 belong to the accessory genome as they were recovered in 22% to 93% of the 228 EHEC strains circulating in France between 2015 and 2016. Among these eight *ivi* genes, three (*mhpR, ascG*, and *mdtM*) were absent only in O117:H7 strains. Three other genes (*yjiR, Z4799*, and *Z4070*) presented a population-based distribution, suggesting gene acquisition by an ancestor and vertical transmission. Distribution of the last two genes (*Z0964* and *Z3135*) is very heterogeneous among EHEC strains and suggests the occurrence of multiple acquisition events by different strains in the course of evolution. We also investigated the co-distribution of the eight *ivi* genes with known virulence genes. All *ivi* genes were more prevalent in strains harboring *stx2* (alone or associated with *stx1*) than in strains carrying only *stx1* gene (*P* < 0.001) and four of them (*mhpR, ascG, mdtM*, and *Z0964*) were more prevalent in *eae*-positive strains than in *eae*-negative strains (*P* < 0.01). Since epidemiological surveys determined that EHEC strains producing Stx2 and T3SS are more associated with the risk of HUS development [10], this raises the question whether one or more of these *ivi* genes contribute to the virulence potential of highly pathogenic EHEC strains. We were not able to make significant associations between *ivi* genes and clinical symptoms using our EHEC collection. However, this may be due to the paucity of clinical information collected in our study and future investigations need to be carried out to address this question. Among the eight *ivi* genes heterogeneously distributed in EHEC, three of them encode well-described or putative transcription factors (TF). As already discussed above, *mhpR* is the transcriptional activator of the *mhpABCDDE* operon involved in 3HPP catabolism. The second TF is *AscG*, a GalR-type TF which represses *ascFB* operon required for the transport and catabolism of arbutin, salicin, and cellobiose [57]. *AscG* has been shown to repress the *prpBCDDE* operon involved in propionate catabolism [58]. Identification of the *mhpR* and *ascG* genes as *ivi* genes in our experiments suggests that EHEC would favour the utilization of 3HPP in the gut while limiting the utilization of arbutin, salicin, cellobiose, and propionate. The third TF, *YjiR*, encodes a putative transcriptional regulator of the MocR family as it possesses a clear Helix-Turn-Helix domain and an aspartate aminotransferase domain. While most regulators of the MocR family are still uncharacterized, the archetypal protein of this family, GabR, has been characterized to induce expression of the *gabTD* operon and thus to confer on *Bacillus subtilis* the ability to utilize γ-amino-butyric acid (GABA) as a nitrogen source [59]. While *YjiR* has only 24% identity to GabR, our RIVET screening also identified the gene *gabT*, which encodes GABA aminotransferase, suggesting the use of GABA by EDL933 in the GIT. However, co-infection assays demonstrated that deletion of *yjiR* does not alter the fitness of EDL933 in the GIT of infected mice. Nonetheless, all these data provide information on several substrates that could constitute specific nutrients used by EHEC to efficiently colonize the GIT. Several other genes belonging to the eight *ivi* genes heterogeneously distributed in EHEC can be correlated with virulence potential. First, gene *Z0964* is located within prophage CP-933K that also includes genes *nleB, nleC, nleD* and *nleH1* encoding several known effectors of the T3SS [60–62]. Second, gene *Z3135* encodes a large protein (2661 amino acids) of unknown function that contains multiple repeated domains with an Ig-like fold. These motifs are largely found in bacterial surface proteins belonging to the inverse autotransporter family such as intimins and invasins and are often involved in bacterial pathogenicity [63]. Finally, *Z4799* encodes a putative DNA-processing protein belonging to a widespread bacterial protein family sharing a highly conserved domain called pfam02481. It is of interest that two proteins of this family, DprA from *Streptococcus pneumoniae* and Smf from *B. subtilis*, have been reported to play a prominent role in natural transformation through binding to exogenous single-stranded DNA and direct interaction with recombinase RecA [64]. Whereas *E. coli* is not a naturally transformable bacterial species, it is tempting to speculate that *Z4799* can participate in activation of the SOS system in the case of DNA damage and thus can modulate the production level of Shiga toxins associated with RecA activation.

In this study, we performed genetic screening to identify genes from the EHEC O157:H7 strain EDL933 that are over-expressed *in vivo* during mouse infection. Among the 31 identified genes, a large proportion is involved in metabolic functions or in stress responses. The remaining *ivi* genes have hypothetic or unknown functions and several of them are potentially associated with EHEC virulence or fitness. We also demonstrated that eight *ivi* genes are heterogeneously
distributed in EHEC strains circulating in France these last years. These genes are more prevalent in strains from the TOP seven priority serotypes, particularly strains carrying established virulence genes such as stx and eae. This work sheds further light on genes encoding niche factors that may increase the potential of STEC strains to cause disease in humans.

Acknowledgments

We thank Cindy Garnier, Luc Planeix, Mélanie Lemor and Valérie Rousseau for excellent technical assistance, Charlotte Renaudat for her help in statistical analysis and David Marsh for language editing. We acknowledge all French medical laboratories participating in the clinical network of Shiga-toxin producing \textit{Escherichia coli}. We also thank the NRL team for providing the food and animal \textit{E. coli} strains. This work was funded by INRA, Institut Pasteur, Sante Publique France and the French Ministry of Agriculture.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Institut Pasteur; Ministère de l’Agriculture et de l’Alimentation; Santé Publique France; INRA.

ORCID

Estelle Loukiadis http://orcid.org/0000-0002-1631-6670
Grégory Jubelin http://orcid.org/0000-0002-9902-0663

References

[1] Schmidt MA. LEWays: tales of EPEC, ATEC and EHEC. Cell Microbiol. 2010;12(11):1544–1552.
[2] Hartland EL, Batchelor M, Delahay RM, et al. Binding of intimin from enteropathogenic \textit{Escherichia coli} to Tir and to host cells. Mol Microbiol. 1999;32(1):151–158.
[3] Ooka T, Seto K, Kawano K, et al. Clinical significance of \textit{Escherichia albertii}. Emerg Infect Dis. 2012;18(3):488–492.
[4] Phillips AD, Frankel G. Intimin-mediated tissue specificity in enteropathogenic \textit{Escherichia coli} interaction with human intestinal organ cultures. J Infect Dis. 2000;181(4):1496–1500.
[5] Zoja C, Buelli S, Morrigi M. Shiga toxin-associated hemolytic uremic syndrome: pathophysiology of endothelial dysfunction. Pediatr Nephrol. 2010;25(11):2231–2240.
[6] Melton-Celsa AR. Shiga Toxin (Stx) classification, structure, and function. Microbiol Spectr. 2014;2(4):EHEC-0024–2013.
[7] Siegler RL, Obrig TG, Pysher TJ, et al. Response to Shiga toxin 1 and 2 in a baboon model of hemolytic uremic syndrome. Pediatr Nephrol. 2003;18(2):92–96.
[8] Bielaszewska M, Friedrich AW, Aldick T, et al. Shiga toxin activatable by intestinal mucus in \textit{Escherichia coli} isolated from humans: predictor for a severe clinical outcome. Clin Infect Dis. 2006;43(9):1160–1167.
[9] Karmali MA, Mascarenhas M, Shen S, et al. Association of genomic O island 122 of \textit{Escherichia coli} EDL 933 with verocytotoxin-producing \textit{Escherichia coli} seropathotypes that are linked to epidemic and/or serious disease. J Clin Microbiol. 2003;41(11):4930–4940.
[10] Boerlin P, McEwen SA, Boerlin-Petzold F, et al. Associations between virulence factors of Shiga toxin-producing \textit{Escherichia coli} and disease in humans. J Clin Microbiol. 1999;37(3):497–503.
[11] Messens W, Bolton D, Frankel G, et al. Defining pathogenic verocytotoxin-producing \textit{Escherichia coli} (VTEC) from cases of human infection in the European Union, 2007–2010. Epidemiol Infect. 2015;143(8):1652–1661.
[12] McWilliams BD, Torres AG. Enterohemorrhagic \textit{Escherichia coli} adhesins. Microbiol Spectr. 2014;2(3).
[13] Lim JY, Yoon J, Hovde CJ. A brief overview of \textit{Escherichia coli} O157: H7 and its plasmid O157. J Microbiol Biotechnol. 2010;20(1):5–14.
[14] Perna NT, Plunkett G 3rd, Burland V, et al. Genome sequence of enterohaemorrhagic \textit{Escherichia coli} O157: H7. Nature. 2001;409(6819):529–533.
[15] Hayashi T, Makino K, Ohnishi M, et al. Complete genome sequence of enterohemorrhagic \textit{Escherichia coli} O157: H7 and genomic comparison with a laboratory strain K–12. DNA Res. 2001;8(1):11–22.
[16] Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in \textit{Escherichia coli} K–12 using PCR products. Proc Natl Acad Sci U S A. 2000;97(12):6640–6645.
[17] Bertin Y, Chaucheyras-Durand F, Robbe-Masselot C, et al. Carbohydrate utilization by enterohaemorrhagic \textit{Escherichia coli} O157: H7 in bovine intestinal content. Environ Microbiol. 2013;15(2):610–622.
[18] Osorio CG, Crawford JA, Michalski J, et al. Second-generation recombination-based in vivo expression technology for large-scale screening for \textit{Vibrio cholerae} genes induced during infection of the mouse small intestine. Infect Immun. 2005;73(2):972–980.
[19] Crepin S, Harel J, Dozois CM. Chromosomal complementation using Tn7 transposon vectors in \textit{Enterobacteriaceae}. Appl Environ Microbiol. 2012;78(17):6001–6008.
[20] Karlyshev AV, Pallen MJ, Wren BW. Single-primer PCR procedure for rapid identification of transposon insertion sites. Biotechniques. 2000;28(6):1078,1080,1082.
[21] Criscuolo A, Brisse S, AlienTrimmer removes adapter oligonucleotides with high sensitivity in short-insert paired-end reads. Commentary on Turner (2014) Assessment of insert sizes and adapter content in FASTQ data from NexteraXT libraries. Front Genet. 2014;5:130.
[22] Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–477.
[23] Beghain J, Bridier-Nahmiás A, Le Nagard H, et al. ClermonTyping: an easy-to-use and accurate in silico method for Escherichia genus strain phylotyping. Microb Genom. 2018;4(7).

[24] Treangen TJ, Ondov BD, Koren S, et al. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol. 2014;15(11):524.

[25] Tamura K, Stecher G, Peterson D, et al. MEGA6: molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30(12):2725–2729.

[26] Letunic I, Bork P. Interactive tree of life (iToL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44(W1):W24–245.

[27] Jones SA, Jorgensen M, Chowdhury FZ, et al. Glycogen and maltose utilization by Escherichia coli O157: h7 in the mouse intestine. Infect Immun. 2008;76(6):2531–2540.

[28] Torres AG, Payne SM. Haem iron-transport system in enterohaemorrhagic Escherichia coli O157:H7. Mol Microbiol. 1997;23(4):825–833.

[29] Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. Science. 1997;278(5338):631–637.

[30] Galperin MY, Makarova KS, Wolf YI, et al. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 2015;43(Database issue):D261–269.

[31] The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2016. Efsa J. 2017;15(12):e05077.

[32] Soysal N, Mariani-Kurkdjian P, Smail Y, et al. Enterohaemorrhagic Escherichia coli hybrid pathotype O80: H2as a new therapeutic challenge. Emerg Infect Dis. 2016;22(9):1604–1612.

[33] De Rauw K, Thiry D, Caljon B, et al. Characteristics of Shiga toxin producing- and enteropathogenic Escherichia coli of the emerging serotype O80: H2isolated from humans and diarrhoeic calves in Belgium. Clin Microbiol Infect. 2018;24(11):e111.e5–111.e8.

[34] Fabich AJ, Jones SA, Chowdhury FZ, et al. Comparison of carbon nutrition for pathogenic and commensal Escherichia coli strains in the mouse intestine. Infect Immun. 2008;76(3):1143–1152.

[35] John M, Kudva IT, Griffin RW, et al. Use of in vivo-induced antigen technology for identification of Escherichia coli O157: H7proteins expressed during human infection. Infect Immun. 2005;73(5):2665–2679.

[36] Le Bihan G, Jubelin G, Garneau P, et al. Transcriptome analysis of Escherichia coli O157: H7grown in vitro in the sterile-filtrated cecal content of human gut microbiota associated rats reveals an adaptive expression of metabolic and virulence genes. Microbes Infect. 2015;17(1):23–33.

[37] Kerangart S, Gournoyer B, Loukiadis E. C-source metabolic profilings of foodborne Shiga-toxin producing E. coli match serogroup differentiations and highlight functional adaptations. Int J Food Microbiol. 2018;266:324–336.

[38] Ferrandez A, Garcia JL, Diaz E. Genetic characterization and expression in heterologous hosts of the 3-(3-hydroxyphenyl)propionate catabolic pathway of Escherichia coli K-12. J Bacteriol. 1997;179(8):2573–2581.

[39] Torres B, Porras G, Garcia JL, et al. Regulation of the mhp cluster responsible for 3-(3-hydroxyphenyl)propionic acid degradation in Escherichia coli. J Biol Chem. 2003;278(30):27575–27585.

[40] Venter H, Mowlà R, Ohene-Agyei T, et al. RND-type drug f 11ux pumps from Gram-negative bacteria: molecular mechanism and inhibition. Front Microbiol. 2015;6:377.

[41] Zhang Y, Xiao M, Horiyama T, et al. The multidrug efflux pump MdtEF protects against nitrosative damage during the anaerobic respiration in Escherichia coli. J Biol Chem. 2011;286(30):26576–26584.

[42] Justino MC, Almeida CC, Goncalves VL, et al. Escherichia coli YtFE is a di-iron protein with an important function in assembly of iron-sulphur clusters. FEMS Microbiol Lett. 2006;257(2):278–284.

[43] Hutchings MI, Mandhana N, Spiro S. The NorR protein of Escherichia coli activates expression of the flavoredbroxin gene norV in response to reactive nitrogen species. J Bacteriol. 2002;184(16):4640–4643.

[44] Shimizu T, Tsutsuki H, Matsumoto A, et al. The nitric oxide reductase of enterohaemorrhagic Escherichia coli plays an important role for the survival within macrophages. Mol Microbiol. 2012;85(3):492–512.

[45] Shimizu T, Hirai S, Yokoyama E, et al. An evolutionary analysis of nitric oxide reductase gene norV in enterohemorrhagic Escherichia coli O157. Infect Genet Evol. 2015;33:176–181.

[46] Brancu P, Matrat S, Vareille M, et al. NsrR, GadE, and GadX interplay in repressing expression of the Escherichia coli O157: H7LEE pathogenicity island in response to nitric oxide. PLoS Pathog. 2014;10(1):e1003874.

[47] Vareille M, de Sablet T, Hindre T, et al. Nitric oxide inhibits Shiga-toxin synthesis by enterohemorrhagic Escherichia coli. Proc Natl Acad Sci U S A. 2007;104(24):10199–10204.

[48] Ichimura K, Shimizu T, Matsumoto A, et al. Nitric oxide-enhanced Shiga toxin production was regulated by Fur and RecA in enterohemorrhagic Escherichia coli O157. MicrobiologyOpen. 2017;6(4).

[49] Camilli A, Mekalanos JJ. Use of recombinase gene fusions to identify Vibrio cholerae genes induced during infection. Mol Microbiol. 1995;18(4):671–683.

[50] Fernandez De Henestrosa AR, Ogi T, Aoyagi S, et al. Identification of additional genes belonging to the LexA regulon in Escherichia coli. Mol Microbiol. 2000;35(6):1560–1572.

[51] Kimmitt PT, Harwood CR, Barer MR. Toxin gene expression by shiga toxin-producing Escherichia coli: the role of antibiotics and the bacterial SOS response. Emerg Infect Dis. 2000;6(5):458–465.

[52] Bogdan C. Nitric oxide synthase in innate and adaptive immunity: an update. Trends Immunol. 2015;36(3):161–178.

[53] Rodionov DA, Viteschak AG, Mironov AA, et al. Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes. J Biol Chem. 2003;278(42):41148–41159.

[54] Motomura K, Hirota R, Ohnaka N, et al. Overproduction of YjbB reduces the level of polyphosphate in Escherichia coli: a hypothetical role of YjbB in phosphate export and polyphosphate accumulation. FEMS Microbiol Lett. 2011;320(1):25–32.
[55] Chekabab SM, Jubelin G, Dozois CM, et al. PhoB activates *Escherichia coli* O157: H7 virulence factors in response to inorganic phosphate limitation. PLoS One. 2014;9(4):e94285.

[56] Cordonnier C, Le Bihan G, Emond-Rheault JG, et al. Vitamin B₁₂ uptake by the gut commensal bacteria *Bacteroides thetaiotaomicron* limits the production of Shiga Toxin by enterohemorrhagic *Escherichia coli*. Toxins (Basel). 2016;8(1).

[57] Hall BG, Xu L. Nucleotide sequence, function, activation, and evolution of the cryptic asc operon of *Escherichia coli* K12. Mol Biol Evol. 1992;9(4):688–706.

[58] Ishida Y, Kori A, Ishihama A. Participation of regulator AscG of the beta-glucoside utilization operon in regulation of the propionate catabolism operon. J Bacteriol. 2009;191(19):6136–6144.

[59] Belitsky BR, Sonenshein AL. GabR, a member of a novel protein family, regulates the utilization of gamma-aminobutyrate in *Bacillus subtilis*. Mol Microbiol. 2002;45(2):569–583.

[60] Marches O, Wiles S, Dziva F, et al. Characterization of two non-locus of enterocyte effacement-encoded type III-translocated effectors, NleC and NleD, in attaching and effacing pathogens. Infect Immun. 2005;73(12):8411–8417.

[61] Holmes A, Lindestam Arlehamn CS, Wang D, et al. Expression and regulation of the *Escherichia coli* O157: H7 effector proteins NleH1 and NleH2. PLoS One. 2012;7(3):e33408.

[62] Gao X, Wang X, Pham TH, et al. NleB, a bacterial effector with glycosyltransferase activity, targets GAPDH function to inhibit NF-kappaB activation. Cell Host Microbe. 2013;13(1):87–99.

[63] Leo JC, Oberhettinger P, Schutz M, et al. The inverse autotransporter family: intimin, invasin and related proteins. Int J Med Microbiol. 2015;305(2):276–282.

[64] Mortier-Barriere I, Velten M, Dupaïgne P, et al. A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. Cell. 2007;130(5):824–836.