Master's thesis

2019

Janani Jeyachandran

Master's thesis

NTNU
Norwegian University of
Science and Technology
Faculty of Medicine and Health Sciences
Department of Clinical and Molecular Medicine

Janani Jeyachandran

Role of putative adherence factor in Shiga toxin producing *Escherichia coli* (STEC)

June 2019

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

Role of putative adherence factor in Shiga toxin producing *Escherichia coli* (STEC)

Janani Jeyachandran

Molecular medicine
Submission date: June 2019
Supervisor: Jan Egil Afset
Co-supervisor: Christina Gabrielsen Ås
Kjersti Haugum

Norwegian University of Science and Technology
Department of Clinical and Molecular Medicine

Abstract

BACKGROUND: Infection with Shiga toxin producing *Escherichia coli* (STEC) may cause severe disease like haemorrhagic uremic syndrome (HUS). In addition to Shiga toxins, the ability to adhere to intestinal epithelial cells is important for STEC to cause severe disease. Such adherence is mostly due to intimate attachment mediated by the *eae* encoded intimin. However, other adhesins may also play a role in adhesion. This master´s project was part of a project designed to study the role of the putative fimbrial operon lpf_{B1} in adhesion and HUS. The aim of this project was to knockout the *eae* gene in a clinical STEC strain using the CRISPR/Cas9 editing tool.

METHODS: We selected a STEC strain from a HUS outbreak, FHI11, which harboured both the *eae* gene and the lpf_{B1} operon but had lost its *stx* genes. The experiments were done according to a two-plasmid CRISPR/Cas9 protocol for *E. coli*, with some modifications. The *cas9* gene was located on a pCas9 plasmid, and the guide DNA (gDNA) was located on a pCRISPR plasmid. A gDNA was designed to guide the Cas9 to insert a double strand break (DSB) at a specific location in the *eae* gene. Two alternative repair DNA templates were designed to introduce a specific deletion at the DSB with a premature stop codon to knockout the *eae* gene. One of these should give a deletion of 201 bp and the other a deletion of 33 bp.

RESULTS: The pCRISPR plasmid containing the gRNA (pCRISPR_T30RNA) was successfully constructed and the assembly was confirmed by PCR and Sanger sequencing. The pCRISPR_T30RNA plasmid and the repair DNA template was then successfully transformed by electroporation into the STEC strain FHI11 harbouring the pCas9 plasmid. Transformants selected using kanamycin and chloramphenicol were then analysed by colony PCR. In the mutagenesis experiment targeted to delete 201 bp, only two colonies found to have an intact *eae* gene were recovered. In contrast, in the mutagenesis experiment targeted to delete 33 bp, 30 colonies were recovered. However, all but two of these had an intact *eae* gene. The two remaining colonies might have undergone a larger deletion than expected. Due to time limitation further analyses of the strains from these two colonies were not performed.

CONCLUSION: The most likely reason that we did not succeed in introducing the specific mutation in the STEC strain to knockout the *eae* gene is that the homologous directed repair mechanism was not as efficient as expected. Further optimization of the method is therefore necessary.

Acknowledgements

This master's thesis in Molecular medicine was carried out during the years 2018-2019. The practical work has been carried out at the Department of Clinical and Molecular Medicine, Faculty of medicine and health sciences, Norwegian University of Science and Technology (NTNU).

First, I would like to express my deepest gratitude to my main supervisor Professor Jan Egil Afset. I have been extremely lucky to have a supervisor who cared so much about my work, and who responded to my questions. Without your guidance and constant feedback this masters' thesis would not have been achievable.

I would also like to thank my co-supervisors Kjersti Haugum and Christina Gabrielson Ås. You both provided excellent guidance and time for discussions and gave suggestions for the development of this project.

I would also like to thank Hilde Lysvand and Kirsti Løseth, at the Department of Clinical and Molecular Medicine, NTNU, for providing the initial laboratory tour, explaining the rules and regulations and always being helpful. I owe my special thanks to Hilde Lysvand for providing necessary demonstrations, helping in ordering the reagents needed for this study, and always being helpful with questions regarding laboratory work. A special thanks to all the staffs at the Department of Microbiology at St. Olav's hospital for helping me with autoclaving and being helpful. I would also like to thank Helene Marie Flatby for sharing her experience she gained during her master's thesis and helping me using Geneious software.

I would like to thank my parents and friends for providing me the support and continuous encouragement throughout my years of study. The journey from Sri Lanka to Norway would not have been possible without their support. Last, but not least, I pay my obeisance to GOD, the almighty to have bestowed upon me good health, courage and inspiration.

Janani Jeyachandran

Trondheim, June 2019

Contents

Abbreviations

A	Adenine
A/E	Attaching and effacing lesion
A-EJ	Alternative end joining
BLAST	Basic local alignment search tool
bp	Base pair
C	Cytosine
CFU	Colony forming unit
CmR	Chloramphenicol resistance
CRISPR	Clustered regularly interspaced short palindromic repeats
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
ds	Double stranded
DSB	Double stranded break
eae	Gene encoding intimin
E. coli	*Escherichia coli*
G	Guanine
gDNA	Guide DNA
gRNA	Guide RNA
HC	Haemorrhagic Colitis
HDR	Homology directed repair
HUS	Haemolytic uremic syndrome
KanR	Kanamycin resistance
Kb	Kilobase
LB	Luria-Bertani
LEE	Locus of enterocyte effacement
Lpf_{B1}	Long polar fimbria-B1
MGW	Molecular grade water
ml	Millilitre
MWM	Molecular weight marker

μl	Microlitre
NCBI	National centre for biotechnology information
NEB	New England Biolabs
NHEJ	Non homologous end joining
nt	Nucleotide/Nucleotides
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
ssDNA	Single stranded DNA
STEC	Shiga toxin producing *Escherichia coli*
Stx	Shiga toxin
stx	Gene encoding Shiga toxin
T3SS	Type III secretion system
T	Thymine
Tir	Translocated intimin receptor
tir	Gene encoding Tir
T_M	Melting temperature

1 Introduction

1.1 *Escherichia coli*

Escherichia coli are gram-negative, rod shaped, facultative anaerobic bacteria that are commonly found as a harmless commensal in the gut of human and warm-blooded animals. This species was first reported by the German microbiologist and paediatrician Theodor Escherich in 1885 [1]. *E. coli* normally reside in the large intestine of warm-blooded animals and in general do not survive well outside the gastrointestinal tract [1]. Therefore, the presence of *E. coli* in environmental samples, food and water is used as an indicator to detect faecal contamination [2].

The common categorization system of *E. coli* is based on serotyping. Traditionally, *E. coli* stains have been serotyped using antisera against the O-, H-, and K- antigens. The O-antigens are heat stable lipopolysaccharides anchored in the cell wall. The H-antigens are heat labile peritrichous flagella and the K-antigens are located in the polysaccharide capsule possessed by some *E. coli*. So far approximately 186 O-antigens and 53 H-antigens have been identified [3].

1.2 Pathogenic *E. coli*

Commensal *E. coli* do normally not have any harmful effects in healthy human individuals. However, such bacteria may cause disease in immunocompromised patients and individuals with anomalies e.g. in the urinary tract or in case the barrier of the gastrointestinal tract is broken [3, 4]. Some strains of *E. coli* that have acquired virulence factors through horizontal transfer on plasmids, transposons, or bacteriophages, or as part of a pathogenicity island. Such strains are capable of causing broad spectrum of diseases, both in the gastrointestinal tract and as extraintestinal infections [3]. *E. coli* that encode virulence factors making them capable of causing infections in humans are assigned to pathotypes depending on virulence repertoire and type of disease. Strains causing diarrhoea are collectively known as diarrhoeagenic *E. coli* (DEC) and the strains causing infection outside the gastrointestinal tract are termed extraintestinal *E. coli* (ExPEC) [3]. Two main pathotypes of ExPEC are recognized: meningitis associated *E. coli* (MNEC) and uropathogenic *E. coli* (UPEC) causing meningitis and urinary tract infection, respectively [3]. DEC are classified in seven pathotypes, based on the pathology they induce during infection and the virulence factors they express;

enteroinvasive *E. coli* (EIEC), enteroaggregative *E. coli* (*EAE*C), diffusely adherent *E. coli* (DAEC), adherent invasive *E. coli* (AIEC), enterotoxigenic *E. coli* (ETEC), enteropathogenic *E. coli* (EPEC), and Shiga toxin producing *E. coli* (STEC) [3].

EIEC are closely related to *Shigella* spp. in their virulence factors and other phenotypic properties. Like *Shigella*, EIEC invade colonic epithelial cells and may cause bacillary dysentery, characterized by the presence of blood, mucus and leucocytes in the stool. EIEC infection is usually transmitted by contaminated food and water via the faecal-oral route [1].

EAEC are the second most common causative agent of traveller's diarrhoea. In addition, EAEC strains cause severe diarrhoea in HIV patients, and acute watery diarrhoea in infants and young children [5, 6]. EAEC are defined by their characteristic stacked brick-like aggregative adherence pattern to the mucosal surface of the host intestine. Following colonization, EAEC produce increased amounts of mucus that encrusts them on the surface of intestinal cells. EAEC also produce enterotoxins and cytotoxins, which stimulate the inflammatory response, mucosal toxicity and intestinal secretion that leads to diarrhoea [6].

DAEC are characterized by a diffuse adherent phenotype, in which the entire cell surface is evenly covered by bacteria due to the action of afimbrial (Afa) or fimbrial adhesins (Afa-Dr). Unlike other categories of DEC, the potential of DAEC to cause diarrhoea remains controversial. Therefore, little is known about the pathogenicity induced by DAEC [7].

AIEC have been detected in 30% of patients who are affected by Crohn's disease. In addition, these bacteria have been associated with diarrhoea in adults and children. AIEC are genetically related to ExPEC strains. AIEC cause pathogenesis by adhering to host intestinal epithelial cells and invade through the cells to reach macrophages, where the bacteria replicate without inducing cell death [1].

ETEC are the most common causative agents of travellers and childhood diarrhoea in developing countries. Colonizing factors and secretion of enterotoxins cause watery diarrhoea, which can range from mild to severe disease. The attachment to host intestinal mucosa is facilitated by surface fimbria, known as colonization factor antigen (CFA).ETEC can produce two types of enterotoxins heat-labile and/or heat-stable enterotoxins [1, 4].

EPEC are associated with infant diarrhoea in developing countries [1]. EPEC are considered to be a non-invasive bacterium and the main hallmarks of pathogenesis are forming attaching and effacing (A/E) lesion and localized adherence to the epithelial cells. In the A/E lesion,

EPEC attach tightly to the intestinal epithelial cell membrane and disrupt the microvillus leading to malabsorption and diarrhoea. Localized adherence is mediated by the bundle forming pilus (BFP) encoded by *bfp* genes located on the EPEC Adherence Factor (EAF) plasmid. Strains of EPEC that harbour the Locus of enterocyte effacement (LEE) pathogenicity island and lack the EAF plasmid are termed atypical EPEC (aEPEC), while strains that harbour both the LEE pathogenicity island and EAF plasmid are termed typical EPEC (tEPEC) [4]. Atypical EPEC are closely related to STEC and share many virulence determinants. Like EPEC, the main virulence property of STEC is the ability to create A/E lesions in intestinal epithelial cells. EPEC is distinguished from STEC by their lack of Shiga toxins [1].

1.3 STEC

E. coli that produce Shiga toxins are known as STEC [1]. STEC may harbour genes expressing either Shiga toxin 1 (Stx1) or Shiga toxin 2 (Stx2), or both. STEC are also known as verocytotoxin producing *E. coli* (VTEC), based on the cytotoxic effect on the Vero cells, a kidney cell line from an African green monkey [8].

STEC have been associated with food-borne gastroenteritis in humans, both sporadic cases and outbreaks. STEC cause disease ranging from asymptomatic carriage to haemorrhagic colitis (HC) and haemorrhagic uraemic syndrome (HUS). STEC which can cause HC and HUS in humans are often termed enterohemorrhagic *E. coli* (EHEC) [9, 10]. HUS is a clinical condition characterized by microangiopathic haemolytic anaemia, thrombocytopenia, and acute renal damage [11]. Most infections start with non-bloody diarrhoea which resolves without treatment. Less commonly, infected patients develop bloody diarrhoea or HC, and 5-15% of patients infected progress to HUS [12]. Children less than 5 years old and elderly persons are most commonly affected by severe complications [10].

The major reservoir of STEC are cattle which harbour the bacteria as part of the normal gastrointestinal flora [13]. In most cases, cattle harbouring STEC remain free of disease, although STEC have been associated with fatal disease in new born calves, as well as pigs and poultry [14, 15].

In addition to the Shiga toxins, STEC strains may harbour a variety of putative virulence factors with varying characteristics. Like in EPEC, many STEC strains contain the LEE with

the adhesin intimin and form A/E lesions. Rarely hybrid strains containing elements from more than one pathotype are seen. Such a strain, a hybrid between EAEC and STEC of serotype O104:H4, which expressed Shiga toxins, caused the largest outbreak with HUS so far, in Germany in 2011 [16]. It has been observed both in outbreak situations and in sporadic cases of HUS that STEC isolates may loose their *stx* genes (suggest to be named EHEC-LST, lost Shiga toxin). Since *stx*-negative STEC fits the definition of aEPEC, it may be difficult to differentiate between aEPEC and EHEC-LST, which often must be inferred from whatever clinical and epidemiological information that may be available. Loss of *stx* genes may occur during infection process, during transport of the sample or during culturing of the strain in the laboratory [17].

1.3.1 Pathogenesis

STEC cause disease through a multi-step process starting with initial colonization, evasion of host defences, multiplication, and host damage due to toxin production (Figure 1) [3]. A variety of virulence genes present on large virulence plasmids or genomic islands have been suggested to encode adherence factors, toxins, proteases, and flagellin involved in STEC pathogenesis [7].

In general, like other strains of *E. coli*, STEC overcome gastric acidity through acid resistance and establish themselves in the intestines [18]. In humans, STEC predominantly bind to the terminal ileum and colon [19]. In contrast, in animal models like gnotobiotic piglets, infant rabbits and neonatal calves STEC colonize the terminal ileum, cecum and colon [14, 20, 21]. It has been hypothesized that STEC binding to Payer's patches in the distal ileum provides an initial site for the bacteria before establishing colonization of other parts of the gut. In comparison, EPEC has been reported to colonize the small intestine and part of the colon [19].

The principle virulence factor associated with HUS is Shiga toxins, which can cause microvascular endothelial injury in the kidneys. Due to the damage on glomerular structure, platelets in the circulation will become activated and initialize the coagulation cascade which subsequently induces an inflammatory response. Formation of localized thrombi (blood clot formed within the blood vessel) can result in thrombocytopenia and microangiopathic haemolytic anaemia. The resulting impairments contribute to acute renal failure characteristic

of HUS [22]. Shiga toxins can also bind to cells of the central nervous system and cause functional disturbances of the CNS including epileptic seizures [23].

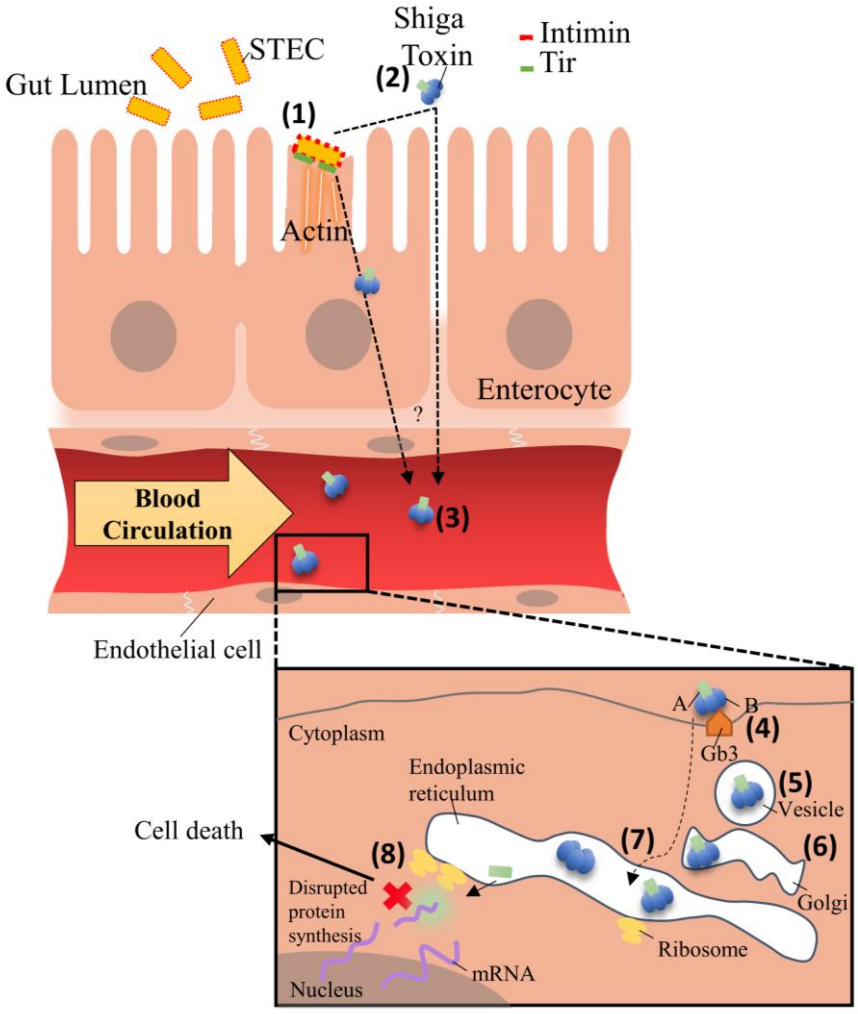


Figure 1. Pathogenesis of STEC infection. (1) STEC interact with the human intestinal epithelial cell via A/E formation. (2) Shiga toxin production occurs in the intestinal tract and (3) translocation of the Shiga toxins across the intestinal lumen to the blood stream results in its distribution to target organs, such as kidneys. In the target organ, (4) the B subunit of Shiga toxin binds to the receptor globotriaosylceramide (Gb3). (5) Shiga toxins enter the cell by endocytosis. (6) The Shiga toxins are reversely translocated into the Golgi bodies and (7) from there toxins get translocated into the endoplasmic reticulum (ER). (8) The A subunit of Shiga toxin inhibits protein synthesis and leads to cell death [24, 25] .

In addition to Shia toxin, factors associated with adherence also play an important role in STEC pathogenesis [26]. STEC require surface-localized adhesins to tightly adhere to the intestine and eventually establish disease. Further, because adhesins bring the bacteria closer to the epithelia, they play such an important role in virulence [27]. STEC contain large number of proteins essential for adhesion [27]. Some well-known adhesins and their locations are shown in table 1, and some of these will be described in more detail in the following sections.

Table 1. Adhesins found in STEC.

Gene	Virulence factor	Function or proposed role	Localization	Reference
eae	Intimin	Adhesin	LEE pathogenicity island	[28]
efa-1/lifA	Efa	Adhesin	OI-122	[29]
hcpA	(Haemorrhagic coli pilus) HCP	Fimbrial adhesin	Chromosome	[30]
iha	IrgA homologue adhesin (Iha)	Adhesin	Tellurite resistance- and adherence conferring island	[31]
lpf	Long polar fimbriae (Lpf)	Fimbrial adhesin	OI -141 and OI-154	[32]
ycbQRST	*E. coli* YcbQ laminin-binding fimbria (ELF)	Fimbrial adhesin	Chromosome	[33]
paa	Porcine attaching and effacing associated factor (Paa)	Adhesin	OI-57	[34]
saa	STEC auto agglutinating adhesin	Adhesin	Plasmid	[35]
sfp	Sfp	Adhesin	pSFO157	[36]
toxB	ToxB	Adhesin	pO157	[37]

1.3.1.1 Shiga toxins

The Shiga toxins are the essential virulence factors in the pathogenesis of HC and HUS [1]. There are two types of Shiga toxins, Stx1 and Stx2, which are encoded by *stx1* and *stx2* genes, respectively. The two major types of Shiga toxin are further classified into several subtypes, Stx1 consists of 3 subtypes (a, c, and d), and Stx2 of 7 subtypes (a to h) [1, 38]. Stx1 is rarely associated with HUS, while for Stx2, Stx2a is the most frequent subtype associated with HUS [39]. The Shiga toxins consist of two subunits: the enzymatic subunit A and a cell binding subunit B, which are organized as a pentamer (Figure 2) [40].

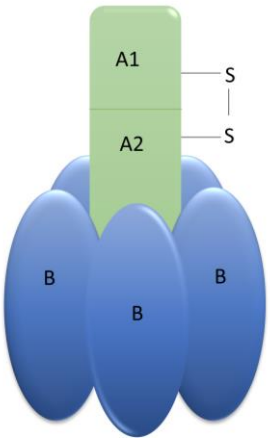


Figure 2. Structure of the Shiga toxin. The toxin is composed of two subunits, an A subunit and a pentameric B subunit [40].

The *stx* genes are located on lambdoid bacteriophages in the bacterial chromosome, introduced through a mechanism called horizontal gene transfer [41]. The Shiga toxins are released in the intestine during a lytic cycle of the phage, and enter and circulate in the blood stream until it reach the target organs, such as kidney [42]. However, the mechanism of Shiga toxin delivery across the intestinal epithelial barrier and to the systemic circulation remains unknown [1, 42]. In the host, Shiga toxins bind to a glycolipid receptor called globotriaosylceramide (Gb3) on the surface of endothelial cells. Following Shiga toxin binding to the Gb3 receptor, the toxin is internalized and trafficked via the retrograde pathway from the Golgi bodies and endoplasmic reticulum (ER). The Shiga toxins inhibit protein synthesis by removing an adenine from the 28S ribosomal RNA which leads to cell death [22].

1.3.1.2 Intimin in colonization of STEC

The two proteins intimin and translocated intimin receptor (Tir), encoded by the *eae* and *tir* genes located on the LEE pathogenicity island, are important for the virulence of STEC. The interaction of intimin with the Tir results in the formation of A/E lesions [27]. The first description of A/E lesions was published by Moon et al. in 1983. They showed that A/E lesions consisted of intimate attachment of the bacterium to the intestinal membrane, effacement of microvilli, and cytoskeleton changes including polymerization of actin that protrude from the membrane and results in pedestal formation (Figure 3) at the attachment site [43].

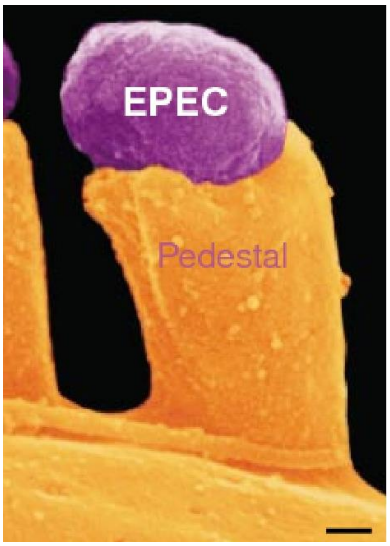


Figure 3. Scanning electron microscopic image of an EPEC bacterium located on a pedestal formed on a cultured epithelial cell. Figure reprinted with permission from publisher [44].

The genes encoding intimin and Tir, as well other genes encoding proteins involved in the formation of A/E lesion, are located on the ~35 Kb LEE pathogenicity island. The LEE consists approximately of 41 genes in total and is organized in five operons named LEE1, LEE2, LEE3, LEE4 and LEE5 [45]. The LEE pathogenicity island harbours genes encoding proteins with a range of functions, including structural components (Esc, Sep and Ces proteins) of a type III secretory system (T3SS), the adhesin intimin and its receptor Tir, LEE encoded regulators (Ler), several proteins (Esp) secreted through the T3SS, and their chaperones (Figure 4). The T3SS is composed of more than 20 proteins which play a major role in delivering the effector proteins into the target host cell [46]. LEE1 to LEE4 harbour

the genes involved in transcriptional regulation (Ler) and structural components of the T3SS, while LEE5 harbours the *eae* and the *tir* genes [45].

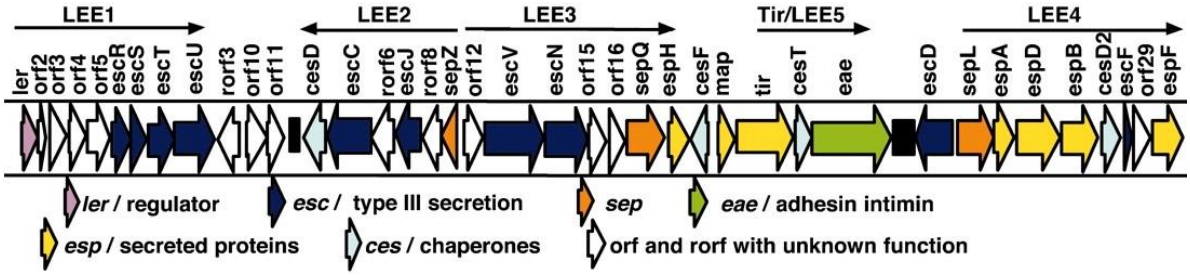


Figure 4. Organization of genes in the LEE pathogenicity island. Figure reprinted with permission from publisher [47].

Intimin is a 94 to 97 kDa outer membrane adhesin that shares homology with invasion proteins of *Yersinia pseudotuberculosis* and *Yersinia enterocolitica* [48]. Intimin consists of two functional regions (Figure 5). The N-terminal region (residue 1-550) contains a signal peptide (SP) region that mediates translocation from the cytoplasm across the inner membrane, and a lysin motif (LysM) region that mediates peptidoglycan binding. From residue 189-550, an internal transmembrane "β-domain" is located that spans the outer membrane. The C-terminal region contains four domains (D0-D3) that interacts with translocated Tir in the host cell membrane (Figure 6) [49]. Among these domains, the β-domain is conserved and predicted to consists of a β-barrel that directs intimin dimerization [50].

All intimin variants currently known are characterized by a conserved N-terminal region and a highly variable C-terminal region. So far, 27 different variants of the intimin protein have been identified [27]. Studies have found that the host and/or tissue tropism of different strains of STEC isolated from human and animal models is influenced by the type of intimin the strains express [48, 51].

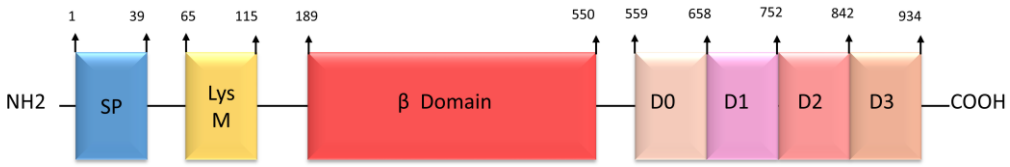


Figure 5. The domains of intimin, withdomain boundaries indicated by amino acid position. Signal peptide (SP) region (1-39), lysin motif (LysM) (65-115), β-domain (189-550), D0-D3 (559-934) [49].

Several studies have observed a strong association between HUS and the presence of the *eae* gene and the *stx2a* gene [52, 53]. Presence of the *eae* gene together with *stx2* gene has been suggested to be an even better marker to identify the STEC strains which have the potential to cause HUS [54]. Likewise, studies conducted using animal models have revealed the importance of intimin in STEC O157:H7 for intimate attachment and disease development *in-vivo* [55, 56]. This result was confirmed by restoring the function of intimin by introducing the wild type *eae* gene into the *eae* mutant strain through a plasmid [56].

Other than intimin, Tir, structural components of T3SS and the effectors secreted through the T3SS are important for the formation of A/E lesions. Tir is first exported by the bacteria through the T3SS into the host cell cytoplasm, where it is integrated into the host cell plasma membrane so that its extracellular binding site can act as a receptor for intimin (Figure 6). Binding of intimin to Tir subsequently leads to clustering of Tir, polymerization of actin and pedestal formation in the host cell [57]. In STEC and EPEC, actin assembly in the host cell cytoplasm is initiated by different signalling pathways. While actin polymerization in EPEC is triggered by phosphorylation of tyrosine 474 (Tir-Nck pathway), Tir will not undergo tyrosine phosphorylation but instead bind to a bacterial effector protein called $EspF_U$ or Tccp (Tir cytoskeleton coupling protein) to trigger actin polymerization in STEC O157:H7 [27]. In addition to Tir, intimin γ has been identified to bind to the host protein nucleolin [58].

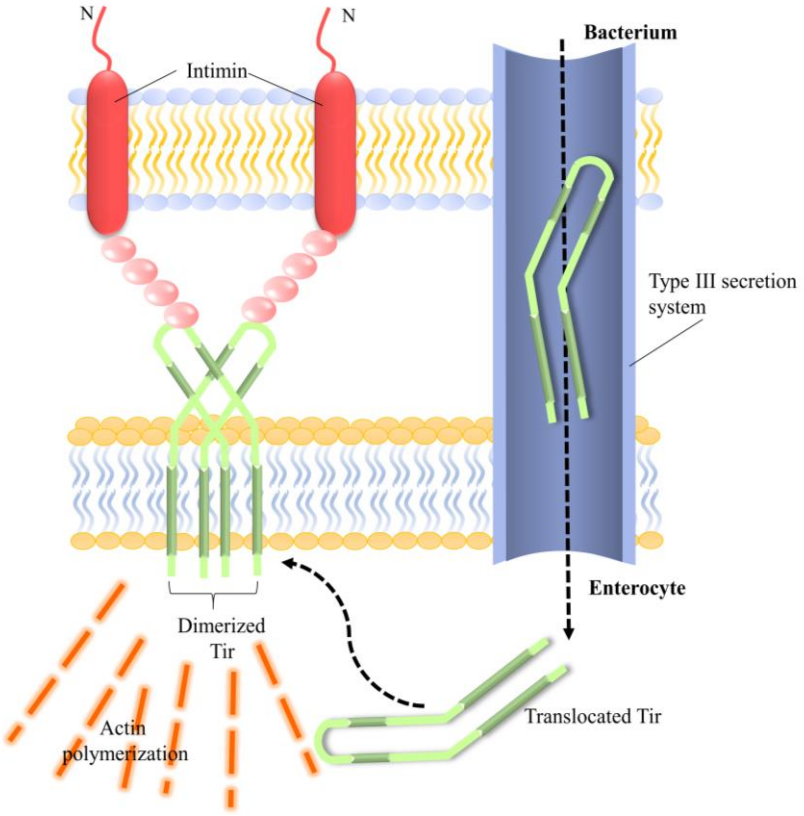


Figure 6. Simplified figure showing Tir translocation and subsequent events in adherence of STEC to intestinal epithelial cells. The bacteria secrete Tir through the T3SS into the host cell. Translocated Tir form a dimer in the host cell membrane where it acts as a receptor for intimin. Binding of intimin to Tir subsequently leads to clustering of Tir, polymerization of actin and pedestal formation in the host cell [25].

Although intimin has been considered for many years to be the primary adhesin in STEC pathogenesis, several lines of evidence indicate that STEC use other mechanisms as well for adhesion to the host cell membrane, which may play a part in induction of severe disease [25]. This became clear when a number of *eae*-negative strains, such as strains of serotypes O113:H21 [59] and O104:H4 [16] were documented as cause of severe disease in sporadic cases and small outbreaks. In addition to this observation, several studies have reported non-LEE encoded putative adhesins which may contribute to STEC colonization [25].

1.3.1.3 Long polar fimbriae (Lpf)

Lpf are involved in colonization, persistence and pathogenesis of STEC [60]. Lpf of STEC O157:H7 are highly similar to the long polar fimbriae of *Salmonella enterica* serovar Typhimurium [61]. Lpf are encoded by two different loci *lpf1* and *lpf2* which are found on O

islands (OI) -141 and OI-154, respectively. To define the same or highly homologous *lpf* loci, several research groups have used different acronyms. For example, *lpf1* and *lpf2* have also been named as $lpfA_{OI\text{-}141}$ and $lpfA_{OI\text{-}154}$, respectively [32]. The *lpf1* operon of STEC consists of 6 genes (*lpfABCC'DE*) [32]. The predicted function of each gene are as follows: *lpfA* encodes the major fimbrial subunit protein, *lpfB* encodes the chaperons needed for proper folding of the protein, *lpfC* encodes the outer membrane usher proteins, and it contains a stop codon that separates the gene into *lpfC* and *lpfC'*, *lpfD* encodes the minor fimbrial subunit, and *lpfE* encodes another fimbrial subunit. The *lpf2* operon consists of 5 genes (*lpfABCDD'*) but lacks the *lpfE* gene, and the *lpfD* gene is duplicated [27]. The expression of the *lpf1* locus is highly controlled by environmental factors and the bacterial transcriptional regulators HNS (Histone like nucleoid structuring protein) and Ler [1].

Torres et al. showed that the *E. coli* O157:H7 *lpf* isogenic mutant adhered slightly less to tissue culture cells compared with the wild type strain [32]. Furthermore, studies done in animal models showed diminished adherence when infected with STEC O157:H7 strains with mutation in one or both *lpf* loci [62, 63]. Another study done on a human small intestine model revealed that mutation in the *lpf* operons led to a shift in the location of colonization from the colon to the small intestine [64].

The *lpfA1* and *lpfA2* genes can be detected among LEE-positive STEC O157:H7 and non-O157:H7 STEC associated with severe disease [65]. The *lpfA* gene has also been detected in LEE-negative STEC strains [25].

1.3.1.4 Other Adhesin proteins

The *E. coli* YcbQ laminin-binding fimbriae (ELF) are encoded by a fimbrial operon termed *ycbQRST*, which is homologous to the F17 fimbrial genes. The name is given due to its ability to bind to the extracellular matrix protein laminin. *In-vitro* studies using an *elfA (ycbQ)* deletion mutant strain showed significant reduction in intestinal adherence comparison to the wild type strain. Along with other known adhesins, ELF has been believed to contribute to the overall adhesive properties of STEC. However, further investigations are needed to reveal their role in vivo [33].

The hemorrhagic coli pilus (HCP) is a type IV pilus, which is found in a wide range of Gram-negative bacteria including EPEC, ETEC and STEC. HCP is encoded by a chromosomal gene

hcpA. Inactivation of the *hcpA* gene in STEC O157:H7 has shown reduction in adherence compared to the wild type strain [66]. Besides adherence, HCP has several other functions, such as formation of biofilm, bacterial aggregation, phage receptors, uptake of DNA, and signal transduction [25].

Other adhesins found in LEE-positive STEC thought to be involved in intestinal adherence are the IrgA homologue adhesin (Iha), Efa-1/LifA and its homolog ToxB [1].

The molecular mechanisms behind cell adhesion of LEE-negative STEC strains have not been as extensively studied as for LEE-positive STEC [10]. However, recently several putative adhesins have been reported. In LEE-negative STEC O113:H21, a novel fimbrial gene cluster (lpf_{O113}) related to Lpf has been identified and shown to be important for adherence in epithelial cells. In contrast to the *lpf2* operon, lpf_{O113} comprises only four open reading frames, *lpfA* to *lpfD*. [67]. The Autoagglutinating adhesin is encoded by the *saa* gene of LEE-negative STEC strains, originally identified in STEC O113:H21 [35]. Furthermore, the *E. coli* immunoglobulin-binding protein G (EibG) binds to human immunoglobulin G and A, and is purported to play a role in bacterial chain-like adherence pattern (CLAP) formation [68].

1.3.2 Classification of STEC

So far, more than 400 STEC serotypes have been identified, but among them only a few are associated with disease in humans [69]. The first *E. coli* serotype recognised as a human pathogen was O157:H7, when a strain of this serotype caused of an outbreak of bloody diarrhoea in the United States in 1982 [70]. In most parts of the world, O157:H7 is the most common STEC serotype isolated in outbreaks and from sporadic cases. However, non-O157:H7 STEC serotypes have also been associated with severe disease in humans, and the most commonly found non-O157 STEC associated with severe disease are STEC of the serogroups O26, O45, O103, O111, O121, and O145 (known as the “Big 6”) [71].

In another classification method, strains of the O157:H7 serotype can be further divided into two categories based on its ability to ferment sorbitol on Sorbitol-MacConkey (SMAC) or CHROM agar. Most STEC O157:H7 strains are sorbitol negative and grow as clear colonies on SMAC plates. However, sorbitol fermenting (SF) strains have been reported to cause severe disease in several countries in Europe. Such strains are commonly non-motile (SF O157: NM) [10, 72].

Karmali et al. categorized STEC strains into five seropathotypes A to E based on the association of the different serotypes with severity of disease and their potential to cause outbreaks [73]. The European food safety agency (EFSA) in 2013 decided on a modified classification in two seropathotypes: HUS-associated and non-HUS associated. The former category comprises all STEC serotypes associated with severe disease (HUS), while the latter comprises all other serotypes [74].

There is enormous genetic diversity within STEC strains [75]. In addition to serotyping, a multitude of methods have been used for typing of STEC strains for epidemiological studies and outbreak investigations, due to the need of discrimination of strains at sub-species and sub-serotype level. Pulsed-field electrophoresis (PFGE), Multiple-locus variable number tandem repeats analysis (MLVA), and whole genome sequencing are examples of such methods. The principle of PFGE is visualization of variations in the size of DNA chromosome fragments between bacterial strains, using rare-cutting restriction enzymes [10]. MLVA is based on polymorphic mini-satellites, known as variable numbers of tandem repeats (VNTRs). PCR is used to amplify the repeat regions, with subsequent (capillary) gel electrophoresis. MLVA is fast and it has a good discriminatory power. Due to these advantages MLVA has replaced PFGE [76].

1.3.3 Epidemiology of STEC

Despite being relatively uncommon in comparison with other zoonotic diseases, STEC infection has become a major focus of health authorities due to the occurrence of food borne outbreaks and the potential for severe complications (HUS) associated with this pathogen, especially in young children. A study conducted by Majowicz et al. estimated that STEC annually causes 2.8 million cases of acute infection, 3,890 cases of HUS, 270 cases of end stage kidney disease, and 230 deaths worldwide [77]. In the literature, the highest worldwide incidence of HUS has been reported in Argentina, particularly in children under the age of 5 [78].

It is believed that STEC is transmitted most commonly through direct or indirect contact with animals colonized with STEC or via consumption of food and water contaminated with faeces from such animals. Generally, outbreaks have been associated with consumption of contaminated food [79, 80]. However, STEC infection has also been documented from person-to-person spread, particularly in sporadic cases [10].

As mentioned previously, it has become apparent that in most parts of the world, non-O157:H7 STEC (Big 6) strains are also associated with severe disease (HC and HUS) in humans [71]. Since the STEC O104:H4 outbreak in Germany in 2011, surveillance of non-O157:H7 STEC has been improved in many countries [16, 81]. In 2011, 463 O157:H7 and 521 non-O157:H7 STEC cases were reported in the United States [82]. Even though STEC infection probably is prevalent in many low-income countries, there are scarcity of reliable data due to lack of surveillance systems and as well as laboratory diagnostics facilities [1].

1.4 Methods for studying gene function

The function of a gene can be studied by altering the DNA sequences using targeted genome editing tools [83]. A wide variety of tools are available for targeted gene editing in eukaryotes and prokaryotes, such as homologous recombination allelic replacement, integrases and directed nucleases [84].

Homologous recombination is the most widely used genome manipulation tool in *E. coli* by which homology arms flanking the repair template specify the site of recombination. The limitation of homologous recombination is that it requires selection markers such as an antibiotic resistance gene inserted between double stranded DNA (dsDNA) obtained from the upstream and downstream region of the target locus leaving scars in the genome [85].

Targeted nucleases such as Zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and Cas9 create double strand breaks (DSBs) in a bacterial genome and induce the cells to undergo DNA repair. These methods make use of DNA repair pathways in cells to make genetic alterations [86].

Several studies have constructed *eae* mutant strains using allelic exchange mutagenesis for eliminating the *eae* gene to study their role in adhesion [87-89]. However, the applications of these techniques are laborious, time consuming and leave scars in the genome [83]. Use of an antibody targeted against the C-terminal domain of intimin is another way that can be used to deactivate the function of intimin [90].

1.4.1 CRISPR/Cas9 system

The Clustered regularly interspaced short palindromic repeats (CRISPR) is an adaptive immune system found in many bacteria and most archaea, which eliminates invading genetic material and protects against infection by bacteriophage and plasmid transfer. The RNA guided CRISPR associated protein (Cas9) has emerged as a powerful tool in genome editing in eukaryotes and prokaryotes. Gene editing tools using ZFNs and TALENs require use of expensive sequence specific protein synthesis while Cas9 requires less-expensive RNA sequences with homology to the given locus [86].

The editing capability of the CRISPR/Cas9 system has also been used for therapeutic purpose e.g. Cas9 programmed to specifically target the antibiotic-resistance genes in virulent strains [91]. Because of its high efficiency and accuracy, this genome editing technology has emerged as a tool in gene therapy applications [92]. Researchers have also used the targeting capability of the CRISPR/Cas9 system to study the genes without altering the DNA sequence. Catalytically dead mutant Cas9 (dCas9) are used for this purpose [93].

1.4.1.1 CRISPR/Cas9 biology

The bacterial CRISPR system consists of short repeated sequences flanking spacer sequences derived from nucleic acids from viruses and plasmids and a set of CRISPR associated (*cas*) genes, that encode proteins which play an essential role in the bacterial immune response [94].

The CRISPR system is classified into five main types (Type I, II, III, V and VI) depending on the type of Cas protein, gene composition, and organization of gene loci. In Type II CRISPR system, the RNA guided Cas9 protein generates a single double strand break at a specific site of interest [95].

The activity of the CRISPR/Cas9 system can be divided into three stages (Figure 7). During the first stage, invading new viral or plasmid derived DNA is incorporated as spacers into the CRISPR locus. In the second stage, the CRISPR array is transcribed into a long precursor crRNA (CRISPR RNA) that is further processed into a mature crRNA. The mature crRNA contains a short sequence that complements a sequence from a foreign genetic element and a repeat sequence. The repeat sequences in the crRNA base pairs with small non-coding RNA, called the trans-activating RNA (tracrRNA) and form a unique dual RNA hybrid structure. In

the final stage, the dual RNA hybrid assembles with Cas9 and searches for the target sequence (protospacer) in the re-exposed invader sequences which is complementary to the short sequence in the crRNA [96].

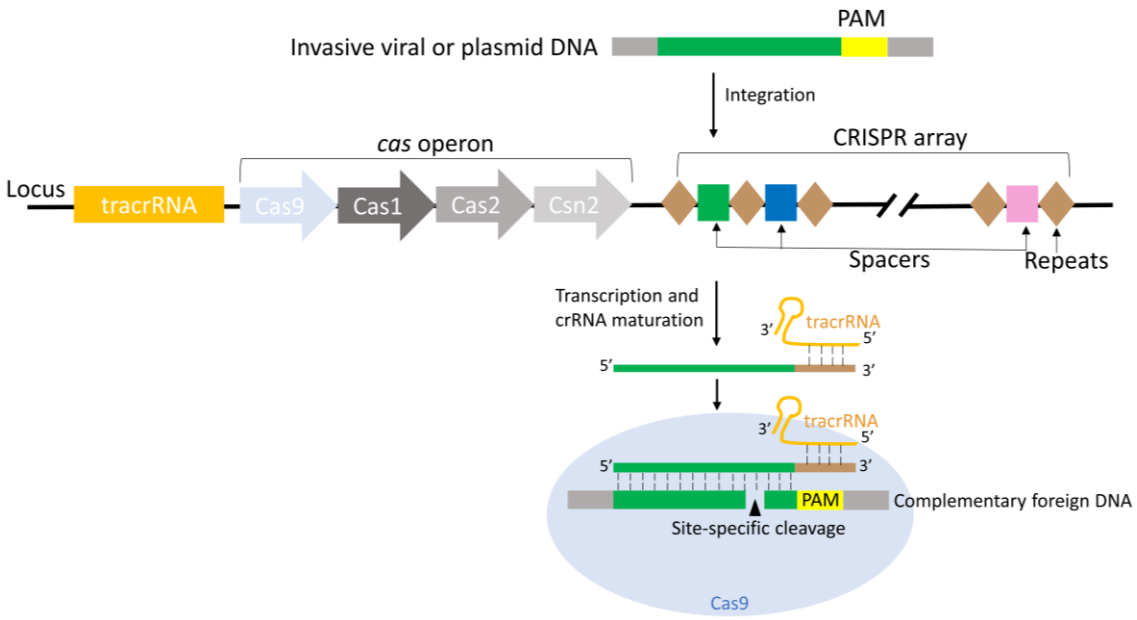


Figure 7. An overview of CRISPR/Cas9 as a bacterial adaptive immune system. During the first exposure to viral or plasmid DNA, new spacers derived from the invaders are added to the end of the CRISPR array. The CRISPR array consists of spacers obtained from the past invaders but does not contains a PAM sequence and is therefore recognized as "self". The CRISPR array is transcribed into long precursor crRNA and subsequently matures into mature crRNA. tracrRNA base pairs with the repeat sequence in the crRNA and the resulting dual-RNA assembles with the Cas9 enzyme to target the foreign DNA based on sequence specificity. The resulting complex recognizes and destroys the invading viral or plasmid DNA upon re-exposure [96].

The recognition of the target protospacer by Cas9 requires a short conserved region known as protospacer adjacent motif (PAM). The PAM sequence varies depending on the bacterial species Cas9 is derived from. The Cas9 nuclease, derived from *Streptococcus pyogenes*, recognizes the PAM sequence 5'-NGG-3' (where N strands for any of the four nucleotides) near the 5' end of the target gene. After the Cas9 has recognized the protospacer with the appropriate PAM sequence, the target DNA will unwind to make complementary base pairing with the crRNA. Upon base pairing, the Cas9 undergoes conformational changes and cleaves both DNA strands approximately 3 bases from the start of the PAM sequence [97].

1.4.1.2 Genome editing with CRISPR/Cas9 system

The CRISPR/Cas9 system used in genome engineering differs from that found in the natural bacterial immune system. In most CRISPR/Cas9 mediated genome editing systems, a 20 bp long targeting oligo, called guide DNA (gDNA), is synthesised, annealed and cloned into a plasmid that contains the sequence encoding the tracrRNA. Function-wise, the gDNA resembles the spacer present in the bacterial CRISPR array. The target of Cas9 can be programmed by changing only the 20 bp long gDNA sequence. Therefore, it is simple and flexible to use as a genome editing tool [86].

The DSB created by Cas9 is repaired either by error prone non-homologous end joining (NHEJ) or by a high-fidelity homology directed repair (HDR) mechanism (Figure 8). In eukaryotes, most DSBs are repaired by the indel-forming NHEJ repair mechanism, which results in small random deletions and/or insertion at the site of DSB. Eukaryotic cells can survive a Cas9 mediated DSB in the absence of a repair template, since the NHEJ repair mechanism does not require a homologous repair template [86]. In eukaryotes, the error prone nature of the NHEJ pathway is used to introduce small indels at the target sites and knockout genes [98]. In contrast, in many bacteria including *E. coli*, DSB introduced by the Cas9 enzyme normally will kill the organism since most bacteria lack the NHEJ pathway. HDR is the predominant pathway used by *E. coli* to repair damaged DNA using a homologous repair DNA template [98].

Genome editing using CRISPR/Cas9 system has been less studied in bacterial genomes compare to eukaryotic genomes [99]. More recently, for the first time Jiang et al. proposed the application of the CRISPR/Cas9 system to introduce a chromosomal mutation in *Streptococcus pneumoniae* and *Escherichia coli* [83]. As mentioned before, in prokaryotes introduction of a DSB into the chromosome can induce recombination on damaged DNA [85, 86]. If a repair DNA template with homology arms was provided, the break could be repaired according to the exogenous template, allowing for precise gene editing. Thus, manipulations in the chromosomal DNA of *E. coli* via DSB-mediated HDR may be achieved by introducing the CRISPR/Cas9 system together with a repair DNA template carrying the desired mutation [83].

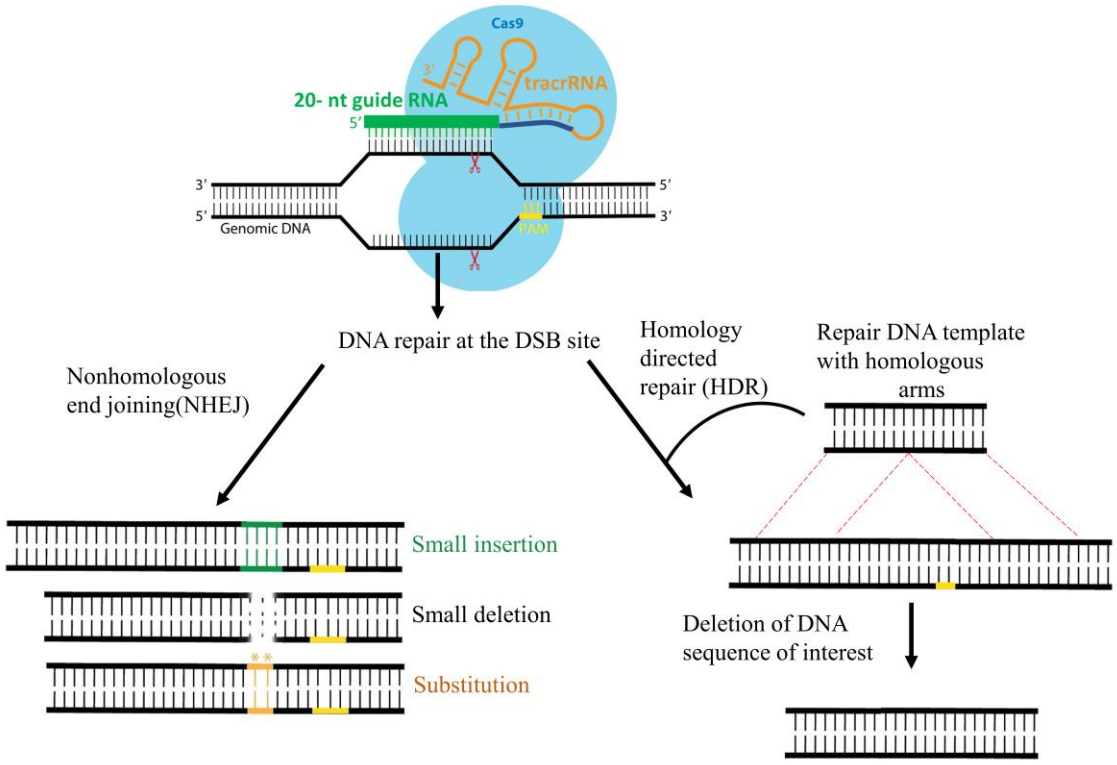


Figure 8. An overview of CRISPR/Cas9 mediated DSB and the DNA repair mechanisms. The RNA guided Cas9 protein generates a single double strand break at a specific site of interest and repair will proceed through NHEJ or HDR depending on the experimental condition [96].

Since a DSB normally will be lethal to the bacterial cell, introducing a marker-less mutation that prevents target recognition can be used to recover the edited clones. The mutagenic repair DNA template harbouring the mutation will recombine in the bacterial genome and eliminate the recognition and cleavage of the protospacer by the Cas9. The cells that escaped the recognition by the CRISPR/Cas9 system by harbouring the desired mutation will be selected. This selection method reduces the number of colonies that need to be screened to identify the edited clones [83].

1.5 Aim of the study

An accurate diagnosis of STEC infection is still hampered by our lack of ability to differentiate between strains that cause HUS from strains that cause no or only mild disease. Many studies have been focused on defining virulence factors and mechanism of pathogenesis which are essential for development of severe disease in humans [9, 100-102].

In a previous study, it was found that an operon with putative fimbrial adherence genes termed *lpf*$_{B1}$ was more highly expressed in non-O157 STEC isolated from children with HUS, compared with strains from children who did not develop HUS. The proteins encoded by this putative operon showed 42-63% of identity to the Lpf2 of *E. coli* O157:H7 Sakai and were found to be located in the same genetic island, the OI-154. These results indicate a role of the *lpf*$_{B1}$ operon in adherence and development of HUS [101].

Experimental verification of the role of the fimbrial operon *lpf*$_{B1}$ in adherence of STEC to epithelial cells is difficult. The strong intimate attachment of the bacterium to intestinal epithelial cells, encoded by the LEE gene *eae* tends to outnumber any additional adherence caused by other adherence factors. To circumvent this problem, we wanted to knockout the *eae* gene in a STEC strain, and the aim of the study was to knockout this gene using the CRISPR/Cas9 system in a HUS-associated STEC strain which harboured the *lpf*$_{B1}$ operon.

If the knockout was successful, the mutant strain could be used to study the impact of the *eae* gene in expression of the *lpf*$_{B1}$. The protocol developed could then be modified to knockout also the *lpf*$_{B1}$ operon. Thereafter the adherence properties of the mutant STEC strains could be compared, with respect to both *eae* and *lpf*$_{B1}$.

2 Materials and methods

This chapter gives a detailed description of the materials and methods that I have used in this project to generate a deletion in the *eae* gene. A two-plasmid system was used to achieve a specific deletion in the target *E. coli* strain termed FHI11, as described previously for *E. coli* [83, 85, 103, 104]. The CRISPR/Cas9 system used in this study consisted of three main components: (1) pCas9 plasmid which encodes Cas9 endonuclease to make DSB, (2) pCRISPR plasmid which encodes the gRNA for guiding the Cas9 to the complementary target region in the *eae* gene, and (3) a homologous DNA repair template, containing a desired mutation, for HDR. As a first step, oligonucleotides (gDNA, repair DNA templates, and primers) were designed and ordered before moving to the wet laboratory procedures.

Experiments were conducted according to a protocol "Multilple stepwise gene knockout using CRISPR/Cas9 in *E. coli*" established by König et al., with some modifications. The major modifications were to use (1) a pCas9 plasmid instead of pCasRed, (2) a pCRISPR plasmid instead of pCRISPR_SacB, and to do (3) cloning by Golden Gate Assembly pCasRed is a modified version of the pCas9 plasmid (Addgene #42876). The pCasRed contains λ-Red genes to increase the editing efficiency. Since we were not able to obtain the plasmids used in the protocol from the relevant laboratory, we decided to use the pCas9 plasmid (without λ-Red genes) available from Addgene. Even though there were also other plasmids available from Addgene which included λ-Red genes, the protocols were not considered detailed enough for us to design our experiments.

2.1 Bacterial strains and plasmids used in this study

Bacterial strains and plasmids used in this study are listed in table 2. The strain FHI11 was received from the Norwegian Institute of Public Health for use in a previous project, was originally isolated from a faecal sample of a patient suffering from HUS.

The strain, which was classified as STEC, had lost its *stx* genes and the ability to express Shiga toxins, either during the infection process or during handling in the laboratory. The strain was available for this project from the strain collection at the Department of Medical Microbiology, St. Olavs hospital. The strain was chosen for this study based on the following criteria: (1) It did not contain the *stx1/2* genes, (2) it contained both the *eae* gene and the lpf_{B1} operon, and (3) it was susceptible to kanamycin and chloramphenicol.

The pCas9 plasmid (Addgene #42876) and pCRISPR plasmid (Addgene #42875) were a gift from Luciano Marraffini [83]. MAX Efficiency® DH5α™ *E. coli* competent cells were used for vector construction [105].

Table 2. List of bacterial strains and plasmids used in this study.

Strains and plasmids		**Characteristics**	**Source**
Strains	STEC FHI11	Clinical isolate from patient with HUS	[102]
	DH5α™ compètent *E. coli* cells	F^- Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rk^-, mk^+) *pho*A *sup*E44 λ^- *thi*$^-$1 *gyr*A96 *rel*A1	Invitrogen
Plasmids	pCas9 (Addgene #42876)	*E. coli* vector containing the *S. pyogenes cas9* gene and a sequence encoding tracrRNA	Addgene and [83]
	pCRISPR (Addgene #42875)	*E. coli* vector containing a crRNA expression cassette for Cas9 targeting	Addgene and [83]

The pCas9 plasmid is derived from a low-copy number plasmid backbone (pACYC184) and the pCRISPR plasmid is derived from a high copy number plasmid (pZE21). The pCas9 plasmid contains sequences for *S. pyogenes* derived Cas9 and tracrRNA driven by respective endogenous promoters, and the chloramphenicol resistance gene CmR (Figure 9).

The pCRISPR plasmid contains the inserting site for gDNA with BsaI restriction sites which can be used for one-step assembly of the gDNA by Golden Gate Assembly, and the kanamycin resistance gene KanR (Figure 10).

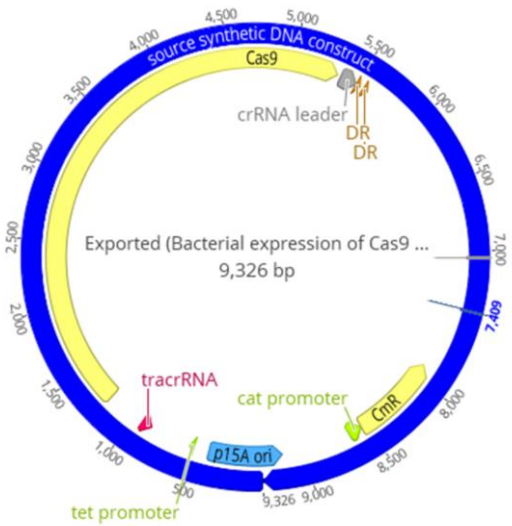


Figure 9. Graphical view of the pCas9 plasmid (Addgene #42876). The plasmid contains two promoters: tet and cat, a chloramphenicol resistance gene (CmR), gene encoding tracrRNA, a Cas9 endonuclease encoding sequence, two direct repeats (DR) with BsaI type IIS restriction sites (not used in this study) as well as a leader sequence.

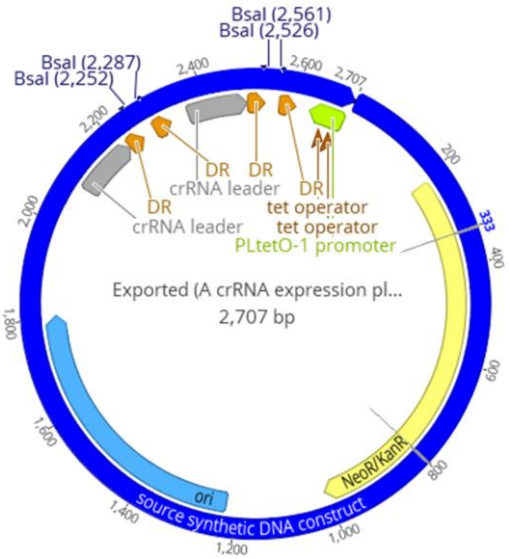


Figure 10. Graphical view of the pCRISPR plasmid (Addgene #42875). The plasmid contains a PLtetO-1 promoter, a kanamycin resistance gene (KanR), two direct repeats with BsaI type IIS sites to facilitate easy insertion of gDNA, and a leader sequence driving the crRNA array. The DR-BsaISpc-DR region was accidently duplicated during construction, but they have reported that this will not interfere with the insertion of gDNA through BsaI digest and oligo cloning [83].

2.2 Methods for cultivation and storage of bacterial strains

The growth conditions maintained for all *E. coli* strains used in this study are given in table 3. All the LB plates were kept upside down to prevent condensed liquid dripping on the surface of the agar. The STEC strain FHI11 was recovered from frozen state in Greaves solution at -80 ˚C from the strain collection at the Department of Medical Microbiology, St. Olavs Hospital.

Table 3. Incubation conditions for the bacterial strains used in this study.

Bacterial strain	**Solid culture**			**Liquid culture**			
	Medium	**Tem**	**Duration**	**Medium**	**Tem**	**Duration**	**Shaking at**
FHI11	LB	37 ˚C	12-18 hrs	LB	37 ˚C	12-18 hrs	200 rpm
E. coli DH5α (pCas9)	LB with Chl	37 ˚C	18-30 hrs	LB with Chl	37 ˚C	18-30 hrs	200 rpm
E. coli DH5α (pCRISPR)	LB with Kan	37 ˚C	12-18 hrs	LB with Kan	37 ˚C	12-18 hrs	200 rpm
E. coli DH5α (pCRISPR_T30RNA)	LB with Kan	37 ˚C	12-18 hrs	LB with Kan	37 ˚C	12-18 hrs	200 rpm
FHI11 (pCas9)	LB with Chl	37 ˚C	18-30 hrs	LB with Chl	37 ˚C	18-30 hrs	200 rpm
FHI11 (pCas9, pCRISPR_T30RNA and repair DNA)	LB with Chl and Kan	37 ˚C	18-30 hrs	LB with Chl and Kan	37 ˚C	18-30 hrs	200 rpm

Tem-Temperature, rpm- Revolution per minute, Chl- Chloramphenicol, Kan-Kanamycin, LB- Luria Bertani. Plasmids harboured by the bacterial stains are shown in brackets.

E. coli DH5α strains harbouring the pCas9 and pCRISPR plasmids were received from Addgene (Teddington, UK, accessed from https://www.addgene.org) as agar stabs. LB agar plates containing appropriate antibiotics were used to get isolated colonies from the agar stab. After incubation on LB agar, a single colony from each plate was inoculated into LB broth

containing appropriate antibiotics and incubated overnight at 37 ˚C with agitation at 200 rpm. Since pCas9 is a low copy number plasmid, it was incubated approximately for 18-30 hrs. In contrast, pCRISPR, which is a high copy number plasmid, was incubated for 12-18 hrs. A stock solution consisting of 500 µl of overnight culture mixed with 500 µl of 50% sterile glycerol solution (Appendix A) was stored at -80 ˚C for use in the study.

MAX Efficiency® DH5α™ *E. coli* competent cells were received from Invitrogen in frozen state in five vials each containing 200 µl of competent cells. These cells were stored at -80 ˚C until use.

2.3 Bioinformatic tools

This section covers the different tools/softwares used to analyse the bacterial genome or genes, and to design oligonucleotides. A brief introduction is given to all the tools/softwares, including NCBI BLAST [106], NCBI Primer-BLAST [107], Geneious R11 [108], and Oligo Primer Analysis Software 7.60 [109]. How these tools were used, and the settings are explained under the presentation of each oligonucleotide.

2.3.1 Geneious Software R11

Geneious R11 is a commercial software used for organization and analysis of biological data, with the focus on molecular sequences [108]. The software enables the visualization and manipulation of DNA and protein sequences. For a given DNA sequence the software contains a variety of tools which can be used to analyse the data, e.g. “Find CRISPR sites” and “design primers”. In this study, this software was used to design the 35 bp long forward (T30RNA_F) and reverse (T30RNA_R) strands of gDNA, the forward primer (InsertID_F) used to verify the pCRISPR_T30RNA construct, as well as the 120 bp long and 70 bp long repair DNA templates.

2.3.2 NCBI Basic Local Alignment Search Tool (BLAST)

The Basic Local Alignment Search Tool (BLAST) website (https://blast.ncbi.nlm.nih.gov/Blast.cgi) at the National Centre for Biotechnology Information (NCBI) was used to search and align nucleotide and protein sequences. BLAST takes a sequence of a nucleotide or a protein (Query sequence) as input and searches it against

the database (Subject sequences) and provides statistics that estimates the likelihood of a match occurring by chance. A similarity score is presented as query coverage, E-value and percent identity. The query coverage measures the percentage of the query sequence that matches the subject sequence. The E-value represents the probability of an alignment occurring by chance when searching a database of a particular size and the percent identity indicates the percent similarity between the query and the subject sequence over the length of the covered area [106].

2.3.3 Tools used in primer designing

The NCBI Primer-BLAST [107] and Oligo Primer Analysis Software 7.60 (Molecular Biology Insights) [109] were used to design all the primers except the primer InsertID_F. Primer-BLAST is a general purpose publicly available tool which can be used to design specific primers for a selected target in one step (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Based on the entry of a specific sequence or a GenBank accession number, Primer-BLAST scans for primers matching the selected parameters.

The Oligo Primer Analysis Software 7.60 is a commercial software used to design PCR primers. The graphical interface of this software makes it easy to design primers based on visual inspection of nucleotide sequences including the melting temperature, internal stability, hairpin and duplex formation [109].

2.4 Oligonucleotides

All the oligonucleotides designed and used in this study are presented in table 4. The oligonucleotides were manufactured by SIGMA Life Science and received in a lyophilized state, except InsertID_F, KOveri_F and KOveri_R, which were manufactured by Invitrogen (Thermo Fisher Scientific). A 100 μM stock solution was made for each oligonucleotide by dissolving the lyophilized powder in molecular grade water (MGW) which was stored at -20 °C until further use. From the stock solution a 10 μM working solution was made for use in PCR, and a 1.6 μM solution for Sanger sequencing, by dilution with MGW before it was stored at -20 °C. For the repair DNA templates, MWG was added to obtain a concentration of 10 μg/μl.

Table 4. List of synthetic oligonucleotide sequences used in this study.

Oligonucleotide Name	**Direction**	**Sequence (5'-3')**	**Size (nt)**	**Purpose**
T30RNA_F	Fw	AAACATGCAGCTTCGTTATCAGTTTGATAAACCGG	35	To construct gDNA
T30RNA_R	Rv	AAAACCGGTTTATCAAACTGATAACGAAGCTGCAT	35	
De201RT_F	Fw	GTTGCTTTGTTTAATTCCGATAAGCTGCAGTCGAATCCTGGTGCGGCGACCGTTGGTTAAATTATTCTGGAGTACAAGAAGCAGGATATTCTTTCTCTGAATATTCCGCATGATATTAAT	120	To construct a dsDNA repair template to introduce a 201 bp deletion and a premature stop codon
De201RT_R	Rv	ATTAATATCATGCGGAATATTCAGAGAAAGAATATCCTGCTTCTTGTACTCCAGAATAATTTAACCAACGGTCGCCGCACCAGGATTCGACTGCAGCTTATCGGAATTAAACAAAGCAAC	120	
Del33RT_F	Fw	TGGTACGGGTAATGAAAATGATCTTCTTTACTCATAATCTCAGCAAATTGAGCCACAGTATGTTAACGAG	70	A ssDNA template to introduce a 33 bp deletion and a premature stop codon
InsertID_R (pBRforEco)	Rv	AATAGGCGTATCACGAGGC	19	Verification of the pCRISPR_T30RNA construct by PCR and Sanger sequencing
InsertID_F	Fw	TAGCTGAGACAAATAGTGCGA	21	
KOveri_F	Fw	CATCATACCCGGCATTAGGC	20	Validate the deletion in the *eae* gene by PCR and Sanger sequencing
KOveri_R	Rv	CGATCCAGACCGTATTTGCT	20	

Fw: Forward primer, Rv: Reverse primer. nt: nucleotides. Bases marked in red show the overhangs.

2.4.1 Design of guide DNA (gDNA)

The gDNA is complementary to a 30 bp protospacer sequence in the *eae* gene and will be transcribed into gRNA that guide the Cas9 to the desired location in the *eae* gene.

Procedure

The oligonucleotides used to construct the gDNA (T30RNA_F and T30RNA_R) were designed according to the guideline established by König et al.[110]. However, the Geneious R11 software was used to find a suitable CRISPR target site in the *eae* gene instead of manually selecting the 30 bases upstream of the PAM sequence and analyse for off-target binding, as recommended by König et al.

The whole genome sequence of strain FHI11 (GCA_000936635.1) was downloaded in GenBank format (.gbk) from the NCBI database and imported into Geneious R11. First, the contig containing the *eae* gene was selected and extracted (Figure 11). Then the "Find CRISPR sites" tool was used to find the most suitable CRISPR/Cas9 target site within the *eae* gene, based on calculated on-target and off-target activity. The gDNA sequence should be 30 bp long, so "N (30)" was entered in the "Target" field, and "NGG" was entered in the "PAM site" field. The on-target activity score is the measure of binding efficiency of the gRNA. The value is given between 0 and 1 based on the position of nucleotides in the sequence and the GC content, where the higher value denotes a higher expected activity [111]. The off-target value is a measure of how many off-target sites the gRNA potentially will bind in a genome of interest. It may be calculated as specificity with value between 0 and 100%, where the higher value denotes a better specificity and less off-target effect for the selected gRNA sequence [112]. The off-target database consisted of the whole genome sequence of strain FHI11, the pCas9 and pCRISPR plasmids, imported in GenBank format (.gbk) from NCBI and the Addgene websites, respectively.

After choosing the most suitable CRISPR target site in the *eae* gene, a 30 nucleotides long DNA sequence without the PAM sequence was extracted using Geneious R11. The extracted sequence, termed T30RNA_F and T30RNA_R (reverse complement), respectively, was chosen so that there should be no BsaI restriction site within the sequence. To facilitate ligation during the cloning step, BsaI compatible cohesive DNA ends were created by adding AAAC to the 5' end and G to the 3' end of the T30RNA_F sequence, and AAAAC to the 5' end of the T30RNA_R sequence (Table 4).

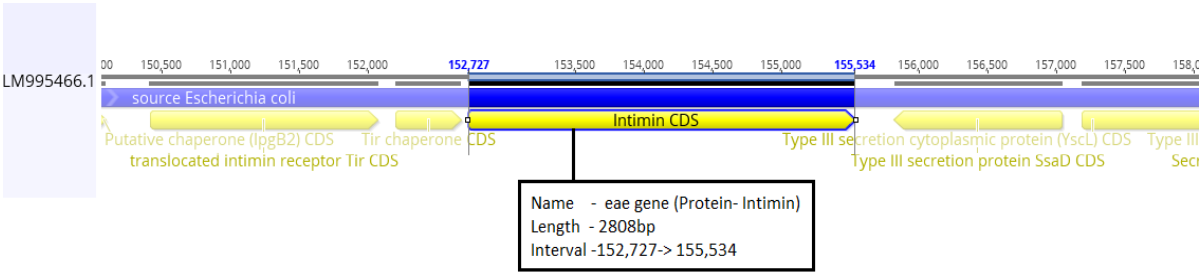


Figure 11. Location of the *eae* gene shown in part of the genome of the strain FHI11 (contig LM995466.1) as visualized by the Geneious R11 software. Gene names in the figure is based on automatic annotation.

2.4.2 Design of a homologous mutagenic repair DNA template

To introduce a specific modification in the bacterial genome via HDR, one must introduce a DNA template carrying a desired mutation [83]. Repair DNA templates consists of homology arms complementary to the sequence upstream and downstream from the deletion. One problem that need to be addressed is that once the bacterial cells have received the CRISPR/Cas9 system, the Cas9 enzyme will keep on making DSBs as long as it recognizes the PAM and protospacer sequence in the target bacterial genome. In this study, this was theoretically solved by designing a repair DNA template so that it would introduce an in-frame stop codon (TAA) and a deletion covering also the protospacer and PAM sequence. This would prevent re-cutting of DNA in cells in which the cut had been repaired through the HDR mechanism. Transformants that survive CRISPR/Cas9 mediated DSB will contain the modification introduced by the repair template.

Procedure

The repair DNA template design was performed according to the guideline established by König et al. [110]. The Geneious R11 software was used to design two different repair DNA templates to generate either a 201 bp or a 33 bp deletion within the *eae* gene. To generate the 201 bp deletion, a 120 bp long dsDNA was designed, while to generate the 33 bp deletion, a 70 bp long single stranded DNA (ssDNA) was designed. Figure 12 provides an overview of how the repair DNA templates were constructed.

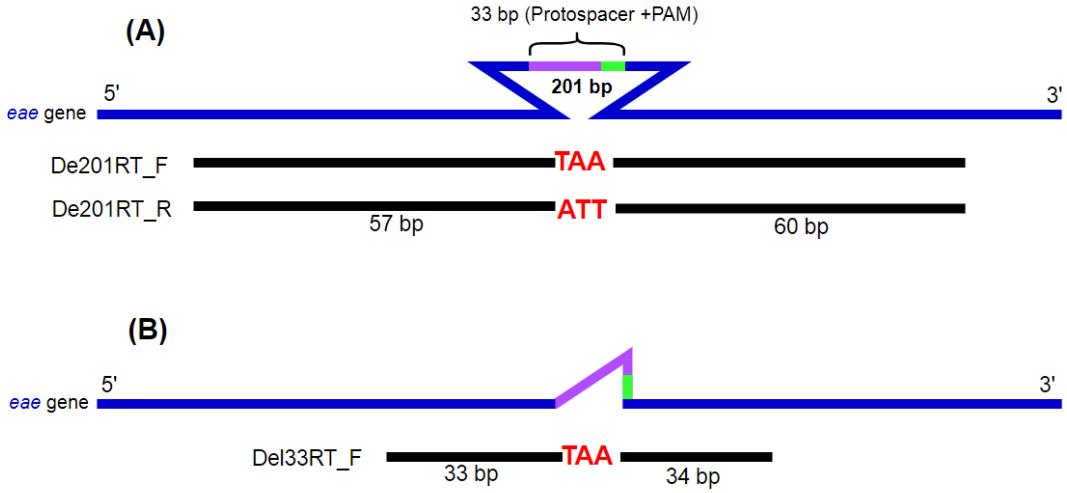


Figure 12. Design of the repair DNA templates. The 5' moiety of the repair template (57 or 33 nucleotides) is complementary to the sequence upstream of the deletion while the 3' moiety of the repair template (60 or 34 nucleotides) is complementary to the sequence downstream the deletion. Stop codon TAA is highlighted in red, the protospacer is shown in purple and the respective PAM sequence is shown in green. (A) 120 bp long leading (De201RT_F) and lagging (De201RT_R) strands were used to generate a 201 bp deletion. (B) A 70 bp long leading strand (Del33RT_F) was used to generate a 33 bp deletion.

The leading strand (De201RT_F) for the dsDNA was constructed by including 57 bases upstream and 60 bases downstream of the 201 bp deletion and adding a stop codon (TAA) in the middle of the repair DNA template. The lagging strand (De201RT_R) was constructed as the reverse complement of De201RT_F. The 70 bp long repair DNA template (Del33RT_F) was obtained by including the 33 bases upstream and 34 bases downstream of the 33 bp deletion, and then adding a stop codon in the middle of repair DNA template.

2.4.3 Design of primers to verify the targeting pCRISPR plasmid construct

After cloning, PCR was used to identify the bacterial colony (hereafter termed colony PCR) that contained the plasmid with the gDNA (pCRISPR_T30RNA) and differentiate these from the bacterial colonies with an intact pCRISPR plasmid or self-ligated plasmids. Primers that were specific to the backbone of the plasmid were designed to anneal to the sites flanking the gDNA insertion site. Gel electrophoresis was used to distinguish the undigested pCRISPR plasmid from the pCRISPR plasmid containing the gDNA.

To achieve a successful PCR reaction with a high efficiency, there are few properties that need to be considered when designing the primers, e.g. as suggested in the Oligo 7 primer analysis software [109].

- The optimal primer length should be between18 bp to 30 bp.
- The difference in melting temperatures between forward and reverse primers should not exceed more than 5 ˚C.
- The primers should have G or C bases within the last 5 bases from the 3' end. This promotes more specific binding at the 3' end of the target sequence.
- Within the primer or primer pair there should not be any complementary regions that promote hairpin or primer-dimer formation and reduce the product yield. PCR primers should be free of significant complementary at their 3' termini.
- The primers should not have false priming site on the target DNA (off-target binding) and it should be very specific to the target site.

Procedure

The forward primer InsertID_F was designed manually using the Geneious R11 software. The reverse primer InsertID_R (pBRforEco) sequence was taken from the Addgene website. Primers available for the specific plasmid was obtained from the "Analyse sequences" link on the plasmid information page. The primers were tested for off-target binding against the host genome and the pCRISPR plasmid by analysing sequence similarity in NCBI BLAST. The same primer pair was used to confirm the PCR results by Sanger sequencing.

2.4.4 Design of primers to validate the genome edit using colony PCR

Once the CRISPR/Cas9 system is introduced into the FHI11 strain, it is important to confirm that the desired deletion is incorporated into FHI11 genome. Gel electrophoresis following colony PCR was used to identify the clone that had undergone deletion from the wild type strain. The same guidelines as mentioned in section 3.4.3 were used to design primers.

Procedure

The primers were generated in NCBI Primer-BLAST and their properties were analysed also using the Oligo Primer Analysis Software. The most suitable primers were tested for off-target binding in the FHI11 genome, and the pCRISPR and pCas9 plasmids by analysing sequence similarity in NCBI BLAST. The primers were designed to flank the 201 bp deletion

in the *eae* gene. Although generating a deletion of 33 bp in the *eae* gene was decided later, the same primer pair was used to identify and compare the strain with a 33 bp deletion with the wild type strain. The PCR product obtained from the mutant strains (201 deletion or 33 bp deletion) should be shorter than that from the wild type strain and would thus migrate faster through the agarose gel.

2.5 Targeted *eae* gene knockout using the CRISPR/Cas9 system

In this study, the CRISPR/Cas9 system was introduced into the strain FHI11 by a two-plasmid system, in which the *cas9* gene and the gDNA encoding the gRNA were separated in the pCas9 and the pCRISPR plasmids, respectively. Firstly, the gDNA (T30RNA) was constructed by phosphorylating and annealing the oligonucleotides T30RNA_F and T30RNA_R. Secondly, the pCRISPR_T30RNA plasmid was constructed by cloning the T30RNA into the pCRISPR plasmid. Then the pCRISPR_T30RNA plasmid and repair DNA template were co-transformed into the FHI1 strain that was already transformed with pCas9 plasmid.

2.5.1 Phosphorylation of T30RNA_F and T30RNA_R

Phosphorylation is a process that requires the T4 polynucleotide kinase to add a phosphate group to the 5’ termini of the oligonucleotides. This enzyme exchanges phosphate from the ATP to the 5’-hydroxyl terminuses of the DNA or RNA, as well as nucleoside 3’-monophosphates. This will provide the necessary terminal modification needed for the subsequent ligation step. This step was performed on the T30RNA_F and T30RNA_R oligonucleotides.

Procedure

The reagents listed in table 5 were prepared in a 1.5 ml Eppendorf tube and mixed well and incubated for 30 min at 37 ˚C using an Eppendorf thermomixer comfort [110].

Table 5. List of reagents used in the phosphorylation of the T30RNA_F and T30RNA_R oligonucleotides.

Material/ Reagents	**Volume (µl)**	**Manufacture**	**Catalog number**
T30RNA_F (100 µM)	1	SIGMA Life Science	-
T30RNA_R (100 µM)	1	SIGMA Life Science	-
10x T4 DNA ligase buffer	5	New England Biolabs	B0202S
T4 polynucleotide kinase	1	New England Biolabs	M0201S
MGW	42	Sigma-Aldrich	W4502
Final volume per reaction	50		

2.5.2 Annealing oligonucleotides

Annealing of T30RNA_F and T30RNA_R was performed to get a dsDNA template (T30RNA) that could be ligated into the pCRISPR plasmid using Golden Gate Assembly. The phosphorylated T30RNA_F and T30RNA_R oligonucleotides were used for this purpose.

Procedure

Annealing was performed according to the protocol established by König et al. [110]. After the phosphorylation step, 2.5 µl of 1 M sterile NaCl was added and the mixture was incubated at 95 °C for 5 min. Then the mixture was allowed to cool down to room temperature by taking the heating block out of the thermomixer for 2 hrs. The annealed oligonucleotide was diluted (1:2) with MGW for the subsequent ligation reaction and the remaining annealed oligonucleotides were stored at -20 °C.

2.5.3 Golden Gate Assembly

Golden Gate Assembly works by utilizing two simultaneous enzymatic activities in a single mixture: BsaI for endonuclease digestion and T4 DNA ligase for ligation. BsaI is a type IIS restriction enzyme that recognizes asymmetric sequences (5'… GGTCTC$(N)_1$… 3' and 3'…CCAGAG$(N)_5$…5') in the target DNA and cut outside of their recognition sequence as shown in figure 13. BsaI generates sticky ends so that the gDNA with complementary overhang can be ligated into the plasmid. The final assembly should no longer contain BsaI recognition sites, therefore no further cutting is possible and successfully assembled products

should accumulate after each cycle. A final incubation step at 60 ˚C for 5 min favours BsaI cutting in the absence of DNA ligation. This step ensures digestion of uncut plasmid containing the BsaI recognition site and reduces background colonies.

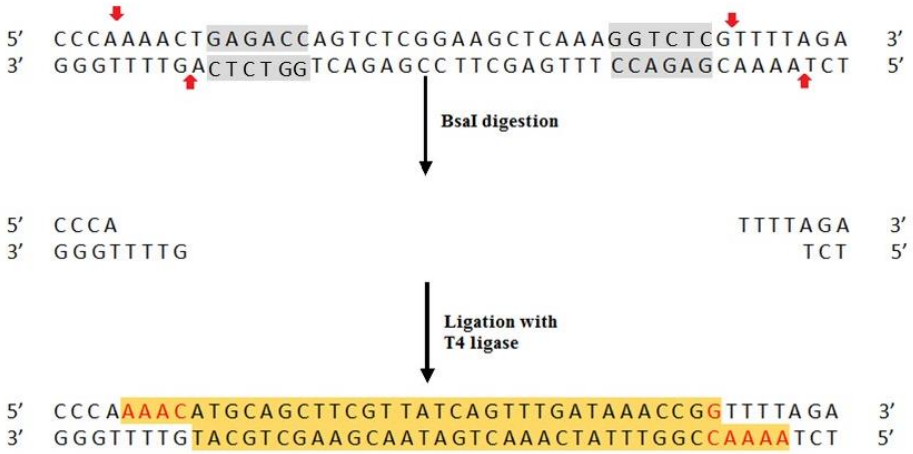


Figure 13. Illustration of the assembly of gDNA (T30RNA) into the pCRISPR plasmid. The BsaI recognition sites are highlighted in grey and T30RNA is highlighted in yellow. Bases shown in read are the overhangs added to the gDNA. Digested pCRISPR plasmid is ligated with gDNA with the help of the T4 ligase enzyme yielding a circular construct.

Procedure

Golden Gate Assembly was performed according to the procedure established by New England Biolabs (NEB) with some modifications [113]. Instead of using the NEB golden Gate enzyme mix, we used 0.5 µl of BsaI-HFv2 and 0.5 µl of T4 ligase enzymes for the reaction mix. The reaction mix containing all the reagents mentioned in table 6 was prepared in a 0.2 ml PCR tube. A 3:1 molar ratio of insert vs vector was used in the assembly. For one insert, New England Biolabs suggests incubating the reaction mixture for 5 min at 37 ˚C followed by 5 min at 60 ˚C. Since the pCRISPR plasmid contains four BsaI recognition sites, a cycling protocol alternating between 37 ˚C and 16 ˚C was selected to ensure complete digestion and successful assembly. A cycling program designed for 5-10 inserts was chosen for assembling (Table 7). The assembly was carried out in Bio-Rad T100 Thermal Cycler.

Table 6. List of reagents used in the Golden Gate Assembly.

Material	**Volume (µl)**	**Manufacturer**	**Catalog number**
pCRISPR plasmid (15 ng/ µl)	5	Addgene	#42875
T30RNA (3 ng/µl) (3:1 insert: vector molar ratio)	1	This project	This project
T4 DNA ligase buffer (x10)	2	New England Biolabs	B0202S
BsaI-HFv2	0.5	New England Biolabs	R3733
T4 DNA ligase	0.5	New England Biolabs	M0202S
MGW	to 20	Sigma-Aldrich	W4502

Table 7. The cycling program used in the Golden Gate Assembly.

Step		**Temperature**	**Time**
30 cycles	Optimal temperature for BsaI digestion within a temperature range for ligase stability	37 ˚C	1 min
	Optimal temperature for T4 ligase	16 ˚C	1min
	Favours BsaI cutting in the absence of DNA ligation	60 ˚C	5 min
	Hold	4 ˚C	∞

2.5.4 Methods for preparation of competent cells

Bacterial cells that can take up extracellular genetic material such as foreign DNA or ligation product from the environment are called competent cells. Bacteria can be made competent by chemical treatment followed by heat shock or by application of an electrical field that creates pores in the cell wall through which DNA can pass [114, 115] .

In this study, heat shock was used to transform the pCRISPR_T30RNA plasmid into MAX Efficiency® DH5α™ competent cells (Invitrogen). To obtain increased transformation efficiency, electroporation was chosen to transform the plasmids and the repair DNA template into electrocompetent FHI11 [110].

2.5.4.1 Making bacterial cells electrocompetent

Preparation of electrocompetent cells includes repeated cycles of washes and centrifugation steps of large volumes (50ml) of culture grown to mid-logarithmic phase of growth using non-ionic sterile cold water and sterile 10% glycerol. During the preparation great care must be taken to keep the cells, washing solutions and the equipment at ice-cold temperature throughout the entire washing procedure.

Procedure

Electrocompetent cells were prepared according to the protocol established by König et al. [110]. In this procedure, all the centrifugation steps were done at 4000 rpm at 4 °C using Eppendorf 5804 R centrifuge. The wild type STEC FHI11 strain was recovered from frozen stock and streaked on LB-agar and incubated at 37 °C overnight. A single colony from overnight culture was inoculated into 5 ml LB broth and incubated overnight at 37 °C with shaking at 200 rpm. The following day, the overnight culture was diluted to an OD_{600nm} of 0.1 in LB broth (50 ml) in an Erlenmeyer flask and incubated at 37 °C with shaking at 200 rpm until an OD_{600nm} of 0.6-0.8 was reached. The OD_{600nm} value was measured every 30 min using a UV-visible spectrophotometer (UV-1700 PharmaSpec). Then the culture was transferred into a pre-chilled 50 ml centrifuge tube and harvested by centrifugation for 20 min. The supernatant was discarded, and the pellet was gently resuspended in 50 ml ice-cold, sterile water. The resuspended pellet was then centrifuged for 10 min. This washing step was repeated 3 times. The resulting pellet was washed with ice-cold, sterile 10% glycerol and centrifuged for 10 min. Finally, the supernatant was discarded, and the pellet was resuspended in 250 µl ice-cold, sterile 10% glycerol and 50 µl aliquots were prepared in pre-chilled cryotubes and stored at -80 °C until future use.

2.5.5 Transformation of the pCRISPR_T30RNA plasmid into competent *E. coli* via heat shock

When competent bacterial cells are kept on ice, the low temperature initiates binding of foreign DNA to the surface of the bacterial membrane. A sudden increase in temperature from 0 °C to 42 °C creates pores in the bacterial cell membrane which allows the foreign DNA to enter the cell. Growing the transformed cells in LB broth without antibiotics allows self-healing. Cells that have taken up a plasmid carrying antibiotic resistance genes are able to grow on LB agar with selective antibiotics [114].

Procedure

Transformation of the pCRISPR_T30RNA plasmid into MAX Efficiency® DH5α™ competent cells was performed according to the manufacturer's protocol [105]. A 50 µl of chemically competent *E. coli* DH5α cells were thawed on ice for 10 min and mixed with 10 µl Golden Gate Assembly mix (section 3.5.3). The reaction mix was kept on ice for 30 min before transferred to an Eppendorf thermomixer comfort instrument set at 42 °C for 45 sec. Then the cells were kept on ice for 2 min before adding 940 µl of S.O.C media (supplied with the competent cells) stored at room temperature and incubated at 37 °C with shaking at 200 rpm for 1 hr. After the 1 hr incubation period, the culture was mixed well by inverting the tube and plated on LB agar containing 50 µg/ml kanamycin. Then, 100 µl and 250 µl of the culture were plated using a L-shaped spreader. The remaining volume was centrifuged to spin down the cells, and where then plated out. The LB agar plates were incubated overnight at 37 °C. The next day single colonies were analysed by colony PCR and Sanger sequencing to identify the positive clones harbouring the pCRISPR_T30RNA plasmid.

2.5.6 Transformation of plasmids and repair DNA into FHI11 via electroporation

Electroporation refers to the process that creates temporary pores in the cell membrane using electrical pulses and enables substances like nucleic acid to enter the cell through the pores [115]. The process requires electrocompetent bacteria, purified plasmid/DNA, a device that delivers the electrical impulse, a cuvette that carry the electrocompetent cells and nucleic acid, and a chamber to hold the cuvette.

Procedure

First, the pCas9 plasmid was transformed into wild type electrocompetent FHI11 cells and then the resulting FHI11 strain harbouring pCas9 was used for the mutagenesis step. Electroporation was performed in a BTX™ ECM™ 630 exponential decay wave electroporator. The settings were as follows: mode-high voltage (HV), 1800V, 25 µF capacitance and 200-ohm resistance.

To transform pCas9 plasmid into the FHI11 strain, 50 µl electrocompetent FHI11 cells were thawed on ice before the cells were gently mixed with 1 µl of the pCas9 plasmid (1ng/µl) and transferred into a pre-chilled 1 mm electroporation cuvette. After electroporation, 950 µl of

room tempered LB media was immediately added and the mixture was gently resuspended. Then the resuspended mixture was transferred into a sterile tube and incubated at 30 ˚C for 3 hrs with agitation at 200 rpm. After incubation, the culture was then mixed well by inverting the tube, and 100 µl and 250 µl of culture was plated using a L-shaped spreader on pre-warmed LB agar plates containing 25 µg/ml chloramphenicol. The plates were incubated overnight at 37 ˚C. One of the colonies obtained from the transformation step and harbouring the pCas9 plasmid was made electrocompetent for the mutagenesis step as described in section 3.5.4.1 (chloramphenicol was added when necessary).

To generate a 201 bp deletion in the *eae* gene, the 120 bp long dsDNA, constructed by annealing 10 µg each of De201RT_F and De201RT_R in a 20 µl volume, was used. To generate a 33 bp deletion in the *eae* gene, 10 µg of Del33RT_F (ssDNA) was used.

For the mutagenesis step, 50 µl electrocompetent FHI11 strain harbouring pCas9 was mixed with 100 ng pCRISPR_T30RNA and 10 µg ssDNA (33 bp deletion) or dsDNA (201 bp deletion) and the mixture was electroporated before it was incubated at 30 ˚C with agitation at 200 rpm for 3 hrs in 950 µl of LB broth. Then the mixture was spread on LB agar containing 25 µg/ml chloramphenicol and 50 µg/ml kanamycin and incubated overnight at 37 ˚C. 100 µl of culture was plated on a selective LB agar before the culture was centrifuged and the cell pallet was suspended in 100 µl of LB broth and plated on LB agar. For each length of deletion 10 replicates were done. Following day, single colonies obtained from the mutagenesis step were analysed by colony PCR. A control was included to confirm the transformation of pCRISPR plasmid. Therefore, 100 ng of empty pCRISPR plasmid was electroporated with 50 µl electrocompetent FHI11 cells harbouring pCas9. and recovered as mentioned before. 1 µl of recovered culture was diluted up to 100 µl in LB broth and plated on on LB agar containing 25 µg/ml chloramphenicol and 50 µg/ml kanamycin and incubated overnight at 37 ˚C. Number of colonies obtained on a plate was used to calculate the transformation efficiency.

Transformation Efficiency Equation:
Transformation Efficiency (CFU/µg) = Colonies / µg / Dilution factor

Colonies = Number of colonies counted on a plate
µg = the amount of plasmid DNA transformed expressed in µg
Dilution factor = The total dilution of the plasmid DNA before plating

2.6 Methods used for DNA isolation

2.6.1 Plasmid isolation

Quantum Prep® Plasmid Miniprep kit (Bio-Rad) uses a silicon dioxide exoskeleton of diatoms as DNA binding matrix [116]. Initial material for plasmid DNA extraction may be obtained by spinning down overnight bacterial culture. Cell resuspension solution is used to resuspend the pellet, which contains RNase to destroy the RNA and at the same time protect DNA in the sample. The lysis solution contains sodium dodecyl sulphate (SDS) and sodium hydroxide (NaOH) which lyse cells and denature proteins. SDS act as a detergent and dissolves phospholipid membranes, which leads to lysis of the bacterial cell wall and releases cellular contents. NaOH increases the pH of the reaction mixture and leads to denaturation of proteins and dsDNA to ssDNA and degrades RNA. Neutralization solution containing potassium acetate shifts the pH from basic to a slightly acidic condition thereby allowing renaturation of plasmid DNA. Subsequent washing steps with washing buffer containing 95% ethanol removes impurities and leaves high quality plasmid DNA bound to the matrix. The bound plasmid DNA can be eluted using low ionic strength TE buffer with pH 8.0 or MGW by centrifugation at top speed.

Procedure

Plasmid DNA was isolated according to the manufacturers' protocol (Quantum Prep Plasmid Miniprep Kit, Bio Rad). In this procedure, all the centrifugation steps were carried out at 13,200 x g at room temperature using an Eppendorf 5415R centrifuge. A single colony was inoculated into 2 ml LB broth (if necessary appropriate antibiotic was added). The culture was incubated overnight at 37 °C with shaking at 200 rpm. Overnight culture was transferred into a microcentrifuge tube and centrifuged for 30 sec. The supernatant was removed and 200 µl resuspension solution was added and mixed well by vortexing until the pellet was completely resuspended. Then the resuspended cells were lysed by adding 250 µl of the cell lysis solution and mixed by gently inverting the capped tube about ten times. The resulting solution was neutralized using 250 µl of the neutralization solution. Thereafter debris in the suspension was pelleted by centrifugation for 5 min. After centrifugation, clear supernatant was transferred into a spin filter, placed on 2 ml microcentrifuge wash tube and mixed well with 200 µl thoroughly suspended Quantum Prep matrix. The suspension was centrifuged for 30 sec. The filtrate was discarded and 500 µl of wash buffer was added and centrifuged for 30 sec. The washing step was repeated using 500 µl wash buffer. After the final washing step, the spin

column was placed in a 1.5 ml collection tube and 50 µl TE buffer/MGW was added into the column to elute the bound plasmid. The mixture was then centrifuged for 1 min at top speed and the eluted plasmid was stored at -20 °C.

2.6.2 DNA extraction by thermal lysis

Heat lysis is a method that can be used for simple lysis of bacterial cells for use in PCR. The heating step will break the cell membrane and causes the release of intracellular components including the bacterial plasmid and chromosomal DNA, so that it can serve as a template in the PCR. Individual colonies are lysed in water with a short heating step and then centrifuged at high speed [117].

Procedure

A 1 µl loop was used to transfer part of a bacterial colony into a 1.5 ml tube containing 100 µl of sterile MGW, and the mixture was then incubated on an Eppendorf thermomixer comfort instrument at 95 °C for 10 min, before the mixture was centrifuged at 13,200 x g for 2 min at room temperature. 1 µl of the supernatant was used for PCR and the remaining volume was transferred into new sterile 1.5 ml Eppendorf tube and stored at -20 °C. Another part of the same colony was inoculated into 2 ml LB broth containing appropriate antibiotics.

2.7 Quantification methods for DNA concentration

Accurate quantification of the concentration of DNA is essential for efficient cloning. The purity of plasmid DNA should be measured since contamination of the sample can inhibit the accuracy of the cloning process.

2.7.1 Qubit

The Qubit 3.0® Fluorometer (Thermo Fisher Scientific) utilizes fluorescence-based quantitative assays to measure the concentration of nucleic acids and proteins. Fluorescent dyes used in the Qubit assay are specific to the target molecules and only emits light when bound to the target molecule, thus avoiding overestimation caused by contaminants. The Qubit® dsDNA HS (High Sensitivity) assay kit includes the Qubit® dsDNA HS reagent,

buffer and two standards (HS 1 and HS 2). The assay kit has a detection range 0.2-100 ng and is selective for dsDNA [118].

Procedure

Quantification of plasmid DNA was performed according to the manufacturer´s protocol for the Qubit 3.0® Fluorometer [118]. Two tubes for the standards and one for each sample were prepared for the assay. For each reaction, thin walled clear 0.5 ml tubes were used. First, the Qubit working solution was prepared by diluting 1 µl x n of Qubit® dsDNA HS reagent in 199 µl x n of Qubit® dsDNA HS buffer (n = numbers of standards plus number of samples). Working solution, standards and DNA samples were aliquoted into each tube according to the table below (Table 8). The resulting mixture was vortexed for 2-3 sec and incubated at room temperature for 2 min. Readings were taken by inserting the tubes in the Qubit® Fluorometer. A 2-point calibration was done using Qubit® dsDNA HS standard 1 and 2 before reading the samples. After successful calibration, tubes containing samples were measured.

Table 8. Preparation of standard and sample reaction tubes for measurement.

	Standard assay tubes	**Sample assay tubes**
Working solution (µl)	190	198
Standard (Low/ High) (µl)	10	-
DNA Sample (µl)	-	2
Total volume (µl)	200	200

2.7.2 Nanodrop

A Nanodrop ND-1000 Spectrophotometer (Thermo Fisher Scientific) is used to measure the concentration and purity of nucleic acids. Nucleic acids such as RNA, ssDNA and dsDNA absorb at 260 nm and contribute to the total absorbance of the sample. The purity of the sample is analyzed by determining the ratio of the absorbance at 260 nm divided by the absorbance at 280 nm (260/280). An A_{260}/A_{280} ratio of about 1.8 is considered to indicate high purity of DNA, and ratio of about 2.0 for high purity of RNA. If the ratio is lower, this may indicate the presence of contamination with e.g. proteins or phenol that absorb at 280 nm. Measuring absorbance around 230 nm is used as a secondary measure of DNA purity. EDTA, carbohydrate, and phenol have a strong absorbance at 230 nm. A ratio of 260 nm to 230 nm is

used to estimate salt carryover, with an expected range of 1.8-2.2 in a mixture of DNA of high purity [119].

Procedure

First the Nanodrop ND1000 instrument was initialized by pipetting 2 to 3 μl deionized water onto the lower optical surface. The lid was then closed to ensure that both surfaces come in contact with the water and initialization was done. Then a blank measurement was done using 1 μl of elution buffer/MGW. After the blanking procedure, samples were measured. The software automatically calculates and displays the nucleic acid concentration and purity ratios. Sample quality was accessed by analyzing the purity ratios and overall spectral quality. Between each measurement, clean, dry, lint-free lab wipe was used to clean the optical surfaces.

2.8 Polymerase Chain reaction (PCR)

PCR is a very useful method to amplify a nucleotide region of interest. PCR utilizes the ability of the DNA polymerase to synthesise new DNA strands complementary to the provided DNA template. PCR contains several components: DNA polymerase, primers, deoxynucleotides (dATP, dCTP, dGTP and dTTP), cations ($MgCl_{2)}$, buffer and a DNA template. PCR is a three-step process (denaturation, annealing, and extension) that is carried out in repeated cycles. Each step has a specific temperature and an incubation period. During denaturation, the dsDNA templates will be desaturated to ssDNA. Once the strands are denatured, the temperature is reduced to the annealing temperature, allowing the primers to base pair to a complementary sequence found in the target DNA. During the extension step, the polymerase uses dNTPs to extend the 3’ end of primers to form a complementary DNA strand. $MgCl_2$ acts as a cofactor that catalyses the DNA polymerase and the buffer maintains the pH of the reaction mix [120].

The sensitivity and the specificity of the PCR reaction depend on several parameters, one of which is the annealing temperature [121]. Gradient PCR can be done for the determination of the optimal annealing temperature. This optimization can be often achieved in a single run.

Procedure

In this study, thermally lysed bacterial colonies (Section 2.6.2) were used as template in the PCR reactions. This technique, often termed colony PCR is a rapid method because it skips the plasmid and/or chromosomal DNA extraction step [117]. In this study colony PCR was used to verify the pCRISPR_T30RNA construction and to validate gene deletion following the mutagenesis step. PCR amplification was carried out on a Bio-Rad T100 Thermal Cycler. After completion of PCR, the generated PCR products were verified by gel electrophoresis (2 % agarose).

2.8.1 Optimization of the annealing temperature for PCR

In this study, gradient PCR was done for 2 sets of primers: (1) primers used to verify the pCRISPR_T30RNA construction (InsertID_F and InsertID_R), and (2) primers used to verify the 201 bp or 33 bp deletion in FHI11 chromosomal DNA (KOveri_F and KOveri_R). For both primer sets, the total reaction volume was 25 µl in a 0.2 ml PCR tube containing the reagents mentioned in table 9.

Table 9. List Reagents used in conventional PCR.

Reagents	**Volume (µl)**	**Manufacturer**
dNTP	1.25	Invitrogen
10x PCR buffer	2.5	Applied Biosystems
Forward primer (10 µM)	1.25	
Reverse primer (10 µM)	1.25	
AmpliTaq Gold (5 Units/µl)	0.25	Applied Biosystems
DNA template	1	This project
MGW	17.5	Sigma-Aldrich
Total volume in each tube	25	

A gradient of annealing temperate was set using the heating block of Bio-Rad T100 Thermal Cycler. Details about the PCR thermocycler conditions for both primer pairs are presented in Appendix B. Seven identical PCR reactions were set up and PCR was done at different annealing temperatures. The annealing temperature that gave the strongest band of correct size and fewest bands of other sizes was chosen for analysis of samples.

2.8.2 Colony PCR to verify the pCRISPR_T30RNA construction

Colony PCR was used to verify the pCRISPR_T30RNA construct in *E. coli* DH5α cells after transformation of the pCRISPR_T30RNA plasmid by heat shock. PCR was done using the primers InsertID_F and InsertID_R targeting the pCRISPR plasmid backbone and flanking the gDNA insertion site in a total reaction volume of 25 µl containing reagents listed in table 9 in a 0.2 ml PCR tube. The PCR thermocycler conditions are presented in table 10. An intact pCRISPR plasmid was used as a negative control. No template PCR reaction was also done in parallel to detect any contamination causing false positive results

Table 10. Thermocycler conditions used in conventional PCR.

Steps	**Temperature**	**Time (Minutes)**
1. Enzyme activation	94 ˚C	10:00
2. Denaturation	94 ˚C	00:30
3. Annealing	63 ˚C	00:30
4. Extension	72 ˚C	01:00
Repeat step 2-4 for 25 cycles		
5. Final extra elongation	72 ˚C	05:00

2.8.3 Colony PCR to verify the deletion in the FHI11 chromosomal DNA

After the mutagenesis step, isolated colonies were analysed by colony PCR. The same reaction condition was used as previously mentioned for PCR amplification of the pCRISPR_T30RNA plasmid. The *eae* gene locus was amplified using the primer pair KOveri_F and KOveri_R for this purpose. Wild type FHI11 strain was used as the control.

2.9 Agarose Gel Electrophoresis

Agarose gel electrophoresis is used to separate DNA and RNA fragments of varying sizes. The separation is based on mass/charge ratio of molecules in an electric field. The resulting negative charge of DNA caused by phosphate backbone allows the DNA to migrate towards a positively charged anode. DNA has a uniform mass/charge ratio which allows DNA molecules to be separated by size so that distance migrated is inversely proportional to the log of its molecular weight. The migration rate of molecules through the gel depends on the following factors: 1) size of the DNA molecule, 2) concentration of agarose in the gel, 3) DNA conformation, 4) applied voltage, 5) type of agarose, 6) presence of stain such as ethidium bromide, and 7) type of buffer used in electrophoresis. The separated DNA molecules can be stained with an appropriate dye and visualized under UV light [122].

Procedure

Agarose gel electrophoresis was used to analyse the products from colony PCR. A 2% agarose gel was prepared as described in Appendix A. The molten agarose was poured carefully into a Mini-Sub Cell GT electrophoresis system (Bio-Rad) with casting gates and comb and left for 30 min to solidify. Once the gel was cooled the casting gates and comb were carefully removed and the gel was covered with 0.5X TBE-buffer (enough to cover both electrodes). To prepare the samples to load into the gel, 1 µl 10 x BlueJuice gel loading buffer (Invitrogen) was added to 10 µl of 1:10 diluted 1 Kb Plus DNA Ladder and PCR amplified samples in a microtiter plate. This buffer contains the colour bromophenol blue, which makes loading and tracking of DNA samples in the agarose gel easier. Then the samples were loaded into the wells. DNA ladder was added in the first well and when necessary an additional well was loaded after the sample wells. Then the gel was covered with a lid and electricity was provided using a PowerPac Basic (Bio-Rad) power source, and the gel was set to run at 75V for 1 hr or at 150V for 2 hrs depending of the electrophoresis cell used. After 1 hr (or 2 hrs) the power supply was turned off, cables were disconnected, and the gel was removed from the apparatus and stained for 30 min using a staining bath containing GelRed (Biotium). The stained gel was transferred into RO-water for brief rinsing before it was inspected in a Bio-Rad Gel DOC^{TM} XR instrument under UV exposure.

2.10 Sanger sequencing

Sanger sequencing was performed to verify that the gDNA was successfully ligated into the pCRISPR plasmid. Primers specific for the pCRISPR plasmid was used for this purpose. PCR was first performed to amplify the specific region flanking the gDNA insertion site as described in section 2.8.2. Then gel electrophoresis was performed to verify the length of PCR products. Only the colonies which have given the expected length of PCR product were used for sequencing. The PCR products were treated with a PCR cleanup enzyme kit (New England Biolabs), which contains Exonuclease I (Exo I, #M0293) and Shrimp Alkaline Phosphatase (rSAP, #M0371). Exo I degrades the remaining PCR primers and rSAP dephosphorylates the dNTPs left from a PCR reaction.

In this study, sequencing was carried out with GenomeLab DTCS (Dye Terminator Cycle Sequencing) Quick Start kit and a fully automated capillary electrophoretic genetic analysis system (Beckman Coulter CEQ 8800). The GenomeLab DTCS Quick Start master mix contains reaction buffer, DNA polymerase, deoxynucleotide triphosphate, and dye-labeled dideoxynucleotide triphosphate terminators. Once amplified DNA template is loaded into the CEQ 8800, labeled DNA fragments will be automatically denatured and then separated by capillary electrophoresis. The fragments are detected by laser-induced fluorescence in four spectral channels. After the separation, base sequences will be automatically base called from the four-channel raw data sets generated from each capillary. Preparation for the DNA sequencing reaction, thermal cycling and ethanol precipitation were performed according to Beckman Coulter's protocol [123].

Procedure

As a first step, the amplification product from the positive clone was treated with a PCR cleanup enzyme as recommended by the manufacturer. The reagents listed in table 11 were added to a 1.5 ml Eppendorf tube and incubated for 15 min at 37 °C. Then the mixture was incubated for 15 min at 80 °C on a Techne Dri-Block® heater to inactivate the enzymes. The concentrations of PCR amplified products following PCR cleanup was measured using the Qubit 3.0® Fluorometer. The amount of DNA template necessary for the sequencing reaction was obtained from a table for estimating the dsDNA concentration from Beckman Coulter. It is important to measure the concentration of DNA template to ensure that a minimum molar ratio of primer to template of 40:1 is obtained and to confirm that not too

much DNA template is used in the reaction. According to Beckman Coulter's protocol, 10 ng (50 fmol) of DNA template (302 bp) was added to the sequencing reaction. The volume of purified PCR product for each sequencing reaction was calculated using the following formula:

$$\text{Volume of purified DNA temlate } (\mu l) = \frac{10 ng}{\text{Concentration of DNA template after purification step } (\frac{ng}{\mu l})}$$

The sequencing reaction was prepared in a 0.2 ml PCR tube with the reagents listed in table 11. DNA sequencing reactions were run on a T100 Thermal Cycler (Bio-Rad) according to the following protocol: 30 cycles of denaturation at 96 °C for 20 sec, annealing at 50 °C for 20 sec and extension at 60 °C for 4 min.

After the amplification reaction, ethanol precipitation was performed to remove the reagents used in amplification reaction. Fresh stop solution was prepared according to table 11 and added into each sequenced sample. Then 60 µl of ice-cold 95% ethanol was added and mixed well and immediately the sample was centrifuged at 14,000 rpm at 4 °C for 15 min using an Eppendorf Centrifuge 5415 R. The supernatant was carefully removed without disturbing the pellet using a micropipette. The pellet was rinsed two times with 200 µl of ice-cold 70% ethanol. For each rinse, the sample was centrifuged at 14,000 rpm at 4 °C for 2 min. The pellet was left to dry at room temperature for 45 min to allow residual ethanol to evaporate. The dried pellet was resuspended in 40 µl of sample loading solution (provided with the kit), transferred to a sample plate and overlaid with one drop of light mineral oil (provided with the kit) to prevent evaporation, protect the dyes and increase their stability.

Table 11. List of reagents used for PCR cleanup, Sanger sequencing reaction and ethanol precipitation.

Method	**Material**	**Volume (µl)**	**Manufacturer**
PCR cleanup	PCR amplified products	5	This project
	Exo I	0.5	New England Biolabs
	rSAP	1	New England Biolabs
	Total volume in each tube	**6.5**	
Preparation of the DNA sequencing reaction	MGW	0-14	Sigma- Aldrich
	DNA template	0.5 - 14.5	This project
	Forward Primer (1.6 µM) or Reverse primer (1.6 µM)	2	Thermo Fisher Scientific SIGMA Life Science
	DTCS Quick start master mix (10x)	2	Genomelab
	Sequence reaction buffer	1.5	GenomeLab
	Total volume in each tube	**20**	
Ethanol precipitation	3 M Sodium Acetate (pH 5.2)	2	Sigma- Aldrich
	100 mM Na2-EDTA (pH 8.0)	2	
	20 mg/ml Glycogen	1	Roche Germany
	Sequence product	20	This project
	Total volume in each tube	**25**	

3 Results

3.1 Choosing a type of mutation to knockout the *eae* gene

In this study the CRISPR/Cas9 editing tool was chosen for knockout of the *eae* gene. The β-domain of the gene was selected as the target. A small deletion and a stop codon within the β-domain of *eae* gene was chosen instead of deleting the complete 2808 bp long *eae* gene, since the efficiency of recombination decreases with the size of the chromosomal deletion [124]. The deletion included the protospacer and PAM region to prevent further cleavage by Cas9, as well as an in-frame stop (TAA) codon which should result in a truncated protein [103]. Two mutagenic DNA templates were designed to generate either a 201 bp deletion or a 33 bp deletion believing that one of the repair templates would generate the expected mutant strain.

3.2 Gene editing using CRISPR/Cas9 system

To inactivate the *eae* gene using the CRISPR/Cas9 system, three main elements were utilized: the pCas9 plasmid encoding Cas9, the pCRISPR plasmid encoding gRNA, and the repair DNA template.

In this study, the type II CRISPR/Cas9 system was used for two purposes; firstly, to introduce a DSB at a specific target site in *eae* gene that subsequently induce the HDR, and secondly to select for cells containing the desired deletion by killing the wild type strain.

First, the forward (T30RNA_F) and the reverse (T30RNA_R) strands of gDNA were annealed, and then cloned into the pCRISPR plasmid by Golden Gate Assembly to construct the pCRISPR_T30RNA plasmid. Then the FHI11 strain was transformed with the pCas9 plasmid to produce cells expressing the Cas9 protein and the resulting strain was co-transformed with the pCRISPR_T30RNA plasmid and the synthetic DNA repair template. The main steps of the experiments as well as the steps where it was possible to verify results, are shown in figure 14.

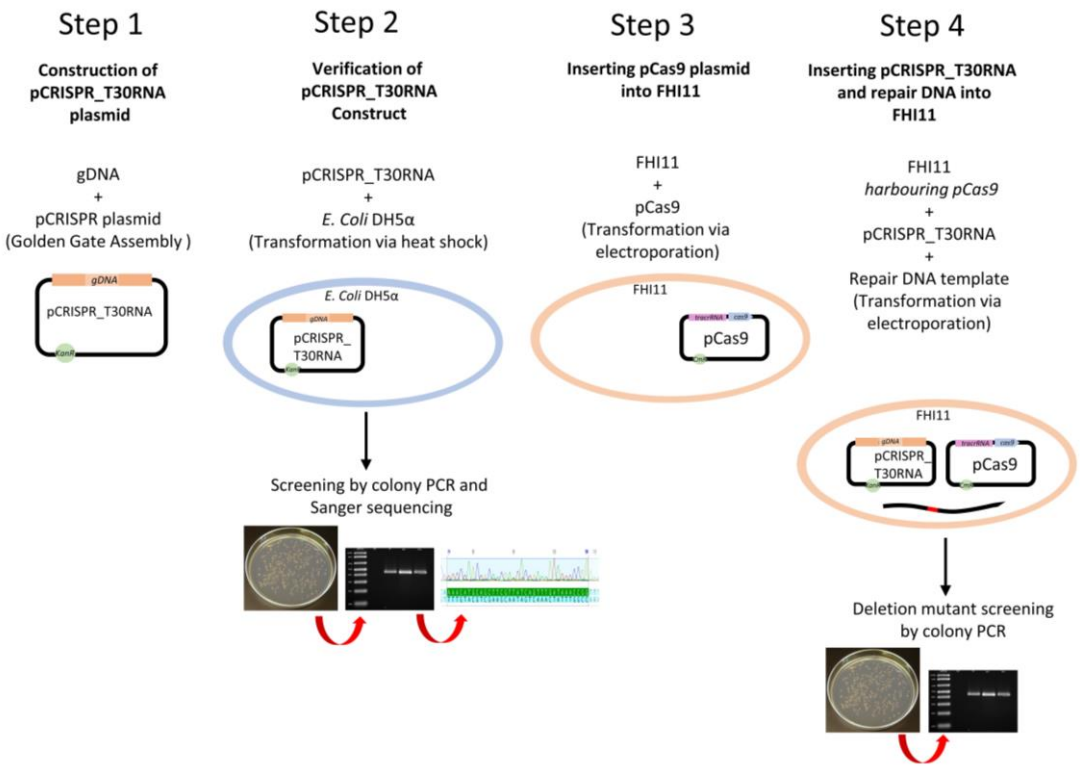
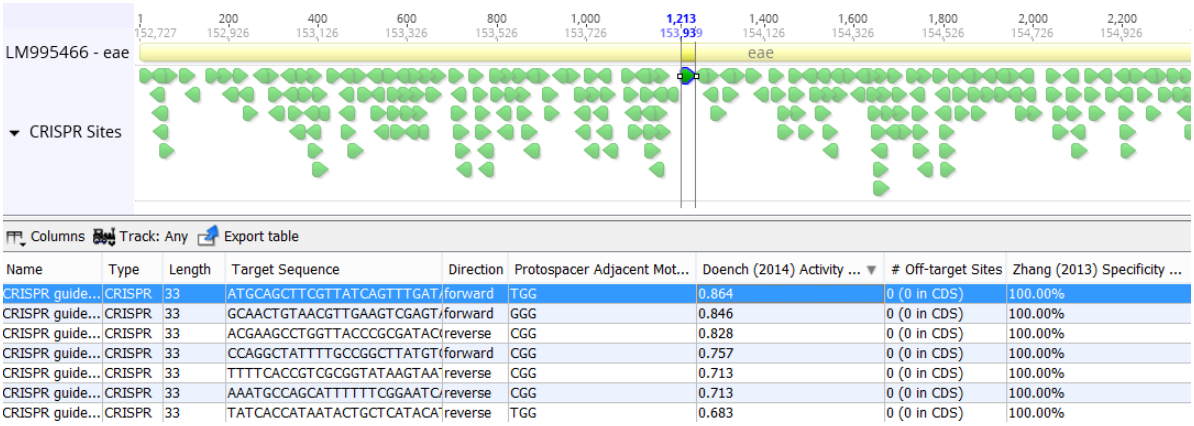


Figure 14. Simplified workflow for the *eae* gene knockout in FHI11 using CRISPR/Cas9 technology. Step 1 describes the construction of the targeting plasmid (pCRISPR_T30RNA) using Golden Gate Assembly. Step 2 includes the transformation of the pCRISPR_T30RNA into *E. coli* DH5α via heat shock and confirming the presence of gDNA in the pCRISPR plasmid by colony PCR and Sanger sequencing (section 3.4). Step 3 includes the transformation of pCas9 into electrocompetent FHI11 cells. Step 4 is the mutagenic step where the pCRISPR_T30RNA plasmid and repair DNA template are electroporated into the electrocompetent FHI11 cells harbouring pCas9 plasmid. The next day colony PCR is performed to identify strains with the deletion mutant (section 3.5).

3.2.1 Design of gDNA for CRISPR/Cas9 system

The guidelines for the gDNA design mentioned in section 2.4.1 were followed in the designing process. The "Find CRISPR sites" tool in Geneious R11 was used to identify a suitable gDNA sequence. This software provided several 30 nt long gDNA with respective PAM sequence (NGG) (Figure 15). A gDNA was chosen based on the location in *eae* gene, as well as calculated on-target activity and off-target effect.



Name	Type	Length	Target Sequence	Direction	Protospacer Adjacent Mot...	Doench (2014) Activity ...	# Off-target Sites	Zhang (2013) Specificity ...
CRISPR guide...	CRISPR	33	ATGCAGCTTCGTTATCAGTTTGAT	forward	TGG	0.864	0 (0 in CDS)	100.00%
CRISPR guide...	CRISPR	33	GCAACTGTAACGTTGAAGTCGAGT	forward	GGG	0.846	0 (0 in CDS)	100.00%
CRISPR guide...	CRISPR	33	ACGAAGCCTGGTTACCCGCGATAC	reverse	CGG	0.828	0 (0 in CDS)	100.00%
CRISPR guide...	CRISPR	33	CCAGGCTATTTTGCCGGCTTATGT	forward	CGG	0.757	0 (0 in CDS)	100.00%
CRISPR guide...	CRISPR	33	TTTTCACCGTCGCGGTATAAGTAA	reverse	CGG	0.713	0 (0 in CDS)	100.00%
CRISPR guide...	CRISPR	33	AAATGCCAGCATTTTTTCGGAATC	reverse	CGG	0.713	0 (0 in CDS)	100.00%
CRISPR guide...	CRISPR	33	TATCACCATAATACTGCTCATACA	reverse	TGG	0.683	0 (0 in CDS)	100.00%

Figure 15. Screenshot from Geneious R11 of the possible CRISPR/Cas9 sites (shown as green bars) found in the *eae* gene. The table below shows the on-target activity and the off-target sites for each CRISPR site.

As I mentioned before, targeting the 5' moiety is often used for knockout of a gene. However, we could not find any CRISPR site near the 5' moiety which has a high on target activity, and the first two CRISPR sites with high on target activity (0.963 and 0.902) were found near the 3' moiety of the *eae* gene. Therefore, a CRISPR/Cas9 target site was selected within the β-domain, at the site with the highest on-target activity and which had no off-target effects. The on-target activity for the chosen gDNA was estimated to be 0.864 (activity score ranging from 0 to 1). The off-target site was estimated to 100% (score ranging from 0 to 100%), indicating that gDNA should not bind to more than one site in the FHI11 genome. The nucleotide sequence of the forward (T30RNA_F) and reverse (T30RNA_R) oligonucleotides, as well as the annealed gDNA is shown in figure 16.

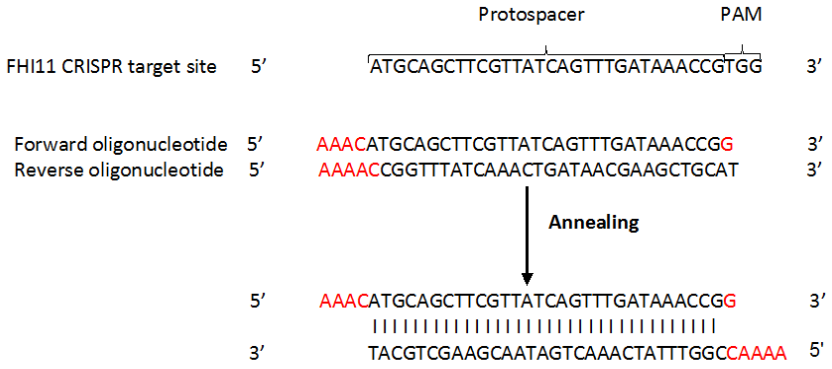


Figure 16. Illustration of the design of the forward (T30RNA_F) and reverse (T30RNA_R) oligonucleotides of the gDNA. Five additional nucleotides included to support the ligation are shown in red. Annealed oligonucleotides (T30RNA or gDNA) are shown at the bottom.

3.2.2 Design of primers for verification of the pCRISPR_T30RNA construct

The guidelines for primer design listed in section 2.4.3 were followed to design primers for the colony PCR to verify the pCRISPR_T30RNA construct. The PCR was designed to generate different sizes of PCR products, depending on whether the pCRISPR plasmid contained the gDNA or not.

Due to the duplicated DR-BsaIspc-DR region in the pCRISPR plasmid, it was not possible to find a forward primer which would not anneal at two identical sites in the intact plasmid. During the cloning step, a 305 bp of the pCRISPR plasmid would be cut in three separate fragments (Figure 17 and 18), and subsequently replaced by a 35 bp long gDNA. After successful digestion with BsaI and insertion of the gDNA, PCR would be expected to produce only one amplification product. Even though the forward primer did not fulfil the requirement in primer design and binds to more than one site on the undigested target plasmid, it was chosen because it would still serve its purpose which was to identify colonies containing a successfully cloned pCRISPR plasmid. The nucleotide sequences of the selected primers are presented in table 4.

Intact/undigested plasmid would produce two PCR products of 576 bp and 302 bp, while PCR of a plasmid which has undergone successful cloning would result in a single PCR product of 302 bp.

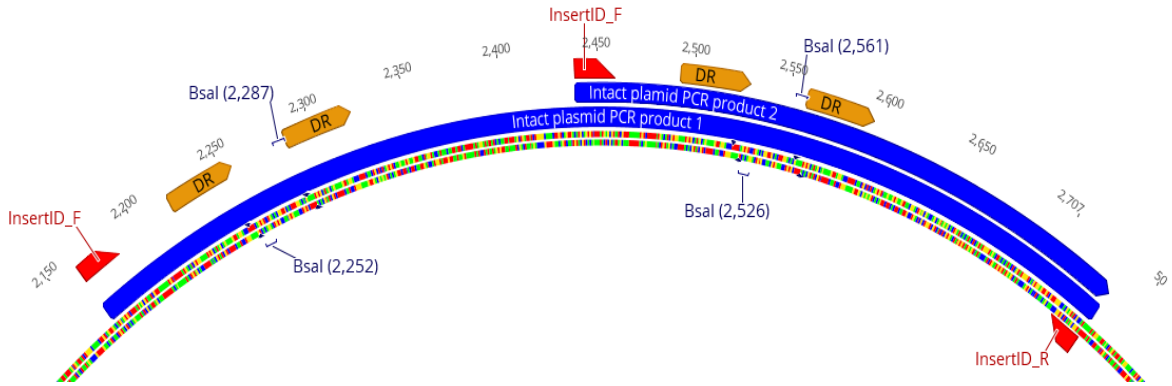


Figure 17. Screenshot from Geneious R11 of part of the pCRISPR plasmid. The InsertID_F and InsertID_R primer binding sites are shown in the intact pCRISPR plasmid. Primers are shown as red bars. Duplicated DR-BsaIspc-DR regions and the four BsaI recognition sites are shown as yellow bars. The two PCR products obtained as a result of the two InsertID_F binding sites on the intact pCRISPR plasmid are shown as blue bars.

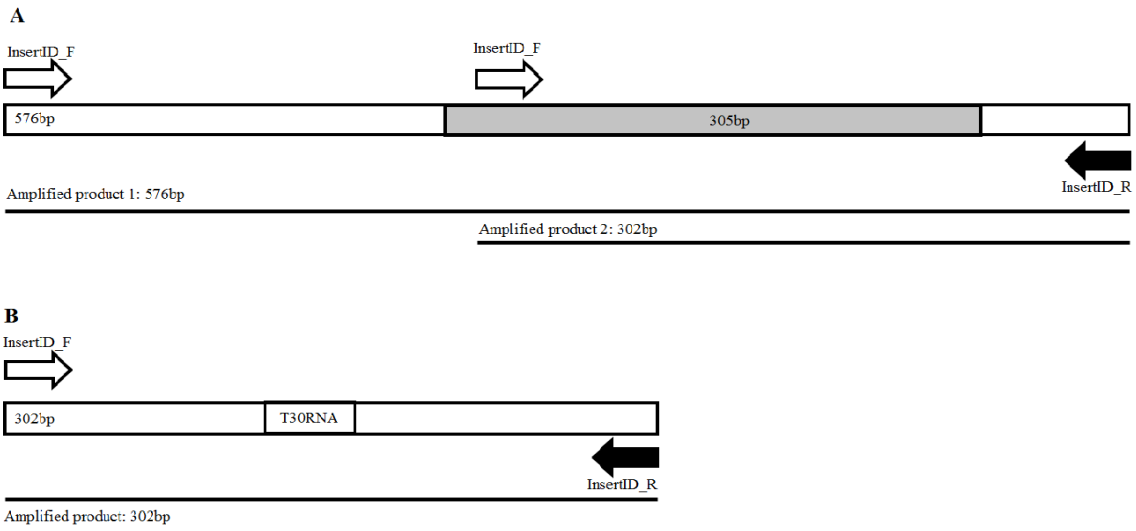


Figure 18. Simplified illustration of the relative position of the primers used for amplification of the region of interest in the pCRISPR plasmid. A) The InsertID_F and InsertID_R primer binding sites are shown in the intact pCRISPR plasmid. The grey area shows the 305 bp sequence that is removed following the digestion step (includes the 2nd InsertID_F primer binding site). B) After the original sequence is replaced with the gDNA in the pCRISPR plasmid, colony PCR using the InsertID_F and InsertID_R primers will produce only one 302 bp long PCR amplification product.

3.3 Optimization of the annealing temperature for PCR

3.3.1 Gradient PCR using the InsertID_F and InsertID_R primers

The intact pCRISPR plasmid was used as DNA template to find an optimal annealing temperature of the colony PCR using the InsertID_F and InsertID_R primers. As mentioned above, this PCR should generate two PCR products (576 bp and 302 bp) when the pCRISPR was used as template. The gradient analysis was done in a single run by setting a 6 ˚C gradient ranging from 60 ˚ to 66 ˚C using the gradient function of the heating block of the PCR instrument. A thermally lysed colony of DH5α *E. coli* harbouring the pCRISPR plasmid was used for this experiment. A PCR master mix without any DNA template was used as a negative control, with MGW added instead of DNA template. Following colony PCR, the amplified products were analysed by gel electrophoresis on a 2% agarose gel. The resulting gel image is presented in figure 19.

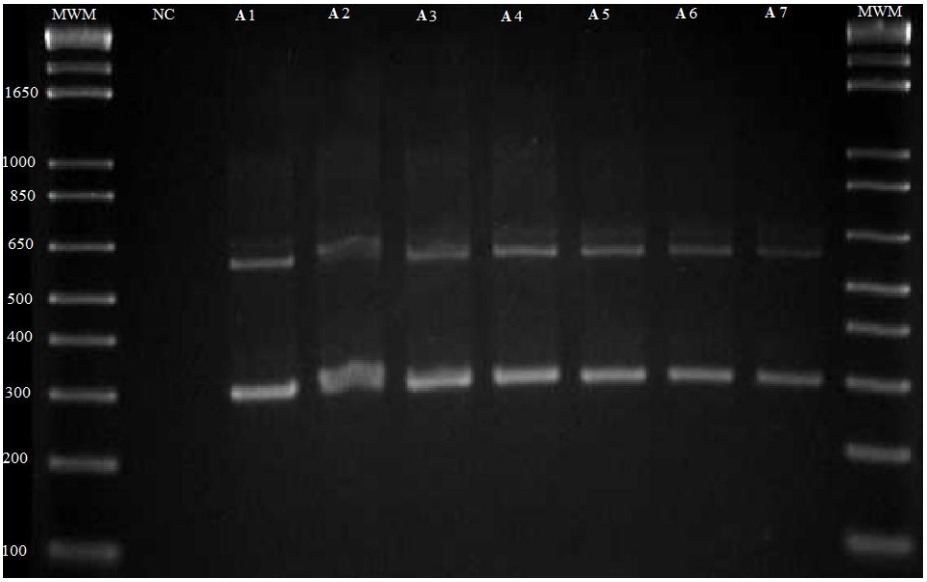


Figure 19. Experimental determination of the optimal annealing temperature for a PCR with the InserID_F and InserID_R primers. All reactions were done in a single run. Seven different annealing temperatures were used for this purpose. A1: 60 ˚C, A2: 61.2 ˚C, A3: 62.3 ˚C, A4: 63.7 ˚C, A5: 64.8 ˚C, A6: 65.6 ˚C, A7: 66 ˚C, NC: negative control (MGW), MWM-1 Kb Plus DNA Ladder molecular weight marker.

Two amplification products were found with sizes between 500 and 650 bp (~576 bp) and at about 300 bp (~302 bp) in the gel. This strongly indicates that the primer pair did bind to the expected sites on the pCRISPR plasmids, as well as to no off-target binding sites on the pCRISPR plasmid or chromosomal DNA of the host cell DH5α *E. coli*. Amplification was most efficient with an annealing temperature between 60 ˚C to 65.6 ˚C, while the yield was lower for both products at 66 ˚C. A temperature of 63˚C was chosen as annealing temperature for this PCR for further use in this study.

3.3.2 Gradient PCR using the KOveri_F and KOveri_R primers

Gradient analysis was also done for the PCR with the KOveri_F and KOveri_R primers. Chromosomal DNA of the wild type FHI11 strain was used as template, expected to generate a product of 436 bp. This PCR was done in a single run by setting a 7 ˚C gradient ranging from 58 ˚ to 65 ˚C. Following colony PCR, the amplified products were analysed by gel electrophoresis on a 2% agarose gel. The resulting gel image is presented in figure 20.

Presence of one amplified product located between the bands of the DNA standard of 400 and 500 bp correlated with the expected size of the amplification product of 436 bp. This

indicated that the primer pair did bind to the expected locus in the *eae* gene and no off-target binding sites on the bacterial genomic or on the plasmid DNA. Since there were no unspecific PCR products, only the yield was considered when selecting an optimal annealing temperature. The yield in amplification product was comparable for all temperatures tested, except at 65 ˚C where a slight reduction in yield was observed. Based on these results an annealing temperature of 63 ˚C was chosen for use also with this PCR.

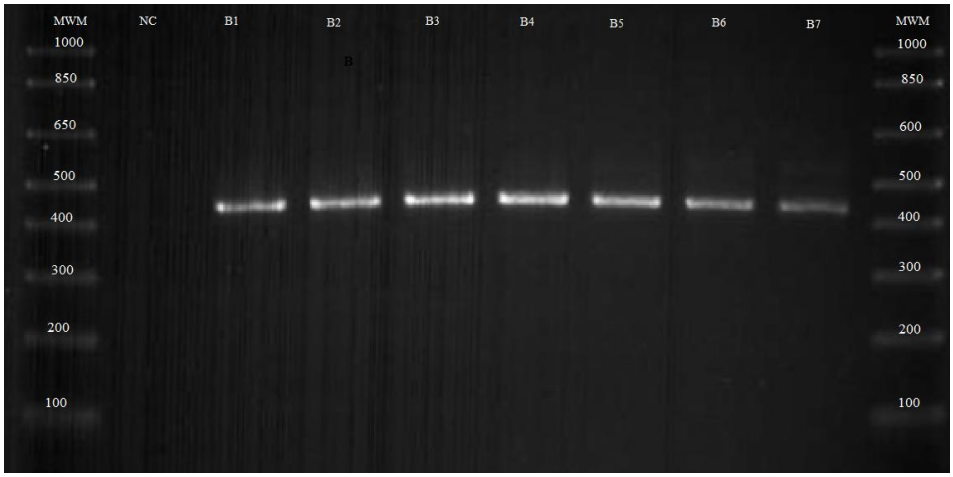


Figure 20. Experimental determination of the optimal annealing temperature for the PCR with the KOveri_F and KOveri_R primers. All reactions were done in a single run. Seven different temperatures were used for this purpose. B1: 58 ˚C, B2: 59.4 ˚C, B3: 60.7 ˚C, B4: 62.3 ˚C, B5: 63.6 ˚C, B6: 64.5 ˚C, B7: 65 ˚C, NC: negative control (MGW), MWM-1 Kb Plus DNA Ladder molecular weight marker.

3.4 Verification of the pCRISPR_T30RNA construct using colony PCR and DNA sequencing

Colony PCR using the InsertID_F and InsertID_R primers was done, as described in section 2.8.2, to verify the construction and transformation of the pCRISPR_T30RNA construct in the *E. coli* DH5α cells following the heat shock method. Figure 21 shows the results of colony PCR on DNA from five well separated single colonies (C1 to C5) picked from overnight growth on LB agar with kanamycin.

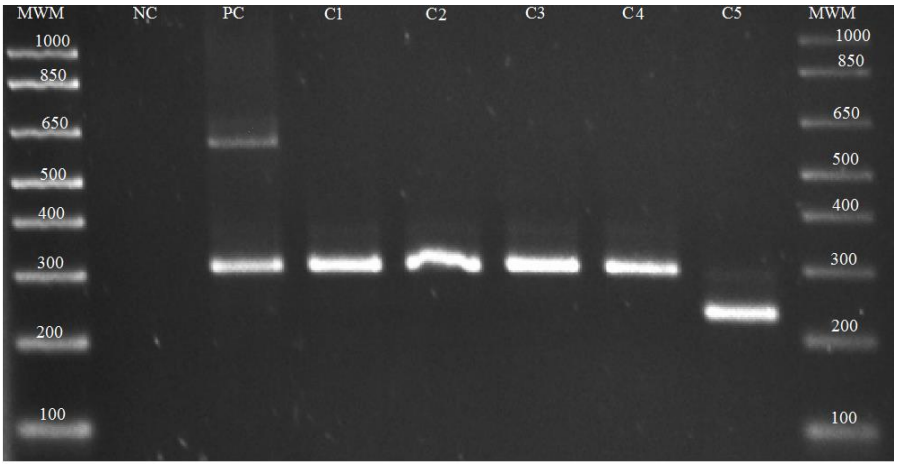


Figure 21. Gel electrophoresis of amplification products from colony PCR with the InsertID_F and InsertID_R primers to verify insertion of the T30RNA in the pCRISPR plasmid. PCR of the intact pCRISPR plasmid generated amplification products of 576 bp and 302 bp, while PCR of the pCRISPR_T30RNA plasmid generated only one product of 302 bp. C1 to C5 are the colonies obtained after heat shock method. NC- negative control (MGW), PC- positive control (intact pCRISPR plasmid), MWM- 1 Kb Plus DNA Ladder molecular weight marker.

Intact pCRISPR plasmid was used as a positive control, and a PCR mixture with MGW instead of DNA template was used as negative control. DNA templates from the colonies labelled C1 to C4 each generated a single amplification product of about 300 bp corresponding to the expected size of 302 bp. This verified that the T30RNA had been successfully incorporated into the pCRISPR plasmid, and that the plasmid had been transformed into the host cell via heat shock. In contrast, the colony labelled C5 generated an unexpected band between 200 and 300 bp.

Two of the colonies (colony C1 and C5) as well as the intact pCRISPR plasmid were also subjected to Sanger sequencing using the InsertID_F and InsertID_R primers. Only the reverse primer was used to sequence the intact pCRISPR, to avoid getting two different amplification products in the sequencing reaction. The sequencing results were imported into Geneious R11 and aligned with the sequence of the intact pCRISPR plasmid.

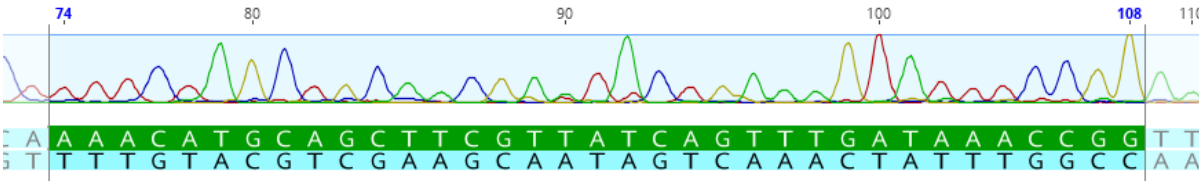


Figure 22. Screenshot from Geneious R11 showing the result of Sanger sequencing of the pCRISPR_T30RNA plasmid from colony C1. The nucleotide sequence matching the T30RNA sequence is highlighted in the chromatogram.

The sequence analysis of the amplification product from colony C1 showed that the identified bases were complementary to the sequence of T30RNA, confirming that the T30RNA was incorporated into the pCRISPR plasmid (Figure 22). Colony C5 did not contain the T30RNA sequence (results not shown). The short PCR product obtained was most likely due to self-ligation following digestion.

After that the pCRISPR_T30RNA construct was verified by PCR and Sanger sequencing, cells from colony C1 were used to make a glycerol stock which was stored at -80 $^{\circ}$C awaiting further use. The pCRISPR_T30RNA plasmid was isolated with Quantum Prep® Plasmid Miniprep kit from (Bio-Rad), and 100 ng of pCRISPR_T30RNA plasmid was used for the following mutagenesis step.

3.5 Mutagenesis step

The pCas9 plasmid was first introduced into FHI11 by electroporation (results not shown). Then, the FHI11 cells harbouring pCas9 were prepared for co-transformation of the pCRISPR_T30RNA and a dsDNA (201 bp deletion) or a ssDNA (33 bp deletion) repair template. The final mutagenesis step was performed as described in section 2.5.6. Transformants capable of survival in the presence of the CRISPR/Cas9 system were selected on LB agar using kanamycin and chloramphenicol, and analysed by colony PCR with the primers KOveri_F and KOveri_R.

To assess the efficiency of electroporation, the pCRISPR plasmid without gDNA was co-transformed into the FHI11 cells harbouring pCas9. Following overnight incubation, approximately 1.04 x 10^6 CFUs μg^{-1} of pCRISPR plasmid DNA were observed on LB agar containing 25 µg/ml chloramphenicol and 50 µg/ml kanamycin indicating an efficient electroporation of the pCRISPR plasmid (Figure 23).

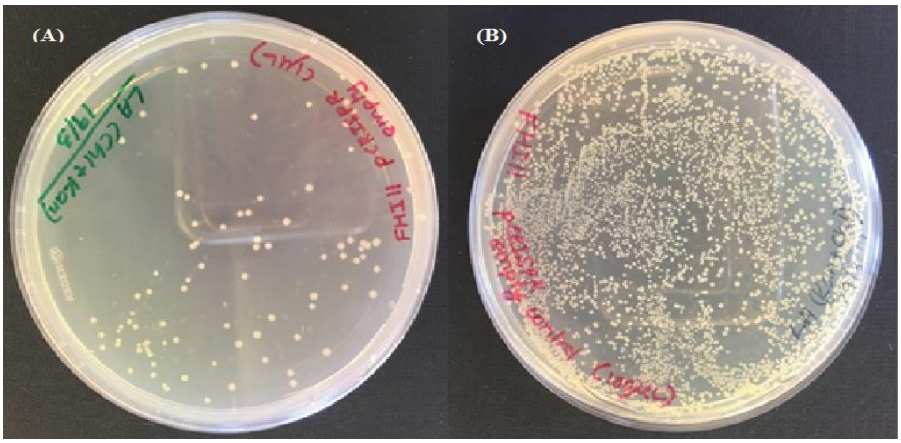


Figure 23. Colonies obtained on LB agar containing 25 μg/ml chloramphenicol and 50 μg/ml kanamycin after transforming an empty pCRISPR plasmid into FHI11 cells harbouring the pCas9 plasmid. A) Colonies obtained from 1μl of recovered culture. B) Colonies obtained from100 μl of recovered culture.

3.5.1 Validating the 201 bp deletion by colony PCR

After mutagenesis, 2 colonies (D1 and D2) were obtained on LB agar containing 25 μg/ml chloramphenicol and 50 μg/ml kanamycin. Both colonies were subjected to colony PCR with the primers KOveri_F and KOveri_R to verify the expected 201 bp deletion that was supposed to be introduced. The wild type FHI11 strain, containing an intact *eae* gene, was used as positive control. A negative control with MGW instead of DNA template was also included. The 201 bp deletion mutant strain with a premature stop codon was expected to give an amplification product of 238 bp.

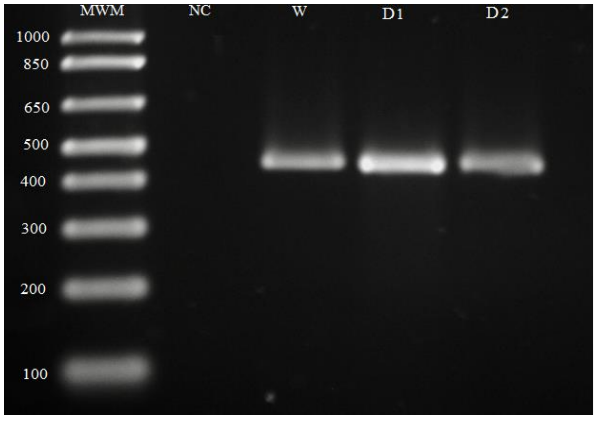


Figure 24. PCR amplification of the locus in the *eae* gene in bacterial cells from two colonies obtained after a mutagenesis experiment, to verify a 201 bp deletion. D1 and D2 are the colonies obtained after the mutagenesis step with 120 bp long dsDNA repair template. NC- negative control (MGW), W- wild type FHI11 strain, MWM- 1 Kb Plus DNA Ladder molecular weight marker.

Contrary to expectation, the mutagenesis step did not generate mutant strains containing a 201 bp deletion. Gel electrophoresis following colony PCR revealed that both colonies generated a PCR product (436 bp) similar to the wild type strain with an intact *eae* gene (Figure 24).

As the CRISPR/Cas9 system induces DSB in wild type cells, theoretically no cells which has received the pCRISPR_T30RNA plasmid should survive if the DSB was not repaired immediately [83]. However, two unedited colonies were recovered after the mutagenesis step. Since none of these colonies did harbour the 201 bp deletion, colony PCR was performed using the InsertID_F and InsertID_R primers to verify that the two colonies contained the pCRISPR_T30RNA plasmid.

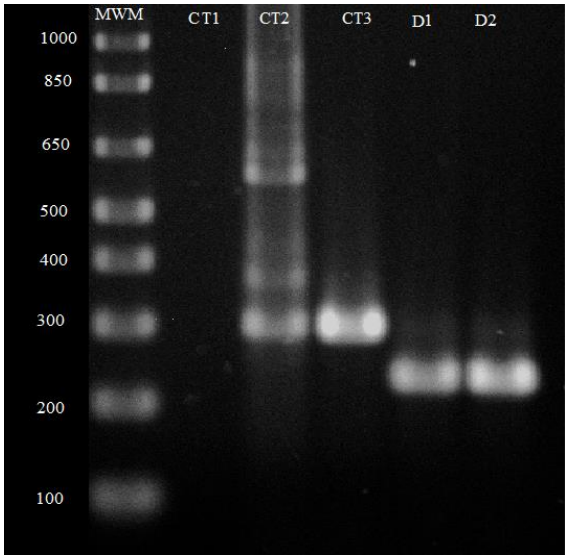


Figure 25. PCR using the InsertID_F and InsertID_R primers of two colonies of FHI11 obtained after mutagenesis. D1 and D2 are the colonies obtained after the mutagenesis step with 120 bp long dsDNA repair template. Three controls were included: CT1- negative control (MGW), CT2-intact pCRISPR plasmid, CT3-pCRISPR_T30RNA plasmid. MWM-1 Kb Plus DNA Ladder molecular weight marker.

The PCR products generated with both colonies confirmed the presence of pCRISPR_T30RNA plasmid (Figure 25). However, the size of the PCR product was shorter than the PCR product generated with the pCRISPR_T30RNA plasmid as template. A possible explanation for this result might be the loss of T30RNA from the pCRISPR plasmid required for target recognition. To confirm this, the PCR product from one of the colonies (colony D1) was analysed by Sanger sequencing, which revealed that the pCRISPR_T30RNA plasmid did not contain the T30RNA (data not shown). Such loss of gDNA from the CRISPR plasmid has been reported before. Jiang et al. termed cells where the gDNA was lost from the plasmid as

“escapers” [83]. Failure to get the desired deletion could also be due to ineffective gRNA. However, no colonies harbouring the pCRISPR_T30RNA plasmid and which were negative for the desired deletion, were cultured after mutagenesis.

The results obtained after the mutagenesis step show that two-plasmid system did support efficient CRISPR/Cas9 mediated DSB. However, mutant strain was not obtained, most likely due to an inefficient HDR mechanism. The two colonies that survived the Cas9 mediated DSB probably did so due to loss of the T30RNA from the pCRISPR_T30RNA plasmid.

3.5.2 Validating the 33 bp deletion by colony PCR

After performing the mutagenesis step with the 70 bp long ssDNA, 30 colonies were recovered on LB agar containing 25 μg/ml chloramphenicol and 50 μg/ml kanamycin. All 30 colonies were analysed by colony PCR using the KOveri_F and KOveri_R primers to verify the 33 bp deletion. Since the deletion is comparatively small, the PCR products were analysed on a 2% agarose gel at 150V for 2 hrs to be able to differentiate the deletion from the wild type strain. The 33 bp deletion mutant strain with a premature stop codon was expected to give an amplification product of 406 bp.

Gel electrophoresis following colony PCR revealed that none of the colonies had obtained a 33 bp deletion in the *eae* gene (Fig. 26). For colony 5 and 24 no amplification products were observed. Colony PCR on these two colonies with primers targeting a different locus (*lpf*$_{B1}$*C*) on the bacterial chromosome gave an amplification product as expected for this target, ruling out that the negative results were due to inadequate DNA template (results not shown).

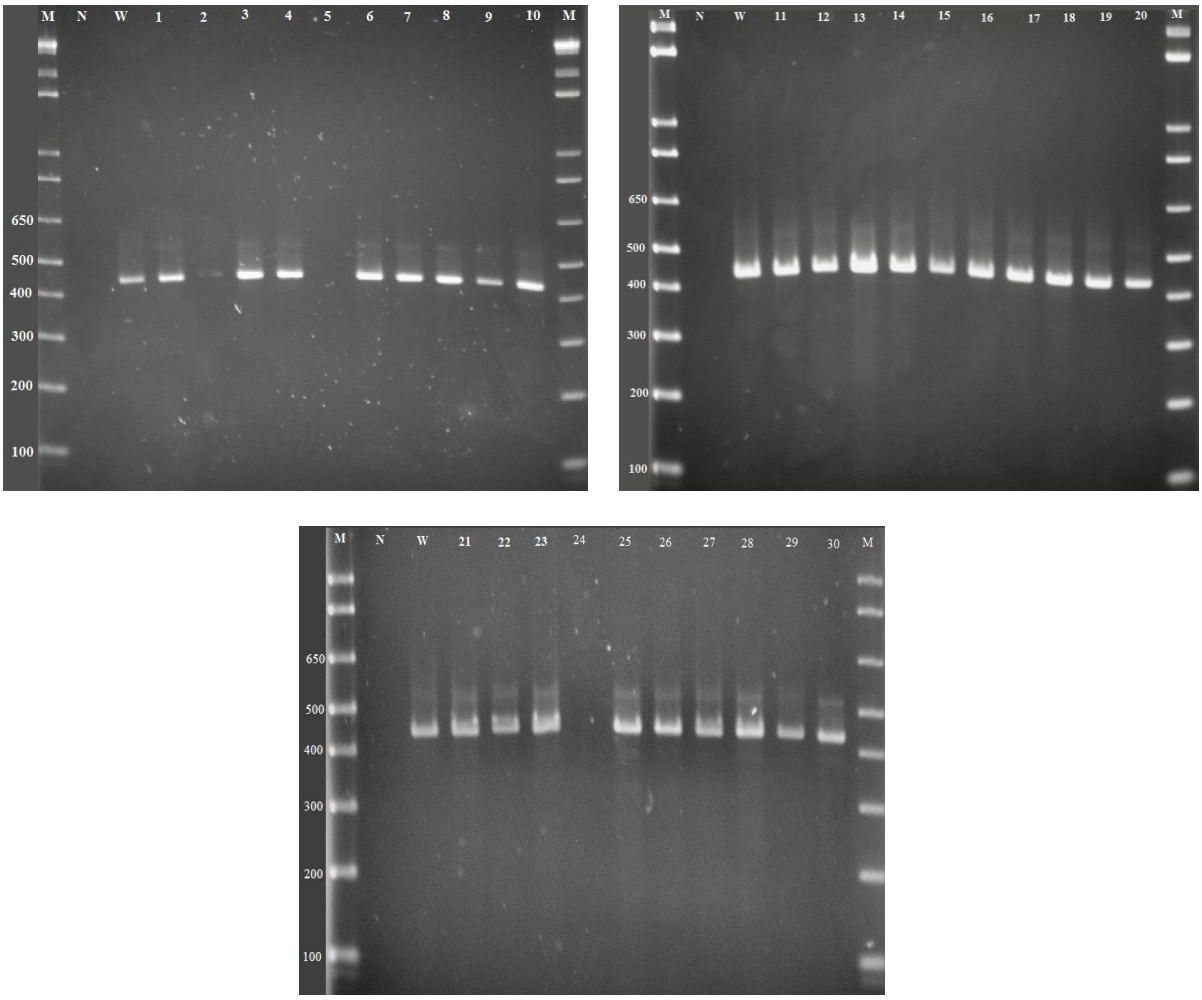


Figure 26. Agarose gel electrophoresis (2% agarose) of PCR products obtained from the strain FHI11 for the *eae* locus. Colony PCR with primers KOveri_F and KOveri_R were used for the verification of a 33 bp deletion in the *eae* gene. 1 to 30 are the colonies obtained after the mutagenesis step with 70 bp long ssDNA repair template. N- negative control (MGW), W- wild type strain, M-1 Kb Plus DNA Ladder molecular weight marker.

The colonies which escaped Cas9 mediated killing and which did not generate any PCR product might have undergone a large DNA deletion (≥ 436 bp) including the primer binding site/sites, as previously reported in *E. coli* [125]. Similar findings have been observed in previous studies, and some authors have hypothesized that the large deletions might be a result of an alternative end-joining (A-EJ) repair mechanism [85, 98, 126]. PCR using primers amplifying a larger sequence within the *eae* gene or whole DNA sequencing comparing the genome sequence before and after mutagenesis could be done to clarify the genomic changes after mutagenesis. Due to limited time this was not carried out.

4 Discussion

Studies have found that production of Shiga toxin alone is insufficient for the STEC strain to cause severe disease such as HC and HUS. Finding specific virulence factors, such as adhesins is important in developing tools to identify and predict the ability of STEC strains to cause severe disease in humans [127]. Adhesion is an important initial step of the pathogenesis and intimin has been considered the primary adhesin of STEC [27]. High expression of the lpf_{B1} operon in STEC strains associated with HUS compared to non-HUS strains indicates a role of this operon in the development of severe disease. The lpf_{B1} operon is 42-63% identical to the *lpf2* operon of *E. coli* O157:H7 Sakai, which has been reported to have a role in adhesion [101]. Most STEC strains adhere strongly to the intestinal epithelial cells via intimin. Therefore, in such strains it has been difficult to study the possible additional role of lpf_{B1} in adhesion. The aim of this study was to knockout the *eae* gene from a STEC strain (FHI11) associated with HUS.

The strain FHI11 was originally isolated in a HUS outbreak, which later had lost its *stx* gene, was chosen for knockout of the *eae* gene. The lack of the *stx* genes made it possible to work with the live strain in a biosafety level 2 (BSL-2) environment. The loss of the *stx* genes from the original strain might have happened during the infection in the patient, during transport or during handling in the laboratory. The strain used in this study was therefore not fully identical to the strain that caused HUS, at least with respect to content of *stx* genes and probably the stx prophage. It is however unlikely that this difference should influence of the adherence properties of the strain, since *stx* has not been found to have any role in EHEC colonization [19, 128].

4.1 Gene editing using CRISPR/Cas9 system

A novel method developed by Jiang et al. was used to generate an *eae* knockout in the strain FHI11. The knockout modelling procedure utilizes the CRISPR/Cas9 system with a synthetic DNA repair template, homologous to the upstream and downstream regions of the targeted deletion in the *eae* gene.

First, wild type FHI11 competent cells were transformed with the pCas9 plasmid. Then a gDNA, termed T30RNA, was designed and constructed by annealing the forward and reverse oligonucleotides T30RNA_F and T30RNA_R, before the gDNA was ligated into a BsaI-

digested pCRISPR plasmid. The FHI11 strain harbouring pCas9 was then co-transformed with the pCRISPR_T30RNA plasmid and a repair DNA template.

The overall success of generating a deletion and a premature stop codon in the *eae* gene depends on the efficiency of three main factors: the efficiency of the gRNA to guide Cas9 to generate a DSB, the rescue efficiency of HDR and mutant selection. These factors will be discussed in the following sections.

4.1.1 Design of an efficient gDNA

A gDNA must be designed before repair DNA template and PCR primers for deletion verification, because design of the latter two depends on where the gDNA guides the Cas9 enzyme to introduce a DSB.

In general, an oligonucleotide with a length of 20 bp is used as gDNA [85]. Use of short gDNA is supported by results from studies on human cells, where it was demonstrated that truncated gDNA of length < 20 bp significantly reduced off-target effect without sacrificing on target activity [129]. In contrast, König et al. recommended the use of a gDNA of 30 bp in their protocol, without giving a reason for their choice of length [110]. In this study 30 bp long gDNA was designed according to guidelines established by König et al., and we found that a gDNA of 30 bp had sufficient efficiency.

The efficiency of a gDNA can be estimated by its ability to significantly reduce bacterial viability in the absence of an exogeneous repair DNA template [103]. Zerbini et al. have suggested that an efficient gDNA should lead to a drop in transformation efficiency between an "empty" pCRISPR plasmid and a pCRISPR plasmid with gDNA of at least three orders of magnitude. They recommend that the efficiency of gRNA should be tested using colony counting, and if the colony count is higher than 10^3 CFU/μg of plasmid DNA, a new gDNA should be designed [103]. In this study, the efficiency of the gDNA oligonucleotide was ensured by choosing a sequence with predicted high on-target and low off-target effect using Geneious R11. When the 120 bp dsDNA repair template was used to generate a 201 bp deletion, no wild type strains harbouring a pCRISPR_T30RNA plasmid were recovered. This shows that an active CRISPR system with the gDNA was lethal to the bacterial cell [98].

4.1.2 Consequences of CRISPR/Cas9 mediated DSB and HDR

In eukaryotic cells, a CRISPR/Cas9 mediated DSB can be repaired via NHEJ pathway. However, most bacteria, except *Mycobacteria* and *Bacillus sublitis*, do not possess an endogenous NHEJ repair system. *E. coli* is believed to rely merely on the HDR mechanism to repair DSBs [130].

Obtaining a mutant strain using the HDR mechanism helps select for cells with the desired deletion since cells with a CRISPR/Cas9 mediated DBS not repaired by HDR will be killed.

In this study we expected that the native HDR system of FHI11 would be able to repair the DSB in the chromosome, created by the CRISPR/Cas9 system. However, no colonies were obtained carrying the desired mutation. This could be due to low frequency of the FHI11 native HDR system. Studies have found that coupling λ-Red recombineering with the CRISPR/Cas9 system increases the editing efficiency [83, 85, 98, 103]. In *E. coli*, this strategy has been demonstrated to generate mutant cells with an efficiency as high as 65% ± 14 % [83]. Analysis of the genome sequence of FHI11 showed hits with 98% identity with the three λ-Red genes *beta*, *gam* and *exo*. Although this finding indicates the potential for an active system in FHI11, the presence of the genes does not prove that the system is biologically active. Analysis by reverse-transcription quantitative PCR (RT-qPCR) of whether the λ-Red genes are expressed or not could have clarified whether the λ-Red system is functional in strain FHI11. We did not have time to do this in this study.

In mammalian cells, different approaches have been developed to enhance HDR, including inhibiting the proteins involved into the NHEJ pathway and inducing the cell cycle into S-phase when HDR is most active [131]. However, it has been shown that using an optimized repair DNA template can improve the HDR rate without regulating the cell cycle [132].

Designing a homologous repair template is as important as designing a gDNA. The size of the deletion is an important factor in deciding the length and the type of repair DNA template. Previous studies have reported that the efficiency of deletion decreased with shorter homologous repair templates [85, 103, 126]. One study has found that it is possible to delete 500 nucleotides with 50 ± 14% of efficiency using a 120 bp dsDNA repair template [103]. Small deletions up to 50 bp or single point mutations can be successfully obtained using ssDNA template, while deletion or insertion more than 50 bp requires dsDNA template [126]. Based on these findings, two different repair DNA templates were designed to obtain 201 bp or 33 bp deletion within the β-domain of *eae* gene with reasonable efficiency. To obtain 201

bp deletion, 120 bp long leading (De201RT_F) and lagging strands (De201RT_R) were designed and annealed to obtain a dsDNA repair template, while for 33 bp deletion, 70 bp long ssDNA repair template (Del33RT_F) was designed.

Zhao et al. obtained a 20.8% ± 8.8% editing efficiency in *E. coli* in the absence of λ-Red genes [126]. This confirms that, in the absence of λ-Red genes HDR mechanism can recover the cells which have underdone DSBs caused by CRISPR/Cas9 system. Although this efficiency was low compared to the efficiency reported when the λ-Red system was used, it would ensure that about 2 successfully edited colonies would be obtained from 10 colonies tested. Since reasonable editing efficiency was obtained with the native HDR system, this study was designed to obtain mutant strain via DSB mediated HDR by the native HDR system of FHI11.

When the 70 bp ssDNA was used to generate a 33 bp deletion in FHI11, in total 30 colonies were recovered. 28 of the 30 colonies generated an amplification product which was showed that no deletion had been taken place in the *eae* gene. Two of these colonies did not generate any PCR products with Koveri_F and KOveri_R primers, indicating that the primer binding sites might have been removed by a large deletion (≥ 436 bp) e.g. by an alternative end-joining repair mechanism (A-EJ) [125]. This is a DSB repair mechanism that can repair DSBs that were not repaired by the HDR mechanism. The A-EJ mechanism, reported to be present in some pathogenic *E. coli*, can generate large deletions up to 40 Kb by extensive DNA restriction and usage of microhomology [98, 125]. Similar finding observed in previous studies further confirms the possibility that A-EJ mechanism can repair DSBs [85, 98, 126]. The possibility to repair CRISPR/Cas9 mediated DSBs through A-EJ in *E. coli* paves the way towards an easy introduction of knockout mutation in *E. coli*, independently of homologous recombination.

Another possible explanation of why the DSB was not repaired by the HDR system might be that the cleaved sites, where proteins involved in DNA repair bind to initiate DNA repair, were blocked. Strenberg et al. reported recently that the Cas9: RNA complex can remain tightly bound to the cleaved products and requires other cellular factors to dissociate from cleaved DNA and promote recycling [133].

It is also worth noting that homologous recombination varies between genomic loci due to secondary structures and sequence dependent effects [134]. The *eae* locus employed in this study might not be a “hot spot”, where HDR take place in high frequency. Another factor that

might play a role in low frequency of HDR is degradation of the linear DNA repair template by restriction enzymes before integration into the chromosome. Low transformation efficiency of the repair template in FHI11 might have led to reduced availability of repair DNA template in the cells needed for HDR. To ensure the availability, the repair DNA template could be constructed on a plasmid and transformed into the bacterial strain. The repair templates used in this study contained strong secondary structures. This could be another factor that affected target binding and reduced the efficiency of HDR. However, a previous study successfully used repair DNA templates with strong secondary structures, which were not considered to limit the efficiency of the HDR mechanism in that study [103].

4.1.3 Selection

Selection of cells which had undergone deletion was done by ensuring that such cell survived, while non-deleted cells died in the mutagenesis experiment. The repair template was designed so that the 33 bp protospacer and PAM sequence were deleted from the edited genome so that they were protected from further DSBs, while the wild type cells with these sequences intact were killed due to the insertion of a DSB by the CRISPR/Cas9 system, and which was not repaired. According to Cui & Richard, if a gRNA efficiently drives Cas9 mediated DSB, no wild type colonies should be recovered [98].

In this study, some wild type colonies were recovered without the desired mutation. When a 120 bp dsDNA repair template was used to obtain a 201 bp deletion in FHI11, two colonies were obtained without the desired deletion. These colonies, which escaped Cas9 mediated cell death, represent “escapers” in which the T30RNA from pCRISPR_T30RNA was lost. A similar finding was observed in a previous study where the gDNA required for target recognition was lost in a cell targeted by Cas9. The same study found that cells harbouring an inactivation mutation in Cas9 also survived the CRISPR/Cas9 mediated DSB. This is a major limitation of CRISPR selection, with the consequence that one has to analyse several colonies to find a subset of genetically modified colonies [83].

Colonies that have undergone a large deletion within *eae* gene would also have escaped the selection mechanism since they would lack the gDNA recognition site. Such cells would have the advantage to escape CRISPR/Cas9 mediated DSB.

4.1.4 Measuring DNA concentration and purity

Purity of samples is essential in cloning since contaminants can inhibit the accuracy of the cloning efficiency. As mentioned in section 2.7.1 and 2.7.2, the Qubit® 3.0 Fluorometer was used to measure the concentration of extracted DNA while Nanodrop was used to measure the purity of samples. Although Nanodrop is commonly used, it may be inaccurate at low concentrations. Since contaminants contribute together with dsDNA, ssDNA and RNA to absorbance at 260 nm, they will contribute to the total measurement at 260 nm and lead to an overestimation of the concentration of dsDNA. Qubit Fluorometric quantification provides more selective, sensitive, and accurate dsDNA concentrations than Nanodrop, but it cannot indicate contamination present in the sample. In contrast, Nanodrop can analyze the contaminants by measuring absorbance at different wavelengths but it cannot measure the concentration of a specific component [135].

5 Conclusion and future perspectives

This master's thesis was designed to knockout the *eae* gene of the STEC strain FHI11 via CRISPR/Cas9 system. A gDNA targeting the β-domain of the *eae* gene was designed to guide the Cas9 to the specific location and create a DSB. Introduction of the specific mutation at the site of DSB was based on the activity of the native HDR mechanism, which was expected to be present in the STEC strain. A synthetic DNA repair template homologous to the upstream and downstream regions of the target deletion in the *eae* gene was provided for the HDR mechanism.

The Construction of the pCRISPR_T30RNA plasmid and introduction of the CRISPR/Cas9 system into the selected STEC strain FHI11 was successful. However, we were not able to generate an *eae* knockout in the STEC strain using the CRISPR/Cas9 editing tool. Further optimization of the method is therefore necessary. Due to time limitation, such optimization was not possible in this study.

There were some limitations which may have affected the outcome of the study. Low frequency of HDR in the selected STEC strain may have reduced incorporation of the repair template after that the DSB had been inserted. The fact that we were not able to obtain a plasmid with the λ-Red system (Exo, Beta and Gam) may have influenced on the efficiency of the HDR system in the STEC strain. It would be interesting to analyse the genomic changes that had occurred after the mutagenesis step with the short repair templates in colonies which did not produce any amplification products with colony PCR. If these STEC strains had undergone a large deletion including the *eae* gene, the mutant strain might be used to study the impact of the *eae* gene in expression of the lpf_{B1} operon. The protocol developed could then be modified to knockout also the lpf_{B1} operon.

6 References

1. Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, Finlay BB: **Recent advances in understanding enteric pathogenic Escherichia coli**. *Clinical microbiology reviews* 2013, **26**(4):822-880.
2. Odonkor ST, Ampofo JK: **Escherichia coli as an indicator of bacteriological quality of water: an overview**. *Microbiology Research* 2013, **4**(1).
3. Kaper JB, Nataro JP, Mobley HL: **Pathogenic Escherichia coli**. *Nature reviews Microbiology* 2004, **2**(2):123-140.
4. Nataro JP, Kaper JB: **Diarrheagenic Escherichia coli**. *Clinical microbiology reviews* 1998, **11**(1):142-201.
5. Huang DB, Nataro JP, DuPont HL, Kamat PP, Mhatre AD, Okhuysen PC, Chiang T: **Enteroaggregative Escherichia coli is a cause of acute diarrheal illness: a meta-analysis**. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2006, **43**(5):556-563.
6. Nataro JP: **Enteroaggregative Escherichia coli pathogenesis**. *Current opinion in gastroenterology* 2005, **21**(1):4-8.
7. Croxen MA, Finlay BB: **Molecular mechanisms of Escherichia coli pathogenicity**. *Nature reviews Microbiology* 2010, **8**(1):26-38.
8. Konowalchuk J, Speirs JI, Stavric S: **Vero response to a cytotoxin of Escherichia coli**. *Infection and immunity* 1977, **18**(3):775-779.
9. Delannoy S, Beutin L, Fach P: **Discrimination of enterohemorrhagic Escherichia coli (EHEC) from non-EHEC strains based on detection of various combinations of type III effector genes**. *Journal of clinical microbiology* 2013, **51**(10):3257-3262.
10. Gyles CL: **Shiga toxin-producing Escherichia coli: an overview**. *Journal of animal science* 2007, **85**(13 Suppl):E45-62.
11. Canpolat N: **Hemolytic uremic syndrome**. *Turk pediatri arsivi* 2015, **50**(2):73-82.
12. Mayer CL, Leibowitz CS, Kurosawa S, Stearns-Kurosawa DJ: **Shiga toxins and the pathophysiology of hemolytic uremic syndrome in humans and animals**. *Toxins* 2012, **4**(11):1261-1287.
13. Hussein HS, Bollinger LM: **Prevalence of Shiga toxin-producing Escherichia coli in beef cattle**. *Journal of food protection* 2005, **68**(10):2224-2241.
14. Dean-Nystrom EA, Bosworth BT, Cray WC, Jr., Moon HW: **Pathogenicity of Escherichia coli O157:H7 in the intestines of neonatal calves**. *Infection and immunity* 1997, **65**(5):1842-1848.
15. Tseng M, Fratamico PM, Manning SD, Funk JA: **Shiga toxin-producing Escherichia coli in swine: the public health perspective**. *Animal health research reviews* 2014, **15**(1):63-75.
16. Frank C, Werber D, Cramer JP, Askar M, Faber M, an der Heiden M, Bernard H, Fruth A, Prager R, Spode A *et al*: **Epidemic profile of Shiga-toxin-producing Escherichia coli O104:H4 outbreak in Germany**. *The New England journal of medicine* 2011, **365**(19):1771-1780.
17. Bielaszewska M, Middendorf B, Kock R, Friedrich AW, Fruth A, Karch H, Schmidt MA, Mellmann A: **Shiga toxin-negative attaching and effacing Escherichia coli: distinct clinical associations with bacterial phylogeny and virulence traits and inferred in-host pathogen evolution**. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2008, **47**(2):208-217.
18. Large TM, Walk ST, Whittam TS: **Variation in acid resistance among shiga toxin-producing clones of pathogenic Escherichia coli**. *Applied and environmental microbiology* 2005, **71**(5):2493-2500.
19. Lewis SB, Cook V, Tighe R, Schuller S: **Enterohemorrhagic Escherichia coli colonization of human colonic epithelium in vitro and ex vivo**. *Infection and immunity* 2015, **83**(3):942-949.

20. Tzipori S, Wachsmuth IK, Chapman C, Birden R, Brittingham J, Jackson C, Hogg J: **The pathogenesis of hemorrhagic colitis caused by Escherichia coli O157:H7 in gnotobiotic piglets**. *The Journal of infectious diseases* 1986, **154**(4):712-716.
21. Sherman P, Soni R, Karmali M: **Attaching and effacing adherence of Vero cytotoxin-producing Escherichia coli to rabbit intestinal epithelium in vivo**. *Infection and immunity* 1988, **56**(4):756-761.
22. Melton-Celsa AR: **Shiga Toxin (Stx) Classification, Structure, and Function**. *Microbiology spectrum* 2014, **2**(4):Ehec-0024-2013.
23. Nathanson S, Kwon T, Elmaleh M, Charbit M, Launay EA, Harambat J, Brun M, Ranchin B, Bandin F, Cloarec S *et al*: **Acute neurological involvement in diarrhea-associated hemolytic uremic syndrome**. *Clinical journal of the American Society of Nephrology : CJASN* 2010, **5**(7):1218-1228.
24. Pacheco AR, Sperandio V: **Shiga toxin in enterohemorrhagic E.coli: regulation and novel anti-virulence strategies**. *Frontiers in cellular and infection microbiology* 2012, **2**:81.
25. Farfan MJ, Torres AG: **Molecular mechanisms that mediate colonization of Shiga toxin-producing Escherichia coli strains**. *Infection and immunity* 2012, **80**(3):903-913.
26. Chen HD, Frankel G: **Enteropathogenic Escherichia coli: unravelling pathogenesis**. *FEMS microbiology reviews* 2005, **29**(1):83-98.
27. McWilliams BD, Torres AG: **Enterohemorrhagic Escherichia coli Adhesins**. *Microbiology spectrum* 2014, **2**(3).
28. Jerse AE, Yu J, Tall BD, Kaper JB: **A genetic locus of enteropathogenic Escherichia coli necessary for the production of attaching and effacing lesions on tissue culture cells**. *Proceedings of the National Academy of Sciences of the United States of America* 1990, **87**(20):7839-7843.
29. Badea L, Doughty S, Nicholls L, Sloan J, Robins-Browne RM, Hartland EL: **Contribution of Efa1/LifA to the adherence of enteropathogenic Escherichia coli to epithelial cells**. *Microbial pathogenesis* 2003, **34**(5):205-215.
30. Ledesma MA, Ochoa SA, Cruz A, Rocha-Ramirez LM, Mas-Oliva J, Eslava CA, Giron JA, Xicohtencatl-Cortes J: **The hemorrhagic coli pilus (HCP) of Escherichia coli O157:H7 is an inducer of proinflammatory cytokine secretion in intestinal epithelial cells**. *PloS one* 2010, **5**(8):e12127.
31. Tarr PI, Bilge SS, Vary JC, Jr., Jelacic S, Habeeb RL, Ward TR, Baylor MR, Besser TE: **Iha: a novel Escherichia coli O157:H7 adherence-conferring molecule encoded on a recently acquired chromosomal island of conserved structure**. *Infection and immunity* 2000, **68**(3):1400-1407.
32. Torres AG, Giron JA, Perna NT, Burland V, Blattner FR, Avelino-Flores F, Kaper JB: **Identification and characterization of lpfABCC'DE, a fimbrial operon of enterohemorrhagic Escherichia coli O157:H7**. *Infection and immunity* 2002, **70**(10):5416-5427.
33. Samadder P, Xicohtencatl-Cortes J, Saldana Z, Jordan D, Tarr PI, Kaper JB, Giron JA: **The Escherichia coli ycbQRST operon encodes fimbriae with laminin-binding and epithelial cell adherence properties in Shiga-toxigenic E. coli O157:H7**. *Environmental microbiology* 2009, **11**(7):1815-1826.
34. Batisson I, Guimond MP, Girard F, An H, Zhu C, Oswald E, Fairbrother JM, Jacques M, Harel J: **Characterization of the novel factor paa involved in the early steps of the adhesion mechanism of attaching and effacing Escherichia coli**. *Infection and immunity* 2003, **71**(8):4516-4525.
35. Paton AW, Srimanote P, Woodrow MC, Paton JC: **Characterization of Saa, a novel autoagglutinating adhesin produced by locus of enterocyte effacement-negative Shiga-toxigenic Escherichia coli strains that are virulent for humans**. *Infection and immunity* 2001, **69**(11):6999-7009.

36. Brunder W, Khan AS, Hacker J, Karch H: **Novel type of fimbriae encoded by the large plasmid of sorbitol-fermenting enterohemorrhagic Escherichia coli O157:H(-)**. *Infection and immunity* 2001, **69**(7):4447-4457.
37. Tatsuno I, Horie M, Abe H, Miki T, Makino K, Shinagawa H, Taguchi H, Kamiya S, Hayashi T, Sasakawa C: **toxB gene on pO157 of enterohemorrhagic Escherichia coli O157:H7 is required for full epithelial cell adherence phenotype**. *Infection and immunity* 2001, **69**(11):6660-6669.
38. Bai X, Fu S, Zhang J, Fan R, Xu Y, Sun H, He X, Xu J, Xiong Y: **Identification and pathogenomic analysis of an Escherichia coli strain producing a novel Shiga toxin 2 subtype**. *Sci Rep* 2018, **8**(1):6756.
39. Byrne L, Vanstone GL, Perry NT, Launders N, Adak GK, Godbole G, Grant KA, Smith R, Jenkins C: **Epidemiology and microbiology of Shiga toxin-producing Escherichia coli other than serogroup O157 in England, 2009-2013**. *Journal of medical microbiology* 2014, **63**(Pt 9):1181-1188.
40. Sandvig K, Bergan J, Dyve AB, Skotland T, Torgersen ML: **Endocytosis and retrograde transport of Shiga toxin**. *Toxicon : official journal of the International Society on Toxinology* 2010, **56**(7):1181-1185.
41. Allison HE: **Stx-phages: drivers and mediators of the evolution of STEC and STEC-like pathogens**. *Future microbiology* 2007, **2**(2):165-174.
42. Schuller S: **Shiga toxin interaction with human intestinal epithelium**. *Toxins* 2011, **3**(6):626-639.
43. Moon HW, Whipp SC, Argenzio RA, Levine MM, Giannella RA: **Attaching and effacing activities of rabbit and human enteropathogenic Escherichia coli in pig and rabbit intestines**. *Infection and immunity* 1983, **41**(3):1340-1351.
44. Frankel G, Phillips AD, Trabulsi LR, Knutton S, Dougan G, Matthews S: **Intimin and the host cell--is it bound to end in Tir(s)?** *Trends in microbiology* 2001, **9**(5):214-218.
45. Schmidt MA: **LEEways: tales of EPEC, ATEC and EHEC**. *Cellular microbiology* 2010, **12**(11):1544-1552.
46. Deibel C, Kramer S, Chakraborty T, Ebel F: **EspE, a novel secreted protein of attaching and effacing bacteria, is directly translocated into infected host cells, where it appears as a tyrosine-phosphorylated 90 kDa protein**. *Molecular microbiology* 1998, **28**(3):463-474.
47. Deng W, Puente JL, Gruenheid S, Li Y, Vallance BA, Vazquez A, Barba J, Ibarra JA, O'Donnell P, Metalnikov P *et al*: **Dissecting virulence: systematic and functional analyses of a pathogenicity island**. *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(10):3597-3602.
48. Fitzhenry RJ, Pickard DJ, Hartland EL, Reece S, Dougan G, Phillips AD, Frankel G: **Intimin type influences the site of human intestinal mucosal colonisation by enterohaemorrhagic Escherichia coli O157:H7**. *Gut* 2002, **50**(2):180-185.
49. Fairman JW, Dautin N, Wojtowicz D, Liu W, Noinaj N, Barnard TJ, Udho E, Przytycka TM, Cherezov V, Buchanan SK: **Crystal structures of the outer membrane domain of intimin and invasin from enterohemorrhagic E. coli and enteropathogenic Y. pseudotuberculosis**. *Structure (London, England : 1993)* 2012, **20**(7):1233-1243.
50. Touze T, Hayward RD, Eswaran J, Leong JM, Koronakis V: **Self-association of EPEC intimin mediated by the beta-barrel-containing anchor domain: a role in clustering of the Tir receptor**. *Molecular microbiology* 2004, **51**(1):73-87.
51. Oswald E, Schmidt H, Morabito S, Karch H, Marches O, Caprioli A: **Typing of intimin genes in human and animal enterohemorrhagic and enteropathogenic Escherichia coli: characterization of a new intimin variant**. *Infection and immunity* 2000, **68**(1):64-71.
52. Haugum K, Brandal LT, Lindstedt BA, Wester AL, Bergh K, Afset JE: **PCR-based detection and molecular characterization of shiga toxin-producing Escherichia coli strains in a routine**

microbiology laboratory over 16 years. *Journal of clinical microbiology* 2014, **52**(9):3156-3163.

53. Melton-Celsa A, Mohawk K, Teel L, O'Brien A: **Pathogenesis of Shiga-toxin producing escherichia coli**. *Current topics in microbiology and immunology* 2012, **357**:67-103.
54. Ethelberg S, Olsen KE, Scheutz F, Jensen C, Schiellerup P, Enberg J, Petersen AM, Olesen B, Gerner-Smidt P, Molbak K: **Virulence factors for hemolytic uremic syndrome, Denmark**. *Emerging infectious diseases* 2004, **10**(5):842-847.
55. Dean-Nystrom EA, Bosworth BT, Moon HW, O'Brien AD: **Escherichia coli O157:H7 requires intimin for enteropathogenicity in calves**. *Infection and immunity* 1998, **66**(9):4560-4563.
56. Donnenberg MS, Tacket CO, James SP, Losonsky G, Nataro JP, Wasserman SS, Kaper JB, Levine MM: **Role of the eaeA gene in experimental enteropathogenic Escherichia coli infection**. *The Journal of clinical investigation* 1993, **92**(3):1412-1417.
57. Frankel G, Phillips AD: **Attaching effacing Escherichia coli and paradigms of Tir-triggered actin polymerization: getting off the pedestal**. *Cellular microbiology* 2008, **10**(3):549-556.
58. Sinclair JF, O'Brien AD: **Cell surface-localized nucleolin is a eukaryotic receptor for the adhesin intimin-gamma of enterohemorrhagic Escherichia coli O157:H7**. *The Journal of biological chemistry* 2002, **277**(4):2876-2885.
59. Paton AW, Woodrow MC, Doyle RM, Lanser JA, Paton JC: **Molecular characterization of a Shiga toxigenic Escherichia coli O113:H21 strain lacking eae responsible for a cluster of cases of hemolytic-uremic syndrome**. *Journal of clinical microbiology* 1999, **37**(10):3357-3361.
60. Ross BN, Rojas-Lopez M, Cieza RJ, McWilliams BD, Torres AG: **The Role of Long Polar Fimbriae in Escherichia coli O104:H4 Adhesion and Colonization**. *PloS one* 2015, **10**(10):e0141845.
61. Baumler AJ, Tsolis RM, Heffron F: **The lpf fimbrial operon mediates adhesion of Salmonella typhimurium to murine Peyer's patches**. *Proceedings of the National Academy of Sciences of the United States of America* 1996, **93**(1):279-283.
62. Jordan DM, Cornick N, Torres AG, Dean-Nystrom EA, Kaper JB, Moon HW: **Long polar fimbriae contribute to colonization by Escherichia coli O157:H7 in vivo**. *Infection and immunity* 2004, **72**(10):6168-6171.
63. Newton HJ, Sloan J, Bennett-Wood V, Adams LM, Robins-Browne RM, Hartland EL: **Contribution of long polar fimbriae to the virulence of rabbit-specific enteropathogenic Escherichia coli**. *Infection and immunity* 2004, **72**(3):1230-1239.
64. Fitzhenry R, Dahan S, Torres AG, Chong Y, Heuschkel R, Murch SH, Thomson M, Kaper JB, Frankel G, Phillips AD: **Long polar fimbriae and tissue tropism in Escherichia coli O157:H7**. *Microbes and infection* 2006, **8**(7):1741-1749.
65. Torres AG, Blanco M, Valenzuela P, Slater TM, Patel SD, Dahbi G, Lopez C, Barriga XF, Blanco JE, Gomes TA *et al*: **Genes related to long polar fimbriae of pathogenic Escherichia coli strains as reliable markers to identify virulent isolates**. *Journal of clinical microbiology* 2009, **47**(8):2442-2451.
66. Xicohtencatl-Cortes J, Monteiro-Neto V, Ledesma MA, Jordan DM, Francetic O, Kaper JB, Puente JL, Giron JA: **Intestinal adherence associated with type IV pili of enterohemorrhagic Escherichia coli O157:H7**. *The Journal of clinical investigation* 2007, **117**(11):3519-3529.
67. Doughty S, Sloan J, Bennett-Wood V, Robertson M, Robins-Browne RM, Hartland EL: **Identification of a novel fimbrial gene cluster related to long polar fimbriae in locus of enterocyte effacement-negative strains of enterohemorrhagic Escherichia coli**. *Infection and immunity* 2002, **70**(12):6761-6769.
68. Merkel V, Ohder B, Bielaszewska M, Zhang W, Fruth A, Menge C, Borrmann E, Middendorf B, Muthing J, Karch H *et al*: **Distribution and phylogeny of immunoglobulin-binding protein G in Shiga toxin-producing Escherichia coli and its association with adherence phenotypes**. *Infection and immunity* 2010, **78**(8):3625-3636.

69. Blanco M, Blanco JE, Mora A, Dahbi G, Alonso MP, Gonzalez EA, Bernardez MI, Blanco J: **Serotypes, virulence genes, and intimin types of Shiga toxin (verotoxin)-producing Escherichia coli isolates from cattle in Spain and identification of a new intimin variant gene (eae-xi)**. *Journal of clinical microbiology* 2004, **42**(2):645-651.
70. Riley LW, Remis RS, Helgerson SD, McGee HB, Wells JG, Davis BR, Hebert RJ, Olcott ES, Johnson LM, Hargrett NT *et al*: **Hemorrhagic colitis associated with a rare Escherichia coli serotype**. *The New England journal of medicine* 1983, **308**(12):681-685.
71. Johnson KE, Thorpe CM, Sears CL: **The emerging clinical importance of non-O157 Shiga toxin-producing Escherichia coli**. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2006, **43**(12):1587-1595.
72. Bielaszewska M, Kock R, Friedrich AW, von Eiff C, Zimmerhackl LB, Karch H, Mellmann A: **Shiga toxin-mediated hemolytic uremic syndrome: time to change the diagnostic paradigm?** *PloS one* 2007, **2**(10):e1024.
73. Karmali MA, Mascarenhas M, Shen S, Ziebell K, Johnson S, Reid-Smith R, Isaac-Renton J, Clark C, Rahn K, Kaper JB: **Association of genomic O island 122 of Escherichia coli EDL 933 with verocytotoxin-producing Escherichia coli seropathotypes that are linked to epidemic and/or serious disease**. *Journal of clinical microbiology* 2003, **41**(11):4930-4940.
74. EFSA: **Scientific Opinion on VTEC-seropathotype and scientific criteria regarding pathogenicity assessment**. *EFSA Journal* 2013, **11**(4).
75. Nielsen EM, Scheutz F, Torpdahl M: **Continuous surveillance of Shiga toxin-producing Escherichia coli infections by pulsed-field gel electrophoresis shows that most infections are sporadic**. *Foodborne pathogens and disease* 2006, **3**(1):81-87.
76. Lindstedt BA, Heir E, Gjernes E, Vardund T, Kapperud G: **DNA fingerprinting of Shiga-toxin producing Escherichia coli O157 based on Multiple-Locus Variable-Number Tandem-Repeats Analysis (MLVA)**. *Annals of clinical microbiology and antimicrobials* 2003, **2**:12.
77. Majowicz SE, Scallan E, Jones-Bitton A, Sargeant JM, Stapleton J, Angulo FJ, Yeung DH, Kirk MD: **Global incidence of human Shiga toxin-producing Escherichia coli infections and deaths: a systematic review and knowledge synthesis**. *Foodborne pathogens and disease* 2014, **11**(6):447-455.
78. Rivero MA, Passucci JA, Rodriguez EM, Parma AE: **Role and clinical course of verotoxigenic Escherichia coli infections in childhood acute diarrhoea in Argentina**. *Journal of medical microbiology* 2010, **59**(Pt 3):345-352.
79. **Multistate outbreak of Escherichia coli O157:H7 infections associated with eating ground beef--United States, June-July 2002**. *MMWR Morbidity and mortality weekly report* 2002, **51**(29):637-639.
80. **Two multistate outbreaks of Shiga toxin--producing Escherichia coli infections linked to beef from a single slaughter facility - United States, 2008**. *MMWR Morbidity and mortality weekly report* 2010, **59**(18):557-560.
81. Severi E, Vial F, Peron E, Mardh O, Niskanen T, Takkinen J: **Community-wide outbreaks of haemolytic uraemic syndrome associated with Shiga toxin-producing Escherichia coli O26 in Italy and Romania: a new challenge for the European Union**. *Euro Surveill* 2016, **21**(49).
82. CDC: **FoodNet Surveillance Report for 2011 (Final Report)**. In: *Atlanta, Georgia: US Department of Health and Human Services.* 2012.
83. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA: **RNA-guided editing of bacterial genomes using CRISPR-Cas systems**. *Nature biotechnology* 2013, **31**(3):233-239.
84. Esvelt KM, Wang HH: **Genome-scale engineering for systems and synthetic biology**. *Molecular systems biology* 2013, **9**:641.
85. Jiang Y, Chen B, Duan C, Sun B, Yang J, Yang S: **Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system**. *Applied and environmental microbiology* 2015, **81**(7):2506-2514.
86. Addgene: **CRISPR 101: A Desktop Resource**. *Addgene* 2017.

87. Donnenberg MS, Kaper JB: **Construction of an eae deletion mutant of enteropathogenic Escherichia coli by using a positive-selection suicide vector**. *Infection and immunity* 1991, **59**(12):4310-4317.
88. Ma J, Ibekwe AM, Yi X, Wang H, Yamazaki A, Crowley DE, Yang CH: **Persistence of Escherichia coli O157:H7 and its mutants in soils**. *PloS one* 2011, **6**(8):e23191.
89. Ogierman MA, Paton AW, Paton JC: **Up-regulation of both intimin and eae-independent adherence of shiga toxigenic Escherichia coli O157 by ler and phenotypic impact of a naturally occurring ler mutation**. *Infection and immunity* 2000, **68**(9):5344-5353.
90. Carvalho HM, Teel LD, Kokai-Kun JF, O'Brien AD: **Antibody against the carboxyl terminus of intimin alpha reduces enteropathogenic Escherichia coli adherence to tissue culture cells and subsequent induction of actin polymerization**. *Infection and immunity* 2005, **73**(4):2541-2546.
91. Bikard D, Euler CW, Jiang W, Nussenzweig PM, Goldberg GW, Duportet X, Fischetti VA, Marraffini LA: **Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials**. *Nature biotechnology* 2014, **32**(11):1146-1150.
92. Martinez-Lage M, Puig-Serra P, Menendez P, Torres-Ruiz R, Rodriguez-Perales S: **CRISPR/Cas9 for Cancer Therapy: Hopes and Challenges**. *Biomedicines* 2018, **6**(4).
93. Rath D, Amlinger L, Rath A, Lundgren M: **The CRISPR-Cas immune system: biology, mechanisms and applications**. *Biochimie* 2015, **117**:119-128.
94. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD: **Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin**. *Microbiology (Reading, England)* 2005, **151**(Pt 8):2551-2561.
95. Anton T, Karg E, Bultmann S: **Applications of the CRISPR/Cas system beyond gene editing**. *Biology Methods and Protocols* 2018, **3**(1).
96. Jiang F, Doudna JA: **CRISPR-Cas9 Structures and Mechanisms**. *Annual review of biophysics* 2017, **46**:505-529.
97. Gasiunas G, Barrangou R, Horvath P, Siksnys V: **Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria**. *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**(39):E2579-2586.
98. Cui L, Bikard D: **Consequences of Cas9 cleavage in the chromosome of Escherichia coli**. *Nucleic acids research* 2016, **44**(9):4243-4251.
99. Zhang C, Quan R, Wang J: **Development and application of CRISPR/Cas9 technologies in genomic editing**. *Human molecular genetics* 2018, **27**(R2):R79-r88.
100. Shaaban S, Cowley LA, McAteer SP, Jenkins C, Dallman TJ, Bono JL, Gally DL: **Evolution of a zoonotic pathogen: investigating prophage diversity in enterohaemorrhagic Escherichia coli O157 by long-read sequencing**. *Microbial genomics* 2016, **2**(12):e000096.
101. Aas CG, Drablos F, Haugum K, Afset JE: **Comparative Transcriptome Profiling Reveals a Potential Role of Type VI Secretion System and Fimbriae in Virulence of Non-O157 Shiga Toxin-Producing Escherichia coli**. *Frontiers in microbiology* 2018, **9**:1416.
102. Haugum K, Johansen J, Gabrielsen C, Brandal LT, Bergh K, Ussery DW, Drablos F, Afset JE: **Comparative genomics to delineate pathogenic potential in non-O157 Shiga toxin-producing Escherichia coli (STEC) from patients with and without haemolytic uremic syndrome (HUS) in Norway**. *PloS one* 2014, **9**(10):e111788.
103. Zerbini F, Zanella I, Fraccascia D, Konig E, Irene C, Frattini LF, Tomasi M, Fantappie L, Ganfini L, Caproni E *et al*: **Large scale validation of an efficient CRISPR/Cas-based multi gene editing protocol in Escherichia coli**. *Microbial cell factories* 2017, **16**(1):68.
104. Kang S, Kim J, Hur JK, Lee SS: **CRISPR-based genome editing of clinically important Escherichia coli SE15 isolated from indwelling urinary catheters of patients**. *Journal of medical microbiology* 2017, **66**(1):18-25.
105. Invitrogen™: **MAX Efficiency® DH5α™ Competent Cells**. *Thermo Fisher Scientific* 2006.

106. Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, Madden TL, Matten WT, McGinnis SD, Merezhuk Y *et al*: **BLAST: a more efficient report with usability improvements**. *Nucleic acids research* 2013, **41**(Web Server issue):W29-33.
107. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL: **Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction**. *BMC bioinformatics* 2012, **13**:134.
108. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C *et al*: **Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data**. *Bioinformatics (Oxford, England)* 2012, **28**(12):1647-1649.
109. Rychlik W: **OLIGO 7 primer analysis software**. *Methods in molecular biology (Clifton, NJ)* 2007, **402**:35-60.
110. König E, Zerbini F, Zanella I, Fraccascia D, Grandi G: **Multiple Stepwise Gene Knockout Using CRISPR/Cas9 in Escherichia coli**. *Bio-Protocol* 2018, **8**(2).
111. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R *et al*: **Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9**. *Nature biotechnology* 2016, **34**(2):184-191.
112. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O *et al*: **DNA targeting specificity of RNA-guided Cas9 nucleases**. *Nature biotechnology* 2013, **31**(9):827-832.
113. New England Biolabs: **NEB® Golden Gate Assembly Kit (BsaI-HF®v2) Instruction manual**. 2018.
114. Hanahan D, Jessee J, Bloom FR: **Plasmid transformation of Escherichia coli and other bacteria**. *Methods in enzymology* 1991, **204**:63-113.
115. Andreason GL, Evans GA: **Introduction and expression of DNA molecules in eukaryotic cells by electroporation**. *BioTechniques* 1988, **6**(7):650-660.
116. Bio-Rad: **Quantum Prep® Plasmid Miniprep kit Instruction manual**.
117. Addgene: **Plasmids 101: A Desktop Resource**. *Addgene* 2017.
118. Thermo Fisher Scientific: **Qubit® 3.0 Fluorometer**. *Thermo Fisher Scientific* 2014.
119. Thermo Scientific: **NanoDrop 1000 Spectrophotometer User's Manual** *Thermo Fisher Scientific* 2010, **V3.8**.
120. Green MR, Sambrook J: **Polymerase Chain Reaction**. *Cold Spring Harb Protoc* 2019, **2019**(6):pdb top095109.
121. Rychlik W, Spencer WJ, Rhoads RE: **Optimization of the annealing temperature for DNA amplification in vitro**. *Nucleic acids research* 1990, **18**(21):6409-6412.
122. Green MR, Sambrook J: **Analysis of DNA by Agarose Gel Electrophoresis**. *Cold Spring Harbor Protocols* 2019, **2019**(1):pdb.top100388.
123. Beckman Coulter: **CEQ™ 8800 Genetic Analysis System User's Guide**. 2004.
124. Pyne ME, Moo-Young M, Chung DA, Chou CP: **Coupling the CRISPR/Cas9 System with Lambda Red Recombineering Enables Simplified Chromosomal Gene Replacement in Escherichia coli**. *Applied and environmental microbiology* 2015, **81**(15):5103-5114.
125. Chayot R, Montagne B, Mazel D, Ricchetti M: **An end-joining repair mechanism in Escherichia coli**. *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**(5):2141-2146.
126. Zhao D, Yuan S, Xiong B, Sun H, Ye L, Li J, Zhang X, Bi C: **Development of a fast and easy method for Escherichia coli genome editing with CRISPR/Cas9**. *Microbial cell factories* 2016, **15**(1):205.
127. Franz E, van Hoek AH, Wuite M, van der Wal FJ, de Boer AG, Bouw EI, Aarts HJ: **Molecular hazard identification of non-O157 Shiga toxin-producing Escherichia coli (STEC)**. *PloS one* 2015, **10**(3):e0120353.

128. Ritchie JM, Thorpe CM, Rogers AB, Waldor MK: **Critical roles for stx2, eae, and tir in enterohemorrhagic Escherichia coli-induced diarrhea and intestinal inflammation in infant rabbits**. *Infection and immunity* 2003, **71**(12):7129-7139.
129. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK: **Improving CRISPR-Cas nuclease specificity using truncated guide RNAs**. *Nature biotechnology* 2014, **32**(3):279-284.
130. Weller GR, Kysela B, Roy R, Tonkin LM, Scanlan E, Della M, Devine SK, Day JP, Wilkinson A, d'Adda di Fagagna F *et al*: **Identification of a DNA nonhomologous end-joining complex in bacteria**. *Science (New York, NY)* 2002, **297**(5587):1686-1689.
131. Srivastava M, Raghavan SC: **DNA double-strand break repair inhibitors as cancer therapeutics**. *Chemistry & biology* 2015, **22**(1):17-29.
132. Song F, Stieger K: **Optimizing the DNA Donor Template for Homology-Directed Repair of Double-Strand Breaks**. *Molecular therapy Nucleic acids* 2017, **7**:53-60.
133. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA: **DNA interrogation by the CRISPR RNA-guided endonuclease Cas9**. *Nature* 2014, **507**(7490):62-67.
134. Yahara K, Didelot X, Ansari MA, Sheppard SK, Falush D: **Efficient inference of recombination hot regions in bacterial genomes**. *Molecular biology and evolution* 2014, **31**(6):1593-1605.
135. Scientific TF: **Comparison of fluorescence-based quantitation with UV absorbance measurements**. 2014.

Appendix A: Recipes

Composition of culture media and antibiotic solution used in this study.

Agar plates

LB (Luria-Bertani) agar was used to culture the wild type FHI11 strain and LB agar containing appropriate antibiotics was used to culture strains containing plasmid/plasmids harbouring an antibiotic cassette. 4 g of LB broth with agar was suspended in 100 ml distilled water. The mixture was sterilized by autoclaving at 15 lbs pressure for 15 min at 121 °C. After autoclaving the medium was poured into sterile petri dishes. When preparing LB agar plates containing antibiotics, the solution was cooled down to 43±2 °C and the appropriate antibiotic was added.

Quality control of agar plates was performed to test for contamination. To make sure the antibiotics function properly, positive and negative controls were plated on agar containing antibiotic/antibiotics. The FHI11 strain was used as negative control. *E. coli* DH5α containing the pCRISPR plasmid was used as positive control for LB containing kanamycin and *E. coli* DH5α containing the pCas9 plasmid was used as positive control for LB agar containing chloramphenicol.

Medium/reagent	**Volume/ Weight**	**Manufacturer**
LB broth with agar	4 g	Sigma-Aldrich
Distilled water	100 ml	

Preparation of stock solutions of antibiotics for bacterial selection

Antibiotic	**Solvent**	**Stock concentration (mg/ml)**	**Working concentration (µg/ml)**	**Manufacturer**
Kanamycin (Kan)	Distilled water	50	50	Gibco® by life technology
Chloramphenicol (Chl)	95% Ethanol	100	25	Sigma-Aldrich

Dry powder formulation of kanamycin was dissolved in sterile distilled water and sterilized using a sterile filter with Millex GP 0.22 µm pore diameter. Chloramphenicol was received as 100 mg/ml stock solution.

Liquid cultures

20 g of LB broth was suspended in 1000 ml distilled water. The mixture was sterilized by autoclaving at 15 lbs pressure for 15 min at 121 °C. When preparing LB broth containing antibiotics, the solution was cooled down to 43±2 °C and the appropriate antibiotic was added.

Quality control of LB broth was performed to test for contamination. 2 ml of LB broth was added into a 15 ml Falcon tube and incubated for 24 hrs at 37 °C, and then tested for any growth. To make sure the antibiotics functioned properly, positive and negative controls were inoculated into LB broth containing antibiotics. Freshly prepared LB broth containing antibiotics were used to grow the strains harbouring plasmids.

LB broth composition

Medium/reagent	**Volume/ Weight**	**Manufacturer**
Tryptone (Pancreatic digest of casein)	10 g	Sigma-Aldrich
Yeast extract	5g	Sigma-Aldrich
Sodium chloride	5g	Sigma-Aldrich
Sterile distilled water	1000 ml	

LB with kanamycin (50 µg/ml)

Medium/reagent	**Volume**	**Manufacturer**
LB medium	5 ml	
Kanamycin (50 mg/ml)	5 µl	Gibco® by life technology

LB with chloramphenicol (25 µg/ml)

Medium/reagent	Volume	Manufacturer
LB medium	5 ml	
Chloramphenicol (100 mg/ml)	1.25 µl	Sigma-Aldrich

LB with kanamycin (50 µg/ml) and chloramphenicol (25 µg/ml)

Medium/reagent	Volume	Manufacturer
LB medium	5 ml	
Chloramphenicol (100 mg/ml)	1.25 µl	Sigma-Aldrich
Kanamycin (50 mg/ml)	5 µl	Gibco® by life technology

Other solutions

0.5 X Tris-Borate-EDTA (TBE) buffer

TBE buffer is used as both a running buffer and gel preparation buffer.

Medium/ reagent	Volume	Manufacturer
10X TBE buffer	100 ml	Sigma-Aldrich
Reverse Osmosis (RO) water	1900 ml	

Agarose Gel

To resolve short fragments of nucleic acid (<1 Kb), 2% agarose was used whereas 1% agarose was used to resolve long fragments of nucleic acid (>1 Kb). Agarose was dissolved in 0.5X TBE-buffer and melted in a microwave oven for 4 min at 650W, swirled in-between to mix. Then the solution was kept in an incubator for 30 min at 60 $^{\circ}$C to reach equilibrium.

Medium/ reagent	Volume/ Weight	Manufacturer
1% or 2% agarose	1.0 g or 2.0 g	Sigma-Aldrich
0.5 X Tris EDTA buffer	100 ml	Sigma-Aldrich

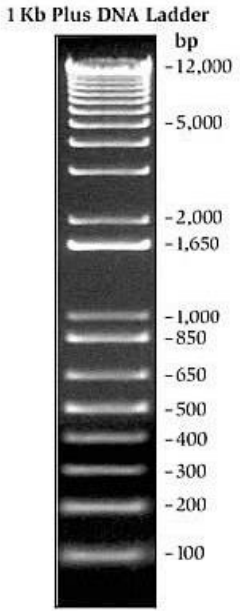
Gel Red

GelRed™ is a fluorescent nucleic acid dye for staining dsDNA, ssDNA or RNA in agarose gels or polyacrylamide gels. Compare to toxic ethidium bromide GelRed™ is environmentally safe and far more sensitive. The prepared stain was stored at room temperature and protected from light exposure.

Medium/ reagent	Volume	Manufacturer
GelRed™ Nucleic Acid Gel Stain 10.000X	150 µl	Biotium
RO water	450 µl	

DNA ladder

Working solution of 1 Kb Plus DNA Ladder was made as 1:10 with TE-buffer.



50% Glycerol

Medium/ reagent	Volume	Manufacturer
Glycerol (99%)	50 ml	Sigma-Aldrich
Sterile distilled water	50 ml	

The prepared solution was autoclaved and stored in the refrigerator until use.

10% Glycerol

Medium/ reagent	Volume	Manufacturer
Glycerol (99%)	10 ml	Sigma-Aldrich
Autoclaved Molecular MGW	90 ml	Sigma-Aldrich

The prepared solution was autoclaved and stored in the refrigerator until use. Sterile and ice-cold solution was used for washing purpose in electrocompetent cell preparation.

1M NaCl

Medium/ reagent	Volume	Manufacturer
5 M NaCl solution	1 ml	Sigma-Aldrich
Molecular grade water	4 ml	Sigma-Aldrich

The prepared solution was sterilized using a sterile filter with Millex GP 0.22μm pore diameter and stored at room temperature.

Appendix B: Thermocycler conditions used in gradient PCR

Thermocycler conditions used in gradient PCR with primers InserID_F and InserID_R.

Steps	Temperature	Time (Minutes)
1. Enzyme activation	94 ˚C	10:00
2. Denaturation	94 ˚C	00:30
3. Annealing	1: 60 ˚C 2: 60.4 ˚C 3: 62.3 ˚C 4: 63.7 ˚C 5: 64.8 ˚C 6: 65.6 ˚C 7: 66 ˚C	00:30
4. Extension	72 ˚C	01:00
Repeat step 2-4 for 25 cycles		
5. Final extra elongation	72 ˚C	05:00

Thermocycler condition used in gradient PCR with primers KOveri_F and KOveri_R.

Steps	Temperature	Time (Minutes)
1. Enzyme activation	94 ˚C	10:00
2. Denaturation	94 ˚C	00:30
3. Annealing	1: 58 ˚C 2: 59.4 ˚C 3: 60.7 ˚C 4: 62.3 ˚C 5: 63.6 ˚C 6: 64.5 ˚C 7: 65.0 ˚C	00:30
4. Extension	72 ˚C	01:00
Repeat step 2-4 for 25 cycles		
5. Final extra elongation	72 ˚C	05:00