A novel locus of enterocyte effacement (LEE) pathogenicity island inserted at \textit{pheV} in bovine Shiga toxin-producing \textit{Escherichia coli} strain O103:H2

Jörg Jores $^a$,*, Leonid Rumer $^a$, Sabine Kießling $^a$, James B. Kaper $^b$, Lothar H. Wieler $^a$

$^a$ Institut für Mikrobiologie und Tierseuchen, Freie Universität Berlin, Philippstraße 13, 10115 Berlin, Germany
$^b$ Center for Vaccine Development, School of Medicine, University of Maryland, 685 W. Baltimore St., Baltimore, MD 21201, USA

Received 1 June 2001; received in revised form 6 August 2001; accepted 6 August 2001

First published online 17 September 2001

Abstract

We describe a locus of enterocyte effacement (LEE) which is part of a new pathogenicity island (PAI) detected in the bovine Shiga toxin-producing \textit{Escherichia coli} strain RW1374 (O103:H2). This PAI is at least 80 kb in size and inserted in the vicinity of the \textit{pheV} tRNA gene at 67 min of the \textit{E. coli} chromosome. Furthermore, the PAI differs from the previously described LEEs by unique flanking regions at both sides, which harbor one copy each of an insertion element in an inverted orientation that is 96% identical to insertion site (IS)629. In addition, a 5-kb PAI-specific sequence downstream of the LEE core region and adjacent to the \textit{E. coli} K12 region is duplicated upstream of the LEE core region as well. The duplicated sequences are more than 80% identical to each other and consist partially of prophage sequences.

Keywords: Pathogenicity island; Locus of enterocyte effacement; Insertion site element; \textit{pheV}; Insertion site

1. Introduction

The locus of enterocyte effacement (LEE) is a pathogenicity island (PAI) required for attaching and effacing (AE) lesions produced on epithelial cells of humans and animals by numerous enteropathogenic (EPEC) and enterohemorrhagic (EHEC) \textit{Escherichia coli} strains and other related bacteria. The LEE encodes a type III secretion system, an adhesin (intimin) responsible for the intimate attachment of the bacteria to the cell and a number of secreted proteins involved in signal transduction [1].

The LEE was first identified by McDaniel et al. (1995) in the EPEC strain E2348/69 (O127:H6) [2]. This LEE is located in the vicinity of the \textit{seIC} tRNA gene and contains 35450 bp [2,3]. In the same chromosomal position of EHEC strain EDL933 (O157:H7) a similar LEE was found, which differed from its EPEC homolog by the inclusion of a cryptic 7.5 kb prophage at one end of the LEE [4]. The prophage genes appear to be uninvolved in the AE phenotype [4]. The same \textit{seIC} location of the LEE was shown for several EPEC and EHEC strains which belong to close evolutionary lineages [5]. At the same time many LEE-harboring strains had intact \textit{seIC} loci indicating a different chromosomal location of the LEE. These strains were evolutionary distant to those with the LEE situated in \textit{seIC} locus [5]. A second LEE insertion site (IS) was identified by Benkel et al. (1997) who showed that the LEE in the bovine Shiga toxin-producing \textit{E. coli} (STEC) strain 413/89-1 (O26:H-) was located in the vicinity of the \textit{pheU} gene encoding tRNA for phenylalanine [6]. Sequence information for only a few internal genes of this LEE is available in the database. Sperandio et al. (1998) showed, that a number of LEE-harboring strains had both the \textit{seIC} and the \textit{pheU} locus intact, which therefore suggested the existence of a third LEE-IS [7]. Recently the LEE of the rabbit diarrheagenic \textit{E. coli} strain (RDEC1) was cloned, sequenced and compared to the other LEEs [8].

In this work we investigated the LEE of the bovine
strain *E. coli* RW1374 (O103:H2) and found that the LEE in this strain is located within a larger PAI, which integrated in the vicinity of the gene *pheV*. The *pheV* gene is absolutely identical to *pheU* but is located in a different chromosomal region.

2. Materials and methods

2.1. Bacterial strains and their properties

STEC strain RW1374 (O103:H2) was isolated from a fecal sample of a hospitalized calf suffering from bloody diarrhea in Hessia/Germany in 1993. The ability of strain RW1374 to cause AE lesions was detected in HEp-2 cells by using the fluorescence actin staining (FAS) test [9]. Intimin type, virulence genes, and intactness of the *selC* and *pheU* locus were determined as described in Section 3. Strain RW1374 was cultivated in LB medium at 37°C.

2.2. DNA preparation

Plasmid DNA and total genomic DNA was extracted using standard protocols [10]. Cosmid DNA was isolated by using the Qiagen® large-construct kit (Qiagen, Hilden, Germany).

2.3. Genomic library of RW1374

Total DNA was partially digested with restriction endonuclease *Sma*I, size-fractionated in a sucrose gradient and dephosphorylated prior to ligation in cosmid SuperCos1 (Stratagene, Amsterdam, the Netherlands) by standard protocols [10]. Cosmids were packaged in Lambda particles with Gigapack III Gold (Stratagene) and strain XL1blue MR’ was infected. All enzymes and kits were purchased from Stratagene.

2.4. PCR

All PCR reactions were carried out in a total volume of 50 µl with Herculase® Enhanced Polymerase Blend (Stratagene) in a GeneAmp® PCR System 2400 (Perkin Elmer, Langen, Germany) thermal cycler. Inverse PCR with the total RW1374 DNA as template and the primers ipheV1 (5’-GACGAGGCGAATCAGGTTTA-3’) and ipheV2 (5’-ATTGTCAGGCAAACCCAGG-3’) was carried out according to standard protocols [11]. Long-term PCR was performed with the total RW1374 DNA as template and the primers sepQ (5’-GGTATGAGGGGCGTTTTTGGAT-3’) and sepL (5’-TTTTTGGGCAGTGATGACTCCTT-3’) as recommended by the manufacturer’s protocols.

![Schematic diagram of cosmid clones derived from *E. coli* RW1374 and a comparison to the diagram of the LEE in strain EDL933. The numbers in the upper part of the diagram correspond to nucleotide numbers of the EDL933 LEE (AF071034), the region between 9214 bp and 44 043 bp corresponds to similarity of the EDL933 LEE and RW1374 LEE.](https://academic.oup.com/femsle/article-abstract/204/1/75/633075)
2.5. DNA–DNA hybridization

Hybridization was done with digoxigenin-labeled probes LEE A, B, C, and D, which are specific for LEE fragments distributed throughout its length [2] using the Roche labeling and detection kit (Roche, Mannheim, Germany). 500 ng of DNA was used for hybridization experiments. Blotting was done using Nylon Membranes, positively charged (Roche, Mannheim, Germany) as recommended by the manufacturer’s protocols.

2.6. DNA sequence analysis

DNA sequence analysis was performed by the chain termination sequencing technique [12] by AGOWA GmbH, Berlin, Germany.

3. Results and discussion

Strain RW1374 was positive in the FAS test, indicating that this strain was able to produce AE lesions on epithelial cells. Positive hybridization (Fig. 1) with the specific LEE probes A, B, C and D [2] indicated the presence of the entire LEE in this strain. Intimin typing by PCR using the system developed by Oswald et al. (2000) revealed an epsilon type intimin [13]. PCR analysis for Shiga toxin genes showed also that the strain harbored an stx1 gene [14]. It also contained a > 70-kb plasmid, which did not hybridize with the LEE-specific probes, but did hybridize with the hlyEHEC-specific probe [15].

In an attempt to localize the IS of the LEE-containing PAI in RW1374, we first tested the possible involvement of the selC and the pheU locus using PCR analysis [5,7]. The selC locus was shown to be intact, while the PCR specific for an intact pheU locus yielded no product suggesting the possible disruption of the pheU locus. Southern blot hybridization of NotI-restriction pattern of the RW1374 chromosome with the LEE probes indicated that the LEE was present on only one NotI-restriction fragment of 900 kb (data not shown).

The cosmid library of RW1374 was screened by hybridization with the LEE-specific probes A, B, C and D [2] (Fig. 1). Three cosmids were identified that gave positive signals in at least one of the hybridization experiments. cIMT1-1 was positive for gene probe A, cIMT1-2 was positive for gene probe A and B, and cIMT1-3 was positive for gene probe D (Fig. 1). None of the cosmids was positive for gene probe C. Restriction endonuclease analysis and cross-hybridization experiments of the clones (data not shown) showed that cIMT1-1 and cIMT1-2 clones overlapped for about 25 kb, while the cIMT1-3 clone did not overlap with either of the two other clones (Fig. 1). We then sequenced about 500 bp on each end of the cloned segments and compared the sequences to the database entries. The sequence of the cIMT1-1 clone was on one flank highly identical to the E. coli MG1655 chromosome (GenBank accession number AE000380), and on the other flank to the open reading frame (orf) L0051 of EDL933 LEE (Fig. 1). In the cIMT1-2 clone one flank was highly identical to the MG1655 chromosome (GenBank accession number AE000380) and the opposite flank contained sequences similar to the LEE gene sepQ (Fig. 1). For the cIMT1-3 clone one flank had no compelling similarity to any database entries while on the other flank high similarity to the LEE gene sepL was found. This sequence information enabled us to design two oligonucleotide primers (sepQ-specific PcIMT1-2 and sepL-specific PcIMT1-3) in order to amplify the region between sepQ and sepL, and to close the gap between the cIMT1-2 and cIMT1-3 clones. By hybridization with the gene probe C we verified that the resulting 9-kb PCR product contained the intimin (eae) gene (data not shown) (Fig. 1). Afterwards 20 kb of the cosmid cIMT1-3 (GenBank accession number AJ303142) and about 11 kb of cosmid cIMT1-1 (GenBank accession number AJ303141) were double-strand sequenced. In each case colinearity of the sequence was determined via PCR or hybridization experiments.

![Diagram](https://academic.oup.com/femsle/article-abstract/204/1/75/633075/fig2)
Sequence analysis of the cosmide cimI-1 showed the insertion of the PAI to be in the pheV tRNA gene (Fig. 2). On the right end of the PAI a stringent and extensive similarity with the MG1655 chromosome started with the sequence TCGATTCCGAGTCC-GGCA homologous (with one internal deletion) to the sequence of the last 23 bp of the 3'-end of the pheV tRNA gene. Downstream of this 22-bp region the identity with the pheV-flanking sequence of MG1655 continued. The very left sequence of the PAI was adjacent to a copy of the intact pheV gene as shown by sequencing of the inverse PCR product (data not shown). The region of the PAI spanned by the cosmid clones and the 9-kb PCR product was about 80 kb. So far we were not able to identify the entire length of the PAI (Fig. 1).

For comparison of the LEE-specific part of the PAI with the LEE of strains EDL933 and E2348/69 we assumed the terminology and orientation offered for the description of the EDL933 LEE [4], but we did not take into account the prophage that is inserted near one end of the EDL933 LEE, but is not present in the E2348/69 LEE. Therefore we refer to the L0016 (Fig. 1) as the first orf of the RW1374 LEE and to L0056 as the last one. The starting point of the similarity between the LEE-specific part of the RW1374 PAI and the EDL933 LEE corresponded to the bp number 9214 of the EDL933 LEE [4] (GenBank accession number AF071034), that is 68 bp upstream of the left end of the L0016 and 656 bp downstream of the left end of the EDL933 LEE (Fig. 1). Downstream of the last orf L0056 (Figs. 1 and 2) the similarity extended for 186 bp and was then interrupted at the point corresponding to bp number 43742 of the EDL933 LEE. After that the similarity resumed 957 bp downstream, spanned 293 bp more and finally terminated at the point corresponding to bp number 44043 of the EDL933 LEE, which is 249 bp upstream of its right end. All the LEE-specific genes of RW1374 are included between the homologs of the first, L0016, and the last, L0056, orfs of the EDL933 LEE and in the same order as in EDL933. The distance between the left border of the L0016 homolog and the right border of the L0056 homolog in RW1374 was about 34 kb. These data implied that the new PAI in RW1374 included a homolog of the EDL933 LEE. In comparison to the EDL933 LEE it lacked the last 249 bp on the right end and the first 656 bp on its left end. We therefore refer to the PAI sequence containing the region between the left and right ends of similarity to the EDL933 LEE as the RW1374 LEE.

The PAI-specific sequence downstream of the right end of the RW1374 LEE (Fig. 2) was determined and was found to contain 7815 bp between the end of the RW1374 LEE and the K12 sequences. Upstream of the left end of the LEE 20 kb were sequenced, but K12 sequences were not present in this 20-kb region. Analysis of the LEE-flanking sequences indicated that the LEE was lying between two homologs of the IS element IS629 originally identified in Shigella sonnei [16]. The right hand element was located 373 bp downstream of the RW1374 LEE right end, and the left hand element 9727 bp upstream of its left end (Fig. 2). The elements were positioned in opposite orientations. ORF analysis revealed that the transposase genes of both homologs were intact. To our knowledge this is the first time that the LEE has been shown to be flanked by two intact IS elements and therefore this region constitutes a putative transposon structure. An additional potential mobilizing element was located downstream of the right end of the RW1374 LEE, where the PAI-specific sequence contains three fragments with similarity to the putative prophage of EDL933 LEE (Fig. 2). The sequence consisting of the last 5 kb of the PAI and the 23-bp homolog of the 3'-terminal pheV stretch was highly similar (78%) to the 5-5 kb PAI fragment located 9.3 kb upstream of the RW1374 LEE left end. We will refer to these two fragments as the left and the right similarity segments respectively. The right end of the similarity fragment was similar to the right end of the similarity to the she PAI of Shigella flexneri 2a [17]. This similarity is interrupted by the IS629 homolog which splits this region in two parts, one with the length of 5837 bp (downstream of the IS element) and the other with a length of 7239 bp (upstream of the IS element). Within each fragment the long stretches of similarity were punctuated by short stretches of no similarity. Joined together the two parts constituted a fragment with an overall similarity of 88% to the region of the she PAI located at nucleotide 34623 bp to nucleotide 47701 bp at the end of the she PAI. This region did not contain any obvious homologs of previously described virulence genes. Upstream of this region there was no compelling similarity with the she PAI based on a 15.3-kb region at the left end of the cimT-3 cosmid clone, whose sequence was determined by single-strand sequencing.

A short stretch of similarity with the MG1655 chromosome in the vicinity of the pheU gene was located 662 bp downstream of the left similarity fragment. The similarity consisted of the last 23 bp of the pheU 3' end and the next 38 bp that included a palindrome of 24 bp (Fig. 2).

Until now no model for integration of the EDL933 LEE in the chromosome has been offered. In the EDL933 LEE no flanking repeat or transposon-, phage-, or plasmid-like elements are found in both flanking regions [1]. In contrast the position of the RW1374 LEE in the structure of the PAI leads to several potential explanations. Beside the possible transposition within the fragment between the two homologs of the IS629, the possibility of site-specific recombination can be considered. It is known that the tRNA genes are frequently used as integration sites of plasmids, phages, and PAIs [18]. The LEE is located within the PAI fragment confined by two homologs of the 23-bp pheV 3'-end terminal stretch. The right homolog terminates the PAI, suggesting that the integration point of the PAI or its precursor lies within this stretch. One pos-
sible scenario therefore is that first the PAI precursor (without LEE), which carried the homolog of the 23-bp pheV terminal stretch, integrated in the pheV gene of the recipient strain by the site-specific recombination, which involved the interaction between the 23-bp pheV terminal stretch and its homolog on the PAI precursor. Then the LEE-containing fragment might have integrated as an autonomous element which also carried a homolog of the 23-bp pheV 3′-end terminal stretch, and integrated into the PAI precursor at the same point and by the similar mechanism, resulting in a copy of the 23-bp pheV 3′-end terminal stretch on each flank of the fragment. It can also not be ruled out that the integration point is situated somewhere in the similarity fragment.

Further characterization and comparative analysis of the LEE PAIs in strains of different evolutionary lineages should elucidate insight into the precise mechanisms of the origin and dissemination of the LEE among *E. coli* pathogens.

Acknowledgements

This work and Leonid Rumer were supported by Grant WI 1436/3-2 from the Deutsche Forschungsgemeinschaft.

References


