

Szent István University
Postgraduate School of Veterinary Science

**Genetic background of the virulence factors of atypical
bovine *Escherichia coli* O157 strains**

PhD dissertation theses

Domonkos László Sváb

2013

Supervisor and member of supervising committee

Supervisor:

István Tóth, DSc

Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian
Academy of Sciences

Member of supervising committee:

Béla Nagy, DVM, DSc, member of the Hungarian Academy of Sciences

Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian
Academy of Sciences

Introduction and aims

In my dissertation titled '**Genetic background of the virulence factors of atypical bovine *Escherichia coli* strains,**' we focused on an unusual group of *E. coli* O157 strains identified from cattle by the enteric bacteriology group of the earlier Veterinary Medical Research Institute (Hungarian Academy of Sciences). Ruminants, mainly cattle are usually the reservoir of this significant human pathogen. Besides numerous enterohemorrhagic (EHEC) and enteropathogenic (EPEC) strains, *E. coli* strains of the O157 serogroup were also identified which represented rare serotypes and had an entirely different set of virulence factors, therefore they were classified as atypical. The majority of these atypical strains produced type V of the cytolethal distending toxin (CDT-V), and carried the operon *lpf2*, encoding type 2 of the putative adhesion factor long polar fimbria (Lpf2).

CDT belongs to the cyclomodulins, the toxins inhibiting the cell cycle, its numerous types are produced by several species of Gram-negative human and animal pathogens. In *E. coli*, five genetic types are known (CDT-I-V), with CDT-V produced mainly by EHEC and strains which cannot be assigned to particular pathotypes. P2-like phage sequences have been identified in the flanking regions of the operon coding CDT-V (*cdt-V*), but only small fragments of sequences have been published up to date.

Lpf is a hair-like cell surface structure widely present in *E. coli* strains. Operons encoding Lpf can be grouped into two types (*lpf1* and *lpf2*) and within those, five and three allelic subtypes, respectively. According to data from the literature, the type *lpf2-1* is widely distributed pathogenic and commensalist *E. coli* strains, which represent numerous sero- and pathotypes. Experimental data also suggest that it has a role in adhesion, implying that Lpf2 is a significant virulence factor, especially in strains which do not have other known adhesins.

The atypical O157 strains gave us an opportunity to thoroughly investigate the genetic characteristics of these two virulence factors, and it became possible to investigate their phylogenetic and evolutionary relation to the 'typical' EHEC and EPEC members of the serogroup. Therefore in the present dissertation we investigated the following:

In the case of Lpf2 (*lpf2*)

- We wished to investigate the distribution of *lpf* operons in *E. coli* O157 strains representing different pathotypes
- By typing and sequencing the *lpf* operons, we wished to determine the diversity of this virulence factor encoding operon

- By analysing the sequence of the flanking regions of the operon, we wished to find the vector of the *lpf* operon and we wanted to monitor the distribution of these vector(s) in *E. coli* strains of different pathotypes

In the case of CDT (*cdt-V*)

- We wished to characterise the *cdt-V* operon and its flanking region in the available atypical *E. coli* O157 strains
- If, as literature suggests up to date, the *cdt-V* operon is indeed carried by a P2-like prophage, we wished to determine the genome sequence of the P2-like prophage in an *E. coli* O157:H43 strain
- In light of the phage genome sequence obtained from our model strain, we wished to acquire more information about the evolution of P2-like prophages by further sequencing and analysis of sequences available in GenBank

Whole genome sequencing

- In order to obtain further, more complete genetic information, we wished to determine the whole genome sequence of the *E. coli* O157:H43 strain T22

Materials and methods

Bacterial strains

We investigated altogether 97 strains, isolated from human disease (n=54) and from healthy cattle (n=43), and the members of the ECOR (*E. coli* Reference Collection, n=72) were investigated for the presence of *lpf* genes. The strains besides the ECOR collection represented three pathotypes: EHEC (n=64), EPEC (n=24) and the atypical strains, which did not carry *stx* nor *eae* genes (n=9). The EHEC and EPEC strains were from serotypes O157:H7 and O157:NM, the serotypes of the atypical strains were O157:H43 (n=3), O157:H37, O157:H12 (n=3), O157:H9 and O157:NM.

Phenotypical methods

In the case of four strains, it was necessary to confirm their serotypes, this was performed with latex agglutination test specific for O157 antigen (Oxoid), and hyperimmune serum specific for the H7 antigen.

Genotyping of strains

lpf allelic types of strains as well as the presence of *lpf* structural genes were investigated with primers available in the literature. Similarly, the presence of *stx1* and *stx2* (encoding

Shiga toxins), *eae* (encoding intimin) as well as genes encoding the heat-stable and heat-labile enterotoxins were investigated with primers published earlier. Phylogenetic grouping of the strains was determined with the triplex PCR system published by Clermont et al.

Presence of characteristic flanking regions of the *lpf2-1* and *cdt-V* operons were investigated by PCR using primers of our own design. This monitoring was performed in the case of 20 strains carrying *lpf2-1* and one *lpf*-negative strain, and in the case of *cdt-V*, we monitored the presence of flanking regions in eighteen strains, seven of which were *cdt-V* positive.

Preparation of cosmid clone library

We prepared a cosmid clone library from the genome of atypical *E. coli* strain T22, which belonged to serotype O157:H43, produced CDT-V and carried the *lpf2-1* operon, using the pWEB TNC Cosmid Cloning Kit (Epicentre). 1000 clones were kept as a clone library, clones carrying the *cdt-V* and *lpf2-1* operons were identified by PCR.

Sequencing

The nucleotide sequence of cosmids containing *cdt-V* and *lpf2-1*, as well as whole genome sequence of strain T22 was determined with the aid of next-generation sequencing platforms SOLiD 4, IonTorrent PGM and 454 Titanium. In order to obtain complete and reliable information, in some cases the dideoxynucleotide method was used as a supplement.

Bioinformatics

Sequence data were assembled using CLC Genomic Workbench 6.0, annotation was performed with the NCBI Prokaryotic Genomes Automatic Annotation Pipeline. Virtual MLST typing of strain T22 was performed using the PubMLST database maintained by University College Cork. In the case of the whole genome sequence of strain T22, the presence of characteristic virulence genes of different *E. coli* pathotypes and their integration sites was specifically investigated.

Investigation of gene expression

We investigated the expression of the *lpfA* gene with reverse transcription coupled PCR in strain T22.

Phage induction experiments

Phage induction experiments were performed on atypical *E. coli* O157 (n=5) and EHEC O157:NM (n=3) strains that carried *cdt-V* and P2-like phage sequences. Mitomycin-C,

norfloxacin and UV light were used as inducing agents. *E. coli* K-12 strains C600 and ER2738, as well as *Shigella sonnei* strain 866-F were used as propagating strains.

Adhesion experiments

In order to obtain information about the role of Lpf2 in adhesion, we performed adhesion experiments with strain T22, using primary cell cultures of bovine kidney and testes.

Results

***lpf* typing**

Out of the 97 investigated strains, 95 carried at least one type of *lpf* operon, and 88 carried both types (*lpf1* and *lpf2*). All the *lpf*-positive strains carried all of the *lpf* structural genes, confirming that they carried whole operons. The EHEC and EPEC strains uniformly carried the allelic types *lpf1-3* and *lpf2-2*. The seven *lpf*-positive atypical strain carried only the *lpf2-1* allelic type. We also determined the *lpf* types of the ECOR collection, 26 strains were *lpf*-positive, seven of them carried only *lpf1*, eleven strains carried only *lpf2*, and eight strains carried both operons. Among these positive strains, there were four which was found *lpf*-negative in earlier studies. There were also nine strains, which were found *lpf*-positive in earlier studies, but we did not detect any of the *lpf* variants in them.

The EHEC and EPEC strains belonged to phylogenetic group B. From the atypical strains six were members of group B1, and three of them were from group A.

Cloning of the *lpf* locus

With the aid of a cosmid clone library, we determined the nucleotide sequence of a 15.3 kb long genomic region of strain T22 containing the *lpf2-1* operon. We also determined the sequence of the *lpf2-1* operons in additional six atypical *E. coli* O157 strains. The GC content of the operons is 44% in contrast to the 52% GC content of the flanking regions. There were only four positions containing amino acid level polymorphism.

Out of the 20 investigated *lpf2-1* positive strains, fifteen carried all nine of the investigated flanking regions, with six of the atypical O157 strains among them.

The reverse transcription coupled PCR confirmed the expression of *lpfA* in the case of strain T22.

In the case of primary cell cultures of bovine kidney and testicles, no specific adhesion of strain T22 was observed.

Cloning the *cdt-V* operon and its flanking regions

The *cdt-V* operon is carried by a 31.2 kb long P2-like prophage integrated into the chromosome. The genes of the prophage show an identity between 94-100% to other P2-like prophage sequences available in GenBank. There are two insertions in the prophage: one is the *cdt-V* operon, with its GC content (41%) significantly different from the average GC content of the prophage (52%) the other is a fragment of a phage tail protein gene similar to a phage tail protein gene from an unknown type of prophage carried by an enterotoxigenic *E. coli* strain. From the 20 investigated characteristic flanking regions, the CDT-V producing atypical O157 strains besides T22 (n=4) were positive for at least 18. The CDT-V producing O157:NM strains (n=3) carried 15-16. Among the strains which did not produce CDT-V (n=11), only two carried 15 and 12 regions respectively, the others carried two regions at most. Our phage induction experiments demonstrated that these P2-like prophages are non-inducible.

Draft genome sequence of the *E. coli* O157:H43 strain T22

The size of the whole genome of *E. coli* O157:H43 strain T22 is 5.039.647 bp. The length of the chromosomal DNA is 4.959.535 bp, and the strain also carries a 80.112 bp long plasmid. The chromosome was assembled into 61, and the plasmid into three contigs, respectively. The chromosome contains 5587 predicted open reading frames (ORF), 5511 of which encodes proteins, 17 of the encodes rRNA, and 59 of them encodes tRNA. The plasmid contains 90 ORFs, 89 of which are protein genes and one of them is a tRNA gene.

The MLST typing based on the sequence of seven housekeeping genes indicates that strain T22 belongs to sequence type 155 (ST155).

Based on the sequence data, strain T22 does not harbour the characteristic virulence genes of EHEC (*stx*, *eae*), EPEC (*eae*, *bfp*), ETEC (*st*, *lt*), nor that of EAEC (*aggR*), neither on the chromosome nor on the plasmid.

GenBank accession numbers

The determined sequences were deposited in GenBank under the following accession numbers:

lpf2-1 operon of *E. coli* O157:H43 strain T22 and its flanking regions: **AHZD01000104**

lpf2-1 operon of *E. coli* O157:NM strain B47: **KC207119**

lpf2-1 operon of *E. coli* O157:H12 strain B54: **KC207120**

lpf2-1 operon of *E. coli* O157:H43 strain T16: **KC207121**

lpf2-1 operon of *E. coli* O157:H9 strain T34: **KC207122**

lpf2-1 operon of *E. coli* O157:H37 strain T49: **KC207123**

lpf2-1 operon of *E. coli* O157:H43 strain T50: **KC207124**

P2-like prophage carried by *E. coli* O157:H43 strain T22: **KC618326.1**

Draft genome sequence of *E. coli* O157:H43 strain T22: **AHZD00000000.2**

Discussion

The phylogenetical typing of *E. coli* strains belonging to the O157 serogroup and the genotyping of their *lpf* operons showed that there are one or more lineages of strains within this serogroup that harbours atypical virulence factors and is separate from the EHEC and EPEC strains of the serogroup. We confirmed earlier data from the literature that there is a correlation between the pathotype and the *lpf* allelic type of the strains. In the case of some ECOR strains, differences of *lpf* positivity compared to data of earlier studies could indicate that there are possibly as of yet unknown allelic types of *lpf*.

The sequence of the *lpf2-1* operon is highly conserved in the case of the atypical O157 strains, and it forms a pathogenicity island (PAI) within a similarly conserved genomic region, and these genes have close homologues widely spread among pathogenic and commensalist *E. coli* strains. Our results suggest that it is likely that the present distribution of *lpf* operons is a result of horizontal gene transfer (HGT) events. Further *in vitro* and *in vivo* experiments are necessary for revealing the potential role of Lpf2 in the adhesion of atypical O157 strains.

Expanding on earlier data from the literature, we confirmed that the *cdt-V* operon is indeed carried by a P2-like prophage. Monitoring the presence of P2-like prophage genes in both CDT-producing and CDT-negative strains indicated that these *cdt-V* carrying prophages have at least three lineages. One is identical or very similar to the one harboured by strain T22 and other atypical O157 strains of *E. coli*. The one carried by O157:NM EHEC strains is different from this line regarding more than one genes. A third line could be a P2-like prophage isolated and reported in the literature, which is inducible, but very little sequence data is available from it. These prophages likely originated from a common ancestor which spread among the different *E. coli* strains via phage transduction. Adaptation of the phages to the different bacterial hosts could have led to their present sequence heterogeneity and also to their temperation.

The nucleotide sequence of the above two virulence operons and their flanking regions was determined in the *E. coli* O157:H43 strain T22, and its whole genome was also sequenced and assembled as a draft genome. Sequence data indicate that this strain indeed represents a new genotype compared to O157 strains with known genomes, and can be classified as atypical. The integration sites of known key virulence factors in its genome are intact, this raises the possibility of the acquisition of these genes by the strain. Similar Stx and intimin negative atypical strains deserve our attention, because they can act as recipients for several more virulence factors, contribute to the evolution of pathogenic and commensalist *E. coli* strains alike, and their zoonotic significance cannot be excluded either.

Investigation of such atypical strains could help to understand the mechanisms underlying the genetic variability of the O157 serogroup, and to better assess the zoonotic potential of strains with different set of virulence factors.

New scientific results

In our works forming the base of my dissertation we achieved the following new scientific results:

1. We determined the genetic types of the *lpf* operons carried by bovine atypical *E. coli* O157 strains. We showed that these atypical strains represent a separate lineage from that of EHEC and EPEC strains of the serogroup regarding their *lpf* operons. While the atypical strains only carry the *lpf2-1* allelic variant, the EHEC and EPEC strains uniformly carry *lpf1-3* and *lpf2-2*.
2. We determined the *lpf* types of strains from the ECOR collection, and detected *lpf* genes which possibly represent new allelic types.
3. We were first to determine the sequence of the *lpf2* operon in bovine atypical *E. coli* O157 strains, showing that the *lpf2* operon forms a conserved PAI in these strains.
4. We were first to determine the nucleotide sequence of the P2-like prophage carrying the *cdt-V* operon, showing that the *cdt-V* operon was likely spread by HGT between different *E. coli* strains.
5. We showed that P2-like prophage genes are present in CDT-V producing *E. coli* strains regardless of sero- or pathotype.
6. We found differences in the distribution of P2-like prophage genes between the CDT-V producing atypical O157 strains and the O157:NM EHEC strains producing the same toxin. While the atypical strains carried at least 18 characteristic regions out of 20, the EHEC strains carried only 15-16 genes, among the missing genes was the regulatory gene C, which has phylogenetic significance in the case of P2-like phages. These differences are possibly the result of the prophages' adaptation to the hosts.
7. We were first to determine the draft genome of an atypical *E. coli* O157:H43 strain. Our results suggest that the *E. coli* strains of this serotype can represent a previously unknown stage in the evolution non-sorbitol-fermenting O157 strains, and as the integration sites of their key virulence genes are intact, they have the potential to become pathogenic.

Publications based on the results of the PhD dissertation

Research papers in peer-reviewed journals

Sváb D., Galli, L., Horváth B., Maróti G., Dobrindt, U., Torres, A.G., Rivas, M., Tóth I.: **The long polar fimbriae operon and its flanking regions in bovine *Escherichia coli* O157:H43 and STEC O136:H12 strains**, Pathog. Dis., 68. 1-7, 2013. (IF: 2,684)

Sváb D., Horváth B., Maróti G., Dobrindt, U., Tóth I.: **Sequence variability of P2-like phage genomes carrying the cytolethal distending toxin V operon in *Escherichia coli* O157**, Appl. Environ. Microbiol., 79. 4958-4964, 2013. (IF: 3,678)

Sváb D., Horváth B., Szűcs A., Maróti G., Tóth I.: **Draft genome sequence of an *Escherichia coli* O157:H43 strain isolated from cattle**, Genome Announc., 1. e00263-13, 2013.

Sváb D., Tóth I.: **Allelic types of long polar fimbriae in bovine and human *Escherichia coli* O157 strains**, Acta Vet. Hung., 60. 1-15, 2012. (IF: 1,173)

Sváb D., Tóth I.: **Citotétális duzzasztó toxinok állatra és emberre patogén *Escherichia coli*-ban**, Magyar Állatorv. Lapja, 135. 367-375, 2013. (IF: 0,146)

Book chapter

TÓTH I., SVÁB D.: Cell Cycle Modulating Toxins Produced by *Escherichia coli*. In: *Pathogenic Escherichia coli: Molecular and Cellular Microbiology*. Szerk.: MORABITO, S. Poole, Nagy-Britannia: Caister Academic Press, 2014. p. 117-138 (nyomdában).

Conference abstracts

Sváb D., Tóth I.: **Genotyping of long polar fimbriae in bovine *Escherichia coli* O157 strains**, Abstracts of the 2nd Central European Forum for Microbiology, October 7-9, 2009, Keszthely, Hungary. Acta Microbiol. Immunol., 56 (Suppl. 2). 245, 2009.

Sváb D., Tóth I.: **Cytolethal distending toxin V (*cdt-V*) operon is flanked by P2-like phage elements in bovine *Escherichia coli* O157**, Abstracts of the 16th International Congress of the Hungarian Society for Microbiology, July 20-22, 2011, Budapest, Hungary. Acta Microbiol. Immunol., 58 (Suppl. 1). 218, 2011.

Sváb D., Horváth B., Maróti G., Tóth I.: **Cloning and sequencing cytolethal distending toxin V (*cdt-V*) operon and its flanking regions in an atypical bovine *Escherichia coli* O157:H43 strain**, Zoonoses and Public Health, 59 (Issue Supplement s1). 46, 2012.

Sváb D., Horváth B., Szűcs A., Maróti G., Tóth I. 2013. **Draft genome sequence of a bovine *Escherichia coli* O157:H43 strain representing a novel genotype**, Abstracts of the 4th Central European Forum for Microbiology, October 16-18, Keszthely, Hungary. Acta Microbiol Immunol Hung. 60 (Suppl. 1). 235, 2013.

Publications not related to the PhD dissertation

Palatinszky M., Nikolausz M., Sváb D., Márialigeti K.: **Preferential ligation during TA-cloning of multitemplate PCR products – a factor causing bias in a microbial community structure analysis**, J. Microb. Methods, 85. 131-136, 2011. (IF: 2,086)

Acknowledgments

I would like to thank my supervisor, **István Tóth**, for helping me with his knowledge, experience, continuous encouragement and a huge amount of patience during the work and the writing of the dissertation.

I am also grateful to **Béla Nagy**, who supported and encouraged me from the beginning of my work, and gave me useful advice on many occasions.

I would also like to thank all the members of our research group, especially **Márta Puruczki** and **Ama Szmolka** for their help in many professional as well as technical matters which were essential for the completion of my work.

I am also thankful for every colleague at the Institute for Veterinary Medical Research who gave me professional or technical assistance.

I would like to thank all my co-authors for their contributions to the experiments and the preparation of the manuscripts, especially **Gergely Maróti** and **Balázs Horváth**, who had a key part in the sequencing projects and the bioinformatics analysis.

Finally, I would like to thank my family for their enduring support and encouragement over these years.

Financial support for our work was provided mainly by OTKA grant number K 81252 and partly by ERC AdG “SYMBIOTICS.”