Infection, Genetics and Evolution 46 (2016) 110–117

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Cytolethal distending toxin producing *Escherichia coli* O157:H43 strain T22 represents a novel evolutionary lineage within the O157 serogroup

Domonkos Sváb [a], Balázs Bálint [b], Gergely Maróti [c], István Tóth [a,*]

[a] Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary
[b] Seqomics Biotechnology Ltd., Mórahalom, Hungary
[c] Institute of Biochemistry, Biological Research Centre, Hungarian Academy of Sciences, Szeged, Hungary

ARTICLE INFO

Article history:
Received 24 June 2016
Received in revised form 14 October 2016
Accepted 3 November 2016
Available online 05 November 2016

Keywords:
Escherichia
O157:H43
CDT
LT-II
Comparative genomics
Phylogeny

ABSTRACT

Enterohemorrhagic *Escherichia coli* (EHEC) O157:H7/NM strains are significant foodborne pathogens intensively studied, while other sero- and pathotypes of the O157 serogroup only began to receive more attention. Here we report the first genome sequence of a cytolethal distending toxin (CDT-V) producing *E. coli* O157:H43 strain (T22) isolated from cattle. The genome consists of a 4.9 Mb chromosome assembled into three contigs and one plasmid of 82.4 kb. Comparative genomic investigations conducted with the core genomes of representative *E. coli* strains in GenBank ($n = 62$) confirmed the separation of T22 from the EHEC and enteropathogenic (EPEC) O157 lineages. Gene content based pangenome analysis revealed as many as 261 T22-specific coding sequences without orthologs in EDL933 EHEC O157 prototypic and two phylogenetically related commensal *E. coli* strains. The genome sequence revealed 10 prophage-like regions which harbor several virulence-associated genes including *cdt* and heat-labile enterotoxin (LT-II) encoding operons. Our results indicate that the evolutionary path of T22 is largely independent from that of EHEC and EPEC O157:H7/NM strains. Thus, the CDT-producing T22 *E. coli* O157:H43 strain represents a unique lineage of *E. coli* O157.

1. Introduction

Enterohemorrhagic *Escherichia coli* (EHEC) O157:H7/NM strains are significant food-borne zoonotic pathogens causing hemorrhagic colitis (HC) and haemolytic uraemic syndrome (HUS) in human (reviewed by Kaper et al., 2004). Typical EHEC strains produce a set of key virulence factors: one or more types of Shiga toxin (Stx) encoded by lambdoid prophages and intimin adhesin encoded by *eae* gene that is part of LEE pathogenicity island (reviewed by Caprioli et al., 2005). Besides the O157:H7/NM serotypes, EHEC strains have been reported from various other serotypes, as well as members of the O157 serogroup which also represent other pathotypes besides EHEC. *E. coli* O157 strains harboring *eae* gene but no *stx* were categorized as enteropathogenic *E. coli* (EPEC), such strains were isolated from human and animal sources (Stephan et al., 2004; Tóth et al., 2009). Additionally, atypical O157 strains, which carry none of the key virulence genes were reported (Sváb et al., 2013b; Tóth et al., 2009).

Because of the significance of EHEC and EPEC *E. coli* O157:H7/NM strains as pathogens, several strains had their whole genome sequences published (Eppinger et al., 2013; Wyrsch et al., 2015). However, less attention has been devoted to atypical O157 strains as well as strains with further serotypes exhibiting different H antigens (Hazen et al., 2013).

Cattle has been considered as a reservoir of Shiga toxin-producing *E. coli* (STEC) strains for a long time (Gyles, 2007). Earlier we reported atypical non-sorbitol fermenting *E. coli* O157 strains with H antigens different from H7 isolated from the milk of healthy cattle, several of these strains produced cytolethal distending toxin (CDT-V; Tóth et al., 2009; Taieb et al., 2015). These findings indicated that similar O157 strains could be persistent in the animal host, and the constant exchange of mobile genetic elements harboring virulence genes may give rise to new combinations of virulence traits. The emergence of novel pathotypes is a permanent potential epidemiological risk, as it was experienced recently in Germany where the largest outbreak of STEC was caused by recombinant enteroaggregative-heamorrhagic *E. coli* O104:H4 (EAHEC) strains (Ahmed et al., 2012).

Earlier, we reported the draft genome of a CDT-V producing atypical *E. coli* O157:H43 strain (T22, Sváb et al., 2013b). In the current study, our goals were to determine the complete genome of T22 and to investigate its phylogenetic relation to publicly available *E. coli* genomes. Comparative genomics revealed the unique genotype and phylogenetic position of T22 among *E. coli* strains.

* Corresponding author at: Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, H-1143 Budapest, Hungária krt. 21., Hungary.
E-mail addresses: svab.domonkos@agrar.mta.hu (D. Sváb), balintb@seqomics.hu (B. Bálint), marotig@brc.hu (G. Maróti), toth.istvan@agrar.mta.hu (I. Tóth).

http://dx.doi.org/10.1016/j.meegid.2016.11.003
1567-1348/

2. Materials and methods

2.1. Bacterial strain

E. coli strain O157:H43 T22 was isolated from the milk of healthy cattle on a Hungarian dairy farm. See Tóth et al. (2009) for further procedure information.

2.2. Whole genome sequencing

Genomic DNA of *E. coli* T22 was isolated with GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich) according to the manufacturer's instructions. Clone library was generated using Illumina Nextera Mate Pair Kit (Cat.Num.: FC-132-1001) per the manufacturer's instructions. Re-sequencing was performed on an Illumina MiSeq machine using V2 sequencing chemistry. Mate-paired reads were pre-processed following the manufacturer's recommendations (Data Processing of Nextera® Mate Pair Reads on Illumina Sequencing Platforms).

2.3. Assembly, annotation and sequence homology search

De novo assembly was performed with CLC Genomic Workbench 8.5.1 (CLC Bio) with contigs being subsequently arranged into scaffolds using SSPACE 3.0 (Boetzer et al., 2011). Gaps in scaffolds were closed with Spades v 3.1.1 (Bankevich et al., 2012) together with an in-house R script (unpublished results).

Annotation was performed with NCBI PGAAP.

Prophage regions in the genome were identified using PHAge Search Tool (PHAST, Zhou et al., 2011), and their automated annotation was manually curated. Sequence similarity searches were carried out using the publicly available tools of NCBI. Codon usage proportions were determined with the Cusp algorithm from the European Molecular Biology Open Software Suit (EMBOSS) package (Rice et al., 2000). The presence of virulence and antimicrobial resistance genes were also checked with Virulence Finder and ResFinder (Kleinheinz et al., 2014). Genes encoding secretion systems and associated effectors were investigated with the Effective database (Jehl et al., 2011).

2.4. Pangenome analysis

Complete genomes ($n = 62$) together with any associated plasmid sequences for representative *E. coli* and *Shigella* strains were downloaded from NCBI to assess the phylogenetic position of T22. Downloaded genomes as well as the T22 contigs were annotated using Prokka 1.10 (Seemann, 2014) to ensure identical gene prediction settings. Get homologues (Contreras-Monera and Vinuesa, 2013) was then applied on the annoted genbank files to identify clusters of orthologous sequences using the COGS clustering algorithm. The obtained pangenome matrix allowed gene content comparison between T22, two closely related commensal strains (IAI1, SE11) as well as the prototypic EHEC O157:H7 EDL933 strain (GenBank no. CU928160.2, AP009240.1and AE005174.2, respectively).

2.5. Phylogenetic analysis of the core genes

Get homologues (Contreras-Monera and Vinuesa, 2013) was used to compute the list of protein coding genes present in all strains (*i.e.* core genes). Clusters with inparalogues were excluded. Core gene DNA sequences were concatenated in the exact same order for each analyzed genome. Concatenated DNA sequences were then aligned using MAFFT (Tabei et al., 2008) with 100 bootstrapped alignment versions concomitantly generated with the "fseqboot" program from the EMBOSS package (Rice et al., 2000). Maximum likelihood phylogenetic trees were calculated from the original as well as from the bootstrapped alignment files using FastTree (Price et al., 2010). Support values based on 100 bootstrap replicates were added with "CompareToBootstrap.pl" script from the FastTree software package.

3. Results and discussion

3.1. Basic features of the E. coli T22 genome

The genome of *E. coli* O157:H43 strain T22 consists of a 4.9 Mb chromosome, and one circular plasmid (pT22) of 82.4 kb (GenBank accession number: LYNF00000000). The general features of the T22 genome are presented in Table 1. The chromosomal architecture of T22 is very similar to a number of other *E. coli* strains including pathogenic, commensals and K-12 laboratory strains. Another important feature of the chromosome is the high proportion (almost 7%) of prophage sequences. The genome of T22 is assembled into three contigs containing 4687 coding sequences (CDS), 75 tRNA and 24 rRNA genes. In addition, the plasmid contains 79 protein coding genes. The present resequencing helped to finalize the previously published draft genome data (Sváb et al., 2013b) regarding the sizes of the chromosome and the plasmid of T22 (Table 1).

The plasmid termed pT22 belongs to the F replicon type, it harbors two genes of a multi-gene iron-di-citrate transport protein system, and carries a *vapB-vapC* toxin-antitoxin system.

3.2. Prophage regions and associated virulence genes

Because of the significant role of prophages as vectors of virulence genes and their dissemination among EHEC O157 strains, we have paid special attention to the prophages harbored by *E. coli* T22. PHAST search and the annotation of the genome revealed that the chromosome of *E. coli* T22 contains 10 prophage regions (Table 2). Similarly to the prototypic EHEC O157:H7 strains (Hayashi et al., 2001; Perna et al., 2001), a great number of prophages are identified and some of them also carry genes with virulence potential. The manual curation of the PHAST results revealed that prophages account for 339 kb in the genome, corresponding to almost 7% of the total chromosome length. Regarding the two EHEC O157:H7 prototypic strains their genomes contain higher number of prophages: Sakai carries 12.2% (Hayashi et al., 2001) and EDL933 11% (Perna et al., 2002) prophage sequences. The size of these individual prophages varies between 13 and 52 kb and they are dispersed in the genome. The GC content of the prophages is between 45.5% and 52.6% and no codon usage bias can be detected when compared to the whole genome. This indicates adaptation of these prophage elements to the T22 host. The main characteristics of the identified prophages are summarized in Table 2, and Fig. 1 shows their schematic representations. We termed the prophages from T22 pp1 to T22 pp10. The architectures of the T22 prophages are different, three of them are lambdoid, three prophages are P2-like, and the remaining four are classified according to the closest homologue sequences (Table 2).

Homologies and their overall structure classify prophages T22pp4, T22pp6 and T22pp7 as P2-like prophages. T22pp6 harbors the operon encoding cytolethal distending toxin V, and has been characterized earlier (Sváb et al., 2013a).

Table 1
Summary of main genomic features of *E. coli* T22.

	Chromosome	Plasmid
Size (bp)	4,875,355	82,444
Predicted coding CDSs	4687	79
GC %	51	49
Coding regions %	88.5	70
Average CDS length (bp)	905	730
tRNA	75	0
rRNA	24	0

Table 2
Main characteristics of predicted prophage regions of *E. coli* T22.

Designation	Contig	Start	End	Length (nt)	GC %	Assigned type	Upstream flanking gene	Downstream flanking gene	Virulence and fitness related genes	Integrase	Repressor
T22pp1	1	1	32,065	32,065	49.4	PhiP27-like	Unknown	tRNA 5′ methoxy-uridine synthase	None identifed	Present	
T22pp2	1	385,293	398,890	13,598	48.8	none	Proline-tRNA	type I secretion system	None identifed	Present	
T22pp3	1	1,480,783	1,531,940	51,158	49.3	*Shigella* serotype converting	Met-tRNA	tmRNA binding protein SmpB	None identifed	Present	cI and cro
T22pp4	1	1,543,880	1,574,799	30,920	52	P2-like	Ribosomal DNA	Yfi operon	None identifed	Present	putative cI
T22pp5	1	1,703,290	1,717,745	14,456	49	*P4*-like	Sugar efflux transporter B	Ser-tRNA	none identifed	Unidentified	
T22pp6	1	2,288,961	2,320,124	31,164	52.6	P2-like	fieF	cpxP	CDT-V	Present	c and cox
T22pp7	1	3,712,670	3,744,641	31,972	52.2	P2-like	TetR transcriptional regulator	Trk potassium-channel	None identifed	Putative int-like protein	cI and cII regulator
T22pp8	1	3,888,837	3,928,433	39,597	49.9	Lambdoid	Putative TEGT family transport protein	Ser-tRNA	Lipoprotein Bor	Unidentified	
T22pp9	2	350,924	396,041	45,118	45.5	Lambdoid	Cytosine tRNA-synthetase	Unknown	Lom-like protein, heat-labile enterotoxin IIc1, *E. coli* artAB-like AB toxin	Present	
T22pp10	3	212,066	258,765	46,700	50.4	Lambdoid	Putative transport protein	Starvation sensing protein rspB	Lom-like protein	Unidentified	

An important difference between these P2-likeprophages is that while T22pp6 has a P2-type regulatory switch consisting of *c* and *cox* genes, prophages T22pp4 and T22pp7 contain the 186-type switch consisting of *cI* and *cII* regulatory genes indicating that they appear to be more closely related to P2-like prophages of Salmonella origin, although P2-like phages are generally hypothesized to be host-specific (Nilsson et al., 2011).

Prophages T22pp8, T22pp9 and T22 pp10 show partial homologies to lambdoid prophages, particularly to Sp8 of the EHEC O157:H7 Sakai strain (Hayashi et al., 2001), in the case of prophage T22pp8. T22pp9 contains a relatively conserved region, which also occurs in Sp3 and Sp10 of the Sakai strain.

T22pp9 also carries the gene cluster with genes encoding heat-labile enterotoxin type IIc1 (LT-IIc1). The region containing the toxin genes is situated downstream of the *Q* antiterminator gene and also contains the *artAB* operon encoding the ADP-ribosyltransferase toxin homologue or pertussis-like toxin, first identified in *Salmonella typhimurium* (Saitoh et al., 2005), and later in *E. coli* strains, where it was termed *ealAB*, for *E. coli artAB*-like AB toxin (Jobling, 2016). These two toxin-encoding operons are separated by three phage genes, encoding a putative membrane protein, a holin, and a hypothetical protein, respectively. It has already been proposed that LT-II is carried by prophages (Jobling and Holmes, 2012) and in the case of LT-IIa, a whole prophage genome was determined, and the carriage of a pertussis-like toxin gene cluster was also reported as prophage Rac-SA53 (Jobling, 2016). In T22pp9 the region carrying the LT-II toxin genes are very similar to that found in prophage Rac-SA53, however, most of the remaining structural regions of T22pp9 are significantly different. The *lt-IIc1* and *ealAB* clusters harbored by T22 are 100% identical to those carried by *E. coli* strain NADC1036 (GenBank KU052040; Jobling, 2016).

Besides the above toxin genes, T22pp9 and alsoT22 pp10 carry an attachment invasion protein precursor gene, encoding the *lom*-like protein, which is thought to have a role in adhesion (Vica Pacheco et al., 1997), however, no recent studies have explored this function in details.

Serum resistance associated lipoprotein Bor (Barondess and Beckwith, 1995) encoding gene is present inT22pp8. This putative virulence factor is encoded by multiple lambdoid prophages found in several pathogenic *E. coli* strains, most notably in Stx-converting phages of EHEC strains (Asadulghani et al., 2009), T22pp1 is integrated into the *yecE-yecD* site, which in some STEC strains contains prophage PhiP27, harboring the *stx2e* operon (Muniesa et al., 2000; Recktenwald and Schmidt, 2002). Within T22pp1 there are genetic stretches showing strong homology to parts encoding replication and packaging functions of PhiP27, suggesting evolutionary relationship between these prophages.

T22pp2, being only 11 kb in size, shows remarkable mosaic structure, but only low level homologies to known prophages. T22pp2 is integrated between a prolin tRNA gene and the operon encoding a type I secretion system. There are no reported examples of this insertion up to date, but the undisrupted region (prolin tRNA and type I secretion system in tandem) can be observed in the genome of several commensal strains.

The T22 prophages are integrated at different sites scattered across the genome, half of them are associated with tRNA genes (Table 2), which is in harmony with the findings of Bobay et al. (2013), who reported that prophages are frequently integrated next to tRNA genes.

The typical integration sites of Stx phages (*wrbA*, *yehV*, *sbcB*, *argW*) are intact in T22, with the exception of *yecN-yecD* containing T22pp1 noted above, providing potential integration sites for Stx phages.

3.3. Additional virulence related genes

Although the prophage pool carried by T22 differs greatly from the prophages carried by typical EHEC O157:H7 strains, these regions also harbor several different (*cdt*, *lt-II*, *lpf*) and identical (*bor*, *lom*) virulence genes (see Section 3.2). Additional virulence genes are also present in the chromosome, representing fimbriae and secretion systems (see below Sections 3.3.1–3.3.2).

3.3.1. Fimbriae

The initial PCR investigation of *E. coli* T22 showed that the strain harbors long polar fimbria (Lpf) genes (Tóth et al., 2009), this surface structure described by Torres et al. (Torres et al., 2002) was found to be harbored by a wide variety of pathogenic *E. coli* in the past decade, and proved to have adhesive functions in EHEC (Lloyd et al., 2012). Subtyping showed that the *lpf* operon harbored by T22 is of allelic type 2-1 (Sváb and Tóth, 2012), which is a widespread variant, but

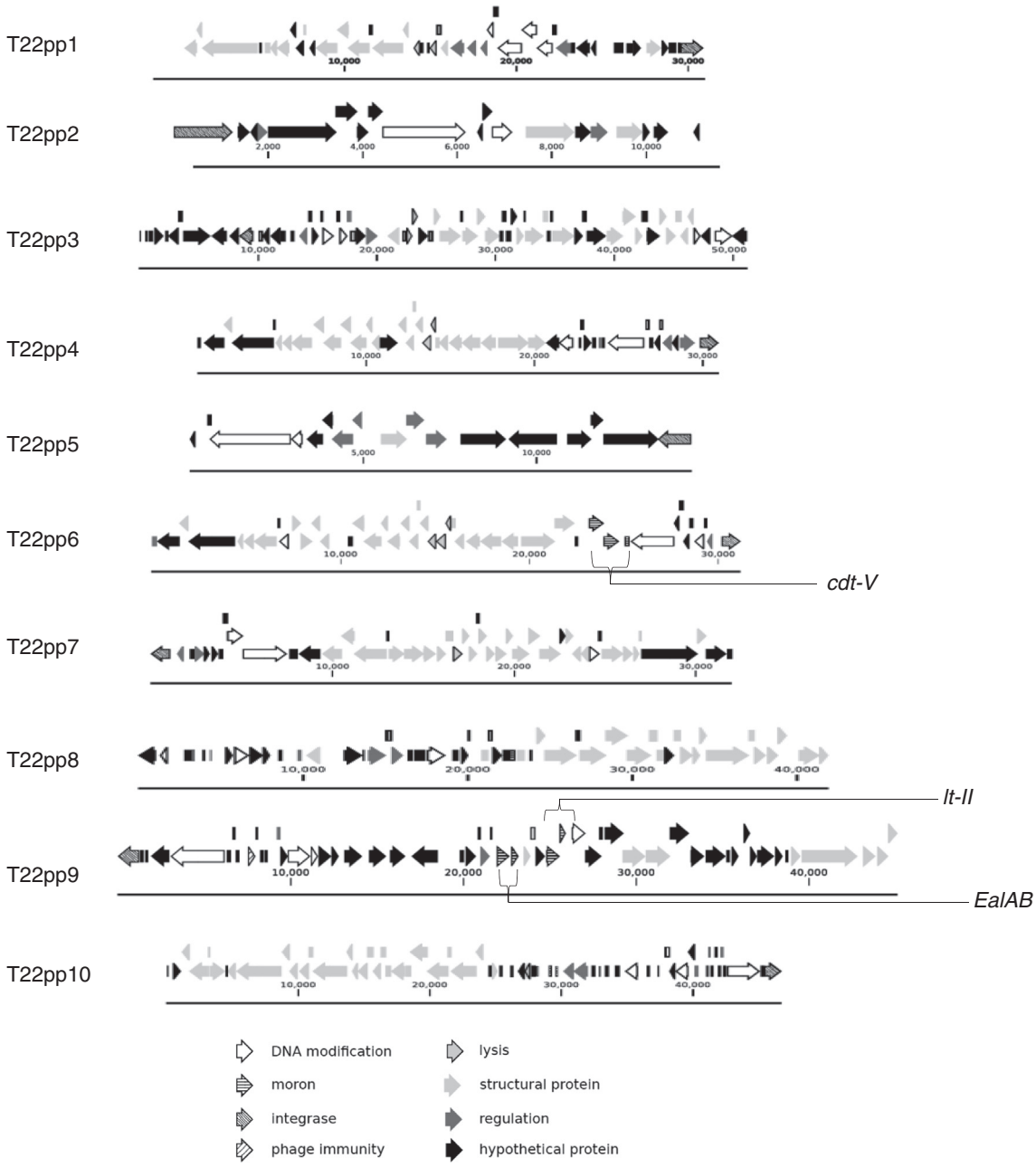

Fig. 1. Schematic representation of prophage regions found in the genome of T22. Arrows filled with the same pattern represent genes of related functions. Note that the size scale of each prophage is in nucleotides and was created individually. The identified toxin-encoding gene clusters are highlighted by brackets. *lt-II:* heat-labile enterotoxin type II, *cdt-V:* cytolethal distending toxin type V, *ealAB*: *Escherichia coli ArtAB*-like toxin.

not characteristic for typical EHEC or EPEC O157 strains (Torres et al., 2009).

E. coli T22 also carries the full operon of type 1 fimbriae encoded by the *fim* gene cluster, which is also a widespread virulence factor of pathogenic *E. coli*, especially uropathogenic *E. coli* (UPEC), where it plays a role in adherence to the urogenital epithelium, and can also be found in several commensal strains (Martinez et al., 2000).

Both fimbrial operons are integrated into their usual site, which is between the *glm* and *pst* gene clusters for *lpf2-1*, and between the *yhj* and *gnt* gene clusters for the *fim* operon, respectively.

3.3.2. Secretion systems

The T22 genome harbors a full type VI secretion system encoding gene cluster. This structure is considered to be a remnant of a bacteriophage tail, utilized by *E. coli* in the transport of various effectors (Leiman et al., 2009). This gene cluster carried by T22 is most similar to that of ETEC E24377A of human origin and several other pathogenic *E. coli* strains. This variant was reported and characterized earlier, termed as subtype 2 (T6SS2) in APEC strains (Ma et al., 2013).

We also identified a truncated version of the type 2 of the type III secretions system (ETT2; (Makino et al., 2003). The *yqe*, *yge*, *epr* and *epa* gene clusters are present but the whole *eiv* gene cluster is missing albeit its flanking pseudogenes are present. (Fig. 2). The above gene clusters are integrated between a gene encoding a serine transporter and a glycine tRNA gene. This putative secretion system is widespread in pathogenic *E. coli*, and is thought to have regulatory functions in virulence in general (reviewed by Zhou et al., 2014).

T22 fulfils all the criteria of a CDT-producing *E. coli* (Tóth and Sváb, 2014) as demonstrated in our previous study (Taieb et al., 2015). CDT-V from T22 caused characteristic morphological changes of HeLa cells

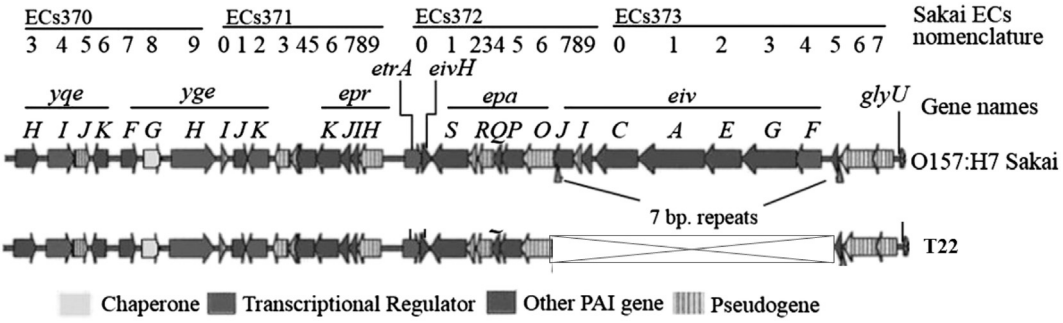

Fig. 2. Comparison of the type 2 type III secretions system (ETT2) gene cluster in T22 with that found in *Escherichia coli* O157:H7 Sakai. The figure is adapted from the work of Zhou et al. (2014) reviewing the ETT2 secretion system, focusing on the gene cluster found in the *E. coli* O157:H7 Sakai genome. Arrows filled with the same pattern represent genes of the same cluster.

in tissue culture assays. Cell cycle arrest of eukaryotic cells at G2-M2 transition was demonstrated by flow cytometry. In addition, the CDT–induced phosphorylation of histone protein H2AX was observed by immunoblot analysis revealing double-stranded DNA damage in the cells (Taieb et al., 2015). Recent studies suggest the distinction of cytolethal distending toxin producing *E. coli* as a pathotype on its own right, abbreviated as 'CTEC' (Hinenoya et al., 2009, 2014). While not being isolated from human source, and thus not having direct evidence for its pathogenicity, *E. coli* T22 can be categorized as CTEC, while also carrying the gene cluster encoding LT-II and the pertussis-like toxin.

3.4. CRISPR regions

Clustered regularly interspaced palindromic repeats (CRISPR) and the proteins encoded by the CRISPR-associated genes (Cas) are utilized by various archaea and bacteria (including *E. coli*) as an adaptive immune system against foreign genetic material reviewed in Bondy-Denomy and Davidson (2014). All four CRISPR regions (CRISPR1–4) identified up to date in *E. coli* are present. A summary of these regions is shown in Table 3.Thus the T22 genome contains a presumably functional CRISPR-Cas system. The CRISPR1 is coupled to the gene cluster encoding the Cas genes, which is most similar to the gene cluster carried by ETEC strain E24377A (Díez-Villaseñor et al., 2010).

The spacers in the CRISPR1 array of T22 are identical to those carried by strain hu24 and k8, isolated from human source (GenBank KF707537.1 and KF707515.1; (Sheludchenko et al., 2015). However, when compared to the repeat regions of typical O157 EHEC strains (Delannoy et al., 2012), no common spacers can be observed.

The CRISPR2 region bears an overall similarity to that found in several K-12 and commensal strains. Comparison with the STEC CRISPR sequence collection deposited by Yin et al. (2013) revealed that the first and fourth spacer of T22 CRISPR2 are shared with the corresponding region of strain 0-0846, but no other matching spacers can be found. On the other hand all spacers of T22 CRISPR1 and CRISPR2 can be found in the corresponding regions of Shiga toxin-producing (STEC) O113:H21 strains determined by Feng et al. (2014), and in several *E. coli* reference (ECOR) strains, albeit in different arrangements. Fig. 3 shows a schematic representation of CRISPR1 and CRISPR2.

It is interesting to note that according to Yin et al., a conservation of CRISPR sequences could be observed between O157 strains carrying the same H antigen (Yin et al., 2013). Recently the draft genome sequence of porcine ETEC O157:H43 strain DEC7A became available (GenBank AIGA01000043.1, Hazen et al., 2012). However, when comparing the CRISPR1 and CRISPR2 regions of T22 to those of DEC7A, no matching spacers are found.

The CRISPR3 and CRISPR4 regions, in harmony with earlier findings (Touchon and Rocha, 2010), forms a combined, short, and apparently conserved array, which shows 100% identity to corresponding regions carried by more than 40 *E. coli* strains according to GenBank, and without containing any spacers (Toro et al., 2014).

Table 3
Main characteristics of CRISPR regions in *E. coli* T22.

Region	Integration site	Number of spacers	Length of spacers (nt)
CRISPR1	Cas2 and alkaline phosphatase	12	32[a]
CRISPR2	*ygcF*, *ygcE*	26	32
CRISPR3	*clpS*, *aat*	0	N/A
CRISPR4	*clpS*, *aat*	0	N/A

[a] One spacer in CRISPR1 is 33 nt long.

3.5. Phylogenetic relations

To investigate the phylogenetic relationships of *E. coli* T22 to other pathogenic and non-pathogenic *E. coli* strains, we selected 62 publicly available whole genomes and compared them by core genome phylogeny. The selected strains represent all the main intestinal pathotypes as well as extraintestinal (ExPEC), uropathogenic (UPEC) and commensal strains. Representatives of the four species of *Shigella* were also included. The accession numbers of the genomes are given in Table A1. All the common conserved genes present in these *E. coli* and *Shigella* genomes were collected. DNA sequence for as many as 1862 common genes were identified and used in genome comparison analysis (Fig. 4).

The core genome phylogeny placed T22 closest to E24377A ETEC strain of O139:H28 serotype, as well as the EHEC O26:H11 and O111:NM strains. The EHEC O103:H2 strain was placed on a neighboring branch. These three EHEC strains (designations 11,368, 11,128 and 120,009, respectively) were isolated from sporadic cases of diarrhea (Ogura et al., 2007), while ETEC E24377A was isolated from drinking water (Tamhankar et al., 2015). The typical EHEC O157:H7 strains and their ancestors, the O55:H7 strains (Feng et al., 2007) form a separate group, as well as the recently emerged O104:H4 enteroaggregative-hemorrhagic (EAHEC) strains. These results indicate that T22 belongs to a separate lineage of intestinal pathogenic strains, divergent from the typical O157 EHEC and also from the recently emerged O104:H4 EAHEC strains. Our phylogenetic analysis also showed that commensal strains like IAI1, SE15 and SMS-3-5 are widely distributed all along the tree. Extraintestinal strains are separate from all the above-mentioned clusters, but they are insterspersed with EPEC and other intestinal and commensal strains. The *Shigella dysenteriae* Sd197 strain was placed close to the EHEC O157:H7 strains, but far from strains of other *Shigella* species (Fig. 4).

Complementary to the phylogenetic relations a further, gene-content based pangenome analysis was performed with the genomes of *E. coli* T22, the commensal *E. coli* strains IAI1, SE11 and the prototypic EHEC O157:H7 EDL933 strain (GenBank no. CU928160.2, AP009240.1, and AE005174.2, respectively). The former two were chosen because on the whole-genome sequence level they are closest to T22 among the commensal strains, and the EDL933strain was included as prototypic EHEC O157 strain. These analyses revealed 3,599CDSs (82% of the T22

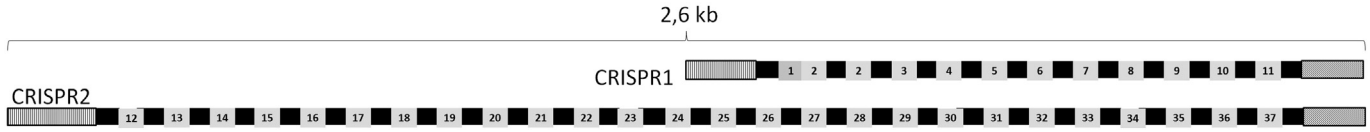

Fig. 3. Genetic structure of the CRISPR1 and CRISPR2 (clustered regularly interspaced repeat) regions of T22. Numbered boxes represent the individual spacers, black boxes represent the repeat sequences, the textured bars represent the leader and closing sequences, respectively. Note that in CRISPR1, spacer 2 is present in two copies.

CDSs) that are common to all four compared *E. coli* genomes, while 261 CDSs (6%) are T22-specific (Fig. 5. and Table A2). T22-specific genes include the gene cluster encoding CDT, LT-II and EalAB, as well as several plasmid and prophage-related structural and regulatory genes in the chromosome. One gene of the O-antigen biosynthesis cluster is also included, which further supports the idea of T22 representing a distinct lineage from other O157 strains, while this gene shows high identity percentage to that carried by the strain *E. coli* PV00-24, and of the O157:H43 serotype (GenBank AB602253.1). This finding also confirms the notion that differences in the O157 biosynthesis genes indicate the separation of lineages within the serogroup (Iguchi et al., 2011). 34 of the 261 unique genes are found on the plasmid pT22. These include the iron di-citrate transport system, the F-pilin as well as the *tra* and *trb* gene cluster related to the conjugative transfer. The predicted functions and closest homologues of the unique genes are summarized in Table A2.

All these results indicate that *E. coli* T22 represents a distinct evolutionary path from the typical STEC and EHEC members of the O157 serogroup, and being divergent from enterotoxigenic *E. coli* (ETEC) strains of O157:H43 serotype, although there is considerably fewer sequence data available for ETEC O157 strains. In a recent study by Sanjar et al. (2015) phylogenetic analysis placed the genomes of O157:non-H7 strains, including T22, apart from typical EHEC and EPEC strains of the serogroup, further suggesting its separate lineage. In harmony with our results Sanjar et al. reported that T22 clustered together with ETEC strain E24377A, commensal *E. coli* strain IAI1, and also with

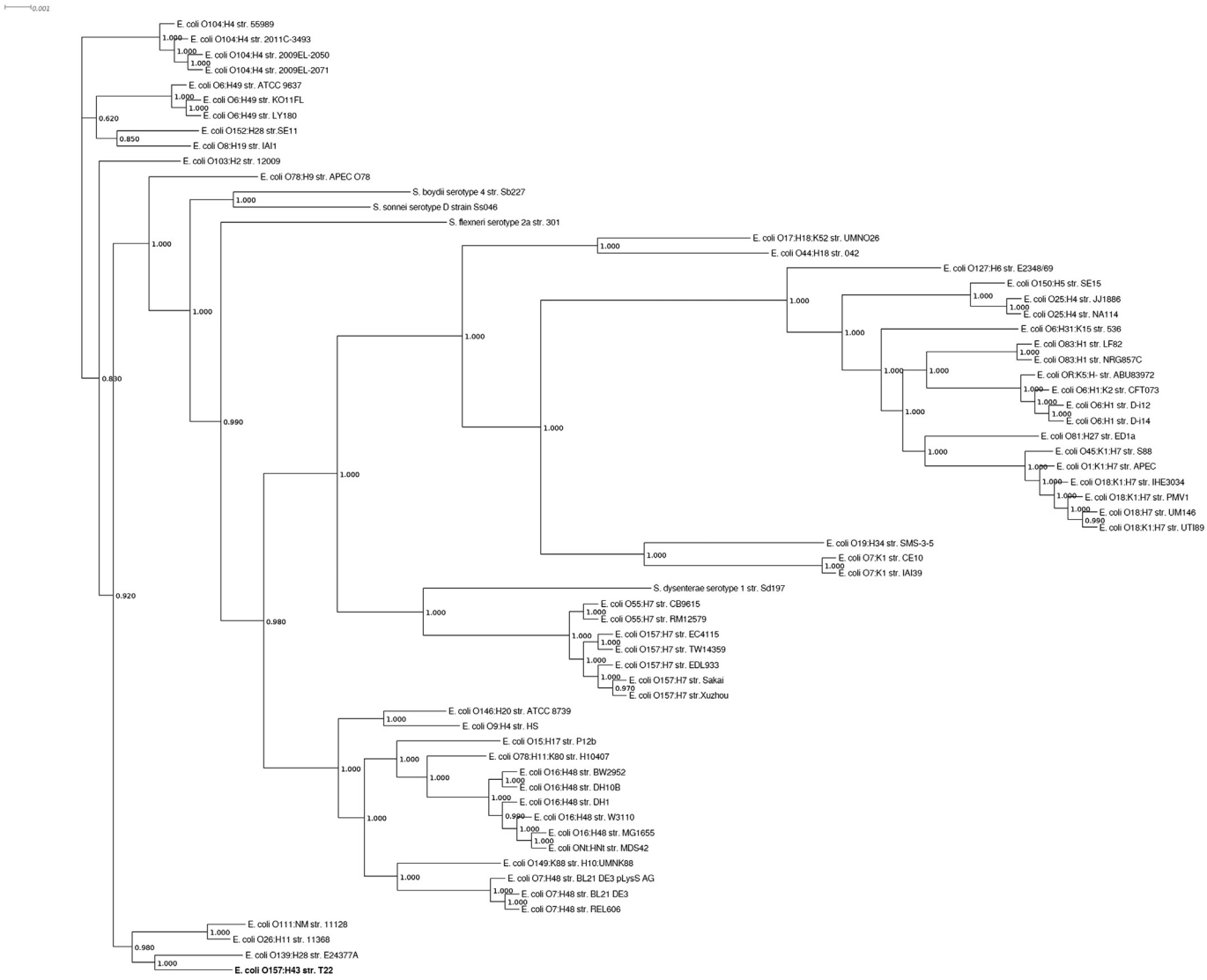

Fig. 4. Core genome phylogeny of T22 and 58 *E. coli* as well as 4 *Shigella* strains with whole genomes available in GenBank. Consensus neighbour-joining supertree based on the sequences of 1862 common genes.

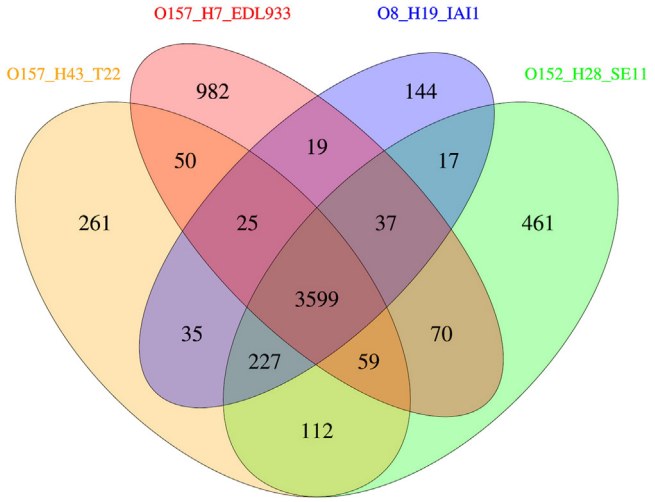

Fig. 5. Comparison of the genetic content of the sequenced *E. coli* O157:H43 strain T22 with the EHEC O157:H7 EDL933 strain and the commensal *E. coli* SE11 and IAI1 strains. The 4 strains share 3599 common genes and 261 specific genes were identified in the T22 genome. The T22 specific genes are listed in Table A2.

EHEC strains 11128, 11368 and 120,009. Together with the results of the pangenome analysis, these data strongly support the idea of T22 representing a unique and so far uncharacterized lineage of pathogenic *E. coli* (Figs. 4 and 5).

Since T22 represents a novel patho- and genotype among *E. coli* O157 strains it is necessary to summarize the relation of T22 to prototypic EHEC O157 strains regarding their phylogenetic markers. Earlier a step-wise evolution model was devised for the emergence of O157:H7/NM EHEC strains (Feng et al., 2007). In this model, the inability to ferment sorbitol is considered a key marker of typical EHEC O157:H7 strains. Interestingly, T22 also shows the non-sorbitol-fermenting (NSF) phenotype (Tóth et al., 2009). In a more recent study, conducted with 400 O157 strains including nine of the O157:H43 strains, pulsed field gel electrophoresis (PFGE) patterns suggested that the O157 serogroup became divergent according to the different H types (Rump et al., 2015). Out of the nine O157:H43 strains investigated by Rump et al. (2015) only two were NSF, and one of theme lacked all EHEC-specific marker genes. Similarly, none of the 13 EHEC-specific marker genes are present in the genome of T22. Because the study was limited to these 13 genes, no further genetic comparison is possible at present between T22 and the O157:H43 strains investigated by Rump et al. (2015).

3.6. Conclusions

EHEC O157:H7/NM strains are well characterized dangerous zoonotic pathogens, but less information is available on *E. coli* O157 strains with different virulence genes and H antigen. Here we present the first comprehensive genomic analysis of a CDT-producing, Stx- and intimin-negative *E. coli* O157 strain, and the first whole genome of the O157:H43 serotype. Our results show that besides CDT, T22 carries an array of additional virulence genes, including those encoding LT-II toxin as well as the type1 fimbria and Lpf as potential adhesins. The genome contains large amount of prophages, one of them carries the *cdt* and another the LT-II encoding genes. To our knowledge, no *E. coli* O157 strain with similar genotype and virulence array has been described up to date in detail.

Regarding the phylogenetic relations of T22, our results indicate that it has a unique position among pathogenic and wild-type commensal *E. coli* strains, representing a separate lineage from the typical EHEC and EPEC O157 strains. The lack of key EHEC virulence genes and the integrity of their typical integration hot spots further support this notion. The unique evolutionary background and genotype of T22 also calls the attention to the continuing emergence of new virulence gene combinations among *E. coli* of serogroup O157.

Acknowledgments

This study was supported by the Hungarian Research Fund (OTKA, grant no. K 81 252).

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.meegid.2016.11.003.

References

Ahmed, S.A., Awosika, J., Baldwin, C., Bishop-Lilly, K.A., Biswas, B., Broomall, S., Chain, P.S.G., Chertkov, O., Chokoshvili, O., Coyne, S., Davenport, K., Detter, J.C., Dorman, W., Erkkila, T.H., Folster, J.P., Frey, K.G., George, M., Gleasner, C., Henry, M., Hill, K.K., Hubbard, K., Insalaco, J., Johnson, S., Kitzmiller, A., Krepps, M., Lo, C.-C., Luu, T., McNew, L.A., Minogue, T., Munk, C.A., Osborne, B., Patel, M., Reitenga, K.G., Rosenzweig, C.N., Shea, A., Shen, X., Strockbine, N., Tarr, C., Teshima, H., van Gieson, E., Verratti, K., Wolcott, M., Xie, G., Sozhamannan, S., Gibbons, H.S., Threat Characterization Consortium, 2012. Genomic comparison of *Escherichia coli* O104:H4 isolates from 2009 and 2011 reveals plasmid, and prophage heterogeneity, including shiga toxin encoding phage stx2. PLoS One 7, e48228. http://dx.doi.org/10.1371/journal.pone.0048228.

Asadulghani, M., Ogura, Y., Ooka, T., Itoh, T., Sawaguchi, A., Iguchi, A., Nakayama, K., Hayashi, T., 2009. The defective prophage pool of *Escherichia coli* O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. PLoS Pathog. 5, e1000408. http://dx.doi.org/10.1371/journal.ppat.1000408.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. J. Comput. Mol. Cell Biol. 19:455–477. http://dx.doi.org/10.1089/cmb.2012.0021.

Barondess, J.J., Beckwith, J., 1995. bor gene of phage lambda, involved in serum resistance, encodes a widely conserved outer membrane lipoprotein. J. Bacteriol. 177, 1247–1253.

Bobay, L.-M., Rocha, E.P.C., Touchon, M., 2013. The adaptation of temperate bacteriophages to their host genomes. Mol. Biol. Evol. 30:737–751. http://dx.doi.org/10.1093/molbev/mss279.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., Pirovano, W., 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinforma. Oxf. Engl. 27:578–579. http://dx.doi.org/10.1093/bioinformatics/btq683.

Bondy-Denomy, J., Davidson, A.R., 2014. To acquire or resist: the complex biological effects of CRISPR-Cas systems. Trends Microbiol. 22:218–225. http://dx.doi.org/10.1016/j.tim.2014.01.007.

Caprioli, A., Morabito, S., Brugère, H., Oswald, E., 2005. Enterohaemorrhagic *Escherichia coli*: emerging issues on virulence and modes of transmission. Vet. Res. 36:289–311. http://dx.doi.org/10.1051/vetres:2005002.

Contreras-Monera, B., Vinuesa, P., 2013. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl. Environ. Microbiol. 79:7696–7701. http://dx.doi.org/10.1128/AEM.02411-13.

Delannoy, S., Beutin, L., Fach, P., 2012. Use of clustered regularly interspaced short palindromic repeat sequence polymorphisms for specific detection of enterohemorrhagic *Escherichia coli* strains of serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19, O145:H28, and O157:H7 by real-time PCR. J. Clin. Microbiol. 50:4035–4040. http://dx.doi.org/10.1128/JCM.02097-12.

Díez-Villaseñor, C., Almendros, C., García-Martínez, J., Mojica, F.J.M., 2010. Diversity of CRISPR loci in *Escherichia coli*. Microbiol. (Read. Engl.) 156:1351–1361. http://dx.doi.org/10.1099/mic.0.036046-0.

Eppinger, M., Daugherty, S., Agrawal, S., Galens, K., Sengamalay, N., Sadzewicz, L., Tallon, L., Cebula, T.A., Mammel, M.K., Feng, P., Soderlund, R., Tarr, P.I., Debroy, C., Dudley, E.G., Fraser, C.M., Ravel, J., 2013. Whole-genome draft sequences of 26 enterohemorrhagic *Escherichia coli* O157:H7 strains. Genome Announc. 1, e0013412. http://dx.doi.org/10.1128/genomeA.00134-12.

Feng, P.C.H., Delannoy, S., Lacher, D.W., Dos Santos, L.F., Beutin, L., Fach, P., Rivas, M., Hartland, E.L., Paton, A.W., Guth, B.E.C., 2014. Genetic diversity and virulence potential of shiga toxin-producing *Escherichia coli* O113:H21 strains isolated from clinical, environmental, and food sources. Appl. Environ. Microbiol. 80:4757–4763. http://dx.doi.org/10.1128/AEM.01182-14.

Feng, P.C.H., Monday, S.R., Lacher, D.W., Allison, L., Siitonen, A., Keys, C., Eklund, M., Nagano, H., Karch, H., Keen, J., Whittam, T.S., 2007. Genetic diversity among clonal lineages within *Escherichia coli* O157:H7 stepwise evolutionary model. Emerg. Infect. Dis. 13:1701–1706. http://dx.doi.org/10.3201/eid1311.070381.

Gyles, C.L., 2007. Shiga toxin-producing *Escherichia coli*: an overview. J. Anim. Sci. 85:E45–E62. http://dx.doi.org/10.2527/jas.2006-508.

Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M., Shinagawa, H., 2001. Complete genome sequence of enterohemorrhagic

Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes 8, 11–22.
Hazen, T.H., Sahl, J.W., Fraser, C.M., Donnenberg, M.S., Scheutz, F., Rasko, D.A., 2013. Draft genome sequences of three O157 enteropathogenic *Escherichia coli* isolates. Genome Announc. 1. http://dx.doi.org/10.1128/genomeA.00516-13.
Hazen, T.H., Sahl, J.W., Redman, J.C., Morris, C.R., Daugherty, S.C., Chibucos, M.C., Sengamalay, N.A., Fraser-Liggett, C.M., Steinsland, H., Whittam, T.S., Whittam, B., Manning, S.D., Rasko, D.A., 2012. Draft genome sequences of the diarrheagenic *Escherichia coli* collection. J. Bacteriol. 194:3026–3027. http://dx.doi.org/10.1128/JB.00426-12.
Hinenoya, A., Naigita, A., Ninomiya, K., Asakura, M., Shima, K., Seto, K., Tsukamoto, T., Ramamurthy, T., Faruque, S.M., Yamasaki, S., 2009. Prevalence and characteristics of cytolethal distending toxin-producing *Escherichia coli* from children with diarrhea in Japan. Microbiol. Immunol. 53, 206–215.
Hinenoya, A., Shima, K., Asakura, M., Nishimura, K., Tsukamoto, T., Ooka, T., Hayashi, T., Ramamurthy, T., Faruque, S.M., Yamasaki, S., 2014. Molecular characterization of cytolethal distending toxin gene-positive *Escherichia coli* from healthy cattle and swine in Nara. Japan. BMC Microbiol. 14:97. http://dx.doi.org/10.1186/1471-2180-14-97.
Iguchi, A., Shirai, H., Seto, K., Ooka, T., Ogura, Y., Hayashi, T., Osawa, K., Osawa, R., 2011. Wide distribution of O157-antigen biosynthesis gene clusters in *Escherichia coli*. PLoS One 6, e23250. http://dx.doi.org/10.1371/journal.pone.0023250.
Jehl, M.-A., Arnold, R., Rattei, T., 2011. Effective–a database of predicted secreted bacterial proteins. Nucleic Acids Res. 39:D591–D595. http://dx.doi.org/10.1093/nar/gkq1154.
Jobling, M.G., 2016. The chromosomal nature of LT-II enterotoxins solved: a lambdoid prophage encodes both LT-II and one of two novel pertussis-toxin-like toxin family members in type II enterotoxigenic *Escherichia coli*. Pathog. Dis. 74. http://dx.doi.org/10.1093/femspd/ftw001.
Jobling, M.G., Holmes, R.K., 2012. Type II heat-labile enterotoxins from 50 diverse *Escherichia coli* isolates belong almost exclusively to the LT-IIc family and may be prophage encoded. PLoS One 7, e29898. http://dx.doi.org/10.1371/journal.pone.0029898.
Kaper, J.B., Nataro, J.P., Mobley, H.L., 2004. Pathogenic *Escherichia coli*. Nat. Rev. Microbiol. 2:123–140. http://dx.doi.org/10.1038/nrmicro818.
Kleinheinz, K.A., Joensen, K.G., Larsen, M.V., 2014. Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and *E. coli* virulence genes in bacteriophage and prophage nucleotide sequences. Bacteriophage:4 http://dx.doi.org/10.4161/bact.27943.
Leiman, P.G., Basler, M., Ramagopal, U.A., Bonanno, J.B., Sauder, J.M., Pukatzki, S., Burley, S.K., Almo, S.C., Mekalanos, J.J., 2009. Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. Proc. Natl. Acad. Sci. U. S. A. 106:4154–4159. http://dx.doi.org/10.1073/pnas.0813360106.
Lloyd, S.J., Ritchie, J.M., Rojas-Lopez, M., Blumentritt, C.A., Popov, V.L., Greenwich, J.L., Waldor, M.K., Torres, A.G., 2012. A double, long polar fimbria mutant of *Escherichia coli* O157:H7 expresses Curli and exhibits reduced in vivo colonization. Infect. Immun. 80:914–920. http://dx.doi.org/10.1128/IAI.05945-11.
Ma, J., Sun, M., Bao, Y., Pan, Z., Zhang, W., Lu, C., Yao, H., 2013. Genetic diversity and features analysis of type VI secretion systems loci in avian pathogenic *Escherichia coli* by wide genomic scanning. Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis. 20:454–464. http://dx.doi.org/10.1016/j.meegid.2013.09.031.
Makino, S.-I., Tobe, T., Asakura, H., Watarai, M., Ikeda, T., Takeshi, K., Sasakawa, C., 2003. Distribution of the secondary type III secretion system locus found in enterohemorrhagic *Escherichia coli* O157:H7 isolates among Shiga toxin-producing *E. coli* strains. J. Clin. Microbiol. 41, 2341–2347.
Martinez, J.J., Mulvey, M.A., Schilling, J.D., Pinkner, J.S., Hultgren, S.J., 2000. Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. EMBO J. 19:2803–2812. http://dx.doi.org/10.1093/emboj/19.12.2803.
Muniesa, M., Recktenwald, J., Bielaszewska, M., Karch, H., Schmidt, H., 2000. Characterization of a shiga toxin 2e-converting bacteriophage from an *Escherichia coli* strain of human origin. Infect. Immun. 68, 4850–4855.
Nilsson, H., Cardoso-Palacios, C., Haggård-Ljungquist, E., Nilsson, A.S., 2011. Phylogenetic structure and evolution of regulatory genes and integrases of P2-like phages. Bacteriophage 1:207–218. http://dx.doi.org/10.4161/bact.1.4.18470.
Ogura, Y., Ooka, T., Asadulghani, T., Nougayréde, J., Kurokawa, K., Tashiro, K., Tobe, T., Nakayama, K., Kuhara, S., Oswald, E., Watanabe, H., Hayashi, T., 2007. Extensive genomic diversity and selective conservation of virulence determinants in enterohemorrhagic *Escherichia coli* strains of O157 and non-O157 serotypes. Genome Biol. 8:R138. http://dx.doi.org/10.1186/gb-2007-8-7-r138.
Perna, N.T., Glasner, J.D., Burland, V., Plunkett III, G., 2002. The genomes of *Escherichia coli* K-12 and pathogenic *E. coli*. In: Donnenberg, M.S. (Ed.), *Escherichia coli*. Virulence Mechanisms of a Versatile Pathogen. Elsevier Science, San Diego, pp. 3–53.
Perna, N.T., Plunkett, G., Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Pósfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E.T., Potamousis, K.D., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A., Blattner, F.R., 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature 409:529–533. http://dx.doi.org/10.1038/35054089.
Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5, e9490. http://dx.doi.org/10.1371/journal.pone.0009490.
Recktenwald, J., Schmidt, H., 2002. The nucleotide sequence of Shiga toxin (Stx) 2e-encoding phage phiP27 is not related to other Stx phage genomes, but the modular genetic structure is conserved. Infect. Immun. 70, 1896–1908.
Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. TIG 16, 276–277.
Rump, L.V., Gonzalez-Escalona, N., Ju, W., Wang, F., Cao, G., Meng, S., Meng, J., 2015. Genomic diversity and virulence profiles of historical *Escherichia coli* O157 strains isolated from clinical and environmental sources. Appl. Environ. Microbiol. 81:569–577. http://dx.doi.org/10.1128/AEM.02616-14.
Saitoh, M., Tanaka, K., Nishimori, K., Makino, S., Kanno, T., Ishihara, R., Hatama, S., Kitano, R., Kishima, M., Sameshima, T., Akiba, M., Nakazawa, M., Yokomizo, Y., Uchida, I., 2005. The artAB genes encode a putative ADP-ribosyltransferase toxin homologue associated with *Salmonella enterica* serovar Typhimurium DT104. Microbiol. (Read. Engl.) 151:3089–3096. http://dx.doi.org/10.1099/mic.0.27933-0.
Sanjar, F., Rusconi, B., Hazen, T.H., Koenig, S.S.K., Mammel, M.K., Feng, P.C.H., Rasko, D.A., Eppinger, M., 2015. Characterization of the pathogenome and phylogenomic classification of enteropathogenic *Escherichia coli* of the O157:non-H7 serotypes. Pathog. Dis. 73. http://dx.doi.org/10.1093/femspd/ftv033.
Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinforma. Oxf. Engl. 30:2068–2069. http://dx.doi.org/10.1093/bioinformatics/btu153.
Sheludchenko, M.S., Huygens, F., Stratton, H., Hargreaves, M., 2015. CRISPR diversity in *E. coli* isolates from Australian animals, humans and environmental waters. PloS One 10, e0124090. http://dx.doi.org/10.1371/journal.pone.0124090.
Stephan, R., Borel, N., Zweifel, C., Blanco, M., Blanco, J.E., 2004. First isolation and further characterization of enteropathogenic *Escherichia coli* (EPEC) O157:H45 strains from cattle. BMC Microbiol. 4:10. http://dx.doi.org/10.1186/1471-2180-4-10.
Sváb, D., Tóth, I., 2012. Allelic types of long polar fimbriae in bovine and human *Escherichia coli* O157 strains. Acta Vet. Hung. 60:1–15. http://dx.doi.org/10.1556/AVet.2012.001.
Sváb, D., Horváth, B., Maróti, G., Dobrindt, U., Tóth, I., 2013a. Sequence variability of P2-like prophage genomes carrying the cytolethal distending toxin V operon in *Escherichia coli* O157. Appl. Environ. Microbiol. 79:4958–4964. http://dx.doi.org/10.1128/AEM.01134-13.
Sváb, D., Horváth, B., Szucs, A., Maróti, G., Tóth, I., 2013b. Draft genome sequence of an *Escherichia coli* O157:H43 strain isolated from cattle. Genome Announc. 1. http://dx.doi.org/10.1128/genomeA.00263-13.
Tabei, Y., Kiryu, H., Kin, T., Asai, K., 2008. A fast structural multiple alignment method for long RNA sequences. BMC Bioinf. 9:33. http://dx.doi.org/10.1186/1471-2105-9-33.
Taieb, F., Sváb, D., Watrin, C., Oswald, E., Tóth, I., 2015. Cytolethal distending toxin A, B and C subunit proteins are necessary for the genotoxic effect of *Escherichia coli* CDT-V. Acta Vet Hung. 63:1–10. http://dx.doi.org/10.1556/AVet.2015.001.
Tamhankar, A.J., Nerkar, S.S., Khadake, P.P., Akolkar, D.B., Apurwa, S.R., Deshpande, U., Khedkar, S.U., Stalsby-Lundborg, C., 2015. Draft genome sequence of enterotoxigenic *Escherichia coli* strain E24377A, obtained from tribal drinking water source in India. Genome Announc. 3, e00225-15. http://dx.doi.org/10.1128/genomeA.00225-15.
Toro, M., Cao, G., Ju, W., Allard, M., Barrangou, R., Zhao, S., Brown, E., Meng, J., 2014. Association of clustered regularly interspaced short palindromic repeat (CRISPR) elements with specific serotypes and virulence potential of shiga toxin-producing *Escherichia coli*. Appl. Environ. Microbiol. 80:1411–1420. http://dx.doi.org/10.1128/AEM.03018-13.
Torres, A.G., Blanco, M., Valenzuela, P., Slater, T.M., Patel, S.D., Dahbi, G., López, C., Barriga, X.F., Blanco, J.E., Gomes, T.A.T., Vidal, R., Blanco, J., 2009. Genes related to long polar fimbriae of pathogenic *Escherichia coli* strains as reliable markers to identify virulent isolates. J. Clin. Microbiol. 47:2442–2451. http://dx.doi.org/10.1128/JCM.00566-09.
Torres, A.G., Giron, J.A., Perna, N.T., Burland, V., Blattner, F.R., Avelino-Flores, F., Kaper, J.B., 2002. Identification and characterization of lpfABCC′DE, a fimbrial operon of enterohemorrhagic *Escherichia coli* O157:H7. Infect. Immun. 70, 5416–5427.
Tóth, I., Sváb, D., 2014. Cell cycle modulating toxins produced by *Escherichia coli*. In: Morabito, S. (Ed.), Pathogenic *Escherichia coli*. Caister Academic Press, Norfolk, pp. 117–138.
Tóth, I., Schmidt, H., Kardos, G., Lancz, Z., Creuzburg, K., Damjanova, I., Pászti, J., Beutin, L., Nagy, B., 2009. Virulence genes and molecular typing of different groups of *Escherichia coli* O157 strains in cattle. Appl. Environ. Microbiol. 75:6282–6291. http://dx.doi.org/10.1128/AEM.00873-09.
Touchon, M., Rocha, E.P.C., 2010. The small, slow and specialized CRISPR and anti-CRISPR of Escherichia and Salmonella. PLoS One 5, e11126. http://dx.doi.org/10.1371/journal.pone.0011126.
Vica Pacheco, S., García González, O., Paniagua Contreras, G.L., 1997. The lom gene of bacteriophage lambda is involved in *Escherichia coli* K12 adhesion to human buccal epithelial cells. FEMS Microbiol. Lett. 156, 129–132.
Wyrsch, E., Roy Chowdhury, P., Abraham, S., Santos, J., Darling, A.E., Charles, I.G., Chapman, T.A., Djordjevic, S.P., 2015. Comparative genomic analysis of a multiple antimicrobial resistant enterotoxigenic *E. coli* O157 lineage from Australian pigs. BMC Genomics 16:165. http://dx.doi.org/10.1186/s12864-015-1382-y.
Yin, S., Jensen, M.A., Bai, J., Debroy, C., Barrangou, R., Dudley, E.G., 2013. The evolutionary divergence of Shiga toxin-producing *Escherichia coli* is reflected in clustered regularly interspaced short palindromic repeat (CRISPR) spacer composition. Appl. Environ. Microbiol. 79:5710–5720. http://dx.doi.org/10.1128/AEM.00950-13.
Zhou, M., Guo, Z., Duan, Q., Hardwidge, P.R., Zhu, G., 2014. *Escherichia coli* type III secretion system 2: a new kind of T3SS? Vet. Res. 45:32. http://dx.doi.org/10.1186/1297-9716-45-32.
Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., Wishart, D.S., 2011. PHAST: a fast phage search tool. Nucleic Acids Res. 39:W347–W352. http://dx.doi.org/10.1093/nar/gkr485.

Web references

Illumina Inc., r. http://www.illumina.com/documents/products/technotes/technote_nextera_matepair_data_processing.pdf (accessed: 8th June 2016).