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

Domonkos Sváb[1], Balázs Bálint[2], Gergely Maróti[3], István Tóth[1*]

[1]Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary. E-mail: svab.domonkos@agrar.mta.hu

[2]Seqomics Biotechnology Ltd., Mórahalom. E-mail: balintb@seqomics.hu

[3]Institute of Biochemistry, Biological Research Centre, Hungarian Academy of Sciences, Szeged. E-mail: marotig@brc.hu

Running title: Comparative genomics of CDT-producing E. coli O157:H43

Keywords: Escherichia, O157:H43, CDT, LT-II, comparative genomics, phylogeny

*Corresponding author: István Tóth, Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, H-1143 Budapest, Hungária krt. 21. Hungary. Phone: +36-1-2522455. Fax:+ 36-1-2521069. E-mail: toth.istvan@agrar.mta.hu

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.

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 (Seeman 2014) to ensure identical gene prediction settings. Get homologues (Contreras-Moreira 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-Moreira 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 82.4kb (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 4,687 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 to 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 T22pp1 to T22pp10. 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).

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 T22pp10 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 alsoT22pp10 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 3.2). Additional virulence genes are also present in the chromosome, representing fimbriae and secretion systems (see below 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 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, utilised 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 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).

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 11368, 11128 and 120009, 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 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 EHEC strains 11128, 11368 and 120009. 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* (Fig. 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).

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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: 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. doi: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. doi: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. doi: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. doi: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. doi:10.3201/eid1311.070381

Gyles, C.L., 2007. Shiga toxin-producing Escherichia coli: an overview. J. Anim. Sci. 85, E45-62. doi: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. doi: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. doi: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. doi: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. doi: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, D591-595. doi: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. doi: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. doi:10.1371/journal.pone.0029898

Kaper, J.B., Nataro, J.P., Mobley, H.L., 2004. Pathogenic Escherichia coli. Nat. Rev. Microbiol. 2, 123–140. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi:10.4161/bact.1.4.18470

Ogura, Y., Ooka, T., Asadulghani, Terajima, J., 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. doi: 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. doi: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. doi: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. doi: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. doi: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. doi:10.1093/femspd/ftv033

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinforma. Oxf. Engl. 30, 2068–2069. doi: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. doi: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. doi:10.1186/1471-2180-4-10

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. doi: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. doi:10.1128/genomeA.00263-13

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. doi:10.1556/AVet.2012.001

Tabei, Y., Kiryu, H., Kin, T., Asai, K. 2008. A fast structural multiple alignment method for long RNA sequences. BMC Bioinformatics. 9, 33. doi: 10.1186/1471-2105-9-33

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. doi: 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. doi: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. doi: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., 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. doi:10.1128/AEM.00873-09

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.

Touchon, M., Rocha, E.P.C., 2010. The small, slow and specialized CRISPR and anti-CRISPR of Escherichia and Salmonella. PloS One 5, e11126. doi: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. doi: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. doi: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. doi: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, W347-352. doi:10.1093/nar/gkr485

Web references

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

Figure legends.

Figure 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.

Figure 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.

Figure 3.

Geneticstructure 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.

Figure 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 1862common genes.

Figure 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.

Tables

Table 1.

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

	Chromosome	**Plasmid**
Size (bp)	4,875,355	82,444
Predicted coding CDSs	4,687	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	32065	32065	49.4	PhiP27-like	unknown	tRNA 5' methoxy-uridine synthase	none identifed	present	
T22pp2	1	385293	398890	13598	48.8	none	proline-tRNA	type I secretion system	none identifed	present	
T22pp3	1	1480783	1531940	51158	49.3	*Shigella* serotype converting	Met-tRNA	tmRNA binding protein SmpB	none identifed	present	cI and cro
T22pp4	1	1543880	1574799	30920	52	P2-like	ribosomal DNA	Yfi operon	none identifed	present	putative cI
T22pp5	1	1703290	1717745	14456	49	P4-like	sugar efflux transporter B	Ser-tRNA	none identifed	unidentified	
T22pp6	1	2288961	2320124	31164	52.6	P2-like	fieF	cpxP	CDT-V	present	c and cox
T22pp7	1	3712670	3744641	31972	52.2	P2-like	TetR transcriptional regulator	Trk potassium-channel	none identifed	putative int-like protein	cI and cII regulator
T22pp8	1	3888837	3928433	39597	49.9	lambdoid	putative TEGT family transport protein	Ser-tRNA	lipoprotein Bor	unidentified	
T22pp9	2	350924	396041	45118	45.5	lambdoid	cytosine tRNA-synthetase	unknown	Lom-like protein, heat-labile enterotoxin IIc1, *E. coli* artAB-like AB toxin	present	
T22pp10	3	212066	258765	46700	50.4	lambdoid	putative transport protein	starvation sensing protein rspB	Lom-like protein	unidentified	

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*
CRISPR2	*ygcF, ygcE*	26	32
CRISPR3	*clpS, aat*	0	N/A
CRISPR4	*clpS, aat*	0	N/A

*One spacer in CRISPR1 is 33 nt long.

Table A1.

Designations and accession numbers of *E. coli* and *Shigella* genomes and plasmids used in core genome phylogeny. EAEC: enteroaggregative *E. coli,* UPEC: uropathogenic *E. coli,* ExPEC: extraintestinal pathogenic *E. coli,* APEC: avian pathogenic *E. coli,* ETEC: enterotoxigenic *E. coli,* MNEC: meningitis associated *E. coli,* AIEC: adherent-invasive *E. coli,* STEC: Shiga toxin producing *E. coli,* EHEC: enterohemorrhagic *E. coli,* EPEC: enteropathogenic *E. coli*

Strain	**Pathogenity**	**Accession number**	**Length**	**Type**
042	EAEC	NC_017626	5241977	genome
042		NC_017627	113346	plasmid
536	UPEC	NC_008253	4938920	genome
55989	EAEC	NC_011748	5154862	genome
ABU_83972	ExPEC	NC_017631	5131397	genome
ABU_83972		NC_017629	1564	plasmid
APEC_O1	APEC	NC_008563	5082025	genome
APEC_O1		NC_009838	241387	plasmid
APEC_O1		NC_009837	174241	plasmid
APEC_O78	APEC	NC_020163	4798435	genome
ATCC_8739	non-pathogenic	NC_010468	4746218	genome
B_REL606	non-pathogenic	NC_012967	4629812	genome
BL21_DE3	non-pathogenic	NC_012971	4558953	genome
BL21_Gold_DE3_pLysS_AG	non-pathogenic	NC_012947	4570938	genome
BW2952	non-pathogenic	NC_012759	4578159	genome
CFT073	UPEC	NC_004431	5231428	genome
clone_D_i14	UPEC	NC_017652	5038386	genome
clone_D_i2	UPEC	NC_017651	5038386	genome

DH1	non-pathogenic	NC_017638	4621430	genome
E24377A	ETEC	NC_009801	4979619	genome
E24377A		NC_009786	79237	plasmid
E24377A		NC_009790	74224	plasmid
E24377A		NC_009788	70609	plasmid
E24377A		NC_009787	34367	plasmid
E24377A		NC_009789	6199	plasmid
E24377A		NC_009791	5033	plasmid
ED1a	commensal	NC_011745	5209548	genome
ETEC_H10407	ETEC	NC_017633	5153435	genome
ETEC_H10407		NC_017724	94797	plasmid
ETEC_H10407		NC_017722	66681	plasmid
ETEC_H10407		NC_017723	5800	plasmid
ETEC_H10407		NC_017721	5175	plasmid
FN649414	ETEC	FN649414	5153435	genome
HS	commensal	NC_009800	4643538	genome
IAI1	commensal	NC_011741	4700560	genome
IAI39	ExPEC	NC_011750	5132068	genome
IHE3034	MNEC	NC_017628	5108383	genome
JJ1886	ExPEC	NC_022648	5129938	genome
JJ1886		NC_022651	110040	plasmid
JJ1886		NC_022650	55956	plasmid
JJ1886		NC_022662	5631	plasmid
JJ1886		NC_022649	5167	plasmid
JJ1886		NC_022661	1552	plasmid
K_12_substr_DH10B	non-pathogenic	NC_010473	4686137	genome
K_12_substr_MDS42	non-pathogenic	NC_020518	3976195	genome
K_12_substr_MG1655	non-pathogenic	NC_000913	4641652	genome
K_12_substr_W3110	non-pathogenic	NC_007779	4646332	genome

KO11FL	non-pathogenic	NC_017660	5021812	genome
KO11FL		NC_016904	103795	plasmid
KO11FL		NC_016903	5360	plasmid
LF82	AIEC	NC_011993	4773108	genome
LY180		NC_022364	4835601	genome
NA114		NC_017644	4971461	genome
Shigella flexneri 2a	Dysentery	NC_004337	4607202	genome
Shigella sonnei Ss046	Dysentery	NC_007384	4825265	genome
Shigella dysenteriae Sd197	Dysentery	NC_007606	4369232	genome
Shigella boydii Sb227	Dysentery	NC_007613	4519823	genome
EDL933	EHEC	NZ_CP008957	5547323	genome
pO157		AF074613.1	92072	plasmid
O103:H2_12009	EHEC	NC_013353	5449314	genome
O103:H2_12009		NC_013354	75546	plasmid
O104:H4_2009EL_2050	EAEC/STEC	NC_018650	5253138	genome
O104:H4_2009EL_2050		NC_018651	109274	plasmid
O104:H4_2009EL_2050		NC_018654	74213	plasmid
O104:H4_2009EL_2050		NC_018652	1549	plasmid
O104:H4_2009EL_2071	EAEC/STEC	NC_018661	5312586	genome
O104:H4_2009EL_2071		NC_018662	75573	plasmid
O104_H4_2009EL_2071		NC_018663	1549	plasmid
O104_H4_2011C_3493	EAEC/STEC	NC_018658	5273097	genome
O104_H4_2011C_3493		NC_018659	88544	plasmid
O104_H4_2011C_3493		NC_018666	74217	plasmid
O104_H4_2011C_3493		NC_018660	1549	plasmid
O111_H_11128	EHEC	NC_013364	5371077	genome
O111_H_11128		NC_013365	204604	plasmid
O111_H_11128		NC_013370	97897	plasmid
O111_H_11128		NC_013366	77690	plasmid

O111_H_11128		NC_013367	8140	plasmid
O111_H_11128		NC_013368	6673	plasmid
O127_H6_E2348_69	EPEC	NC_011601	4965553	genome
O127_H6_E2348_69	EPEC	NC_011603	97978	plasmid
O127_H6_E2348_69	EPEC	NC_011602	6147	plasmid
O157_H7 Sakai	EHEC	NC_002695	5498450	genome
O157_H7	EHEC	NC_002128	92721	plasmid
O157_H7	EHEC	NC_002127	3306	plasmid
O157_H7_EC4115	EHEC	NC_011353	5572075	genome
O157_H7_EC4115		NC_011350	94644	plasmid
O157_H7_EC4115		NC_011351	37452	plasmid
O157_H7_EDL933	EHEC	NC_002655	5528445	genome
O157_H7_EDL933		NC_007414	92077	plasmid
O157_H7_TW14359	EHEC	NC_013008	5528136	genome
O157_H7_TW14359		NC_013010	94601	plasmid
O26_H11_11368	EHEC	NC_013361	5697240	genome
O26_H11_11368		NC_013369	85167	plasmid
O26_H11_11368		NC_013362	63365	plasmid
O26_H11_11368		NC_013363	5686	plasmid
O26_H11_11368		NC_014543	4073	plasmid
O55_H7_CB9615	EPEC	NC_013941	5386352	genome
O55_H7_CB9615		NC_013942	66001	plasmid
O55_H7_RM12579	EPEC	NC_017656	5263980	genome
O55_H7_RM12579		NC_017653	94015	plasmid
O55_H7_RM12579		NC_017657	66078	plasmid
O55_H7_RM12579		NC_017654	12068	plasmid
O55_H7_RM12579		NC_017658	6211	plasmid
O55_H7_RM12579		NC_017655	5954	plasmid
O7_K1_CE10	MNEC	NC_017646	5313531	genome

O7_K1_CE10		NC_017647	54289	plasmid
O7_K1_CE10		NC_017648	5163	plasmid
O7_K1_CE10		NC_017649	4197	plasmid
O7_K1_CE10		NC_017650	1549	plasmid
O83_H1_NRG_857C	AIEC (adherent-invasive)	NC_017634	4747819	genome
O83_H1_NRG_857C		NC_017659	147060	plasmid
P12b	commensal	NC_017663	4935294	genome
PMV_1	ExPEC	NC_022370	4984940	genome
PMV_1		NC_022371	98864	plasmid
S88	MNEC	NC_011742	5032268	genome
S88		NC_011747	133853	plasmid
SE11	commensal	NC_011415	4887515	genome
SE11		NC_011419	100021	plasmid
SE11		NC_011413	91158	plasmid
SE11		NC_011416	60555	plasmid
SE11		NC_011407	6929	plasmid
SE11		NC_011408	5366	plasmid
SE11		NC_011411	4082	plasmid
SE15	commensal	NC_013654	4717338	genome
SE15		NC_013655	122345	plasmid
SMS_3_5	commensal	NC_010498	5068389	genome
SMS_3_5		NC_010488	130440	plasmid
SMS_3_5		NC_010485	8909	plasmid
SMS_3_5		NC_010486	4074	plasmid
SMS_3_5		NC_010487	3565	plasmid
T22_genome_and_plasmid		LYNF00000000	4957799	genome and plasmid
UM146	AIEC (adherent-invasive)	NC_017632	4993013	genome
UM146		NC_017630	114550	plasmid
UMN026	UPEC	NC_011751	5202090	genome

UMN026		NC_011749	122301	plasmid
UMN026		NC_011739	33809	plasmid
UMNK88	ETEC	NC_017641	5186416	genome
UMNK88		NC_017645	160573	plasmid
UMNK88		NC_017642	90868	plasmid
UMNK88		NC_017639	81883	plasmid
UMNK88		NC_017640	81475	plasmid
UMNK88		NC_017643	65549	plasmid
UTI89	UPEC	NC_007946	5065741	genome
UTI89		NC_007941	114230	plasmid
W	non-pathogenic	NC_017664	4897452	genome
W		NC_017637	102536	plasmid
W		NC_017636	5360	plasmid
Xuzhou21	EHEC	NC_017906	5386223	genome
Xuzhou21		NC_017907	92728	plasmid
Xuzhou21		NC_017903	37785	plasmid

Table A2.

T22-specific genes when compared to the genetic content of *E. coli* EDL933, SE11, and IAI1. Note that four genes are present in two copies.

Description	Contig	Start	Stop	Strand	Closest homologue(s)	Identity (%)	GenBank no.
hypothetical protein	1	4993	5334	-	*E. coli* H1827/12	340/342 (99%)	CP013031.1
Caudovirus prohead protein	1	9580	10230	-	*E. coli* UMNK88, conserved hypothetical protein	646/651 (99%)	CP002729.1
hypothetical protein	1	11449	11631	-	*E. coli* UMNK88, conserved hypothetical protein	176/183 (96%)	CP002729.1
phage terminase small subunit	1	13400	13882	-	Present in more than 10 *E. coli* strains, terminase	479/483 (99%)	
hypothetical protein	1	17155	17643	-	*E. coli* 1303, hypothetical protein	487/489 (99%)	CP009166.1
hypothetical protein	1	21195	22103	-	*E. coli* CFSAN029787, DNA binding protein	896/909 (99%)	CP011416.1
Ash protein family peptidase	1	22320	23180	-	*E. coli* GB089, peptidase	823/843 (98%)	CP013663.1
hypothetical protein	1	23198	23407	-	*E. coli* 2012C-4227, hypothetical protein	210/210 (100%)	CP013029.1
ORF6 domain protein	1	23493	24260	-	*E. coli* 08-00022, antirepressor	765/769 (99%)	CP013662.1
hypothetical protein	1	24357	24476	+	*E. coli* 08-00022, hypothetical protein	120/120 (100%)	CP013662.1
HTH type transcriptional regulator	1	24770	25459	+	*E. coli* O157:H7 FRIK944, transcriptional regulator	646/648 (99%)	CP016625.1
transcriptional repressor	1	25606	26058	+	*E. coli* CFSAN029787, DNA binding protein	424/432 (98%)	CP011416.1
hypothetical protein	1	27358	27453	-	*E. coli* GB089, hypothetical protein	96/96 (100%)	CP013663.1
putative hydrolase	1	109272	109700	-	Present in more than 10 *E. coli* strains, metal-dependent phosphohydrolase	429/429 (100%)	
invasin	1	126526	127458	+	E. *coli*O145:H28 str. RM12581, adherence invasion outer membrane protein	837/837 (99%)	CP007136.1
hypothetical protein	1	147204	147320	+	Presen in more than 10 *E. coli* strains, two-component sensor hisditine kinase	117/117 (100%)	

rbsK	1	147516	148730	+	Present in more than 10 *E. coli* strains, ribokinase	1213/1215 (99%)	
rbsK	1	147516	148730	+	Present in more than 10 *E. coli* strains, ribokinase	1213/1215 (99%)	
hypothetical protein	1	148744	149502	+	*E. coli* UMNK88, conserved hypothetical protein	759/759 (100%)	CP002729.1
hypothetical protein	1	148744	149502	+	*E. coli* UMNK88, conserved hypothetical protein	759/759 (100%)	CP002729.1
hypothetical protein	1	149559	150167	+	Present in more than 10 E. coli strains, hypothetical protein	609/609 (100%)	
hypothetical protein	1	150261	150524	+	Present in more than 10 E. coli strains, hypothetical protein	264/264 (100%)	
hypothetical protein	1	150527	151561	-	Present in more than 10 *E. coli* strains, phosphotriesterase	1035/1035 (100%)	
transposase	1	152213	153592	-	Present in more than 10 *E. coli* strains, transposase	1378/1380 (99%)	
hypothetical protein	1	153876	154286	-	Present in more than 10 *E. coli* strains, two-component sensor histidine kinase	411/411 (100%)	
hypothetical protein	1	154274	154675	-	Present in more than 10 E. coli strains, hypothetical protein	402/402 (100%)	
hypothetical protein	1	154935	155504	-	*E. coli* SEC540, hypothetical protein	570/570 (100%)	CP013962.1
hypothetical protein	1	155704	155904	-	Present in more than 10 E. coli strains, hypothetical protein	201/201 (100%)	
hypothetical protein	1	157331	157933	+	Present in more than 10 E. coli strains, hypothetical protein	603/603 (100%)	
hypothetical protein	1	162772	163923	+	Present in more than 10 E. coli strains, hypothetical protein	1152/1152 (100%)	
hypothetical protein	1	169598	170281	+	*E. coli* EC590, hypothetical protein	684/684 (100%)	CP016182.1
hypothetical protein	1	170297	170707	+	Present in more than 10 E. coli strains, hypothetical protein	411/411 (100%)	
propanediol utilization protein pduV	1	174754	174978	+	*E. coli* EC590	225/225 (100%)	CP016182.1
putative transporter component	1	180514	180699	-	*E. coli* O157:H43 PV00-24 O-antigen biosynthesis cluster, pseudogene, C-terminal part of	186/186 (100%)	AB602253.1

					YeeE		
hypothetical protein	1	386651	386893	+	*E. coli* 08-00022, hypothetical protein	231/243 (95%)	CP013662.1
hypothetical protein	1	387808	387987	-	*E. coli* ZH063, hypothetical protein	178/180 (99%)	CP014522.1
hypothetical protein	1	388719	389171	+	*E. coli* FHI29	388/402 (97%)	LM995858.1
hypothetical protein	1	389173	389406	+	Present in more than 10 *E. coli* strains, hypothetical protein	225/234 (96%)	
Poxvirus D5 protein	1	389708	391462	+	*E. coli* GB089, DNA primase	1722/1725 (99%)	CP013663.1
prophage CP4-57 regulatory protein	1	394122	394466	+	*E. coli* GB089, AlpA family transcriptional regulator	252/307 (82%)	CP013663.1
hypothetical protein	1	430123	430536	-	Present in more than 10 *E. coli* strains, hypothetical protein	414/415 (99%)	
mrpA	1	681130	681960	+	*E. coli* S51, hydantoin racemase	831/831 (100%)	CP015995.1
hydrogenase 4 subunit D	1	681993	682361	+	*E. coli* ACN001, ACN002, YD786, FHI65, FHI98, APEC O78, hydrogenase 4 subunit D	369/369 (100%)	CP007491.1 CP013112.1 CP007442.1 LM996539.1 LM997368.1 CP004009.1
hycG	1	686817	687338	+	*E. coli* SEC470, formate hydrogenylase	522/522 (100%)	CP013962.1
fhlA1	1	688010	688960	+	*E. coli* S51, hydrogenase	951/951 (100%)	CP015995.1
fhlA2	1	688967	690022	+	*E. coli* S51, hydrogenase	1056/1056 (100%)	CP015995.1
hypothetical protein	1	832127	832288	+	Present in more than 10 E. coli strains, hypothetical protein	162/162 (100%)	
hypothetical protein	1	1062900	1063769	-	*E. coli* CI5, membrane protein	870/870 (100%)	CP011018.1
Oligogalacturonate-specific porin KdgM precursor	1	1063783	1064478	-	*E. coli* CI5, porin	696/696 (100%)	CP011018.1
hypothetical protein	1	1064613	1065737	-	*E. coli* S51, 3-octaprenyl-4-hydroxybenzoate carboxy-lyase	1125/1125 (100%)	CP015995.1
rbsK2	1	1065767	1066681	-	*E. coli* S51, carbohydrate kinase	915/915 (100%)	CP015995.1
glucosamine--fructose-6-phosphate aminotransferase	1	1066686	1067684	-	*E. coli* CI5, porin, iron dicitrate transport regulator FecR	999/999 (100%)	CP011018.1

					ADP-ribosylglycohydrolase		
ADP-ribosyl-glycohydrolase	1	1067684	1068760	+	*E. coli* S51, ADP-ribosyl-glycohydrolase	1077/1077 (100%)	CP015995.1
HTH-type transcriptional repressor DasR	1	1068774	1069526	-	*E. coli* CI5, porin, iron dicitrate transport regulator FecR ADP-ribosylglycohydrolase	999/999 (100%)	CP011018.1
yicJ	1	1069813	1071144	+	*E. coli* CI5, permease	1332/1332 (100%)	CP011018.1
hypothetical protein	1	1273239	1273337	+	*E. coli* E24377A	99/99 (100%)	CP000800.1
lsrK	1	1369620	1370141	+	*E. coli* Co6114, hypothetical protein	522/522 (100%)	CP016034.1
hypothetical protein	1	1484487	1486739	-	*E. coli* O104:H4 str. C227-11, hypothetical protein	2253/2253 (100%)	CP011331.1
hypothetical protein	1	1488390	1489169	-	*E. coli* H1827/12, hypothetical protein	778/780 (99%)	CP013031.1
hypothetical protein	1	1490344	149055	-	*E. coli* H1827/12, integrase	207/207 (100%)	CP013031.1
hypothetical protein	1	1490610	1490825	-	*E. coli* H1827/12, hypothetical protein	213/216 (100%)	CP013031.1
hypothetical protein	1	1490822	1491184	-	*E. coli* Co6114, hypothetical protein	360/363 (99%)	CP016034.1
hypothetical protein	1	1491175	1491711	-	*E. coli* FHI75	526/535 (98%)	LM996644.1
hypothetical protein	1	1491840	1492664	-	*E. coli* H1827/12, hypothetical protein	817/825 (99%)	CP013031.1
hypothetical protein	1	1492730	1493092	-	*E. coli* H1827/12, hypothetical protein	362/363 (99%)	CP013031.1
hypothetical protein	1	1493530	1493775	-	*E. coli* GB089, phage tail protein	246/246 (100%)	CP013663.1
hypothetical protein	1	1493693	1493989	-	*E. coli* GB089, phage tail protein	297/297 (100%)	CP013663.1
helix-turn-helix type protein	1	1495057	1495317	+	*E. coli* FORC_028, phage DNA binding protein	261/261 (100%)	CP012693.1
hypothetical protein	1	1495310	1495861	+	*E. coli* FORC_028, phage DNA binding protein	550/552 (99%)	CP012693.1
hypothetical protein	1	1495858	1496196	+	*E. albertii* KF1, hypothetical protein	339/339 (100%)	CP007025.1
Bacterial regulatory proteins, gntR family	1	1496206	1497147	+	*E. coli* CI5, GntR family transcriptional regulator	752/753 (99%)	CP011018.1
PerC transcriptional regulator	1	1497150	1497638	+	*Shigella* phage Sfv, hypothetical	486/489 (99%)	KC814930.1

					protein		
DNA N-6-adenine methylase	1	1497638	1498291	+	*E. coli* H1827/12, metyhltransferase	646/654 (99%)	CP013031.1
lexA	1	1498288	1498614	+	*E. coli* O177:H21, LexA family transcriptional regulator	326/372 (99%)	CP016546.1
rusA	1	1498611	1499000	+	*E. coli* O157:H7 FRIK944, crossover junction endodeoxyribonuclease	351/365 (96%)	CP009166.1
KilA-N domain protein	1	1499020	1499817	+	*Shigella* phage SfIV, replication protein	939/942 (99%)	KC814930.1
hypothetical protein	1	1499825	1500814	+	*E. coli* K15-KW01, hypothetical protein	967/990 (98%)	CP016358.1
hypothetical protein	1	1501233	1501682	-	APEC O78, hypothetical protein	447/450 (99%)	CP004009.1
hypothetical protein	1	1501704	1502645	-	APEC O78, hypothetical protein	936/942 (99%)	CP004009.1
hypothetical protein	1	1504293	1505051	+	none		
phage terminase	1	1505607	1506101	+	*E. coli* UMNK88, phage terminase small subunit	491/495 (99%)	CP002729.1
hypothetical protein	1	1507843	1508025	+	*Salmonella* Anatum, hypothetical protein	181/183 (99%)	CP014659.1
Caudovirus prohead protein	1	1509244	1509894	+	Present in more than 10 *E. coli* strains, primosome assembly protein PriA	650/651 (99%)	
hypothetical protein	1	1511164	1511364	+	*E. coli* H1827/12, hypothetical protein	201/201 (100%)	CP013031.1
hypothetical protein	1	1512072	1512578	+	Present in more than 10 E. coli strains, hypothetical protein	504/507 (99%)	
hypothetical protein	1	1512575	1513135	+	*E. coli* FHI42	554/561 (99%)	LK999942.1
phage tail tube protein	1	1514794	1515150	+	*E. coli* FHI89	357/357 (100%)	LM997040.1
hypothetical protein	1	1515150	1515419	+	*E. coli* H1827/12, hypothetical protein	270/270 (100%)	CP013031.1
phage related minor protein	1	1515561	1517384	+	Enterobacteria phage SfI, tail protein	1806/1821 (99%)	JX509734.1
hypothetical protein	1	1525165	1525935	+	*E. coli* O157:H16 strain Santai	409/415 (99%)	CP007592.1
Phage Tail Collar Domain protein	1	1526973	1527446	-	*E. coli* MG1655 phage tail protein, 61% coverage	225/238 (95%)	CP012868.1
Type-2 restriction enzyme EcoRII	1	1530732	1531940	-	*E. coli* G749, restriction endonuclease	1204/1209 (99%)	CP014488.1

phage P2 gpU	1	1545440	1545925	-	*E. coli* SF-468 phage tail protein	486/486 (100%)	CP012625.1
phage tail protein	1	1548528	1548863	-	*E. coli* SF-468 hypothetical protein	336/336 (100%)	CP012625.1
phage tail tube protein	1	1548927	1549445	-	*E. coli* SF-468 phage tail protein	519/519 (100%)	CP012625.1
Caudovirales tail fiber	1	1550771	1551367	-	*E. coli* SF-468 phage tail protein	491/599 (82%)	CP012625.1
Phage Tail Collar Domain protein	1	1551367	1552380	-	*E. col*i SF-468, phage tail protein, 74% coverage	675/760 (89%)	CP012625.1
phage tail protein	1	1552377	1552985	-	*Serratia marcescens* SM39, putative phage baseplate assembly protein	579/609 (95%)	AP013063.1
baseplate wedge subunit	1	1553893	1554240	-	*Salmonella* Weltevreden C2346, baseplate assembly protein	333/348 (96%)	LN890520.1
hypothetical protein	1	1555009	1555734	+	none		
phage virion protein	1	1555809	1556258	-	*Salmonella* phage SEN1, tail completion protein	400/453 (88%)	KT630644.1
P2 phage tail completion	1	1556251	1556718	-	*Serratia marcescens* SM39, putative phage tail completion protein R	464/468 (99%)	AP013063.1
hypothetical protein	1	1556681	1556839	-	*E. coli* SF-468 phage lysis protein	154/159 (97%)	CP012625.1
phage lysis regulatory protein, LysB family	1	1556826	1557239	-	*E. coli* SF-468, phage lysis protein, partially covering the original	387/414 (93%)	CP012625.1
phage lysozyme	1	1557236	1557733	-	*E. coli* SF-468 lysozyme	477/498 (96%)	CP012625.1
Phage holin family 2	1	1557720	1558016	-	*E. coli* SF-468, holin	296/297 (99%)	CP012625.1
phage tail protein	1	1558020	1558223	-	*E. coli* SF-468 phage tail protein	201/204 (99%)	CP012625.1
phage head completion protein	1	1558223	1558732	-	*E. coli* SF-468 phage head completion/stabilization protein	505/510 (99%)	CP012625.1
terminase endonuclease	1	1558826	1559575	-	Bacteriophage 186, R protein	733/750 (98%)	U32222.1
capsid scaffolding protein	1	1560723	1561577	-	*E. coli* SF-468 precorrin-8W decarboxylase	825/855 (97%)	CP012625.1
hypothetical protein	1	1564596	1565345	-	none		
hypothetical protein	1	1565348	1566175	-	none		
hypothetical protein	1	1566822	1567253	+	Present in more than 10 *Salmonella* Newport strains, hypothetical protein	406/432 (92%)	
hypothetical protein	1	1567322	1567531	-	*E. coli* SF-468	202/210 (96%)	CP012625.1
hypothetical protein	1	1567965	1568147	-	*E. coli* SF-468, TumA	182/183 (99%)	CP012625.1
bacteriophage replication protein	1	1568261	1570390	-	*E. coli* SF-468, replication protein	2042/2130 (96%)	CP012625.1
hypothetical protein	1	1570706	1570933	-	*E. coli* PCN033, hypothetical	214/228 (94%)	CP006632.1

					protein		
hypothetical protein	1	1571000	1571338	-	*Salmonella* Typhimurium D23580, predicted phage protein	330/339 (97%)	FN424405.1
hypothetical protein	1	1571302	1571502	-	*Salmonella* Weltevreden C2346, bacteriophage protein	183/201 (91%)	LN890520.1
Phage regulatory protein CII (CP76)	1	1571510	1572019	-	*Salmonella* Weltevreden str. 1655, hypothetical protein	460/510 (90%)	CP014996.1
hypothetical protein	1	1572052	1572423	-	Salmonella Weltevreden 1655, Cro/CI family transcriptional regulator	344/372 (92%)	CP014996.1
Bacteriophage CI repressor helix-turn-helix domain protein	1	1572544	1573380	+	*Kluyvera intermedia* strain CAV1151, ornithine decarboxylase repressor	757/837 (90%)	CP011602.1
hypothetical protein	1	1573389	1573727	+	*Salmonella* phage SEN1, hypothetical protein	303/337 (90%)	KT630644.1
hypothetical protein	1	1706665	1707120	-	*Salmonella* Typhimurium CDC 2011K-1702, hypothetical protein	456/456 (100%)	CP014967.1
hypothetical protein	1	1707113	1707400	-	*Salmonella* Typhimurium CDC 2011K-1702, derepression protein	288/288 (100%)	CP014967.1
DNA binding transcription regulator	1	1709530	1710030	+	*E. coli* FHI72	501/501 (100%)	LM996868.1
hypothetical protein	1	1711102	1712391	+	*E. coli* FHI28	1288/1290 (99%)	LM995781.1
hypothetical protein	1	1712493	1713866	-	*E. coli* FHI28	1374/1374 (100%)	LM995781.1
hypothetical protein	1	1929013	1929384	+	*E. coli* FAP1, hypothetical protein	372/372 (100%)	CP009578.1
phage tail protein	1	2293510	2293785	-	*E. coli* GB089, hypothetical protein	270/276 (98%)	CP013663.1
hin2	1	2295718	2296212	-	*E. coli* LY180, DNA invertase	480/499 (96%)	CP006584.1

Phage Tail Collar Domain protein	1	2296362	2296850	+	*E. coli* UMNK88, phage tail fiber protein and hypothetical protein (2 genes)	454/455 (99%), 93% coverage	CP002729.1
hypothetical protein	1	2297888	2298124	-	*E. coli* O157:H7 FRIK944, phage tail protein	237/237 (100%)	CP016625.1
hypothetical protein	1	2303027	2303185	+	*E. coli* G749, phage lysis protein	158/159 (99%)	CP014488.1
phage lysis regulatory protein, LysB family	1	2303172	2303597	-	*E. coli* G749, protein lysB	424/426 (99%)	CP014488.1
hypothetical protein	1	2303585	2304010	-	*E. coli* G749, protein lysA	424/426 (99%)	CP014488.1
Phage holin family 2	1	2304522	2304803	-	Present in more than 10 *E. coli* and phages, and also *E. albertii* strains, phage holin	282/282 (100%)	
Cytolethal distending toxin subunit A precursor	1	2312111	2312887	+	*cdtA,* present in more than 10 *E. coli* strains and phages	777/777 (100%)	
Deoxyribonuclease CdtB	1	2312884	2313693	+	*cdtB, E. coli* strain B21	810/810 (100%)	AB839657.1
Cytolethal distending toxin subunit C	1	2313708	2314253	+	*cdtC,* present in 6 *E. coli* strains besides T22	234/234 (100%)	
hypothetical protein	1	2316616	2316891	-	E. coli G749, hypothetical protein	274/276 (99%)	CP014488.1
DksA-like zinc finger protein	1	2316888	2317112	+	*E. coli* MNCRE44, hypothetical protein	223/225 (99%)	CP010876.1
hypothetical protein	1	2317112	2317414	-	*E. coli* O177:H21, hypothetical protein	295/303 (97%)	CP016546.1
hypothetical protein	1	2317414	2317638	-	*E. coli* S51, hypothetical protein	224/225 (99%)	CP015995.1
hypothetical protein	1	2317702	2318202	-	Present in 10 *E. coli* strains, replication gene B protein	501/501 (100%)	

Regulatory phage protein cox	1	2318372	2318644	-	Present in more than 10 *E. coli* and also *Shigella* strains and phages, *cox*	273/273 (100%)	
L-rhamnose operon transcriptional activator RhaR	1	2327137	2327736	-	*E. coli* O177:H21, transcriptional regulator	599/600 (99%)	CP016546.1
L-rhamnose operon transcriptional activator RhaR	1	2327703	2327987	-	*E. coli* O177:H21, transcriptional regulator	283/285 (99%)	CP016546.1
htrC	1	2410164	2410493	-	*E. coli* 268-78-1, stationary phase growth adaptation protein	329/330 (99%)	CP014092.1
hypothetical protein	1	2483320	2483826	+	*E. coli* FAP1, hypothetical protein	506/507 (99%)	CP009578.1
hypothetical protein	1	2483823	2484032	+	Present in more than 10 *E. coli* strains, hypothetical protein	209/210 (99%)	
Type I restriction enzyme EcoR124II R protein	1	2713494	2716610	-	*E. coli* CI5, DEAD/DEAH box helicase	3117/3117 (100%)	CP011018.1
Type I restriction enzyme specificity protein MPN_089	1	2716732	2717898	-	*E. coli* CI5, restriction endonuclease subunit S	1167/1179 (99%)	CP011018.1
putative type 1 restriction endonuclease	1	2717895	2719451	-	*E. coli* CI5, restriction endonuclease, M subunit	1557/1557 (100%)	CP011018.1
hypothetical protein	1	2720660	2721055	+	*E. coli* CI5, S-adenosylhomocysteine hydrolase	396/396 (100%)	CP011018.1
hypothetical protein	1	2721048	2721995	+	*E. coli* CI5, hypothetical protein	948/948 (100%)	CP011018.1
putative restriction endonuclease	1	2722597	2723247	+	*E. coli* CI5, HNH endonuclease	651/651 (100%)	CP011018.1
rhsC	1	3094034	3095224	+	*E. coli* ec45	1190/1191 (99%)	AH006361.2
hypothetical protein	1	3124086	3124361	-	*E. coli* 2011C-3911, hypothetical protein (partial coverage)	166/166 (100%)	CP015240.1
hypothetical protein	1	3412180	3412890	-	none		
hypothetical protein	1	3560553	3560708	+	Present in more than 10 *E. coli* strains, hypothetical protein	154/154 (100%)	
hypothetical protein	1	3713911	371410	-	Present in more than 10 *E. coli* strains, integrase	192/192 (100%)	
Bacteriophage CI repressor helix-turn-helix domain protein	1	3714118	3714687	-	Present in 9 *E. albertii, Salmonella* and *E. coli* strains, phage repressor protein	329/330 (99%)	
Phage regulatory protein CII (CP76)	1	3715067	3715576	+	Present in 7 *Salmonella* and 1 *E. coli* strain, hypothetical protein	509/510 (99%)	

hypothetical protein	1	3715751	3715975	+	*E. coli* RS218, hypothetical protein (partial coverage)	224/225 (99%)	CP007149.1
hypothetical protein	1	3715998	3716339	+	*Salmonella* Typhi B/SF/13/03/195, hypothetical protein	342/342 (100%)	CP012151.1
hypothetical protein	1	3716407	3716640	+	Present in more than 10 E. coli and E. albertii strains, hypothetical protein	233/234 (99%)	
DksA-like zinc finger protein	1	3716640	3716867	+	*E. coli* MNCRE44, hypothetical protein	223/225 (99%)	CP010876.1
dam1	1	3716864	3717721	+	*E. coli* MS6198, DNA adenine methylase	855/858 (99%)	CP015834.1
hypothetical protein	1	3720286	3720474	+	*E. coli* MS6198, hypothetical protein	189/189 (100%)	CP015834.1
hypothetical protein	1	3720823	3721722	-	none		
terminase ATPase subunit	1	3723153	3723860	-	*Klebsiella pneumoniae* Kp52.145, unnamed protein product	601/729 (82%)	FO834906.1
hypothetical protein	1	3729098	3729313	+	APEC IMT5155, putative secretory protein	216/216 (100%)	CP005930.1
hypothetical protein	1	3729808	3730185	+	Present in more than 10 E. coli strains, putative membrane spanning protein	378/378 (100%)	
phage lysis regulatory protein, LysB family	1	3730182	3730610	+	*E. coli* FHI40, P12b, S88, putative regulatory protein	429/429 (100%)	LM996157.1 CP002291.1 CU928161.2
hypothetical protein	1	3734069	3735205	+	*E. coli* W, putative variable tail fiber protein	1049/1060 (99%)	CP002967.1
Caudovirales tail fiber	1	3735502	3735909	+	*E. coli* O157:H16 strain Santai	400/408 (98%)	CP007592.1
phage tail protein E	1	3739247	3739549	+	*E. coli* GB089, hypothetical protein	270/276 (98%)	CP013663.1
phage-related minor protein	1	3739676	3742753	+	*E. albertii,* NIAH_Bird_3, predicted tail tape measure protein	2994/3078 (98%)	AP014855.1
cell division protein ftsK	1	3789945	3794051	+	*E. coli* S51 DNA translocase FtsK	4107/4107 (100%)	CP015995.1
exonuclease	1	3890143	3890598	-	*E. coli* 2012C-4227, exonuclease	451/456 (99%)	CP013029.1
hypothetical protein	1	3892498	3892932	+	*E. coli* EDL933-1, hypothetical protein	246/273 (90%)	CP015855.1
hypothetical protein	1	3895574	3896368	+	*E. coli* FAP1, hypothetical protein	327/345 (95%), 43% coverage	CP009578.1
hypothetical protein	1	3897033	3897311	-	APEC O78, putative SinR-like	271/279 (97%)	CP004009.1

					protein		
hypothetical protein	1	3897364	3897543	-	APEC O78, hypothetical protein	179/180 (99%)	CP004009.1
hypothetical protein	1	3898032	3898724	-	APEC O78, hypothetical protein	679/693 (98%)	CP004009.1
hypothetical protein	1	3899054	3899845	-	none		
hypothetical protein	1	3899845	3900063	-	none		
hypothetical protein	1	3900675	3901325	+	*E. coli* CFSAN004177, hypothetical protein	647/651 (99%)	CP014670.1
transcriptional repressor	1	3901306	3902409	+	*E. coli* CFSAN004177, DNA binding protein	1094/1104 (99%)	CP014670.1
hypothetical protein	1	3905607	3905843	+	Present in 8 *E. coli* strains, hypothetical protein	233/237 (98%)	
hypothetical protein	1	3907964	3908242	+	none		
hypothetical protein	1	3908846	3908983	+	*E. coli* GB089, TciA	105/113 (93%), 84% coverage	CP013663.1
hypothetical protein	1	3909685	3910077	+	*E. coli* 08-00022, holin	320/386 (83%)	CP013662.1
hypothetical protein	1	3910067	3910345	+	none		
putative peptidoglycan	1	3910347	3910892	+	*E. coli* SA53, putative phage lysozyme	466/528 (88%)	JQ031711.1
hypothetical protein	1	3910958	3911059	+	*E. coli* 06-00048, hypothetical protein	93/100 (93%)	CP015229.1
hypothetical protein	1	3911016	3911240	-	none		
phage minor tail protein	1	3926222	3926920	+	*E. coli* FORC_028, phage minor tail protein	682/699 (98%)	CP012693.1
hypothetical protein	2	43	576	+	Present in more thane 10 *E. coli* strains, hypothetical protein	484/523 (93%)	
hypothetical protein	2	537	1757	+	*E. coli* strain 06-00048, hypothetical protein	491/551 (89%), 45% coverage	CP015229.1
fhuE1	2	119324	120055	-	Present in more thane 10 *E. coli* strains, putative OMR family iron-siderophore receptor precursor	732/732 (100%)	
fhuE2	2	120072	121535	-	Present in more thane 10 *E. coli* strains, ferric-rhodoturic acid / ferric coprogen receptor FhuE	1442/1464 (98%)	
ycfZ	2	140940	141278	-	*E. coli* S51, hypothetical protein	339/339 (100%)	CP015995.1
hypothetical protein	2	330803	330985	-	Present in more than 10 E. coli strains, oxydoreductase	183/183 (100%)	

rcbA	2	352476	352760	-	Present in more thane 10 E. coli strains, hypothetical protein	262/283 (93%)	
RecT family protein	2	352914	353966	-	*Salmonella* Typhi CT18, hypothetical prophage protein	1014/1053 (96%)	AL513382.1
recE	2	353978	357100	-	*E. coli* SEC470, exonuclease VIII	2964/3125 (95%)	CP013962.1
hypothetical protein	2	358444	358872	+	Present in more thane 10 *E. coli* strains, superinfection exclusion protein B	414/429 (97%)	
hypothetical protein	2	359156	359434	-	none		
dnaT	2	360753	362042	+	*E. coli* GB089, phage replisome organizer	578/591 (98%), 45% coverage	CP013663.1
hypothetical protein	2	363273	363695	+	*E. coli* LY180, hypothetical protein	376/423 (89%)	CP006584.1
hypothetical protein	2	364014	365039	+	none		
hypothetical protein	2	365451	366428	+	*Edwardsiella piscicida* C07-087, hypothetical protein	879/977 (90%)	CP004141.1
hypothetical protein	2	366472	366651	+	UPEC 26-1, two-component sensor histidine kinase	155/174 (89%)	CP016497.1
hypothetical protein	2	366669	367541	+	UPEC 26-1, hypothetical protein	825/872 (95%)	CP016497.1
hypothetical protein	2	367931	369439	-	none		
hypothetical protein	2	369989	370093	+	*E. coli* 55989	90/105 (86%)	CU928145.2
NAD-dependent ADP-ribosyltransferase	2	372850	373575	+	*E. coli* strain NADC1036, putative pertussis-like toxin catalytic subunit, *ArtAB* subunit A	726/726 (100%)	KU052040.1
subtilase cytotoxin subunit B-like protein	2	373658	374101	+	*E. coli* strain NADC1036, putative exported protein toxin binding subunit, *ArtAB* subunit B	426/426 (100%)	KU052040.1
hypothetical protein	2	374422	374814	+	*E. coli* SA31, holin	393/393 (100%)	KU052039.1
hypothetical protein	2	374804	375082	+	*E. coli* NADC1036	278/279 (99%)	KU052040.1
putative peptidoglycan	2	375084	375629	+	*E. coli* NADC1036, endolysin	546/546 (100%)	KU052040.1
Heat-labile enterotoxin IIB, A chain precursor	2	375705	376484	+	*E. coli* NADC1036, LT-IIc subunit A	780/780 (100%)	KU052040.1
hypothetical protein	2	376956	377063		*E. coli* 357900	108/108 (100%)	JQ031705.1
Ash family protein	2	388187	388729	+	*E. coli* O177:H21, Ash family protein	484/555 (87%)	CP016546.1
hypothetical protein	2	389561	389722	+	none		
ail	2	389880	390416	+	none		

hypothetical protein	3	31734	32228	+	*E. coli* S51, hypothetical protein	494/495 (99%)	CP015995.1
transposase DDE domain	3	94564	95202	+	*E. coli* S51, hypothetical protein	639/639 (100%)	CP015995.1
ynjI	3	449922	450659	-	*E. coli* S51, hypothetical protein	738/738 (100%)	CP015995.1
putative autotransporter	pT22	540	2738	-	pVir68 plasmid, hypothetical protein	2145/2196 (98%)	CP001162.1
Glycosyltransferase TibC	pT22	5686	6906	-	*E. coli* pVir68 plasmid, hypothetical protein	1208/1221 (99%)	CP001162.1
hypothetical protein	pT22	18205	18570	-	*E. coli* 53638 plasmid p53638_226, transposase	360/365 (99%)	CP001064.1
hypothetical protein	pT22	18551	18823	-	*Shigella sonnei* FORC_011 plasmid pFORC11.1, hypothetical protein	273/273 (100%)	CP010830.1
fec operon regulator FecR	pT22	21437	22390	-	*E. coli* VR50, iron-dicitrate trasmembrane sensor protein RNA polymerase sigma factor FecI	951/954 (99%)	CP011134.1
putative RNA polymerase sigma factor FecI	pT22	22387	22908	-	*E. coli* ST540, RNA polymerase sigma factor FecI	519/522 (99%)	CP007391.1
tyrosine recombinase xerC	pT22	23593	24189	+	*E. coli* 210205630 tyrosine recombinase xerC	897/897 (100%)	CP015912.1
Iron-sulfur cluster ParA-like protein	pT22	26348	26980	+	*Shigella sonnei* FDAARGOS 90 unnamed plasmid 1, hypothetical protein	618/633 (98%)	CP014096.1
hypothetical protein	pT22	26980	27354	+	*E. coli* S51 plasmid pS51_1, hypothetical protein	365/375 (97%)	CP015996.1
DinI-like family protein	pT22	29047	29226	-	*E. coli* plasmid pCoo, p53638_75	236/240 (98%)	AY536429.1 CR942285.1 CP001065.1
plasmid SOS inhibition protein psiA	pT22	29791	30024	+	*E. coli* UMNK88 plasmid pUMNK88, SOS inhibition protein psiA	232/232 (100%)	CP002732.1
plasmid SOS inhibition protein psiA	pT22	30040	30510	+	*E. coli* UMNK88 plasmid pUMNK88, SOS inhibition protein psiA	471/471 (100%)	CP002732.1
hypothetical protein	pT22	31168	31809	+	*E. coli* ECC-1470 plasmid pECC-1470_100	617/643 (96%)	CP010345.1
hypothetical protein	pT22	31844	32131	+	Present in more than 10 E. coli strains, hypothetical protein	283/288 (98%)	

conjugal transfer transcriptional regulator TraJ	pT22	34861	35502	+	*Shigella dysenteriae* p80-547, E. coli pECC-1470_100, p1303_109, pEB2, pVir68, pMAR2 hypothetical protein	642/643 (99%)	KT754160.1 CP010345. CP009167. CP006000.1 CP001162.1 FM180569.1
Relaxosome protein TraY	pT22	35640	35867	+	Found in 10 *E. coli* plasmids, IncF plasmid conjugative transfer protein TraY	228/228 (100%)	
F-pilin	pT22	35901	36266	+	*E. coli* pEB2, pilin subunit	364/366 (99%)	CP006000.1
conjugal transfer protein TraP	pT22	39312	39902	+	TraP, found in more than 10 *E. coli* plasmids	589/591 (99%)	
conjugal transfer protein	pT22	39889	40086	+	Present in more than 10 *E. coli* strains, conserved hypothetical protein	198/198 (100%)	
conjugal transfer protein TrbG	pT22	40098	40349	+	TrbG, found in 9 *E. coli* plasmids	252/252 (100%)	
hypothetical protein	pT22	41210	41314	+	*E. coli* ECwhn14 plasmid pECwhn14	105/105 (100%)	CP012197.1
conjugal transfer protein TrbE	pT22	48484	48741	+	conjugal transfer proteinTrbE, found in more than 10 *E. coli* plasmid	254/258 (98%)	
conjugal transfer protein	pT22	49493	49840	+	*E. coli* MRSN388634 plasmid pMR0516mcr, trbA	345/348 (99%)	KX276657.1
hypothetical protein	pT22	49842	50117	-	*E. coli* 1303, p1303_109, hypothetical protein	276/276 (100%)	CP009167.1
conjugal transfer pilin chaperone TraQ	pT22	50198	50482	+	TraQ, p1303_109	285/285 (100%)	CP009167.1
conjugal transfer protein TrbJ	pT22	51004	51285	+	TrbJ, p1303_109	282/282 (100%)	CP009167.1
conjugal transfer protein TrbF	pT22	51272	51664	+	TrbF, p1303_109	393/393 (100%)	CP009167.1
conjugal transfer entry exclusion protein TraS	pT22	55870	56391	+	TraS, found in more than 10 *E. coli* plasmid	520/522 (99%)	
hypothetical protein	pT22	68039	68242	+	*E. coli* FAM22871 plasmid pFAM22871_1, yigA	204/204 (100%)	KU355873.1
small toxic polypeptide	pT22	70154	70360	+	*E. coli* 1303, p1303_109, modulation of SrnB protein	207/207 (100%)	CP009167.1

hypothetical protein	pT22	72033	72185	-	Present in more than 10 *E. coli* and *Salmonella* plasmids	153/153 (100%)	
hypothetical protein	pT22	73161	73250	-	pVir68 plasmid and *E. coli* 1303 p1303_109	90/90 (100%)	CP001162.1 CP009167.1
eamA	pT22	74555	75502	-	pVir68 plasmid, pagO	945/948 (99%)	CP001162.1
rob1	pT22	80673	81509	-	pVir68 plasmid, hypothetical protein	827/838 (99%)	CP001162.1

662

Figure 1.

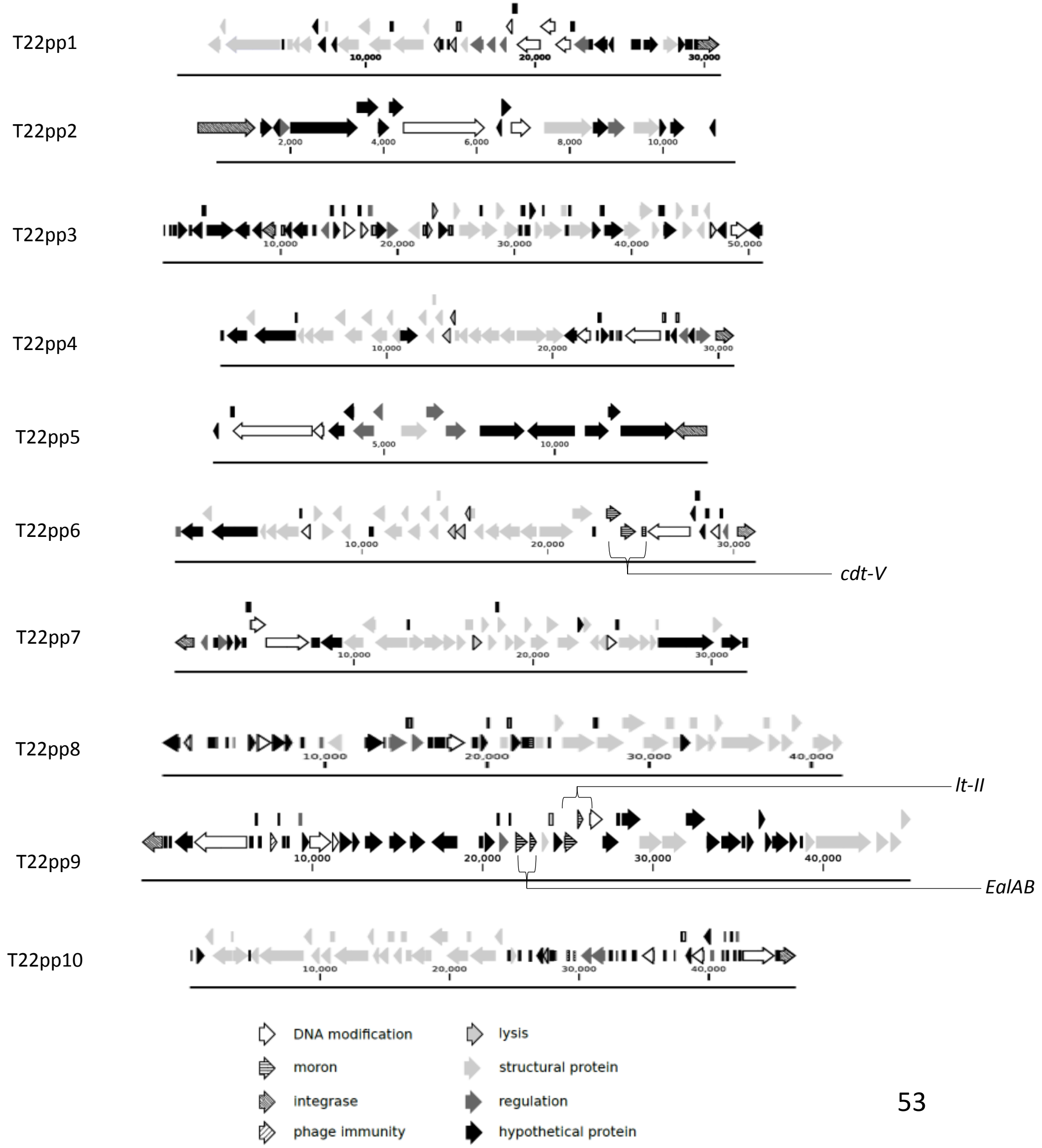

T22pp1
10,000
20,000
30,000
T22pp2
2,000
4,000
6,000
8,000
10,000
T22pp3
10,000
20,000
30,000
40,000
50,000
T22pp4
10,000
20,000
30,000
T22pp5
5,000
10,000
T22pp6
10,000
20,000
30,000
cdt-V
T22pp7
10,000
20,000
30,000
T22pp8
10,000
20,000
30,000
40,000
lt-II
T22pp9
10,000
20,000
30,000
40,000
EalAB
T22pp10
10,000
20,000
30,000
40,000
DNA modification
moron
integrase
phage immunity
lysis
structural protein
regulation
hypothetical protein

Figure 2.

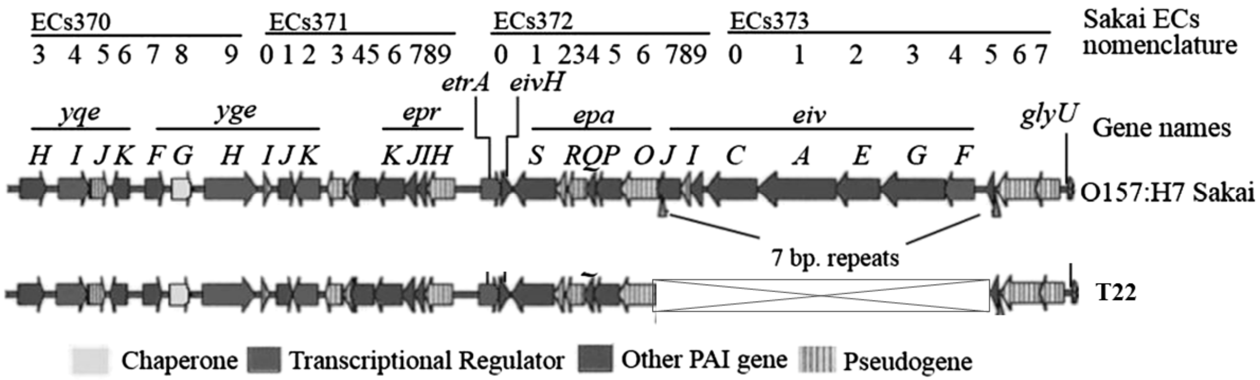


Figure 3.

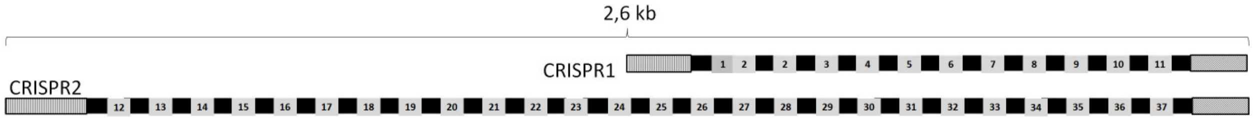


Figure 4.

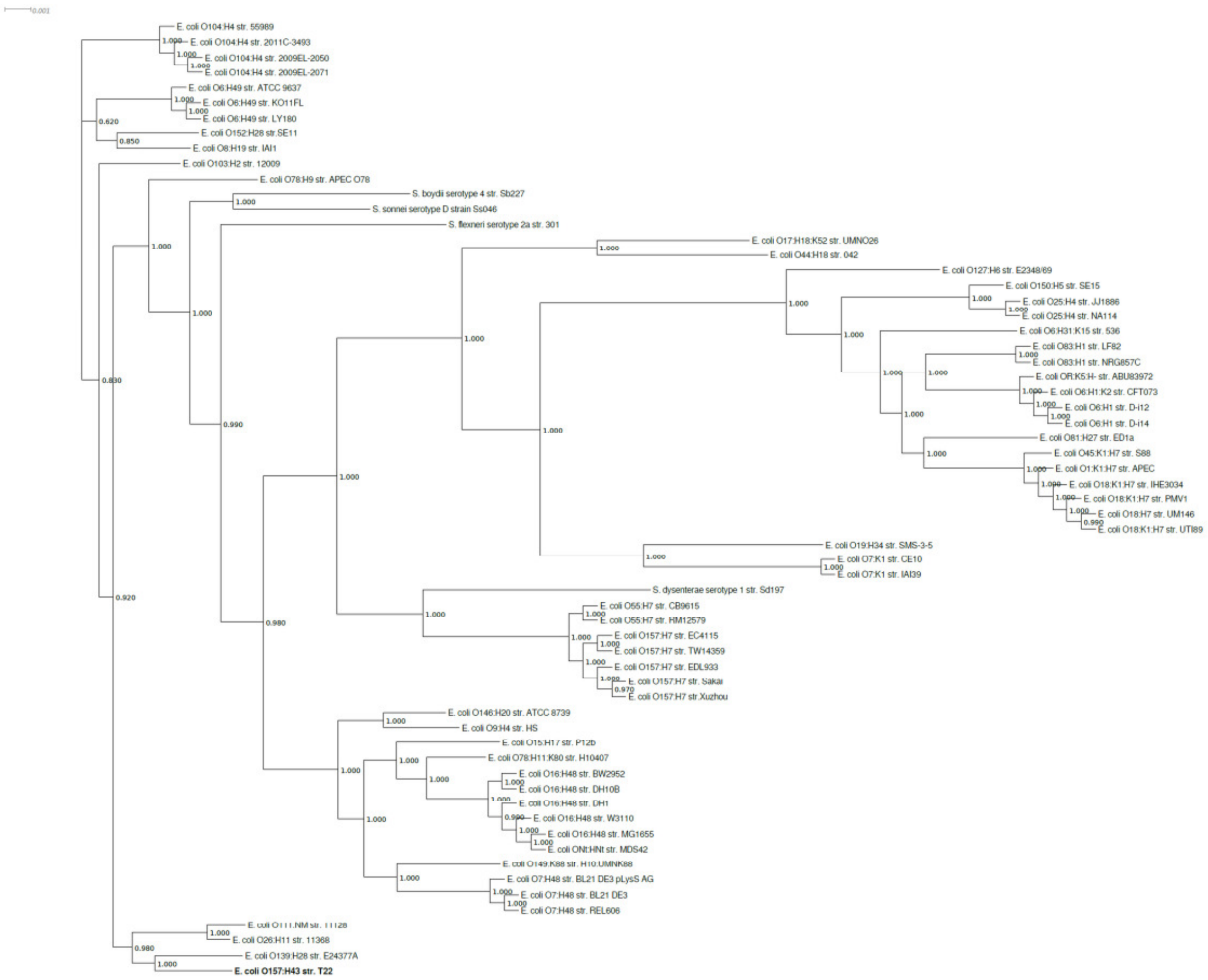
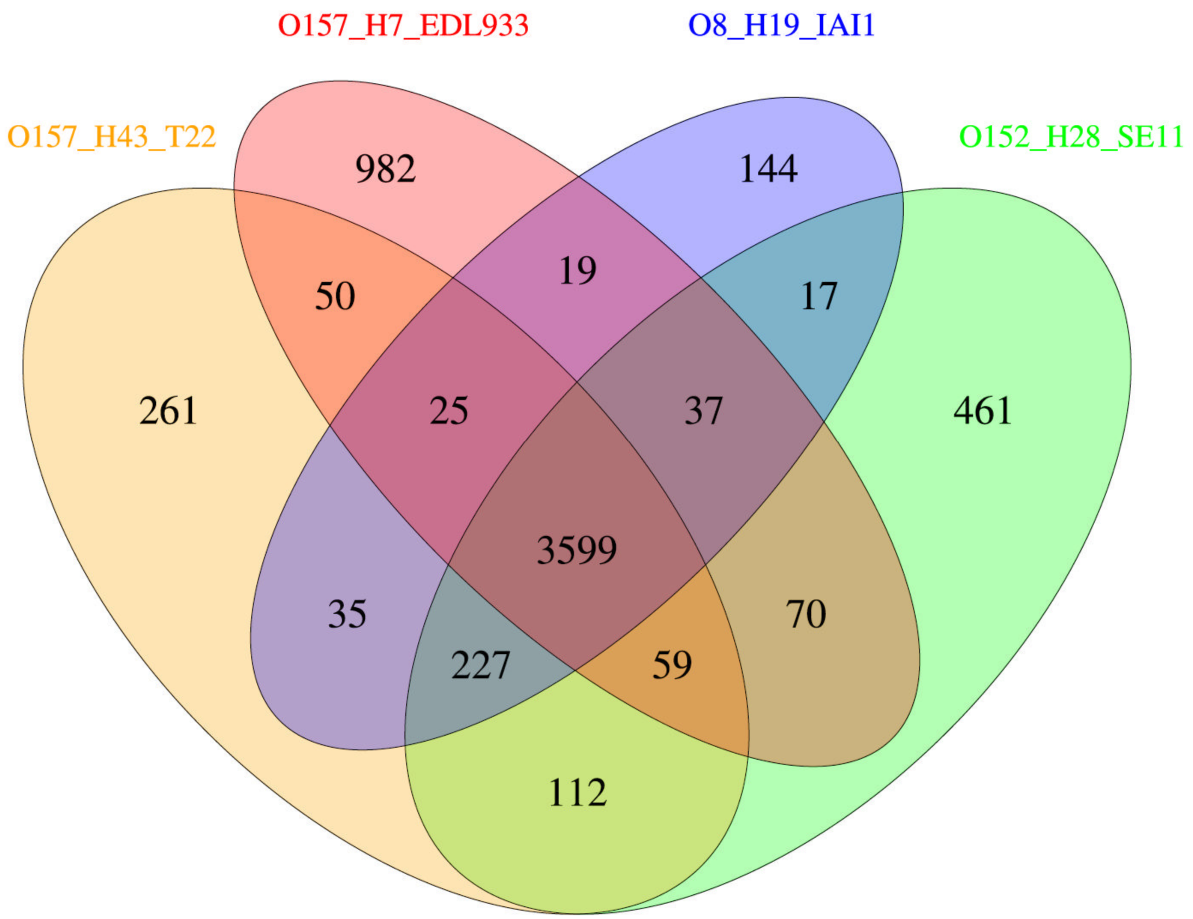

0.001
E. coli O104:H4 str. 55989
1.000
E. coli O104:H4 str. 2011C-3493
1.000
E. coli O104:H4 str. 2009EL-2050
1.000
E. coli O104:H4 str. 2009EL-2071
E. coli O6:H49 str. ATCC 9637
1.000
E. coli O6:H49 str. KO11FL
1.000
E. coli O6:H49 str. LY180
0.620
E. coli O152:H28 str.SE11
0.850
E. coli O8:H19 str. IAI1
E. coli O103:H2 str. 12009
E. coli O78:H9 str. APEC O78
S. boydii serotype 4 str. Sb227
1.000
S. sonnei serotype D strain Ss046
S. flexneri serotype 2a str. 301
E. coli O17:H18:K52 str. UMNO26
1.000
1.000
E. coli O44:H18 str. 042
E. coli O127:H6 str. E2348/69
E. coli O150:H5 str. SE15
1.000
E. coli O25:H4 str. JJ1886
1.000
E. coli O25:H4 str. NA114
1.000
1.000
E. coli O6:H31:K15 str. 536
E. coli O83:H1 str. LF82
1.000
E. coli O83:H1 str. NRG857C
1.000
1.000
E. coli OR:K5:H- str. ABU83972
1.000
E. coli O6:H1:K2 str. CFT073
1.000
E. coli O6:H1 str. D-i12
1.000
E. coli O6:H1 str. D-i14
0.830
1.000
1.000
0.990
1.000
E. coli O81:H27 str. ED1a
1.000
E. coli O45:K1:H7 str. S88
1.000
E. coli O1:K1:H7 str. APEC
1.000
E. coli O18:K1:H7 str. IHE3034
1.000
E. coli O18:K1:H7 str. PMV1
1.000
E. coli O18:H7 str. UM146
0.990
E. coli O18:K1:H7 str. UTI89
E. coli O19:H34 str. SMS-3-5
1.000
E. coli O7:K1 str. CE10
1.000
E. coli O7:K1 str. IAI39
S. dysenterae serotype 1 str. Sd197
0.920
E. coli O55:H7 str. CB9615
1.000
E. coli O55:H7 str. HM12579
1.000
0.980
1.000
E. coli O157:H7 str. EC4115
1.000
E. coli O157:H7 str. TW14359
1.000
E. coli O157:H7 str. EDL933
1.000
E. coli O157:H7 str. Sakai
0.970
E. coli O157:H7 str.Xuzhou
E. coli O146:H20 str. ATCC 8739
1.000
E. coli O9:H4 str. HS
E. coli O15:H17 str. P12b
1.000
E. coli O78:H11:K80 str. H10407
1.000
1.000
E. coli O16:H48 str. BW2952
1.000
E. coli O16:H48 str. DH10B
1.000
E. coli O16:H48 str. DH1
0.990
E. coli O16:H48 str. W3110
1.000
1.000
E. coli O16:H48 str. MG1655
1.000
E. coli ONt:HNt str. MDS42
E. coli O149:K88 str. H10:UMNK88
1.000
E. coli O7:H48 str. BL21 DE3 pLysS AG
1.000
E. coli O7:H48 str. BL21 DE3
1.000
E. coli O7:H48 str. REL606
E. coli O111:NM str. 11128
1.000
E. coli O26:H11 str. 11368
0.980
E. coli O139:H28 str. E24377A
1.000
E. coli O157:H43 str. T22

Figure 5.


O157_H7_EDL933
O8_H19_IAI1
O157_H43_T22
O152_H28_SE11
982
144
50
19
17
261
25
37
461
3599
35
70
227
59
112