We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Horizontal Gene Transfer and the Diversity of *Escherichia coli*

Maryam Javadi, Saeid Bouzari and Mana Oloomi

Additional information is available at the end of the chapter

Abstract

Escherichia coli (E. coli) strains are normal flora of human gastrointestinal tract. The evolution encoded by horizontally-transferred genetic (HGT) elements has been perceived in several species. E. coli strains have acquired virulence potential factors by attainment of particular loci through HGT, transposons or phages. The heterogeneous nature of these strains is because of HGT through mobile genetic elements. These genetic exchanges that occur in bacteria provide the genetic diversity.

A diverse enterobacterial species of E. coli is classified into (i) commensal nonpathogenic, (ii) intestinal and (iii) extraintestinal pathogenic strains. This is related to the presence or absence of regions which are associated with certain pathotypes. The genetic information belongs to the flexible E. coli genome and it has been horizontally acquired by plasmids, bacteriophages and genomic islands. The rapid evolution of E. coli variants results from the genomic regions, contribute to frequently rearrangements, excision and transfer as well as acquisition of additional genome thus resulting to the creation of new (pathogenic) variants.

HGT is a key step in the evolution of bacterial pathogens. This study is focused on determining the common virulence factors as a signature at the genetic level to use for classification, diversity and evolution of E. coli.

Keywords: *Escherichia coli* (*E. coli*), horizontal gene transfer (HGT), pathogenicity islands (PAIs), evolution, bacteriophages

1. Introduction

Escherichia coli (*E. coli*) are commensal organisms and are a part of human and animal micro-flora. Although most strains of *E. coli* are harmless, some isolates have the potential to cause severe diseases. Among the various nonpathogenic *E. coli* strains, some are able to acquire virulence determinants through the horizontal transfer of virulence genes. Based on the site of the infection and disease caused by *E. coli*, pathogenic strains are divided into major groups: extraintestinal pathogenic *E. coli* (ExPEC) and intestinal pathogenic *E. coli* (InPEC). This variability and adaptability reinforces the necessity of novel approaches to overcome pathogenic *E. coli*. Antibiotic resistance among pathogenic strains is considerable and is due to uncontrolled usage of antibiotics in human and veterinary field. Consequently, focus on modern and reverse vaccination, besides comparative genome analysis, is the most useful approach to control disease [1].

Genome evolution is the process by which the content and organization of genetic informa-tion of a species change over time. This process includes different forms of changes: point mutation and gene conversions, rearrangement (inversion or translocation), and deletion and insertion of foreign DNA (plasmid integration and transposition). These mechanisms seem to be the primary forces behind the genetic adaptation of bacterial organisms to novel environments and by which bacterial populations diverge and form separate, evolutionarily distinct species. Mechanisms of horizontal gene flux include the transmission of mobile genetic elements such as conjugative plasmids, bacteriophages, transposons, insertion ele-ments, and genomic islands, as well as the mechanism of recombination of foreign DNA into host DNA [2].

Point mutations and genetic rearrangements only lead to evolutionary development, primar-ily without creation of novel genetic determinants, while horizontal gene transfer (HGT) pro-duces extremely dynamic genomes. Thus, HGT can effectively alter the life-style of bacterial species. This is particularly true for bacterial pathogens, where virulence is linked to acquisi-tion of virulence determinants by HGT [3].

A major driving force of evolution and diversification in pathogenic bacteria compared with modification of the existing DNA is the acquisition of virulence determinants through succes-sive horizontal gene transfer [4]. The evolution of pathogenic bacteria with a strong lineage dependency often results from integration, retention, and expression of foreign DNA with a specific genomic background. In fact, parallel evolution from strains with the same genomic pathotype has occasionally emerged from multiple lineages, although the genetic mecha-nisms are not fully understood [4].

The modification of old functions and the development of new ones are required for bacte-rial evolution. The most frequent events are nucleotide exchange, insertion, and deletion. Mutation rates, per nucleotide per generation, are generally in the range of 10^{-6}–10^{-9} in bac-teria. Moreover, gene disruption, deletions, and module exchange between different genes occur at appreciable frequency. These mechanisms are common in all living organisms. They allow modification of existing functions for optimizing in a niche or adapting to a new

niche. Bacteria have no sexual life cycles, in contrast to higher organisms, to facilitate the exchange of alleles within a population. This function is fulfilled by horizontal gene transfer in bacteria; in this way the entire functional genomic unit can be imported from other sources that are not restricted by species. The DNA is transferred from less than 1 to more than 100 kb, in size. It can encode entire metabolic pathways or complex surface structures. These genes can be taken up as naked DNA or transferred in the form of plasmids, transposons, or phages [5].

2. Horizontal gene transfer (HGT)

2.1. Horizontal gene transfer (HGT) and pathogenicity islands

Subgroups of genomic islands which have a pivotal role in HGT are pathogenicity islands (PAIs). The concept of PAI was founded in the late 1980s by Jörg Hacker and colleagues in Werner Goebel's group at the University of Würzburg, Germany [2, 3].

2.1.1. The genetic features of PAIs

One or more virulence genes are carried by PAIs. There are also genomic or metabolic islands with genomic elements and characteristics similar to PAI, but lacking virulence genes. They are not present in the genome of a nonpathogenic species or a closely related species, but they are present in the genome of the same pathogenic bacterium [3].

Large genomic regions are relatively occupied by PAIs. They often differ from the core genome and the majority of PAIs are in the range of 10–200 kb in their base compositions and they also show different codon usage. It is considered that the horizontally acquired PAI still has the base composition of the donor species. On the other hand, it is also observed that the horizontally acquired DNA base composition will tend to the base composition of the recipient's genome during evolution. Further factors such as DNA topology or specific codon usage of the virulence genes in PAI may also account for the maintenance of the divergent base composition [3].

PAIs are frequently adjacent to tRNA genes. tRNA genes serve as anchor points for insertion of foreign DNA that has been acquired by horizontal gene transfer through recombination process. They are frequently associated with mobile genetic elements and they are often flanked by direct repeat (DR) sequences. PAIs delete with distinct frequencies and they are often unstable. PAI virulence functions are lost with a frequency higher than the normal rate of mutation. Integrases, transposases, and insertion sequence (IS) elements have been identified as elements that contribute to the mobilization and instability of PAIs [2].

PAIs often represent mosaic-like structures rather than the homogeneous nature of horizontally acquired DNA [2].

The islands are divided into different subtypes based on their genetic composition and also on their effects in a specific ecological niche, within a particular organism. Therefore, the same islands may have different functions [2].

2.1.2. Evolution, transfer, regulation, and integration sites of PAIs

The observation that important virulence factors are present in very similar forms in different bacteria may be explained by horizontal gene transfer. PAIs transfer via three major paths, including (i) natural transformation, (ii) plasmids, and (iii) transduction. PAI integration into the bacterial chromosome is a site-specific process. PAIs are mostly inserted at the 3′ end of tRNA loci. Furthermore, phage attachment sites are frequently located in this region. Specific genes and intergenic regions have been used by PAIs, in operons. For instance, *selC* locus is an insertion site frequently used by *E. coli* [3].

PAI genes respond to environmental signals by gene expression. PAIs are part of complex regulatory networks that include regulators encoded by the PAI itself or by other PAIs, and other global regulators elsewhere in the chromosome or by plasmids. PAI regulators can be involved in the regulation of genes located outside the PAI. Regulators mostly belong to the AraC/XylS family or to the two-component response family [2].

2.2. Horizontal gene transfer and transduction by phages

2.2.1. Evolution of bacterial pathogens by phages

Phages play an important role in the evolution and virulence of many pathogens. From common virulence factors encoded by phages in *E. coli* strains, we could mention Shiga toxin, enterohemolysin, cytolethal distending toxin, superoxide dismutase, and some outer membrane protein (OMP) [5].

The analysis of bacterial genome sequencing revealed that phages affect the bacterial genome architecture [5]. In addition, phages are important vehicles for horizontal gene exchange between different bacterial species and account for a good share of the strain-to-strain differences within the same bacterial species. In fact, two-third of gamma proteobacteria and low G+C Gram-positive bacteria harbor prophages [4, 5].

The early studies had indicated that some prophages carry additional cargo genes (lysogenic conversion genes) which are not required for the phage life cycle. Instead, a lot of DNA from phages or morons (more DNA) from prophages in pathogenic bacteria encode virulence factors. Lysogenic conversion is thought to have a great impact on the evolution of pathogenic bacteria and results in a very interesting situation of bacterium-phage coevolution.

2.2.2. Phage-mediated gene transfer

Phage-mediated horizontal gene transfer occurs via transduction or lysogenic conversion. In addition, bacterial gene disruption can occur by prophage integration into the bacterial genome.

2.2.2.1. Transduction

The phage DNA must be packed, after the phage heads are completed. In a limited frequency, DNA fragments of the host genome are packed instead of the phage DNA; however, the transduction process is quite accurate. In this process, fully functional phage particles can result, which in return deliver the packed DNA into other suitable bacteria. On the one side of transduction, the absence of phage DNA does not harm the bacterium. Instead, the injected foreign bacterial DNA can be incorporated into the genome. This is a typical example of phage-mediated horizontal gene transfer. These phages have been observed in many bacteria, for instance *Salmonella spp* [5].

2.2.2.2. Lysogenic conversion

Phages can play an important role in the emergence of new pathogens. This was recognized for Shiga toxin of *E. coli,* which is phage-encoded. Moron genes are thought to enhance phage replication when the temperate phage is residing as a prophage in the chromosome of a bacterium [5]. Moron-encoded functions enhance the fitness of the lysogen and improve the fitness of the phage. Phage-mediated horizontal transfer of virulence factors between bacteria is a very significant mechanism for evolutionary pathways. Several investigations indicate that pathogenic *E. coli* strains harbor different prophages, including P2, Mu, and lambda prophages. In comparison, phage-possessing strains grow quicker than non-lysogenic *E. coli* [5].

2.2.2.3. Gene disruption

Acquiring virulence genes is not the only mechanism by which pathogenicity develops. Pathogenic bacteria also develop from commensal bacteria by loss of genes. *Shigella* is an example of virulence by loss of *E. coli*-specific genes, including flagellar genes and *cadA*. On a small scale, prophage can cause single-gene loss when integrated into host genes. A common locus for prophage integration evolves from the tRNA genes [5].

2.3. Phages, PAIs, and plasmids in the evolution of pathogenic *E. coli* through horizontal gene transfer (HGT)

2.3.1. PAIs and E. coli pathotypes

The best understood genomic islands are PAIs, to date, which carry a cluster of virulence genes. The virulence gene products contribute to the pathogenicity of bacterium. In the case of *E. coli,* bacteria have adopted pathogenic islands to cause disease in specific environments by acquiring a foreign DNA from ancient nonpathogenic *E. coli* strains (e.g., a normal inhabitant of the gut). In enterohaemorrhagic *E. coli* (EHEC), due to the specific adaptation to different environments, a virulence-associated plasmid, one Stx-converting phage, and several PAIs have been acquired and maintained. Genomic islands may be involved in the development of specific diseases, such as diarrhea and hemolytic-uremic syndrome in a specific environment. For instance, colonization in the large intestine (EHEC), watery diarrhea in small intestine (enteropathogenic *E. coli*, EPEC), and survival and colonization in the bladder (uropathogenic

E. coli, UPEC). Such events probably led to the development of specific pathotypes of *E. coli* as mentioned above. In the case of EHEC, *stx* gene (transfer by phages), OI (O-island), and LEE (PAI) were acquired through HGT [6]. The acquisition of LEE island and *espC* gene (*E. coli* secreted protease gene) by ancient core genome led to the emergence of EPEC pathotype. Several PAIs, such as PAI-I, -II, -III, and -IV are present in the UPEC genome, indicating the occurrence of horizontal gene transfer in distinct evolutionary pathways in this particular pathotype [2].

LEE (the locus of enterocyte effacement) was initially described in EPEC strains, the causative agents of infant diarrhea in developing and industrial countries. EPEC strains are able to cause attaching-and-effacing (A/E) lesions of the microvillus brush border of enterocytes [6]. All of the genes necessary for this phenotype are located on a PAI, termed LEE. LEE is horizontally transferred and contains 41 ORFs, integrated adjacent to either the *selC*, *pheU*, or *pheV* loci. EHEC strains also possess a LEE which has 54 ORFs; 41 are common to the LEE of both EPEC and EHEC. The remaining 13 ORFs belong to a putative prophage, designated 933L (P4-like prophage), which is located close to the *selC* locus [7]. In contrast to the LEE of EPEC, cloned EHEC LEE is not able to induce A/E lesions. It has been suggested that the different phenotypes produced by LEE of EPEC and EHEC are based on natural selection for adaptation with the host microenvironment or for evasion of the host immune system [2, 6].

HPI (high pathogenicity island) was first described in *Yersinia spp*. Then it was shown in enteroinvasive (EIEC), enterotoxigenic (ETEC), enteropathogenic (EPEC), and Shiga toxin-producing (STEC) *E. coli* and extraintestinal *E. coli*. The widespread presence of HPI in different species and pathotypes also implies an efficient mechanism of horizontal transfer [6].

E. coli is the most prevalent bacterial causative agent of urinary tract infection (UTI). UPEC strains also express an array of virulence factors, which are often encoded by PAI. Alpha-hemolysin belongs to the RTX toxins with pore-forming activities in erythrocytes and other eukaryotic cells. The alpha-hemolysin operon (*hly* gene) of UPEC can be located either on a plasmid or on the chromosome [7].

In the *E. coli* species, CFT073, 536 (O6: K15), J96 (O4: K6) strains can cause urinary tract infection. When related PAI encoding alpha-hemolysin is deleted, it can lead to nonpathogenic strains from pathogenic ones. In the case of *E. coli* C5, *hly* loci is located on PAI-I and inserted in *leuX* loci. However, in other UPEC strains, several PAIs including PAI-I, PAI-II, PAI-IV, and PAI-V could harbor *hly* genes [7, 8].

Regions encoding hemolysin and P-fimbriae are named as PAI-I, -II, which are located at centisomes 82 and 97 in *E. coli* chromosome. Moreover, tRNA loci *leuX* and *selC* are also located in this region. Direct repeats (DR) 16 and 18 are adjacent to these PAIs [9]. In *E. coli* strains, besides hemolysin toxin, there is another toxin named as enterohemolysin which is encoded by a large virulence plasmid. Both of the related operons contain *hlyA*, *hlyB*, *hlyC*, and *hlyD* genes. Hemolysin and enterohemolysin are different toxins in terms of the genes that encode them and in terms of their immunological features. In contrast to hemolysin, enterohemolysin is a cell-associated toxin and it is also a member of pore-forming toxin family [9].

Since most virulence factors must be exposed on the surface of a bacterium or be secreted, many bacteria develop secretion pathways. An example of a T1SS (type 1 secretion system) encoded by a PAI is the *hly* operon of UPEC which is responsible for synthesis, activation, and transport of α-hemolysin. Genes encoding the T2SS (type 2 secretion system) belong to the core gene set; however, a large number of substrate proteins for T2SS are encoded by genes within PAI. Some of these proteins are important for pathogenesis. The gene cluster encoding T3SS (type 3 secretion system) can be found on virulence plasmid. PAI encoding T3SS includes LEE in enteropathogenic *E. coli*. Various passenger domains secreted by T5SS (type V secretion system) are referred to as autotransporters, e.g., the immunoglobulin G proteases and the VacA toxin. Examples of T5SS encoded by PAI are LPA and the EspC PAI of pathogenic *E. coli* [9].

Finally, there are many identified PAIs in different *E. coli* pathoypes which could transfer via horizontal gene transfer process and consequently led to emergence of new clinical isolates with distinct characteristics [2].

2.3.2. The role of phages in E. coli pathogenicity and evolution

Phages were acquired through horizontal gene transfer by an old nonpathogenic *E. coli* strain and led to the development of specific pathotypes. However, the genomic background has a pivotal role in evolutionary pathways. Pathogenic *E. coli* strains are classified based on repertoires of virulence factors and the most common diseases associated with them [10].

2.3.2.1. Shiga toxins

Shiga toxins (Stxs) are a family of related toxins with two major groups, Stx1 and Stx2, which are expressed by genes considered to be horizontally acquired by bacteriophages. Shiga toxin encoded by *stx1* and *stx2* genes is an A-B type toxin that inhibits protein synthesis and causes hemorrhagic colitis and hemolytic-uremic syndrome. The *stx* genes are located in the genome of heterogeneous lytic (*stx2*) or cryptic (*stx1*) lambdoid phages.

E. coli strain O157:H7 is a very prominent example of phage acquisition. *E. coli* O157:H7 is a mucosal pathogen and produces numerous pathogenic factors, of which the most significant one is phage-encoded Shiga toxin [10].

The second important factor is T3SS encoded by LEE, a PAI which is adjacent to 933L prophage. Shiga toxin is a causative agent of severe diarrhea, hemorrhagic colitis syndrome (HC), and hemolytic-uremic syndrome (HUS) in STEC strains. The Shiga toxin released by bacteria residing in the intestinal lumen is thought to be responsible for all of these symptoms. The toxin crosses the intestinal epithelial barrier, and then enters the bloodstream, damages colon vascular cells, kidneys, and the central nervous system [5, 10].

O157:H7 *E. coli* is a group of closely related strains which has diverged from enteropathogenic *E. coli* O55: H7 during the evolutionary process. The current model of emergence of toxigenic *E. coli* O157:H7 from its non-toxigenic, less virulent progenitor, *E. coli* O55:H7 relies on a number of genetic events. Key events in this process were the replacement of the *rfb* gene

region, followed by the sequential acquisition of first bacteriophage *stx2* and then phage *stx1*. Genome diversity occurred through random drift and bacteriophage-mediated events. Many strains possess only the *stx2* prophage. The *rfb* region encodes the enzymes necessary for the synthesis of the O side chains of the bacterial lipopolysaccharides. *rfb* locus is the target for frequent recombination and horizontal gene transfer [5, 11].

A large number of STEC serotypes are known. Although O157:H7 is the most important, four non-O157 STEC serotypes such as O26:H11, O103:H2, O145:H28, and O111:H8, have emerged as leading causes of infection. The major virulence factor of STEC is Shiga toxins (Stxs). STEC strains carry Stx1, Stx2, or both. Stx1 and Stx2 are divided into three (a, c, and d) and seven (a–g) subtypes, respectively. Stx phages can insert their DNA into specific chromosomal sites during infection of *E. coli* cells. They remain silent, allowing their bacterial hosts to survive as lysogenic strains [11].

Stx phages' insertion of DNA into genes can be in the basic genetic elements of the *E. coli* chromosome, in contrast to many genetic elements that are frequently integrated within tRNA genes. Nine Stx phage-insertion sites have been described. *wrbA*, codes for a tryptophan repressor-binding protein; *yehV*, which codes for a transcriptional regulator; *yecE*, whose function is unknown; *sbcB*, produces an exonuclease; *Z2577*, which codes for an oxidoreductase; *ssrA*, a tmRNA; *prfC*, that encodes peptide chain release factor 3; *argW*, a tRNA-Arg; and *torS-torT*, which is the intergenic region. The *ssrA* gene is also carries type III effector for EspK-encoding gene, which is an insertion site for EspK phages [12].

Prophages of non-O157 EHEC strains were also shown to be remarkably divergent in their structure and integration sites from those of EHEC O157 (Sakai strain). Consequently, among STEC O26:H11 strains isolated from dairy products, cattle, and human patients, a diverse range of genetic patterns was observed. Different *stx* subtypes and various insertion sites were identified among the Stx phages. Lysogenic STEC O26:H11, between human strains and strains from food and cattle was observed with some differences. These results confirm the existence of different clones of STEC O26:H11 with various levels of pathogenicity [12].

Some outbreaks in the world were noticeable because they highlighted the potential contributions of rapid whole-genome sequencing for understanding the phylogenetic origins of a new pathogen, its transmission and epidemiology, and the genetic basis for its pathogenicity. One of them is related to *E. coli* O104:H4, which ended in early July 2011 after the O104:H4 outbreak; sporadic diarrhea/HUS cases linked to *E. coli* O104:H4 have been reported. It is unknown whether these sporadic cases are derived from the new outbreak strain or it is continued transmission which was indicated. Variation in the panel of virulence factors can appear even in closely related *E. coli*. For instance, it is not clear that the pathogenicity of the O104:H4 outbreak is contributed to what extent by shared virulence factors of the sporadic isolates. Some isolates from HUS patients were compared and the most significant differences in their mobile elements were indicated. Besides variation in plasmid content, there is also substantial variation in the number and content of the prophages and genomic islands. The number of predicted prophages varies across the O104:H4 isolates, illustrating the dynamics of phage gain and loss over the relatively short evolutionary time. A key event in the evolution of fully virulent *E. coli* O104:H4 capable of causing HUS was acquisition of O104:H4-G

(the *stx2*-containing prophage). Moreover, there is a unique antibiotic resistance profile in each isolate [13].

On the whole, investigations have indicated that, the Stx phages provide an example of the rapid exchange of moron cassettes between phages from different *E. coli* strains. The two sequenced O157 strains, Sakai and EDL933, are closely related on the DNA sequence organization of the Stx2 phages, sp5 and 933W, respectively. The genome organization of phage lambda is integrated into the same chromosomal site, *wrbA*. In contrast, the Stx1 phages, sp15 and 933V from the two sequenced O157 strains, share DNA sequence only across the non-structural genes. The Stx2-encoding bacteriophage P27 from a clinical *E. coli* isolate clearly differs from the corresponding prophages in the Sakai and EDL933 strains, and it was integrated into a different *E. coli* gene *yecE*. In fact, the Stx phages provide strong evidence for the shuffling of phage modules and morons between phages from different *E. coli* strains [12].

2.3.2.2. Cytolethal distending toxins (CDTs)

First, CDTs were recognized as bacterial toxins that block the eukaryotic cell cycle, suppress cell proliferation, and eventually lead to cell death [14]. The holotoxin is a heterotrimer of three protein subunits, CdtA, CdtB, and CdtC. These are encoded by three adjacent and sometimes overlapping genes. CdtB is the active subunit, possessing DNase activity and sharing homology with the mammalian DNaseI. Genes encoding CDTs are widely disseminated among Gram-negative pathogenic bacteria, including *E. coli* strains. The production of CDTs has been associated with several pathotypes, for instance enterohemorrhagic *E. coli* (EHEC) and enteropathogenic *E. coli* (EPEC) [15].

Five different CDTs (I–V) have been reported for *E. coli* so far, and they were designated in order of publication. CDT production has been associated with pathogenic *E. coli*. *cdt-III* gene is localized into a large conjugative virulence plasmid (pVir). *cdt-I* and *cdt-IV* are encoded by lambdoid prophages, while the *cdt-V* operon is flanked by P2-like phage sequences: *repA* (replication gene A) and *Q* (a capsid gene) [16].

CDT-I and CDT-II were identified in EPEC serotype O86:H34 and O128: NM strains, respectively. CDT-III was detected in *E. coli* O15:H21 [14].

CDT-IV was detected in human and animal pathogenic *E. coli* strains of intestinal and extraintestinal origin (*E. coli* 28C, APEC O1). *cdt-IV* operon of strain 28C is flanked on both sides by prophage-related ORFs that was indicated by a DNA homology search and it encodes putative prophage-related proteins. Phage putative proteins include a lambdoid prophage host-specificity protein (*orf1*), a Lom-like protein (*orf2*), a putative tail fiber protein (*orf3*), a putative CP-933-like protease (*rorf1*), and a putative OmpT-like outer membrane protease (*rorf2*) [17]. The genome sequence of APECO1 strain revealed that the *cdt-IV* operon is framed by two prophages and pathogenicity island-associated DNA sequences, including integrase and tRNA genes. Also, genetic mosaic organization among the sequenced *cdt-I* and *cdt-IV* alleles was observed [16]. The phenomenon could also be a result of extensive genetic exchanges among different phages, which might even occur in the mammalian intestine. Similar to virulence genes, morons such as orf4, orf5, orf6, *cdt*, rorf1, and rorf2 might have spread by horizontal

gene transfer [17]. *cdt*-I and *cdt*-IV genes might have been acquired by phage transduction from a common ancestor and evolution of the CDT-encoding phages in different bacterial host-generated differences in the *cdt* genes and their flanking DNA contents [16].

The presence of CDT-V in Shiga-toxigenic *E. coli* (STEC) strains of various serotypes has been reported with clinical and nonclinical origin. CDT-V has also been associated with other strains in different serotypes and pathotypes associated with human diarrhea. The P2-like prophage sequences seem to be characteristic for the CDT-V-positive strains, with few differences that can be attributed to the adaptation processes in various hosts. The common acquisition of the *cdt*-V operon in O157: NM EHEC strains has been proposed by the presence of P2-like phage sequences after lineage divergence from the O157:H7 strains. However, there is a strong association between the presence of CDT-V and O157: NM strains [14].

These findings indicate that during evolution, while the *cdt-V* genes are rather conserved and the carrier P2-like phages became diverse, in most cases it may have resulted in loss of their mobility. Therefore, the evolutionary history of the *cdt-V* operon and its P2-like carrying prophages is proposed. Within more variable and potentially inactivated bacteriophage genomes, the highly conserved *cdt-V* operon may cause selective pressure to maintain a functional *cdt* gene cluster. Stabilization of this cargo determinant was done by inactivation of this bacteriophage genome. It is clear that further investigations of flanking regions and P2-like prophage sequences in CDT-V-positive strains will help to clarify the evolutionary background of the distribution of these variants [14–17].

The presence of *cdt* genes in different bacterial species and the analysis of DNA in the vicinity of the *cdt* genes suggest that the toxin has been acquired from heterogenic species by horizontal gene transfer. However, the probable phylogenetic origin (or ancestor) has still remained elusive. Interestingly, the phage and the corresponding insertion sequence remnants were found nearby the *E. coli cdt* genes. All these data suggest that *cdt* genes were acquired by horizontal transfer events and evolved separately since then. *stx* gene, and some types of *cdt* genes are the examples of horizontally acquired genes by phages in *E. coli*. CDT production which has been associated with some pathogenic *E. coli*, isolated from clinical diarrheal patients, suggests that the *cdt* genes are acquired independently in a number of *E. coli* lineages, possibly as a result of HGT.

2.3.3. HGT and plasmids harboring virulence genes in pathogenic *E. coli*

It is now evident that some virulent genes are located on a large plasmid (pO157) in pathogenic *E. coli*. Some of these genes are the extracellular serine protease gene (*espP*), catalase-peroxidase gene (*katP*), and type II secretion pathway protein D (*etpD*). Different sizes of this plasmid were reported that may contain some of these three genes. pO157 plasmid is mainly associated with EHEC and ETEC strains. In fact, pO157 plasmid was detected in clinical EHEC O157 isolates in 1983 for the first time. This plasmid could be detected in other atypical human EPEC. Atypical EPEC strain lacks EAF plasmid and *bfp* gene. In addition, the genome content of this plasmid in Shiga toxin-producing strains is quite divergent. However, virulence-associated genes' profile is serotype dependent [18].

EspP can be grouped into the autotransporter proteins family and characterized by catalytically active serine residue in the active center. EspP cleaves pepsin and human coagulation factor V. EspP was detected in EHEC O157 and O26 strains. Catalase-peroxidase gene (*katP*) encodes a protein which shows bi-functional catalase and peroxidase activity. This enzyme is expressed by pathogenic strains, and has been thought to protect these pathogens from oxidative damage caused by reactive oxygen molecules produced by phagocytes or other host cells during the infection process. Type II secretion pathway protein D, encoded by *etpD* gene, is another pathogenic factor encoded on the mentioned plasmid. Analysis of ~14 kb DNA derived from plasmid pO157 of EHEC strain EDL933 has indicated 13 ORFs (*etpC-etpO*) operon which had great similarity to type II secretion pathway. *etp* genes are separated from the EHEC-*hlyC* genes with an *IS911*-like insertion element. The existence of IS elements indicates that *etp* gene cluster could be exchangeable between EHEC plasmids [18].

In pathogenic bacteria, there are S-fimbriae and P-fimbriae as different types of fimbriae in uropathogenic *E. coli,* which can cause attachment to host cells. The *pap* gene cluster includes 14 genes encoding P-fimbriae. A cluster of six genes, termed *sfp* including *sfpA* gene, is located on a large plasmid, pSFO157, in some of the pathogenic *E. coli* strains. This genomic cluster mediates mannose-resistant hemagglutination and expression of a novel type of fimbriae, Sfp fimbriae, which is 3–5 nm in diameter, the major subunit of which is SfpA. Sorbitol-fermenting (SF) EHEC O157: H⁻, which is one of the HUS disease agents, harbors pSFO157 plasmid. The *sfp* gene cluster is surrounded by IS elements, and the origin of plasmid replication proves that these genes might be acquired by HGT. There are nine ORFs (open Reading frames) in the *sfp* gene cluster. The upstream region of *sfpA* is homologous to *IS*1294 and *IS*2 insertion sequences besides Tn2501-like sequence and *IS*100, which is integrated in *IS*3. Moreover, in iso-*IS*1-like sequence, *IS*3 and *IS*100 regions possess DR (directed repeat) sequences indicating the regular transposition process in plasmids [18–21].

3. Diversity of *E. coli* strains

The coevolution of bacterial pathogens related to genetic elements, including pathogenicity islands and phages encoding virulence factors, has been observed in several species. *E. coli* is a member of the normal intestinal microflora of humans and animals. However, certain *E. coli* strains have acquired virulence-associated factors via horizontal gene transfer, enabling the bacteria to colonize its host and cause disease. Pathogenic *E. coli* utilize particular strategies to penetrate into host cells and tissues. Special enzymes and pathogenic factors are known as unique virulence determinants of each bacterium. However, each bacterium has a unique genomic background of its own chromosome, such as *fliC* and *fimH* genes that encode the main subunit of flagella and fimbriae type I, respectively. Both of them are known as virulence-associated factors and interfere in pathogenicity of *E. coli* [19].

Investigations have revealed, *hly* gene which is located on a PAI and encodes alpha-hemolysin, is frequently detected in *cdt*-III- and *cdt*-IV-producing human and animal pathogenic *E. coli* strains.

The prevalence of *hly* gene in *cdt*-type III, IV, and V was more than other isolates. These results demonstrate that, possibly, there could be a relationship between the existence of *hly* gene and the type of *cdt* gene in clinical *E. coli* isolates. *fliC* and *fim* are two chromosomally located genes that can be defined as the genomic background of *E. coli* strains. The *fliC* gene encodes the flagellin subunit; type 1 fimbriae are also encoded by the chromosomally located *fim* gene cluster. The presence of *fim* DNA sequences is common among *E. coli* strains. In fact, the majority of clinical isolates, both virulent and non-virulent, could be induced to express type 1 fimbriae [19].

Among pathogenic *E. coli*, the existence of a large virulent plasmid (pO157) has also been observed. The *etpD*, *katP*, and *espP* genes are located on this plasmid. The pO157 plasmid is mainly associated with EHEC and ETEC strains. There is a relation between the occurrence of *stx* genes and these virulent plasmid-associated genes. Moreover, PCR analysis revealed a close relationship between the occurrence of plasmid-born *katP* gene and *stx* gene in pathogenic *E. coli*. Most of the *katP*+ strains belong to Shiga toxin-producing *E. coli*. The *katP* gene is mostly present in CDT-I- and CDT-II-producing strains. EspP, which possesses human coagulation factor V and pepsin A proteolytic activity, is a significant marker of virulence in Shiga toxin-producing strains. In CDT-III-producing isolates, high frequency of *espP gene* is considerable. Alpha-hemolysin is frequently associated with human uropathogenic *E. coli* (UPEC); furthermore, related encoding PAI is also unstable and the operon could be located on either a plasmid or the chromosome. Besides, urinary tract infection (UTI) is caused predominantly by type 1fimbriated UPEC and initial binding is mediated by the FimH adhesin of the mentioned fimbriae. Investigation showed that most of the *hly*+ strains harbor *fimH* gene. In addition, all *hly*+ strains possess one or more of plasmid pO157 genes, including *etpD*, *katP*, and *espP*. These genes plus the *stx* gene are among the EHEC and STEC characteristics, although the *espP* gene is common in EPEC and EHEC. Simultaneous presence of these genes indicates that clinical isolates obtain *hly* operon and relevant PAI. In addition, in evolutionary pathways, isolates improve their pathogenicity by achieving the *cdt* genes. Studies demonstrate that virulence genes from CDT-producing strains belong to the heterogeneous group. Strains which are clustered as particular groups have similar characteristics, while possessing their own unique genotype and genomic content. For instance, each distinct *cdt*-type group, by possessing a particular *cdt* gene as genomic backbone, has an approximately similar pattern based on other virulence genes [19].

This evidence further confirms that horizontal gene transfer could occur among pathogenic strains. Moreover, findings indicate that CDT-producing strains may have originated from a common ancestor during their evolution by HGT, and they departed from each other [17].

CDT-producer strains did not show particular phylogenomic relation and pattern. Indeed, they might carry the same or similar virulence gene sets, but remarkably possess their own divergent genomic structure. This is probably because of their complex and distinct evolutionary pathways, indicating independent acquisition of mobile genetic elements that have driven from their evolution [19]. Furthermore, it was shown that there are different types of CDTs that are encoded by prophages, plasmids, and/or pathogenicity islands that result in different types of CDTs through HGT in different origins [7, 17, 19, 22].

In the recent years, whole-genome sequences for many bacteria have become accessible. It improves our understanding about virulence-associated genes and horizontal gene transfer from the emergence of new pathogens aspects. Some pathogens like *E. coli* could acquire virulence genes via HGT. On the other hand, from the diagnostic point of view, virulent gene examination could improve our knowledge about different pathotypes' detection and classification.

Phage-related and virulence-associated factors transferred by phages were found to be prevalent signature proteins. The signature proteins identified include several individual phage proteins (holins, nucleases, terminases, and transferases) and multiple members of different protein families (the lambda family, phage-integrase family, phage-tail tape protein family, putative membrane proteins, regulatory proteins, restriction-modification system proteins, tail fiber-assembly proteins, base plate-assembly proteins, and other prophage tail-related proteins).

4. Conclusions and the way forward

The heterogeneous nature of strains could be because of the HGT through mobile genetic elements. The genetic exchanges that occur in bacteria provide genetic diversity and versatility. Plasmids, bacteriophages, and genomic islands belong to the flexible *E. coli* genome and their genetic information can be horizontally acquired. The rapid evolution of *E. coli* variants contributes to these genomic regions as they are subject to rearrangements, excision, and transfer frequently. The creation of new pathogenic variants is the result of further acquisition of additional genome.

The accumulating amount of sequence information generated in the era of "genomics" helps to increase our understanding of factors and mechanisms that are involved in diversification of this new bacterial species, as well as in those that may direct host-specificity. From a comparative genomic aspect, a significant challenge is to utilize bulky amount of datasets to distinguish and conceptualize specific sequence signatures that scientifically or diagnostically are applicable traits. By comparing more sequence data from different strains, new signature biomarkers will be recognized for use as vaccines or as diagnostic factors in future. Signature conserved proteins in a wide range of pathogenic bacterial strains can potentially be used in modern vaccine-design strategies.

Author details

Maryam Javadi, Saeid Bouzari and Mana Oloomi*

*Address all correspondence to: manaoloomi@yahoo.com

Molecular Biology Department, Pasteur Institute of Iran, Tehran, Iran

References

[1] Moriel DG, Rosini R, Seib KL, Serino L, Pizza M, Rappuoli R. *Escherichia coli*: Great diversity around a common core. MBio. 2012;**3**(3):e00118-12

[2] Schmidt H, Hensel M. Pathogenicity islands in bacterial pathogenesis. Clinical Microbiology Reviews. 2004;**17**(1):14-56

[3] Schubert S, Darlu P, Clermont O, Wieser A, Magistro G, Hoffmann C, Weinert K, Tenaillon O, Matic I, Denamur E. Role of intraspecies recombination in the spread of pathogenicity islands within the *Escherichia coli* species. PLoS Pathogens. 2009;**5**(1):e1000257

[4] Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, Kodama T, Abe H, Nakayama K, Kurokawa K, Tobe T, Hattori M, Hayashi T. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America. 2009;**106**(42):17939-17944

[5] Brüssow H, Canchaya C, Hardt WD. Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. Microbiology and Molecular Biology Reviews. 2004;**68**(3):560-602

[6] Mohammadzadeh M, Oloomi M, Bouzari S. Genetic evaluation of locus of enterocyte effacement pathogenicity island (LEE) in enteropathogenic *Escherichia coli* isolates (EPEC). Iran Journal of Microbiology. 2013;**5**(4):345-349

[7] Oloomi M, Bouzari S. Molecular profile and genetic diversity of cytolethal distending toxin (CDT)-producing *Escherichia coli* isolates from diarrheal patients. APMIS. 2008;**116**(2):125-132

[8] Bidet P, Bonacorsi S, Clermont O, De Montille C, Brahimi N, Bingen E. Multiple insertional events, restricted by the genetic background, have led to acquisition of pathogenicity island IIJ96-like domains among *Escherichia coli* strains of different clinical origins. Infection and Immunity. 2005;**73**(7):4081-4087

[9] Taneike I, Zhang HM, Wakisaka-Saito N, Yamamoto T. Enterohemolysin operon of Shiga toxin-producing *Escherichia coli*: A virulence function of inflammatory cytokine production from human monocytes. FEBS Letters. 2002;**524**(1-3):219-224

[10] Gyles CL. Shiga toxin-producing *Escherichia coli*: An overview. Journal of Animal Sciences. 2007;**85**(13 Suppl):E45-62

[11] Kyle JL, Cummings CA, Parker CT, Quiñones B, Vatta P, Newton E, Huynh S, Swimley M, Degoricija L, Barker M, Fontanoz S, Nguyen K, Patel R, Fang R, Tebbs R, Petrauskene O, Furtado M, Mandrell RE. *Escherichia coli* serotype O55:H7 diversity supports parallel acquisition of bacteriophage at Shiga toxin phage insertion sites during evolution of the O157:H7 lineage. Journal of Bacteriology. 2012;**194**(8):1885-1896

[12] Bonanno L, Loukiadis E, Mariani-Kurkdjian P, Oswald E, Garnier L, Michel V, Auvray F. Diversity of Shiga toxin-producing *Escherichia coli* (STEC) O26:H11 strains examined via Stx subtypes and insertion sites of Stx and EspK bacteriophages. Applied Environmental Microbiology. 2015;**81**(11):3712-3721

[13] Grad YH, Godfrey P, Cerquiera GC, Mariani-Kurkdjian P, Gouali M, Bingen E, Shea TP, Haas BJ, Griggs A, Young S, Zeng Q, Lipsitch M, Waldor MK, Weill FX, Wortman JR, Hanage WP. Comparative genomics of recent Shiga toxin-producing *Escherichia coli* O104:H4: Short-term evolution of an emerging pathogen. MBio. 2013;**4**(1):e00452-12

[14] Sváb D, Horváth B, Maróti G, Dobrindt U, Tóth I. Sequence variability of P2-like pro-phage genomes carrying the cytolethal distending toxin V operon in *Escherichia coli* O157. Applied Environmental Microbiology. 2013;**79**(16):4958-4964

[15] Tóth I, Nougayrède JP, Dobrindt U, Ledger TN, Boury M, Morabito S, Fujiwara T, Sugai M, Hacker J, Oswald E. Cytolethal distending toxin type I and type IV genes are framed with lambdoid prophage genes in extraintestinal pathogenic *Escherichia coli*. Infection and Immunity. 2009;**77**(1):492-500

[16] Asakura M, Hinenoya A, Alam MS, Shima K, Zahid SH, Shi L, Sugimoto N, Ghosh AN, Ramamurthy T, Faruque SM, Nair GB, Yamasaki S. An inducible lambdoid prophage encoding cytolethal distending toxin (Cdt-I) and a type III effector protein in entero-pathogenic *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America. 2007;**104**(36):14483-14488

[17] Javadi M, Oloomi M, Bouzari S. Genotype cluster analysis in pathogenic *Escherichia coli* isolates producing different CDT types. Journal of Pathogens. 2016;**2016**:9237127

[18] Brunder W, Khan AS, Hacker J, Karch H. Novel type of fimbriae encoded by the large plasmid of sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:H(−). Infections and Immunity. 2001;**69**(7):4447-4457

[19] Oloomi M, Javadi M, Bouzari S. Presence of pathogenicity island-related and plasmid encoded virulence genes in cytolethal distending toxin producing *Escherichia coli* iso-lates from diarrheal cases. International Journal of Applied Basic Medical Research. 2015;**5**(3):181-186

[20] Brunder W, Schmidt H, Karch H. KatP, a novel catalase-peroxidase encoded by the large plasmid of enterohaemorrhagic *Escherichia coli* O157:H7. Microbiology. 1996;**142**(Pt11):3305-3315

[21] Brunder W, Schmidt H, Karch H. EspP, a novel extracellular serine protease of entero-haemorrhagic *Escherichia coli* O157:H7 cleaves human coagulation factor V. Molecular Microbiology. 1997;**4**(4):767-778

[22] Bouzari S, Oloomi M, Oswald E. Detection of the cytolethal distending toxin locus cdtB among diarrheagenic *Escherichia coli* isolates from humans in Iran. Research in Micro-biology. 2005;**156**(2):137-144