JOURNAL OF BACTERIOLOGY, Nov. 2010, p. 5822–5831 Vol. 192, No. 21
0021-9193/10/$12.00 doi:10.1128/JB.00710-10


# A Commensal Gone Bad: Complete Genome Sequence of the Prototypical Enterotoxigenic *Escherichia coli* Strain H10407▽†

Lisa C. Crossman,[1]‡§ Roy R. Chaudhuri,[2]§ Scott A. Beatson,[3] Timothy J. Wells,[4] Mickael Desvaux,[4]¶ Adam F. Cunningham,[4] Nicola K. Petty,[1] Vivienne Mahon,[5] Carl Brinkley,[6] Jon L. Hobman,[7] Stephen J. Savarino,[8] Susan M. Turner,[4] Mark J. Pallen,[9] Charles W. Penn,[9] Julian Parkhill,[1] A. Keith Turner,[1] Timothy J. Johnson,[10] Nicholas R. Thomson,[1] Stephen G. J. Smith,[5] and Ian R. Henderson[4]*

*The Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridge, United Kingdom[1]; Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom[2]; School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane, Australia[3]; School of Immunity and Infection[4] and School of Biosciences,[9] University of Birmingham, Birmingham, United Kingdom; Department of Clinical Microbiology, School of Medicine, Trinity College Dublin, Dublin, Ireland[5]; Department of Enteric Infections, Walter Reed Army Institute of Research,[6] and Enteric Disease Department, Naval Medical Research,[8] Silver Spring, Maryland; School of Biosciences, The University of Nottingham, Sutton Bonington, United Kingdom[7]; and Department of Veterinary and Biomedical Sciences, University of Minnesota, Saint Paul, Minnesota[10]*

Received 18 June 2010/Accepted 17 August 2010

**In most cases, *Escherichia coli* exists as a harmless commensal organism, but it may on occasion cause intestinal and/or extraintestinal disease. Enterotoxigenic *E. coli* (ETEC) is the predominant cause of *E. coli*-mediated diarrhea in the developing world and is responsible for a significant portion of pediatric deaths. In this study, we determined the complete genomic sequence of *E. coli* H10407, a prototypical strain of enterotoxigenic *E. coli*, which reproducibly elicits diarrhea in human volunteer studies. We performed genomic and phylogenetic comparisons with other *E. coli* strains, revealing that the chromosome is closely related to that of the nonpathogenic commensal strain *E. coli* HS and to those of the laboratory strains *E. coli* K-12 and C. Furthermore, these analyses demonstrated that there were no chromosomally encoded factors unique to any sequenced ETEC strains. Comparison of the *E. coli* H10407 plasmids with those from several ETEC strains revealed that the plasmids had a mosaic structure but that several loci were conserved among ETEC strains. This study provides a genetic context for the vast amount of experimental and epidemiological data that have been published.**

Current dogma suggests the Gram-negative motile bacterium *Escherichia coli* colonizes the infant gut within hours of birth and establishes itself as the predominant facultative anaerobe of the colon for the remainder of life (3, 59). While the majority of *E. coli* strains maintain this harmless existence, some strains have adopted a pathogenic lifestyle. Contemporary tenets suggest that pathogenic strains of *E. coli* have acquired genetic elements that encode virulence factors and enable the organism to cause disease (12). The large repertoire of virulence factors enables *E. coli* to cause a variety of clinical manifestations, including intestinal infections mediating diarrhea and extraintestinal infections, such as urinary tract infections, septicemia, and meningitis. Based on clinical manifestation of disease, the repertoire of virulence factors, epidemiology, and phylogenetic profiles, the strains causing intestinal infections can be divided into six separate pathotypes, *viz*., enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC), enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), diffuse adhering *E. coli* (DAEC), and enterotoxigenic *E. coli* (ETEC) (33, 35, 39).

ETEC is responsible for the majority of *E. coli*-mediated cases of human diarrhea worldwide. It is particularly prevalent among children in developing countries, where sanitation and clean supplies of drinking water are inadequate, and in travelers to such regions. It is estimated that there are 200 million incidences of ETEC infection annually, resulting in hundreds of thousands of deaths in children under the age of 5 (55, 64). The essential determinants of ETEC virulence are traditionally considered to be colonization of the host small-intestinal epithelium via plasmid-encoded colonization factors (CFs) and subsequent release of plasmid-encoded heat-stable (ST) and/or heat-labile (LT) enterotoxins that induce a net secretory state leading to profuse watery diarrhea (20, 62). More recently, additional plasmid-encoded factors have been implicated in the pathogenesis of ETEC, namely, the EatA serine protease autotransporter (SPATE) and the EtpA protein, which acts as an intermediate in the adhesion between bacterial flagella and host cells (23, 32, 42, 46). Furthermore, a

* Corresponding author. Mailing address: School of Immunity and Infection, University of Birmingham, Birmingham, B15 2TT, United Kingdom. Phone: 44 121 4144368. Fax: 44 121 4143599. E-mail: i.r.henderson@bham.ac.uk.

‡ Present address: Genome Analysis Centre, Norwich, United Kingdom.

¶ Present address: INRA, UR454 Microbiology, F-63122 Saint-Genès Champanelle, France.

§ L.C.C. and R.R.C. contributed equally to this investigation.

† Supplemental material for this article may be found at http://jb.asm.org/.

▽ Published ahead of print on 27 August 2010.

TABLE 1. General characteristics of three sequenced *E. coli* chromosomes

| Characteristic | Value | | |
|---|---|---|---|
| | H10407 | K-12 | HS |
| Etiology | Pathogen | Laboratory strain | Commensal |
| Length (bp) | 5,153,435 | 4,643,538 | 4,686,137 |
| GC content (%) | 50.8 | 50.8 | 50.8 |
| Total no. of CDSs | 4,746 | 4,384 | 4,200 |
| tRNA genes | 87 | 86 | 86 |
| rRNA genes | 7 | 7 | 7 |

number of chromosomal factors are thought to be involved in virulence, e.g., the invasin Tia; the TibA adhesin/invasin; and LeoA, a GTPase with unknown function (14, 21, 22). *E. coli* H10407 is considered a prototypical ETEC strain; it expresses colonization factor antigen 1 (CFA/I) and the heat-stable and heat labile toxins. Loss of a 94.8-kb plasmid encoding CFA/I and a gene for ST enterotoxin from *E. coli* strain H10407 leads to reduced ability to cause diarrhea (17).

Here, we report the complete genome sequence and virulence factor repertoire of the prototypical ETEC strain H10407 and the nucleotide sequence and gene repertoire of the plasmids from ETEC strain E1392/75, and we describe a novel conserved secretion system associated with the sequenced ETEC strains.

## MATERIALS AND METHODS

**Bacterial strains and sequencing.** The ETEC O78:H11:K80 strain H10407 was isolated from an adult with cholera-like symptoms in the course of an epidemiologic study in Dacca, Bangladesh, prior to 1973 (19) and was shown to cause diarrhea in adult volunteers (6, 17). The *E. coli* H10407 isolate that was sequenced was from the Walter Reed Army Institute of Research (WRAIR) cGMP stock manufactured in February 1998 as lot 0519. The whole genome was sequenced to a depth of 8× coverage from pUC19 (insert size, 2.8 to 5 kb) and pMAQ1b (insert size, 5.5 to 10 kb) small-insert libraries. Sanger sequencing was carried out using Amersham Big Dye (Amersham, United Kingdom) terminator chemistry on ABI3700 sequencing machines. End sequences from larger-insert plasmid (pBACe3.6; 20- to 30-kb insert size) libraries were used as a scaffold. Sequence reads were assembled into contigs with Phrap (P. Green, unpublished data) and finished using GAP4, as described previously (33). The plasmids from the ETEC O6:H16:K15 strain E1392/75, which was isolated from a patient in Hong Kong with diarrhea, express the CFA/II (CS1 and CS3) colonization factors and produce the ST and LT toxins and were also sequenced using a similar approach (7, 50, 60). Plasmid DNA for ETEC E1392/75 was provided by Acambis United Kingdom.

**Gene prediction, annotation, and comparative analysis.** Annotation was carried out using the genome viewer Artemis (47). Coding sequences were predicted using the gene prediction programs Orpheus (26), Glimmer2 (11), and Glimmer3 (10) and then manually curated. Protein domains were marked up using Pfam (48), and transmembrane domains and signal sequences were predicted using TMHMM and SignalP, respectively (15, 37). Annotation was transferred from previously annotated *E. coli* genomes to orthologous genes and manually curated. A homologue was considered to be present if a hit was found with >60% identity over at least 80% of the length of the query protein. Regions of difference (ROD) and plasmids were annotated and curated manually.

**Nucleotide sequence accession numbers.** The annotated genome sequence of ETEC H10407 and the plasmids from ETEC H10407 and E1392/75 have been deposited in the EMBL databases (accession number FN649414 for the complete ETEC H10407 chromosome; Tables 1 and 2 list the general features of the nucleotide sequences and accession numbers for the plasmids).

## RESULTS AND DISCUSSION

**Structure and general features of the ETEC H10407 chromosome.** The ETEC H10407 genome consists of a circular chromosome of 5,153,435 bp and four plasmids designated pETEC948, pETEC666, pETEC58, and pETEC52. The general features of the ETEC H10407 chromosome are presented in Table 1 and the plasmids in Table 2. We identified 4,746 protein-coding genes (CDSs) in the chromosome, 33 (0.67%) of which did not have any match in the database, while 579 (11.67%) encoded conserved hypothetical proteins with no known function and 503 (10.14%) were genes associated with mobile elements, such as integrases or transposases, or were phage related. We have identified 25 ROD that occur in the ETEC H10407 genome and are differentially distributed among the other sequenced *E. coli* chromosomes (Fig. 1; see Table S1 in the supplemental material). The combined size of these ROD is 755,359 bp (14.7% of the chromosome) and includes nine prophages, designated ETP29, -33, -86, -128, -216, -284, -295, -468, and -507, where the numeric designations denote their approximate positions (times 10,000 bp) on the chromosome. None appeared to carry cargo genes related to virulence.

**Comparative genomics of the ETEC H10407 chromosome.** Previously, a phylogeny was constructed based on the concatenated sequences of 2,173 genes that are conserved in all *E. coli* strains and in *Escherichia albertii* and *Escherichia ferguso-*

TABLE 2. General characteristics of the plasmids from ETEC strains H10407 and E1392/75

| Characteristic | Value in *E. coli*: | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | H10407 | | | | E1392/75 | | | | |
| Plasmid | pETEC948 | pETEC666 | pETEC58 | pETEC52 | pETEC1018 | pETEC746 | pETEC557 | pETEC75 | pETEC62 |
| Accession no. | FN649418 | FN649417 | FN649416 | FN649415 | FN822745 | FN822748 | FN822746 | FN822749 | FN822747 |
| Size (bp) | 94,797 | 66,681 | 5,800 | 5,175 | 101,857 | 74,575 | 55,709 | 7,497 | 6,222 |
| No. of CDSs | 115 | 88 | 7 | 6 | 165 | 117 | 73 | 9 | 13 |
| Rep | RepFIIA | RepFIIA | ColE2 | ColE1 | RepFIIA | RepI1 | RepFIB/RepI1 | ColE1 | ND[a] |
| Stability genes | StbAB, PsiAB, SopAB, YacAB, RelE | StbAB, PsiAB, Mok/Hok | | | StbAB, PsiAB, CcdAB | StbAB, NikAB | SopAB, PsiAB | | |
| Insertion elements | IS*1*, IS*2*, IS*3*, IS*66*, IS*91*, IS*100*, IS*629*, IS*911*, IS*1414*, IS*Ec10*, IS*Ec12*, IS*Sfl4*, Tn*3* | IS*1*, IS*21*, IS*66*, IS*600*, IS*1294*, IS*Ec8* | | | IS*1*, IS*2*, IS*3*, IS*21*, IS*30*, IS*66*, IS*91*, IS*100*, IS*629*, IS*630*, IS*639*, IS*911*, IS*1414*, IS*Shdy1* | IS*2*, IS*100*, IS*186*, IS*1328* | IS*1*, IS*30*, IS*66*, IS*100*, IS*911*, IS*Shdy1* | IS*100* | IS*CR2* |

[a] ND, not determined. pETEC62 has a gene conserved among many small plasmids that is annotated as a "probable replication initiation protein," but no experimental evidence exists for this function.

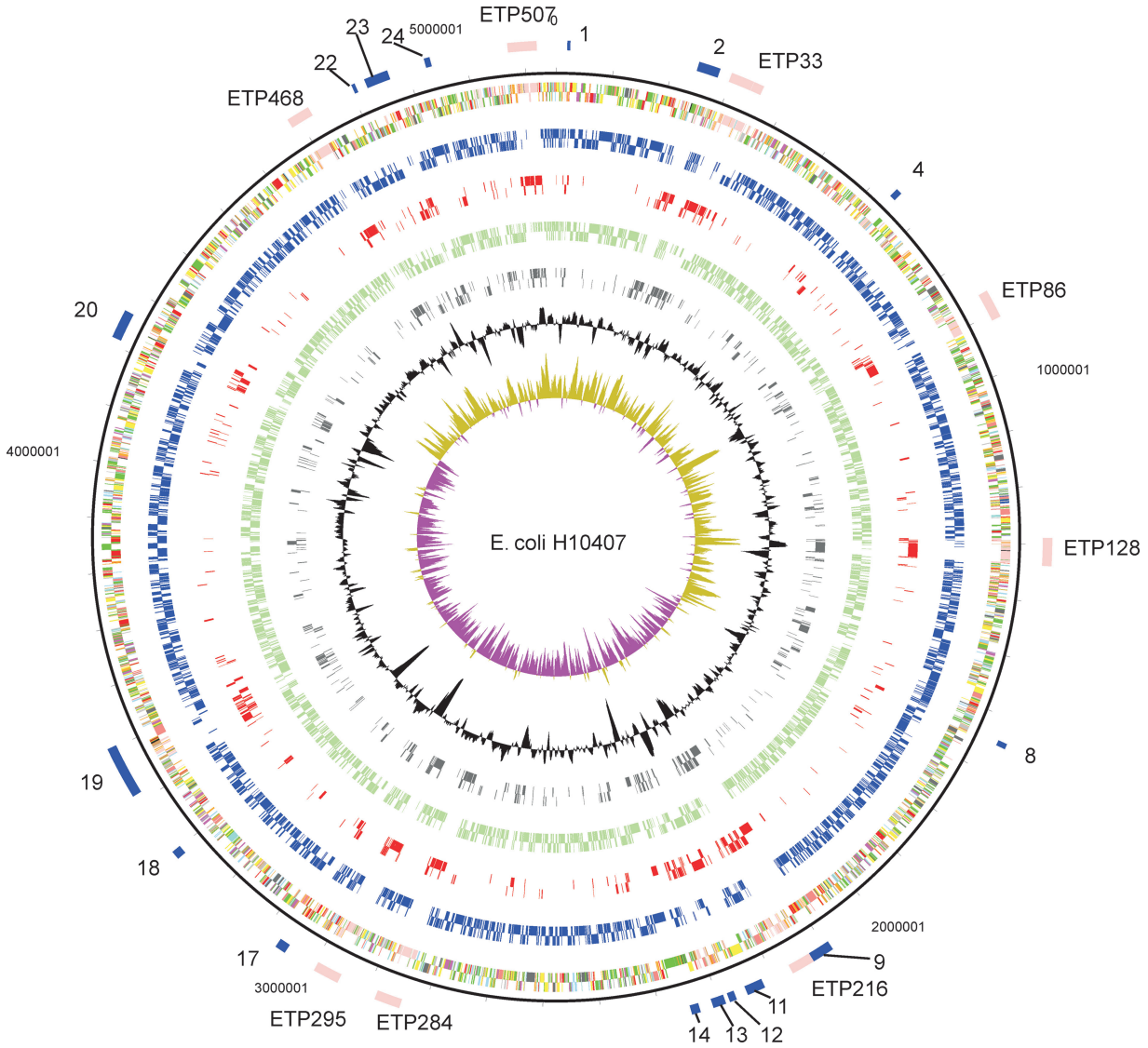


FIG. 1. Circular representation of the *E. coli* H10407 chromosome. From the outside in, the outer circle 1 marks the positions of regions of difference (mentioned in the text), including prophage (light pink), as well as regions differentially present in other *E. coli* strains (blue) (see table S1 in the supplemental material). Circle 2 shows the sizes in bp. Circles 3 and 4 show the positions of CDSs transcribed in clockwise and counterclockwise directions, respectively. Genes in circles 3 and 4 are color coded according to the functions of their gene products: dark green, membrane or surface structures; yellow, central or intermediary metabolism; cyan, degradation of macromolecules; red, information transfer/cell division; cerise, degradation of small molecules; pale blue, regulators; salmon pink, pathogenicity or adaptation; black, energy metabolism; orange, conserved hypothetical; pale green, unknown; brown, pseudogenes. Circles 5 and 6 and circles 9 and10 show the positions of *E. coli* H10407 genes that have orthologues (by reciprocal FASTA analysis) in *E. coli* K-12 MG1655 (blue) or *E. coli* 042 (green), respectively. Circles 7 and 8 and circles 11 and 12 show the positions of genes unique to *E. coli* H10407 compared to *E. coli* K-12 MG1655 (red) or *E. coli* 042 (gray), respectively. Circle 13 shows a plot of G+C contents (in a 10-kb window). Circle 14 shows a plot of GC skew ([G − C]/[G + C], in a 10-kb window).

*nii*, which were included as outgroup sequences (4). The established *E. coli* subgroups (A, B1, B2, D, and E) are all monophyletic, with the exception of group D, which is divided at the root. In agreement with previous optical-mapping experiments (5), *E. coli* H10407 is located in the A subgroup with the nonpathogenic laboratory strains *E. coli* K-12 and C and the nonpathogenic commensal isolate *E. coli* HS. The majority of commensal strains of bacteria belong to the A subgroup (59).

Comparison of *E. coli* H10407 with the closely related nonpathogenic *E. coli* K-12, C, and HS strains revealed that these chromosomes are largely colinear (see Fig. S1 in the supplemental material) and that the *E. coli* H10407 chromosome contains 599 CDSs not present in the nonpathogenic strains (Fig. 2; see Table S2 in the supplemental material). The majority (528) of these are clustered in the 25 ROD and are predicted to represent prophage genes and other mobility factors. Several genes comprise previously described loci specifically associated with ETEC virulence, *viz*., *leoA* (ROD 20), *tia* (ROD 20), and *tib* (ROD 13) (13, 14, 22). Other genes comprise loci previously noted in ETEC H10407, including the degenerate ETT2 locus (ROD 18) (45), antigen 43 (ROD 23)

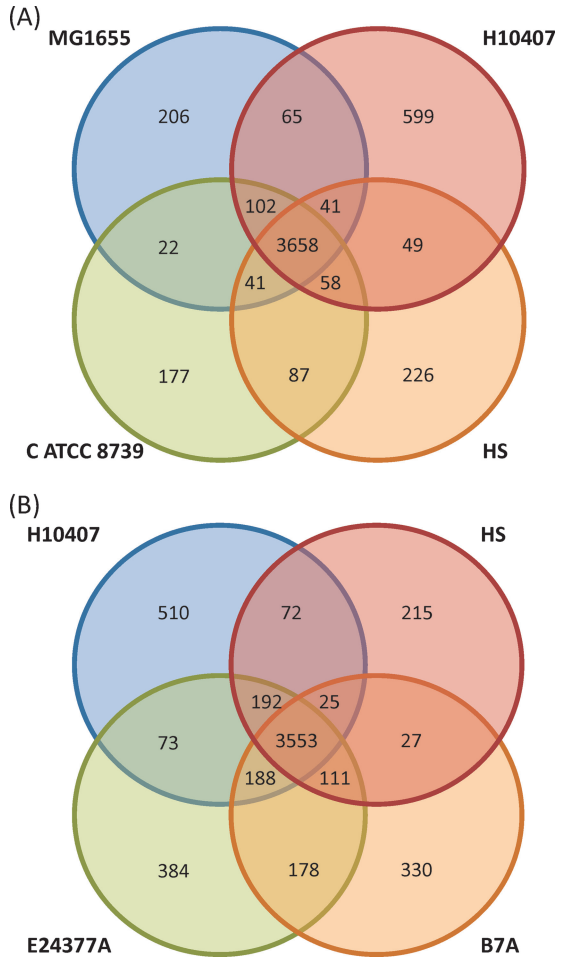


FIG. 2. Comparison of the genetic contents of the *E. coli* H10407 chromosome with those of the chromosomes of other sequenced strains of *E. coli*. (A) Comparison of *E. coli* H10407 with the three nonpathogenic *E. coli* strains HS, C, and K-12 revealed that the four strains share a large proportion of common genes. Only 599 *E. coli* H10407-specific genes were identified. The *E. coli* H10407-specific CDSs are not thought to be associated with virulence (see the text for details). (B) Comparison of *E. coli* H10407 with the genome-sequenced ETEC strains E24377A and B7A. The four strains possess 3,553 genes in common; however, the ETEC strains share only 188 genes not present in the commensal strain *E. coli* HS. The latter genes are not unique to ETEC; they are widely distributed among *E. coli* strains and are largely present among nonpathogenic strains of *E. coli*, such as *E. coli* K-12.

(63), a type 2 protein secretion locus found in many strains of *E. coli* (ROD 19) (4), and the *ecpP* fimbrial gene cluster also found in many *E. coli* strains (ROD 1) (4). Other ROD encode the Sil/Pco efflux system that confers silver/copper resistance (ROD 2) and yersiniabactin (ROD 11) and comprise the O78 serotype O antigen biosynthetic locus (ROD 14). The *sil* operon is closely related to *sil* from the IncH2 plasmid pMG101 (30, 38, 53) and is adjacent to a partially interrupted copper resistance operon similar to *pco* from plasmid pRJ1004 (2). The *sil-pco* locus is flanked by insertion sequence (IS) elements and phage-related sequences, suggesting horizontal transfer of these genes. The yersiniabactin iron acquisition locus is widely distributed in *E. coli* and other members of the *Enterobacteriaceae* (49). The remaining *E. coli* H10407-specific CDSs, which are not present on a ROD and do not encode prophage or mobility factors, encode the H11 flagellin subunits (CDS 2029 to 2033) and an additional copy of antigen 43 (CDS 2119) and comprise several pseudogenes (CDS 427, 1476, and 1573). These data largely agree with previously published subtractive-hybridization studies (5).

If a particular protein plays an important role in ETEC-mediated disease, then one would expect the gene encoding it to have a wide distribution among ETEC strains. To determine if there were any chromosomal genes specific to ETEC strains, comparisons were made with *E. coli* strains E24377A and B7A, the only other ETEC strains for which genome sequence data are available (44). Unlike *E. coli* H10407, both *E. coli* strains E24377A and B7A belong to the B1 subgroup of the *E. coli* phylogeny, a subgroup from which many commensals, but also a number of pathogens, are derived (4, 59). Comparison of *E. coli* H10407 with the sequenced ETEC strain E24377A revealed that the chromosomes are largely colinear (see Fig. S2 in the supplemental material). The genome of ETEC B7A is not finished, but experience with other *E. coli* genomes and comparison of the 198 finished ETEC B7A contigs suggest that the chromosome is also largely colinear with the other sequenced ETEC genomes (see Fig. S2 in the supplemental material). Analyses of the gene contents of all three strains revealed 3,741 genes conserved in all the strains, only 188 of which are not present in the commensal *E. coli* HS (Fig. 2B; see Table S3 in the supplemental material). The 188 genes identified through this comparison included loci encoding xanthine dehydrogenase (CDSs 0339 to 0343), the Mat fimbriae (CDSs 0348 to 0352), conserved proteins with unknown functions (CDSs 0673 to 0678), a flavoprotein electron transfer system (CDSs 1730 to 1734), the colanic exopolysaccharide biosynthetic machinery (CDSs 2171 to 2202), the Fec iron citrate uptake system (CDSs 3161 to 3166), a cellulose synthase system (CDSs 3776 to 3779), and a putative sugar utilization system (CDSs 4145 to 4154), all of which are present in the nonpathogen *E. coli* K-12 and are widely distributed among other *E. coli* strains (data not shown). The remainder of the 188 genes encode prophage or other mobility factors that are predicted to have no role in virulence. Of the 599 *E. coli* H10407-restricted genes identified through comparisons with the nonpathogenic *E. coli* strains mentioned above (Fig. 2A), 47 were conserved among the three pathogenic ETEC isolates. However, these genes were all related to mobile elements, and no putative virulence factors were identified. Notably, no significant homologues of *leoA*, *tibC*, *tibA*, or *tia* were detected in either *E. coli* E24377A or B7A, strongly suggesting these genes are not essential for ETEC-mediated disease. In conclusion, these data agree with previous observations that the chromosome of *E. coli* H10407 is most closely related to those of nonpathogenic *E. coli* strains and that the factors mediating diarrhea are not chromosomally encoded, indicating that the essential virulence factors are encoded on the plasmids (61).

**Potential virulence genes carried on the ETEC plasmids.** Since chromosomal comparisons revealed that no chromosomal CDS was unique to all three ETEC strains, we next examined the CDSs present on the four plasmids of ETEC H10407. The general characteristics of the plasmids are shown in Table 2. The two larger plasmids (pETEC948 and

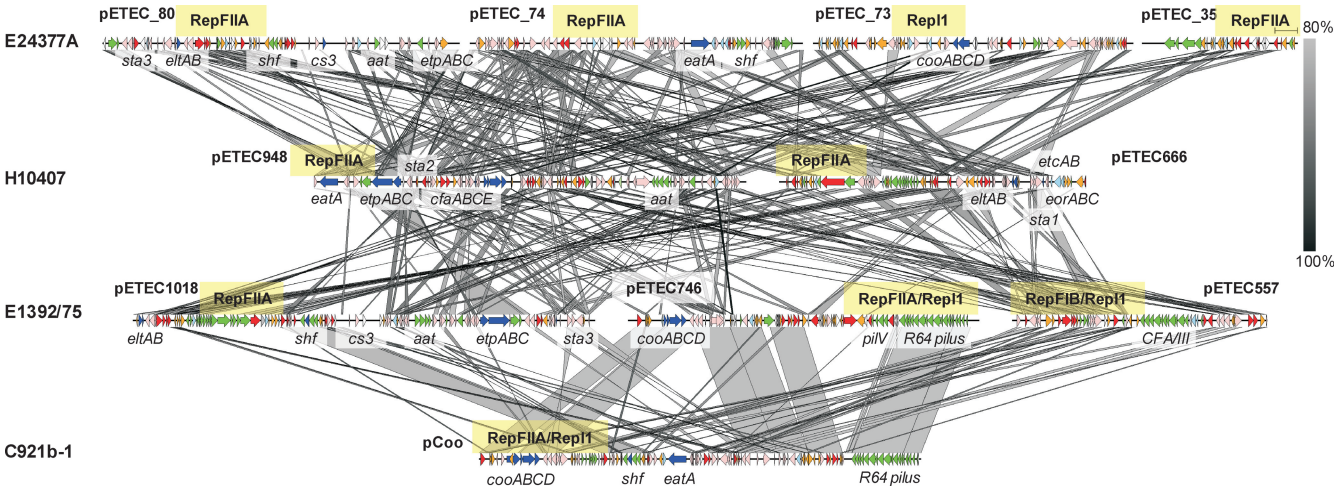


FIG. 3. Nucleotide sequence comparison of large conjugative-like plasmids from ETEC strains. Plasmid sequences from each strain were concatenated and compared using BLASTn. BLAST matches longer than 250 bp are shown as gray blocks in a comparison between plasmids from E24377A (pETEC_80, pETEC_74, pETEC_73, and pETEC_35), H10407 (pETEC948 and pETEC666), E1392/75 (pETEC1018, pETEC746, and pETEC557), and C921b-1 (pCoo). The shading of the gray blocks is proportional to the BLAST match (minimum, 80% nucleotide identity; maximum, 100% nucleotide identity). Each plasmid is denoted as a black line; the identity of each plasmid is noted above the line, and the source ETEC strain from which the plasmids are derived is given on the left side of the diagram. Coding sequences are depicted by arrows and are colored according to known or predicted functions: blue, virulence related; red, plasmid-related protein; green, outer membrane related (includes conjugal transfer loci); pink, transposase/insertion element related; light blue, regulatory protein; orange, conserved hypothetical protein; uncolored, hypothetical protein. The positions of genes encoding known or predicted virulence-related proteins are denoted by white boxes containing the gene names. In addition, the locus encoding the R64 conjugative pilus and the variant PilV tips is also depicted. The putative origin of replication associated with each of the plasmids is highlighted within yellow-shaded boxes. The chimeric nature of the plasmids is clearly visible, with recombination between plasmids a frequent occurrence. The unlabeled figure was prepared using a custom script (M. J. Sullivan and S. A. Beatson, unpublished data).

pETEC666) are reminiscent of conjugative plasmids that are often associated with the carriage of virulence factors, whereas the two smaller plasmids (pETEC58 and pETEC52) are homologous to mobilizable plasmids frequently encountered in a variety of bacterial species (24, 34). The latter plasmids have been shown to be mobilizable in the presence of IncF and other plasmid transfer systems (51). The majority of the CDSs on all four plasmids encode plasmid maintenance and transfer functions or were pseudogenes, genes with unknown functions not predicted to be involved in virulence, and transmissible elements (Table 2). An exhaustive list of the genetic content is unwarranted here, as a complete annotation of the plasmids is provided in the EMBL databases. Nevertheless, there are several noteworthy CDSs, described below, that can be termed "cargo" genes that have a known or putative role in pathogenesis. Thus, analyses revealed that *E. coli* H10407 pETEC948 possesses cargo genes encoding the previously described EatA SPATE (*eatA*), heat-stable enterotoxin STa2 (*sta2*), CFA/I fimbriae and associated regulator (*cfaABCD*), and the Etp two-partner secretion system and associated glycosyltransferase (*etpABC*) (Fig. 3) (18, 23, 42, 66). Analyses of the *E. coli* H10407 pETEC666 plasmid revealed that it contains the cargo genes encoding the previously described heat-stable enterotoxin STa1 (*sta1*) and the two subunits of LT enterotoxin (*eltA* and *eltB*) (Fig. 3) (8, 65). In addition, the plasmids contain several loci not previously associated with ETEC strains. ETEC H10407 pETEC948 possesses genes comprising a type I secretion locus similar to the dispersin secretion locus (*aatABCDP*) described for *E. coli* 042 (Fig. 4) (52). Associated with this locus is a gene encoding CexE, a previously described secreted protein of ETEC (43), which bears homology to the *E. coli* 042 dispersin protein (Fig. 4). Furthermore, pETEC666 carries genes encoding a two-component sensor kinase, here designated *etcA* and *etcB* (*E. coli* two-component), and a three-gene locus (here designated *eor* for *E. coli* oxidoreductase) encoding a protein with homology to cytochrome *b*-type subunit oxidoreductase protein (*eorA*), a protein with homology to an oxidoreductase molybdopterin binding domain protein (*eorB*), and a periplasmic protein with unknown function (*eorC*). In addition, ETEC H10407 pETEC58 encodes a putative deoxycytidylate deaminase (pETEC58_0005).

As mentioned above, if a particular protein plays an important role in ETEC-mediated disease, then one would expect it to have a wide distribution among ETEC strains. To determine whether the genes encoding the putative and known virulence factors of the ETEC H10407 plasmids, which we identified above, were conserved among ETEC strains, we next examined their prevalence among the available sequenced strains. To aid in this process, we determined the sequences of the plasmids from ETEC strain E1392/75. *E. coli* E1392/75 possesses five plasmids: three large conjugative plasmids designated pETEC1018, pETEC746, and pETEC557 and two mobilizable plasmids termed pETEC75 and pETEC62 (Table 2 lists their general characteristics). Included in the prevalence investigations were the ETEC strains E24377A and B7A and the plasmid pCoo from ETEC strain C921b-1, all of which were sequenced in other projects (28, 44). As the ETEC B7A genome is incomplete and no plasmids were resolved and pCoo is the only plasmid sequenced from ETEC C921b-1, we can confirm only the presence of genes among the available DNA se-

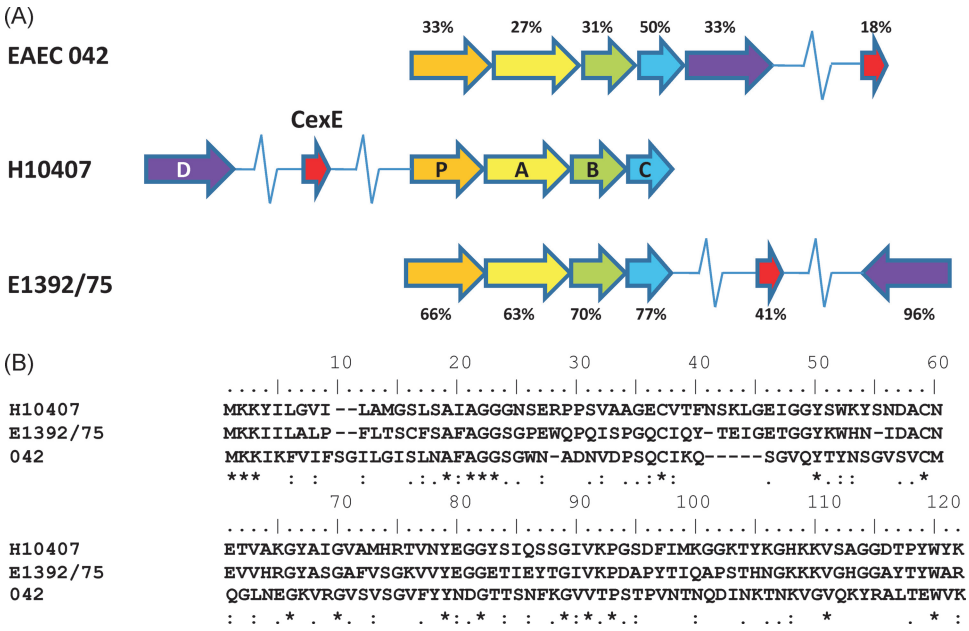


FIG. 4. Comparison of the EAEC *aat-aap* locus with the *aat-cexE* loci of ETEC strains. (A) The genetic organizations of the *aat* and *cexE* loci are depicted. The level of amino acid identity for each component of the *aat-cexE* system is shown; the percentages represent comparison with the *E. coli* H10407 orthologues. Orthologues are colored coded for ease of identification. Genes that are not juxtaposed are depicted with a blue line separating them. (B) Amino acid sequence alignment of ETEC CexE proteins with the EAEC 042 dispersin. All three proteins possess a signal sequence that is cleaved after the amino acid at position 21 in the alignment. There is limited conservation in the sequences; however, two cysteine residues that are disulfide bonded in dispersin are conserved. Based on the structure of dispersin, the remainder of the conserved residues appear to represent hydrophobic core residues required for structural integrity of the molecule. Asterisks indicate positions of amino acid identity; periods and colons show positions of low and high amino acid similarity.

quences and not the absence of particular genes from these strains. The distributions and locations of the cargo genes encoding known or putative virulence factors among the sequenced ETEC plasmids is depicted in Fig. 3 and is also shown in Table S4 in the supplemental material. Comparative analyses revealed that, like ETEC H10407, the ETEC strains E1392/75, B7A, and E24377A possess the ST and LT enterotoxins (none were identified for *E. coli* C921b-1, but previous analyses showed that the strain harbors LT and ST) (54). The EtpABC two-partner secretion system was identified in ETEC E1392/75 and E24377A. Homologues may exist in ETEC strains B7A and C921b-1, but their existence or nonexistence in these strains could not be resolved due to the lack of complete sequence data; however, other studies have not demonstrated a universal association of the *etpABC* locus with ETEC strains (23). Unlike ETEC strains H10407, E24377A, and C921b-1, the autotransporter-encoding *eatA* gene was not present on the ETEC E1392/75 plasmids. A homologue annotated as EatA is found in *E. coli* B7A; however, further analyses of this protein revealed that it is more closely related to SepA, a homologous SPATE protein from *Shigella flexneri* (1). No equivalents of ETEC H10407 *etcAB* or *eorABC* or of the gene encoding the putative deoxycytidylate deaminase were detected in any of the other ETEC strains.

Like *E. coli* H10407, the ETEC strains E24377A, E1392/75, and C921b-1 encode dispersin-like proteins previously designated CexE (43). Further analyses revealed that CexE is present in ETEC strains 27D and G427 (two CFA/I$^+$ strains) (43) and ETEC O167:H5, a CS6- and CS5-encoding strain (9). For EAEC, dispersin is secreted via the Aat type I secretion system; associates noncovalently with the extracellular face of the outer membrane, preventing collapse of the AAF/II fimbriae onto the bacterial cell surface by alteration of the surface charge; and is required for colonization (31, 40, 52). Analyses of the nucleotide sequences from ETEC strains B7A, E24377A, and E1392/75 revealed the presence of loci encoding type I secretion systems bearing striking homology to the Aat dispersin secretion system (Fig. 4). The cooccurrence of *cexE* genes with *aat* loci suggests that the CexE proteins are substrates for the Aat-like secretion systems of ETEC. Since, plasmid-borne fimbrial loci are inextricably linked to ETEC-mediated disease (18), CexE may play a role similar to that of dispersin by maintaining the CFs in such a manner that they can interact with epithelial receptors. However, further studies are required to investigate the function and distribution of CexE and to identify other relatives of this protein not yet recognized.

As mentioned above, adherence via plasmid-encoded fimbrial systems is a crucial step in ETEC pathogenesis (62). *E. coli* H10407 pETEC948 possesses the CFA/I chaperone-usher system (Fig. 3). ETEC E24377A possesses two chaperone-usher fimbrial systems located on pETEC_80 and pETEC_73, encoding the CS3 and CS1 fimbriae, respectively (44). Similarly, *E. coli* E1392/75 possesses CS3- and CS1-encoding loci on plasmids pETEC1018 and pETEC746, respectively, whereas pCoo possesses the CS1 cluster, all of which have been described previously (28, 57, 58). In addition, *E. coli* E1392/75 pETEC557 also encodes the CFA/III-type IV fimbria (29). To

determine whether fimbrial systems other than those mentioned above might play a crucial role in ETEC pathogenesis, we investigated conservation of putative fimbrial loci among the available *E. coli* sequences. ETEC H10407 contains 12 additional loci predicted to encode fimbriae, all of which are chromosomally located (see Table S5 in the supplemental material). Four of these loci (*mat*, *sfm*, *ycb*, and *yde*) contain pseudogenes and were considered nonfunctional. We sought to establish if *E. coli* H10407 harbored ETEC-specific fimbrial loci that might not be expressed by commensal *E. coli*, *E. coli* K-12, or enteroaggregative *E. coli.* The vast majority of fimbrial operons identified are also located in commensal and laboratory strains, with notable exceptions. The *yqi* and *stf-mrf* fimbrial loci are present in *E. coli* H10407 but contain pseudogenes in commensal or laboratory *E. coli* strains. However, an apparently functional *yqi* operon is also present in enteroaggregative *E. coli* strain 042, and thus, a functional *yqi* locus does not appear to be ETEC specific. Indeed, the *yqi* operon does not appear to be present in ETEC B7A (4). With regard to the *stf-mrf* operon, the *mrfC* gene is a pseudogene in *E. coli* K-12 but not in ETEC H10407. This six-gene cluster (*smfA-mrfCD-stfEFG*) is present in ETEC E24377A and EAEC 042, though with some divergence in the *stf* genes.

Finally, the ETEC E1392/75 pETEC62 plasmid possesses CDSs encoding a type II dihydropteroate synthase gene conferring sulfonamide resistance and CDSs encoding streptomycin phosphotransferase genes conferring streptomycin resistance. The plasmid possesses 99% nucleotide identity with the ETEC E24377A pETEC_6 plasmid and shares high levels of identity with plasmids from a variety of *E. coli* strains, including the *Shigella sonnei* pKKTET7 and the EPEC pE2348-2 plasmids However, this plasmid has no homologue in ETEC H10407 and no detectable homology among the ETEC B7A sequences, suggesting it may not be widespread among ETEC strains and thus is not essential for ETEC-mediated diarrhea.

In conclusion, the putative and known virulence genes identified on the plasmids of *E. coli* H10407 have differential distributions among the sequenced ETEC strains. In all cases, the ETEC strains possess genes encoding the ST and/or LT toxins (*sta* and/or *eltAB*, respectively), a chaperone-usher fimbrial biogenesis locus (e.g., the *cfa* locus), and components of an *aat-cexE* dispersin-like type I secretion system. Thus, despite the variation in individual plasmid gene contents, comparison of the entire plasmid complement of the sequenced ETEC strains suggests that there is a conserved core of genes contained on the plasmids that are predicted to be involved in virulence and may be essential for the establishment of ETEC-mediated disease.

**ETEC plasmids demonstrate a mosaic structure.** To determine whether the virulence factors identified above were encoded on a specific plasmid, or repertoire of plasmids, we examined the nucleotide sequence identity shared by the ETEC plasmids. The nucleotide sequences of the conjugative plasmids from each of the ETEC strains H10407, E1392/75, and E24377A were concatenated and compared using BLASTn. The levels of nucleotide sequence identity between pCoo and the other ETEC plasmids were determined in a similar manner. These comparisons revealed that while the plasmids all belong to a narrow subset of incompatibility groups (see below), extensive rearrangements and recombination events have occurred, resulting in individual plasmids that vary in their repertoires of virulence genes (Fig. 3; see Table S4 in the supplemental material). Such recombination can be seen by examining the distribution of the *eatA* gene. Thus, the *eatA* gene is not present in ETEC strain 1392/75, and in ETEC strain E24377A, the *eatA* gene is located on pETEC_74 and the *eltAB*, *aatPABC*, and *etpABC* loci are located on pETEC_80. In contrast, in ETEC strain H10407, the *eatA* gene is collocated with *etpABC* and *aatPABC* on pETEC948, whereas the *eltAB* locus is located on pETEC666. The *eatA* gene is present on ETEC C921b-1 pCoo, along with *cooABCD*; however, in ETEC strain E24377A, *cooABCD* is located on a separate plasmid (pETEC_73) (Fig. 3; see Table S4 in the supplemental material). Other virulence-associated genes also display such differential distributions (see Table S4 in the supplemental material), suggesting that the extrachromosomal components of the ETEC genome are in a state of flux (34, 44). Notably, the plasmids contain an extensive repertoire of IS elements and transposons (Table 2) (34); it is likely that the mobility of these genetic elements, or recombination between the elements, gives rise to the observed mosaic structure of the ETEC plasmids.

Similar comparisons of the small mobilizable plasmids of the ETEC strains did not demonstrate recombination between the mobilizable plasmids. Furthermore, there did not appear to be any significant exchange of genetic material between the conjugative plasmids and the small mobilizable plasmids (data not shown).

**Plasmid stability and maintenance functions of the ETEC plasmids.** To determine whether the virulence factors described above were encoded on self-transmissible plasmids, we examined the CDSs encoding the plasmid maintenance and transfer functions of each ETEC plasmid. A complete description of *E. coli* H10407 pETEC666 has been published (41), and the complete repertoire of genes for each ETEC plasmid are given in the EMBL databases (see Table 2 for accession numbers); thus, only the most salient features are described here. Plasmid nomenclature utilizes a system based on incompatibility groupings; plasmids of the same incompatibility group should not coexist within the same bacterial cell because of the similarity of their replication systems (34). However, sequence analyses of the CDSs encoding the plasmid replication functions of the repertoire of ETEC plasmids revealed that the large conjugative-like plasmids of *E. coli* strains H10407, E1392/75, and E24377A belong to a narrow subset of incompatibility groups and comprise multiple plasmids with the same replication mechanism (Fig. 3 and Table 2). Thus, the *E. coli* H10407 plasmids pETEC948 and pETEC666 belong to the RepFIIA (IncFIIA) subset of incompatibility groupings and have RepA1 proteins that share 94% identity (95% similarity), whereas the *E. coli* E1392/75 plasmids pETEC746 and pETEC557 harbor RepI1 (IncI1) replication functions (*E. coli* E1392/75 pETEC557 is an apparent cointegrate of RepF1B and RepI1 plasmids; such cointegration has previously been noted for *E. coli* C921b-1, where pCoo represents a cointegrate between a RepFIIA and a RepI1 plasmid [28]), with the corresponding RepZ proteins sharing 94% identity (95% similarity). Similarly, the previously described ETEC strain E24377A (44) possesses three plasmids with RepFIIA functions. The

basis for these antidogmatic observations is not understood and requires further in-depth investigation.

Analyses of the nucleotide sequences of the repertoire of large conjugative-like plasmids revealed that they possess a number of plasmid stability systems, including postsegregation killing systems and active-partitioning systems. The distribution of these systems among the plasmids sequenced in this study is given in Table 2. These stability systems have been described previously (25, 56).

Previous studies have noted that the large plasmids encoding the toxins of ETEC are in some cases self-transmissible and in other cases not transmissible (27). To investigate whether the plasmids sequenced in this study possessed transmissibility functions, we examined the transfer regions of the conjugative-like plasmids. As noted previously, *E. coli* H10407 pETEC666 has a transfer region that is interrupted by several IS*cE8* elements, severely diminishing the ability of the system to function efficiently (41). In contrast, *E. coli* H10407 pETEC948 possesses only remnants of the conjugation apparatus and is presumably not self-transmissible. In addition, the *E. coli* E1392/75 pETEC1018 plasmid also contains an incomplete conjugation apparatus, which is presumed to be ineffective at promoting conjugation; however, *E. coli* E1392/75 pETEC746 possess an intact conjugation system that is 100% identical to the region encoding the functional R64-like conjugative pilus of pCoo of *E. coli* C921b-1, and thus, it is presumed to be functional. *E. coli* E1392/75 pETEC557 lacks CDSs encoding the R64 conjugative pilus and possesses remnants of an F-like conjugation system.

ETEC strains H10407, E1392/75, and E24377A all contain similar small mobilizable plasmids (pETEC52, pETEC75, and pETEC_5, respectively) with *mob* and *rep* regions displaying 100% identity. The *E. coli* E1392/75 pETEC75 plasmid contains an IS*100* element not present in the other two plasmids. The distribution of these plasmid types among the sequenced ETEC strains suggests that they might be common components of ETEC genomes. This plasmid type has been found in a number of other *E. coli* strains and has been shown to increase the fitness of certain *E. coli* host strains (16). Therefore, multiple selective advantages might be conferred on the ETEC strains possessing these small plasmids. The *rep* and *mob* regions (3,058 bp) of the ETEC H10407 pETEC58 plasmid, which encodes the putative deoxycytidylate deaminase, demonstrates 81% identity with plasmid pHW66 from *Rahnella* sp. strain WMR66; the putative deoxycytidylate deaminase is lacking in pHW66. In contrast to the other ETEC plasmids, there are no plasmids homologous to ETEC H10407 pETEC58 among the other genome-sequenced ETEC isolates.

**The *E. coli* E1392/75 pETEC746 plasmid contains a pilin shufflon.** As mentioned above, ETEC E1392/75 pETEC746 contains regions homologous to the *Salmonella enterica* serovar Typhimurium RepI1 plasmid R64 that are also present in *E. coli* C921b-1 pCoo and that have been shown to be functional in that system (28). As sequencing of the ETEC genome was being completed, dideoxy sequencing of the region from bp 56253 to 59961 of pETEC746 from *E. coli* E1392/75 identified a nucleotide region undergoing dynamic alteration. The region of DNA consisted of a shufflon similar to that of R64 (36). PilV is a component of a conjugative pilus that expresses

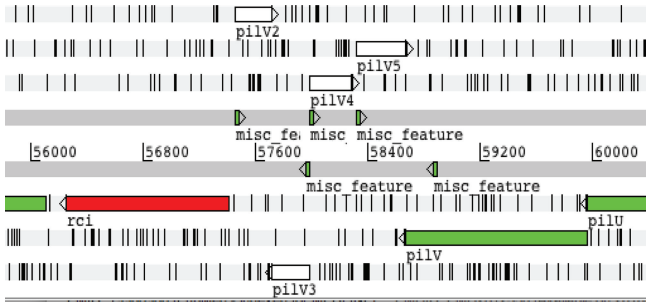


FIG. 5. Arrangement of the *pilV* shufflon region of *E. coli* E1392/75 pETEC746. Annotation of the *pilV* region is shown using the Artemis sequence viewer (1). Sequence blocks encoding C-terminal fragments of PilV are found in both orientations between *pilV* and the *rci* recombinase gene. Identical 13-bp repeats (GTGCCAATCCGGT) are shown as miscellaneous features and mark the predicted sites of recombination between the C-terminal fragments and the *pilV* gene.

different tips involved with attachment to cells. The tips are regulated via a DNA shufflon mechanism involving recombination at particular repeating sites. Recombination is mediated by the *rci* recombinase linked to this region. Alternative tip adhesins are involved in attachment to different strains and species and have been elucidated experimentally in *S.* Typhimurium (36). Evidence that the shufflon is functional in the *E. coli* E1392/75 plasmid pETEC746 is provided in the sequences from a small-insert library. Within the sequences are examples of *pilV* with alternative C-terminal tips, implying that the plasmids sequenced represented a population in genetic flux. There is direct evidence for sequences of *pilV* with tips *V1*, *V3*, and *V4* (Fig. 5). There are also regions of DNA sequence equivalent to tips *shuC1*, *shuC′*, and *shuC2* from *S.* Typhimurium. However, these were present only in a small subpopulation of pETEC746 plasmids and have been omitted from the complete finished sequence.

**Conclusions.** This study provides a genomic context for the vast amount of experimental and epidemiological data published thus far and provides a template for future diagnostic and intervention strategies. Evidence presented here suggests that the prototypical ETEC isolate *E. coli* H10407 was a commensal isolate that acquired a number of plasmids containing a limited repertoire of virulence genes and thereby gained the ability to cause disease. Furthermore, comparisons of the genetic content of *E. coli* H10407 with those of other ETEC strains has revealed only a limited number of conserved genes, suggesting that to become pathogenic, *E. coli* need only acquire (i) toxins (ST, LT, or both) to elicit net secretion from enterocytes; (ii) a fimbrial system that mediates attachment to the intestinal epithelium, e.g., CFA/I; and (iii) a novel type I secretion system, the substrate of which (CexE) maintains the fimbriae in the correct physical organization. These data suggest that ETEC vaccine strategies should focus on these plasmid-encoded virulence factors. However, given the relative plasticity of the *E. coli* genome, molecular epidemiological studies are essential to determine whether these factors are widely distributed among ETEC strains from geographically diverse locations.

## ACKNOWLEDGMENT

This work was supported by project grant BB/C510075/1 from the BBSRC to I.R.H., M.J.P., C.W.P., J.P., and N.R.T.

## REFERENCES

1. **Benjelloun-Touimi, Z., P. J. Sansonetti, and C. Parsot.** 1995. SepA, the major extracellular protein of Shigella flexneri: autonomous secretion and involvement in tissue invasion. Mol. Microbiol. **17:**123–135.
2. **Brown, N. L., S. R. Barrett, J. Camakaris, B. T. Lee, and D. A. Rouch.** 1995. Molecular genetics and transport analysis of the copper-resistance determinant (pco) from Escherichia coli plasmid pRJ1004. Mol. Microbiol. **17:**1153–1166.
3. **Chang, D. E., D. J. Smalley, D. L. Tucker, M. P. Leatham, W. E. Norris, S. J. Stevenson, A. B. Anderson, J. E. Grissom, D. C. Laux, P. S. Cohen, and T. Conway.** 2004. Carbon nutrition of Escherichia coli in the mouse intestine. Proc. Natl. Acad. Sci. U. S. A. **101:**7427–7432.
4. **Chaudhuri, R. R., M. Sebaihia, J. L. Hobman, M. A. Webber, D. L. Leyton, M. D. Goldberg, A. F. Cunningham, A. Scott-Tucker, P. R. Ferguson, C. M. Thomas, G. Frankel, C. M. Tang, E. G. Dudley, I. S. Roberts, D. A. Rasko, M. J. Pallen, J. Parkhill, J. P. Nataro, N. R. Thomson, and I. R. Henderson.** 2010. Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative Escherichia coli strain 042. PLoS One **5:**e8801.
5. **Chen, Q., S. J. Savarino, and M. M. Venkatesan.** 2006. Subtractive hybridization and optical mapping of the enterotoxigenic Escherichia coli H10407 chromosome: isolation of unique sequences and demonstration of significant similarity to the chromosome of E. coli K-12. Microbiology **152:**1041–1054.
6. **Coster, T. S., M. K. Wolf, E. R. Hall, F. J. Cassels, D. N. Taylor, C. T. Liu, F. C. Trespalacios, A. DeLorimier, D. R. Angleberger, and C. E. McQueen.** 2007. Immune response, ciprofloxacin activity, and gender differences after human experimental challenge by two strains of enterotoxigenic Escherichia coli. Infect. Immun. **75:**252–259.
7. **Cravioto, A.** 1980. Ph.D. thesis. University of London, London, United Kingdom.
8. **Dallas, W. S.** 1990. The heat-stable toxin I gene from Escherichia coli 18D. J. Bacteriol. **172:**5490–5493.
9. **de Haan, L. A., G. A. Willshaw, B. A. van der Zeijst, and W. Gaastra.** 1991. The nucleotide sequence of a regulatory gene present on a plasmid in an enterotoxigenic Escherichia coli strain of serotype O167:H5. FEMS Microbiol. Lett. **67:**341–346.
10. **Delcher, A. L., K. A. Bratke, E. C. Powers, and S. L. Salzberg.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23:**673–679.
11. **Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27:**4636–4641.
12. **Duriez, P., O. Clermont, S. Bonacorsi, E. Bingen, A. Chaventre, J. Elion, B. Picard, and E. Denamur.** 2001. Commensal Escherichia coli isolates are phylogenetically distributed among geographically distinct human populations. Microbiology **147:**1671–1676.
13. **Elsinghorst, E. A., and D. J. Kopecko.** 1992. Molecular cloning of epithelial cell invasion determinants from enterotoxigenic Escherichia coli. Infect. Immun. **60:**2409–2417.
14. **Elsinghorst, E. A., and J. A. Weitz.** 1994. Epithelial cell invasion and adherence directed by the enterotoxigenic Escherichia coli tib locus is associated with a 104-kilodalton outer membrane protein. Infect. Immun. **62:**3463–3471.
15. **Emanuelsson, O., S. Brunak, G. von Heijne, and H. Nielsen.** 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. **2:**953–971.
16. **Enne, V. I., P. M. Bennett, D. M. Livermore, and L. M. Hall.** 2004. Enhancement of host fitness by the sul2-coding plasmid p9123 in the absence of selective pressure. J. Antimicrob. Chemother. **53:**958–963.
17. **Evans, D. G., T. K. Satterwhite, D. J. Evans, Jr., and H. L. DuPont.** 1978. Differences in serological responses and excretion patterns of volunteers challenged with enterotoxigenic Escherichia coli with and without the colonization factor antigen. Infect. Immun. **19:**883–888.
18. **Evans, D. G., R. P. Silver, D. J. Evans, Jr., D. G. Chase, and S. L. Gorbach.** 1975. Plasmid-controlled colonization factor associated with virulence in Esherichia coli enterotoxigenic for humans. Infect. Immun. **12:**656–667.
19. **Evans, D. J., Jr., and D. G. Evans.** 1973. Three characteristics associated with enterotoxigenic Escherichia coli isolated from man. Infect. Immun. **8:**322–328.
20. **Fleckenstein, J. M., P. R. Hardwidge, G. P. Munson, D. A. Rasko, H. Sommerfelt, and H. Steinsland.** 2010. Molecular mechanisms of enterotoxigenic Escherichia coli infection. Microbes Infect. **12:**89–98.
21. **Fleckenstein, J. M., D. J. Kopecko, R. L. Warren, and E. A. Elsinghorst.** 1996. Molecular characterization of the tia invasion locus from enterotoxigenic Escherichia coli. Infect. Immun. **64:**2256–2265.
22. **Fleckenstein, J. M., L. E. Lindler, E. A. Elsinghorst, and J. B. Dale.** 2000. Identification of a gene within a pathogenicity island of enterotoxigenic Escherichia coli H10407 required for maximal secretion of the heat-labile enterotoxin. Infect. Immun. **68:**2766–2774.
23. **Fleckenstein, J. M., K. Roy, J. F. Fischer, and M. Burkitt.** 2006. Identification of a two-partner secretion locus of enterotoxigenic Escherichia coli. Infect. Immun. **74:**2245–2258.
24. **Francia, M. V., A. Varsaki, M. P. Garcillan-Barcia, A. Latorre, C. Drainas, and F. de la Cruz.** 2004. A classification scheme for mobilization regions of bacterial plasmids. FEMS Microbiol. Rev. **28:**79–100.
25. **Friehs, K.** 2004. Plasmid copy number and plasmid stability. Adv. Biochem. Eng. Biotechnol. **86:**47–82.
26. **Frishman, D., A. Mironov, H. W. Mewes, and M. Gelfand.** 1998. Combining diverse evidence for gene recognition in completely sequenced bacterial genomes. Nucleic Acids Res. **26:**2941–2947.
27. **Froehlich, B., E. Holtzapple, T. D. Read, and J. R. Scott.** 2004. Horizontal transfer of CS1 pilin genes of enterotoxigenic Escherichia coli. J. Bacteriol. **186:**3230–3237.
28. **Froehlich, B., J. Parkhill, M. Sanders, M. A. Quail, and J. R. Scott.** 2005. The pCoo plasmid of enterotoxigenic Escherichia coli is a mosaic cointegrate. J. Bacteriol. **187:**6509–6516.
29. **Gomez-Duarte, O. G., S. Chattopadhyay, S. J. Weissman, J. A. Giron, J. B. Kaper, and E. V. Sokurenko.** 2007. Genetic diversity of the gene cluster encoding longus, a type IV pilus of enterotoxigenic Escherichia coli. J. Bacteriol. **189:**9145–9149.
30. **Gupta, A., K. Matsui, J. F. Lo, and S. Silver.** 1999. Molecular basis for resistance to silver cations in Salmonella. Nat. Med. **5:**183–188.
31. **Harrington, S. M., J. Sheikh, I. R. Henderson, F. Ruiz-Perez, P. S. Cohen, and J. P. Nataro.** 2009. The Pic protease of enteroaggregative Escherichia coli promotes intestinal colonization and growth in the presence of mucin. Infect. Immun. **77:**2465–2473.
32. **Henderson, I. R., F. Navarro-Garcia, and J. P. Nataro.** 1998. The great escape: structure and function of the autotransporter proteins. Trends Microbiol. **6:**370–378.
33. **Iguchi, A., N. R. Thomson, Y. Ogura, D. Saunders, T. Ooka, I. R. Henderson, D. Harris, M. Asadulghani, K. Kurokawa, P. Dean, B. Kenny, M. A. Quail, S. Thurston, G. Dougan, T. Hayashi, J. Parkhill, and G. Frankel.** 2009. Complete genome sequence and comparative genome analysis of enteropathogenic Escherichia coli O127:H6 strain E2348/69. J. Bacteriol. **191:**347–354.
34. **Johnson, T. J., and L. K. Nolan.** 2009. Pathogenomics of the virulence plasmids of Escherichia coli. Microbiol. Mol. Biol. Rev. **73:**750–774.
35. **Kaper, J. B., J. P. Nataro, and H. L. Mobley.** 2004. Pathogenic Escherichia coli. Nat. Rev. Microbiol. **2:**123–140.
36. **Komano, T., S. R. Kim, and T. Yoshida.** 1995. Mating variation by DNA inversions of shufflon in plasmid R64. Adv. Biophys. **31:**181–193.
37. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:**567–580.
38. **McHugh, G. L., R. C. Moellering, C. C. Hopkins, and M. N. Swartz.** 1975. Salmonella typhimurium resistant to silver nitrate, chloramphenicol, and ampicillin. Lancet **1:**235–240.
39. **Nataro, J. P., and J. B. Kaper.** 1998. Diarrheagenic Escherichia coli. Clin. Microbiol. Rev. **11:**142–201.
40. **Nishi, J., J. Sheikh, K. Mizuguchi, B. Luisi, V. Burland, A. Boutin, D. J. Rose, F. R. Blattner, and J. P. Nataro.** 2003. The export of coat protein from enteroaggregative Escherichia coli by a specific ATP-binding cassette transporter system. J. Biol. Chem. **278:**45680–45689.
41. **Ochi, S., T. Shimizu, K. Ohtani, Y. Ichinose, H. Arimitsu, K. Tsukamoto, M. Kato, and T. Tsuji.** 2009. Nucleotide sequence analysis of the enterotoxigenic Escherichia coli Ent plasmid. DNA Res. **16:**299–309.
42. **Patel, S. K., J. Dotson, K. P. Allen, and J. M. Fleckenstein.** 2004. Identification and molecular characterization of EatA, an autotransporter protein of enterotoxigenic Escherichia coli. Infect. Immun. **72:**1786–1794.
43. **Pilonieta, M. C., M. D. Bodero, and G. P. Munson.** 2007. CfaD-dependent expression of a novel extracytoplasmic protein from enterotoxigenic Escherichia coli. J. Bacteriol. **189:**5060–5067.
44. **Rasko, D. A., M. J. Rosovitz, G. S. Myers, E. F. Mongodin, W. F. Fricke, P. Gajer, J. Crabtree, M. Sebaihia, N. R. Thomson, R. Chaudhuri, I. R. Henderson, V. Sperandio, and J. Ravel.** 2008. The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. J. Bacteriol. **190:**6881–6893.
45. **Ren, C. P., R. R. Chaudhuri, A. Fivian, C. M. Bailey, M. Antonio, W. M. Barnes, and M. J. Pallen.** 2004. The ETT2 gene cluster, encoding a second type III secretion system from Escherichia coli, is present in the majority of strains but has undergone widespread mutational attrition. J. Bacteriol. **186:**3547–3560.
46. **Roy, K., G. M. Hilliard, D. J. Hamilton, J. Luo, M. M. Ostmann, and J. M. Fleckenstein.** 2009. Enterotoxigenic Escherichia coli EtpA mediates adhesion between flagella and host cells. Nature **457:**594–598.
47. **Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16:**944–945.

48. **Sammut, S. J., R. D. Finn, and A. Bateman.** 2008. Pfam 10 years on: 10,000 families and still growing. Brief Bioinform. **9:**210–219.
49. **Schubert, S., A. Rakin, and J. Heesemann.** 2004. The Yersinia high-pathogenicity island (HPI): evolutionary and functional aspects. Int. J. Med. Microbiol. **294:**83–94.
50. **Scotland, S. M., N. P. Day, and B. Rowe.** 1983. Acquisition and maintenance of enterotoxin plasmids in wild-type strains of Escherichia coli. J. Gen. Microbiol. **129:**3111–3120.
51. **Selvaratnam, S., and M. A. Gealt.** 1993. Transcription of ColE1Ap mbeC induced by conjugative plasmids from twelve different incompatibility groups. J. Bacteriol. **175:**6982–6987.
52. **Sheikh, J., J. R. Czeczulin, S. Harrington, S. Hicks, I. R. Henderson, C. Le Bouguenec, P. Gounon, A. Phillips, and J. P. Nataro.** 2002. A novel dispersin protein in enteroaggregative Escherichia coli. J. Clin. Invest. **110:**1329–1337.
53. **Silver, S., A. Gupta, K. Matsui, and J. F. Lo.** 1999. Resistance to ag(i) cations in bacteria: environments, genes and proteins. Met Based Drugs **6:**315–320.
54. **Smyth, C. J.** 1982. Two mannose-resistant haemagglutinins on enterotoxigenic Escherichia coli of serotype O6:K15:H16 or H-isolated from travellers' and infantile diarrhoea. J. Gen. Microbiol. **128:**2081–2096.
55. **Steffen, R., F. Castelli, H. D. Nothdurft, L. Rombo, and N. J. Zuckerman.** 2005. Vaccination against enterotoxigenic Escherichia coli, a cause of travelers' diarrhea. J. Travel Med. **12:**102–107.
56. **Summers, D. K., and D. J. Sherratt.** 1985. Bacterial plasmid stability. Bioessays **2:**209–211.
57. **Svennerholm, A. M., and C. Ahren.** 1982. Serological subtypes of Escherichia coli colonization factor antigen II. Eur. J. Clin. Microbiol. **1:**107–111.
58. **Tacket, C. O., R. H. Reid, E. C. Boedeker, G. Losonsky, J. P. Nataro, H. Bhagat, and R. Edelman.** 1994. Enteral immunization and challenge of volunteers given enterotoxigenic E. coli CFA/II encapsulated in biodegradable microspheres. Vaccine **12:**1270–1274.
59. **Tenaillon, O., D. Skurnik, B. Picard, and E. Denamur.** 2010. The population genetics of commensal Escherichia coli. Nat. Rev. Microbiol. **8:**207–217.
60. **Turner, A. K., T. D. Terry, D. A. Sack, P. Londono-Arcila, and M. J. Darsley.** 2001. Construction and characterization of genetically defined aro omp mutants of enterotoxigenic Escherichia coli and preliminary studies of safety and immunogenicity in humans. Infect. Immun. **69:**4969–4979.
61. **Turner, S. M., R. R. Chaudhuri, Z. D. Jiang, H. DuPont, C. Gyles, C. W. Penn, M. J. Pallen, and I. R. Henderson.** 2006. Phylogenetic comparisons reveal multiple acquisitions of the toxin genes by enterotoxigenic Escherichia coli strains of different evolutionary lineages. J. Clin. Microbiol. **44:**4528–4536.
62. **Turner, S. M., A. Scott-Tucker, L. M. Cooper, and I. R. Henderson.** 2006. Weapons of mass destruction: virulence factors of the global killer enterotoxigenic Escherichia coli. FEMS Microbiol. Lett. **263:**10–20.
63. **van der Woude, M. W., and I. R. Henderson.** 2008. Regulation and function of Ag43 (Flu). Annu. Rev. Microbiol. **62:**153–169.
64. **Wennerås, C., and V. Erling.** 2004. Prevalence of enterotoxigenic Escherichia coli-associated diarrhoea and carrier state in the developing world. J. Health Popul. Nutr. **22:**370–382.
65. **Yamamoto, T., T. Tamura, and T. Yokota.** 1984. Primary structure of heat-labile enterotoxin produced by Escherichia coli pathogenic for humans. J. Biol. Chem. **259:**5037–5044.
66. **Yamamoto, T., and T. Yokota.** 1980. Cloning of deoxyribonucleic acid regions encoding a heat-labile and heat-stable enterotoxin originating from an enterotoxigenic Escherichia coli strain of human origin. J. Bacteriol. **143:** 652–660.