HHS Public Access

Author manuscript

Pathog Dis. Author manuscript; available in PMC 2015 January 02.

Published in final edited form as:

Pathog Dis. 2013 October ; 69(1): 66–70. doi:10.1111/2049-632X.12059.

Draft genome sequences of five recent human uropathogenic *Escherichia coli* isolates

Sargurunathan Subashchandrabose[a], **Tracy H. Hazen**[b], **David A. Rasko**[b], and **Harry L. T. Mobley**[a,#]

[a]University of Michigan Medical School, Department of Microbiology and Immunology, Ann Arbor, Michigan, USA [b]University of Maryland School of Medicine, Institute for Genome Sciences, Department of Microbiology and Immunology, Baltimore, Maryland

Abstract

This study reports the release of draft genome sequences of five isolates of uropathogenic *Escherichia coli* (UPEC), isolated from patients suffering from uncomplicated cystitis in 2012 in Ann Arbor, Michigan. Phylogenetic analyses revealed that these strains belonged to *E. coli* phylogroups B2 and D, and are closely related to known UPEC strains. Comparative genomic analysis revealed that more conserved proteins were shared between these recent isolates and UPEC strains causing cystitis than those causing pyelonephritis. Additional genomic comparisons identified that three isolates encode a type III secretion system (T3SS) and a putative T3SS effector gene cluster along with an invasin-like outer membrane protein. Presence of T3SS genes is a rare occurrence among UPEC strains. These genomes further substantiate the heterogeneity of the gene pool of UPEC and provide a foundation for comparative genomic studies using recent clinical isolates.

Keywords

Uropathogenic *E. coli*; draft genomes; type III secretion system

Urinary tract infections are one of the most common bacterial infections afflicting humans (Russo & Johnson, 2003). Uropathogenic *Escherichia coli* (UPEC) is the etiological agent of a majority of cases of uncomplicated urinary tract infections (UTIs) in otherwise healthy individuals (Hooton, 2012). UPEC are a heterogeneous group of bacteria that are closely related to avian pathogenic *E. coli* and neonatal meningitis *E. coli* (Russo & Johnson, 2000). UPEC is believed to exhibit a commensal-like lifestyle within the gastrointestinal tract and induces pathological changes only upon entry into extraintestinal sites such as the urinary

Address correspondence to: Harry L.T. Mobley, Ph.D. Frederick G. Novy Professor and Chair, Department of Microbiology and Immunology, University of Michigan Medical School, 5641 Medical Science Bldg. II, 1150 West Medical Center Dr., Ann Arbor, Michigan 48109-0620, hmobley@med.umich.edu, Tel: (734) 764-1466, Fax: (734) 763-7163.

Nucleotide sequence accession numbers. This Whole Genome Shotgun sequencing project has been deposited at DDBJ/EMBL/GenBank under the accession numbers APNW00000000, APNU00000000, APNY00000000, APNX00000000, and APNV0000000 corresponding to the UPEC isolates HM26, HM27, HM46, HM65, and HM69, respectively. The versions described in this paper are the first version, APNW01000000, APNU01000000, APNY01000000, APNX01000000, and APNV01000000.

tract and bloodstream (Russo & Johnson, 2000). Available genomes of UPEC strains reveal a complex mosaic structure encompassing a core genome interrupted by regions that carry the hallmarks of horizontally transferred genetic elements (Welch, *et al.*, 2002, Brzuszkiewicz, *et al.*, 2006). Frequently, such islands contain genes that contribute to uropathogenesis and can be considered as pathogenicity islands (Lloyd, *et al.*, 2007, Lloyd, *et al.*, 2009).

UPEC remains a major burden on human health and is becoming increasingly recalcitrant to routinely used therapeutic agents (Gupta, *et al.*, 2011, Hooton, 2012). Several virulence factors, such as type 1 and P fimbriae, flagella, capsule, and toxins, have been identified in UPEC (Brumbaugh & Mobley, 2012). Multiple fitness mechanisms, including co-opting metabolic enzymes to enable survival and colonization in the mammalian urinary tract have been delineated in UPEC (Alteri & Mobley, 2012). However, to translate the knowledge on UPEC pathogenesis to develop novel therapeutic and prophylactic agents, a comprehensive understanding of the cues encountered by UPEC during human infection is required (Hagan, *et al.*, 2010). In an effort to better define those cues, we are currently profiling the transcriptomes of UPEC derived directly from patients with clinical urinary tract infection. Due to the genetic heterogeneity observed among UPEC strains, we sequenced the genomes of an additional five clinical isolates. The majority of UPEC reference genomes have been derived from isolates that are decades old, and it is possible that human activity, including both intended and unintended exposure to antibiotics, has changed the selective pressures on this bacterium in recent years.

UPEC isolates (HM26, HM27, HM46, HM65 and HM69) were isolated from urine of female patients diagnosed with cystitis at the University of Michigan Health Service clinic. The age of patients ranged from 18 to 25 years, with a median age of 22 years. Briefly, urine samples were cultured in MacConkey agar and lactose-fermenting colonies were screened using a Vitek2 system to conclusively identify *E. coli*. Quantitative cultures of urine samples revealed high levels of UPEC bacteriuria (>10^5 CFU/ml) in all samples. All samples, except HM26, were obtained from patients suffering from isolated instances of UTIs; HM26 was isolated from a patient suffering from recurrent UTI (four episodes in the six months preceding sample collection). Antimicrobial susceptibility profile for each isolate was determined using a Vitek2 system and HM69 was found to be resistant to trimethoprim/sulfamethoxazole, the primary therapeutic agent for uncomplicated cystitis (Gupta, *et al.*, 2011). None of these isolates were resistant to ciprofloxacin, another commonly used to treat UTIs (Gupta, *et al.*, 2011). Somatic (O) and flagellar (H) antigen types were determined at the *E. coli* reference center at Pennsylvania State University and are as follows: HM26 (O2:H18), HM27 (O4:H5), HM46 (O166:H15), HM65 (O2:H6/41), and HM69 (O15:H18). These isolates represent a typical diversity of *E. coli* isolated from uncomplicated cystitis in humans in all characteristics.

Genomic DNA was extracted from bacteria grown in lysogeny broth using DNeasy kit (Qiagen). The genome sequence of each isolate was generated at the Institute for Genome Sciences Genome Resource Center (http://www.igs.umaryland.edu/research/grc/intro.php) on Illumina HiSeq2500 using paired-end libraries with 300 bp inserts [Table 1]. The draft genomes were assembled using both the Velvet assembler (Zerbino & Birney, 2008) with

kmer values determined using VelvetOptimiser v2.1.4 (http://bioinformatics.net.au/software.velvetoptimiser.shtml), and the Edena v3 assembler (Hernandez, *et al.*, 2008). Contigs from the two assemblies were merged using Minimus (Sommer, *et al.*, 2007) and contigs longer than 200 bp were used for further analysis. The resulting genome assemblies contained an average of 125 contigs per genome (range 43–285) [Table 1]. Nucleotide sequences were annotated using the RAST server (Aziz, *et al.*, 2008). The numbers of predicted genes from the draft genomes [Table 1] were similar to the previously sequenced *E. coli* genomes with an average of 5,165 genes per genome (range 4,904–5,420). The presence of select known urovirulence factors in these isolates can be found in Table 2.

We probed the phylogenetic relationship between our recent isolates with a collection of representative *E. coli* and *Shigella* strains [Fig. 1] using a whole genome phylogeny-based approach as previously described (Sahl, *et al.*, 2011). Briefly, draft genome sequences were aligned to sequenced reference strains [Fig. 1] using Mugsy (Angiuoli & Salzberg, 2011). Aligned regions were extracted and a maximum-likelihood phylogenetic tree with 100 bootstrap replicates was inferred from the aligned regions using RAxML v7.2.8 (Stamatakis, 2006) and visualized using FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree). Phylogroups B2 and D encompass most known UPEC strains (Russo & Johnson, 2000) and the isolates identified in this study are also members of these two phylogroups. Three isolates (HM26, HM46 and HM69) cluster with a cystitis strain, UMN026 and an enteroaggregative *E. coli* strain 042 in phylogroup D [Fig. 1], whereas HM27 and HM65 cluster with extensively studied prototypical UPEC strains, CFT073, UTI89 and 536, which are members of phylogroup B2 [Fig. 1]. Based on the whole genome phylogeny the recent UPEC isolates appear to be similar to previously identified UPEC isolates.

Initial screening of the draft genome sequences for features, not found in previously sequenced strains, revealed that the isolates in phylogroup D, HM26, HM46 and HM69, all contain a significant number of genes encoding proteins involved in plasmid conjugation and transfer functions, suggesting that these isolates harbor plasmids. All three isolates contain an IncF-type machinery, and HM26 and HM69 encode an additional IncI1machinery, suggesting that multiple plasmids may be present in these isolates. Further analysis of the genes present in these plasmids is required to elucidate whether these plasmids contain genes involved in antibiotic resistance and virulence.

Blast score ratio (BSR), an *in silico* approach to conduct comparative proteomic analyses based on proteins predicted to be encoded in a genome (Rasko, *et al.*, 2005), was used to compare the proteins encoded in the newly sequenced strains with well characterized UPEC strains. The BSRs were calculated as the ratio of raw BLASTP score for the query to the raw BLASTP score of the reference strain. BSR cut-offs of ≥ 0.8 and < 0.8 to > 0.4 were used to determine whether a gene is conserved or divergent, respectively. A BSR value of 0.8 corresponds to ~85–90% identity over 90% length of a protein sequence, indicating a highly conserved sequence (Rasko, *et al.*, 2005). An average of 3133 proteins were conserved and 1697 proteins were divergent between the recent UPEC isolates and the established UPEC strains CFT073, 536, F11, UTI89, and UMN026 [Table 1]. When compared to cystitis strain F11, these isolates contained 3342 and 1477 proteins that were conserved and divergent, respectively. 3236 proteins were conserved and 1783 proteins were divergent between these

isolates and pyelonephritis strain CFT073. Taken together, BSR analysis indicates that the isolates sequenced during this study are more closely related to cystitis strains than pyelonephritis strains.

Type III secretion system (T3SS) is used by bacteria to inject effectors directly into host cells (Ren, *et al.*, 2004). T3SS has been the subject of extensive investigation in enteropathogenic and enterohemorrhagic strains of *E. coli* (Wong, *et al.*, 2011). However, T3SS genes are not commonly found in UPEC isolates; a previous study revealed that three out of 76 cystitis isolates, collected in Japan, had genes encoding components of a T3SS (Miyazaki, *et al.*, 2002). In contrast, in this study including only five isolates, three isolates (HM26, HM46 and HM69) revealed genes that encode the structural components of a T3SS near tRNA *glyU*. These phylogroup D UPEC isolates also contain a putative effector island (*eip* island) adjacent to tRNA *selC*, that encodes potential T3SS effectors and an invasin-like outer membrane protein. Regions near *glyU* and *selC* tRNAs are common sites for insertion of horizontally transferred genetic elements. Gene encoding the invasin-like protein is unusually large for a bacterial gene (10,548 bp) and the predicted protein contains 19 repeats of bacterial immunoglobulin-like domains. A PSORTB search indicates that this protein possibly localizes in the outer membrane. Both T3SS structural genes and the *eip* island are reminiscent of the ETT2 locus and the *eip* island found in an enteroaggregative *E. coli* strain 042 (Chaudhuri, *et al.*, 2010, Ren, *et al.*, 2004, Sheikh, *et al.*, 2006) and in a cystitis strain UMN026 (Lescat M *et al.,* 2009). The ETT2 locus is distinct from the T3SS found in the locus of enterocyte effacement (*lee*) pathogenicity island in enteropathogenic *E. coli* (Ren, et al., 2004). ETT2 genes have been implicated in the pathogenesis of sepsis caused by *E. coli* (Ideses, *et al.*, 2005 and Ayres J *et al.*, 2012). Surprisingly, a UPEC strain was determined as the cause of hemolytic uremic syndrome (Tarr *et al.*, 1996) in a patient and that isolate exhibited a phenotype typically associated with T3SS-specific effectors. Efforts are under way to test the role of genes encoding T3SS structural and effector proteins and the invasin-like protein in uropathogenesis.

In summary, we present the genome sequences for five recent isolates of UPEC. Many of the genes previously implicated in the pathogenesis of UPEC were identified in these isolates. Our results also reveal that ETT2 genes are found in three out of five UPEC strains sequenced during this study. The availability of these additional genome sequences will be a valuable resource to the UPEC research community for further comparative genomic analyses.

Acknowledgments

S.S is a recipient of the Research Scholars Fellowship from the North Central section of American Urological Association administered by the Urology Care Foundation. D.A.R and T.H.H were supported by start-up funds from the State of Maryland and NIH grants 1U19AI090873 and RAI092828A. This work was supported by Public Health Service grants to H.L.T.M (AI059722, AI043363 and DK094777) from the National Institutes of Health.

The authors thank Ariel Brumbaugh, Stephanie Himpsl, Dr. Rob Ernst and the staff at University Health Service clinic at the University of Michigan, for help with sample collection. Samples were collected in accordance to the protocol approved by Institutional Review Board at the University of Michigan.

References

Alteri CJ, Mobley HL. *Escherichia coli* physiology and metabolism dictates adaptation to diverse host microenvironments. Curr Opin Microbiol. 2012; 15:3–9. [PubMed: 22204808]

Angiuoli SV, Salzberg SL. Mugsy: fast multiple alignment of closely related whole genomes. Bioinformatics. 2011; 27:334–342. [PubMed: 21148543]

Ayres JS, Trinidad NJ, Vance RE. Lethal inflammasome activation by a multi-drug resistant pathobiont upon antibiotic disruption of the microbiota. Nat Med. 2012; 18:799–806. [PubMed: 22522562]

Aziz RK, Bartels D, Best AA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008; 9:75. [PubMed: 18261238]

Brumbaugh AR, Mobley HL. Preventing urinary tract infection: progress toward an effective *Escherichia coli* vaccine. Expert Rev Vaccines. 2012; 11:663–676. [PubMed: 22873125]

Brzuszkiewicz E, Bruggemann H, Liesegang H, et al. How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains. Proc Natl Acad Sci U S A. 2006; 103:12879–12884. [PubMed: 16912116]

Chaudhuri RR, Sebaihia M, Hobman JL, et al. Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative *Escherichia coli* strain 042. PLoS ONE. 2010; 5:e8801. [PubMed: 20098708]

Gupta K, Hooton TM, Naber KG, et al. International clinical practice guidelines for the treatment of acute uncomplicated cystitis and pyelonephritis in women: A 2010 update by the Infectious Diseases Society of America and the European Society for Microbiology and Infectious Diseases. Clin Infect Dis. 2011; 52:e103–120. [PubMed: 21292654]

Hagan EC, Lloyd AL, Rasko DA, Faerber GJ, Mobley HL. *Escherichia coli* global gene expression in urine from women with urinary tract infection. PLoS Pathog. 2010; 6:e1001187. [PubMed: 21085611]

Hernandez D, Francois P, Farinelli L, Osteras M, Schrenzel J. De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. Genome Res. 2008; 18:802–809. [PubMed: 18332092]

Hooton TM. Clinical practice. Uncomplicated urinary tract infection. N Engl J Med. 2012; 366:1028–1037. [PubMed: 22417256]

Ideses D, Gophna U, Paitan Y, Chaudhuri RR, Pallen MJ, Ron EZ. A Degenerate Type III Secretion System from Septicemic *Escherichia coli* Contributes to Pathogenesis. J Bacteriol. 2005; 187:8164–8171. [PubMed: 16291689]

Lescat M, Calteau A, Hoede C, et al. A module located at a chromosomal integration hot spot is responsible for the multidrug resistance of a reference strain from *Escherichia coli* clonal group A. Antimicrob Agents Chemother. 2009; 53:2283–2288. [PubMed: 19364861]

Lloyd AL, Rasko DA, Mobley HL. Defining genomic islands and uropathogen-specific genes in uropathogenic *Escherichia coli*. J Bacteriol. 2007; 189:3532–3546. [PubMed: 17351047]

Lloyd AL, Henderson TA, Vigil PD, Mobley HL. Genomic islands of uropathogenic *Escherichia coli* contribute to virulence. J Bacteriol. 2009; 191:3469–3481. [PubMed: 19329634]

Miyazaki J, Ba-Thein W, Kumao T, Akaza H, Hayashi H. Identification of a type III secretion system in uropathogenic *Escherichia coli*. FEMS Microbiol Lett. 2002; 212:221–228. [PubMed: 12113938]

Rasko DA, Myers GS, Ravel J. Visualization of comparative genomic analyses by BLAST score ratio. BMC Bioinformatics. 2005; 6:2. [PubMed: 15634352]

Ren CP, Chaudhuri RR, Fivian A, Bailey CM, Antonio M, Barnes WM, Pallen MJ. 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. 2004; 186:3547–3560. [PubMed: 15150243]

Russo TA, Johnson JR. Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli*: ExPEC. J Infect Dis. 2000; 181:1753–1754. [PubMed: 10823778]

Russo TA, Johnson JR. Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. Microbes Infect. 2003; 5:449–456. [PubMed: 12738001]

Sahl JW, Steinsland H, Redman JC, Angiuoli SV, Nataro JP, Sommerfelt H, Rasko DA. A comparative genomic analysis of diverse clonal types of enterotoxigenic *Escherichia coli* reveals pathovar-specific conservation. Infect Immun. 2011; 79:950–960. [PubMed: 21078854]

Sheikh J, Dudley EG, Sui B, Tamboura B, Suleman A, Nataro JP. EilA, a HilA-like regulator in enteroaggregative *Escherichia coli*. Mol Microbiol. 2006; 61:338–50. [PubMed: 16762026]

Sommer DD, Delcher AL, Salzberg SL, Pop M. Minimus: a fast, lightweight genome assembler. BMC Bioinformatics. 2007; 8:64. [PubMed: 17324286]

Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006; 22:2688–2690. [PubMed: 16928733]

Tarr PI, Fouser LS, Stapleton AE, Wilson RA, Kim HH, Vary JC Jr, Clausen CR. Hemolytic–uremic syndrome in a six-year-old girl after a urinary tract infection with shiga-toxin–producing *Escherichia coli* O103:H2. N Engl J Med. 1996; 335:635–638. [PubMed: 8687518]

Welch RA, Burland V, Plunkett G 3rd, et al. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc Natl Acad Sci U S A. 2002; 99:17020–17024. [PubMed: 12471157]

Wong AR, Pearson JS, Bright MD, et al. Enteropathogenic and enterohaemorrhagic *Escherichia coli*: even more subversive elements. Mol Microbiol. 2011; 80:1420–1438. [PubMed: 21488979]

Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008; 18:821–829. [PubMed: 18349386]

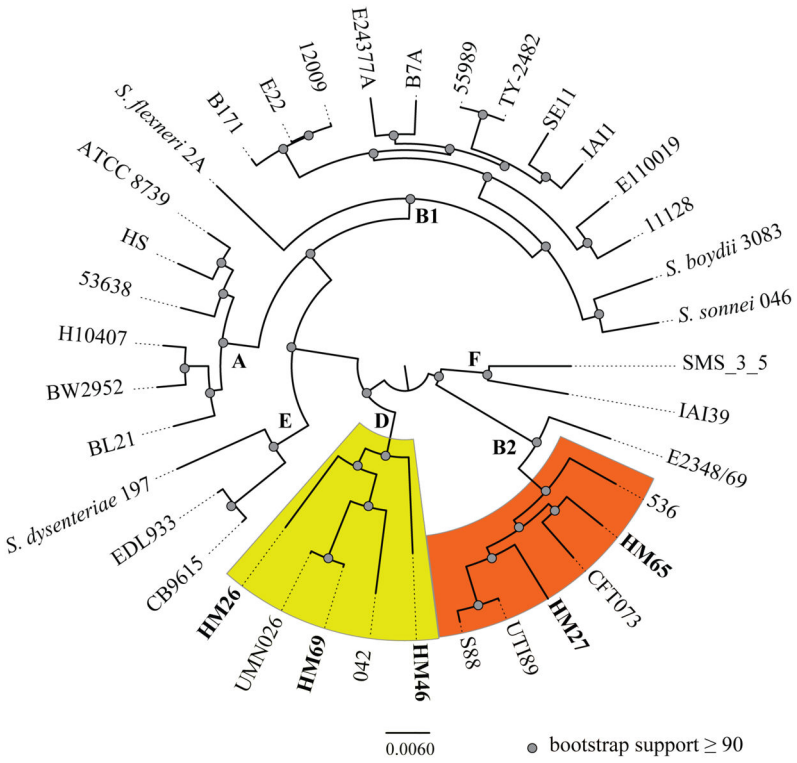

Fig. 1.

Phylogenomic analysis of the five recent UPEC isolates sequenced in this study compared to a collection of diverse *E. coli* and *Shigella* genomes available in the public domain. Genomes used for whole genome alignments are available in GenBank at the following GenBank Accession Numbers: EDL933 (NC_002655), CB9615 (NC_013941), 11128 (NC_013364), E2348/69 (NC_011601), B171 (AAJX00000000), 12009 (NC_013353), E22 (AAJV00000000), E110019 (AAJW00000000), 53638 (AAKB00000000), HS (NC_009800), ATCC 8739 (NC_010468), BL21 (NC_012947), BW2952 (NC_012759), H10407 (NC_017633), TY-2482 (AFOG01000000), *S. flexneri* 2A 2457T (NC_004741), *S. sonnei* 046 (NC_007384), *S. boydii* 3083 (NC_010658), B7A (AAJT02000000), E24377A (NC_009801), IAI1 (NC_011741), SE11 (NC_011415), 55989 (NC_011748), CFT073 (NC_004431), S88 (NC_011742), UTI89 (NC_007946), 536 (NC_008253), 042 (FN554766), UMN026 (NC_011751), *S. dysenteriae* Sd197 (NC_007606), IAI39 (NC_0117500 and SMS-3-5 (NC_010498). The genomes were aligned using Mugsy (Angiuoli & Salzberg, 2011), and a maximum-likelihood phylogeny with 100 bootstrap replicates was inferred using RAxML v7.2.8 (Stamatakis, 2006). Bootstrap support at all branch points were ≥90.

Table 1

Sequencing statistics and genome characteristics

Isolate	Phylogroup	No. reads	Genome size (bp)	%GC	No. genes[a]	Sequence Coverage	Conserved Genes[b]	Divergent Genes[c]
HM26	D	11,660,136	5,271,678	50.64	5199	223	3,137	1,677
HM27	B2	12,778,406	5,166,851	50.51	5150	250	3,158	1,640
HM46	D	12,952,066	4,967,159	50.78	4904	263	3,084	1,456
HM65	B2	11,907,652	5,162,282	50.49	5154	233	3,177	1,741
HM69	D	11,722,448	5,374,332	50.70	5420	220	3,109	1,973
Ave		12,204,142	5,188,460	50.62	5165	238	3,133	1,697

[a] Gene numbers are based on RAST (Rapid Annotation using Subsystem Technology)

[b] Number of genes that are conserved (blast score ratio of ≥ 0.8) among UPEC strains CFT073, 536, F11, UTI89, UMN026 and all the HM isolates

[c] Number of genes that are divergent (blast score ratio of < 0.8 to > 0.4) among UPEC strains CFT073, 536, F11, UTI89, UMN026 and all the HM isolates

Table 2

Prevalence of select urovirulence factors

Virulence Factor	HM 26	HM 27	HM 46	HM 65	HM 69
Iron Uptake					
Enterobactin R	+	+	+	+	+
Salmochelin R	+	+	−	+	−
Aerobactin R	−	−	−	−	+
Yersiniabactin R	+	+	−	+	+
ChuA, Heme R	+	+	+	+	+
Hma, Heme R	−	+	−	−	−
Toxins					
Hemolysin A	−	+	−	+	−
Cnf	−	−	−	+	−
Fimbriae					
Type 1	+	+	+	+	+
Pap	−	+	−	+	−
F1C	−	+	−	+D	−

R, receptor; D, disrupted; +, found in genome sequence; and −, not present in genome sequence