Genotypic analyses of uropathogenic *Escherichia coli* based on *fimH* single nucleotide polymorphisms (SNPs)

Sara Y. Tartof,¹ Owen D. Solberg² and Lee W. Riley^{1,3}

Correspondence

Lee W. Riley

lw Riley@berkeley.edu

¹Division of Epidemiology, School of Public Health, University of California, Berkeley, CA, USA

²Department of Integrative Biology, University of California, Berkeley, CA, USA

³Division of Infectious Diseases, University of California, Berkeley, CA, USA

The application of genotyping techniques for subtyping uropathogenic *Escherichia coli* has contributed to better understanding of the epidemiology of community-acquired urinary tract infection (UTI). However, the current techniques are hampered by limited reproducibility, poor discriminatory power, labour-intensive performance or high cost. A screening test that is sequence-based would provide an inexpensive, reproducible way to subtype *E. coli* isolates. Such a test, if also discriminatory, would be highly useful for epidemiological studies. The discriminatory ability of 12 putative virulence genes (*fimH*, *fliD*, *fliM*, *iha*, *motA*, *papA/H*, *kpsMTII*, *fepE*, *fimA*, *flgA*, *malG*, *purD*) was evaluated based on single nucleotide polymorphisms (SNPs) in nine uropathogenic *E. coli* isolates, all previously found to belong to a single multilocus sequence type (MLST) complex (ST69). An additional 25 epidemiologically well-characterized *E. coli* isolates belonging to 12 distinct MLST clonal complexes were analysed for *fimH* SNP. None of the 12 genes except *fimH* were able to further discriminate the nine ST69-complex strains. Isolates belonging to the 12 non-ST69 MLST groups were separated into 10 *fimH* SNP subgroups. While *fimH* SNP analysis may not be an appropriate phylogenetic method, it offers discriminatory power similar to that of MLST and could be used as a simple, inexpensive screening test for epidemiological studies of uropathogenic *E. coli*.

Received 28 February 2007

Accepted 25 June 2007

INTRODUCTION

The burden of urinary tract infections (UTIs) in the United States is substantial. From 2001 to 2002, UTIs accounted for 7.88 million ambulatory care visits (Schappert & Burt, 2006). UTI also represents one of the most common hospital-acquired infections. The predominant pathogen in both complicated and uncomplicated UTIs is *Escherichia coli* (Valiquette, 2001).

Genotyping tests have enhanced our understanding of the epidemiology of UTI caused by uropathogenic *E. coli*, including helping to characterize modes of transmission, source and risk factors for infection. For example, strain typing techniques have demonstrated that UTIs can occur as community-wide outbreaks, which had not been commonly considered (Phillips *et al.*, 1988). Manges *et al.* (2001) subtyped uropathogenic *E. coli* isolates by serotype, Enterobacterial Repetitive Intergenic Consensus-PCR (ERIC-PCR), PFGE and virulence factor profile, and found

that a single *E. coli* clonal group (CgA) accounted for nearly half of community-acquired trimethoprim-sulfamethoxazole-resistant UTIs in women in one university community (Manges *et al.*, 2001). This and other studies have suggested that a proportion of community-acquired UTIs may be caused by *E. coli* disseminated from one or more point sources, such as a contaminated food product (Johnson *et al.*, 2002, 2003; Manges *et al.*, 2006).

Both ERIC-PCR and PFGE are based on comparison of gel electrophoresis band patterns, which are often subject to inter-laboratory variability and not amenable to simple standardization. Sequence-based genotyping methods are becoming increasingly used in epidemiological studies of infectious diseases, including UTIs. Multilocus sequence typing (MLST), a sequence-based genotyping technique first introduced by Maiden *et al.* (1998), provides reproducibility, comparability and transferability between laboratories. MLST is based on sequencing 'housekeeping' genes, which are under stabilizing selection. Therefore, this system may not be optimal for distinguishing highly closely related strains. For *E. coli*, a standardized MLST was shown to be more discriminatory than the ERIC-PCR-based typing method, but less discriminatory than PFGE (Tartof

Abbreviations: ERIC-PCR, Enterobacterial Repetitive Intergenic Consensus-PCR; ExPEC, extraintestinal pathogenic *Escherichia coli*; MLST, multilocus sequence type; SNP, single nucleotide polymorphism; UTI, urinary tract infection.

et al., 2005). PFGE is labour-intensive, time-consuming, and suffers drawbacks associated with comparison of band patterns in electrophoresis gels. However, MLST based on multiple gene sequencing would not be amenable for screening a large number of isolates.

It has been suggested that the *E. coli* strains that cause most UTIs constitute genetically distinct groups of *E. coli*, characterized by the presence of multiple virulence factors (reviewed by Johnson, 1991; Mobley *et al.*, 1994). These virulence factors include adhesins, siderophores, toxins, polysaccharides and other products (Eden *et al.*, 1976; Guyer *et al.*, 2002; Johnson & Kuskowski, 2000; Johnson & Stell, 2000; Johnson, 2003; Russo & Johnson, 2000; Russo *et al.*, 2001; Timmis *et al.*, 1985). These properties help the bacteria circumvent host defences to allow entry into a normally sterile site, replicate, and stimulate an inflammatory response. Virulence factors are frequently under selective pressure driven by host innate and acquired immunity. This may contribute to antigenic diversification of proteins directly exposed to host defence systems. Therefore, genes encoding such factors are more likely to undergo mutations over a short period of time compared with genes necessary for basic metabolic processes, such as the 'housekeeping' genes used in MLST. We reasoned that such genes may be more useful for studying epidemiological events that occur over a short time period or geographical distance.

Other investigators have applied this idea to distinguish between so-called extraintestinal pathogenic *E. coli* (ExPEC) and non-ExPEC strains. These applications have generally made dichotomous determinations (presence or absence) of these targets by DNA hybridization probes or by PCR (Johnson *et al.*, 2001, 2005b). In this study, we sought to determine whether a sequence-based comparison of these so-called virulence factor genes could further discriminate uropathogenic and other ExPEC organisms, which may then be applied to better elucidate modes of transmission of community-acquired *E. coli* infections.

METHODS

Selection of genes. Twelve genes (*fimH*, *fliD*, *fliM*, *iha*, *motA*, *papA/H*, *kpsMTII*, *fepE*, *fimA*, *flgA*, *malG*, *purD*) were selected as targets for sequence analysis. Four were previously identified as putative virulence genes of uropathogenic *E. coli* (*fimH*, *iha*, *papA/H*, *kpsMTII*) (Johnson *et al.*, 2000; Johnson & Stell, 2000; Sokurenko *et al.*, 2004), five genes were associated with flagella or type 1 fimbriae shown to contribute to the virulence of uropathogenic *E. coli* (*fliD*, *fliM*, *motA*, *fimA*, *flgA*) (Connell *et al.*, 1996; Francis *et al.*, 1994; Iino *et al.*, 1988; Ikeda *et al.*, 1985; Lane *et al.*, 2005; Macnab, 1987) and three additional genes were included based on their relatively low level of nucleotide sequence conservation observed in alignments of these genes in strains CFT073 and F11 (*fepE*, *malG*, *purD*). Genes were also selected based on the likelihood of their being present as a single copy in uropathogenic strains of *E. coli*. Eight new primer pairs were designed for this study (*fepE*, *fimA*, *flgA*, *fliD*, *fliM*, *malG*, *motA*, *purD*). Primers were designed to amplify a product of approximately 650–1100 bp, and to include the highest number of observed single nucleotide polymorphisms (SNPs) in a sequence alignment of K12,

CFT073 and O157:H7 *E. coli* reference strains. Four additional primer pairs were previously reported (Johnson *et al.*, 2000; Johnson & Stell, 2000). Nucleotide sequence and citation information for the primers used to amplify the selected gene segments and the expected PCR product lengths are shown in Table 1.

Bacterial isolates and analysis. Six human uropathogenic isolates (102, 220, PY3, A925, A1707, G), two human blood isolates (W55291, X19714) and one animal *E. coli* isolate (An559), all previously shown to belong to clonal group CgA by ERIC-PCR and as ST69 complex by MLST, comprised a core group; they were sequenced for each of the 12 selected genes (Table 2). The CgA ERIC-PCR electrophoretic pattern has been described previously as four predominant bands of approximately 1145, 1029, 908 and 720 bp (Manges *et al.*, 2001). These isolates were selected from varying geographical and host sources to maximize the potential to detect differences. Six of these isolates (102, W55291, 220, PY3, X19714, A925) were identical by MLST (designated ST69) (Tartof *et al.*, 2005). Three isolates (A1707, An559, G) differed by one nucleotide in one of the seven MLST genes. All nine isolates grouped together in an SNP dendrogram based on concatenated sequences of the MLST alleles (Tartof *et al.*, 2005). All isolates had been previously typed by PFGE, and all were distinguishable by this method (Manges *et al.*, 2001; Ramchandani *et al.*, 2005). Sequence information for CFT073 and K12 was obtained from GenBank.

***fimH*.** *fimH* was analysed for an additional 25 previously well-characterized *E. coli* isolates. This target was specifically selected for analyses of all the study isolates for the following reasons: (a) it is present in both commensal and pathogenic variants of *E. coli*, thus satisfying the typeability criterion for its use for genotyping; (b) sequences of its alleles are well-represented in GenBank; and (c) it is known to be under strong selective pressure and thus likely to show a high degree of sequence heterogeneity. The 25 isolates (including K12 and CFT073) included strains which had been previously typed by ERIC-PCR and MLST (Table 2). The isolates were from the following sources: (a) eight isolates (six CgA, two non-CgA) from women from California, USA, with symptoms of UTI, who were seen at a university health service (consecutively enrolled into a study between 11 October 1999 and 31 January 2000), and five trimethoprim-sulfamethoxazole-resistant isolates (two CgA, three non-CgA) from Minnesota, USA (obtained from students who were seen at university health services with uncomplicated cystitis and were enrolled in a study between June 1998 and August 1999); (b) five animal isolates belonging to CgA by ERIC2 PCR provided by the Gastroenteric Disease Center at Pennsylvania State University (University Park, PA, USA); (c) nine human bacteraemia isolates (six CgA, three non-CgA) from San Francisco General Hospital provided by Dr Francoise Perdreau-Remington; and (d) five other isolates recovered from humans (three CgA, two non-CgA) for geographical comparison and provided by James R. Johnson, MD of Minneapolis VA Medical Center, University of Minnesota, USA. Sequence information for CFT073 and K12 was obtained from GenBank.

DNA isolation and PCR. *E. coli* isolates were cultured and DNA was extracted as previously described (Tartof *et al.*, 2005). Amplifications were carried out in a total volume of 25 µl with 1 µl template DNA, 2 µl of each 10 µM primer (Table 1), 2.5 µl 10× buffer, 15 mM MgCl₂, 2.5 µl 2 mM dNTP mix, and 0.1 µl AmpliTaq Gold (Applied Biosystems). Reaction conditions used were 2 min denaturation at 95 °C, 33 cycles of 30 s denaturation at 94 °C, 30 s annealing at 58 °C, 1 min extension at 72 °C, and a final additional 7 min at 72 °C in an Applied Biosystems GeneAmp PCR System 2400 thermocycler.

Sequencing and sequence analysis. All PCR products were purified for sequencing with the Qiagen QIAquick PCR Purification kit. The forward and reverse PCR primers were used for sequencing of

Table 1. Reference information for genes and PCR primers

Gene	Gene product/function	Accession no.	ORF length (bp)	PCR amplicon (bp)	Primers (5'–3')	Primer reference
<i>fepE</i>	Ferric enterobactin transport protein FepE	NP_752605	1134	1106	TGTCATCACTGAATATTAAACAGGG ATGGCATCCAGAAAACAGGA	Present study
<i>fimA</i>	Type-1 fimbrial protein, A chain precursor	NP_757241	606	693	TGCTTTCCTCTATGAGTCAA GATACTGAACCTTGAAGGTC	Present study
<i>fimH</i>	FimH protein precursor	NP_757248	912	506	TCGAGAACGGATAAGCCGTGG GCAGTCACCTGCCCTCCGGTA	Johnson & Stell (2000)
<i>flgA</i>	Flagellar basal body P-ring biosynthesis protein	NP_753252	660	907	GTCACACTGGTGTGCTGGTGA GCGGACAGGTACAATTACAG	Present study
<i>fliD</i>	Flagellar hook-associated protein	NP_754231	1407	693	CAAATCAGCAATCGTCGTTT ATGGCGACGTTGTTGACTGT	Present study
<i>fliM</i>	Flagellar motor switch protein	NP_754254	1050	550	ATGGGCGATAGTATCTTTCTC CGAACGCACGTACTCAACTT	Present study
<i>iha</i>	Outer-membrane receptor: ferrienterochelin and colicins	NP_755485	2091	827	CTGGCGGAGGCTCTGAGATCA TCCTTAAGCTCCCGCGGCTGA	Johnson <i>et al.</i> (2000)
<i>kpsMTII</i>	Group II capsular polysaccharide synthesis	NP_755573	777	272	GCGCATTTGCTGATACTGTTG CATCCAGACGATAAGCATGAGCA	Johnson & Stell (2000)
<i>malG</i>	Maltose transport system permease protein MalG	NP_756854	891	831	GGCAAAGTCACCCACAGGT TCAAACCTTCATTCGCGTGCC	Present study
<i>motA</i>	Flagellar motor protein	NP_754197	888	691	GCCTTGAGCACTCTATCAA GGTGCGTAACCGTTCAGATT	Present study
<i>papA/H</i>	PapA and PapH proteins	NP_755467, NP_755466	600, 588	736	ATGGCAGTGGTGTCTTTGGTG CGTCCCACCATAACGTGCTCTTC	Johnson & Stell (2000)
<i>purD</i>	Phosphoribosylamine–glycine ligase	NP_756817	1290	1028	GCCGAGAGAAGCGCGTTCAT CATTAATGGCGGAAGTGCGC	Present study

Table 2. Sequence types (ST) and ST complexes

Isolate*	Source	Serotype	CgA	MLST ST designation	MLST complex designation
102†	UTI-California	O11	Yes	ST69	ST69
160‡	UTI-California	O11	Yes	ST69	ST69
403†	UTI-California	O11	Yes	ST69	ST69
431†	UTI-California	O11	Yes	ST69	ST69
220‡	UTI-California	O77	Yes	ST69	ST69
44‡	UTI-California	O77	Yes	ST69	ST69
A1707‡	UTI-Maryland	O17, O77	Yes	ST407	ST69
A925‡	UTI-Barcelona	O77	Yes	ST69	ST69
PY3‡	UTI-Los Angeles	O17, O77	Yes	ST69	ST69
G‡	UTI-Minnesota	O17	Yes	NA	ST69
Y‡	UTI-Minnesota	O77	Yes	ST69	ST69
An559‡	Cow	O17	Yes	ST408	ST69
An298‡	Cow	O11	Yes	ST101	ST101
An45	Pig	O11	Yes	ST101	ST101
An300	Cow	O11	Yes	ST101	ST101
An437	Water	O77	Yes	ST394	ST69
H42937	Blood	O15	Yes	ST393	ST31
M32569	Blood	O15	Yes	ST393	ST31
H3708‡	Blood	O17	Yes	ST69	ST69
S48427‡	Blood	O77	Yes	ST69	ST69
W55291‡	Blood	O77	Yes	ST69	ST69
X19714‡	Blood	O86	Yes	ST69	ST69
SEQ 111	UTI-California	NA	No	NA	NA
SEQ 329	UTI-California	NA	No	ST355	ST73
A219	UTI-Toulouse	NA	No	ST23	ST23
A563	UTI-Munich	NA	No	ST10	ST10
Jo J	UTI-Minnesota	NA	No	ST73	ST73
Jo K	UTI-Minnesota	NA	No	ST73	ST73
Jo L	UTI-Minnesota	NA	No	ST73	ST73
W38035	Blood	NA	No	ST144	No complex
F591	Blood	NA	No	ST131	No complex
W18993	Blood	NA	No	ST12	No complex

NA, Not available.

*Isolates listed in bold-type sequenced for all 12 virulence genes.

†Indistinguishable by PFGE

‡Typed by PFGE.

the *fimH* sequences. Sequencing was performed at the University of California at Berkeley Sequencing Facility. The facility runs a 25 cycle sequencing reaction with the following program: 96 °C for 10 s, 50 °C for 5 s, 60 °C for 4 min.

Forward and reverse strand DNA sequence traces were visually inspected and edited with Chromas and BioEdit version 7.0.1. Finished sequences were aligned with CLUSTAL W. Nucleotide sequences from two complete *E. coli* genomes, K12 (NC_000913) and CFT073 (NC_004431), were trimmed and included in the alignment.

Dendrogram construction. The 422 bp alignment of a portion of the *fimH* gene for 32 pathogenic *E. coli* strains and of the two reference sequences was imported into MEGA version 3.1 (Kumar *et al.*, 2004) for phylogenetic analysis. The *fimH* sequence used to create the dendrogram is located at position 4 547 233–4 547 654 on the K12 genome. For the same strains, an MLST alignment was made from the 3423 bp concatenation of sequences of the seven genes used in

standardized *E. coli* MLST (*adk*, *fumC*, *icd*, *purA*, *gyrB*, *recA*, *mdh*) (Tartof *et al.*, 2005). For both alignments, neighbour-joining trees were produced by the Kimura two-parameter model of nucleotide substitution with uniform rates. Bootstrap values were computed by MEGA with 500 replications.

RESULTS AND DISCUSSION

The first part of the study was to find sequence variability in a selected set of genes for a panel of closely related clinical isolates from various sources. Nine *E. coli* isolates (102, W55291, 220, PY3, X19714, A925, A1707, G, An559), identified as CgA by ERIC-PCR and as ST69-complex members by MLST but as distinguishable by PFGE, were analysed for sequence variation in the 12 candidate genes (*fimH*, *fliD*, *fliM*, *iha*, *motA*, *papA/H*, *kpsMTII*, *fepE*, *fimA*,

Fig. 1. Neighbour-joining trees constructed from the sequences of the *fimH* gene subsegment (a) and concatenated sequences of the seven standard MLST genes (b) described in the text. The 22 isolates previously identified as CgA are shown with shaded backgrounds. Bootstrap values over 50 % are indicated on the corresponding nodes. *An559 strain fell out of the ST69 complex when analysed for *fimH* SNP.

flgA, *malG*, *purD*) (Table 1). We found that the selected gene segments, except *fimH*, had identical sequences among the nine strains, indicating that these gene targets did not contain any SNPs that could discriminate these strains beyond those discriminated by the standardized MLST procedure.

At the *fimH* locus, we compared nucleotide sequence variation in *fimH* in 34 (32 clinical and 2 reference strains) epidemiologically well-characterized isolates. These strains consisted of 16 ST69-complex and 18 non-ST69-complex members from various sources. This set of 34 strains was previously typed by the standardized seven-gene MLST procedure and reported by Tartof *et al.* (2005). We found that the SNP analysis of *fimH* could separate the 34 isolates into subgroups similar to those separated by the seven-gene-based MLST procedure.

In the *fimH*-sequence-based dendrogram, 10 distinct terminal groups representing 10 unique sequences were found. By comparison, the dendrogram constructed from seven concatenated MLST genes distinguished 16 distinct terminal groups. However, three of those sequences differed by only one SNP from the ST69 sequence type and are included in the same MLST complex. The 16 isolates identified as ST69 complex by MLST were mostly

retained as a single clade group in the *fimH* dendrogram (Fig. 1). The differences included a cow isolate (An559) which left the group, and a UTI isolate (no. 111) and the K12 strain which joined the group. In both dendrograms, three animal CgA isolates constituted a distinct branch (An300, An298, An45), Minnesota UTI CgA *E. coli* isolates (J and K) remained clustered, and two San Francisco human bacteraemia CgA isolates (H42937 and M32569) formed another cluster. With the exception of the animal isolate (An559), all CgA isolates which grouped with ST69 by MLST constituted a clade by the *fimH* SNP analysis.

FimH is a mannose-binding subunit protein located at the tip of type 1 fimbriae (Abraham *et al.*, 1988; Klemm & Christiansen, 1987; Krogfelt *et al.*, 1990). Phenotypic variants of FimH are predominantly the product of SNPs in *fimH* (Hommais *et al.*, 2003; Schembri *et al.*, 2000; Sokurenko *et al.*, 1997, 1998). We selected *fimH* for our initial analysis for several reasons. (a) FimH is a critical determinant of tropism for the urinary tract and vaginal epithelium for extraintestinal *E. coli* (Connell *et al.*, 1996; Hung *et al.*, 2002; Klumpp *et al.*, 2002). A urovirulent phenotype is associated with genetic variants of this protein and hence is potentially clinically relevant (Hommais *et al.*, 2003). (b) The vast majority of both intestinal and extraintestinal *E. coli* express type 1 fimbriae. In one study,

among 31 putative virulence factor genes examined in a collection of 63 *E. coli* urine isolates from patients with cystitis, pyelonephritis and prostatitis, *fimH* was the most prevalent (Johnson *et al.*, 2005a). Thus *fimH* satisfies the typeability criteria of a genotyping test. (c) It has been reported that non-synonymous mutations accumulate at the *fimH* locus at a high rate (Sokurenko *et al.*, 2004). Therefore, the level of discrimination at this locus is likely to be sufficiently high enough for studying uropathogenic *E. coli* UTI epidemiological events that occur over short time periods or in restricted geographical settings.

Of course, the other 11 genes examined in this study may offer a similar level of discrimination as did *fimH* for the same 34 isolates. However, not all of the other 11 genes are necessarily present in both commensal and ExPEC isolates. Unlike *fimH*, some of the other genes are not restricted to the members of the *Enterobacteriaceae* (e.g. *fliD*, *fliM*, *motA*, *flgA*, *malG*, *purD*). Finally, within a clonal complex, *fimH* acquires point mutations more frequently than *fimA* or any of the genes used in MLST (Weissman *et al.*, 2006), and many *fimH* sequences from a variety of *E. coli* isolates are deposited in the nucleic acid sequence database for comparison. Since none of the 11 genes studied were able to further discriminate the ST69-complex strains, there is no added advantage of using the other genes.

Interestingly, by *fimH* SNP analysis, strain An559 isolated from a cow in 1988 in the USA fell into a distinct cluster group outside of the ST69 complex, whereas by MLST it is considered a member of the ST69 complex (ST408) (Fig. 1b). This strain was also previously shown to be 94% similar by PFGE to one of the human UTI CgA isolates (Ramchandani *et al.*, 2005). Thus it appears that in some of the strains, *fimH* SNP analysis provides more information than MLST.

One limitation of this study is that the *fimH* allele is prone to horizontal transfer, as evidenced by the comparison of the two dendrograms (Fig. 1). Thus *fimH* SNP-based groupings would not be considered phylogenetically meaningful taxa as are the groupings based on MLST. However, this *fimH* SNP analysis is not meant to provide yet another phylogenetic tool. Rather, the main goal of this study was to develop a simple, sequence-based screening test that could be applied to a large number of UTI *E. coli* isolates to characterize recent epidemiological events. Plasmids are also subject to frequent horizontal transfers, and yet plasmid profile analyses have been shown to be highly useful for epidemiological studies of enteric pathogens (Holmberg *et al.*, 1984).

The results from this study suggest that this *fimH* SNP-based genotyping method could serve as a relatively inexpensive, highly reproducible screening test for epidemiological studies of uropathogenic *E. coli*. A large collection of *E. coli* isolates could initially be screened for *fimH* sequence. Those that are indistinguishable can then be tested further by MLST or PFGE. This test could replace

ERIC-PCR for screening *E. coli* isolates, especially in laboratories with access to a sequencing facility.

ACKNOWLEDGEMENTS

We thank the staff of Tang Health Center, James R. Johnson, Francoise Perdreau-Remington, Amee Manges and Chitrita DebRoy for providing isolates. We thank Remi Ajiboye for technical support. This work was supported by the NIH grant number RO1 AI059523.

REFERENCES

- Abraham, S. N., Sun, D., Dale, J. B. & Beachey, E. H. (1988). Conservation of the D-mannose-adhesion protein among type 1 fimbriated members of the family Enterobacteriaceae. *Nature* **336**, 682–684.
- Connell, I., Agace, W., Klemm, P., Schembri, M., Marild, S. & Svanborg, C. (1996). Type 1 fimbrial expression enhances *Escherichia coli* virulence for the urinary tract. *Proc Natl Acad Sci U S A* **93**, 9827–9832.
- Eden, C. S., Hanson, L. A., Jodal, U., Lindberg, U. & Akerlund, A. S. (1976). Variable adherence to normal human urinary-tract epithelial cells of *Escherichia coli* strains associated with various forms of urinary-tract infection. *Lancet* **i**, 490–492.
- Francis, N. R., Sosinsky, G. E., Thomas, D. & DeRosier, D. J. (1994). Isolation, characterization and structure of bacterial flagellar motors containing the switch complex. *J Mol Biol* **235**, 1261–1270.
- Guyer, D. M., Radulovic, S., Jones, F. E. & Mobley, H. L. (2002). Sat, the secreted autotransporter toxin of uropathogenic *Escherichia coli*, is a vacuolating cytotoxin for bladder and kidney epithelial cells. *Infect Immun* **70**, 4539–4546.
- Holmberg, S. D., Wachsmuth, I. K., Hickman-Brenner, F. W. & Cohen, M. L. (1984). Comparison of plasmid profile analysis, phage typing, and antimicrobial susceptibility testing in characterizing *Salmonella typhimurium* isolates from outbreaks. *J Clin Microbiol* **19**, 100–104.
- Hommais, F., Gouriou, S., Amorin, C., Bui, H., Rahimy, M. C., Picard, B. & Denamur, E. (2003). The FimH A27V mutation is pathoadaptive for urovirulence in *Escherichia coli* B2 phylogenetic group isolates. *Infect Immun* **71**, 3619–3622.
- Hung, C. S., Bouckaert, J., Hung, D., Pinkner, J., Widberg, C., DeFusco, A., Auguste, C. G., Strouse, R., Langermann, S. & other authors (2002). Structural basis of tropism of *Escherichia coli* to the bladder during urinary tract infection. *Mol Microbiol* **44**, 903–915.
- Iino, T., Komeda, Y., Kutsukake, K., Macnab, R. M., Matsumura, P., Parkinson, J. S., Simon, M. I. & Yamaguchi, S. (1988). New unified nomenclature for the flagellar genes of *Escherichia coli* and *Salmonella typhimurium*. *Microbiol Rev* **52**, 533–535.
- Ikeda, T., Asakura, S. & Kamiya, R. (1985). “Cap” on the tip of *Salmonella* flagella. *J Mol Biol* **184**, 735–737.
- Johnson, J. R. (1991). Virulence factors in *Escherichia coli* urinary tract infection. *Clin Microbiol Rev* **4**, 80–128.
- Johnson, J. R. (2003). Microbial virulence determinants and the pathogenesis of urinary tract infection. *Infect Dis Clin North Am* **17**, 261–278.
- Johnson, J. R. & Kuskowski, M. (2000). Clonal origin, virulence factors, and virulence. *Infect Immun* **68**, 424–425.
- Johnson, J. R. & Stell, A. L. (2000). Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. *J Infect Dis* **181**, 261–272.

- Johnson, J. R., Russo, T. A., Tarr, P. I., Carlino, U., Bilge, S. S., Vary, J. C., Jr & Stell, A. L. (2000). Molecular epidemiological and phylogenetic associations of two novel putative virulence genes, *iha* and *iroN_{E. coli}* among *Escherichia coli* isolates from patients with urosepsis. *Infect Immun* **68**, 3040–3047.
- Johnson, J. R., Delavari, P., Kuskowski, M. & Stell, A. L. (2001). Phylogenetic distribution of extraintestinal virulence-associated traits in *Escherichia coli*. *J Infect Dis* **183**, 78–88.
- Johnson, J. R., Manges, A. R., O'Bryan, T. T. & Riley, L. W. (2002). A disseminated multidrug-resistant clonal group of uropathogenic *Escherichia coli* in pyelonephritis. *Lancet* **359**, 2249–2251.
- Johnson, J. R., Murray, A. C., Gajewski, A., Sullivan, M., Snippes, P., Kuskowski, M. A. & Smith, K. E. (2003). Isolation and molecular characterization of nalidixic acid-resistant extraintestinal pathogenic *Escherichia coli* from retail chicken products. *Antimicrob Agents Chemother* **47**, 2161–2168.
- Johnson, J. R., Kuskowski, M. A., Gajewski, A., Soto, S., Horcajada, J. P., Jimenez de Anta, M. T. & Vila, J. (2005a). Extended virulence genotypes and phylogenetic background of *Escherichia coli* isolates from patients with cystitis, pyelonephritis, or prostatitis. *J Infect Dis* **191**, 46–50.
- Johnson, J. R., Owens, K., Gajewski, A. & Kuskowski, M. A. (2005b). Bacterial characteristics in relation to clinical source of *Escherichia coli* isolates from women with acute cystitis or pyelonephritis and uninfected women. *J Clin Microbiol* **43**, 6064–6072.
- Klemm, P. & Christiansen, G. (1987). Three fim genes required for the regulation of length and mediation of adhesion of *Escherichia coli* type 1 fimbriae. *Mol Gen Genet* **208**, 439–445.
- Klumpp, D. J., Forrestal, S. G., Karr, J. E., Mudge, C. S., Anderson, B. E. & Schaeffer, A. J. (2002). Epithelial differentiation promotes the adherence of type 1-piliated *Escherichia coli* to human vaginal cells. *J Infect Dis* **186**, 1631–1638.
- Krogfelt, K. A., Bergmans, H. & Klemm, P. (1990). Direct evidence that the FimH protein is the mannose-specific adhesin of *Escherichia coli* type 1 fimbriae. *Infect Immun* **58**, 1995–1998.
- Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**, 150–163.
- Lane, M. C., Lockatell, V., Monterosso, G., Lamphier, D., Weinert, J., Hebel, J. R., Johnson, D. E. & Mobley, H. L. (2005). Role of motility in the colonization of uropathogenic *Escherichia coli* in the urinary tract. *Infect Immun* **73**, 7644–7656.
- Macnab, R. M. (1987). Flagella. In *Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biology*, pp. 70–83. Edited by F. C. Neidhardt, K. B. Low, J. Ingraham, B. Magasanik, M. Schaechter & H. E. Umbarger. Washington, DC: American Society for Microbiology.
- Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K. & other authors (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**, 3140–3145.
- Manges, A. R., Johnson, J. R., Foxman, B., O'Bryan, T. T., Fullerton, K. E. & Riley, L. W. (2001). Widespread distribution of urinary tract infections caused by a multidrug-resistant *Escherichia coli* clonal group. *N Engl J Med* **345**, 1007–1013.
- Manges, A. R., Natarajan, P., Solberg, O. D., Dietrich, P. S. & Riley, L. W. (2006). The changing prevalence of drug-resistant *Escherichia coli* clonal groups in a community: evidence for community outbreaks of urinary tract infections. *Epidemiol Infect* **134**, 425–431.
- Mobley, H. L., Island, M. D. & Massad, G. (1994). Virulence determinants of uropathogenic *Escherichia coli* and *Proteus mirabilis*. *Kidney Int Suppl* **47**, S129–S136.
- Phillips, I., Eykyn, S., King, A., Gransden, W. R., Rowe, B., Frost, J. A. & Gross, R. J. (1988). Epidemic multiresistant *Escherichia coli* infection in West Lambeth Health District. *Lancet* **1**, 1038–1041.
- Ramchandani, M., Manges, A. R., DebRoy, C., Smith, S. P., Johnson, J. R. & Riley, L. W. (2005). Possible animal origin of human-associated, multidrug-resistant, uropathogenic *Escherichia coli*. *Clin Infect Dis* **40**, 251–257.
- Russo, T. A. & Johnson, J. R. (2000). Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli*: ExPEC. *J Infect Dis* **181**, 1753–1754.
- Russo, T. A., Carlino, U. B. & Johnson, J. R. (2001). Identification of a new iron-regulated virulence gene, *ireA*, in an extraintestinal pathogenic isolate of *Escherichia coli*. *Infect Immun* **69**, 6209–6216.
- Schappert, S. M. & Burt, C. W. (2006). Ambulatory care visits to physician offices, hospital outpatient departments, and emergency departments: United States, 2001–02. *Vital Health Stat* **13** **159**, 1–66.
- Schembri, M. A., Sokurenko, E. V. & Klemm, P. (2000). Functional flexibility of the FimH adhesin: insights from a random mutant library. *Infect Immun* **68**, 2638–2646.
- Sokurenko, E. V., Chesnokova, V., Doyle, R. J. & Hasty, D. L. (1997). Diversity of the *Escherichia coli* type 1 fimbrial lectin. Differential binding to mannoses and uroepithelial cells. *J Biol Chem* **272**, 17880–17886.
- Sokurenko, E. V., Chesnokova, V., Dykhuizen, D. E., Ofek, I., Wu, X. R., Krogfelt, K. A., Struve, C., Schembri, M. A. & Hasty, D. L. (1998). Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin. *Proc Natl Acad Sci U S A* **95**, 8922–8926.
- Sokurenko, E. V., Feldgarden, M., Trintchina, E., Weissman, S. J., Avagyan, S., Chattopadhyay, S., Johnson, J. R. & Dykhuizen, D. E. (2004). Selection footprint in the FimH adhesin shows pathoadaptive niche differentiation in *Escherichia coli*. *Mol Biol Evol* **21**, 1373–1383.
- Tartof, S. Y., Solberg, O. D., Manges, A. R. & Riley, L. W. (2005). Analysis of a uropathogenic *Escherichia coli* clonal group by multi-locus sequence typing. *J Clin Microbiol* **43**, 5860–5864.
- Timmis, K. N., Boulnois, G. J., Bitter-Suermann, D. & Cabello, F. C. (1985). Surface components of *Escherichia coli* that mediate resistance to the bactericidal activities of serum and phagocytes. *Curr Top Microbiol Immunol* **118**, 197–218.
- Valiquette, L. (2001). Urinary tract infections in women. *Can J Urol* **8** (Suppl. 1), 6–12.
- Weissman, S. J., Chattopadhyay, S., Aprikian, P., Obata-Yasuoka, M., Yarova-Yarova, Y., Stapleton, A., Ba-Thein, W., Dykhuizen, D., Johnson, J. R. & Sokurenko, E. V. (2006). Clonal analysis reveals high rate of structural mutations in fimbrial adhesins of extraintestinal pathogenic *Escherichia coli*. *Mol Microbiol* **59**, 975–988.