JCM Accepts, published online ahead of print on 9 November 2011
J. Clin. Microbiol. doi:10.1128/JCM.00640-11
Copyright © 2011, American Society for Microbiology and/or the Listed Authors/Institutions. All Rights Reserved.

JCM Accepts published online ahead of print

1 **Title:** Identification of *Escherichia coli* genes associated with urinary
2 tract infections
3
4 **Running title:** *E. coli* genes associated with urinary tract infections

5 **Author Names:** Bin-Hsu Mao[1], Yung-Fu Chang[2], Joy Scaria[2], Chih-Ching Chang[3],
6 Li-Wei Chou[4], Ni Tien[5], Jiunn-Jong Wu[6, 10], Chin-Chung Tseng[7], Ming-Cheng Wang[7],
7 Chao-Chin Chang[8], Yuan-Man Hsu[9], Ching-Hao Teng[1, 10 *]
8
9 **Authors' Affiliations (the address of the institutions at which the work was**
10 **perform):**
11 Institute of Molecular Medicine, National Cheng Kung University Medical College,
12 Tainan City, Taiwan.[1]
13 Department of Population Medicine and Diagnostic Sciences, College of Veterinary
14 Medicine, Cornell University, Ithaca, NY, USA.[2]
15 Department of Environmental and Occupational Health, National Cheng Kung
16 University Medical College, Tainan City, Taiwan.[3]
17 Department of Physical Medicine and Rehabilitation, China Medical University
18 Hospital, Taichung City, Taiwan.[4]
19 Department of Laboratory Medicine, China Medical University Hospital, Taichung

JCM Accepts published online ahead of print

20 City, Taiwan.[5]

21 Department of Medical Laboratory Science and Biotechnology, National Cheng Kung

22 University Medical College, Tainan City, Taiwan. [6]

23 Division of Nephrology, Department of Internal Medicine, National Cheng Kung

24 University Hospital, Tainan City, Taiwan. [7]

25 Graduate Institute of Microbiology and Public Health, School of Veterinary Medicine,

26 National Chung Hsing University, Taichung City, Taiwan. [8]

27 Department of Biological Science and Technology, China Medical University,

28 Taichung City, Taiwan.[9]

29 Center of Infectious Disease and Signaling Research, National Cheng Kung

30 University, Tainan City, Taiwan.[10]

31 ***Corresponding author:**

32 Mailing address: Institute of Molecular Medicine, National Cheng Kung

33 University Medical College, 4th F, 367 Sheng Li Road, Tainan City,

34 Taiwan.

35 E-mail: chteng@mail.ncku.edu.tw

36 Phone: 886-6-2353535 ext. 4595

37 FAX: 886-6-2095845

38

JCM Accepts published online ahead of print

39 **Abstract:**

40 *Escherichia coli* is the most common cause of urinary tract infections (UTIs). *E. coli*
41 genes epidemiologically associated with UTIs are potentially valuable in developing
42 strategies for treating and/or preventing such infections as well as differentiating
43 uropathogenic *E. coli* from non-uropathogenic *E. coli*. To identify *E. coli* genes
44 associated with UTIs in humans, we combined microarray-based and polymerase
45 chain reaction-based analyses to investigate different *E. coli* source groups derived
46 from feces of healthy humans and from patients with cystitis, pyelonephritis, or
47 urosepsis. The *cjrABC-senB* gene cluster, *sivH*, *sisA*, *sisB*, *eco274*, and *fbpB*, were
48 identified to be associated with UTIs. Of these, *cjrABC-senB*, *sisA*, *sisB*, and *fbpB* are
49 known to be involved in urovirulence in the mouse model of ascending UTI. Our
50 results provide evidence to support their roles as urovirulence factors in human UTIs.
51 In addition, the newly identified UTI-associated genes were mainly found in members
52 of phylogenetic groups B2 and/or D.

53

54

55

56

57 ## Introduction

JCM Accepts published online ahead of print

58 *Escherichia coli* is the most common cause of urinary tract infections (UTIs),
59 including acute cystitis, pyelonephritis, and urosepsis, three common and clinically
60 distinct UTI syndromes. It is widely accepted that uropathogenic *E. coli* (UPEC)
61 originate from the distal gut microbiota (8, 13, 15). To cause ascending UTI, UPEC
62 needs to overcome and adapt to different distinct host environments, such as the
63 bladder, the kidneys, and even the blood stream. Accordingly, UPEC tends to be
64 distinct from the commensal *E. coli* strains in the intestinal tract in having extra
65 virulence genes allowing their successful transition from the intestinal tract to the
66 urinary tract.

67 An epidemiological association between an *E. coli* gene and UTIs may suggest
68 that the gene itself encodes a factor contributing to urovirulence or has a genetic
69 linkage to such a gene. Therefore, the genes associated with UTIs are potentially
70 valuable in differentiating UPEC from non-uropathogenic *E. coli* and in the
71 development of strategies for managing and preventing this particular type of disease.

72 Palaniappan *et al.* have developed an oligonucleotide spotted microarray
73 containing probes representing 342 *E. coli* genes to differentiate *E. coli* pathotypes
74 (27). A majority of the genes are derived from the UPEC strain CFT073, the
75 enterohemorrhagic *E. coli* strain EDL933, and the commensal *E. coli* K12 strain
76 MG1655. The remaining genes are derived from other *E. coli* strains capable of

JCM Accepts published online ahead of print

77 causing intestinal infections. The association of the majority of the 342 genes in the
78 array with UTIs has not yet been investigated, except for 36 uropathogenic genes
79 included among them (27).
80 To identify *E. coli* genes associated with UTIs, we used the microarray
81 developed by Palaniappan *et al.* to screen for the genes potentially associated with
82 UTIs and then performed a polymerase chain reaction (PCR)-based analysis with a
83 larger bacterial sample size to confirm these genes' epidemiologic associations with
84 UTIs. One gene cluster and 5 individual genes (hereafter the gene cluster and 5
85 individual genes are referred to as MIGs, abbreviation of "microarray-identified
86 genes") were associated with UTIs. Of these, the gene cluster and 3 of the individual
87 genes have recently been shown to be involved in urovirulence in the mouse model of
88 UTI (6, 20, 22). In addition, we analyzed the phylogenetic distribution of the MIGs,
89 and assessed the correlations between these MIGs, as well as between these genes and
90 15 known virulence genes.
91
92 **Materials and Methods**
93 ***E. coli* isolates and patients**
94 The UTI-associated isolates in this study, cystitis, pyelonephritis, and urosepsis
95 isolates, were collected from two hospitals in Taiwan - the China Medical University

JCM Accepts published online ahead of print

96 Hospital (CMUH) at Taichung city in central Taiwan, and National Cheng Kung
97 University Hospital (NCKUH) at Tainan city in southern Taiwan. The UTI-associated
98 isolates used in this study were a subset of the 2,206 *E. coli* strains isolated from the
99 urine specimens submitted to the diagnostic laboratories of the two hospitals between
100 Jun 2006 and April 2007 (1,619 isolates from CMUH and 587 isolates from NCKUH).
101 According to the diagnostic criteria mentioned below, among the 1,619 isolates from
102 CMUH, 696 isolates were from cystitis patients, 421 isolates from pyelonephritis
103 patients, and 372 from urosepsis patients. Among the 587 isolates from NCKUH, 229
104 isolates were from cystitis patients, 141 isolates from pyelonephritis patients, and 94
105 isolates from urosepsis patients. Together, there were 925 cystitis isolates, 562
106 pyelonephritis isolates, and 466 urosepsis isolates. Out of each type of UTI isolates
107 we randomly selected 67 (7.2%) cystitis isolates, 72 (12.8%) pyelonephritis isolates
108 and 64 (13.8%) urosepsis isolates for this study. The biliary tract infection
109 (BTI)-associated bacteremia *E. coli* strains (n=24) were obtained from the blood
110 specimens of BTI patients with bacteremia at NCKUH between September 2004 and
111 November 2007. In addition, 115 commensal fecal isolates were collected from the
112 feces of healthy donors between Jun 2006 and April 2007. Each bacterial isolate in
113 this study was derived from a different patient or healthy donor.

114 According to the diagnostic criteria of UTIs previously described (35), the

JCM Accepts published online ahead of print

115 prerequisite for patients with UTIs was that their fresh urine samples contained
116 bacterial counts of ≥ 10^5 cfu/ml. Cystitis was defined by the presence of dysuria,
117 urinary frequency, and/or lower abdominal pain. Pyelonephritis was based on the
118 presence of body temperature ≥ 38.3 °C, and flank pain and/or costovertebral angle
119 tenderness, with or without the syndrome of cystitis. Urosepsis was defined by the
120 presence of bacteremia in addition to UTI syndromes.
121 The diagnostic criteria for BTI were fever, abdominal pain in the right upper
122 quadrant, and/or jaundice, with imaging demonstrating the presence of acute
123 cholecystitis or acute cholangitis (37).
124 **DNA microarray analysis**
125 For genomic DNA preparation, a single colony of bacteria were inoculated into
126 LB broth and was incubated at 37°C for 12 hrs to the concentration of around 3×10^9
127 CFU/ml. 4ml of the O/N bacterial culture was then subjected to genomic DNA
128 extraction with Qiagen DNeasy kit (QIAGEN, Valencia, CA, USA.) according to
129 manufacturer's instructions. The integrity of the genomic DNA was verified by 1%
130 agarose gel electrophoresis stained with ethidium bromide. 2ug the genomic DNA
131 was used for the enzyme digestion and labeling processes, and then concentrated to a
132 volume of 12 ul for the following hybridization with the microarray. Microarray
133 printing, bacterial genomic DNA labeling, microarray hybridization, and data

JCM Accepts published online ahead of print

134 acquisition were performed as previously described (27). Each slide had triplicate
135 spots of each gene. Similarly, the derived data were analyzed as previously described
136 but with a modification in the cutoff criteria determining the positive signals. For each
137 gene the background-subtracted median fluorescence intensities of each triplicate spot
138 were averaged and $\log_2$ transformed, designated as L values. A gene was considered
139 positive when its L value was greater than the value of the mean (M) substracted by
140 the standard deviation (SD) of the L values derived from 48 genes commonly present
141 in the *E. coli* strains MG1655, CFT073, and EDL933 (27).

142 Fifteen of the 342 *E. coli* gene probes were selected as positive control for the
143 microarray experiments, because the target genes they hybridize with were identified
144 in all the current available *E. coli* strains with complete genome sequences (some of
145 them have been demonstrated as *E. coli* essential genes) (Supplement table 1). Twenty
146 *Salmonella enterica* serovar Typhimurim *LT2*-specific gene probes and the
147 autoblanks (Spots of dimethyl sulfoxide without any probes) were used as the
148 negative controls (Supplement table 1) (27).

149 **PCR-based genotyping and phylogenetic typing**

150 The frequencies of the genes screened in by the microarray analysis were
151 determined by PCR-based analysis. The primers were designed to target the
152 conserved regions of the MIGs (Table 1). The PCR reactions were heated to 95°C in

JCM Accepts published online ahead of print

153 an automated thermal cycler for 5 min, followed by 30 cycles of denaturation (95°C,
154 45 sec), annealing (59°C, 45 sec), and extension (72°C, 50 sec). Taq polymerase was
155 used in the reactions. The phylogenetic groups of the 342 *E. coli* isolates were
156 determined based on the PCR-based method described previously (5). The
157 frequencies of the 15 selected known virulence genes of extraintestinal pathogenic *E.*
158 *coli* were determined by PCR-based assays, using primers and PCR conditions as
159 described previously (3, 16).

160 The following *E. coli* strains were used as controls for the PCR-based analysis
161 of the MIGs and the known virulence genes. The *E. coli* strains which served as
162 positive controls included CFT073 (*sivH*, *shiA*, *sisA*, *sisB*, *fbpB*, *papGII*, *chuA*, *ompT*,
163 *sat*, *iha*, *usp*, *ireA*, *iroN*, and *hlyA*), UTI89 (*cjrA*, *cjrB*, *cjrC*, *senB*, *cnf1*, *sfaS*, and
164 *ibeA*), EDL933 (*eco274*), J96 (*papGI* and *papGIII*), and one of the UTI-associated
165 clinical isolates A53, which was identified to harbor *afa/dra* by sequencing the PCR
166 product amplified by the *afa/dra* specific primers (*afa/dra*). MG1655 was used as a
167 negative control for all the genes except *ompT*. An *ompT* deletion mutant of *E. coli*
168 strain RS218 was constructed and served as the negative control of *ompT*.

169 The PCR amplification was done in a 25µl reaction mixture. Amplifications
170 were carried out in Eppendorf Mastercycler[@] gradient thermal cycler (Eppendorf,
171 Hamburg, Germany). The PCR products were electrophoresed in 1.5% agarose gels,

JCM Accepts published online ahead of print

172 stained with ethidium bromide, and photographed using an AlphaImager @ HP system
173 (Cell Biosciences, Inc, Santa Clara, CA, USA). The sizes of the products were
174 determined by comparing them with a 100-bp DNA ladder (Fermentas Inc., Glen
175 Burnie, MD, USA) run on the same gel. All PCR tests were performed 3 times with
176 independently prepared boiled lysates. Additional investigations were further
177 conducted, if discrepancies between the independent assays occurred.

178 **Statistical analysis**

179 Comparisons involving the frequencies of a given gene in different groups were
180 measured by using two-tailed Fisher's exact test. *P* value < 0.05 was arbitrarily set as
181 the threshold for statistical significance. Correlations between genes were measured
182 by using the Fisher's exact test (two tailed). Because of multiple comparisons, *P* value
183 < 0.01 was arbitrarily set as the threshold for statistical significance, with *P* value <
184 0.05 as borderline statistical significance (12, 14).

185 **Results**

186 ***E. coli* genes potentially associated with UTIs**

187 To screen for *E. coli* genes that are potentially associated with UTIs, a
188 microarray-based pilot study was conducted by using 40 *E. coli* isolates which were
189 divided into 4 source groups according to the clinical syndromes (or conditions) with
190 which they were associated; fecal isolates from healthy humans (n=8 isolates), cystitis

JCM Accepts published online ahead of print

191 isolates (n=12), pyelonephritis isolates (n=10), and urosepsis isolates (n=10) (Table 2).
192 The bacterial isolates were subjected to gene profiling using the DNA microarray
193 developed by Palaniappan *et al.* (27). With the gene profiles derived from the
194 microarray data, we determined the frequencies of each microarray-detectable gene in
195 the 4 source groups of *E. coli* (Supplement table 1). Based on the results of this pilot
196 study, eight genes (*cjrA*, *cjrB*, *cjrC*, *SenB*, *sivH*, *shiA*, *eco274*, and *fbpB*) were selected
197 for a further study, since they exhibited higher frequencies in the UTI-associated
198 source groups than those in the fecal source group. While these differences in
199 frequencies were consistently observed between UTI-associated and fecal source
200 groups, only the *sivH* in the pyelonephritis isolates reached statistically significance
201 (Table 2). Based on BLAST and literature searches, their potential functions were
202 predicted (Table 3). Among them, *cjrA*, *cjrB*, *and cjrC* are located in the operon
203 *cjrABC*, and *senB* is located downstream of the operon with its 5' end partially
204 overlapping with *cjrC* (32). Thus, *cjrABC* and *senB* are referred to as *cjrABC-senB* in
205 this study. According to recent studies using the mouse model of ascending UTI,
206 *cjrABC-senB*, *shiA*, and *fbpB* are involved in the virulence of UPEC (6, 20, 22).
207 However, this is the first time their association with human UTIs has been studied.
208 **PCR confirmed associations between the MIGs and UTIs**

JCM Accepts published online ahead of print

209 To further confirm whether the genes identified from the microarray-based pilot
210 study are associated with UTIs, a PCR-based analysis with a larger sample size was
211 performed. Here, another source group containing *E. coli* isolates causing biliary tract
212 infection (BTI)-associated bacteremia (BTI-associated isolates) was included in
213 addition to the original 4 source groups. The 5 source groups contained a total of 342
214 *E. coli* isolates (Table 4). We compared the distributions of 6 known virulence genes
215 (*papGII*, *cnfI*, *hlyA*, *chuA*, *iroN*, and *usp*) in the UTI-associated source groups with
216 those in the fecal isolates, all the virulence genes exhibited significantly higher
217 frequencies in the UTI-associated source groups then the fecal isolates (Supplement
218 table 2).

219 According to the PCR analysis with all the *E. coli* isolates, 107 isolates had at
220 least one gene of *cjrABC-senB*. A total of 104 out of the 107 isolates (97%) contained
221 all 4 genes, suggesting that their coexistence is common in *E. coli*. Therefore, we
222 investigated the distribution of the intact gene cluster, *cjrABC-senB,* by detecting the
223 4 genes separately with different primer pairs specific to each gene (Table 1). Two
224 *shiA* homologs, *sisA* and *sisB*, are identified in *E. coli* (22), but the microarray used in
225 this study could not differentiate between them. SisA and SisB share 86% identity at
226 the level of amino acid sequences, with their N-termini being the most divergent parts.
227 The distributions of *shiA*, *sisA*, and *sisB* were investigated separately by using *shiA*

JCM Accepts published online ahead of print

228 primers able to detect the sequence common to *sisA* and *sisB*, and primers specific to
229 *sisA* and *sisB*, in the PCR-based analysis (Table 1).

230 Overall, the results confirmed that these MIGs are associated with UTIs. The
231 frequencies of most of the MIGs in each of the UTI-associated source groups (cystitis,
232 pyelonephritis, and urosepsis) were significantly higher than that in the fecal source
233 group (Table 4). Although the frequencies of eco274 in the cystitis and fecal isolates
234 were not significantly different, its frequencies in the pyelonephritis and urosepsis
235 isolates were significantly higher than that in the fecal isolates.

236 When the three UTI-associated source groups were compared, *shiA* and *sisA*
237 showed significantly higher frequencies in the pyelonephritis isolates than in the
238 cystitis isolates (Table 4). When the UTI-associated groups were compared with the
239 BTI-associated group, the distributions of *cjrABC-senB*, *sivH*, *shiA*, *sisA*, and *fbpB*
240 markedly favored the UTI-associated bacterial isolates (Table 4). *sisB* tended to
241 exhibit higher frequencies in the UTI-associated isolates than in the BTI-associated
242 isolates although the differences did not reach statistical significance. In addition,
243 when the BTI-associated isolates were compared with the fecal isolates, the
244 frequencies of the MIGs were not significantly different (data not shown).

245 **Phylogenetic distribution of the MIGs**

JCM Accepts published online ahead of print

246 The MIGs were mainly concentrated within phylogenetic groups B2 and/or D
247 (Table 5). *cjrABC-senB* and *eco274* showed significantly greater frequencies in group
248 D than in group B2, while *sivH* and *fbpB* were present with significantly higher
249 frequencies in group B2 than in group D. *sivH* and *fbpB* were almost entirely confined
250 to group B2 and were not found in group A or B1. The frequencies of *shiA*, *sisA*, and
251 *sisB* in group B2 and in group D isolates were not significantly different.

252 **Stratification of the MIGs by phylogeny**

253 To determine whether the associations of the MIGs with UTIs were still present
254 in individual phylogenetic groups, we further evaluated the distributions of the genes
255 in the fecal and UTI-associated isolates with stratification of individual phylogenetic
256 groups. Since the genes were mainly concentrated in groups B2 and/or D, and were
257 relatively rare in groups A and B1 (Table 6), only the group B2 and group D strains
258 were assessed.

259 In group B2, the distributions of *shiA*, *sisA*, and *sisB* still markedly favored the
260 UTI-associated isolates over the fecal isolates (Table 6). The frequencies of *sivH* in
261 the cystitis and pyelonephritis isolates of B2 bacteria were higher than those in the B2
262 fecal isolates, while the frequencies of *fbpB* in the cystitis and urosepsis isolates of B2
263 bacteria were higher than those in the B2 fecal isolates (Table 6).

JCM Accepts published online ahead of print

264 In group D, when the UTI-associated source groups were compared with the
265 fecal source group, significant difference in frequencies favoring the cystitis and
266 pyelonephritis isolates was still detected for *cjrABC-senB*, *shiA*, and *sisA*.

267 **Correlations of the MIGs with one another and with known virulence genes**

268 We performed pairwise comparisons of the MIGs with one another and with the
269 15 selected known virulence genes among the 203 UTI-associated *E. coli* isolates (the
270 total isolates in the source groups of pyelonephritis, cystitis, and urosepsis) (Table 7).

271 When the MIGs were compared with one another, both positive and negative
272 associations were detected. The positive associations seemed to be related to the
273 phylogenetic distribution of the *E. coli* strains which carried these genes. *sivH* and
274 *fbpB*, which were mainly concentrated in group B2 (Table 5), were positively
275 associated with each other. *cjrABC-senB* and *eco274*, whose frequencies in group D
276 were significantly higher than those in group B2 (Table 5), demonstrated a positive
277 association. *sisA* and *sisB*, whose frequencies in groups B2 and D were not
278 significantly different (Table 5), exhibited a marginal positive statistical association
279 (P=0.014; data not shown). In addition, negative associations were detected when
280 *sivH* was compared with *cjrABC-senB* and *eco274*.

281 As to the associations between MIGs with the 15 known virulence genes,
282 positive and negative associations were also detected (Fig. 1). Of note, the association

JCM Accepts published online ahead of print

283 patterns of *sivH* and eco274 with some of the known virulence genes were opposite.

284 *sivH* was positively associated with *cnf1*, *usp*, *ireA*, *iroN*, *sfaS*, while *eco274* was

285 negatively associated with these genes. In addition, *sivH* was negatively associated

286 with *sat* and *iha*, while *eco274* was positively associated with these genes (Table 7).

287 **Discussion**

288 This study is the first to identify the epidemiological associations of the *E. coli* genes,

289 *cjrABC-senB*, *sivH, sisA*, *sisB, eco274*, and *fbpB,* with UTIs by utilizing the

290 microarray and PCR-based analyses on fecal isolates and three distinct UTI

291 syndrome-associated isolates. Most of the MIGs were associated with all three distinct

292 UTI syndromes, cystitis, pyelonephritis, and urosepsis, while eco274 was associated

293 only with pyelonephritis and urosepsis. *cjrABC-senB*, *sisA*, *sisB*, and *fbpB* have

294 recently been found to be involved in the virulence of UPEC in the mouse model of

295 ascending UTI (6, 20, 22). Consistently, our results show these genes' associations

296 with human UTIs, well supporting their roles as *E. coli* virulence factors in human

297 UTIs. As for *eco274* and *sivH*, their roles in urovirulence have not yet been assessed

298 in the mouse model. However, they are likely to be urovirulence genes themselves or

299 have genetic linkage with such genes, based on their associations with UTIs. Thus,

300 these MIGs could be potential targets for developing preventive, and/or therapeutic

301 strategies for UTIs as well as potential markers of UPEC.

JCM Accepts published online ahead of print

302 *cjrABC-senB* has been shown to be present on the plasmids of enteroinvasive *E.*
303 *coli* (EIEC) and two UPEC strains, UTI89 and UMN026 (26, 32, 34). Interruption of
304 this gene cluster in UTI89 decreases the bacterium's ability for bladder colonization in
305 the early stage of the mouse model of UTI (6). The *cjrABC* operon is predicted to be
306 involved in iron acquisition, which may contribute to urovirulence (6). *senB* encodes
307 for the TieB protein, which may have some role in enterotoxicity of EIEC (26).
308 However, its role in UTIs is not yet clear. *senB* and the gene encoding the ShET-2
309 toxin, named *senA* (or *sen*), are located on a plasmid of an enteroinvasive *E. coli*
310 strain EI37 (26). It has been proposed that the polar effect caused by interruption of
311 *senB* may affect expression of *senA* (26). However, Soto *et al.* showed that *senA* was
312 not present in all of the 170 UPEC clinical isolates they examined (33), therefore it
313 seems unlikely that the association of *senB* with UTIs is strictly based on its affect on
314 *senA* expression.

315 *shiA* is primarily identified in the SHI-2 pathogenicity island (PAI) of *Shigella*
316 *flexneri* (25, 36). This gene has been shown to be involved in the down regulation of
317 inflammatory responses in both the rabbit ileal loop and mouse lung models of
318 shigellosis (10, 11). The *shiA* homologs, *sisA* and *sisB*, in the UPEC strain CFT073
319 have been shown to be involved in suppressing the host immune response, facilitating
320 bacterial colonization of the bladder and kidney during the initial stage of UTI in mice

JCM Accepts published online ahead of print

321 (22).

322 *fbpB* is located in the *fbp* locus, which contains the genes *fbpABCD* and is
323 potentially involved in iron acquisition (28). *fbpB* is predicted to encode a periplasmic
324 siderophore-binding protein (28). CFT073 contains two identical copies of *fbpABCD*
325 located in distinct genomic islands, PAI-CFT073-aspV and GI-CFT073-cobU (21).
326 The CFT073 mutant with deletion of the two *fbp* loci is significantly outcompeted by
327 the wild-type strain in the bladders and kidneys of mice (20).

328 *sivH* was originally identified in the CS54 island of *Salmonella enterica*
329 serotype Typhimurium (24). The *sivH* of *Salmonella* is known to be involved in the
330 colonization of the Peyer's patches in mice (17). Deletion of the genomic island RDI
331 13, which contains *sivH*, in the meningitis-associated *E. coli* strain RS218 decreases
332 the bacterium's ability to adhere to and invade human brain microvascular endothelial
333 cells (39). Whether SivH contributes to the urovirulence of UPEC is unknown.

334 *eco274* is classified as an EDL933-specific gene in the DNA microarray used in
335 our initial screening (27). This gene is located in the O island #148 of EDL933 (29).
336 Its role in urovirulence is unknown. When the three UTI-associated groups were
337 compared with the fecal isolates, it was noted that *eco274* was not associated with
338 cystitis, but rather only with pyelonephritis and urosepsis. Thus, *eco274* is likely to
339 encode a virulence factor involved only in pyelonephritis and urosepsis, or have a

JCM Accepts published online ahead of print

340 genetic linkage with a gene encoding such a virulence factor.

341 The distributions of several MIGs in the UPEC strains revealed in this study are
342 consistent with results from other related studies. The frequencies of *cjrABC-senB* in
343 the three UTI-associated groups ranged from 36% to 43% in the present study, which
344 is similar to the findings of Cusumano *et al.* that *senB* exists in 8 of 18 UPEC isolates
345 (44%) (6). Similarly, the higher frequencies of *sisA* in the UTI-associated groups
346 compared to those of *sisB* in our study (67%-80% versus 28%-31%) support the
347 hypothesis of Lloyd *et al* (20). that *sisA* is more prevalent than *sisB* in UPEC isolates.
348 Their hypothesis is based on the findings that in all sequenced bacterial species, *sisA*
349 is mainly distributed in extraintestinal pathogenic *E. coli* strains, while *sisB* is mainly
350 distributed in enteric strains (22). Also, our finding that no significant difference in
351 the distributions of *fbpB* between the cystitis and pyelonephritis isolates is consistent
352 with the findings of Parham *et al.* that the distributions of the *fbp* locus in these two
353 types of isolates are not significantly different. However, the *fbpB* frequencies in the
354 cystitis and pyelonephritis isolates in our study (45% and 44%, respectively) are
355 lower than those of the *fbp* locus in the same types of isolates in the study of Parham
356 *et al.*(58% and 59%, respectively). In addition, Lloyd *et al.* classified *fbpB* as
357 UPEC-specific based on an investigation of 11 UPEC and 4 fecal or commensal *E.*
358 *coli* strains, showing that *fbp*B is present in all the UPEC strains but not in the

JCM Accepts published online ahead of print

359 fecal/commensal strains (21). However, in our study *fbpB* was detected in 18 out of
360 the 115 fecal isolates (Table 4), suggesting that *fbpB* is not a UPEC specific gene,
361 although its frequencies in fecal isolates was significantly lower than those in the
362 UTI-associated isolates.

363 The significantly higher frequency of *shiA* in the pyelonephritis group than that
364 in the cystitis group may be due to the distribution of *sisA*. This is because the
365 frequency of the *shiA* distribution was the composite of the *sisA* and *sisB* distributions,
366 and only *sisA* but not *sisB* exhibited higher frequencies in the pyelonephritis group
367 than in the cystitis group. In addition, the significantly higher frequencies of *sisA* in
368 the pyelonephritis group may suggest that *sisA* plays a more important role in
369 pyelonephritis than in cystitis.

370 The distributions of the MIGs, which were mainly concentrated in groups B2
371 and/or D, are similar to those of most extraintestinal virulence genes, concentrated in
372 groups B2 and/or D as well (12, 15). Such accordance is supportive to our assertion
373 that the MIGs are potential virulence genes or have genetic linkage to such genes.

374 The associations of the MIGs with extraintestinal infections may be
375 syndrome-dependent (i.e. BTI versus UTI), because these genes were correlated with
376 the UTI-associated isolates, but not with the BTI-associated isolates, when compared
377 with the fecal isolates. Wang *et al.* have shown that *E. coli* strains responsible for BTI

JCM Accepts published online ahead of print

378 mainly belong to the phylogenetic groups A and B1(37), unlike the other
379 extraintestinal pathogenic strains, including UPEC, mainly belonging to phylogenetic
380 group B2 and, to a lesser extent, group D (7, 16, 30). Similarly, 50% (12/24) of the
381 BTI-associated isolates in this study belong to phylogenetic group A (data not shown).
382 Given our finding that these MIGs were mainly concentrated in phylogenetic groups
383 B2 and/or D, such distinct phylogenetic distribution of BTI-associated isolates might
384 be responsible for the syndrome-dependent associations. However, the possibility that
385 these genes are specifically associated with UTIs, but not other types of extraintestinal
386 infections, can not be excluded. A detailed study to assess more types of
387 extraintestinal infections caused by *E. coli* may be required.

388 The associations of *cjrABC-senB*, *sivH*, and *fbpB* with the UTI syndromes may
389 not necessarily be phylogenetic group dependent, although these associations were
390 observed exclusively in one of the investigated phylogenetic groups, B2 and D (Table
391 6). These observations may be due to the decreased sample size after stratification by
392 phylogeny.

393 Two genes having genetic linkage, such as their co-localization in the same
394 plasmid or genomic island, may result in a positive association. However, the MIGs
395 positively associated with each other were not found in the same genomic islands or
396 plasmids, according to the BLAST search on the completely sequenced *E. coli* strains.

JCM Accepts published online ahead of print

397 These observations suggest that such associations are not due to genetic linkages but
398 due to a process of co-selection, which may facilitate the pathogenesis in UTIs. As an
399 example, *cjrABC-senB* is located in plasmids of *E. coli* strains (26, 32, 34), while its
400 positively associated genes, *eco274* and *sisA*, are located in the chromosome (34, 38).
401 Also, *sivH* is located in a three-gene genomic island (named RDI 13 in RS218, as
402 mentioned) which is always inserted between *yfgJ* and *xseA* in the chromosome (2, 4,
403 34, 38), while *sisA* and *fbpB*, which are positively associated with *sivH*, are located in
404 other PAIs (2, 34, 38).

405 The positive association of *sisA* with the known virulence genes, *papGII, iha,*
406 and *sat*, may be due to genetic linkages among them (Table 7), because *sisA* and these
407 known virulence genes are located in a PAI, PAI-CFT073-pheV, in CFT073 (21).
408 However, the associations are not absolute. In addition, *sisA* was not associated with
409 *hly* which is also located in PAI-CFT073-pheV (21). These observations demonstrate
410 that the genetic linkages between virulence factors in a PAI are not constant,
411 supporting the suggestion of Johnson *et al.* that virulence genes may be transferred
412 horizontally, independent of the PAIs where they were originally located, in addition
413 to being transferred with the entire PAI (14).

414 It is known that UPEC requires multiple virulence genes in together to cause
415 infection (23). The virulence gene combination of a UPEC strain may determine the

JCM Accepts published online ahead of print

416 pathogenesis process employed by this strain to cause infection. Johnson *et al.* have
417 identified two groups of urovirulence genes. The member genes in the same group
418 exhibit positive associations, but in general negatively associate with genes in the
419 other group (14). Also, we found that *eco274* and *sivH*, which were negatively
420 associated with each other, exhibited opposite association patterns with a portion of
421 the known virulence genes (as mentioned in the Results section). These findings
422 imply that sets of virulence genes, with members that are for the most part discrete,
423 may exist among UPEC to direct bacteria through distinct pathways to cause UTIs.
424 However, a further study with more virulence factors to determine their distributions
425 and co-occurrence, is necessary to test this inference.

426 The expression of a bacterial gene in an environment may reflect the role of this
427 gene in bacterial adaptation to this particular environment (31). Thus, transcriptome
428 analyses of UPEC genes during UTIs may provide clues whether or not and how a
429 bacterial gene is involved in pathogenesis of UTIs. According to a recent microarray
430 analysis of *E. coli* global gene expression in 8 urine samples from different women
431 with UTIs, expression of *sisA* is detected in most of the urine samples (9), consistent
432 with the notion that this gene is a virulence factor in human UTIs. However, the
433 expression of *sisB* and *fbpA*, which is located upstream of *fbpB* in the *fbp* locus, was
434 not detected in these urine samples (9). This may be because *sisB* and the genes in the

JCM Accepts published online ahead of print

435 *fbp* locus are only transiently required for pathogenesis of human UTIs and the urine
436 samples only represent a stage of the infection that these genes are not involved in.
437 Alternatively, since most urovirulence genes only exist in a portion of UPEC strains,
438 the *E .coli* strains in these urine samples may not have harbored these potential
439 virulence genes.

440 In conclusion, the MIGs are potential targets for developing preventive and/or
441 therapeutic strategies to manage UTIs as well as potential markers for differentiating
442 UPEC from non-uropathgenic *E. coli*. Virulence factors of UPEC are good targets for
443 prevention and treatment of UTIs. For example, the FimH adhesin of type 1 fimbria is
444 responsible for colonization of UPEC on the uroepithelium of the bladder. FimH
445 antagonists have been developed as anti-adhesive drug for oral treatment of UTIs (18).
446 Also, FimH and iron receptors, such as IreA, Hma, and IutA are able to induce
447 protective immune response against UPEC infections (1, 19). Thus, *cjrABC-senB*,
448 *sisA*, *sisB*, and *fbpB*, which are involved in urovirulence of UPEC, are potential
449 therapeutic and/or preventive targets. In addition, all the MIGs are potential markers
450 for UPEC. Such markers may be valuable in public health for monitoring biological
451 threats, such as outbreaks of UTIs caused by *E. coli* and emergency of new virulence
452 *E. coli* strains, and also in basic microbiology research, such as studies in evolution
453 and classification of pathogenic *E. coli*. However, so far, none of the known

JCM Accepts published online ahead of print

454 urovirulence factors alone is sufficient to account for the virulence properties of
455 UPEC and most of the urovirulence genes only exist in a portion of UPEC strains.
456 Therefore, to develop effective and widely usable preventive and/or therapeutic
457 strategies to manage *E. coli*-caused UTIs, a combination of multiple urovirulence
458 genes to serve as the targets may be necessary. Accordingly, the properties of the
459 MIGs revealed in this study, including their prevalence, phylogenetic distribution, and
460 correlation patterns with other known virulence genes, may be beneficial for
461 designing such a gene combination for controlling these *E. coli*-caused diseases.

JCM Accepts published online ahead of print

462 Table 1. Primer sequences used in this study

Genes	Primer Sequence (5'-3')	Amplicon size (bp)
cjrA	AAAGGGTGGTCCTGGGAGAT	223
	ACGTCAGTTGCTGGCTTTCA	
cjrB	CGAAGTTCAGCCCGCTATGT	397
	GCTTTCCCAAGATGCCTCAG	
cjrC	AAACCTCAGCGCAAAATCGT	518
	AGGCTTCAGGAATGGGTTCA	
senB	CCGTTGAAAGATCCGAGACC	312
	GTTTGGGTAGACCGGCATGT	
sivH	TACAGCACGCGTAAACCGTA	866
	TGGCAGTACAGTTCCGATCA	
shiA	TCACCTTACTGGTATGAACTC	451
	TCCAGGGCCAGACATATTCA	
sisA	TTGCCCGACAGGAGAATGAC	360
	GCAGTATATGGCGTGCCTGT	
sisB	GAACGATAGATTATGCTTTG	518
	TCAGTACACTGAAGGCTCGC	
eco274	TTGACAAAGCCTGCCTGACC	207
	CCTCCAACCCGTGTTTTTGC	
fbpB	GCAAATCGCGCAGGATAAAG	821
	ACGCACAAGGAGGTGCGTAT	

463

464

JCM Accepts published online ahead of print

465 Table 2. The microarray-analysis-derived frequencies of the genes which were

466 potentially associated with UTIs

Gene	No. (%) of *E. coli* isolates			
	Fecal isolates (n=8)	Cystitis isolates (n=12)	Pyelonephritis isolates (n=10)	Urosepsis isolates (n=10)
cjrA	2 (25)	5 (42)	6 (60)	7 (70)
cjrB	2 (25)	5 (42)	6 (60)	7 (70)
cjrC	2 (25)	5 (42)	6 (60)	7 (70)
senB	2 (25)	5 (42)	6 (60)	7 (70)
sivH	2 (25)	5 (42)	8 (80)*	7 (70)
shiA	4 (50)	8 (67)	9 (90)	9 (90)
eco274	2 (25)	8 (67)	7 (70)	6 (60)
fbpB	4 (50)	8 (67)	8 (80)	8 (80)

467 * $P < 0.05$, pairwise comparisons between the indicated UTI-associated source group

468 with the fecal source group.

469

470 Table 3. The potential functions of the MIGs

Gene	Designation in microarray[a]	Potential function of the gene product[b]	Accession no. in the representative UPEC strains[c]	Completely genome-sequenced UPEC strains harboring the genes
cjrA	*cjrA*	Putative inner membrane protein	YP_538626	UTI89, UMN026
cjrB	*cjrB*	TonB-like protein	YP_538627	UTI89, UMN026
cjrC	*cjrC*	Putative TonB-dependent receptor	YP_538628	UTI89, UMN026
senB	*senB*	Enterotoxin Tie protein	YP_538629	UTI89, UMN026
sivH [d]	*eco293*	Putative intimin or invasin protein	NP_754913	UTI89, CFT073, UMN026, 536, IAI39
shiA	*eco294*	Potential suppressor of innate immune response	NP_755432 and NP_756354	CFT073, UMN026, IAI39,
eco274	*eco274*	Potential transcriptional regulator	YP_002410135	UMN026, IAI39
fbpB	*eco288*	Potential iron-chelating protein	NP_752239	UTI89, CFT073, 536

471

472 [a] The gene's designation used in the microarray described previously (27).

473 [b] The potential functions of all the genes are based on the BLAST search, except for that of *shiA*, which is based on the finding of Lloyd *et*

474 *al*.(22).

475 [c] The accession no. indicates UTI89-derived *cjrA*, *cjrB*, *cjrC*, and *senB*; CFT073-derived *shiA*, *sivH*, and *fbpB*;

476 IAI39-derived *eco274,* respectively.

477 [d] *sivH* is also named *sinH*.

JCM Accepts published online ahead of print

JCM Accepts published online ahead of print

478 Table 4. Distributions of the MIGs among 342 *Escherichia coli* isolates in different source groups

Gene	No. (%) of *E. coli* isolates					P[a]			
	Fecal isolates (*n*=115)	Cystitis isolates (*n*=67)	Pyelonephritis isolates (*n*=72)	Urosepsis isolates (*n*=64)	BTI isolates[b] (*n*=24)	Fecal vs Cystitis	Fecal vs Pyelonephritis	Fecal vs Urosepsis	Fecal vs BTI
cjrABC-senB	23 (20)	24 (36)	32 (44)	23 (36)	2 (9)	0.023	0.001	0.031	–
sivH	21 (18)	36 (54)	37 (51)	30 (47)	4 (13)	<0.001	<0.001	<0.001	–
shiA	40 (35)	47 (70)	62 (86)	54 (84)	6 (25)	<0.001	<0.001	<0.001	–
sisA	35 (30)	45 (67)	61 (85)	51 (80)	5 (22)	<0.001	<0.001	<0.001	–
sisB	10 (9)	19(28)	16(22)	20 (31)	3 (13)	0.001	0.016	<0.001	–
eco274	31 (27)	22 (33)	31 (43)	30 (47)	7(38)	–	0.026	0.009	–
fbpB	18 (16)	30 (45)	32 (44)	32 (50)	3 (9)	<0.001	<0.001	<0.001	–

479

480 *Table 4 continued on next page*

481

482

483 *Table 4, continued*

Gene	P^a					
	Cystitis vs Pyelonephritis	Cystitis vs Urosepsis	Pyelonephritis vs Urosepsis	BTI vs Cystitis	BTI vs Pyelonephritis	BTI vs Urosepsis
cjrABC-senB	–	–	–	0.016	0.001	0.015
sivH	–	–	–	0.002	0.004	0.013
shiA	0.025	–	–	<0.001	<0.001	<0.001
sisA	0.017	–	–	<0.001	<0.001	<0.001
sisB	–	–	–	–	–	–
eco274	–	–	–	–	–	–
fbpB	–	–	–	0.006	0.006	0.001

484

485 [a] Only *P* values of <0.05 (by Fisher's exact test) are shown.

486 [b] "BTI isolates" indicates BTI-associated bacteremia isolates.

487

JCM Accepts published online ahead of print

JCM Accepts published online ahead of print

488 Table 5. Phylogenetic distribution of the MIGs in the 342 *E. coli* isolates from all the source groups

Gene	No.(%) of *E. coli* isolates				P[a]				
	Group A (n=61)	Group B1 (n=15)	Group B2 (n=185)	Group D (n=81)	A vs B2	A vs D	B1 vs B2	B1 vs D	B2 vs D
cjrABC-senB	6 (10)	0 (0)	60 (32)	38 (47)	< 0.001	< 0.001	0.006	< 0.001	0.028
sivH	0 (0)	0 (0)	126 (68)	2 (2)	< 0.001	–	< 0.001	–	< 0.001
shiA	14 (23)	1 (7)	138 (75)	56 (69)	< 0.001	< 0.001	< 0.001	< 0.001	–
sisA	7 (11)	0 (0)	137 (74)	53 (65)	< 0.001	< 0.001	< 0.001	< 0.001	–
sisB	8 (13)	1 (7)	40 (22)	19 (23)	–	–	–	–	–
eco274	2 (3)	0 (0)	53 (29)	66 (81)	< 0.001	< 0.001	0.013	< 0.001	< 0.001
fbpB	0 (0)	0 (0)	111 (60)	4 (5)	< 0.001	–	< 0.001	–	< 0.001

489

490 [a] Only P values of <0.05 (by Fisher's exact test) are shown.

491

492

493 Table 6. Prevalence of the MIGs stratified by phylogeny

Gene	No.(%) of *E. coli* isolates				P^a		
	Fecal isolates	Cystitis isolates	Pyelonephritis isolates	Urosepsis isolates	Cystitis vs Fecal	Pyelonephritis vs Fecal	Urosepsis vs Fecal
Group B2	*n*=42	*n*=45	*n*=48	*n*=45			
cjrABC-senB	10 (24)	14 (31)	18 (38)	17 (38)	–	–	–
sivH	21 (50)	35 (78)	36 (75)	30 (67)	0.008	0.017	–
shiA	19 (45)	34 (76)	45 (94)	39 (87)	0.005	< 0.001	< 0.001
sisA	19 (45)	33 (73)	45 (94)	39 (87)	0.009	< 0.001	< 0.001
sisB	1(2)	14 (31)	12 (25)	13 (29)	< 0.001	0.002	0.001
eco274	15(36)	10 (22)	12 (25)	15 (33)	–	–	–
fbpB	17 (40)	29 (64)	30 (63)	32 (71)	0.032	–	0.005
Group D	*n*=25	*n*=13	*n*=20	*n*=16			
cjrABC-senB	8 (32)	10 (77)	14 (70)	6 (38)	0.016	0.017	–
sivH	0 (0)	1 (8)	1 (5)	0 (0)	–	–	–
shiA	12 (48)	12 (92)	17 (85)	13 (81)	0.012	0.013	–
sisA	11 (44)	12 (92)	16 (80)	12 (75)	0.005	0.018	–
sisB	4 (16)	4 (31)	4 (20)	5 (31)	–	–	–
eco274	16 (64)	12 (92)	17 (85)	15 (94)	–	–	–
fbpB	1(4)	1 (8)	2 (10)	0 (0)	–	–	–

494

495 [a] Only *P* values of < 0.05 (by Fisher's exact test) are shown.

496

JCM Accepts published online ahead of print

JCM Accepts published online ahead of print

497 Table 7. Associations of the MIGs with one another and with the selected known

498 virulence genes in the 231 UTI-associated *E. coli* isolates.

Identified	*cjrABC-senB*	*eco274*	*sivH*	*fbpB*	*sisA*	*sisB*
cjrABC-senB	NA	++	(++)	-	+	-
eco274	++	NA	(++)	-	-	-
sivH	(++)	(++)	NA	++	++	-
fbpB	-	-	++	NA	+	+
sisA	+	-	++	+	NA	-
sisB	-	-	-	+	-	NA
Known						
papG I	-	-	-	-	-	-
papG II	-	-	-	-	++	-
papG III	-	-	-	+	-	-
chuA	++	-	++	++	++	-
ompT	++	++	-	++	-	+
afa/draBC	-	-	-	-	-	-
sat	++	++	(++)	-	++	++
iha	++	++	(++)	-	++	++
cnf1	-	(++)	++	++	-	++
usp	-	(++)	++	++	++	-
ireA	(++)	(++)	++	-	++	-
iroN	(++)	(++)	++	++	-	-
sfaS	-	(+)	++	++	-	-
ibeA	-	-	+	+	(+)	-
hlyA	-	-	+	++	-	++

499

500 Note. Significant codes: -, $P \geq 0.01$; +, $P<0.01$; ++, $P<0.001$. Because of multiple
501 comparisons, P value < 0.01 was arbitrarily set as the threshold for statistical
502 significance, with P value < 0.05 as borderline statistical significance (12, 14).
503 Parentheses indicate negative associations. "Identified" and "Known" indicate the
504 MIGs and selected known virulence genes in this study, respectively.

505 **References**

506 1. **Alteri, C. J., E. C. Hagan, K. E. Sivick, S. N. Smith, and H. L. Mobley.**
507 2009. Mucosal immunization with iron receptor antigens protects against
508 urinary tract infection. PLoS Pathog **5:**e1000586.

509 2. **Brzuszkiewicz, E., H. Bruggemann, H. Liesegang, M. Emmerth, T.**
510 **Olschlager, G. Nagy, K. Albermann, C. Wagner, C. Buchrieser, L. Emody,**
511 **G. Gottschalk, J. Hacker, and U. Dobrindt.** 2006. How to become a
512 uropathogen: comparative genomic analysis of extraintestinal pathogenic
513 *Escherichia coli* strains. Proc Natl Acad Sci U S A **103:**12879-12884.

514 3. **Chapman, T. A., X. Y. Wu, I. Barchia, K. A. Bettelheim, S. Driesen, D.**
515 **Trott, M. Wilson, and J. J. Chin.** 2006. Comparison of virulence gene
516 profiles of *Escherichia coli* strains isolated from healthy and diarrheic swine.
517 Appl Environ Microbiol **72:**4782-4795.

518 4. **Chen, S. L., C. S. Hung, J. Xu, C. S. Reigstad, V. Magrini, A. Sabo, D.**
519 **Blasiar, T. Bieri, R. R. Meyer, P. Ozersky, J. R. Armstrong, R. S. Fulton, J.**
520 **P. Latreille, J. Spieth, T. M. Hooton, E. R. Mardis, S. J. Hultgren, and J. I.**
521 **Gordon.** 2006. Identification of genes subject to positive selection in
522 uropathogenic strains of *Escherichia coli*: a comparative genomics approach.
523 Proc Natl Acad Sci U S A **103:**5977-5982.

524 5. **Clermont, O., S. Bonacorsi, and E. Bingen.** 2000. Rapid and simple
525 determination of the *Escherichia coli* phylogenetic group. Appl Environ
526 Microbiol **66:**4555-4558.

527 6. **Cusumano, C. K., C. S. Hung, S. L. Chen, and S. J. Hultgren.** 2010.
528 Virulence plasmid harbored by uropathogenic *Escherichia coli* functions in
529 acute stages of pathogenesis. Infect Immun **78:**1457-1467.

JCM Accepts published online ahead of print

JCM Accepts published online ahead of print

530 7. **Goullet, P., and B. Picard.** 1986. Highly pathogenic strains of *Escherichia*
531 *coli* revealed by the distinct electrophoretic patterns of carboxylesterase B. J
532 Gen Microbiol **132:**1853-1858.
533 8. **Gruneberg, R. N.** 1969. Relationship of infecting urinary organism to the
534 faecal flora in patients with symptomatic urinary infection. Lancet **2:**766-768.
535 9. **Hagan, E. C., A. L. Lloyd, D. A. Rasko, G. J. Faerber, and H. L. Mobley.**
536 2010. *Escherichia coli* global gene expression in urine from women with
537 urinary tract infection. PLoS Pathog **6:**e1001187.
538 10. **Ingersoll, M. A., J. E. Moss, Y. Weinrauch, P. E. Fisher, E. A. Groisman,**
539 **and A. Zychlinsky.** 2003. The ShiA protein encoded by the *Shigella flexneri*
540 SHI-2 pathogenicity island attenuates inflammation. Cell Microbiol
541 **5:**797-807.
542 11. **Ingersoll, M. A., and A. Zychlinsky.** 2006. ShiA abrogates the innate T-cell
543 response to *Shigella flexneri* infection. Infect Immun **74:**2317-2327.
544 12. **Johnson, J. R., P. Delavari, M. Kuskowski, and A. L. Stell.** 2001.
545 Phylogenetic distribution of extraintestinal virulence-associated traits in
546 *Escherichia coli*. J Infect Dis **183:**78-88.
547 13. **Johnson, J. R., N. Kaster, M. A. Kuskowski, and G. V. Ling.** 2003.
548 Identification of urovirulence traits in *Escherichia coli* by comparison of
549 urinary and rectal *E. coli* isolates from dogs with urinary tract infection. J Clin
550 Microbiol **41:**337-345.
551 14. **Johnson, J. R., T. T. O'Bryan, M. Kuskowski, and J. N. Maslow.** 2001.
552 Ongoing horizontal and vertical transmission of virulence genes and *papA*
553 alleles among *Escherichia coli* blood isolates from patients with
554 diverse-source bacteremia. Infect Immun **69:**5363-5374.

JCM Accepts published online ahead of print

555 15. **Johnson, J. R., and T. A. Russo.** 2005. Molecular epidemiology of
556 extraintestinal pathogenic (uropathogenic) *Escherichia coli*. Int J Med
557 Microbiol **295:**383-404.
558 16. **Johnson, J. R., and A. L. Stell.** 2000. Extended virulence genotypes of
559 *Escherichia coli* strains from patients with urosepsis in relation to phylogeny
560 and host compromise. J Infect Dis **181:**261-272.
561 17. **Kingsley, R. A., A. D. Humphries, E. H. Weening, M. R. De Zoete, S.**
562 **Winter, A. Papaconstantinopoulou, G. Dougan, and A. J. Baumler.** 2003.
563 Molecular and phenotypic analysis of the CS54 island of *Salmonella enterica*
564 serotype typhimurium: identification of intestinal colonization and persistence
565 determinants. Infect Immun **71:**629-640.
566 18. **Klein, T., D. Abgottspon, M. Wittwer, S. Rabbani, J. Herold, X. Jiang, S.**
567 **Kleeb, C. Luthi, M. Scharenberg, J. Bezencon, E. Gubler, L. Pang, M.**
568 **Smiesko, B. Cutting, O. Schwardt, and B. Ernst.** 2010. FimH antagonists
569 for the oral treatment of urinary tract infections: from design and synthesis to
570 in vitro and in vivo evaluation. J Med Chem **53:**8627-8641.
571 19. **Langermann, S., S. Palaszynski, M. Barnhart, G. Auguste, J. S. Pinkner, J.**
572 **Burlein, P. Barren, S. Koenig, S. Leath, C. H. Jones, and S. J. Hultgren.**
573 1997. Prevention of mucosal *Escherichia coli* infection by
574 FimH-adhesin-based systemic vaccination. Science **276:**607-611.
575 20. **Lloyd, A. L., T. A. Henderson, P. D. Vigil, and H. L. Mobley.** 2009.
576 Genomic islands of uropathogenic *Escherichia coli* contribute to virulence. J
577 Bacteriol **191:**3469-3481.
578 21. **Lloyd, A. L., D. A. Rasko, and H. L. Mobley.** 2007. Defining genomic
579 islands and uropathogen-specific genes in uropathogenic *Escherichia coli*. J

JCM Accepts published online ahead of print

580 Bacteriol **189:**3532-3546.

581 22. **Lloyd, A. L., S. N. Smith, K. A. Eaton, and H. L. Mobley.** 2009.
582 Uropathogenic *Escherichia coli* Suppresses the host inflammatory response
583 via pathogenicity island genes *sisA* and *sisB*. Infect Immun **77:**5322-5333.

584 23. **Marrs, C. F., L. Zhang, and B. Foxman.** 2005. *Escherichia coli* mediated
585 urinary tract infections: are there distinct uropathogenic *E. coli* (UPEC)
586 pathotypes? FEMS Microbiol Lett **252:**183-190.

587 24. **McClelland, M., K. E. Sanderson, J. Spieth, S. W. Clifton, P. Latreille, L.**
588 **Courtney, S. Porwollik, J. Ali, M. Dante, F. Du, S. Hou, D. Layman, S.**
589 **Leonard, C. Nguyen, K. Scott, A. Holmes, N. Grewal, E. Mulvaney, E.**
590 **Ryan, H. Sun, L. Florea, W. Miller, T. Stoneking, M. Nhan, R. Waterston,**
591 **and R. K. Wilson.** 2001. Complete genome sequence of *Salmonella enterica*
592 serovar Typhimurium LT2. Nature **413:**852-856.

593 25. **Moss, J. E., T. J. Cardozo, A. Zychlinsky, and E. A. Groisman.** 1999. The
594 *selC*-associated SHI-2 pathogenicity island of *Shigella flexneri*. Mol Microbiol
595 **33:**74-83.

596 26. **Nataro, J. P., J. Seriwatana, A. Fasano, D. R. Maneval, L. D. Guers, F.**
597 **Noriega, F. Dubovsky, M. M. Levine, and J. G. Morris, Jr.** 1995.
598 Identification and cloning of a novel plasmid-encoded enterotoxin of
599 enteroinvasive *Escherichia coli* and *Shigella* strains. Infect Immun
600 **63:**4721-4728.

601 27. **Palaniappan, R. U., Y. Zhang, D. Chiu, A. Torres, C. Debroy, T. S.**
602 **Whittam, and Y. F. Chang.** 2006. Differentiation of *Escherichia coli*
603 pathotypes by oligonucleotide spotted array. J Clin Microbiol **44:**1495-1501.

604 28. **Parham, N. J., S. J. Pollard, R. R. Chaudhuri, S. A. Beatson, M. Desvaux,**

JCM Accepts published online ahead of print

605 **M. A. Russell, J. Ruiz, A. Fivian, J. Vila, and I. R. Henderson.** 2005.
606 Prevalence of pathogenicity island IICFT073 genes among extraintestinal
607 clinical isolates of *Escherichia coli*. J Clin Microbiol **43:**2425-2434.
608 29. **Perna, N. T., G. Plunkett, 3rd, V. Burland, B. Mau, J. D. Glasner, D. J.**
609 **Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J.**
610 **Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W.**
611 **Davis, A. Lim, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S.**
612 **Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R.**
613 **Blattner.** 2001. Genome sequence of enterohaemorrhagic *Escherichia coli*
614 O157:H7. Nature **409:**529-533.
615 30. **Picard, B., J. S. Garcia, S. Gouriou, P. Duriez, N. Brahimi, E. Bingen, J.**
616 **Elion, and E. Denamur.** 1999. The link between phylogeny and virulence in
617 *Escherichia coli* extraintestinal infection. Infect Immun **67:**546-553.
618 31. **Seed, P. C., and S. J. Hultgren.** 2005. Blueprinting the regulatory response of
619 *Escherichia coli* to the urinary tract. Trends Microbiol **13:**246-248.
620 32. **Smajs, D., and G. M. Weinstock.** 2001. The iron- and temperature-regulated
621 *cjrBC* genes of *Shigella* and enteroinvasive *Escherichia coli* strains code for
622 colicin Js uptake. J Bacteriol **183:**3958-3966.
623 33. **Soto, S. M., E. Guiral, J. Bosch, and J. Vila.** 2009. Prevalence of the *set-1B*
624 and *astA* genes encoding enterotoxins in uropathogenic *Escherichia coli*
625 clinical isolates. Microb Pathog **47:**305-307.
626 34. **Touchon, M., C. Hoede, O. Tenaillon, V. Barbe, S. Baeriswyl, P. Bidet, E.**
627 **Bingen, S. Bonacorsi, C. Bouchier, O. Bouvet, A. Calteau, H. Chiapello, O.**
628 **Clermont, S. Cruveiller, A. Danchin, M. Diard, C. Dossat, M. E. Karoui, E.**
629 **Frapy, L. Garry, J. M. Ghigo, A. M. Gilles, J. Johnson, C. Le Bouguenec,**

JCM Accepts published online ahead of print

630 **M. Lescat, S. Mangenot, V. Martinez-Jehanne, I. Matic, X. Nassif, S.**
631 **Oztas, M. A. Petit, C. Pichon, Z. Rouy, C. S. Ruf, D. Schneider, J. Tourret,**
632 **B. Vacherie, D. Vallenet, C. Medigue, E. P. Rocha, and E. Denamur.** 2009.
633 Organised genome dynamics in the *Escherichia coli* species results in highly
634 diverse adaptive paths. PLoS Genet **5:**e1000344.

635 35. **Tseng, C. C., J. J. Huang, W. C. Ko, J. J. Yan, and J. J. Wu.** 2001.
636 Decreased predominance of *papG* class II allele in *Escherichia coli* strains
637 isolated from adults with acute pyelonephritis and urinary tract abnormalities.
638 J Urol **166:**1643-1646.

639 36. **Vokes, S. A., S. A. Reeves, A. G. Torres, and S. M. Payne.** 1999. The
640 aerobactin iron transport system genes in *Shigella flexneri* are present within a
641 pathogenicity island. Mol Microbiol **33:**63-73.

642 37. **Wang, M. C., C. C. Tseng, A. B. Wu, J. J. Huang, B. S. Sheu, and J. J. Wu.**
643 2009. Different roles of host and bacterial factors in *Escherichia coli*
644 extra-intestinal infections. Clin Microbiol Infect **15:**372-379.

645 38. **Welch, R. A., V. Burland, G. Plunkett, 3rd, P. Redford, P. Roesch, D.**
646 **Rasko, E. L. Buckles, S. R. Liou, A. Boutin, J. Hackett, D. Stroud, G. F.**
647 **Mayhew, D. J. Rose, S. Zhou, D. C. Schwartz, N. T. Perna, H. L. Mobley,**
648 **M. S. Donnenberg, and F. R. Blattner.** 2002. Extensive mosaic structure
649 revealed by the complete genome sequence of uropathogenic *Escherichia coli*.
650 Proc Natl Acad Sci U S A **99:**17020-17024.

651 39. **Xie, Y., V. Kolisnychenko, M. Paul-Satyaseela, S. Elliott, G. Parthasarathy,**
652 **Y. Yao, G. Plunkett, 3rd, F. R. Blattner, and K. S. Kim.** 2006. Identification
653 and characterization of *Escherichia coli* RS218-derived islands in the
654 pathogenesis of *E. coli* meningitis. J Infect Dis **194:**358-364.

655

656

JCM Accepts published online ahead of print