Int J Mol Epidemiol Genet 2011;2(4):339-353
www.ijmeg.org /ISSN1948-1756/IJMEG1108002

Original Article

Epidemiologic Investigation of Extra-intestinal pathogenic E. coli (ExPEC) based on PCR phylogenetic group and *fim*H single nucleotide polymorphisms (SNPs) in China

Kamal Said Abdallah[1], Yang Cao[1], Dian-Jun Wei[1,2]

[1]Department of Clinical Laboratory, Second Hospital of Tianjin Medical University, 300211, Tianjin, [2]Tianjin Institute of Infection Diseases, Second Hospital of Tianjin Medical University, 300211, Tianjin, China.

Received August 23, 2011; accepted November 6, 2011; Epub November 25, 2011; Published December 15, 2011

Abstract: Using the combination method with PCR phylogrouping and *fim*H SNPs analysis, this study investigates the epidemiology of Extra-intestinal pathogenic E. coli in China. 116 E. coli strains including (74 from Urine, 39 from other extra-intestinal sources and 3 references strains) were collected. The bacteria Genomic DNA were extracted; phylogroup and the *fim*H gene amplifications were determined by two-step triplex PCR-based phylogrouping and simple PCR amplification assay respectively. Finally the *fim*H SNPs analysis and phylogenetic analysis and construction of tree were carried out using DNAMAN Version 6.0.3.93 and MEGA4, ClustalW and CLC Bio software respectively for 50 E. coli strains isolated from clinical sample and 3 references; K-12 E. coli strain was used as reference comparison. For E. coli strains phylogroup, 25% (28/113) were observed to belong to the group A, 15% (17/113) to the group B1, 14% (16/113) to the group B2, and 46% (52/113) to the group D. 75% (85/113) were *fim*H positive. *fim*H SNPs analysis for 50 isolated from clinical sample and 3 references found 60 SNPs at 57 polymorphic sites. The number of amino-acid variants and silent SNPs were observed more in UPEC strains than in other extra-intestinal E. coli strains. Most of the UPEC strains with the same amino-acid variants were belong to the same phylogroup. This combination method could serve as a rapid, highly reproducible typing test for epidemiological studies of ExPEC. Large collection data could be compared with other clinical laboratories that the sequence data are accessible.

Keywords: Polymerase chain reaction (PCR), single nucleotide polymorphisms (SNPs), Extra-intestinal E.coli (ExPEC), epidemiology

Introduction

The E. coli pathotypes that cause extra-intestinal infections are collectively called Extraintestinal pathogenic E. coli (ExPEC). They are phylogenetically and epidemiologically distinct from commensal and diarrheagenic stains. Compared to human commensal strains, which usually derive from phylogenetic groups A and B1, most of ExPEC strains belong to the B2 and D groups and harbor various virulence factors which allow them to induce diseases in both healthy and compromised hosts [1, 2]. Based on the prevalence of various virulence factors in clinical E. coli isolates from extraintestinal infections, ExPEC were currently defined as E. coli isolates containing two or more of the following genes: papA (P fimbriae structural subunit) and/or papC (P fimbriae assembly), sfa/foc (S and F1C fimbriae subunits), afa/dra (Dr-antigen-binding adhesins), kpsMT II (group 2 capsular polysaccharide units), and iutA (aerobactin receptor) [3]. Unlike diarrheagenic strains, ExPEC are incapable of causing gastrointestinal disease in human. However, ExPEC can cause diverse infections at different anatomical locations outside the human intestinal tract [4, 5]. E. coli are the major cause of extraintestinal infections such as Urinary Tract Infections (UTIs), neonatal meningitis, gram-negative bacteremia, and prostatitis [6]. E. coli also occur in intra-abdominal infections and nosocomial pneumonia, and occasionally are involved in other extraintestinal infections such as osteo-

myelitis, cellulitis, and wound infections. Uropathogenic differ from non-pathogenic E. coli and from other E. coli pathotype by producing specific virulence factors, which enable the bacteria to adhere to uroepithelial cells and to establish urinary tract infections [7]. The abilities of Uropathogenic E. coli (UPEC) to grow extra-intestinally may enable them to cause a variety of diseases not just Urinary Tract Infections. E. coli strains can be assigned to one of the main phylogenetic groups: A, B1, B2 or D [8]. In bacterial pathogens, a variety of SNPs has been discovered that confer a selective advantage during the course of a single infection, epidemic spread or long-term evolution of virulence [9, 10]. *fim*H is a specific adhesion located at the tip of type 1 fimbriae that determines mannose-sensitive binding of bacteria to target cells [9]. The combination of PCR phylogrouping and *fim*H SNP analysis may be a useful method to type a large collection of clinical E. coli isolates for epidemiologic studies [11]. In this study, we use this combination method to describe the epidemiological evolution of Extra-intestinal pathogenic E. coli i in China.

Materials and methods

Total 116 E. coli strains were used in this study. 113 Extraintestinal pathogenic E. coli strains were isolated from clinical patients with one of the following diseases: Urethritis, Cholecystitis, Pyelonephritis, bacteremia, neonatal meningitis prostatitis and other extra-intestinal infection sources at the Tianjin second Hospital of Tianjin Medical University (49 strains of E. coli which were isolated in Jun 2010 conserved in a centrifuge tube containing Luria Broth (LB) medium plus15% glycerol and stored at -70°C were collected, 64 were isolated between the end of 2010 and the middle of 2011). 3 selected data of reference strains were downloaded from [GenBank] (CFT037 sequence NC_004431, UTI89 sequence NC_007946, and K-12 sequence NC000913). Among the 113 E. coli strains, 74 were isolated from urine sample of patients (urinary tract infection), 39 were isolated from other sources of extra-intestinal of patients. All the 113 E. coli strains are from women and men patients that are hospital visiting and or hospitalized. The ages were not determined. Only one isolated per patient was included. The common laboratory strain K-12 was studied for comparison. Strains were stored at -70°C in Luria Broth plus 15% glycerol until ready for use.

E. coli strains identification

E. coli strains identification were confirmed according to the Clinical Laboratory Procedures. Patient specimens were inoculated into Blood and MacConkey agar plates, and were incubated at 37°C for 24h in an incubator. Suspected E. coli were selected from pure culture for biochemical with API and serological identification with standard antiserum typing kits.

DNA isolation

DNA template preparation was performed by the boiling method as follows: the E. coli clinical isolates were inoculated into 5 ml tube containing LB medium (tryptone 10g/l, Yest 5g/l, NaCl 10g/l) and incubated for 18 hours at 37°C with shaking at 160r/min. Cells from 1.5 ml of an overnight culture were harvested by centrifugation at 6000r/min for 10min. After the supernatant was decanted, the pellet was resuspended in 200 μl of distilled water. The cells were lysed by heating them at 95°C for 10 min. After heating they were put immediately in the ice for 5 min. Cellular debris was removed by centrifugation at 12000r/min for 15 min. Added equal volume of chloroform to wash it, mixed by vortexing. After centrifuging at 12000r/min for 2 min at 4°C, the supernatant was transferred to a fresh centrifuge tubes and kept at -20°C. The supernatant was used as a source of template for amplification.

Two-step triplex PCR-based phylogrouping

Phylogenetic grouping of the 113 E. coli isolates was assessed by a previously reported triplex-PCR-based assay [12, 13]. All amplification procedures were repeated at least three times. The primer pairs used for PCR amplification were: ChuA.1 (5'-GACGAACCAACGGTCAGGAT-3') and ChuA.2 (5'-TGCCGCCAGTACCAAAGACA-3'), YjaA.1 (5'-TGAAG TGTCAGG AGACGCTG-3') and YjaA.2 (5'-ATGGA GAATGCG TTCCT CAAC-3'), and TspE4C2.1 (5'-GAGTA ATGTCG GGGC ATTCA-3') and TspE4 C2.2 (5'-CG CGCCAACAAAGTATTACG-3'), which generate 279-, 211-, and 152-bp fragments, respectively. Briefly, the amplifications were carried out in a total volume of 25 μl, each reaction mixture contained 12.5 μl premix, 1 μl each forward primer, and each reverse primer, 3.5 μl distilled H_2O, and 3 μl of DNA template. The reaction conditions were 4 min of initial denaturation at 94°C followed by 30 5-sec cycles of denaturation at 94°C; 10 sec of anneal-

Table 1. Comparison of phylogenic group of ExPEC strains isolated from Urine and from other sources

ExPEC samples	Phylogroups determination				
	group A	group B1	group B2	group D	Total
UPEC	13 (17.5%)	13 (17.5%)	9 (12%)	39 (53%)	74 (100.0%)
Other sources	15 (38.5%)	4 (10.3%)	7 (17.9%)	13 (33.3%)	39 (100.0%)
Total	28 (25%)	17 (15%)	16 (14%)	52 (46%)	113 (100.0%)

Abbreviations: Extra-intestinal pathogenic E. coli (ExPEC), Uropathogenic E. coli (UPEC); Note: group B1 Vs group A: X^2=3.91, P<0.05. Group D Vs group A: X^2=6.53, P<0.05. Other comparisons P all > 0.05.

Table 2. Extra-intestinal *fim*H gene determination

ExPEC samples	*fim*H positive (n =85)	*fim*H negative (n =28)	Total
UPEC	59 (80%)	15 (20%)	74 (100%)
Other sources	26 (66.67%)	13 (33.33%)	39 (100%)
Total	85 (75%)	28 (25%)	113 (100%)

Abbreviations: Extra-intestinal pathogenic E. coli (ExPEC), Uropathogenic E. coli (UPEC)

ing at 59°C and a final extension step of 5 min at 72°C. The results allowed the classification of isolates into one of the four major phylogroups (A, B1, B2, or D).

PCR amplification of fimH gene

The 113 E. coli strains *fim*H PCR were performed as reported previously [14]. The primers used for PCR amplification and partial *fim*H gene sequencing were FimH-f (5'-CGAGTTATTACCCTGTTTGCTG-3') and FimH-r (5'-ACGCCAATAATCGATTGCAC-3'). Briefly, the amplifications were carried out in a total volume of 50 μl, each reaction mixture contained 25 μl of premix, 1 μl each forward primer, and each reverse primer, 20 μl distilled ionized sterilized H_2O, and 3 μl of DNA template. The reaction conditions were 2 min of initial denaturation at 95°C followed by 33 cycles, 30 sec of denaturation at 94°C; 30 sec of annealing at 58°C, 1min of extension at 72°C and a final extension step of 7 min at 72°C. Both strands of the 878-bp PCR-amplified fragment (located at bp 7 to 884 of E. coli sequence NC000913 [GenBank]) were sequenced. After visual inspection and editing with the Beijing Invitrogen company and DNAMAN version 6.0.3.93), fragments of 674-bp *fim*H sequences (located at bp 119 to 792 of E. coli K-12 sequence NC000913 [GenBank]) were compared to the sequence of E. coli K-12. All amplification procedures were repeated at least three times.

fimH gene sequencing statistical analysis of SNPs

The PCR results of 50 selected E. coli strains among the 85/113 strains were sent to a sequencing company (Invitrogen) in Beijing for sequencing, the Forward- and reverse-strand DNA sequence traces were reported in Pubmed for BLAST. These two Forward- and reverse-strand DNA sequence traces were spliced by using Sequence Scanner version 1.0. The finished spliced sequences were aligned for the flowing process: Sequence, Alignment and then Multiple Sequence Alignment in DNAMAN Version 6.0.3.93 for SNPs analysis. Fragments of 674-bp *fim*H sequences (located at bp 119 to 792 of E. coli K-12 of the 50 selected E.coli strains plus CFT037 and UTI89 were compared to the whole *fim*H sequence (903-bp) of E. coli K-12 in DNAMAN. The 674-bp *fim*H gene fragments alignment of the 50 clinical E. coli isolates plus CFT037, UTI89 and E. coli K-12 reference strains were imported into MEGA4 [15] for phylogenetic analysis and construction of trees. The significance of the branching order was evaluated by bootstrap analysis with 500 replicates. The evolutionary distances were computed by the Kimura 2-parameter method [16].

Result

Phylogroup distribution

The totalities of 113 E.coli strains were analyzed. The phylogenetic group of each strains was determined by reference methods [17, 18], were as follows: ChuA (+), YjaA (+), TSPE4.C2 (+/-) was assigned to group B2; ChuA (+), YjaA (-), TSPE4.C2 (+/-) to group D; ChuA (-), YjaA (+/-), TSPE4.C2 (+) to group B1; and chuA (-), YjaA (+/-), TSPE4.C2 (-) to group A. Among the 113 E. coli strains analyzed (**Table 1**), twenty-eight (25%) were observed to belong to the group A, seventeen (15%) were observed to belong to the group B1, sixteen (14%) were observed to belong to the group B2, and fifty-two (46%) were observed to belong to the group D.

The comparison of UPEC strains and other ExPEC strains in different phylogroups (**Table 1**) show that group A is statistical significant than B1 in the other ExPEC strains ($P<0.05$), group D is statistical significant than group A in UPEC strains ($P<0.05$). No significant found between other comparisons.

fimH SNPs result

We first examined the *fim*H gene of all E. coli isolated strains by PCR to get more information for this gene (**Table 2**). Twenty-eight (25%) of the 113 E. coli isolated in our study were *fim*H negative (20% [15/74] UPEC, and (33.33% [13/39] other sources of extra-intestinal E.coli.

For the *fim*H SNPs analyzes, 53 E. coli strains were used. We selected only 50 E. coli isolated strains among the 85/113 with *fim*H positive for sequencing (**Table 3** and **Figure 1**) 27 UPEC strains and 23 other sources of extra-intestinal E.coli with the addition of the 3 reference sequence strains (CFT037, UTI89 and E. coli K-12 reference strains). The strain annotations and they accession numbers in GenBank are shown in (**Table 6**). All the 674-bp *fim*H DNA sequences were compared to that of E. coli-12 (**Figures 2, 3, 4, 5,** and **6**). Total 60 SNPs were observed at 57 polymorphic sites (**Table 3** and **Figure 1**). 10% (6/60) E. coli strains (B626, C1230, A564, B39, C1357, and A2861) were observed with no SNPs. Except B3595 strain mutations C602T, A603G, and C1354 strain mutations GC545, 546CT, all mutations were point substitutions (SNPs), 28% (17/60) were transversions,72 % (43/60) were transitions. 23 SNPs resulted to 22 amino-acid changes and 37 SNPs were silent substitutions. Only 6 among the amino-acid replacement were transversions. 15 SNPs were singletons (observed in only one *fim*H type, C1352 strain itself result 8 singletons), among them 12 were with amino-acid replacements. No amino-acids changes were observed to The C1839 strains (group A) SNPs. 63% (38/60) of E. coli strains show amino-acid changes independently on the same position (V48A).

The distribution of *fim*H variants was analyzed by the building of an unrooted phylogram (**Figure 1**). The 53 strains *fim*H SNPs analyzed were classified into 25 terminal groups. 14 terminal groups were distinct allelic variants (B3595 [B1], A528 [A], B599 [B1], B678 [B1], C1119 [B2], C1354 [B2], C1839 [A], B3626 [D], C2689 [D], B3623 [D], C 2586 [B2], UTI89 [B2], B459 [D], and A2794 [B2]). Most of them belong to the group D, B2, and B1. The remaining 11 terminal groups were classified and almost belong independently on the same phylogroup respectively, small quantity of this terminal group strains were not respect this role: B42 and B41 belong to the group B2, C1230 and K-12 belong to the group A, B460 belong to the group B2, A774 belong to the group A, A208 belong to the group B2, while most of the strains of these terminal groups belong to the group D, D, D, B1, and D respectively. Amino-acid variants and silent SNPs were compared between UPEC and other extra-intestinal source strains (**Table 4** and **Table 5** respectively). Except some singletons (P106S, V175A, C182S, R187C, L193P, V206L, T95I, and A127V), the numbers of strains with the same amino-acid variants (**Table 4**) were observed more in UPEC than the other extra-intestinal strains. 23/37 silent SNPs (**Table 5**) were observed in UPEC strains. In additional most of the UPEC strains with the same amino-acid variants (**Figure 1**) were belong to the same phylogroup.

Discussion

E. coli accounts for approximately 70 to 95% of community-acquired cases and 50% of all hospital acquired infections [19]. These organisms are responsible for significant social and economic costs for both communities and public health resources. The population biology of Extraintestinal E. coli is poorly understood. The

Table 3. Single-nucleotide polymorphism(s) position with amino-acid replacements and the strains phylogroup

Strains	Phylogroup	*fim*H SNP(s)[a]
K-12	A	None (Reference sequence)
B626	D	None
C1230	A	None
A564	D	None
B39	D	None
C1357	D	None
A2861	D	None
C636	D	C367T (P123S)
B3736	D	C367T (P123S)
B42	B2	C374T (P125L)
C2669	D	C374T (P125L)
B41	B2	C374T (P125L)
B639	D	C374T (P125L)
B457	D	C374T (P125L)
B513	A	T143C (V48A), A224G (D75G)
B493	A	T143C (V48A), A224G (D75G)
B514	A	T143C (V48A), A224G (D75G)
B547	A	T143C (V48A), A224G (D75G)
A528	A	C141T, T143C (V48A), G171A, T318G,
C1839	A	C141T, G171A, T318G, C339G, T714A, A717G
C1119	B2	T143C (V48A), G171A, T534C, C546T, T714A, A717G
C601	B1	C141T, T143C (V48A), G171A, T318G, C339G, T714A, A717G
A774	A	C141T, T143C (V48A), G171A, T318G, C339G, T714A, A717G
C592	B1	C141T, T143C (V48A), G171A, T318G, C339G, T714A, A717G
C2668	B1	C141T, T143C (V48A), G171A, T318G, C339G, T714A, A717G
A745	B1	C141T, T143C (V48A), G171A, T318G, C339G, T714A, A717G
B3744	B1	C141T, T143C (V48A), G171A, T318G, C339G, T714A, A717G
B3595	B1	C141T, T143C (V48A), G171A, T318G, G420A, C546T, C602T and A603G (P201L)
B599	B1	T143C (V48A), G171A, T318G, C546T, G645A, G702C, T714A, A717G
B583	D	C141T, T143C (V48A), G171A, G259T (G87C), T318G, G420A, C546T, A603G
C561	D	C141T, T143C (V48A), G171A, G259T (G87C), T318G, G420A, C546T, A603G
B678	B1	C141T, T143C (V48A), G171A, T318G, C546T, G645A, G702C, T714A, A717G
B595	A	T143C (V48A), T207C, C225T, T318G, C327T, C339G, C419T (A140V), T534C, A603G
C93	A	T143C (V48A), T207C, C225T, T318G, C327T, C339G, C419T (A140V), T534C, A603G
B586	A	T143C (V48A), T207C, C225T, T318G, C327T, C339G, C419T (A140V), T534C, A603G
B3626	D	T143C (V48A), T207C, C225T, T318G, C327T, C339G, C419T (A140V), T534C, A603G, C742A (R248S)
B575	D	C141T, T143C (V48A), G171A, T318G, C321T, C339G, C411T, G560 (R187H), T714A, A717G

B3672	D	C141T, T143C (V48A), G171A, T318G, C321T, C339G, C411T, G560 (R187H), T714A, A717G
B460	B2	C141T, T143C (V48A), G171A, T318G, C321T, C339G, C411T, G560 (R187H), T714A, A717G
C2689	D	T143C (V48A), G171A, T207C, C246T, T318G, C327T, G333A, C588T, T591C, A597G, A603G, A647T(Y216F), T714A, A717G
B566	D	T143C (V48A), G171A, T207C, C339G, G408T, T450A, C546T, C588T, T591C, A597G, A603G, A647T(Y216F), T714A, A717G
B3674	D	T143C (V48A), G171A, T207C, C339G, G408T, T450A, C546T, C588T, T591C, A597G, A603G, A647T(Y216F), T714A, A717G
C1354	B2	T143C (V48A), G171A, C316T and T318C (P106S), T524C (V175A), T534C, C543A, G545C and C546T(C182S), C559T (R187C), T578C (L193P), G616C (V206L), T714A, A717G
A208	B2	C141T, T143C (V48A), G171A, C300T, C327T, G408T, C432T, C489T, T534C, A603G, G702C, T714A, A717G
A675	D	C141T, T143C (V48A), G171A, C300T, C327T, G408T, C432T, C489T, T534C, A603G, G702C, T714A, A717G
A2893	D	C141T, T143C (V48A), G171A, C300T, C327T, G408T, C432T, C489T, T534C, A603G, G702C, T714A, A717G
B3623	D	C141T, T143C (V48A), T207C, C246T, G259T(G87C), A272G (N91S), G296A (S99N), A312T, T315C, C321T, C339G, C411G, G414A, T429A, T534C, A603G, T714A, A717G
C2586	B2	C141T, T143C (V48A), T207C, C246T, A272G (N91S), C284T(T95I), G296A (S99N), A312T, T315C, C321T, C339G, C411G, G414A, T429A, T534C, T551C (V184A), A603G, T714A, A717G
B459	D	C141T, T143C (V48A), T207C, C246T, A272G (N91S), G296A (S99N), A312T, T315C, C321T, C339G, T396G, C411G, G414A, T534C, C546T, A603G, C654A, T714A, A717G
UTI89	B2	C141T, T143C (V48A), C171A, T207C, G210A, C246T, T247G (S83A), A272G (N91S), G296A (S99N), A312T, T315C, C321T, C339G, T396G, C411G, G414A, T534C, A603G, T714A, A717G
A2794	B2	C141T, T143C (V48A), C171A, T207C, G210A, C246T, A272G (N91S), G296A (S99N), A312T, T315C, C321T, C339G, C380T (A127V), T396G, C411G, G414A, T534C, C546T, A603G, C654A, T714A, A717G
B624	B2	C141T, T143C (V48A), C171A, T207C, G210A, C246T, A272G (N91S), G296A (S99N), A312T, T315C, C321T, C339G, T396G, C411G, G414A, T534C, C546T, T551C (V184A), A603G, C654A, T714A, A717G
CFT073	B2	C141T, T143C (V48A), C171A, T207C, G210A, C246T, A272G (N91S), G296A (S99N), A312T, T315C, C321T, C339G, T396G, C411G, G414A, T534C, C546T, T551C (V184A), A603G, C654A, T714A, A717G

Abbreviation: Single nucleotide polymorphisms (SNPs). DNA nucleotides abbreviations: Thymine (T), Adenine (A), Cytosine (C), Guanine (G). Amino-acid abbreviations: Alanine (A), Arginine (R), Aspargine (N), Aspartic acid (D), Cysteine (C), Glutamic acid (E), Glutamine (Q), Glycine (G), Histidine (H), Isoleucine (I), Leucine (L), Lysine (K), Methionine (M), Phenylalanine (F), Pronine (P), Serine (S), Theonine (T), Tryptophan (W), Tyrosine (Y), Valine (V). [a]*fim*H sequences were analysed by comparison with the sequence of E. coli K-12 (sequence NC_000913: GenBank). All the SNPs of our sequenced genes that located at bp 119 to 792 (674 bp of length) of E. coli K-12 are shown. The underlined *fim*H SNPs are nucleotide replacement found in only one study isolate (singleton). The left nucleotides of the underlined and non-underlined are the nucleotides of E.coli K-12, and the right nucleotides are the nucleotides change in this study. The number is the SNPs position. Amino acid replacements are shown in parentheses.

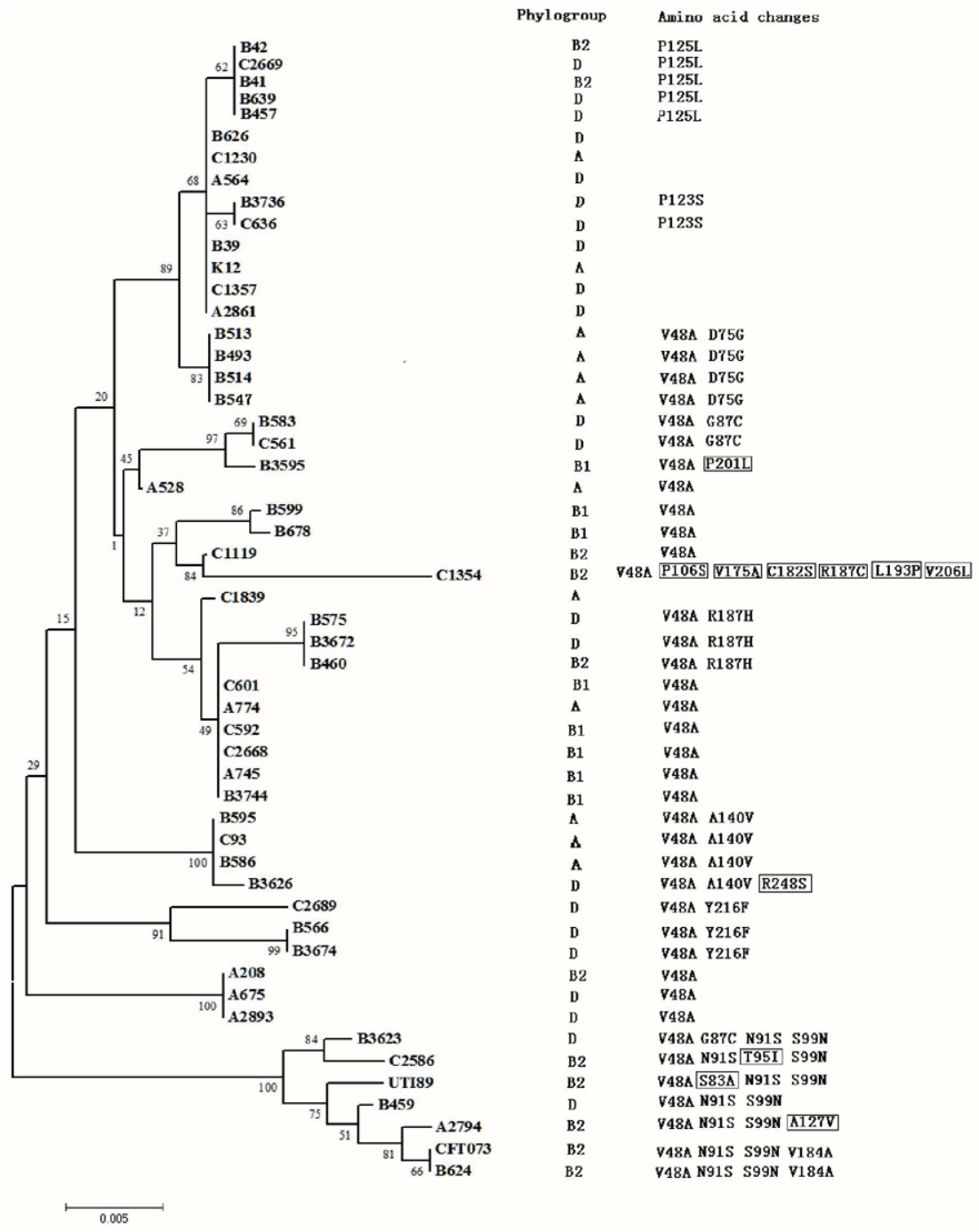

Figure 1. Phylogenetic tree derived from 50 *Fim*H sequence variants selected from 85 *fim*H positive clinical E. coli isolates and E. coli reference K-12, CFT073, and UTI89 were determined by MEGA 4, using the Kimura 2-parameter model [15]. *fim*H sequences were analyzed by comparison with the sequence of E. coli K-12. Boxes contain amino-acid replacements found in only one strain study (singleton) that resulted in 11 *fim*H protein variants deduced from the analysis of nucleotide sequences. Non boxes amino-acid are not singleton, the change of the amino-acid in the same position independently can be observed in different strains. The left capital letter (A, B and C) of the E. coli strains determined the source of the strains: B: strains are isolated from urine of the patients (UPEC), the references CFT073 and UTI89 are also urine isolated. A and C strains are isolated from other extra-intestinal sources. The phylogroup capital letters (A, B1, B2, and D) are determined by PCR [12, 13]. Numbers within the tree indicate the occurrence (%) of the branching order 500 bootstrapped trees. The value shows are between 1 and 100.

Figure 2. Sequences from 119 to 269 bp and 10 polymorphic sites of SNPs analysis were shown. They pick images were boxed above.

Table 4. Comparison of *fim*H SNPs amino-acid variants observed between UPEC and other extra-intestinal sources

Amino-acid variants	Number of UPEC strains	Number of Other ExPEC strains	Total number of each amino-acid variants
P125L	4	1	5
P123S	1	1	2
D75G	4	0	4
G87C	2	1	3
P201L	1	0	1
V48A	22	16	38
P106S, V175A, C182S, R187C, L193P, V206L	0	1	1
R187H	3	0	3
A140V	3	1	4
R248S	1	0	1
Y216F	2	1	3
T95I	0	1	
S93A	1	0	1
A127V	0	1	1
V184A	2	0	2
N91S, S99N	5	2	7

Abbreviations: Extra-intestinal pathogenic E. coli (ExPEC), Uropathogenic E. coli (UPEC); Total numbers of strains having specific amino-acid changes are shown on the right. Underlined amino-acid change were singleton (just observed in one study isolated).

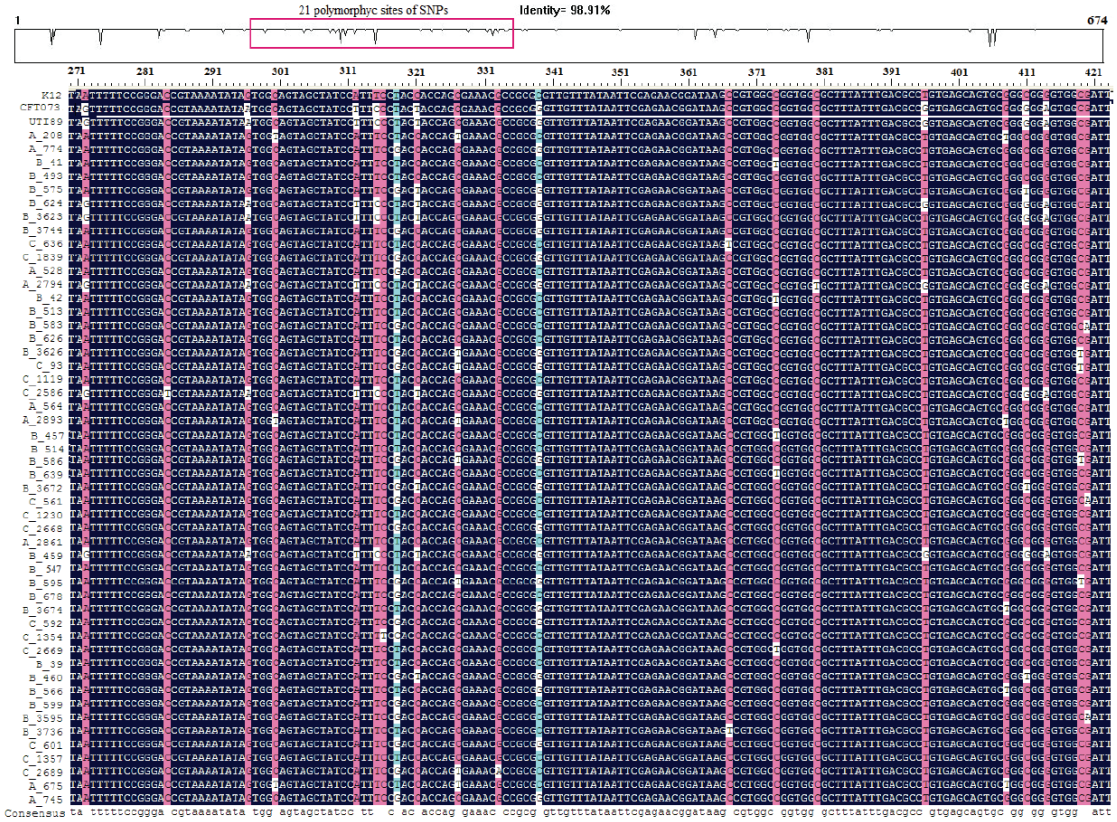

Figure 3. Sequences from 270 to 423 bp and 21 polymorphic sites of SNPs analysis were shown. They pick images were boxed above.

abilities of Uropathogenic E. coli (UPEC) to grow extraintestinally may enable them to cause a variety of diseases not just Urinary Tract Infections. It is now recognized that there are a subset of fecal E. coli having these virulence factors which can colonize periurethral area, enter urinary tract and cause symptomatic disease. These are currently defined as UPEC [20]. Phenotypic variants of FimH are predominantly the product of SNPs in *fim*H [18]. SNPs that contribute to the ability of pathogens to cause disease confer a selective advantage during the course of a single infection, epidemic spread or long-term evolution of virulence [9, 10, 21]. In a previous study, Tartof et al. explored *fim*H single-nucleotide polymorphism (SNP) analysis as a screening test for the epidemiological study of Uropathogenic E. coli (UPEC) [22]. Insignificant mutation in C-terminal leads to the formation of truncated *fim*H, which can affect the attachment of non-pathogenic feces strains to urotheilial cell; however, attachment factor has no role in Uropathogenic E. coli pathogenic by itself [23]. *fim*H variants from E. coli strains of human fecal origin differ in receptor specificity from those from UTI isolates. Indeed, *fim*H variants from commensally isolates primarily recognize oligomannose-like receptors, while those from UTI isolates preferentially bind monomannose-like receptors and provide an adaptive advantage for bacterial colonization of the urinary tract [17]. We selected *fim*H gene for describing the epidemiological evolution of UPEC and other sources E. coli in China for several reasons. (a) The vast majority of both intestinal and extraintestinal E. coli expresses type 1 fimbriae [24]. (b) *fim*H is a critical determinant of tropism for the urinary tract and vaginal epithelium for extraintestinal E. coli [14, 25]. An urovirulent phenotype is associated with genetic variants of this protein and hence is potentially clinically relevant [18]. (c) It has been reported that non-synonymous mutations accumulate at the *fim*H locus at a high rate [26]. Therefore, the level of discrimination at this locus is likely to be sufficiently high enough for studying uropatho-

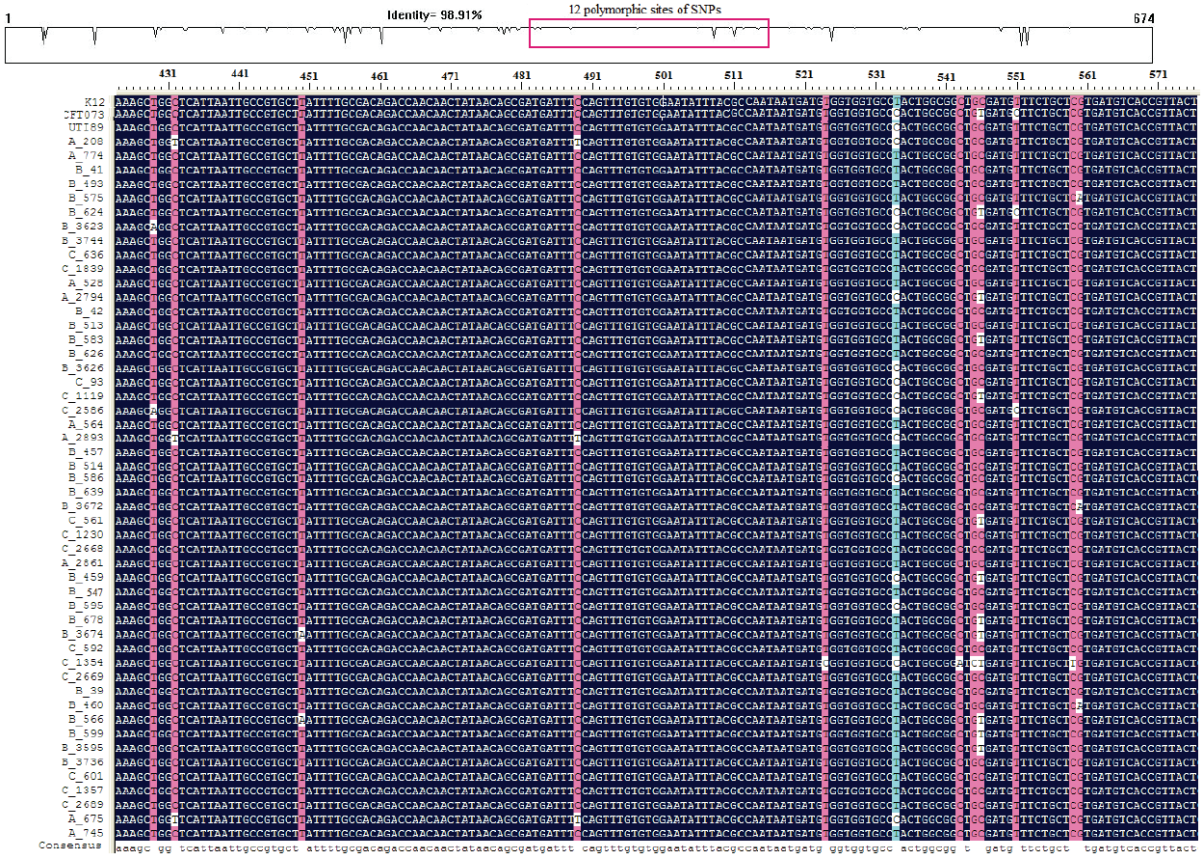

Figure 4. Sequences from 424 to 576 bp and 12 polymorphic sites of SNPs analysis were shown. They pick images were boxed above.

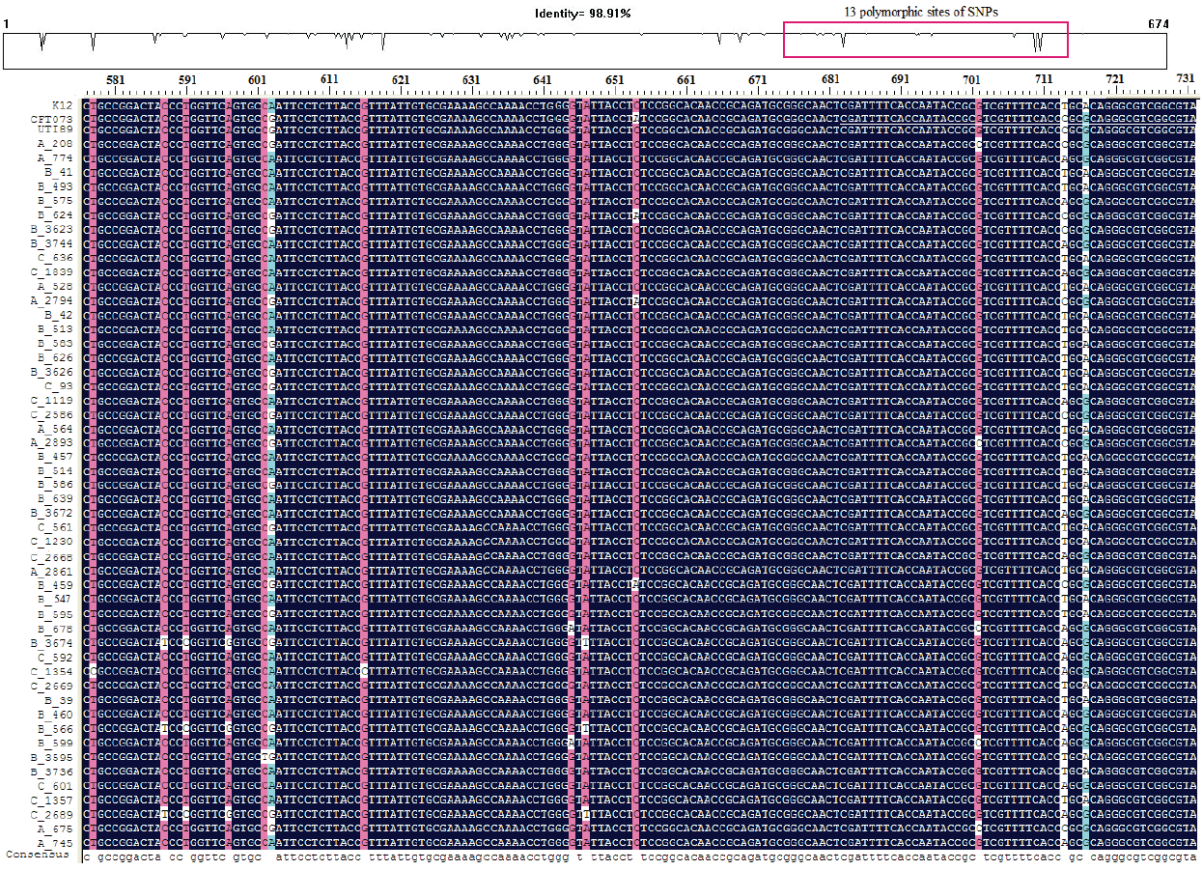

Figure 5. Sequences from 477 to 732 bp and 13 polymorphic sites of SNPs analysis were shown. They pick images were boxed above.

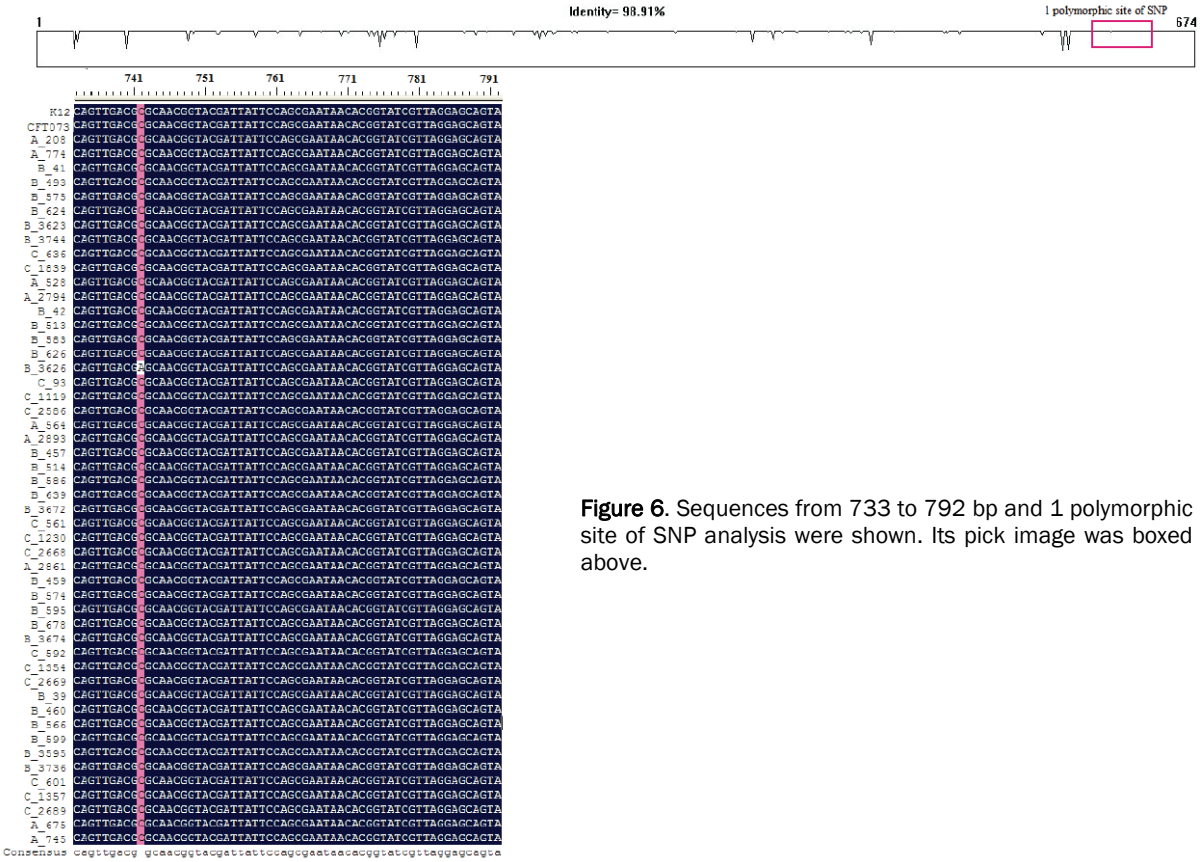

Figure 6. Sequences from 733 to 792 bp and 1 polymorphic site of SNP analysis were shown. Its pick image was boxed above.

genic E. coli UTI and other sources epidemiological events that occur over short time periods or in restricted geographical settings.

In this study we found that the comparison of UPEC strains and other extra-intestinal E.coli strains in different phylogroups shown a statistical significant of group A than group B1 in the other extra-intestinal E.coli strains ($P<0.05$), group D is statistical significant than group A in UPEC strains ($P<0.05$). No significant found between other comparisons. Except some singletons, the numbers of strains with the same amino-acid variants (**Table 4)** were observed more in UPEC than the other extra-intestinal strains. 23/37 silent SNPs (**Table 5**) were observed in UPEC strains. Most of the UPEC strains with the same amino-acid variants (**Figure 1**) were belong to the same phylogroup. Compared with the previous study results that shown the SNPs located at bp 401 to 824 of E. cokli-12 [11] (the SNPs shown in this study is located at bp 119 to 792 of E. coli K-12), interestingly we found that 4 amino-acid variants: V184A, Y216F, A140V, and R187H were observed in reference study and belong to the same phylogroup B2, D, A and D, D and B2, respectively, the reference strains compared were: f-8 strain (f-8=CFT073 in this study), f-28 strain, f-4, and f-9 strain. Except f-4 strains, and f-8 strain, all were distinct terminal groups representing distinct allelic variant with the strains isolated in this study. 14 silence SNPs: C411G, G414A, T534C, A603G, A717G, C654A, G420A, T429A, T591C, A597G, G408T, T450A, C432T, and C489T found in this study were also observed in reference study [11].

The results from this study allowed us to conclude that this combination method could serve as a rapid, relatively inexpensive, highly reproducible typing test for epidemiological studies of Extra-intestinal pathogenic E. coli in the future. The data could be compared with other clinical laboratories to more understand the epidemiology investigation of pathogenic E. coli strains

Table 5. Comparison of *fim*H no amino-acid variants (37 silent SNPs) observed between UPEC and other extra-intestinal sources

Nucleotides change with Amino-acid no variants	Number of UPEC	Number of Other sources of ExPEC	Total number of each Amino-acid no variants
C141T	12	13	25
G171A	10	14	24
T318G	11	10	21
C339G	15	9	24
T714A	13	14	27
A717G	13	14	27
T534C	8	9	17
C546T	9	4	13
G420A	2	1	3
A603G	11	8	19
G645A	2	0	2
G702C	1	3	4
T207C	10	4	14
C225T	3	1	4
C327T	3	5	8
C321T	8	1	10
C411T	3	0	3
C246T	5	3	8
G333A	0	1	1
C588T	2	1	3
T591C	2	1	3
T597G	2	1	3
G408T	2	3	5
T450A	2	0	2
C543A	0	1	1
C300T	0	3	3
C432T	0	3	3
C489T	0	3	3
A312T	5	2	7
T351C	5	2	7
C411G	4	2	6
G414A	4	2	6
T429A	0	1	1
C654A	3	1	4
C171A	3	1	4
G210A	3	1	4
T396G	3	1	4

Abbreviations: Extra-intestinal pathogenic E. coli (ExPEC), Uropathogenic E. coli (UPEC).
Total numbers of strains having a specific nucleotide changes are shown on the right. Underlined nucleotide change were singleton.

between different regions. A large collection of E. coli isolates could initially be screened for *fim*H sequence. Those that are indistinguishable can then be tested further by MLST or PFGE. This test could replace ERIC-PCR for typing E. coli isolates, especially in laboratories with access to a sequencing facility.

Acknowledgements

This work was supported by grant of international students studying in China under Chinese government scholarship program. I am grateful to Dr WEI Dian-jun for his expert assistance support to finish this work. I express my sincere

Table 6. Strain annotations and they accession numbers in GenBank

Strain annotations in text		Strain Annotation in GenBank	Genbank accession no.
1	A208	TMUSHA208	JN408528
2	A528	TMUSHA528	JN408529
3	A564	TMUSHA564	JN408530
4	A675	TMUSHA675	JN408531
5	A745	TMUSHA745	JN408532
6	A774	TMUSHA774	JN408533
7	A2494	TMUSHA2494	JN408534
8	A2861	TMUSHA2861	JN408535
9	A2893	TMUSHA2893	JN408536
10	B39	TMUSHB39	JN408537
11	B41	TMUSHB41	JN408538
12	B42	TMUSHB42	JN408539
13	B457	TMUSHB457	JN408540
14	B459	TMUSHB459	JN408541
15	B460	TMUSHB460	JN408542
16	B493	TMUSHB493	JN408543
17	B513	TMUSHB513	JN408544
18	B514	TMUSHB514	JN408545
19	B566	TMUSHB566	JN408546
20	B547	TMUSHB547	JN408547
21	B575	TMUSHB575	JN408548
22	B583	TMUSHB583	JN408549
23	B586	TMUSHB586	JN408550
24	B595	TMUSHB595	JN408551
25	B599	TMUSHB599	JN408552
26	B624	TMUSHB624	JN408553
27	B626	TMUSHB626	JN408554
28	B639	TMUSHB639	JN408555
28	B678	TMUSHB678	JN408556
30	B3595	TMUSHB3595	JN408557
31	B3623	TMUSHB3623	JN408558
32	B3626	TMUSHB3626	JN408559
33	B3672	TMUSHB3672	JN408560
34	B3674	TMUSHB3674	JN408561
35	B3736	TMUSHB3736	JN408562
36	B3744	TMUSHB3744	JN408563
37	C93	TMUSHC93	JN408564
38	C561	TMUSHC561	JN408565
39	C592	TMUSHC592	JN408566
40	C601	TMUSHC601	JN408567
41	C636	TMUSHC636	JN408568
42	C1119	TMUSHC1119	JN408569
43	C1230	TMUSHC1230	JN408570
44	C1354	TMUSHC1354	JN408571
45	C1357	TMUSHC1357	JN408572
46	C1839	TMUSHC1839	JN408573
47	C2586	TMUSHC2586	JN408574
48	C2668	TMUSHC2668	JN408575
49	C2669	TMUSHC2669	JN408576
50	C2689	TMUSHC2689	JN408577
51	UTI89[a]	UTI89	NC_007946
52	CFT073[b]	CFT073	NC_004431
53	K-12 (reference)[c]	K-12	NC_000913

[a]*fim*H gene is located from 4913555 to 4914457 bp; [b]*fim*H gene is located from 5143529 to 514440 bp; [c]*fim*H gene is located from 4546831 to 4547733 bp.

thanks to all the teachers of the Department of Clinical Laboratory of Second Hospital of Tianjin Medical University for their Technical Help. I would also like to thank also Cao Yang, Men Kun for their outstanding contributions to the understanding some machine manipulations.

Address correspondence to: Dr. WEI Dian-Jun, Department of Clinical Laboratory, Second Hospital of Tianjin Medical University, 300211, Tianjin; Tianjin Institute of Infection Diseases, Second Hospital of Tianjin Medical University, 300211, Tianjin, China. E-mail: weidianjun@gmail.com

References

[1] Russo TA, Johnson JR. Medical and economic impact of extra-intestinal infections due to Escherichia coli: focus on an increasingly important endemic problem. Microbes Infect 2003; 5: 449-56.

[2] Russo T A, Johnson JR. Proposal for a new inclusive designation for extraintestinal pathogenic isolates of Escherichia coli: ExPEC. J Infect Dis 2000; 181: 1753-4.

[3] Johnson JR, Russo TA. Molecular epidemiology of extraintestinal pathogenic (uropathogenic) Escherichia coli. Int J Med Microbiol 2005; 295: 383-404.

[4] Kucheria R, Dasgupta P, Sacks SH, Khan MS, Sheerin NS. Urinary tract infections: new insights into a common problem. Postgrad Med J 2005; 81: 83-6.

[5] Pong A, Bradley JS. Bacterial meningitis and the newborn infant. Infect Dis Clin North Am 1999; 13: 711-33.

[6] Eisenstein BI, Jones GW. The spectrum of infections and pathogenic mechanisms of Escherichia coli. Adv Intern Med 1988; 33: 231-52.

[7] Johnson JR. Virulence factors in Escherichia coli urinary tract infection. Clin Microbiol Rev 1991; 4: 80-128.

[8] Bingen E, Picard B, Brahimi N, Mathy S, Desjardins P, Elion J, Denamur E. Phylogenetic analysis of Escherichia coli strains causing neonatal meningitis suggests horizontal gene transfer from a predominant pool of highly virulent B2 group strain. J Infect Dis 1998; 177: 642-650.

[9] Srinivas V Ramaswamy, Robert Reich, Shu-Jun Dou, Linda Jasperse, Xi Pan, Audrey Wanger, Teresa Quitugua and Edward A Gravis. Single nucleotide polymorphisms in genes associated with isoniazid resistance in Mycobacterium tuberculosis. Antimicrobial Agents and Chemotherapy 2003; 47: 1241-1250.

[10] Sokurenko EV, Hasty DL and Dykhuizen DE. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. Trends Microbiol 1999; 7: 191-195.

[11] Rubens CS Dias, Beatriz M Moreira, and Lee W Riley. Use of fimH Single-Nucleotide Polymorphisms for Strain Typing of Clinical Isolates of Escherichia coli for Epidemiologic Investigation. J Clin Microbiol 2010; 48: 483-488.

[12] Clermont O, Bonacorsi S and Bingen E. Rapid and simple determination of the Escherichia coli phylogenetic group. Appl Environ Microbiol 2000; 66: 4555-4558.

[13] Gordon DM, Clermont O, Tolley H and Denamur E. Assigning Escherichia coli strains to phylogenetic groups: multilocus sequence typing versus the PCR triplex method. Environ Microbiol 2008; 10: 2484-2496.

[14] Hung CS, Bouckaert J, Hung D, Pinkner J, Widberg C, DeFusco A, Auguste CG, Strouse R, Langermann S, Waksman G, Hultgren SJ. Structural basis of tropism of Escherichia coli to the bladder during urinary tract infection. Mol Microbiol 2002; 44: 903-915.

[15] Tamura K, Dudley J, Nei M and Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 2007; 24: 1596-1599.

[16] Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 1980; 16: 111-120.

[17] Sokurenko EV, Chesnokova V, Dykhuizen DE, Ofek I, Wu XR, Krogfelt KA, Struve C, Schembri MA, Hasty DL. Pathogenic adaptation of Escherichia coli by natural variation of the FimH adhesin. Proc Natl Acad Sci USA 1998; 95: 8922-8926.

[18] Hommais F, Gouriou S, Amorin C, Bui H, Rahimy MC, Picard B and Denamur E. The FimH A27V mutation is pathoadaptive for urovirulence in Escherichia coli B2 phylogenetic group isolates. Infect.Immun 2003; 71: 3619 3622.

[19] Kucheria R, Dasgupta P, Sacks SH, Khan MS and Sheerin NS. Urinary tract infections: new insights into a common problem. Postgrad Med J 2005; 81: 83-86.

[20] Warren JW. Host parasite interactions and host defence mechanisms, Chapter Infection Diseases of the kidney. Edited by Schrier RW, Gottschalk CW. Eds, Little Brown, London, 1997; pp. 873-894.

[21] Weissman SJ, Moseley SL, Dykhuizen DE, Sokurenko EV. Enterobacterial adhesins and the case for studying SNPs in bacteria. Trends Microbiol 2003; 11: 115-117.

[22] Tartof SY, Solberg OD, Riley LW. Genotypic analyses of uropathogenic Escherichia coli based on fimH single nucleotide polymorphisms (SNPs). J. Med. Microbiol 2007; 56: 1363-1369.

[23] Fallah Mehrabadi J, Ghoraba H MSc, Imani Fooladi AA. Rohaninejad H MSc, Amini Hosseini SS MSc. Comparing of fimH gene variation in normal flora and uropathogenic Escherichia coli. Kowsar Med J 2010; 15: 65-69.

[24] Johnson JR, Kuskowski MA, Gajewski A, Soto S, Horcajada JP, Jimenez de Anta, MT and Vila J. Extended virulence genotypes and phylogenetic background of Escherichia coli isolates from patients with cystitis, pyelonephritis, or prostatitis. J Infect Dis 2005; 191: 46-50.
[25] Klumpp DJ, Forrestal SG, Karr JE, Mudge CS, Anderson BE and Schaeffer AJ. Epithelial differentiation promotes the adherence of type 1-piliated Escherichia coli to human vaginal cells. J Infect Dis 2002; 186: 1631-1638.
[26] Sokurenko EV, Feldgarden M, Trintchina E, Weissman SJ, Avagyan S, Chattopadhyay S, Johnson JR and Dykhuizen DE. Selection footprint in the FimH adhesin shows pathoadaptive niche differentiation in Escherichia coli. Mol Biol Evol 2004; 21: 1373-1383.