



GENOMIC INSIGHTS INTO UROPATHOGENIC *E. COLI* IN CLINICAL SAMPLES

Al-Hassan, Omar Kareem Abdulrahman

College of Medicine, Al-Muthanna University, Samawah, Iraq

Abstract:

Uropathogenic E. coli is the primary causal agent responsible for urinary tract infections, particularly in women. We obtained a total of 30 samples from the Atatürk University Research Hospital in Erzurum, Türkiye, from January 2021 to April 2021. We established the presence of E. coli using routine identification and biochemical assays. We discovered certain isolates through whole genome sequencing. This investigation revealed that a total of 84 genes, coding for virulence factors (VF), were present in five isolates. Our isolates include the following virulence factors: 12 genes encoding adhesins, 9 genes encoding iron acquisition systems, 8 genes involved in protection and invasion, 4 genes encoding toxins, and several additional miscellaneous proteins. The isolates contained in B1 and B2 belong to phylogenetic group classes one and four, respectively. Isolates two, three, and five are all part of the same phylogenetic group. Sequence type (ST) 69 contains isolate one, while ST 162 now contains isolate two. We isolated three individuals in ST 405, four in ST 131, and five in ST 69.

Keywords: *E. coli, WGS, Urinary tract infection and MLST.*

Introduction

Launching commensal bacteria within a new host forte frequently promotes the transition from commensalism to pathogenicity.

Extraintestinal pathogenic *Escherichia coli* (ExPEC) are diverse pathovars that can live in two ways: they can live in the gut as valuable bacteria or get out and cause diseases in other parts of the body. Zlatkov, (2019) divides ExPEC into uropathogenic *E. coli* (UPEC), newborn meningitis-causing *E. coli* (NMEC), and sepsis-associated *E. coli* (SEPEC) based on the illness they are allied with.

Consequently, it was whispered that UPEC causes the patient's UTI to initiate from their GI tract but translocate to the urinary tract by colonizing the periurethral zone. Numerous virulence factors (VFs), which are exceptional components or

enzymes of the bacteria, make this colonization of the urinary tract sense. It takes these VFs to fight off the body's natural strong immune response to UPEC infection and assistance the bacteria endure in the punitive environment of the human urinary tract (Meerman, 2016). Functional groups such as adhesins, poisons, iron acquisition mechanisms, and protectins categorize UPEC VFs. Only chromosomes contain some VF-producing genes, like *pap* and *hly*, while plasmids connect others, like *iss* and *traT*, to them. Others, such as *Afa*, could be a combination of both. Therefore, the transmission

of VFs can occur either vertically or horizontally, complicating the understanding of the role of specific VF genes in the development of UTIs (Kudinha, 2017).

Moreover, studying these VFs and their corresponding genes can yield precise insights into the interplay between these factors in enteric *E. coli* pathotypes and host proteins at the molecular level. This knowledge can shed light on the mechanisms by which these factors contribute to diseases and enable the development of effective preventive measures (Kaper *et al.*, 2004).

Whole genome sequencing (WGS) examination of the complete pathogen genome has the potential to offer unparalleled precision in distinguishing even closely related lineages of bacteria. This has the potential to revolutionize the analysis of outbreaks in hospitals. However, clinicians have been reluctant to use WGS in outbreak analyses since the early sequencing platforms were costly and difficult to use. Advancements in sequencing technologies and analysis tools have significantly enhanced the efficiency and speed of WGS while also reducing the associated costs (Quainoo *et al.*, 2017). Multilocus sequence typing (MLST) disclose that *E. coli* and countless other bacterial species have a clonal structure, and that some clones are prevalent (Woodford *et al.*, 2011). Diverse lineages have split off from a common forebear of all *E. coli*, giving rise to four main phylogenetic groups: A, B1, B2, and D (O. Clermont *et al.*, 2000). Current sequence-based phylogeny of the species has verified the division of these lineages into numerous clonal groups, which further subdivide into manifold sublineages (Banerjee & Johnson, 2014 ; Nicolas-Chanoine *et al.*, 2014).

This study aimed to isolate and identify *E. coli* from UTI as a causative agent, as well as its distribution across sexes and age groups. This detection is based on morphological and selective cultural media. WGS technology aids in the investigation of virulence factor genes, phylogenetic groups, and Multilocus sequence typing, in addition to identifying the connections between the previous objectives, thereby enhancing our understanding of all aspects of our research.

Materials and methods 1. Sample collection

Between January and April 2021, at Atatürk University Research Hospital (Erzurum-Türkiye) collected midstream urine specimens from patients assumed of having UTIs in sterile ampules. We punctually sent the samples to the lab for microbiological examination. The collected thirty samples of both sexes, immediately shifted them to the labs, and kept them in a cooling container until further ID. **2. *E. coli* identification**

We cultured all specimens on MacConkey and blood agar for one night at 37 °C. *E. coli* selective medium (Eosin methylene blue EMB/Oxoid Ltd., Basingstoke, UK) encourages our isolates to grow after incubation at 37°C for 22–24 hours. The colonies we precisely identified are green and metallic in color. Based on the indole, methyl-red, Voges-Proskauer, and citrate reactions (IMViC), as well as sugar fermentation, the biochemical test deep-rooted the look of the culture medium (Mahe *et al.*, 2021).

3. Sequencing of the whole genome

We selected five isolates for whole-genome seq. First, DNA extraction kit steps at the beginning, followed by library preparations using the Miseq reagent V3 kit, according to the illumine company instructions at the BM laboratory in Ankara, Türkiye (<https://www.bmlabosis.com/>). the isolates 1, 2, 3, 4, and 5. step-in silico platform was performed by Miseq, San Diego, California, USA. From Miseq machine the outputs in Fastq file format, and the FastQC analysis confirmed their quality. The OmicsBox platform tools (via ABySS) assemble the reads using a *de novo* substant, following the

company's guide. The second option alignment via NCBI Blast bottom confirmed that NZ-CP017669.1 would be a proper reference strain in December 2021.

4. Virulence factors

E. coli, as a pathogenic species, has many virulence factors and enzymes that help in the colonization and pathogenicity processes. We investigated these factors using a genomic traits finder in the web-based Center for Genomic Epidemiology (CGE) at <http://cge.cbs.dtu.dk/services/VirulenceFinder>, which includes species selection, a threshold identity of 80%, and a minimum length of 60%. Finally, we uploaded the data in the form of assembled contigs.

5- Phylogeny

We screened our isolates in the Wen-interface to identify the phylogroups within the *E. coli* strains. Clermon typing serves the purpose of <http://clermonttyping.iame-research.center/index.php>. It is an in-silico method that tries to replicate the outcomes of the PCR phylotyping of *E. coli* that was first suggested by O. Clermont et al. (2000 and updated in 2013 by Olivier Clermont et al., 2013), as well as the different PCR tests created to identify the *Escherichia* clades (Olivier Clermont et al., 2013) as well as the various PCR assays developed to identify the *Escherichia* clades.

6- Multilocus Sequence Typing

In this step, the analysis of seven specific genes (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA*) in *E. coli* determines the identification of sequence types (STs) and their classification. These genes act as conserved indicators to determine the STs and their respective groups. The valuable webbased tool <https://cge.food.dtu.dk/services/MLS> provides MLST recognition by species selection and minimum depth allele, which work with genomic contigs to decide the area of MLSTs (Larsen *et al.*, 2012).

7. Available data

At NCBI we uploaded all the sequenced strains under Bioprojects (PRJNA835465 and read archive SRR1955197 for strain no.1 and PRJNA846072 and SRR 19543626, 19543627, 19543628 and 19543629 for stains 2, 3,4,5).

Results and discussions

We collected 30 midstream urine samples from four-month-old patients hospitalized at Erzurum Teaching Hospital, highly suspected of having UTIs. Figure (1) illustrated the age ranges: less 15 to over 56 years with a mean age 40,46; 20 (66,66%) of the study contributors were females, while percentages of men (33,33%). With statistic differences $p < 0.05$ and the age group 56–64 (33.33%) had the uppermost rate of bacterial UTIs out of whole age group 15–19 (23.3% of all samples)

Came in second. While ages 36-46 represent only two specimens in total.

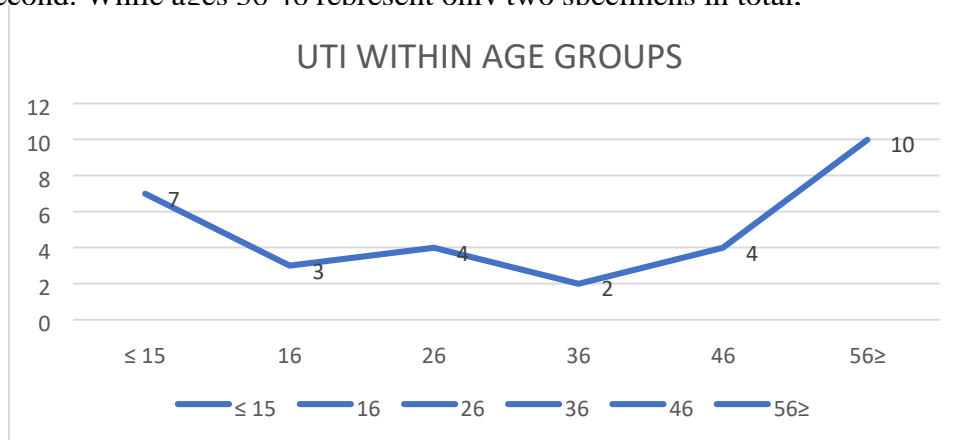


Figure 1: Illustrate infections distributed among age groups.

Nearly 10% of postmenopausal women report having a UTI in the past year, and 50% to 60% of adult women will experience at least one UTI (Alós, 2005). Recurrence within six months is common, and the risk of UTIs upsurges with age; the proportion in women over 65 is nearly twice that of the female populace overall and younger women; increased sexual activity is a major risk factor for UTIs (Medina & Castillo-Pino, 2019). In addition, semi of the Women suffer UTI thru their lives because of anatomical variances and physical causes in contrast with men (Al-Badr & Al-Shaikh, 2013; Fazly Bazzaz *et al.*, 2021).

2- Virulence-factor genes

The platform may identify virulence genes using two methods: a BLAST search using assembled genome data or k-mer alignment (KMA) of raw reads (Clausen *et al.*, 2018) against a FASTA database containing the virulence genes. The goal of developing VirulenceFinder was to provide a solution for rapid WGS-based virulence gene discovery and allele-level typing (Joensen *et al.*, 2014). In this investigation, we found a total of 84 genes that code for virulence factors in five isolates of *E. coli*. These genes have various activities that allow *E. coli* to adhere to, colonize, and invade many host cells during the course of urinary tract infection pathogenicity. The most including VFs genes in isolate 5 (23) (27,38%) then isolate 4 (21) (25%). Strains one, two and three represent (15,13,12) (17,85; 15,47; 14,28%) correspondingly. The VFs present in our isolates include 12 adhesin genes (*iha*, *afaA*, *afaC*, *afaD*, *hra*, *papAF43*, *papA-F11*, *papC*, *cea*, *ipfA*, *nfaE*, *yfcV*), 9 genes related to iron acquisition (*fuyA*, *iutA*, *sitA*, *iroN*, *iucC*, *irp*, *ipr2*, *chu*, *chuA*), 8 genes involved in protection and invasion (*traT*, *iss*, *terC*, *kpsE*, *kpsMII-K96*, *kpsMII-K5*, *kpsMIII-K96*, *hlyC*), 4 toxin genes (*sat*, *sata*, *cia*, *cib*), and 7 other miscellaneous protein genes (*gad*, *etsC*, *air*, *eilA*, *capU*, *mcmA*, *ompT*). Several genes, such as *gad* (3), *iss* (2), and *etsC* (2), are present in multiple copies in isolate no. 1. In isolate 2, *gad* (2), *sitA* (2), and *terC* (2) are also found in duplicate within the genome. In addition, isolate 3 included 3 copies each of *gad*, *terC*, *eilA*, *fuyA*, *hra*, *irp2*, *kpsE*, *kpsMII-K5*, *sitA*, and *traT*, whereas isolate 4 contained 2 copies each of *terC* and *traT*. Finally, *gad* (2), *iss* (2), *terC* (2), and *traT* (2) in the genome composition of isolate 5.

The sequencing initiatives uncovered significant genome variation within ExPEC, highlighting the presence genes of toxins, adhesins, iron acquisition systems, polysaccharide capsules, lipopolysaccharides (LPS), invasins and proteases. these factors often transferred via plasmids (Gomes *et al.*, 2010; Köhler & Dobrindt, 2011; Lu *et al.*, 2011). Moreover, several ExPEC virulence genes, ex:

hlyA, *fyuA*, *traT*, and *iutA* are sovereign predictors for pathogenicity. Particularly *iutA* and *traT* are independent predictors for pathogenicity (Lee *et al.*, 2010). Another finding is the detection of about 30 VF genes from *E. coli* as the causative agent of bacteremia, and rectal swabs (fecal samples) include many genes that support the facts of our plans (Sannes *et al.*, 2004). *E. coli* isolates that transmit *chuA*, *fyuA*, *vat*, and *yfcV* genes efficiently take over the urinary tract and elevation the uropathogenicity (Crossman *et al.*, 2010; Spurbeck *et al.*, 2012), that cornerstone for differentiate commensal from pathogenic strains in presence *chu* and *fyu* genes. Current data from isolate one revealed absence *chu* gene to exploit iron from blood and the reason is this isolate belongs to phylogroup B1 (see next section) via analyzing many genomes of *E. coli* pathogroups Pap gene encodes P fimbriae (pili of pyelonephritis) are related with uropathogenic strains and aim glycosphingolipids (Mulvey, 2002).

proteins of Iron acquisition been recognized as being linked to virulence due to the fact that iron is a crucial nutrient that is limited in availability for pathogenic bacteria within the host. This scarcity of iron stimulates pathogenicity, particularly in relation to heme intake (Subashchandrabose & Mobley, 2015; Wilks & Heinzl, 2014). Our study detected nine iron acquisition virulence genes, enriching the pathogenic isolates to survive. Iss gene encodes proteins that attack the host complement system through an immune defenses arms, presented in isolate five, among ExPEC and other pathotypes because they are distributed sporadically and prophage encoded (Lynne *et al.*, 2006).

3- Phylogenetic groups

In 2000s Clermont and associates created a triplex experiment an *E. coli* strain divides into one of the main phylogroups, A, B1, B2, or D depending on the attendance or absence of these three genes (Clermont *et al.*, 2000). The quadruplex polymerase chain reaction (PCR) was developed in 2013 by Clermont and colleagues to sort *E. coli* into phylogroups A, B1, B2, C, D, E, F, and clade I. PCR was supplemented with a gene target, *arpA* (Clermont *et al.*, 2013). Table (1) explains our findings in phylogenetic group classes represented, on the one hand, isolates no.1 and no.4 located in B1 and B2, correspondingly; on the other hand, isolates no.2, no.3, and no.5 occupied the same phylogenetic group D.

Table 1: phylogenetic groups of *E. coli* sequenced isolates

Isolate No.	Quadruplex genes	Supp.	Phylogroup	Mash group
1	+ - - +	fdm, trpAgpC, ybgD	B1	B1
2	+ + - -	trpAgpC, ybgD	D	D
3	+ + - +	trpAgpC, ybgD	D	D
4	- + + +	trpAgpC	B2	B2
5	+ + - -	trpAgpC, ybgD	D	D

Within all groups A, B1, B2, and D, *E. coli* belong to phylogroups B2 and D are more likely to cause extraintestinal infection and possess the consistent virulence genes than A and B1 (Johnson & Stell,

2000; Picard et al., 1999). Besides, phylogroup B2 was positively allied with the UPEC pathotype. In contrast, the D group was more connected with the other ExPEC and less with the UPEC pathotypes (Hutton et al., 2018). Tenaillon et al., 2010 found that the majority of extraintestinal pathogenic *E. coli* isolates belong to phylogenetic groups B2 and D. In our study, these groups were present in 26% and 32% of the isolates, respectively. The overrepresentation of members from the phylogroups B2 and D in ExPECs may be attributed to the upregulation of iron acquisition genes. Recent studies have demonstrated that the presence of these genes enhances the inherent pathogenicity of the bacteria (Id et al., 2020). This relation supports our entitlements of spreading those genes in subjected phylogroups.

4- Sequence typing

MLST or also referred to as Sequence typing (ST) based on several genes or housekeeping genes of an organism's genome. Currently, seven genes were detected as a golden mark of *E. coli* sequence typing: *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* and *recA*. Data revealed that all isolates had different areas of STs with 100% of identity and without any gaps. Here, isolate one is set in ST 69 but isolate two let fall in ST 162. The others isolated three in ST 405, isolate four in ST 131 and isolate five settled in ST 69.

The Afa fimbrial adhesin is produced by the *afa* genes and is commonly seen in ExPEC, UPEC, and broadly adhering *E. coli* strains (Dhakal et al., 2008; Servin, 2005). Most strains covering genes *afa* related to ST131 clonal group (Clark & Maresso, 2021). *Sat* toxin is a vacuolating cytotoxin concerned in uropathogenesis (Guyer et al., 2002). *FimH*, *fyuA*, *kpsMII*, *iha*, *iutA*, *tratT*, *ompT* and *usp* are all golden marks for detecting ST131 in extra-intestinal (NicolasChanoine et al., 2008; Shaik et al., 2017). It is also detected in ST131 and ExPEC strains from the B2 phylogroup beside B1 and D (Guyer et al., 2000; Hojabri et al., 2020); these elicit ST131 of isolate four expressed *sat*, *afaA*, *afaC*, and *afaD* gene and other VFs in database analysis under investigation. ST131 strains are frequently shown for generate extended-spectrum β -lactamases, specifically CTX-M₁₅ (Lau et al., 2008), and nearly all are fluoroquinolones resistant. ST131 is related with thirty percent of all ExPEC, 60-90% of fluoroquinolones resistant ExPEC, and about 80% of ESBL-resistant ExPEC (Pitout & DeVinney, 2017). ST405 associated with the production of CTX-M₁₅ ESBL; ST131 and ST405 are extremely virulent and have resistant genes in an inconsistent manner (Mihaila et al., 2010; Smet et al., 2010). Most UTIs ST69 were MDR and naturally port a single arrangement of a gene cassette (*dfrA17*–*aadA5*) encoding dihydrofolate reductase for trimethoprim/sulfamethoxazole and aminoglycoside adenylyltransferase, respectively, on a class I integron (Ajiboye et al., 2009; Solberg et al., 2006). Serious public health and clinical anxieties are the broadminded acquisitions of drug resistance in clonal lineages, mainly global distribution lineages, including ST131, ST393, ST69, ST73, and ST95 (Riley, 2014). Outcomes of Ludden et al., (2021) resemblance with ours in the most identified STs were ST131, ST10, and ST69 from clinical cohorts of infected patients with pathogenic *E. coli*.

Conclusions

E. coli was isolated and identified as the causative agent for urinary tract infection in females other than males, with a prevalence mainly in age 56 and older. Virulence factors distributed within uropathogenic *E. coli* while Phylogenetic group D was recorded as the most typical of the three isolates, while B1 and B2 came after in data analysis. Sequence typing revealed ST69 in two isolates, ST405, ST162, and ST131.

References

Ajiboye, R. M., Solberg, O. D., Lee, B. M., Raphael, E., Debroy, C., & Riley, L. W. (2009). Global spread of mobile antimicrobial drug resistance determinants in human and animal *Escherichia coli* and

salmonella strains causing community-acquired infections. *Clinical Infectious Diseases*, 49(3), 365–371.

Al-Badr, A., & Al-Shaikh, G. (2013). Recurrent urinary tract infections management in women: A review. *Sultan Qaboos University Medical Journal*, 13(3), 359–367.

Alós, J. I. (2005). Epidemiology and etiology of urinary tract infections in the community. Antimicrobial susceptibility of the main pathogens and clinical significance of resistance.

Banerjee, R., & Johnson, J. R. (2014). A new clone sweeps clean: The enigmatic emergence of *Escherichia coli* sequence type 131. *Antimicrobial Agents and Chemotherapy*, 58(9), 4997–5004.

Clark, J. R., & Maresso, A. M. (2021). *Comparative Pathogenomics of Escherichia coli: Polyvalent Vaccine Target Identification through Virulome Analysis*. February, 1–42.

Clausen, P. T. L. C., Aarestrup, F. M., & Lund, O. (2018). Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics*, 19(1), 1–8.

Clermont, O., Bonacorsi, S., & Bingen, E. (2000). Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Applied and Environmental Microbiology*, 66(10), 4555–4558.

Clermont, Olivier, Christenson, J. K., Denamur, E., & Gordon, D. M. (2013). The Clermont *Escherichia coli* phylo-typing method revisited: Improvement of specificity and detection of new phylogroups. *Environmental Microbiology Reports*, 5(1), 58–65.

M. J., Penn, C. W., Parkhill, J., Turner, A. K., Johnson, T. J., Thomson, N. R., Smith, S. G. J., & Henderson, I. R. (2010). A commensal gone bad: Complete genome sequence of the prototypical enterotoxigenic *Escherichia coli* strain H10407. *Journal of Bacteriology*, 192(21), 5822–5831.

Dhakal, B. K., Kulesus, R. R., & Mulvey, M. A. (2008). Mechanisms and consequences of bladder cell invasion by uropathogenic *Escherichia coli*. *European Journal of Clinical Investigation*, 38(SUPPL.2), 2–11.

Fazly Bazzaz, B. S., Darvishi Fork, S., Ahmadi, R., & Khameneh, B. (2021). Deep insights into urinary tract infections and effective natural remedies. *African Journal of Urology*, 27(1).

Gomes, D., Bertoldi, I., Spagnuolo, A., Marchi, S., Rosini, R., Nesta, B., Moriel, D. G., Bertoldi, I., Spagnuolo, A., Marchi, S., Rosini, R., Nesta, B., Pastorello, I., Mariani Coreia, V. A., Torricelli, G., Cartocci, E., Savino, S., Scarselli, M., Dobrindt, U., ... Serino, L. (2010). Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 107(20), 9072–9077.

Guyer, D. M., Henderson, I. R., Nataro, J. P., & Mobley, H. L. T. (2000). *Identification of Sat , an autotransporter toxin produced by uropathogenic Escherichia coli*. 38.

Guyer, D. M., Radulovic, S., Jones, F., Mobley, H. L. T., Guyer, D. M., Radulovic, S., Jones, F., & Mobley, H. L. T. (2002). *Sat , the Secreted Autotransporter Toxin of Uropathogenic Escherichia coli , Is a Vacuolating Cytotoxin for Bladder and Kidney Epithelial Cells Sat , the Secreted Autotransporter Toxin of Uropathogenic Escherichia coli , Is a Vacuolating Cytotoxin for Bla.*

Hojabri, Z., Darabi, N., Mirmohammadkhani, M., Rahimi, H., Hemmati, R., Saeedi, Z., Roustaei, K., Leekitcharoenphon, P., Pajand, O., & Aarestrup, F. M. (2020). Expansion of a Subset within the C2 Subclade of Escherichia coli Sequence Type 131 (ST131) Is Driving the Increasing Rates of Aminoglycoside Resistance.

Hutton, T. A., Innes, G. K., Harel, J., Garneau, P., Cucchiara, A., Schifferli, D. M., & Rankin, S. C. (2018). Phylogroup and virulence gene association with clinical characteristics of Escherichia coli urinary tract infections from dogs and cats. *Journal of Veterinary Diagnostic Investigation*, 30(1), 64–70.