

Uropathogenic *Escherichia coli*: An Ideal Resource for DNA Microarray Probe Designing

Payam Behzadi^(✉) and Elham Behzadi

Department of Microbiology, College of Basic Sciences, Shahr-e-Qods Branch,
Islamic Azad University, Tehran, Iran
behzadipayam@yahoo.com, elham_behzadi@hotmail.com

Abstract. Uropathogenic *Escherichia coli* (UPEC) is the most important bacterial agent causing urinary tract infections (UTIs) in patients around the world. The UTIs rank second among different types of infectious diseases. So, there is an urgent need to have a rapid and accurate diagnostic method for detecting UTIs. DNA microarray is an advanced pan-genomic technique which can be used as a rapid and accurate diagnostic method with high specificity and sensitivity. Designing a collection of DNA microarray probes enables us to have a sharp methodology for detection and identification of UPEC pathotypes. For this reason, the authors of the present review literature have tried to represent a vast range of availabilities regarding UPEC recognition by DNA microarray probe designing.

UPEC encompasses a wide range of virulence genes which are contributed to UTIs. This characteristic makes UPEC an ideal resource for DNA microarray probe designing to have a reliable UTIs diagnostic method.

Besides, UPEC possesses different pathogenic strains which are distributed worldwide. And individual strain has its particular conserved sequences within the virulence genes. In other word, DNA microarray probes can be designed for specific UPEC strains within different geographical regions.

In conclusion, UPEC is a living accessible genomic treasure which can be used for designing a diversity of DNA microarray probes with a high specificity and sensitivity.

Keywords: Microarray · Urinary tract infection · Uropathogenic *Escherichia coli* · Probe · Pangenome

1 Introduction

Urinary tract infections (UTIs) as the second most important infectious diseases are occurred by the well-known pathogenic microorganism of Uropathogenic *Escherichia coli* (UPEC). As we know, the UTIs are classified into lower and upper parts with asymptomatic or symptomatic, complicated or uncomplicated and acute or chronic characteristics. In addition to human disposing factors for UTIs, the pathogenomic properties of microbial causative agent play an important role for occurring UTIs. Among a wide range of UTIs causative agents, UPEC has a unique situation. Among

patients with UTIs around the world, the incidence of UTIs in women is four-folded in comparison with men. This feature is referred to anatomical characteristics in females. Several global reports confirm the key role of intestinal strains of *Escherichia coli* (*E.coli*) as significant resources for recurrent UTIs in women [1–6].

The phylogenetic studies show that, all the pathotypes of *E.coli* are originated from the commensal and non-pathogenic, strains in their evolutionary pathways. By the time, the intestinal and extra-intestinal pathotypes of *E.coli* are divided into 5 classes of A, B1, B2, D and E. The A, B1 and E classes are recognized as intestinal strains and the classes of B2 and D involve UPEC strains for the most. However, the B2 group dominates D category. According to previous studies, the UPEC strains are the most well-known etiologic bacterial agents of nosocomial (up to 50%) and community acquired (up to 95%) UTIs [3, 7].

Pangenomics has revealed a huge diversity of genes in non-pathogenic and pathogenic *E.coli* strains; so the pathogenomics of *E.coli* may help us to detect and identify different pathotypes of UPEC by advanced molecular and pangenomic technology like DNA Microarray [3, 8].

2 Pangenomics of UPEC

The scientific term of Pangenomics was firstly coined in 2005 by Tettelin and colleagues. Indeed, pangenome denotes common chromosomal sections (known as core or backbone genome) and supplementary genetic elements (adaptive or flexible genome) in a Genus. The core genome encompasses intra-chromosomal gene families including housekeeping and other functional genes which are directly related to normal cellular activities; but the adaptive or supplementary genome comprising mobile genes in the forms of integrons, pathogenicity islands (PAIs), phages, plasmids, retrotransposons and transposons which are in association with vital defense systems. For the most, the flexible genetic repertoire encompasses antimicrobial resistance, heavy metal resistance and virulence genes that pertaining to microorganism environmental adaptation. Today, the microbial Pangenomics is an effective means for detection and identification of microbial causative agents of infectious diseases. Among a wide range of infections, UTIs as common global infectious diseases have an urgent need for an accurate diagnostics and a definite treatment. So, the use of Pangenomics and advanced pangenomic diagnostic techniques is unavoidable. The pangenome (genetic repertoire) of a microorganism like UPEC varies from one strain to another. In accordance with previous studies, the pangenome concept of *E.coli* varies between 4.5 to 5.5 Mb. The least boarder is recognized in commensal strains of *E.coli*, while the pathotypes including UPEC strains possess more genomic contents. The precise pangenomic concept for each *E.coli* strain is accessible through the National Center for Biotechnology Information (NCBI) database (<https://www.ncbi.nlm.nih.gov/>). NCBI is an important and easy-to-work database which provides us the latest information regarding deciphered microbial pangenomes. Until today, 5363 genome assemblies, 342 sequence reads and 47198 genes are registered for *E.coli* strains in NCBI [3, 9–16].

Although, there are a large number of genes which are valuable for DNA microarray detection and identification, the unrelated genes (orphan genes) among a

mass of gene families within flexible gene pool are appropriate sequences as target goals. The orphan gene sequences prepare an appropriate specificity for influent detection and identification by DNA microarray technology [10, 17, 18].

3 DNA Microarray Technology

Microarray technology is an advanced pangenomic diagnostics which can be used as a reliable, flexible, rapid, accurate, sensitive and specific tool. DNA microarray technique as a miniaturized lab-on-a-chip is appropriate and cost effective when the number of specimens is high. The technique of DNA microarray is principally classified into two separate sections which each part has an operative effect on the other. These sections include dry lab (known as *in silico*) and wet lab (known as *in vitro*). The probe designing section involves *in silico* activities and is completely dependent on databases, servers, tools and software. Probe designing process determines the accuracy, sensitivity and specificity of the method. Employing a skillful bioinformatician is an urgent need to minimize the probable biases. The use of proper analyzers, tools and software affects the final outcomes. Besides, utilization of suitable probe sets is an essential need, too. Noticeably, probe designer software represents several microarray probes which the best ones must be chosen and rechecked for final application. On the other hand, probe spotting, DNA target labeling, hybridization and scanning are the wet lab portion of DNA microarray technique in which the related procedures must be performed in accordance with proper protocols. In brief, all of the procedures and sections must be achieved in correct and precise manner to have an acceptable result [3, 19–24].

In parallel with aforementioned procedures, the more registered recognized microbial genomes in databases such as NCBI, European Molecular Biology Laboratory (EMBL, <http://embl.org/>) and DNA Data Bank of Japan (DDBJ, <http://www.ddbj.nig.ac.jp/>), the more accurate and effective designed microarray probes. This is why, we can design very effective and sharp microarray probes for E.coli pathotypes like UPEC [9, 10, 12, 19, 25, 26].

The UPEC pathotypes encompass a wide range of genes (e.g. virulence genes) with a significant diversity. In other word, UPEC is an incredible treasure for designing a huge number of well-designed microarray probes. These properties make DNA microarray an appropriate diagnostic technique for detection and identification of UPEC. The virulence genes may be placed on main chromosome or plasmids. The use of chromosomal genes is suitable for recognition of UPEC pathotypes by phylogenetic properties while the use of plasmid genes is proper for multidrug-resistance UPEC strains [3, 27].

4 Chromosomal Virulence Genes (CVGs)

In accordance with several studies, the presence of virulence genes in UPEC genetic pool makes it easy for the bacterium to adapt itself with peripheral factors. This property enables UPEC to be a successful pathogenic bacterium causing UTIs. Adhesins, invasins, siderophores, iron uptake systems, secretion systems, toxins and etc. are favorable

Table 1. Categorization of different virulence genes relating to adhesins in UPEC strains

Virulence genes	Length (base pair (bp))	Virulence factors	Characteristics	Activity	Location	Type of UTI
<i>afaA</i>	306	Afimbrial adhesion AFA-I	Cluster genes	Adhesin	Exterior	Chronic and recurrent, Lower and Upper UTIs
<i>afaB</i>	798					
<i>afaC</i> (highly conserved)	2583					
<i>afaD</i> (highly conserved)	444					
<i>afaE-I</i>	485					
<i>draA</i>	305	Afimbrial adhesion Dr	The proteins are the same as <i>afa</i> genes products	Adhesin, invasin (by activating signal transduction cascade)	Exterior	Chronic and recurrent, Lower and Upper UTIs
<i>draB</i>	743					
<i>draC</i>	2579					
<i>draD</i>	443					
<i>draE</i>	482					
<i>draP</i>	173					
<i>cgsA</i>	459	Curli fimbriae	Fimbrial shaped	Adhesin	Exterior	All
<i>cgsB</i>	483					
<i>cgsC</i>	333					
<i>cgsD</i>	651					
<i>cgsE</i>	390					
<i>cgsF</i>	417					
<i>cgsG</i>	834					
<i>cfaA</i>	764	CFA/I fimbriae	Pilus	Adhesin	Exterior	Lower and recurrent UTIs
<i>cfaB</i>	547					
<i>cfaC</i>	2914					
<i>cfaD</i>	1217					
<i>cfaE</i>						
<i>fimA</i>	612	Type 1 fimbriae	Bacterial peripheral surrounding fimbriae, cluster genes, conserved genes, chaperone-usher (CU) secretion fimbriae	Adhesin	Exterior	All
<i>fimB</i>	603					
<i>fimC</i>	726					
<i>fimD</i>	2844					
<i>fimE</i>	597					
<i>fimF</i>	544					
<i>fimG</i>	512					
<i>fimH</i>	912					
<i>fimI</i>	692					
<i>focA</i>	543	FIC fimbriae	Chaperone-usher (CU) secretion fimbriae, Cluster genes, in association with cluster genes of S fimbriae	Adhesin	Exterior	All
<i>focB</i>	329					
<i>focC</i>	696					
<i>focD</i>	2679					
<i>focF</i>	528					
<i>focG</i>	504					
<i>focH</i>	1011					
<i>focI</i>	540					

(continued)

Table 1. (continued)

Virulence genes	Length (base pair (bp))	Virulence factors	Characteristics	Activity	Location	Type of UTI
<i>sfaA</i>	556	S fimbriae	Cluster genes, the S fimbriae compartments are variable in different environmental factors	Adhesin	Exterior	Severe UTIs (mostly in upper UTIs)
<i>sfaB</i>	330					
<i>sfaC</i>	285					
<i>sfaD</i>	549					
<i>sfaE</i>	696					
<i>sfaF</i>	2679					
<i>sfaG</i>	528					
<i>sfaH</i>	969					
<i>sfaS</i>	492					
<i>papA</i>	636	P fimbriae	Gene clusters, molecular clock for recognition of B2 and D strains, Chaperone-usher (CU) secretion fimbriae,	Adhesin	Exterior	Acute lower and upper (especially nephritis) UTIs
<i>papB</i>	315					
<i>papC</i>	2532					
<i>papD</i>	735					
<i>papE</i>	555					
<i>papF</i>	534					
<i>papG</i>	1034					
<i>papH</i>	591					
<i>papI</i>	368					
<i>papJ</i>	582					
<i>papK</i>	540					
<i>papX</i>	555					
<i>elfA</i>	552	Escherichia coli laminin binding fimbriae	Gene cluster, molecular clock for recognition of B1 strains, Recognized in Enterohemorrhagic <i>E. coli</i> (EHEC) strains for the most	Adhesin	Exterior	Mostly recurrent UTIs in women
<i>elfC</i>	2601					
<i>elfD</i>	702					
<i>elfG</i>	1072					
<i>hcpA</i>		Hemorrhagic <i>E. coli</i> pilus	Recognized in EHEC strains for the most	Adhesin	Exterior	Mostly recurrent UTIs in women
<i>hcpB</i>						
<i>hcpC</i>						

characteristics for UPEC. The related genes are invaluable treasures for DNA microarray probe designing. Table 1 shows several numbers of chromosomal virulence genes relating to adhesins recognized in a wide range of UPEC pathotypes [3, 27–34]:

As Table 1 shows there are a huge number of adhesins with different genetic characteristics recognized in UPEC strains. The diversity of virulence genes including adhesins enables us to have a vast range of patterns for designing specific, sensitive and flexible DNA microarray probes.

5 Conclusion

E.coli bacteria have divided into two groups of pathogens including Intra-intestinal pathogenic E.coli (InPEC) and Extra-intestinal pathogenic E.coli (ExPEC). The InPEC pathotypes involving Enterohemorrhagic E.coli (EHEC), Enteropathogenic E.coli (EPEC), Enteroinvasive E.coli (EIEC), Enterotoxigenic E.coli (ETEC) etc. encompass different virulence genes and factors. In addition to InPEC, the ExPEC strains including UPEC possess a variety of virulence genes and factors which are contributed to bacterial pathogenicity and adaptation. According to Table 1, there are wide ranges of genes which produce different types of adhesins with different functions and activities. Moreover, the size of genes and their genetic properties are different. However, some of them have huge similarities with each other. So, these genes give us a great opportunity to design an abundance of microarray probes; but, the phylogenetic and structural properties of the genes and their similarities and dissimilarities have a deep influence on the accuracy of the probes. Awareness of the genes characteristics is an essential need for designing sharp, specific, sensitive and flexible microarray probes. Of course, simultaneously the presence of skillful bioinformaticians makes it easier to use proper probe sets and platforms. Best designed microarray probes guarantee an accurate, rapid and reliable diagnosis for a definite treatment.

6 Conflict of Interest

There is no conflict of interest for the authors.

Acknowledgements. We are obliged to Shahr-e-Qods Branch, Islamic Azad University as the financial supporter of this paper for presenting in IWBBIO 2017, Granada, Spain.

We also have special thanks to Profs. Ignacio Rojas and Francisco Ortuño for their kind helps and collaborations.

References

1. Behzadi, P., Behzadi, E.: The microbial agents of urinary tract infections at central laboratory of Dr. Shariati Hospital, Tehran, Iran. *Turkiye Klinikleri J. Med. Sci.* **28**(4), 445–449 (2008)
2. Behzadi, P., Behzadi, E., Ranjbar, R.: Urinary tract infections and *Candida albicans*. *Cent. Eur. J. Urol.* **68**(1), 96–101 (2015)
3. Jahandeh, N., et al.: Uropathogenic *Escherichia coli* virulence genes: invaluable approaches for designing DNA microarray probes. *Cent. Eur. J. Urol.* **68**, 452–458 (2014)
4. Subashchandrabose, S., Mobley, H.L.T.: Virulence and fitness determinants of Uropathogenic *Escherichia coli*. *Microbiology Spectr.* **3**(4) (2015)
5. Matuszewski, M.A., et al.: Uroplakins and their potential applications in urology. *Cent. Eur. J. Urol.* **69**(3), 252–257 (2016)
6. Behzadi, E., Behzadi, P.: The role of toll-like receptors (TLRs) in urinary tract infections (UTIs). *Cent. Eur. J. Urol.* **69**, 404–410 (2016)

7. Torres, A.G.: *Escherichia coli* in the Americas. Springer, Cham (2016)
8. Behzadi, P., Behzadi, E., Ranjbar, R.: IL-12 family cytokines: general characteristics, pathogenic microorganisms, receptors, and signalling pathways. *Acta Microbiol. Immunol. Hung.* **63**(1), 1–25 (2016)
9. Snipen, L.G., Ussery, D.W.: A domain sequence approach to pangenomics: applications to *Escherichia coli*. *F1000Research* **1**, 19 (2013)
10. Mira, A., et al.: The bacterial pan-genome: a new paradigm in microbiology. *Int. Microbiol.* **13**(2), 45–57 (2010)
11. Sharma, A.: Can “Pan-Genomics” demystify the designation of “Bacterial Species”. *SM J. Bioinform. Proteomics* **1**(1), 1001 (2016)
12. Behzadi, P., Behzadi, E., Ranjbar, R.: *Basic Modern Molecular Biology*, 1st edn. Persian Science & Research Publisher, Tehran (2014)
13. Behzadi, P., Behzadi, E.: *Environmental Microbiology*, 1st edn. Niktab Publisher, Tehran (2007)
14. Mira, A., Klasson, L., Andersson, S.G.: Microbial genome evolution: sources of variability. *Curr. Opin. Microbiol.* **5**(5), 506–512 (2002)
15. Genome/*Escherichia coli* (2017). <https://www.ncbi.nlm.nih.gov/genome/?term=Escherichia+coli>. Accessed 10 Feb 2017
16. Gene/*Escherichia coli*: National Center for Biotechnology Information, U.S. National Library of Medicine, Bethesda, USA (2017)
17. Hu, P., et al.: Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. *PLoS Biol.* **7**(4), e1000096 (2009)
18. Cho, B.-K., et al.: The transcription unit architecture of the *Escherichia coli* genome. *Nat. Biotechnol.* **27**(11), 1043–1049 (2009)
19. Behzadi, P., et al.: Microarray long oligo probe designing for *Escherichia coli*: an in-silico DNA marker extraction. *Cent. Eur. J. Urol.* **69**, 105–111 (2016)
20. Behzadi, P., Behzadi, E., Ranjbar, R.: Microarray probe set: biology, bioinformatics and biophysics. *Alban. Med. J.* **2**, 78–83 (2015)
21. Behzadi, P., Behzadi, E., Ranjbar, R.: Microarray data analysis. *Alban. Med. J.* **4**, 84–90 (2014)
22. Ranjbar, R., Behzadi, P., Mammina, C.: Respiratory Tularemia: *Francisella tularensis* and microarray probe designing. *Open Microbiol. J.* **10**, 176–182 (2016)
23. Behzadi, P., Ranjbar, R., Alavian, S.M.: Nucleic acid-based approaches for detection of viral hepatitis. *Jundishapur J. Microbiol.* **8**, e17449 (2015)
24. Behzadi, P., Ranjbar, R.: Microarray long oligo probe designing for Bacteria: an in silico pan-genomic research. *Alban. Med. J.* **2**, 5–11 (2016)
25. Lukjancenko, O., Wassenaar, T.M., Ussery, D.W.: Comparison of 61 sequenced *Escherichia coli* genomes. *Microb. Ecol.* **60**(4), 708–720 (2010)
26. Mao, B.-H., et al.: Identification of *Escherichia coli* genes associated with urinary tract infections. *J. Clin. Microbiol. JCM* **50**(2), 449–456 (2011)
27. MOH Key Laboratory of Systems Biology of Pathogens, Institute of Pathogen Biology, CAMS&PUMC, Beijing, China: Virulence factors of pathogenic bacteria, 28 October 2016. <http://www.mgc.ac.cn/cgi-bin/VFs/compvfs.cgi?Genus=Escherichia>. Accessed 12 Feb 2017
28. Baby, S., Karnaker, V.K., Geetha, R.: Adhesins of Uropathogenic *Escherichia coli* (UPEC). *Int. J. Med. Microbiol. Trop. Dis.* **2**(1), 10–18 (2016)
29. Zalewska, B., et al.: A surface-exposed DraD protein of Uropathogenic *Escherichia coli* bearing Dr fimbriae may be expressed and secreted independently from DraC usher and DraE adhesin. *Microbiology* **151**(7), 2477–2486 (2005)
30. Cordeiro, M.A., et al.: Curli fimbria: an *Escherichia coli* adhesin associated with human cystitis. *Braz. J. Microbiol.* **47**(2), 414–416 (2016)

31. Kwak, M.-J., et al.: Genome sequence of *Escherichia coli* NCCP15653, a group D strain isolated from a diarrhea patient. *Gut Pathog.* **8**(1), 7 (2016)
32. Wurpel, D.J., et al.: Comparative proteomics of Uropathogenic *Escherichia coli* during growth in human urine identify UCA-like (UCL) fimbriae as an adherence factor involved in biofilm formation and binding to uroepithelial cells. *J. Proteomics* **131**, 177–189 (2016)
33. Archer, C.T., et al.: The genome sequence of *E. coli* W (ATCC 9637): comparative genome analysis and an improved genome-scale reconstruction of *E. coli*. *BMC Genom.* **12**(1), 9 (2011)
34. Roos, V., Klemm, P.: Global gene expression profiling of the asymptomatic bacteriuria *Escherichia coli* strain 83972 in the human urinary tract. *Infect. Immun.* **74**(6), 3565–3575 (2006)

Bioinformatics and Biomedical Engineering
5th International Work-Conference, IWBBIO 2017,
Granada, Spain, April 26–28, 2017, Proceedings, Part II
Rojas, I.; Ortuño, F. (Eds.)
2017, XXXVI, 741 p. 275 illus., Softcover
ISBN: 978-3-319-56153-0