Step I: White Paper Application

Application Guidelines

1. *The application should be submitted electronically per requirements via the web site of any of the NIAID Genomic Sequencing Centers for Infectious Diseases. Include all attachments, if any, to the application.*
2. *There are no submission deadlines; white papers can be submitted at anytime.*
3. *GSC personnel at any of the three Centers can assist / guide you in preparing the white paper.*
4. *Investigators can expect to receive a response within 4-6 weeks after submission.*
5. *Upon approval of the white paper, the NIAID Project Officer will assign the project to a NIAID GSC to develop a management plan in conjunction with the participating scientists.*

White Paper Application

Project Title: *E. coli* urinary tract infection: Importance of antimicrobial peptides and bacterial factors of the rectal flora and UTI pathogenic *E. coli*.

Authors:
Karen Leth Nielsen, **M.Sc. Eng. (Biotech.). Statens Serum Institut, Copenhagen**
Niels Frimodt-Møller, **MD DMSc. Statens Serum Institut, Copenhagen**
Paal Skytt Andersen, **PhD. Statens Serum Institut, Copenhagen**
Dr. Michael Feldgarden. The Broad Institute

Primary Investigator Contact:

Name	Karen Leth Nielsen
Position	PhD student
Institution	Statens Serum Institut
Address	Artillerivej 5, Copenhagen S, Denmark
State	
ZIP Code	2300
Telephone	+4532683646
Fax	+4532683231
E-Mail	KTE@ssi.dk

1. Executive Summary *(Please limit to 500 words.)*

Provide an executive summary of the proposal.

E. coli is the main pathogen in urinary tract infection. In most cases the bacteria originate from the patients' own intestinal flora and spreads via the perineum to the vagina, where it contaminates the area around the meatus externa urinae. Depending on well known virulence factors but less understood host factors, the bacteria ascend via the urethra to the bladder, where they infect the bladder epithelium causing UTI. Antimicrobial peptides (AMPs) are produced by epithelial cells and leucocytes; and accessory glands usually coat the mucosal surface of the urinary tract, thereby preventing microorganisms from gaining access. UTI may therefore be caused by failure to produce AMPs or resistance mechanisms against AMPs in the infecting pathogens.

In a clinical study we have collected *E. coli* isolates from the urine and the rectum of 50 female patients with UTI and from the rectum of 50 healthy controls. Concentrations of antimicrobial peptides (AMPs), thought to be critical host immune factors that are involved in keeping the urine sterile, were measured in all urine samples to compare levels at infection with AMP levels in patients and between patients and controls.

We propose sequencing a collection of 192 strains prospectively collected from 50 female patients with *E. coli* UTI and 50 female controls, who have never suffered from UTI. Through this work we will provide data to support further studies in the following areas:

1) Predictors of disease: What is the correlation between *E. coli* clones cultured from urine vs. those originating in the intestinal flora? Is there a difference in urinary and

rectal isolates between UTI patients and controls?
2) Host pathogen interactions: This effort will provide the first study to determine whether there are correlations among infecting strains with concentrations of AMPs and susceptibility among the strains.

2. Justification

Provide a succinct justification for the sequencing or genotyping study by describing the significance of the problem and providing other relevant background information.

This section is a key evaluation criterion.

1. *State the relevance to infectious disease for the organism(s) to be studied; for example the public health significance, model system etc.*

2. *Are there genome data for organisms in the same phylum / class / family / genus? What is the status of other sequencing / genotyping projects on the same organism including current and past projects of the NIAID GSC? Provide information on other characteristics (genome size, GC content, repetitive DNA, pre-existing arrays etc.) relevant to the proposed study. Have analyses been performed on the raw data already generated/published? If additional strains are proposed for a species, please provide a justification for additional strains?*

3. *If analyses have been conducted, briefly describe utility of the new sequencing or genotyping information with an explanation of how the proposed study to generate additional data will advance diagnostics, therapeutics, epidemiology, vaccines, or basic knowledge such as species diversity, evolution, virulence, etc. of the proposed organism to be studied.*

Urinary tract infections are one of the most common bacterial infections in primary care. Acute uncomplicated urinary tract infection, or cystitis, is common among women, accounting for ~7 million office visits annually in the United States (1) and affecting half of women at least once during their lifetime (2). A recent study of sexually active young women found that the incidence of cystitis was 0.5% to 0.7% per year (3). In aggregate, the direct costs of these infections have been estimated at $1 billion yearly in the United States (4).

Furthermore, UTI is the most common nosocomial infection associated with use of foreign bodies such as bladder catheters and invasive urological procedures (11). In approximately 20% of treated uncomplicated UTIs, patients suffer recurrence and further complications, such as pyelonephritis and bacteremia. In fact, *E. coli* accounts for 30% of bacteremias, of which 50% of the cases originate in the urinary tract, carrying a mortality of 15-20%(10).

Antimicrobial peptides (AMPs) are part of the innate immune system and play a role in keeping the urinary tract sterile. AMPs of the urinary tract are composed of Human β-defensin 1 and 2, α-defensins, cathelicidin (LL-37) and hepcidin. HBD-1 is excreted

constitutively from the kidneys and is upregulated approximately 3 times during pyelonephritis (5). Furthermore, *defb-1*$^{-/-}$ mice, without the gene encoding the mouse ortholog of this peptide, have 30% more bacteria in the urine than wild type mice. *Defb-1* has 3 SNPs, which are correlated to expression of the peptide, and these have been correlated to several conditions (asthma, cystic fibrosis, sepsis etc.)(6,7). Cathelicidin, as for HBD-1, expression is increased when bacteria are encountered and *CRAMP*$^{-/-}$ (encoding the mouse ortholog of LL-37) mice have significantly more *E. coli* attached to bladder cells than wild type mice (8). HBD-1 and cathelicidin should therefore be further investigated in correlation to urinary tract infection.

Urinary tract infection (UTI) is primarily caused by *E. coli*, and infection is often caused by *E. coli* from the patients own faecal flora (9). To date, it is not known whether patients with predisposition to urinary tract infection are colonised with more virulent *E. coli* in the faecal flora, as compared to healthy controls, who have never had UTI, or whether there are differences in sensitivity towards AMP, which enable certain *E. coli* to invade the bladder.

A range of studies have identified a number of virulence genes in *E.coli* of importance in causing UTI, such as genes encoding adhesion type 1 adherence to bladder epithelium, type B for adhesion to ureteral and renal pelvic epithelium, biofilm formation, iron uptake mechanisms and toxins like hemolysins and others (12,13). Furthermore, phylogrouping (A, B1, B2, D) has revealed that B2 and D are prevalent among urinary pathogens, as well as in human intestinal flora, and this is related to a higher content of virulence genes among these two phylotypes.

E. coli-associated UTI originates from the patient's own rectal flora, from where it contaminates the perineum and tissue adjoining the meatus externae urinae such as the vagina in women and glans penis in men. Little is known, however, about which virulence- or other mechanisms *E. coli* use to colonise the intestines, the perineum or the vagina. We know little about the relationship between gut colonising *E. coli* and those isolates that can also spread to the urinary tract, or how long this spread takes. Furthermore, the exact mechanism of how the bacteria gain access to the urethra and circumvent the local immune system in order to ascend to the bladder is so far unknown. In contrast, there is now detailed understanding of the pathogenic process in the bladder; how *E. coli* adheres to the bladder mucosa, penetrates the panet cells of the bladder epithelium, produces intracellular bacterial communities, which apopotose and lead to reinfection of additional bladder mucosal cells.

To understand the bacterial factors underlying the pathogenesis in UTI, we must compare *E. coli* isolates that are part of the rectal flora in UTI patients to both the urinary pathogens found in the same patients and to *E. coli* from persons never experiencing UTI. If bacterial factors are important, such a study would reveal significant differences between such strain collections. Previous studies have compared rectal and UTI *E. coli* isolates from the same female host (14,15,16). Prevalence in the rectal flora was one factor that determined if that clone would occur in the urinary tract. Known virulence factors were also identified as promoting UTI. However, one of the strongest indications of UTI was phylogenetic membership (i.e., belonging to Group B2 *E. coli*). This suggests that there are more, as yet unidentified, virulence factors involved in pathogenesis. Not only will genome sequencing

identify UTI-specific genes, but, unlike the previous studies, it will also identify allelic variation that can contribute to UTI (e.g., 17).

The Broad Institute has recently completed sequencing of over eighty five commensal *E. coli*. A major goal of that work was to provide a commensal, non-pathogenic context in order to better understand the genomics of *E. coli* pathogenesis. In addition, there are over forty finished *E. coli* genomes publicly available. Analysis of the proposed UTI and rectal isolates, in conjunction with those commensal isolates, will enable us to identify pathogenesis-specific loci. Other GSCIDs have focused on other pathogenic groups of *E. coli*, such as the O157:H7 serotype. A comparison with those data will further refine our understanding of how genetic variation contributes to specific *E. coli*-related disease syndromes, including UTI.

1. Stamm WE, Hooton TM. Management of urinary tract infections in adults. *NEJM.* 1993;329:1328 –1334.
2. Kunin CM. *Urinary Tract Infections: Detection, Prevention, and Management,* 5th ed. Baltimore: Williams and Wilkins, 1997.
3. Hooton TM, Scholes D, Hughes JP, et al. A prospective study of risk factors for symptomatic urinary tract infection in young women. *NEJM.* 1996;335:468–474.
4. Johnson JR, Stamm WE. Diagnosis and treatment of acute urinary tract infections. *Infect Dis Clin North Am.* 1987;1:773–791.
5. Zasloff, M. 2007. Antimicrobial Peptides, Innate Immunity, and the Normally Sterile Urinary Tract. *J Am Soc Nephrol* 18:2810–2816.
6. Cagliani, R., Fumagalli, M., Riva, S., Pozzoli, U., Comi, G.P., Menozzi, G., Bresolin, N., Sironi, M. 2008. The signature of long-standing balancing selection at the human defensin β-1 promoter.*Genome Biology.* 9: R143-143.11
7. Levy, H., Raby, B.A., Lake, S., Tantisira, K.G., Kwiatkowski, D., Lazarus, R., Silverman, E.K., Richter, B., Klimecki, W.T., Vercelli, D., Maritnez, F.D., Weiss, S.T. 2005. Association of defensin b-1 gene polymorphisms with asthma. *J Allergy Clin Immunol*. 115: 252-258.
8.Chromek, M., Slamová, Z., Bergman, P., Kovács, L., Podracká, L., Ehrén, I., Hökfelt, T., Gudmundsson, G.H., Gallo, R.L., Agerberth, B., Brauner, A. 2006. The antimicrobial peptide cathelicidin protects the urinary tract against invasive bacterial infection. *Nature Medicine*. 12: 636-641.
9. Moreno, E., Andreu, A., Picrau, C., Kuskowski, M.A., Johnson, J.R., Prats, G. 2008. Relationship between *Escherichia coli* Strains Causing Acute Cystitis in Women and the Fecal *E. coli* Population of the Host. Journal og Clinical Microbiology. 46:2529–2534
10. Søgaard, M., Nørgaard, M., Dethlefsen, C., Schønheyder, H.C. 2011. Temporal Changes in the Incidence and 30-Day Mortality associated with Bacteremia in Hospitalized Patients from 1992 through 2006: A Population-based Cohort Study. CID. 52:61-69.
11. Tambyah, P.A, Maki, D.G. 2000. Catheter-associated urinary tract infection is rarely symptomatic: a prospective study of 1,497 catheterized patients.Arch Intern Med. 160:678-82.
12. Johnson, J.R. 1991. Virulence factors in Escherichia coli urinary tract infections. Clin. Microbial. Rev. 4:80-128
13. Mühldorfer, I. and Hacker, J. 1994. Genetic aspects of Eschrichia coli virulence. Microb. Pathog. 16:171-181.
14. Moreno, E., Andreu, A., Pérez, T., Sabaté, M., Johnson, J.R., and Prats. G. 2006. Relationship between *Escherichia coli* strains causing urinary tract infection in women and the dominant faecal flora of the same hosts. Epidemiol. Infect. 134:1015-1025.
15. Moreno, E., Andreu, Pigrau, C., Kuskowski, M.A., M., Johnson, J.R., and Prats. G. 2008. Relationship between *Escherichia coli* strains causing acute cystitis in women and the fecal *E. coli* population of the host. J. Clin. Microbiol. 46:2529-2534.
16. Moreno, E., Johnson, J.R., Pérez, T., Prats, G., Kuskowski, M.A., and Andreu, A. 2009. Structure and urovirulence characteristics of the fecal *Escherichia coli* population among healthy women. Microbes and Infection 11:274-280.
17. Sokurenko, E.V., Feldgarden, M., Trintchina, E., Weissman, S.J., Avagyan, S., Johnson, J.R. and Dykhuizen, D.E. 2004. Selection Footprint in the FimH Adhesin Shows Pathogenicity-Adaptive Niche Differentiation in *Escherichia coli*. Mol. Biol. Evol. 21: 1373-83.

3. Rationale for Strain Selection

4. *Provide the rationale behind the selection of strains and the number of strains proposed in the study. The focus of the program is on potential agents of bioterrorism or organisms responsible for emerging or re-emerging infectious diseases. Non-select agents or non-pathogenic organisms will be considered when they can provide insight into these scientific areas.*

We prospectively studied 50 female patients with *E. coli* UTI and 50 female controls, who have never had UTI. From the UTI patients we collected the infecting strain and a rectal swab from the time of infection; 4 UTI patients still had infections at a later follow up visit ('post-treatment'), and *E. coli* were isolated from these long-term infections (4

isolates). From controls we collected a rectal swab. From all rectal swabs 20 colonies of *E. coli* were typed with RAPD to identify different clones, and one colony belonging to every RAPD type per swab was selected (typically, 1-2 clones per swab).

In total, 248 *E. coli* strains would be sequenced: 56 from urine and 192 from rectal swabs. Isolates in this study:

	Urine isolates baseline	Urine isolates after treatment	Rectal swab
UTI patients (50)	52	4	108
Controls (51)	0	0	84

Concentrations of AMPs will be measured in all urine samples to compare levels at infection with AMP levels in patients after infection and between patients and controls. MIC/IC50 levels of the AMPs will be determined in all clones of *E. coli* isolates.

4a. Approach to Data Production: **Data Generation**

5. *State the data and resources planned to be generated. (e.g draft genome sequences, finished sequence data, SNPs, DNA/protein arrays generation, clone generation etc.)*

We will produce draft genomes sequences for all *E. coli* isolates described in this project. As we wish to identify novel genetic elements that are potentially unique to certain disease states, persistence, and spread, we will generate draft *de novo* genome sequences for all proposed strains using the Illumina platform. Our experience in sequencing numerous *E. coli* strains suggests that a major issue is that many genes of interest, such as those associated with virulence and antibiotic resistance, are associated with transposable elements and other repetitive motifs. To overcome this problem, we will use large jumping library insert sizes (~5-7 kb) that should enable us to scaffold over the repeated elements.

4b. Approach to Data Production: **Data Analysis**

6. *Briefly describe the analysis (value-add) envisioned to be performed subsequently by the community and the potential to develop hypotheses driven proposals given the datasets and resources produced by this work.*

The following represents some of the questions, which can be studied from analysing sequences from the 192 *E. coli* isolates:

1) In how many cases is the same *E. coli* isolate found in urine and feces, (i.e. which proportion of *E. coli* cultured from urine originate in the intestinal flora)?
2) Are there specific clones that repeatedly colonize the urinary tract?
3) How many of the clones found in the intestinal flora have the ability to cause UTI?

4) Is there a difference in *E. coli* clones from feces between UTI patients and controls?
5) Is there a difference in *E. coli* isolates found in controls vs. patients, i.e. would it be possible by a rectal swab culture to predict, whether a person has an UTI risk?
6) If there are differences in AMP-susceptibility among the strains, does this correlate to specific genes?
7) Is it possible to discern between *E. coli* clones from the intestinal flora regarding colonization or pathogenic ability?
8) Are there antibiotic resistance genes present in isolates that were not found by phenotypic methods?
9) What roles do plasmid composition and presence play in answering the previous questions?

Answering these questions will enable us to better define the relative contributions of bacterial genotype and host factors.

5. Community Support and Collaborator Roles:

7. *Provide evidence of the relevant scientific community's size and depth of interest in the proposed sequencing or genotyping data for this organism or group of organisms. Please provide specific examples.*

To our knowledge, a similar sequencing analysis of pathogenic strains from infections, from the same patients' rectal flora and from the rectal flora of controls has never been performed for *E. coli*. This has crucial importance for the understanding of the relationship between carrier strains and pathogenic strains, and for the pathogenesis of UTI, as well as to identify potential reservoirs of pathogenesis.

8. *List all project collaborators and their roles in the project*

Karen Leth Nielsen (Dept. for Microbiological Surveillance and Research, Statens Serum Institut, Copenhagen, Denmark). Karen Leth Nielsen is performing the PhD-project regarding antimicrobial peptides and urinary tract infection. She will perform analysis and interpretation on the *E. coli* genomes, along with the other analysis of this project.

Paal Skytt Andersen (Dept. for Microbiological Surveillance and Research, Statens Serum Institut, Copenhagen, Denmark). Paal Skytt Andersen is supervisor on Karen Leth Nielsen's PhD-project, and has experience in whole genome sequencing, annotation, alignment and will assist and supervise in the analysis of the genomes.

Niels Frimodt-Møller (Dept. for Microbiological Surveillance and Research, Statens Serum Institut, Copenhagen, Denmark). Niels Frimodt-Møller is supervisor for Karen Leth Nielsen on her PhD-project and is the major developer of this project. He will supervise Karen during the analysis and assist with interpretation of results.

Michael Feldgarden (The Broad Institute). Dr. Feldgarden has organized several large-scale, GSC-funded population genomics projects, including one focused on ~100 *E. coli*

commensal genomes. He is also involved with the analysis of the data from these projects, and will oversee the Broad Institute's contribution to data analysis.

James R. Johnson (VA Medical Center/University of Minnesota). Dr. Johnson has published extensively on extraintestinal pathogenic *E. coli*, and has conducted numerous studies comparing rectal and urinary tract *E. coli*. He assisted in development of this white paper.

Thomas O'Brien (Brigham & Women's Hospital). Dr. O'Brien has had extensive experience in infectious disease, including publishing one of the seminal studies demonstrating epidemic resistance plasmid spread. He served as an advisor on the Broad Institute's previous *E. coli* sequencing project, and he assisted in the development of this white paper.

9. *List availability of other funding sources for the project.*

Danish National Research Foundation (grant: DANCARD) and EU-DG Research FP 7 programme "PAR" have provided necessary funding of salary and materials to complete collection and DNA preparation of strains used in this study.

6. Availability & Information of Strains:

10. *Indicate availability of relevant laboratory strains and clinical isolates. Are the strains/isolates of interest retrospectively collected, prepared and ready to ship?*

Note: If samples are prospectively prepared the GSC can provide protocols and recommendation based on the Centers past experiences. The samples must however meet minimum quality standards as established by the Center for the optimal technology platform (sequencing/ genotyping) to be used in the study.

All isolates are (prospectively) collected and ready to ship

11. *Attach relevant information, if available in an excel spreadsheet for multiple samples: e.g.*
    - *Name*
    - *Identifier*
    - *Material type (DNA/RNA/Strain)*
    - *Genus*
    - *Species*
    - *Specimen / Strain*
    - *Isolation source*
    - *Isolated from*
    - *Select agent status*
    - *International permit requirement*
    - *BEIR/ATCC repository accession number*

- *Other public repository location*
- *Other public repository identifier*
- *Sample provider's name*
- *Sample provider's contact*

12. *What supporting metadata and clinical data have been collected or are planned on being collected that could be made available for community use?*

Site of isolation (UTI, rectal), gender, date of collection

7. Compliance Requirements:

7a. Review NIAID's Reagent, Data & Software Release Policy:

NIAID supports rapid data and reagent release to the scientific community for all sequencing and genotyping projects funded by NIAID GSC. It is expected that projects will adhere to the data and reagent release policy described in the following web sites.

http://www3.niaid.nih.gov/research/resources/mscs/data.htm

http://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-013.html

<Each Center to include their website that describes/points to the guidelines>

Once a white paper project is approved, NIAID GSC will develop with the collaborators a detailed data and reagent release plan to be reviewed and approved by NIAID.

Accept x☒ Decline ☐

7b. Public Access to Reagents, Data, Software and Other Materials:

13. *State plans for deposit of starting materials as well as resulting reagents, resources, and datasets in NIAID approved repositories. Sequencing projects will not begin until the strain is deposited into NIAID funded BEI repository (http://www.beiresources.org/). This includes web based forms are completed by the collaborator and received by the NIAID BEI (http://www.beiresources.org/).*

All strains will be deposited at Broad and further available for research purposes.
All data generated will be published.

7c. Research Compliance Requirements

Upon project approval, NIAID review of relevant IRB/IACUC documentation is required prior to commencement of work. Please contact the GSC Principal Investigator(s) to ensure necessary documentation are filed for / made available for timely start of the project.

Investigator Signature:

Investigator Name: **Date:**

Blank Last Page