**Background:**

Uropathogenic *Escherichia coli* (UPEC) is known to inhabit the gut, where it lives as a commensal1, but is also capable of causing disease in the bladder during a urinary tract infection(UTI). As yet, no investigation has defined the community structure of the gut microbiota or the relative abundance of UPEC in the gut during a UTI episode. Genes that have been shown to be important for virulence in the bladder, such as *hlyA*, *papC*, and *iutA*, have also been correlated with enhanced persistence over time and a reduced rate of turnover in the gut2,3. Additionally, dominance within the subpopulation of *E. coli* in the gut has been correlated with the presence of known urovirulence genes4-6. With these factors in mind, ***I hypothesize that there is a bloom of uropathogenic E. coli in the gut that coincides with the bloom of uropathogenic E. coli in the bladder during the onset of an acute UTI.*** To test this hypothesis, I will compare the community structure of the gut microbiota during UTI and after recovery using a combination of metagenomic sequencing and targeted qPCR. Together, these methods will allow for an accurate measurement of the representation of UPEC in the gut microbiota during the onset of a UTI. Because the gut can act as a reservoir for UPEC7,8, understanding the community structure of this opportunistic pathogen in its reservoir will aid our understanding of UTI susceptibility and progression of the disease.

**Sub-aim1: Compare the community structure of the gut microbiota during acute UTI and after recovery.**

*Rationale:* While the human microbiome is generally stable through time9,10, changes in gut community structure between healthy and disease states have been identified11. Changes in the periutheral microbiota have also been found, as the prevalence of *E. coli* increases in the days preceding the onset of a UTI12. This is concordant with the rectal-perineal-urethral hypothesis, which states that the gut contains a reservoir of UPEC, which escape the gut and colonize the periurethral area before ascending into the bladder to cause a UTI8. While a bloom of growth in the gut would increase transmission to the periurethral area, no study has investigated changes in the relative abundance of UPEC in the gut microbiota during the onset of a UTI episode. However, at least one population susceptible to UTIs, pregnant women13, undergo shifts in the ecology of their gut microbiome during the course of their pregnancy, including an expansion of proteobacteria14. Thus, ***I hypothesize that there is a greater representation of E. coli in the gut microbiota of a patient at the onset of a UTI than at times when the patient is healthy.*** To test this hypothesis, I will measure the relative abundance of Enterobacteriaceae, the Family taxon that contains *E. coli*, in fecal samples collected from patients during UTI and after their recovery. If overgrowth of the gut reservoir of UPEC coincides with the onset of a UTI, then the abundance of Enterobaceriaceae will be higher in the patient during UTI than when the patient is healthy. This information will be very useful in determining if there is a reservoir for *E. coli* in the gut, which may impact UTI susceptibility and disease progression.

*Experimental methods:* In collaboration with Case Western University and their clinical facilities, women aged 18-41 years who present with symptoms of an uncomplicated, lower UTI will be open to enroll in a new cohort if they pass the entry requirements, which reduce the chance of enrolling patients with anatomical or functional abnormalities. Patients will supply three fecal samples, one at the time of entry, before the administration of antibiotics, one after completing the antibiotic regimen at 14 days after enrollment, and one at 28 days after enrollment. These samples will be labeled as UTI, Treated, and Recovered, respectively (Figure 2a), and will be subjected to further testing, as outlined (Figure 2b). Fifty patients will be enrolled with the expectation that 30 will submit all the samples, be free from recurrent UTIs during the study, and have confirmed cases of UTI caused by UPEC. A cohort of 30 patients will give the study the power to detect an effect size of 1.40 (Figure 2c).

Genomic DNA from the fecal samples collected from cohort will be extracted and used as template 16S rRNA gene using 454 pyrosequencing. Sequence data from this will then be used to estimate the community structure of the gut microbiota, as has been done previously9. Binning of the amplicons into phylotypes using the QIIME software package will allow for the representation of those phylotypes in the gut microbiota to be measured at the Family level15. Statistical analysis will include application of the Shannon diversity index to identify intra- and inter-host differences in community structure, principle component analysis (PCA) to measure the tendency of the samples to cluster to their sample groups, and Mann-Whitney non-parametric t-tests to identify changes in the representation of Enterobacteriaceae during the UTI and after recovery.

Antibiotic treatment has been shown to dramatically effect the composition of the gut microbiota. While the majority of taxons regenerate after 28 days

16-20, some genera may take months to reappear21. This will reduce the overall species richness of the gut microbiota, and may artificially reduce the divergence between the UTI and Recovered groups. As an additional control against this error, data on the gut community structure of healthy adults available from the Human Microbiome Project (HMP) will be included in these analyses in order to identify differences between healthy microbiomes and microbiomes that may be altered after antibiotic treatment, as has been done in other studies14.

*Expected results:* Analysis using the Shannon diversity index will measure the differences between the sample groups and will show that the samples from the different groups will show more inter-group diversity than intra-group diversity, indicating that the groups have shared features of their community structure that differ between the groups. This analysis will be supported by PCA which will show that communities from the Recovered group will cluster with the data from the HMP, while the UTI group and the Treated group will each cluster separately. Finally, the Mann-Whitney t-test will show that the UTI group will have a statistically significant higher representation of Enterobacteriaceae than the Recovered group, which is an indication that gut *E. coli* carriage is higher during UTI than after recovery. Taken together, these data will show that the gut microbiota during UTI is in an altered state distinct from a healthy state, and that the abundance of Enterobacteriaceae is higher in this altered state than in the healthy state.

*Anticipated challenges:* Recurrence of a UTI episode within 6-12 months after an initial UTI caused by UPEC occurs in approximately 25-30% of cases. The chance of recurrence is greatest within the first 3 months following the initial UTI, and is often caused by the same UPEC strain as the one that caused the initial UTI22. The chance of another UTI episode caused by the same strain drops after 6 months. This indicates that there may be a short-lived reservoir of UPEC in the patient that survives antibiotic treatment and immune system clearance. This reservoir may be located in the gut, which would complicate analysis of the gut carriage of UPEC during an acute UTI episode. As a result, any patients who self-reported another UTI episode within the 6 months following enrollment in the cohort will be excluded from the final analyses. This self-reporting is cost-effective, non-invasive, and will reduce the confounding influence of recurrent UTIs on this analysis.

It is possible that there are members of the Enterobacteriaceae that could be have increased representation in the gut at the onset of a UTI episode due to host factors, such as immunodeficiency. If there are large discrepancies between the estimated abundance of Enterobacteriaceae and *E. coli* specific markers identified in sub-aim 2, then the relative representation of *E. coli* in the gut microbiome will be re-estimated using qPCR targeting *E. coli* specific housekeeping genes and comparing those results to the results from a qPCR targeting a control region conserved in the family Enterobacteriaceae. This will identify the proportion of *E. coli* in the fecal sample relative to the abudance of the Enterobacteriaceae family, which can then be used to calculate the abundance of *E. coli* relative to the total gut microbiota.

**Sub-aim 2: Compare the relative abundance of urovirulence genes in the gut microbiome during acute UTI and after recovery.**

*Rationale:* Analysis of the *E. coli* population structure in the gut has shown that dominance in the gut is highly correlated with the presence of urovirulence genes during a UTI episode4,5; however, the relative representation of urovirulence genes in the gut microbiome during UTI and in a healthy state within the same patient has yet to be compared. Given the connection between urovirulence factors and gut fitness2,3, ***I hypothesize that there is a greater relative abundance of known urovirulence genes in the gut microbiome of a patient at the onset of a UTI than when the person is healthy.*** While a bloom in the population of UPEC in the gut would increase the relative abundance of urovirulence genes, it is also possible that an increase in the relative abundance of *E. coli* in the gut microbiota is not needed for an increase in the relative abundance of UPEC urovirulence genes because an increased number of urovirulence genes in the dominant strain of *E. coli* is correlated with reduced *E. coli* strain diversity in the gut4,6. In this situation, a single strain of UPEC carrying many urovirulence genes may outcompete other *E. coli* species in the gut, resulting in their local extinction, without changing the total abundance of *E. coli* in the gut. In this scenario, the relative abundance of urovirulence genes may increase without affecting the relative abundance of *E. coli* in the gut microbiota. No single set of urovirulence genes necessary and sufficient for pathogenesis has been defined23, therefore, many urovirulence genes must be assayed to measure the prevalence and abundance of UPEC. Thus, in order to measure the relative abundance of urovirulence genes in the gut microbiome, I will use a series of quantitative polymerase chain reactions (qPCRs) targeting known urovirulence genes and compare the results to results from qPCRs targeting housekeeping genes. If there is an increase in the relative abundance of urovirulence genes in the gut microbiome at the onset of a UTI, then the abundance of those genes will be higher at the onset of UTI relative to housekeeping genes common to all microbial taxa. This information is critical to understanding the population dynamics of UPEC in the gut, as the current definitions of UPEC rely extensively on the presence or absence of known urovirulence factors{Russo:2000vr, Kaper:2004bm}.

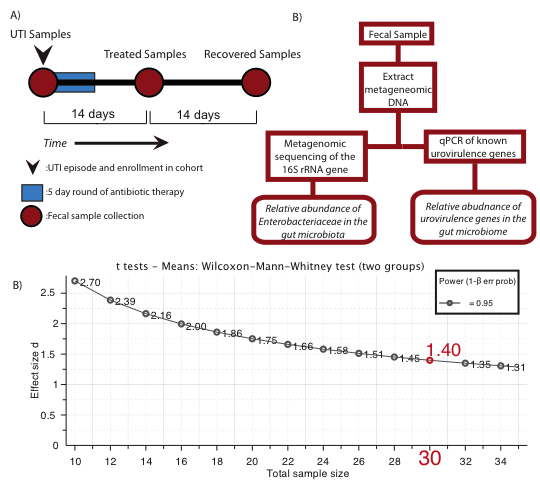
*Experimental methods:* Genomic DNA extracted in sub-aim 1 will be used as template for quantitative polymerase chain reactions (qPCRs) targeting known urovirulence genes from gene clusters listed in Table 1. Additional qPCRs will be performed that target a conserved region of the 16S rRNA gene common to all bacteria as well as a region of the single copy housekeeping gene, rpoB, which is divergent in E. coli.

Comparison of the qPCR results from the urovirulence genes to the qPCR results from the 16S rRNA and *rpoB* genes can be used to estimate the relative abundance of the urovirulence genes to the total gut microbiota population and the E. coli sub-population, respectively. Mann-Whitney non-parametric t-tests will be used to identify statistically significant differences between sample groups in uroviruelnce gene abundance relative to the total microbiota and to E. coli specifically.

*Expected results:* In healthy adults, *E. coli* are found at an abundance of >0.1% in 15% of microbiomes and are detectable in a total of 61% of fecal samples9. Given the prevalence of urovirulence genes in dominant strains of *E. coli* in the gut4,6,24, I expect that Recovered samples will have an abundance of urovirulence genes relative to the *E. coli rpoB* gene of 4 to 95%, depending on the urovirulence gene. Thus, given the low representation of *E. coli* in the gut, I expect the abundance urovirulence genes to be very low compared to the abundance of the bacterial 16S rRNA genes in the gut microbiota. However, in samples collected at the onset of a UTI episode, I expect that the proportion of urovirulence genes relative to both the *E. coli* *rpoB* gene and the bacterial 16S gene to be significantly higher in the UTI sample set as compared to the Recovered set. Taken together, these data will define the reservoir of UPEC urovirulence genes that exist in the gut microbiome during the onset of a UTI episode. If this reservoir of virulence genes is higher at the onset of a UTI than when the patient is healthy, then this is a strong indication

*Anticipated challenges*: Given the low abundance of *E. coli* in the healthy genome, standard qPCR may not be able to detect urovirulence genes accurately given technical limitations, such as off-target priming and analog output. Digital PCR{Vogelstein:1999ve} has been used in the place of qPCR to detect and quantify the presence of pathogens, such as Human Immunodeficiency Virus due to its lower limit of detection and digital output. If qPCR fails to identify the relative abundance of targeted urovirulence genes, then Digital PCR can be used to determine if the abundance of urovirulence genes or confirm their absence. This information can be substituted for the relative abundance of known urovirulence genes, and instead be used to quantify the number of copies of known urovirulence genes per copy of the *E. coli rpoB* gene or 16S rRNA gene. Both qPCR and Digital PCR will enable the comparison of urovirulence gene abundance at the onset of UTI to the abundance at after recovery, providing a necessary insight into the repertoire of genes available to UPEC.

**Conclusions:** Description of the gut community structure and abundance of urovirulence genes at the onset of acute UTI will aid in our understanding of the reservoir of UPEC that exists in the gut. Population dynamics in this reservoir may mediate susceptibility to acute UTI; however, relatively little attention has been paid to this sub-population of UPEC. The identification of a bloom in the gut population of *E. coli* and an increase in the representation of urovirulence genes in the gut microbiome that precedes the onset of a UPEC UTI would be significant support to the idea hypothesis that urovirulence genes also mediate gut fitness. This hypothesis, if supported, is a significant insight into the evolutionary pressures that shape this opportunistic pathogen and will aid our understanding of how virulence has evolved in UPEC – an important step in learning to control the diseases it causes.



**Figure 2. A)** Diagram of the sample collection procedure. Three samples will be collected, one at enrollment, one at 14 days post enrollment, and one at 28 days post enrollment. A standard five day course of trimethorpim-sulfa will be given between the first two collections to treat the urinary tract infection (UTI). **B)** Outline of experimental design for sub-aims. **C)** Power analysis curve constructed using G\*Power. The power, or one minus the rate of a Type II error, of a Mann-Whitney t-test was plotted using the total sample size and effect size, d, to calculate the effect size of a cohort consisting of 30 patients.