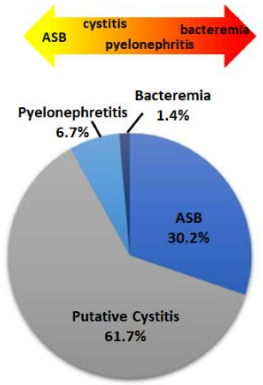
Developing a Computational Pipeline for Characterization of Uropathogenic *Escherichia coli*.

GRACE MORALES[1], BENJAMIN ABELSONS[2],GERALD VAN HORN[1], DOUGLASS CLAYTON[2], MARIA HADJIFRANGISKOU[1,2,3], JONATHAN SCHMITZ[1,2,3]
DEPARTMENTS OF: [1]PATHOLOGY, MICROBIOLOGY & IMMUNOLOGY; [2]UROLOGY AND [3]VANDERBILT INSTITUTE FOR INFECTION, IMMUNOLOGY, AND INFLAMMATION, VANDERBILT UNIVERSITY MEDICAL CENTER
NASHVILLE, TN, USA.

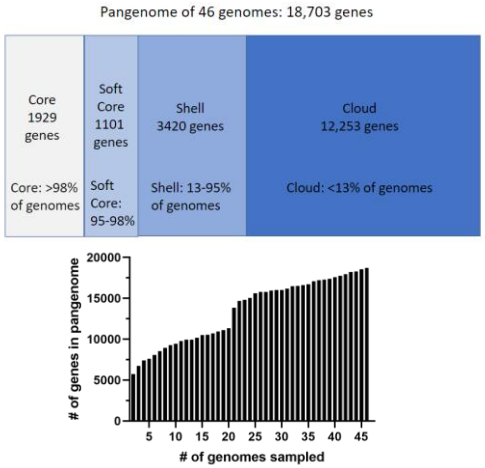
BACKGROUND

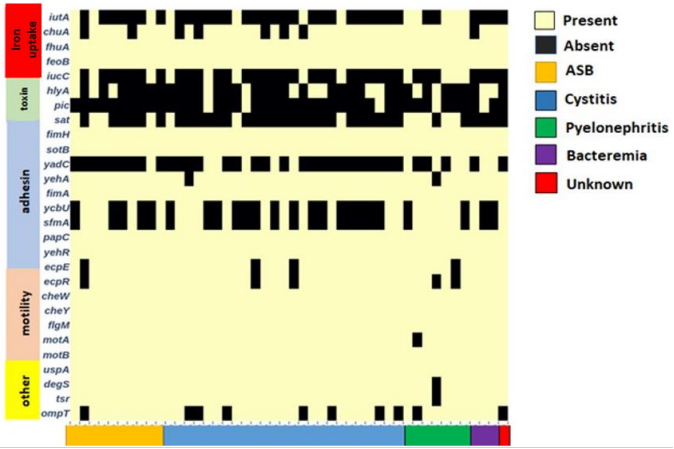


Bacteriuria is a spectrum of clinical presentations. Uropathogenic *E. coli* (UPEC) is the most prevalent cause of urinary tract infections. UPEC are very diverse, and there is no traditional molecular marker to differentiate symptomatic UPEC from asymptomatically colonizing *E. coli*.

RESULTS

UPEC exhibit an open pangenome, with most of the diversity coming from the accessory genome. Percentages indicate amount of genomes that had indicated genes, i.e: 1929 genes in the core, with 98% of genomes having those genes.

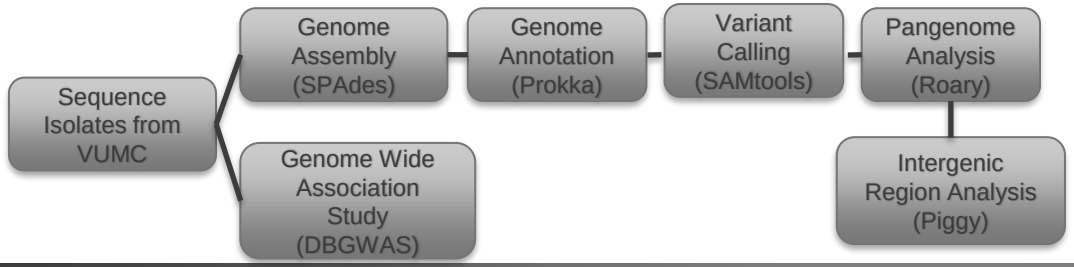




A heat map representative of 46 UPEC genomes showing the diverse nature how UPEC carry virulence factors. Clinical outcome (x-axis) does not correlate with factor carriage, showing a discrete gene presence might not be feasible to differentiate UPEC.

METHODS

A workflow of the developing pipeline is shown below. Briefly, urinary isolates will be sequenced, then their genomes assembled using SPAdes software. The coding regions will be annotated using Prokka, and variants will be identified using SAMtools. Pangenome analysis will be done using Roary. Files from Roary will also be used by the software Piggy to analyze the intergenic regions. Finally, a GWAS will be preformed to statistically cross correlate bacterial genome features to patient clinical outcomes.



CONCLUSIONS

The current pipeline development allows for clinical isolates to be characterized in a high-throughput manner that includes both coding and non-coding regions. By characterizing both coding and non-coding regions, we might be able to identify an alternative genetic signature associated with asymptomatic isolates or symptomatic UPEC. This pipeline is innovative in its use of repurposing of large-scale clinical data, its use of a microbial GWAS linking pathogens with host response, and its goal of characterizing the UPEC pathotype. The pipeline has future implications for other clinical isolates and translational discovery.

NIH/NIDDK: 1P20DK123967-01 * NIH T32: 5T32AI112541-05