Heather Price

Gen Data

Assembly Report

*Escherichia coli* Strain 131 Assembly.

**Introduction:**

While *Escherichia coli* is a native gut bacterium that is generally harmless to the organism that it inhabits, this coliform bacterium can mutate into different strains of itself, producing new colonies which are pathogenic or harmful to the host in question. The strain that was examined was a mutated form of *E. coli* which has proven to be multi-drug resistant, in that it produces a form of beta-lactamase which blocks cell wall inhibition of penicillin like drugs (1,2). The need to learn more about the genome of these mutant strains of bacteria as well as treatment options comes from the variety of infections that can plague the host organism. In the case of humans, this can be simply gastrointestinal upset to sepsis (1).

Even with current technology and social hygiene standards, there are still many areas which lack the basic necessities to combat pathogenic microbes on the most basic levels. As the pathogens in question, such as the mutated strain of *E. coli* can travel via vectors such as flies (2) and other pests, it is an important and primary concern for the safety of the population in general to perform specimen collection and genome assembly to ascertain any available information on the mutated strains.

**Methods:**

Data would be collected via vector capture and cultivation of colony specimens (2) performed in accordance with lab set protocols by the primary investigator. The samples should be collected and processed to produce high concentrations of bacteria to be analyzed so that data of the pathogen genome can be collected. Raw data was processed into a FASTQ file from an Illumina seq protocol. The files ascertained were single end, or single read sequencing, was provided in the FASTQ file to be analyzed and assembled. The original file was provided through the #curl command as a .tgz file compression type and was unzipped through a bash shell script using the Spruce Knob cluster with the line of code in Appendix: 1. Inspection of the given FASTQ file was performed with the line of code in Appendix: 2.

Several attempts were made at assembling the given single end read first using Velvet assembly with variations of the code given in Appendix; 3, as well as Abyss in Appendix; 4. Different parameters were set for the combination of nodes and processors, all combinations yielded a total of 36 giga bytes of dedicated memory. Returned errors included running out of memory to proceed with the assembly, as well as errors noting a non-zero return value that terminated the assembly on the cluster side.

**Discussion:**

Had the given assembly work, the completed genome would have been submitted to the RAST (Rapid Annotation using Subsystem Technology) server to compare to known and annotated genomes. As the bacteria in question is pathogenic in nature, areas of interest to explore would be genetic code corresponding to toxin production, such as genetic expression of shiga toxin in the classification of either STX1 or STX2(3) which produces gastrointestinal distress such as diarrhea, or cytholethal distending toxin, which lead to cell apoptosis (4). Another area of interest would be production of any enzymatic activity which may interfere with or block the mechanisms of action in various used drugs for treating the infections caused by these bacteria, such as beta-lactamase production as stated previously. Both areas are primarily of interest for the identification of symptomatic patients as well as exploring treatment options in a given patient.

**Appendix:**

1. Code used for unzipping FASTQC files:
   1. tar -zxvf SRR6982909\_1.fastq.tgz
2. Run and inspect for fastqc file code:
   1. fastqc SRR6982909\_1.fastq
3. Code used to attempt assembly; Velvet:
   1. velveth assembly\_31\_hp 31 -fastq -short SRR6982909\_1.fastq
   2. velvetg assembly\_31\_hp -cov\_cutoff 4 -min\_contig\_lgth 50
   3. velveth assembly\_23\_hp 23 -fastq -short SRR6982909\_1.fastq
   4. velvetg assembly\_23\_hp -cov\_cutoff 4 -min\_contig\_lgth 50
   5. velveth assembly\_manyK 33,49 -fastq -short SRR6982909\_1.fastq
   6. velvetg assembly\_manyK -cov\_cutoff 4 -min\_contig\_lgth 50
4. Code used to attempt assembly; Abyss:
   1. abyss-pe np=6 k=57 se=SRR6982909\_1.fastq name=Assembly\_23take1 "unitigs"

**References:**

1. Alizade H. (2018). Escherichia coli in Iran: An Overview of Antibiotic Resistance: A Review Article. Iranian journal of public health, 47(1), 1–12.

2. Heiden, S.E., Kurz, M.S., Bohnert, J., Bayingana, C., Ndoli, J.M., Sendegeya, A., Gahutu, J.B., Eger, E., Mockenhaupt, F.P., Schaufler, K. (2020) Flies from a tertiary hospital in Rwanda carry multidrug-resistant Gram-negative pathogens including extended-spectrum beta-lactamase-producing *E. coli* sequence type 131. *Antimicrob Resist Infect Control* **9,**34 (2020). https://doi.org/10.1186/s13756-020-0696-y

3. Melton-Celsa A. R. (2014). Shiga Toxin (Stx) Classification, Structure, and Function. Microbiology spectrum, 2(4), 10.1128/microbiolspec.EHEC-0024-2013. <https://doi.org/10.1128/microbiolspec.EHEC-0024-2013>

4. Croxen, M. A., Law, R. J., Scholz, R., Keeney, K. M., Wlodarska, M., & Finlay, B. B. (2013). Recent advances in understanding enteric pathogenic Escherichia coli. Clinical microbiology reviews, 26(4), 822–880. https://doi.org/10.1128/CMR.00022-13