Assemble your genome using ABySS and SPAdes.

**1) From the ABySS output, create a table for the unitigs, contigs, and scaffolds with the number of each, N50 for each, and predicted genome length.**



**2)**[**https://github.com/bcgsc/abyssLinks to an external site.**](https://github.com/bcgsc/abyss)**This is the link to the documentation for ABySS. In your own words, please summarize the function of each of the commands (e.g., abyss-pe, k, B, etc) that you included in your code.**

# Abyss on paired end (pe) reads, name the files ‘microbe’, Kamer 96 base pair no bigger than 2G on sample genome files

abyss-pe name=microbe k=96 B=2G in='Reads1.fastq.gz Reads2.fastq.gz'

**3) Using either output, perform a BLAST search to identify your species. Write your species name here:** *Klebsiella pneumoniae*

BLAST Results: Klebsiella pneumoniae strain 31285 chromosome, complete genome

**4) Perform quality assessment using QUAST. You need find a reference genome and reference annotation to upload to QUAST for the best quality check. Which assembler gave you the higher quality output? How do you know?**

N50 is indicative of the assembly’s quality; the higher the value, the more quality the assembly. SPAdes has the higher quality output because the N50 output is 193741; whereas ABySS had at most 160025.

**5) Describe what BUSCO is used for. What were the BUSCO values for your assembly?**

BUSCO is used to quantify the genome assembly completeness (%) and partial completeness (%). Within SPAdes, the complete BUSCO was 97.97%, while the ABySS complete BUSCO was 97.30%. On the other hand, both the SPAdes and ABySS partial BUSCO was 0.00%.

**6) Perform a genome annotation using Prokka. Find 3 of the 5 genes/features in your results file and create a table of those results: recA, gyrA, 16S rRNA, rpsB, dnaA.**



 



**7)**[**https://github.com/tseemann/prokkaLinks to an external site.**](https://github.com/tseemann/prokka)**Here is the documentation for prokka. In your own words, what is the function of each of the commands in your line of code?**

# run prokka - output directory name of file ‘prokkaanotation’ –-look for prefix ‘microbe’ in the folder AbyssOutput, the name of the file is microbe-8.fa

prokka -outdir prokkaannotation --prefix microbe AbyssOutput/microbe-8.fa

**8) What is the function of the genes/features you chose?**

The recA gene expresses the recA protein, which is involved in the maintenance and repair of DNA.

16S rRNA is involved in the synthesis of the 16S Subunit of the ribosomal protein

The rpsB gene expresses the 30S ribosomal protein S2, which discriminates against aminoacyl tRNAs that do not match the codon of mRNA, ensuring accuracy during translation.

**9) Find those same genes/features in your RAST annotation. What information did you learn about them from RAST?**

RecA – To prevent the deleterious effects of recA overexpression, recX is expressed to regulate the protein. A screenshot of a computer

Description automatically generated

16S rRNA – there’s an enzyme responsible for transferring a methyl group in the presence of cytosine.



rpsB – From the Prokka annotation, we saw that the rpsB gene is 726 bp long; however, on RAST, I was unable to find a corresponding gene. However, I did find a gene that was <726 that is involved in Translation such as rpsB is.

A screenshot of a computer

Description automatically generated

10) Upload the folder of this information to your GitHub in your Bioinfomatics Repository. Please share the link to your repository.

<https://github.com/jmande1/bioinformatics>