1. Summarize the difference in outputs between ABySS and SPAdes

SPAdes and ABySS both assembled the genome of Salmonella successfully. However, the ABySS output was much more complicated to interpret than the SPAdes output. There were fifty-one items in my folder once the assembly had finished. Knowing which ones to use in further analysis is essential. In contrast, SPAdes only gave four files that were very easy to use. I think the interface of SPAdes is more user friendly than the interface of ABySS.

1. Write a brief summary of the important output metrics from QUAST

Some of the important metrics from QUAST include the genome completeness, L50, and N50 values. The genome completeness uses BUSCO (benchmarking universal single copy orthologs) which looks for genes that are common in all members of the taxon examined. By comparing the assembled genes to this set of common genes, BUSCO can determine the completeness of the assembled genome. The L50 is the number of contigs/scaffolds needed to have approximately 50% of the genome. A low L50 number corresponds to a higher quality genome because that means there are not a lot of little pieces that could be arranged in different ways. The N50 number means that 50% of contigs/scaffolds are this long. A higher N50 number corresponds to a higher quality genome, for similar reasons because it means there are a lot of big pieces, not a large quantity of small fragments. All fo these values from QUAST are important in determining the quality of an assembled genome.

Below is a picture of a pathway using the RAST Seed Viewer:

A screenshot of a computer

Description automatically generated