

CS425
ASSIGNMENT 5 (DUE APRIL 20, 2017)
SEQUENCE ASSEMBLY

Computer Science Department
Colorado State University

April 13, 2017

Genome Assembly Hands-on Experience [100 pts].

Download Illumina sequencing reads from the course website and assemble them using Velvet. You need to download Velvet from <https://www.ebi.ac.uk/~zerbino/velvet/> and compile with `MAXKMERLENGTH=96`, i.e. by `'make MAXKMERLENGTH=96'`.

1. **Assembly with multiple k -mer sizes [50 pts].** Assemble the input reads using 'velveth' to build the hash table and then 'velvetg' to construct, condense, and correct the de Bruijn graph and output contigs. Assemble the same dataset in single-end mode with different values of $k = 11, 19, 33, 55, 75, 95$ and the default coverage cutoff. Use the provided script on the course website to report N50 and total size of each assembly. Draw two plots: (i) N50 (y axis) and (ii) total assembly size (y axis) in terms of k (x axis). What do you learn from those plots?
2. **Assembly with multiple cutoffs [50 pts].** Now, assemble the input dataset above in single-end mode with $k = 55$ and different values of coverage cutoff. Pick a range of cutoffs that spans the entire spectrum of 'almost no error removal' to 'removal of almost all useful information'. Use the provided script on the course website to report N50 and total size of each assembly and draw two plots of N50 (y axis) and total assembly size (y axis) in terms of the cutoff (x axis). What do you learn from those plots?

Upload a report including the results of each task onto Canvas.