Bioinformatics Computing
CSE40532/60532
Homework #4

Homework problems: (due 10/31)

1. We will reuse the two Anthrax strains from GenBank, the gold standard "Ames ancestor" that is virulent (NC_007530) and the non-virulent lab strain "Ames" (NC_003997). There will also be sample data for assembly (see below).

2. Visit the MUMmer (http://mummer.sourceforge.net/) and AMOS (http://amos.sourceforge.net) websites.

3. Read main pages to get a feel for the project, the players, and the goals. Specifically, focus on the sections/help on "nucmer" and on genome alignments in general for the MUMmer part of this homework.

4. Download and install the MUMmer package and AMOS 3.1.0 (click "Download" on the quick links). For this homework, you do not need to install Qt required for some of the visualization tools. Minimus will work without it.

5. Run nucmer on the two strains from #1 using default parameters and with the "Ames ancestor" as the reference. Place the resulting ".delta" file in your drop box (5 points).

6. Summarize SNPs and indels between these two strains using the "show-snps" utility with the "-C" option. Save the ".snps" file in your drop box (5 points) HINT: You may want to also run mummerplot with the --filter option to help with the rest of this assignment or the more comprehensive "dnadiff".

7. Submit a brief write up in your drop box summarizing the results. Hypothesize what the differences may be between the strains? Are there a lot of SNPs? Potential structural differences? (10 points)

8. Read the minimus documentation, and look at the example projects in amos-3.1.0/test/minimus. Get a basic idea if the input, what the bank is, and perform a test assembly (influenza). Submit contigs for this assembly in your dropbox (5 points)

9. Download the test dataset from the course website, and assemble it using minimus. Submit the resulting contigs in your dropbox (5 points)

10. Compare your minimus assembly to the reference available from the course webpage using nucmer. Place the resulting delta file for the bacteria genome in your dropbox (5 points).

11. Provide a brief summary of the results (how many contigs, average contig size, etc.) for grading and do the same sort of comparisons as #7 (show-coords, show-snps, etc). Do you think these are good or bad assemblies and why? Submit in your dropbox this summary and any relevant supporting files (10 points)