Biol5705
Module: Gene Sequence Analysis
Assignment 2
Email results to morgan.g.i.langille@gmail.com by start of class next week.

PSI-BLAST

- Access PSI-BLAST from the BLAST page http://www.ncbi.nlm.nih.gov/BLAST/
  - Paste the gi number “4505641” in the query box.
  - Use the default parameters, except for choosing the RefSeq database, setting Expect Threshold (E-value) to 0.1, and setting maximum target sequences to 1000. Ensure that the PSI-BLAST Threshold is 0.005. Run PSI-BLAST.
  - The hits are divided into two sections. The hits with better statistical significance than the *e*-value threshold, 0.005, are listed first. Those with *e*-values worse than threshold, but have an *e*-value better than that selected on the query page, 0.1, are listed further down the page. Hits with *e*-values better than the threshold are used in forming the profile that will be used in subsequent PSI-BLAST iterations.

1) What is the name of the query sequence?
2) What is the length of the query sequence?
3) How many sequences are worse then your PSI-BLAST Threshold (but better than your Expect Threshold)?
   - Run PSI-BLAST iteration 2 with default selected sequences (above psi-blast threshold)
   - Note that newly found sequences are marked high-lighted in yellow.
   - Keep running PSI-BLAST until you find a very distant homolog from Bacillus smithii (note: try using ctrl-F to search for organism name, instead of manually scanning).
4) What iteration of PSI-BLAST did you find the B.smithii homolog?
5) What is the name of the homolog
6) What is the e-value and identity of the hit?

Genome Alignment with Mauve

- Download 4 genomes files:
- http://morganlangille.com/teaching/biol5705/assignment_2_genomes.zip
- Download Mauve (http://gel.ahabs.wisc.edu/mauve/download.php)

Start Mauve by clicking on “Mauve” file
Align your 4 genomes using “progressive mauve” with default parameters (this will take awhile)

7) Export your alignment as an image (Tools->Export->Export Image). Put the alignment picture embedded with your answers or send as separate attachment with your answer document.

8) What genome contains a very large inversion?

Search for PLES_53841 by name using View->Go to->Find Features

9) What genome is this gene in? And what is the gene name? Does this gene appear to be an insertion or deletion?

10) What is the name of the flanking genes (appearing directly on the left and right) of PLES_53841? Are these conserved (appearing in other genomes) or variable (not present in some of the genomes)?