Session 3 Recitation

1) Log into the HPCC (and a development node) and go to your home directory. In the directory that you created in Session 2 (CMSE890Sec303), create a subdirectory Session3 (no spaces).
2) Go to Ensembl and use wget to copy the primary genome assembly for worm (Caenorhabditis_elegans.WBcel235.dna.toplevel.fa.gz) to the Session3 directory on the HPCC that you just created. Use **gunzip** to unzip the file and more or less to look through the file.
3) Go to D2L and download the worm genome (1) file that is posted there to your laptop. Then use the transfer method of your choice (scp, rsync, mounting your home directory, globus, etc.) to copy it to the Session3 directory. How long did it take to transfer? Use **ls -lah** to check how big the file is when zipped. Use **gunzip** on it. Now use **ls -lah** to check the file size. Look up the **diff** command on Wikipedia and use it to compare the file that you copied using wget to the one you transferred from your laptop.
4) Go to D2L and download the worm genome (2) file that is posted there to your laptop. Use the same transfer method to copy it to the Session3 directory. How long did it take to transfer? Use **gunzip** on it. Use the diff command to compare it to the other two files (Note: you may want to redirect the output to a new file rather than the screen). What is wrong?
5) Use a different file transfer method (of your choice) and repeat #3. How long did that take to transfer? How did it compare to the file from wget?
6) Open the worm genome 2 file in either vi (Vim is fine) or emacs. Use Google to find a command to move immediately from one chromosome header to the next.
7) Use Google/Wikipedia to find a command for your text editor to put "chr" in front of the chromosome Roman numeral in each chromosome header. Do NOT try to do this by hand! Save your file.
8) Open the worm genome 1 file in your text editor of choice. How long did that take compared to worm genome 2?
9) Use **gzip** to zip back up the .fa files.