Author: Christopher Tran

Goal: Create bash script to print number of occurrence of genes and place information in .txt file. Create fasta(10) files of genes with all corresponding sequences. Primary objective is to find the highest number of occurrences in the bacterial genome.

Usage : week2.sh r\_bifella.fasta interesting\_motifs.txt

Files created: readme week2.sh motif\_count.txt motifs AGGTTCCA.fasta ATTATTATTA.fasta CGCGCGCGCG.fasta GGGGG.fasta TTATC.fasta ATGATGATG.fasta ATTTTTGCA.fasta GCAGT.fasta TAGT.fasta TTTTCTTTT.fasta

Instructions: Obtained interesting\_motifs.txt and r\_bifella.fasta from /home/rbif/week2\_assignment by coping files with cp command and moved to /home/tranc/week2. Created "readme" in /home/tranc/week2 to give step by step instructions. Using the "interesting\_motifs.txt" and "r\_bifella.fasta" to obtain goal in finding which gene had the highest occurences."week2.sh" was created. With the script, variables were set and loops were created by using both "interesting\_motifs.txt" and "r\_bifella.fasta". grep command was used to pull corresponding sequences and occurences within "interesting\_motifs.txt" and "r\_bifella.fasta". Information was transfered to "motif\_count.txt" which was created after grep command was performed. A print statement off the occurence will show up after running the script. This .txt file will also show the occurences of sequences within "r\_bifella.fasta". Location of "motif\_count.txt is /home/tranc/week2.

.Fasta files were then generated by using loop script. Corresponding genes with each sequences was specifically pulled using a loop and grep command. Information was pulled and transfered into (targetsequence).fasta.

Directory "motifs" was created in /home/tranc/week2. Using a loop, to move all the (targetsequence).fasta into motifs directory. .fasta files are located in /home/tranc/week2/motifs.

Please review week2.sh for coding scripts.

Highest number occuring sequence : TAGT occurrences : 394