PLPTH813, Homework 1

Due on 2/7/2019

Problem 1. Two sequences were enclosed. Use regular expression in TextWrangler or Notepad++ to tackle each of the following tasks. Each task is an independent job. Please report the regular expressions you use.

1. Remove the text after the first space in each sequence name
2. Remove the text after the second space in each sequence name
3. Remove sequences and only keep sequence names

>gi|330443715:175410-179348 Saccharomyces cerevisiae S288c chromosome XIV, complete sequence

ATGAGCGCTATCTATAAATTATCTATTCAGGGCATACGGTCTTTTGACTCCAATGATAGGGAAACTATTG

AATTTGGCAAGCCTCTGACTTTAATAGTCGGCATGAATGGTTCAGGTAAAACGACTATCATCGAATGTTT

AAAGTACGCTACCACAGGTGATCTACCGCCCAACAGCAAGGGAGGAGTATTCATTCATGACCCGAAGATA

ACTGGTGAAAAGGACATTAGAGCTCAGGTCAAACTGGCGTTTACGAGTGCCAATGGACTCAATATGATTG

>gi|330443743:1026843-1028048 Saccharomyces cerevisiae S288c chromosome XV, complete sequence

ATGCGAATCAACAGTGAGCTAGCGAACAAGTTTTCTGCCTCAACGGTGCACTTAGAACATATCACAACTG

CTTTAAGTTGTTTAACACCTTTTGGTTCTAAAGACGATGTGCTTATATTCATTGATGCTGATGGGCTGTC

ATTTGTCAGGGAGAATAATCATGTGATAAAAATCCAACTACTGTTATCTCGGGAGCTATTTATGTCTTAT

Problem 2. Describe what the following regular expressions mean.

1. AT.T
2. [AGCT]+
3. 12\.[0-9][0-9]
4. A{2}
5. A{2,4}
6. [BREAD]{2,4}
7. ^>
8. ^$
9. ^[A-Z]\*$
10. .\*

Problem 3.

Report the Unix commands for the following tasks:

1. Create a directory with the name of “HW1”
2. Create a file (two.fasta) using vi. This file needs contain the sequences from the problem 1.
3. Copy the file of two.fasta to two.backup.fasta
4. Rename the file two.backup.fasta to two.2.fasta
5. Delete two.2.fasta
6. Report the number of lines in the file of two.fasta
7. Grab both the 5th and 6th lines and report them using head and tail
8. Use grep to extract the names of two sequences
9. Use grep to extract the sequence name from chromosome XV