

Problems 2

Malay Basu (malay@uab.edu)

Problem 1

Write an R script that takes two arguments: (1) a fasta file name, (2) a sequence ID. The script should print out the sequence matching the id in FASTA format to the terminal.

Problem 2

Using `wget` download BLOSUM62 matrix from NCBI FTP server (`ftp://ftp.ncbi.nih.gov/blast/matrices/BLOSUM62`). Process it in anyway you can and read it in R as a matrix, a dataframe, or a list. You should store the data such a way that you can call the score given two amino acids as key as a fast lookup table. Read the accompanied `ex_align.fas` file and calculate the score of the given alignment. Consider each indel has score 0. The alignment file is in aligned fasta format.