

Homework 4: Python and data analyses

1. Using python write a tool which generates a random subsampling tool for sequences. Given a FASTA sequence database file, which has 100,000 sequences, generate a new file which is a random subset these sequences selecting only 10% of them. Make this 10% an option in the program so it is easy to change to 20%, etc.

See the script 'rand_shuffle_seqs.py' in the homework template

2. BLAST
- 3.