

Exercise 11 - Bioinformatics applications

As usual begin this week, by picking a partner and forking the TA's Exercise 11 Github repo. Then provide collaborative access to the other partner. Clone the forked repo so that you have the required files. Be sure to commit regularly to show how you arrived at your solutions and demonstrate coding effort by both partners.

1. In this problem we will practice the skills we used on Wednesday this week. In your forked repo, you should have a **Problem1** directory with eight bacterial proteomes (all with the file extension **.fasta**) and three sets of reference sequences (all with the file extension **.ref**). Write a Unix shell script that creates alignments and profile Hidden Markov Models (HMMs) for each reference set of sequences, searches for each protein type (sigma factor, sugar transporter, and spore formation component) in each bacterial proteome. Finally, the script should generate a single results file that contains a row for each match from all HMM-proteome combinations. Be sure this file contains columns indicating the proteome and HMM resulting in the match and the e-value associated with each match.

2. This problem will allow you to practice your regular expression skills and **Python/R** coding. In the **Problem2** directory, is a **.fasta** file with 90 sequences in it. Some subset of those sequences have the sequence motif 'AKKPRVZE' (motif1). Another subset of the 90 sequences have the motif 'AAQWWRNYGG' (motif2). Write a **R** or **Python** script to sort the initial fasta file into three new fasta files (one containing motif1, one containing motif2, and a third containing any sequences that do not contain either motif).

Devise a plan for splitting up the work and generating the required code. Do this in parallel, not sequentially. Don't forget to check and edit each other's code. Remember to frequently **add-commit** locally and **push-pull** to GitHub to avoid conflicts. Also, remember you don't have to be in the same place at the same time to work on this collaboratively thanks to GitHub!!!

Turning in your assignment via GitHub

Once you have committed all changes to your local Git repos and pushed all of those commits to the forked repo on GitHub, you can "turn in" your assignment using a **pull request**. This can be done from the GitHub repo website. When viewing the forked repo, select "Pull requests" in the upper middle of the screen, then click the green "New pull request" button in the upper right. You'll then see a screen with a history of commits for you and your collaborator, select the green "Create pull request button". In the text box next to your user icon near the top of the page, remove whatever text is there and add "owner's last name - collaborator's last name submission", but obviously substitute your last names. If I and Ann Raiho worked on the project together the text would read "jones-raiho submission". Then click the green "Create pull request" button. **Only one of you will need to create a pull request.**