Eric Mauritzen

Github: https://github.com/ericiscool/BIMM185

Introduction

This week we spent time finishing the homology exercise. In this exercise, I used the Bi-Directional Best Hit method to determine orthologs between the E. coli and A. tumefaciens proteomes. I also spent time last week iteratively improving my database by modifying the types of some fields. I was also very new to the MySQL library for python and spent time learning this.

BLAST

For the homology exercise I wrote a shell script to create the blast databases, 'MakeBlastDB.sh'. I then wrote a script to blast each entry in the database against each other. Next, I modified this output to include the 'send' and 'scov' columns by using simple arithmetic operations on the other columns in 'ParseBlastOutput.ipynb'. I then used this table to determine the orthologs. This process is outlined in 'GetOrthologs.ipynb'.

Bayesian Model to Predict Operons

Last week we were given the exercise to predict operons based on Bayesian methods. I am still working through this exercise on my own.

Discussion

In summary, the past week was spent cleaning up and becoming more familiar with the SQL database and implementing the homology exercise. I have been doing a lot of research (googling) on MySQL and interfacing it with python.

Suggestions

Firstly, thank you for your patience. BIMM185 is probably the worst class to teach in the bioinformatics curriculum. I believe that most of the students find it unusual to have such a loosely structured class. I think the entire class would benefit from having clearer expectations of what is expected.

In most other classes assignments are released to us as pdf documents that breakdown what is expected for each part of the assignment. These assignments are distinct from lecture materials. By combining both lecture and assignment in a single PowerPoint, confusion can arise.