

Biological Data Analysis (CSE 182) : Assignment 4

Logistics

Submit a hard copy containing the code and results. Create a compressed file containing the code and output as separate files, and email Julio. Do not deviate significantly from the suggested file names. The assignment is due by 12/7, midnight.

Computational Mass Spectrometry

- (25 pts.) Compute the Isotope Pattern of the peptide SLAMMER. Submit a plot of the the first 4 peaks, as well as your calculations. Assume the following:
 - The peptide has charge = 1.
 - The height of the highest peak is 100 (Note: we only care about the relative heights).
 - Only two isotopes are important $C - 13$, with abundance 2%, and $O - 18$, with abundance 5%. Note that these abundances are given for illustration purposes. In reality, the natural abundance of isotopes is smaller.
- (25 pts.) Mass spectra are often written in a simple text format, such as .dta. An example of a tandem mass spectrum S_1 in (.dta format) for an unknown peptide P is given on the web-site.
 - (a) Describe the format of the file S_1 in words. What is the parent mass and charge of P ? (Note: Use the web to understand the .dta format.)
 - (b) Identify the peptides in S_1 and S_2 using Mascot MS/MS ion search (http://www.matrixscience.com/search_form_select.html). Note that S_2 is a difficult example, so please use caution. Once the peptide is 'identified', give your own reasons for the identification being correct. Such reasons could include percentage of b-y ions matched, percentatge of total intensity explained, and so on.
- (48pts.) Design and implement an algorithm *perf* which takes as input a binary character matrix, and constructs a perfect phylogeny if one exists. The output of the program should be a tree (described as a set of parent child relationships). Test your algorithm on the two data-sets provided.
- (2pts.) What language did you use? How much time did you take to do the assignment? Who did you discuss your homework with?