Pseudo Code, Snapshots, and Assumptions for Hw #3

Bridget Mohn and Eric Wilson

For every pair of sequences in the given multiple sequence alignment

Count the number of time each amino acid occurs

Add the count to the proper place in the Counts dictionary

Count the number of times each aligned pair occurs

Add the count to the proper place in the Counts dictionary

Calculate the P() of each amino acid

Calculate the P() of each aligned amino acid pair

Calculate/Create the Amino Acid Substitution Matrix

Compare our matrix with BLOSUM 62 Amino Acid Substitution Matrix

Calculate the Correlation Coefficient between the two matrices

**Eric’s Pseudocode**  
Use 2D dictionary to count number of occurrences of each amino acid pair

Read each aligned sequence into big list

Create 2D dictionary with key for each possible character in each row and column

For each amino acid in all rows of list

Increment corresponding value in count list, increment total

For each pair between all amino acids (n^2)

Increment corresponding value in 2D dictionary, increment total

For each item in list of counts calculate probability

Value = value / total

For each item in 2D dictionary calculate probability

Value = value/total

For each item in 2D dictionary calculate substitution score

Value = lg(value/(probA \* probB))

Output matrix to file