

Pseudo Code and Intermediate Step Snapshots

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

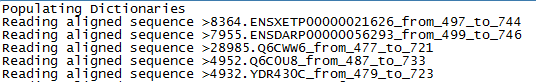
For each item in 2D dictionary calculate probability

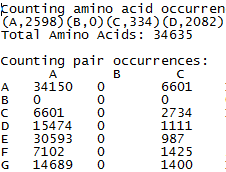
Value = value/total

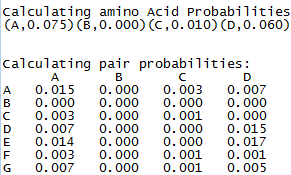
For each item in 2D dictionary calculate substitution score

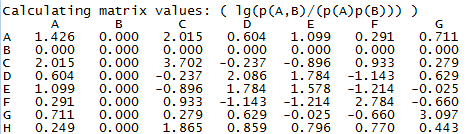
Value = lg(value/(probA \* probB))

Output matrix to file









Example Input and Output Snapshots

