Carson Sorenson

A01988754

CS5050

Testing and estimate of maximum size

To Approximate the size of the solution matrix where the O(n^2) algorithm becomes impractical I first found the amount of RAM I had on my computer. Building the solution cache for this solution I used integers which have a size of 4 bytes. My computer has 16 GB of RAM or $16 * 10^9$ bytes. I applied the equation n = sqrt(RAM / sizeof(data type)). Solving the solution for my computer this came out to be about 63,000. This means in my case the algorithm could handle up to two 63,000-bit long strings.

Final summary discussion

Observing this algorithm, it was very efficient with time and only took a few seconds to run a sequence all the way through. However, it is not efficient when it comes to memory use. The algorithm computes the solution to every subproblem and stores all the data. When computing the optimal solution, we do not use every solution, meaning we are storing the memory for no reason. This algorithm is very fast though and is very good at computing the optimal DNA matching sequence. This algorithm is practical for many applications and algorithms like this have done a lot for science.

It was interesting to run all the data and see how closely related we are to everything. There were three studies ran using all the sequences located in the "sequences" folder. There are three other folders. The folder "neandertal_study" includes a text document with the alignment score and the actual alignment. The actual alignment was 81,377 meaning that the human and neandertal's are very similar DNA strands. The next folder is "human_study". This folder contains a text document which was produced by the program which includes all the pairwise distances. I transferred this information to a table included in the same folder. I found that this study produced final scores that were very close to each other. The max score was 82,729 which was Armenia vs France, and the minimum score was 82,457 which was American vs Samoan. Lastly, there is the "great_ape_study" folder. This folder contains 6 text documents, each containing the alignments between the prototypical human and the 6 great apes. I found that the highest score was human against bonobo with a score of 67,848. The lowest score being human against baboon with a score of 56,897. There is also an excel sheet that has the scores of all pairwise distances between the apes.