

Comparative Study of Motif Algorithms

John Gervasoni A12441121

Jisoo Kim A15638045

Using python, I will create a tool that will implement the most commonly used motif algorithms. Each algorithm will be compared on the bases of time and space. The input of DNA sequences will vary in length and base frequency and the output of each algorithm will be compared for similarities and differences. Finally, I will be comparing the effect of different DNA and motif lengths on time complexity and memory management. For each algorithm. The goal is to definitively state which algorithm is best for which cases. Each algorithm will be independently rated by time elapsed and peak memory usage. The cases are as follows:

- (a) Short DNA sequence and motif
- (b) Short DNA sequence and long motif
- (c) Long DNA sequence and short motif
- (d) Long DNA sequence and long motif

Finally, if there is substantial difference between the algorithm's outputs and public DNA motif databases, two additional cases will be added:

- (e) Widest motif output
- (f) Narrowest motif output

To benchmark, both tracemalloc and time will be imported to monitor peak memory usage and time elapsed, respectively. My tool will make each algorithm available with easy-of-use flag modifiers. Given data that I have received from testing motif algorithms and a list of DNA sequences and motifs, my tool will suggest which algorithm to use and the estimated runtime.

The majority of data used will be from <https://jaspar.genereg.net/> and <https://motifmap.ics.uci.edu/>. Comparisons will be made between the motif algorithms and the non redundant transcription factor binding profiles generated from Jasper. The species will be yeast sacCer2 as well as those used previously in class.