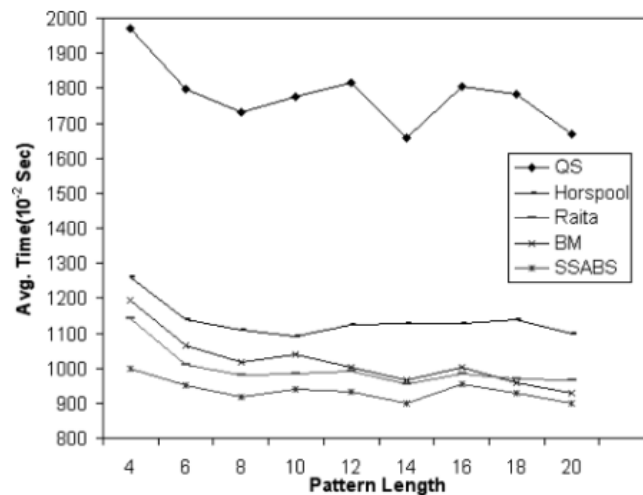


## Lab assignment 2: Exact string matching (KMP vs Boyer Moore)

- Given two DNA sequences **seq\_x** and **seq\_y**, find if one is a subsequence of other using KMP algorithm. If so, identify the *text* and *pattern* sequences as determined the pattern matching.

	Test case 1	Test case 2
<b>Input</b>	ATGCATTGC TGCATTG	TGCATTG ATGCATTGC
<b>Output</b>	text: ATGCATTGC pattern: TGCATTG number of matches: 1	text: ATGCATTGCATTGCATTGCG pattern: TGCATTG number of matches: 2

- Solve the above problem using Boyer-Moore method.
- Solve the above problem using Boyer-Moore-Horspool method.
- Plot the time complexity as seen through your experiments for the approaches listed in Lab assignments 1 and 2. Use the complete genome of *E. coli* ([Escherichia coli str. K-12 substr. MG1655](#)) for the benchmarking study. A sample plot for such a study, taken from *A FAST pattern matching algorithm* by S S Sheik *et al.* (doi: 10.1021/ci030463z), is shown below. Specify the list of patterns and their lengths used for the experiments and use 5 different patterns for each length.



**Figure 1.** Comparison of the proposed algorithm with the well-known algorithms available in the literature. The database used is the gene sequences comprised of nucleotides ( $\sigma = 4$ , Table 1). The graph clearly depicts the performance of various algorithms considered in the present study.

**Note:** The solutions should be submitted as assignment along with proper write up containing discussion of results in IEEE article format.