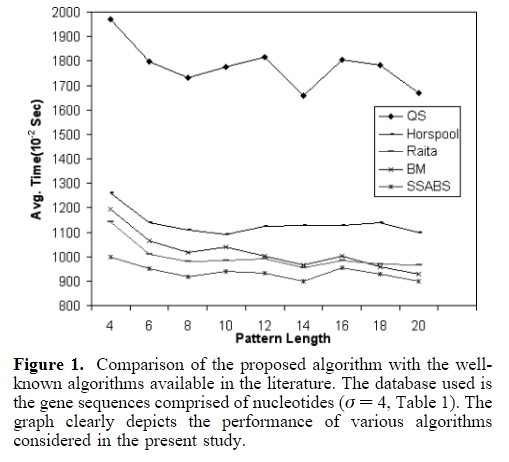
**Lab assignment 2: Exact string matching (KMP *vs* Boyer Moore)**

1. 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** |
| **Input** | ATGCATTGC  TGCATTG | TGCATTG  ATGCATTGC |
| **Output** | text: ATGCATTGC  pattern: TGCATTG  number of matches: 1 | text: ATGCATTGCATTGCATTGCG  pattern: TGCATTG  number of matches: 2 |

1. Solve the above problem using Boyer-Moore method.
2. Solve the above problem using Boyer–Moore–Horspool method.
3. 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](https://www.ebi.ac.uk/ena/browser/view/U00096)) 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.



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