**Mansoura University**



**Faculty of Computers and Information**

**Department of Computer Science**

**Project Proposal**

# Arabic Title

# 

### 

##### English Title

### 

### **Submitted by:**

#### 

|  |  |  |
| --- | --- | --- |
| Student Name | Student Email | Section |
| Huda Abdelfatah Elsayed |  | 1 |
| Hadeer Neiazy Elsayed |  | 1 |
| Naira aAbdul mawgod Metwally |  | 1 |

Project Abstract:

Counting the k-mers in a DNA sequence. k-mer is just a sequence of k characters in a string (or nucleotides in a DNA sequence). Now, it is important to remember that to get all k-mers from a sequence you need to get the first k characters, then move just a single character for the start of the next k-mer and so on. Effectively, this will create sequences that overlap in k-1 positions.

So, by way of example, the sequence ATCGATCAC contains the following 3-mers (k-mer of size 3): output--🡪 (ATC , TCG , CGA , GAT , ATC , TCA , CAC). Counting K-mers for a DNA sequence means finding frequencies of K-mers for the entire sequence. In bioinformatics, K-mer counting is used for genome and transcriptome assembly, metagenomic sequencing, and error correction of sequence reads

Project Objectives:

K-mer counting is an important step in many application for example, genome assemblers using the overlap-layout-consensus paradigm such as (Celera , Arachne).and use k-mer shared by read as seeds to find overlaps. K-mer count statistics are also used to estimate the genome size. The count k-mer are also used to seed fast multiple sequence alignment

Used k-mer frequencies with large k(20<= k <=100) to study the mechanisms of sequence duplication in genomes. We consider the k-mer counting problem in the context where the input string S is either one DNA sequence or concatenation of many DNA sequence, and the alphabet is {A,C,G,T}.The main application to which we apply our new k-mer counting algorithms. Finally, k-mer distribution can produce new biological insights directly.

Who are the project **competitive**? and how will your project be **different**?

\*Jellyfish --🡪 its commend line tool for counting k-mer in DNA sequence.Output of it as file

\*The different our project is GUI , Faster

Tools, Hardware and Software Resources:

**Tools :-**

**1-Python 3.10**

**2- Brackets**

**Software:-**

**Hardware:-**

**1-pc**

SCHEDULING PHASES:

|  |  |  |
| --- | --- | --- |
| **From** | **To** | **Activity** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

References:

[**https://bioinfologics.github.io/post/2018/09/17/k-mer-counting-part-i-introduction/**](https://bioinfologics.github.io/post/2018/09/17/k-mer-counting-part-i-introduction/)

[**https://academic.oup.com/gigascience/article/7/12/giy125/5140149**](https://academic.oup.com/gigascience/article/7/12/giy125/5140149)