4/23/2021

SEQUNCE CORRECTION TOOL

Team Member :

- Ahmed Emorshdy
- Amaal Ahmed
- Mohamed Elmorsi
- Mohamed Amged
- Mahmoud Abbas
- Seif Mohamed
- Samir Saad

ERROR CORRECTION PROJECT

Project Abstract:

Recent advancements in sequencing machines have rapidly improved our ability to study genomic material at an unprecedented scale. Despite substantial improvements in sequencing technologies, errors present in the data still risk confounding downstream analysis and limiting the applicability of sequencing technologies. Computational error correction promises to eliminate sequencing errors, but the relative accuracy of error correction algorithms remains unknown. So, we aim to develop software dealing with different biological data using artificial intelligence technology to detect a specific error resulting about machine sequence and try to correct it to get best accuracy

Project Objectives:

- Detecting error in sequencing data in many different laboratories rapidly using artificial intelligence technology, These errors resulting from many reasons such as sequencing machine error or contamination.
- … is also valuable in a number of research centers which dealing with:
    I. Sequencing new data.
    II. Coverage
    III. Data accuracy

Software Tools:

- Google Colab
- Python Library
- Vs Code
- XXamp

The tools that compete our tool in Error Corrections are:

- Blue: it is a fast, accurate short-read error-correction tool based on k-mer consensus and context and It will correct both Illumina and 454-like data, and accepts sequence data files in both FASTQ and FASTA formats and is written in C# and runs natively on Windows, and with mono on Linux.

- BLESS: A Bloom-filter-based Error correction Solution for high-throughput Sequencing reads is a novel algorithm that produces accurate correction results with much less memory compared with previous solutions.

- Echo: an application for detection and analysis of oscillators identifies metabolic regulation on genome-wide circadian output.

- HI TEC: an algorithm that provides a highly accurate, robust and fully automated method to correct reads produced by high-throughput sequencing methods. Our approach provides significantly higher accuracy than previous methods. It is time and space efficient andworks very well for all read lengths, genome sizes and coverage levels.

- Quake: Quake is a package to correct substitution sequencing errors in experiments with deep coverage (e.g. >15X), specifically intended for Illumina sequencing reads

- RACER: A software program for correcting errors in sequencing data. RACER has better error-correcting performance than existing programs, is faster and requires less memory. To support our claims, we performed extensive comparison with the existing leading programs on a variety of real datasets.

Our Tool:

- Get high accuracy for biological data
- Using AI not traditional algorithm to dealing with data
- Software not depending on specific type of data but dealing with DNA, RNA and protein sequencing
  - ✓ Simple UI
  - ✓ Easy to use for Bioinformatics Students
  - ✓ Efficient memory