Midterm Project Update

Jacob Morrison
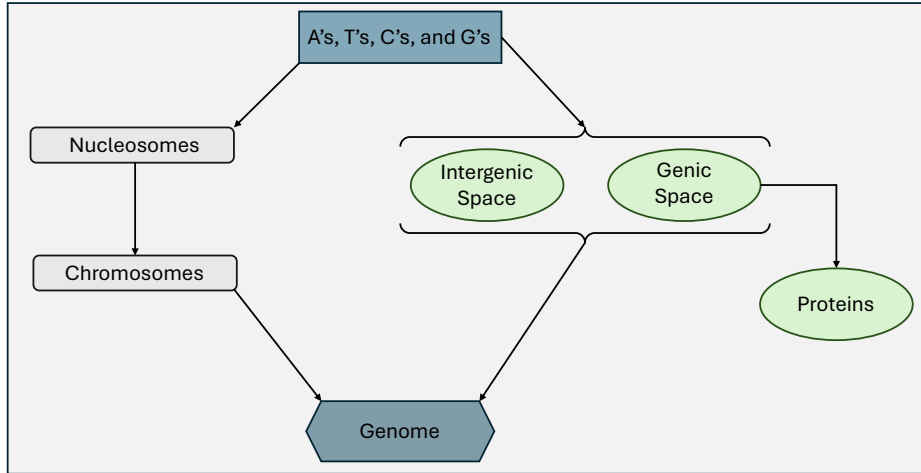
Grand Valley State University
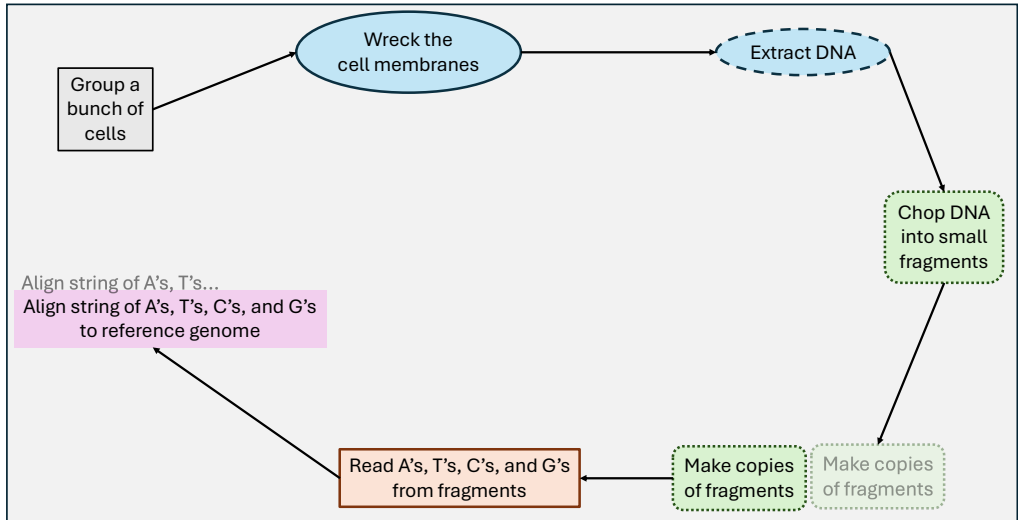
16 October 2024

Introduction

- Group Name: Packing the Bits

- Ground Members: Jacob Morrison

- A little bit about myself
  - Bioinformatics Research Scientist at Van Andel Institute
  - Work in a lab that studies the origins of ovarian cancer
  - Develop and maintain software for analyzing DNA sequencing data

- Project: Create a small tool to improve the speed of one of the tools I maintain for work

Genetics

DNA Sequencing

Project Motivation

- After alignment, an initial quality control is performed
  - Verifies successful sequencing experiment
  - Understand potential biases of data before analysis
- Part of QC is checking "coverage"
- Current process is slow
- Create tool to improve speed and reduce number of times the same file is read
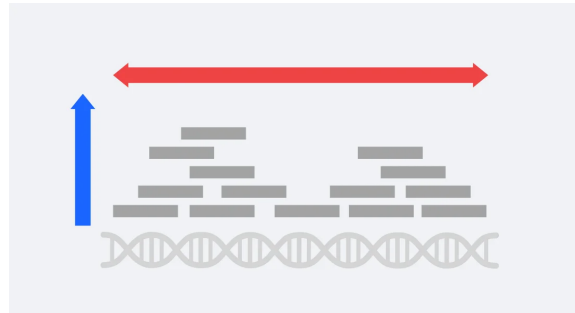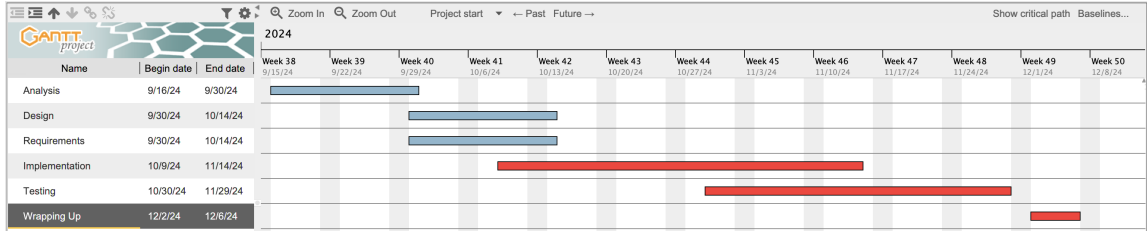

Image from 3billion.io (link)

Project Timeline

- Finished Analysis, Design, and Requirements

- Started implementation at the end of last week

- Group of one, so it's all on me

Project Demonstration

CLI

```
(base)  (src) $ ./coverage

Usage: coverage [options] <in.bam>

Options:
    -o STR    Output filename
    -r STR    BED file of regions of interest
    -s INT    Step size of windows [100000]
    -@ INT    Number of threads [1]
```

Output

```
chr1    0       290278    0
chr1    290278  290320    1
chr1    290320  290428    2
chr1    290428  290470    1
chr1    290470  842029    0
chr1    842029  842179    1
chr1    842179  842271    0
chr1    842271  842421    1
chr1    842421  955309    0
chr1    955309  955371    1
```

Thank You