BGR

Bacterial Genomic Reconstruction

Department of Electrical Engineering
香港城市大學
City University of Hong Kong

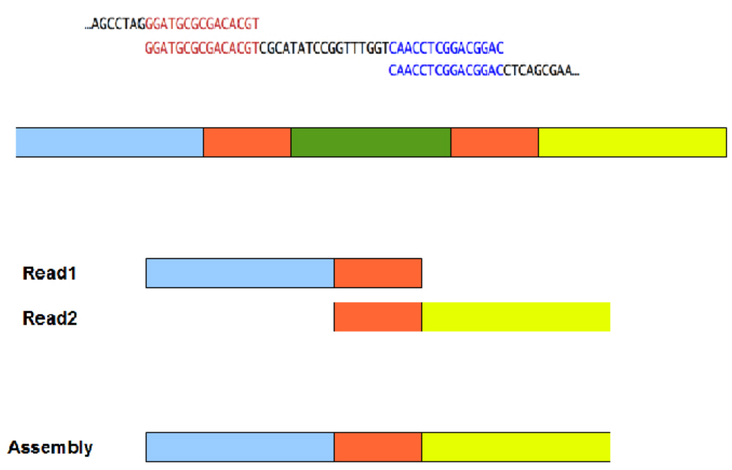
NAME: SIVAKUMAR SRINIVAS MAJOR: CDE

SUPERVISOR: DR. SUN, YANNI

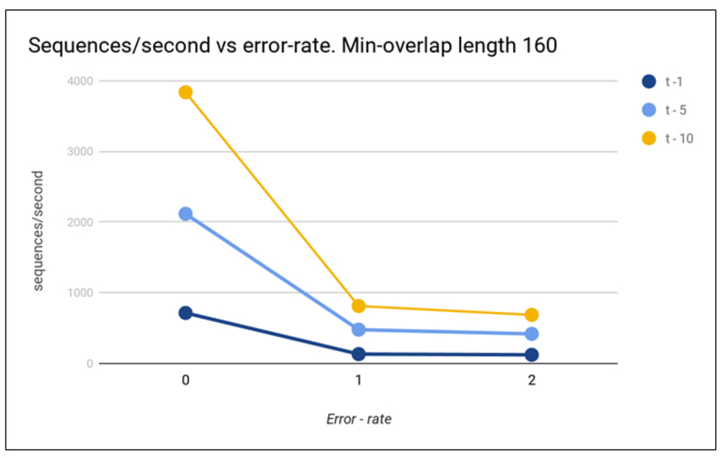
Objective/Background

- Pre-processing of viral metagenomic data.
  - Viral data is plagued with bacterial genomes
  - Problem, we do not possess references for all Bacteria.
  - Sequencing data contains errors.

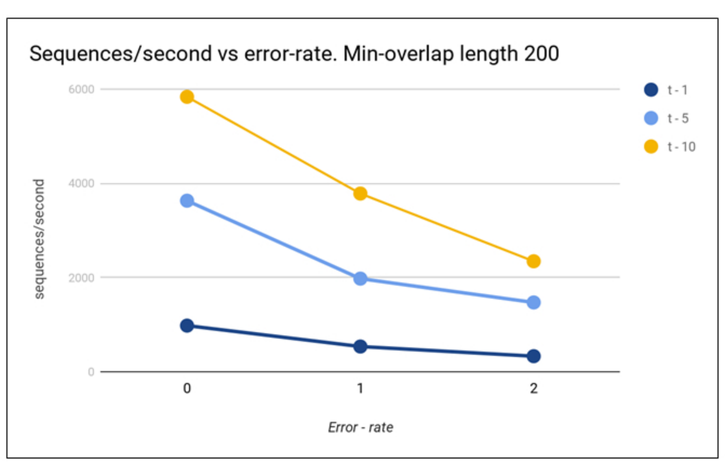
Read alignment



Overlap assembly



Dataset 1



Dataset 2

Methodology

This approach combines alignment and assembly.

1. Use 16s rRNA genes as a reference genome for alignment (forms the seed).
2. The seed is then extended using overlaps.

Results

Two datasets used to test the program.

Reads were of length 250.

1. Simulated using ART.
2. With 1000+ viral and 20 bacteria genomes to see if the program scales.

Conclusion

1. A memory-efficient, multi-threaded and scalable approach was successfully developed.
2. The final implementation can be found at github.com/srinivas9804