ALGORITHM

Project Report

Prof. Joo Jong Hwa
2018.06.12
2016112177 Seo Bomi

ⅰ) Project Objectives :

Recover a genome of N length when M short reads with L length are given.

ⅱ) Design Overview

MyGenome: Original Genome of N length.

ReferenceGenome: Comparison genome of length N with 5% discrepancy with MyGenome.

Short Reads: String of length L from any location in MyGenome.

Mismatch: Reference Genome and MyGenome are not exactly the same, allowing for some discrepancy in the search for short results from reference Genome.

ⅲ) Algorithm used

- Benchmark Algorithm:

○ Trivial Search: Search with trivial method.

- My Algorithm:

○ BWT string matching: The traditional BWT algorithm wastes memory as it creates the N x N matrix when the string length of a given string is N and begins searching after alignment. I thought it was only necessary to have a partial string for alignment, so I created and sorted the N x 30 matrix.

-Time Complexity: O(nm)

○ Indexing Algorithm: The original indexing algorithm was used to use the Phone Book method, which also thought the Phone Book was heavily occupied by memory and there were too many types of orders for the need. Therefore, short reads were used as sequences, not as sequences of all possible combinations.

-Time Complexity: O(log n)

ⅳ) Results

N: 10000, mismatch=3

short read	Trivial	BWT	Hashing
long = 20 num = 1000	132.348 87.21%	87.651 86.93%	0.205 86.29%
long = 30 num = 1000	133.234 94.47%	92.47 93.56%	0.272 91.92%
long = 30	193.576	130.581	0.184

num = 1500	98.57%	98.42%	93.22

N: 100000, mismatch=3

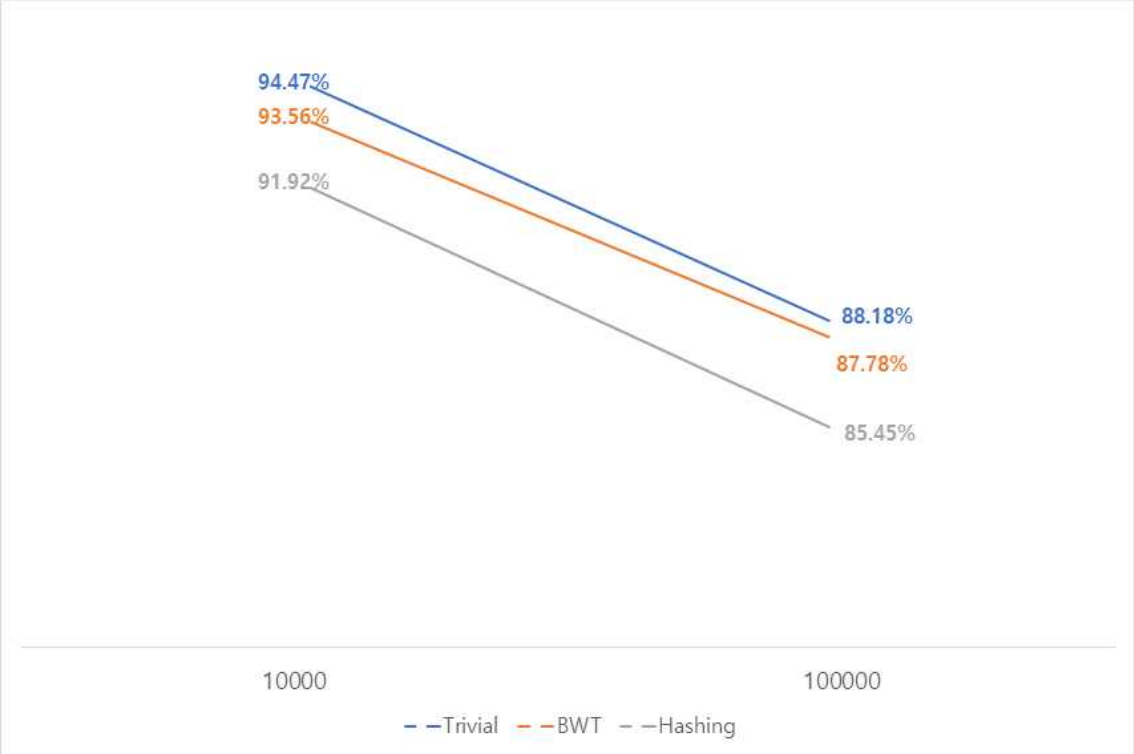
short read	Trivial	BWT	Hashing
long = 30	14277.439	9821.89	1.991
num = 10000	88.182%	87.784%	85.452%

N: 1000000, mismatch=3

	Hashing
short read long = 35	21.366
short read num = 100000	97.4212%
short read long = 40	25.708
short read num = 100000	98.1856%

Time Comparison

Accuracy Comparison



v) Development Environment

CPU: Intel(R) Core(TM) i5-7500 CPU @ 3.40GHz
RAM: 8.00GB
OS: 64bit Window10
Language: C++
Platform: Microsoft Visual Studio 2017