ALGORITHM PROJECT
DNA RECONSTRUCTION

2016112185 컴퓨터공학과 박지수

Your Date Here

PROBLEM DEFINITION

Short Reads

Reconstruct

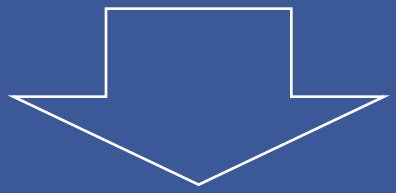
Original Sequence

DATA EXPLANATION

Length of Genome = 197000
Where?
Randomly generated 197000 A, T, G, Cs using function rand()

INPUT & OUTPUT

shuffled Short Reads
Length = 800
Overlap = 400
Number = 492



reconstructed DNA Sequence
Length = 197000

ALGORITHM

REFERENCE

DENOVO

ALGORITHM : EULERIAN PATH

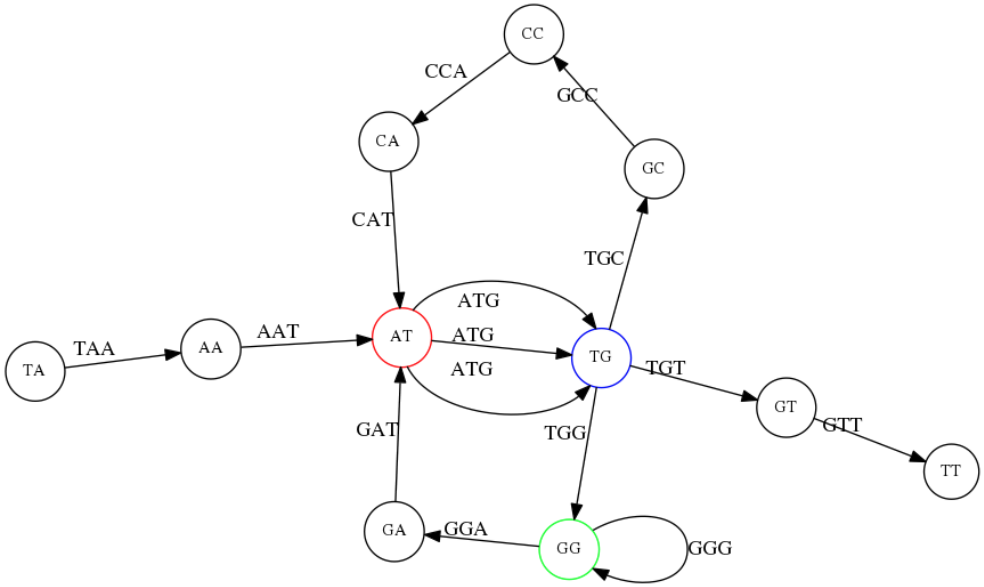
Split My genome into
Short Reads & Shuffle

Split Short Reads into L-Node & R-Node

Generate Debruijn Graph
(Directed Graph) using Nodes

Find Eulerian Path

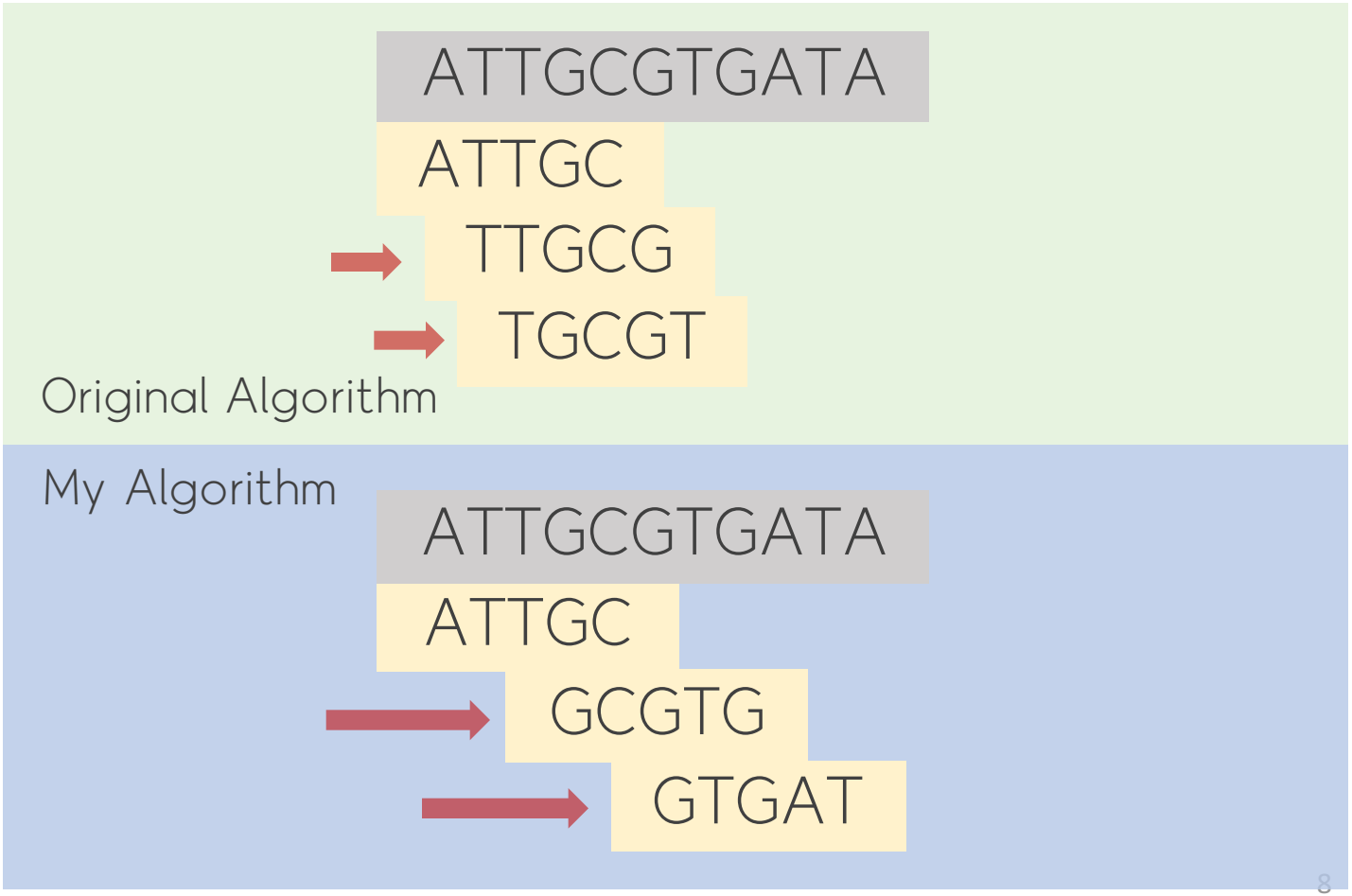
ALGORITHM : EULERIAN PATH



Short Node =
Edge of Graph

ALGORITHM : EULERIAN PATH

Divide Short Reads into Smaller Nodes



ALGORITHM : EULERIAN PATH

Finding Starting Vertex of Eulerian Path

Mid-Vertices : In-Degree = Out-Degree

Start-Vertices : In-Degree < Out-Degree

Start-Vertices : In-Degree > Out-Degree

ALGORITHM : EULERIAN PATH

Selecting Correct Path

Longest Path

MY MACHINE


macOS High Sierra
버전 10.13.4
MacBook Air (13-inch, Early 2014)
프로세서 1.4 GHz Intel Core i5
메모리 4GB 1600 MHz DDR3
그래픽 Intel HD Graphics 5000 1536 MB
일련 번호 C1MP56AZG085
시스템 리포트...
소프트웨어 업데이트...


디바이스 이름 DESKTOP-9VK8THQ
프로세서 Intel(R) Core(TM) i5-4260U CPU @ 1.40GHz 2.00 GHz
설치된 RAM 2.00GB
디바이스 ID 50CE3AD6-711C-4A3C-82A6-5B0452F687ED
제품 ID 00328-00252-49792-AA719
시스템 종류 64-bit operating system, x64-based processor
펜 및 터치 이 디스플레이에 사용할 수 있는 펜 또는 터치 식 입력이 없습니다.

TIME & SPACE COMPLEXITY

Time Complexity = O(V)

Space Complexity = O(V^2)
Directed Graph = n x n matrix

COMPARE WITH BENCHMARK

Brute-Force Graph Search

Time : 15~20 minutes
Time Complexity : O(V!)
DFA method of enumerating all possible paths of a Graph

Eulerian Graph Search

Time : 12 seconds
Time Complexity : O(V)
Only one start vertex

PROS & CONS

advantages
- Linear time : O(V)
- High Accuracy

difficulties
- Tangled graph
- Storage requirements (huge)
    - Long Short reads
    - Short Genome

FUTURE IMPROVEMENT

Adjacency Matrix
→ Adjacency List
Space complexity = O(m+n)

From	To		
1	2	3	5
2	3	5	
3	2		
4	2	5	
5			