

---

## 9조 활동 보고서

6월 02일 활동 보고서

---



과	목	알고리즘 3분반
교	수	주종화 교수님
제	출 일	5월 26일
조	이 름	9조
조	장	2022110151 이주연
조	원	2023111033 김태은
		2021112504 박지우
		2022113556 정태호

## 목차

### I. 향후 일정 정리

### II. 문제 재정의

## 1. 향후 일정 정리

6월 3일 (화)	개별 구상한 BWT 기반 변형 알고리즘 공유
6월 6일 (금)	코드 구현 완료
6월 7일 (토)	PPT 제작 <ul style="list-style-type: none"> <li>- 각 구현한 <b>알고리즘 구조</b> 설명 ( pseudo code 혹은 다른 방법 이용)</li> <li>- 기존 알고리즘(BWT)에서 <b>변형한 부분</b> 설명</li> <li>- 3가지 척도를 활용한 알고리즘 <b>성능 분석</b>:               <ul style="list-style-type: none"> <li>: 1) 최악의 시간복잡도 2) 제자리성 3) 안전성</li> </ul> </li> <li>- 구현한 알고리즘의 장/단점               <ul style="list-style-type: none"> <li>: BWT/직선적 알고리즘(벤치마킹 알고리즘)과 <b>비교했을 때의 장/단점</b></li> </ul> </li> </ul>
6월 8일 (일)	PPT 및 코드/파일 제출
6월 11일(수)	발표

## 2. 문제 재정의

**Problem: Given 10,000 number of short reads of length 32, reconstruct the original sequence of length 3,000,000 that those shorts reads come from.**

항목	설정값	설명
Reference 길이 N (부모 sequence 길이)	3,000,000	부모 시퀀스(reference)의 길이, 현실적으로 메모리 허용 범위 내에서 알고리즘별 시간 차이가 유의미한 길이로 설정
Short Read 수 M (한 아이로부터 추출된 short read 개수)	10,000	아이로부터 추출한 short read의 개수
Short Read 길이 L	32	Short read 하나당 길이
허용 mismatch 수 D	3개 이하	돌연변이를 감안해 허용할 오류값 개수
레퍼런스	실제 유전자의 일부	아이의 유전자를 통해 부모의 유전자를 알아낸다

## 문제 상황 정의

“한 아이의 m개 short reads와 부모의 sequence를 비교해  
아이가 부모의 친자식인지 판별하고, 아이의 sequence를 복원한다.”

### 부모 DNA Sequence

AGCTTAGTGATCTTTAGCCC  
TAGTCACCTAGTCTCCATCA  
TGCTACGGCTAAGCATATTA  
TCATGATACAGGTATTC.....

길이 : N = 3,000,000

.....CATG

### 아이 한 명의 DNA sequence로부터

랜덤 추출된 M=10,000개의 short reads

8개 | 8개 | 8개 | 8개

1<sup>st</sup> short read

8개 | 8개 | 8개 | 8개

2<sup>nd</sup> short read

8개 | 8개 | 8개 | 8개

3<sup>rd</sup> short read

.

.

.

8개 | 8개 | 8개 | 8개

M=10,000<sup>th</sup> short read

허용 mismatch 수  $D \leq 3$ 개

( short read를 4파트로 나눴을 때 4파트 중 한 파트는 완전히 일치하는 파트)