
9조 활동 보고서

5월 26일 활동 보고서



| | | |
|---|-----|----------------|
| 과 | 목 | 알고리즘 3부반 |
| 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. 사용 데이터 정의

II. 문제 정의

III. Benchmark 알고리즘 설정

IV. 향후 계획

0. 조장 지정

2022110151 이주연

1. 사용 데이터 정의

| 항목 | 설정값 | 설명 |
|-----------------|---------------|---|
| 레퍼런스 길이 N | 100,000 (10만) | 현실적으로 확인이 가능하면서 알고리즘별 시간차이가 유의미한 길이로 설정 |
| Short Reads 수 M | 10,000 | 탐색이 끝난 short reads 수 |
| Short Read 길이 L | 15 | Short read 길이 |
| 허용 mismatch 수 D | 3개 이하 | 돌연변이를 감안해 허용할 오류값 개수 |
| 레퍼런스 | 실제 유전자의 일부 | 아이의 유전자를 통해 부모의 유전자를 알아낸다 |

2. 문제 정의

Short Read Sequencing Problem

길이 N의 부모 유전자 배열을 알아내기 위해

길이 L의 short reads(유전체의 일부분을 나타내는 짧은 염기서열) M개를 이용해 알고리즘으로 전체 유전자 배열을 추론한다.

허용 mismatch 수(D)는 3개로 설정한다.

3. Benchmark 알고리즘

BWT 기반 Matching

Reference 시퀀스에 대해 BWT 변환 수행

예상 시간 복잡도 : $O(M \cdot L)$

4. 향후 계획

random read 생성기 구현

BWT 변환 함수 구현

각자 BWT 알고리즘 개선 방안 설정 후 구현

각 알고리즘의 소요 시간 비교 및 분석