



# 컴퓨터 알고리즘과 실습



---

학번 : 2016110056

학과 : 불교학부

이름 : 박승원

날짜 : 2017년 5월 24일

---

# 문제

- practice 7 의 문제 2번에서 A/C/G/T로 이루어진 길이  $N=1,000,000$  짜리 sequence 를 random 하게 생성하는 것을 해보았다. 이 sequence 를 MyGenomeSeq 이라고 하자.
- 문제 1. MyGenomeSeq를 임의의 위치에서 길이  $L=50$  짜리로 잘라서 short read 를 만들어 보자. 이와 같은 short read 를  $M=100,000$  개를 만들어 reads.txt 파일에 저장하는 프로그램을 작성하여 코드와 reads.txt를 제출하시오.

```
#include<random>
#include<fstream>
#include<cstring>
#define M 1000000
#define N 100000
#define L 48
#define D 2
using namespace std;

char gene[M], person[M], short_read[N][L];

int main()
{
    uniform_int_distribution <> di{0, M-48}, percent{0, 99}, code{0, 3};
    random_device rd;
    char gene_code[4] = {'G', 'T', 'C', 'A'};

    // generate human gene
    ofstream f1("human_gene.txt");
```

```

for(int i=0; i<M; i++) {
    char c = gene_code[code(rd)];
    gene[i] = c;
    f1 << c;
    if(i % 100 == 99) f1 << endl;
}

// write gene of a person to diagnose
ofstream f3("person_gene.txt");
for(int i=0; i<M; i++) {
    if(! percent(rd)) { // 1% mutation rate
        char c = gene_code[code(rd)];
        f3 << c;
        person[i] = c;
    } else {
        f3 << gene[i];
        person[i] = gene[i];
    }
    if(i % 100 == 99) f3 << endl;
}

// generate short reads
ofstream f4("short_read.txt");
for(int i=0; i<N; i++) strncpy(*(short_read + i), person + di(rd), L);
for(int i=0; i<N; i++) {
    for(int j=0; j<L; j++) f4 << short_read[i][j];
    f4 << endl;
}
}

```

문제2. 평균 커버리지..

$(9+9+11+8+11+11+14+13+15+14)/40$