

Quake

Сайт: <http://www.cbcb.umd.edu/software/quake/>

На сервере программа находится в /storage/labnas/NGS/4/Quake

Запуск bin/quake.py

Четвертое домашнее задание

1. Оценка качества работы Quake

Исправить риды с помощью Quake (используя $k=15$), и приложив их и исходные риды к геному, оценить:

- False positives — количество ошибок в изначальных данных, которые не были исправлены
- False negatives — количество правильных нуклеотидов, замененных на неверные
- True negative — количество исправленных ошибок

Мутациями пренебречь

2*. Улучшить потребление памяти Quake

Снова необязательное бонусное задание

В файле correct.cpp есть строка:

```
bithash *trusted = new bithash(k);
```

Она вызывает конструктор bithash (bithash.cpp, bithash.h), который создает бит сет размером 4^k . Для $k=19$ это уже 32Гб. Попробуйте изменить класс bithash так, чтобы память использовалась более экономно, при этом сильно не замедлив выполнение программы.

Данные:

Папка: /storage/labnas/NGS/4/

Данные: SRR350492_1.fastq

Тестовые данные: test.fastq

Референсный геном: MG1655-K12.fasta

P.S. Презентация с основами matplotlib, может кому пригодится <http://mit.spbau.ru/files/scipy.pdf>