

BWA

Сайт: <http://bio-bwa.sourceforge.net/>

На сервере программа находится в /storage/labnas/NGS/3/bwa-0.6.1

Возможно, в процессе работы пригодится визуализатор Tablet <http://bioinf.scri.ac.uk/tablet/download.shtml>

Третье домашнее задание

0. Разобраться с BWA

Для прикладывания длинных ридов используйте BWA-SW.

1. Статистика ридов 454.

С помощью BWA посчитать следующие статистики для ридов:

- Покрытие генома (график, процент покрытого генома, среднее покрытие)
- Распределение длин indel'ов
- Распределение зависимости количества indel'ов от длины гомополимерного участка (участок состоящий из одного типа нуклеотидов, например AAAAAAAAA)
- Зависимость качества нуклеотидов
 - при ошибочно вставленном нуклеотиде
 - при замене
- Таблица частот замен/вставок/удалений

Read\Reference	A	C	G	T	-
A	-				
C		-			
G			-		
T				-	
-					-

Статистики реализуйте так, как вы считаете разумным.

2*. Исправление вставок и удаление в ридов 454 по ридам Illumina.

Задание, которое можно сделать вместо 1.

С помощью BWA и ридов Illumina попытаться исправить вставки и удаления в ридах 454. Оценить качество исправлений с помощью референсного генома.

Для тестирования удобно пользоваться референсным геномом, но итоговая программа должна принимать на вход только файлы с ридами.

Данные:

Папка: /storage/labnas/NGS/3/

Данные: 454.fastq

Тестовый геном: test.fasta

Референсный геном: reference.fasta

*Illumina: illumina_left.fastq, illumina_right.fastq