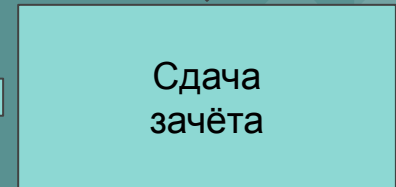
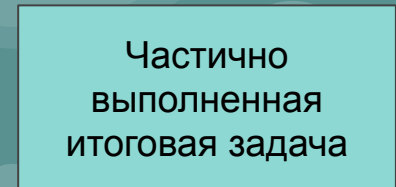
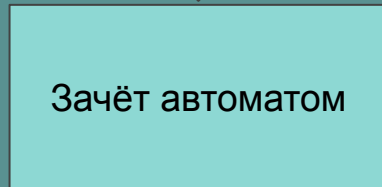
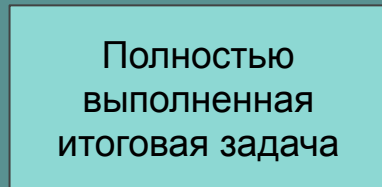
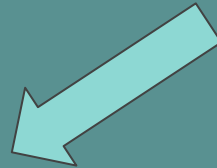
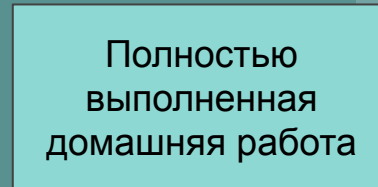
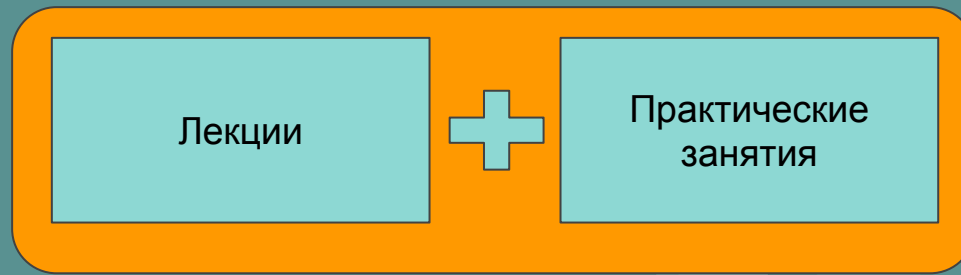




Структура курса “Основы программирования”

Преподаватели:

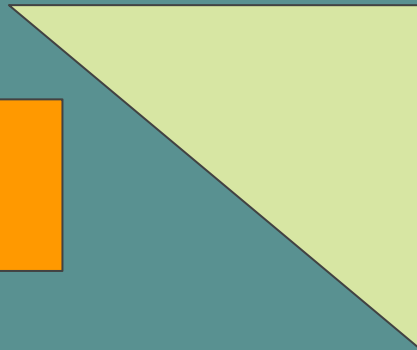
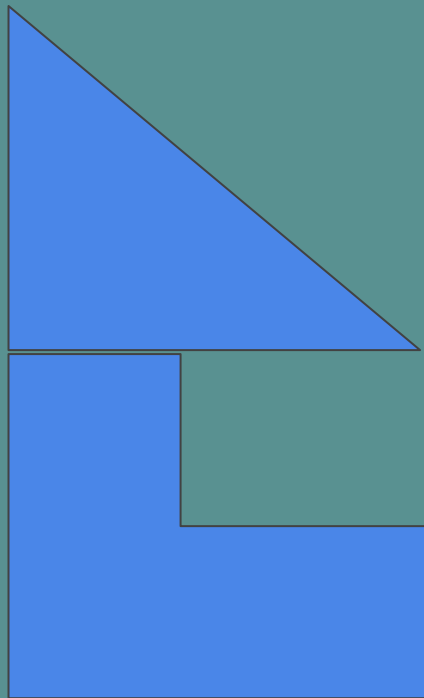
- Колпаков Ф.А.
- Колмыков С.К.
- Куляшов М.А.



Структура курса “Основы программирования”

Практические
занятия

Домашние работы



Структура курса “Основы программирования”



Итоговая задача

Итоговая задача курса

Общее описание: необходимо написать программу, которая принимает fastq файл и на выходе даёт набор базовых статистик и графиков.

Чтение файла:

- прочитать и расшифровать качество прочтений из FASTQ файла;

Расчет базовой статистики:

- имя файла;
- количество последовательностей в fastq файле;
- самая часто встречающаяся длина последовательности;
- средний GC состав в файле;
- средний нуклеотидный состав;

Итоговая задача курса

Общее описание: необходимо написать программу, которая принимает fastq файл и на выходе даёт набор базовых статистик и графиков.

Работа с адаптерными последовательностями:

- спросить пользователя размер минимального вхождения адаптерной последовательности. Если вводимое пользователем число меньше 3, то выдавать сообщение, что это слишком короткое вхождение и устанавливать все равно минимальное 3.
- после расчета базовой статистики, предложите пользователю удалить адаптеры. Если ответ положительный, то необходимо найти адаптерные последовательности (они будут вам даны отдельным файлом) с учетом того, что адаптер может быть представлен не полностью
- Посчитать количество удаленных адаптерных последовательностей
- сохраните все измененные последовательности в новый fastq файл

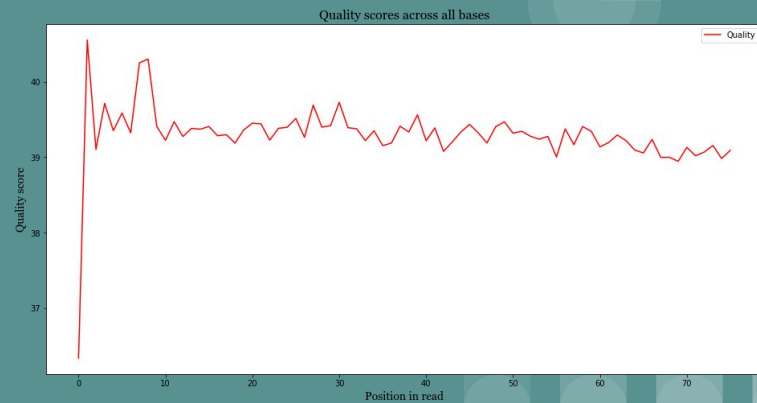
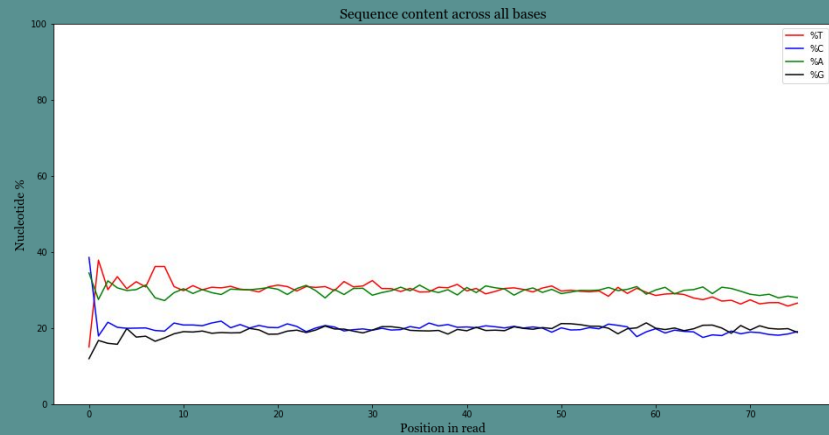
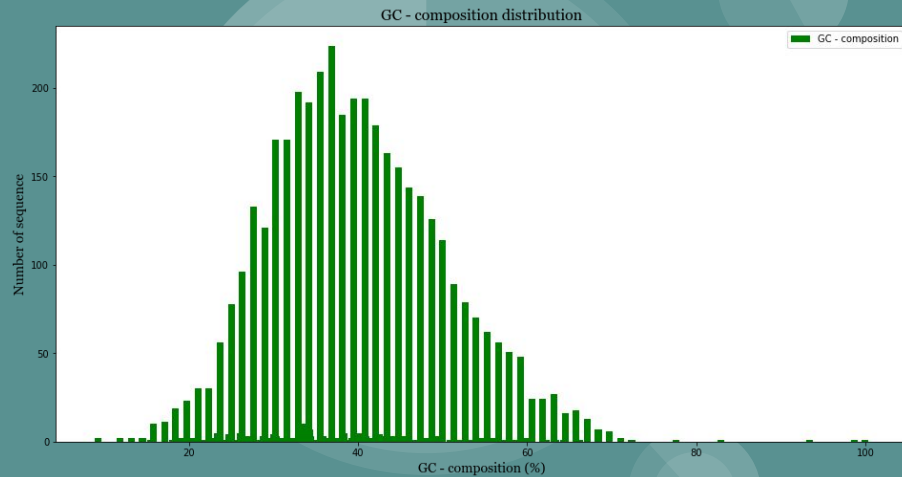
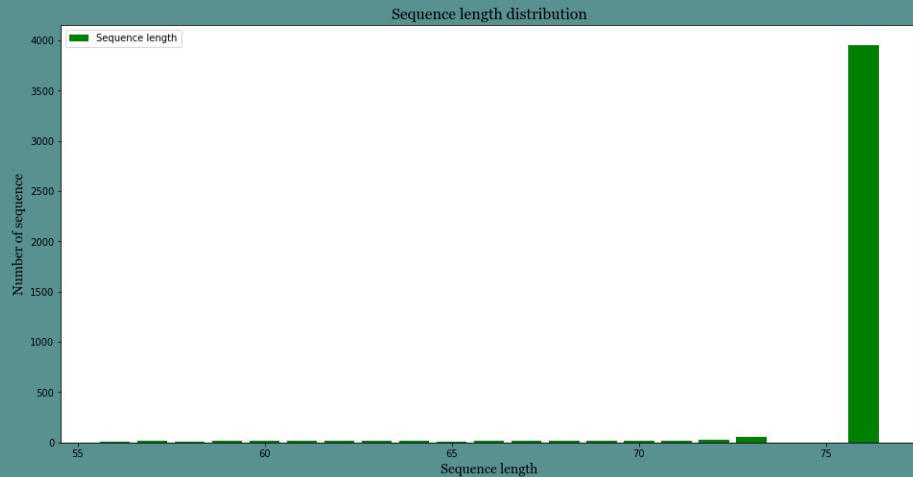
Итоговая задача курса

Общее описание: необходимо написать программу, которая принимает fastq файл и на выходе даёт набор базовых статистик и графиков.

Графическая визуализация:

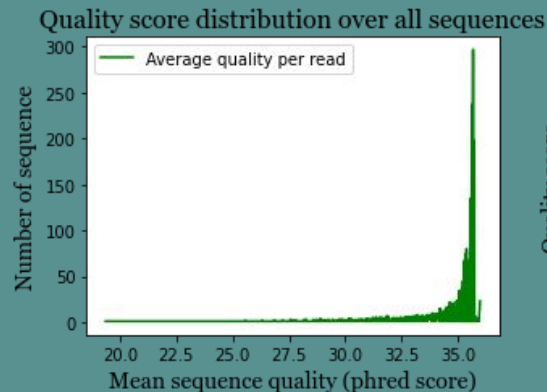
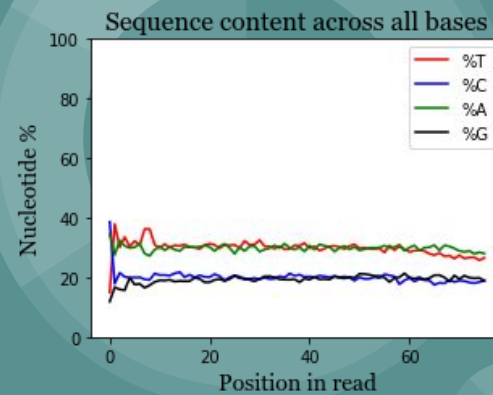
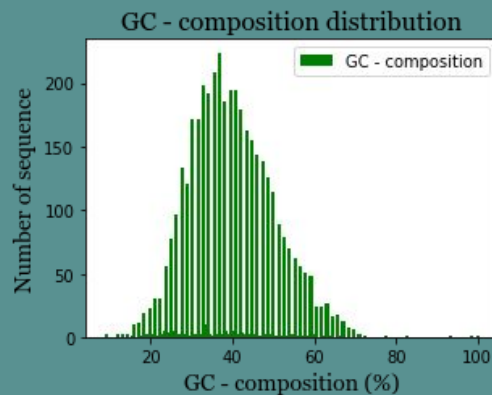
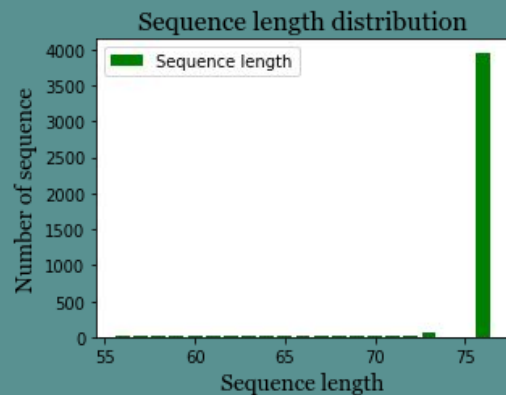
- добавить возможность выбора качества сохраняемых графиков
- Распределение длин последовательностей
- общий (единый, не subplot) график распределение GC состава последовательностей
- график частоты нуклеотидов в каждой позиции в последовательностях
- Рассчитать для каждой последовательности средний phred score и визуализировать распределения этих значений на графике
- построить график среднего значения Phred-score для позиции каждого нуклеотида в последовательностях

Итоговая задача курса



Итоговая задача курса

All visual statistic:



Итоговая задача курса

Общее описание: необходимо написать программу, которая принимает fastq файл и на выходе даёт набор базовых статистик и графиков.

Вывод статистики

- создать таблицу DataFrame со всей статистикой
- если производилось удаление адаптерных последовательностей, то добавить их в статистику
- сохранить эту таблицу также в отдельный файл
- сохранить графики по отдельности
- добавить возможность создания subplots из всех графиков

Итоговая задача курса

Общее описание: необходимо написать программу, которая принимает fastq файл и на выходе даёт набор базовых статистик и графиков.

Визуализация в собственном геномном браузере

- Для 10 ридов, провести визуализацию рида, а также визуализировать адаптерную последовательность там, где она находилась у рида до удаления;

Итоговая задача курса

Общее описание: необходимо написать программу, которая принимает fastq файл и на выходе даёт набор базовых статистик и графиков.

Получение информации из базы данных GEO

- Получить и распарсить карточку из базы данных GEO;
- Из HTML распарсить поля в которых будет описан вид, обработка в эксперименте, тип эксперимента, генотип, ссылка на сырые данные, ссылка на саму базу данных;

Итоговая задача курса

Общее описание: необходимо написать программу, которая принимает fastq файл и на выходе даёт набор базовых статистик и графиков.

Запись результатов в базу данных SQL

- Создать базу данных для хранения информации;
- В базе данных создать 2 таблицы, одну таблицу с рассчитанной статистикой по ридам, вторую таблицу с полученной метаинформацией из GSE;
- Сделать таблицы таким образом, чтобы они были связаны между собой, т.е. в обеих таблицах был общий идентификатор, по которому данные из первой таблицы можно сопоставить с данными во второй таблице;