

Мини-обзор бактерии или археи, генома и протеома

Электронные таблицы для получения и
представления результатов

Содержание практикума

1. Про мини-обзор
2. Темы для выбора и включения в миниОбзор
3. Гистограмма
4. Демонстрация.
 1. Построение гистограммы длин белков.
 2. Разное – выделение строк по свойствам
5. Задание для выполнения в классе. Построить гистограмму длин белков своей бактерии или археи
6. Подсказки по построению гистограммы
7. Введение в мини-обзоре
 1. Wiki eng.
 2. Pubmed интро.
8. ДЗ
9. Инструкции по ДЗ.
 - Словарик
 - Список формул и методов.

1. Мини-обзор про бактерию или архею с исследованием её генома и протеома

Мини-обзор это мини-мини проект, строящийся аналогично курсовым: Связный и понятный текст со своими результатами. В тексте обзора

1. НАЗВАНИЕ и АВТОР (студент)
2. Немножко про выбранный организм, что сумеете найти (ВВЕДЕНИЕ)
3. РЕЗУЛЬТАТЫ полученные вами в виде текста, таблиц и рисунков. **НЕ МЕНЕЕ ТРЁХ**
4. Краткое перечисление МЕТОДОВ
 1. электронные таблицы
 2. программы пакета EMBOSS и/или сервисы on-line
 3. программы на python
5. Сопроводительные материалы – ссылки на
 1. большие таблицы Google sheets
 2. собственные программы на python

Зачёт мини-обзора обязателен для зачёта биоинформатики за семестр

1. Мини-обзор про бактерию или архею с исследованием её генома и протеома

Результаты делятся на

1. Входящие в обязательные задания практикумов
2. Выбранные студентом из предложенных тем
3. Придуманные студентом самостоятельно – ценятся ВЫШЕ остальных.

Что ценится проверяющими

1. Творчество студента, интересные наблюдения и гипотезы даже в обязательных заданиях
2. Простота и понятность изложения обзора.

Контрольные точки по мини-обзору

- 3 ноября 2023 – обязательное задание по мини-обзору в блоке ЭТ (гистограмма длин белков)
- Создание Google doc для мини обзора
- **10** ноября - введение к мини обзору
- **Начало** декабря – первый вариант мини-обзора. Он проверяется и преподаватель указывает ошибки и комментирует недочёты, требующие исправления
- декабря – сдача мини обзора на проверку без потери баллов
- зачётная сессия, последний шанс получить зачёт за мини обзор

2. Темы исследований для мини-обзора

1. Гистограмма длин белков – **обязательное**. Какие белки самые короткие и самые длинные?
2. Распределение генов белков и разных типов РНК по репликонам.
3. Исследование межгенных промежутков (!)
4. Исследование пересекающихся генов белков (!!!)
5. Предсказание и исследование оперонов (!)
6. Исследование старт-кодонов генов белков
7. Определение частот стоп-кодонов генов
8. Исследование состава нуклеотидов А, Т, G, С по репликонам
9. Исследование частоты динуклеотидов по репликонам.
Динуклеотид – два нуклеотида подряд.
10. Мини-обзор литературы о выбранном организме (Введение)
11. Сравнение с геномами других штаммов того же вида или рода (!)

Что есть в хромосомной таблице

- Список генов белков
 - ID гена в геноме: locus_tag
 - Координаты гена в репликоне
репликон это молекула ДНК: хромосома ИЛИ плаزمиды ИЛИ другая ДНК
 - Идентификатор белка в базе данных о последовательностях белков (пример WP_012619398.1)
 - Название белка. Короткий код белка, может быть полезен для сравнения с белками близких штаммов или видов того же рода
 - Примечания (например, “псевдоген”)
- Список генов РНК
 - ID гена в репликоне: locus_tag
 - Координаты гена в репликоне
 - Класс РНК
 - Иногда – короткое название

Чего нет в хромосомной таблице

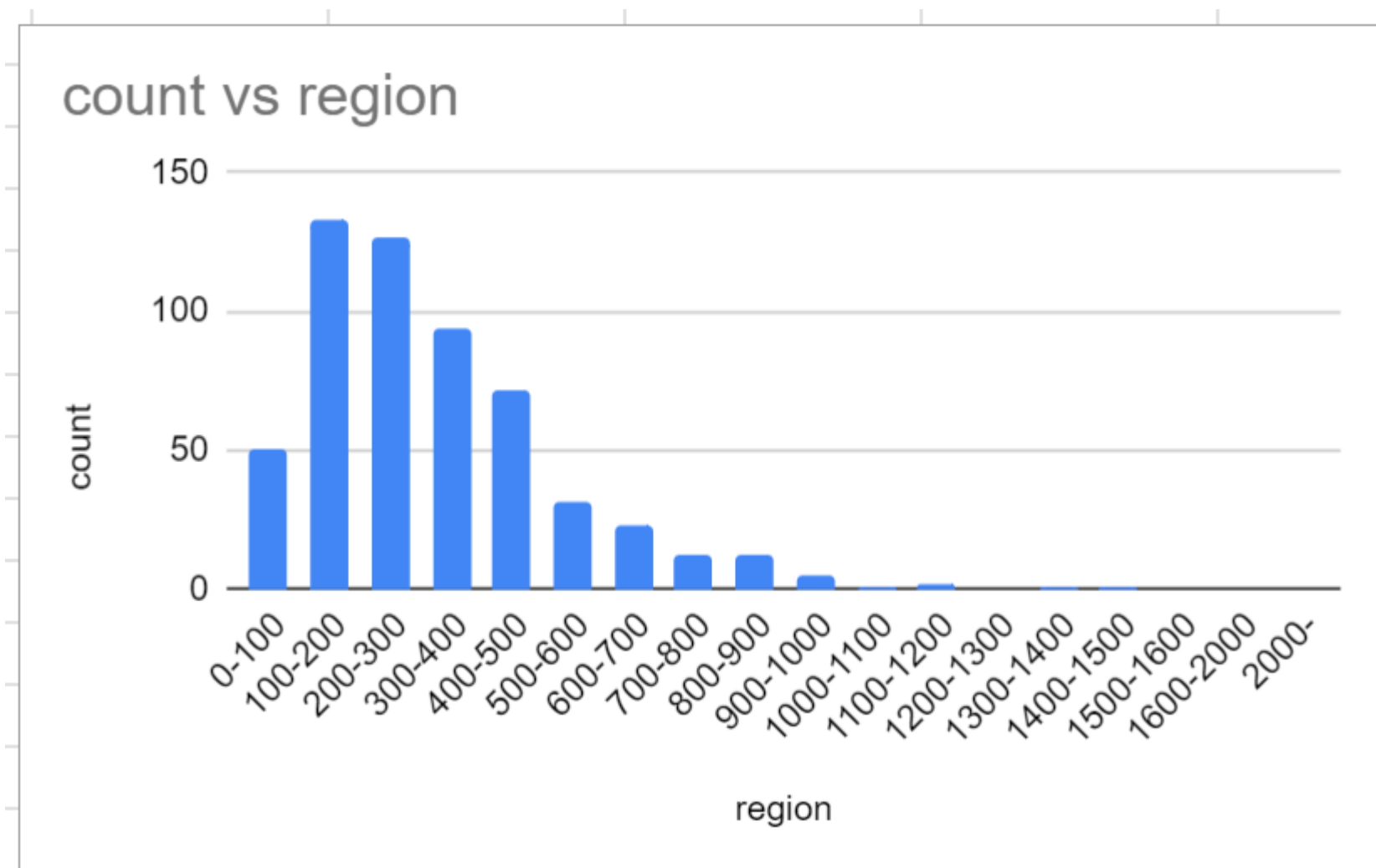
- НИЧЕГО, кроме перечисленного выше
 - Нет последовательности гена
 - Нет последовательности белка, закодированного в гене
 - Нет оперонов

ОПЕРОН это последовательность генов бактерии или археи, транслирующаяся в одну мРНК. Значит, гены оперона (i) расположены на одной и той же цепочке (ii) межгенные интервалы невелики (iii) функции белков связаны

База оперонов в геномах бактерий <https://operondb.jp/>

Программа для предсказания оперонов в прокариотической геномной ДНК – репликоне https://biocomputo.ibt.unam.mx/operon_mapper/

3. Гистограмма длин белков по хромосомной таблице генома *Buchnera aphidicola* str. JF99



4. Демонстрация

1. Построения гистограммы длин белков
 2. Разного

5. Задание для выполнения в классе

Построить гистограмму длин белков из протеома своей бактерии или археи

6. Подсказка по выполнению (на основе демонстрации)

- Размер кармана – константа, которую можно менять

- Адресация с фиксированием при распространении
=C2+\$B\$2

- соединение текстов =C2&"-"&C3

- Счёт с несколькими условиями

=COUNTIFS(proteins_CDS!U:U,">"&C2, proteins_CDS!U:U,"<="&C3)

рус =СЧЁТЕСЛИМН ()

	B	C	D	E
region_size				
border				
region				
	100	0	0-100	51
		100	100-200	133
		200	200-300	127

Получить рис. Как на предыдущей странице:

Insert => Chart => Customize (для подписей)

7. Введение в мини-обзор

Про организм на уровне штамма или вида (wiki, eng., rus.)

1. Таксономия: вид, род, семейство

2. Где обитает

3. Чем примечательна

4. Введение заканчивается описанием того, какие интересные результаты получены автором. Пишется после получения и описания результатов.

ИСТОЧНИКИ информации:

- Wiki eng. сначала, потом rus
- Pubmed (см. Pubmed wiki)
- NCBI (National Center for Biotechnology Information)
All databases, поиск по виду.

Пока вам будет трудно ориентироваться, но таксономию найдёте.

7.2. Pubmed (<https://pubmed.ncbi.nlm.nih.gov/>) Intro

- Простой поиск: operon mapper
- Поиск по названиям статей: Buchnera aphidicola[ti]
- Полезные фильтры (слева на странице поиска):
 - Free full text
 - Review
- Рекомендуемое имя скачанной статьи в файле .pdf:
 - XXXXXXX-NNNN(WWWW)
 - XXXXXXX – фамилия первого автора
 - NNNN - год выхода статьи
 - WWWWWWWWWW – о чем, чтобы самому вспомнить
 - Пример: Taboada-2018(operon_mapper).pdf
- Поиск по отдельным полям. Найдём статью из примера:
 - Taboada[1au] 2018[dp]
 - К автору можно добавить 1ю букву имени Taboada B[1au]
 - Убрать требование, что он 1й: Taboada B[au]
 - Расширить диапазон дат: 2017:2019[dp]
 - Указания других полей (tags) см.
<https://pubmed.ncbi.nlm.nih.gov/help/#using-search-field-tags>

Конец презентации

Приёмы для статистики старт и СТОП КОДОНОВ

- Названия столбцов
- Нумерация строк исходной таблицы
- Пометка строк
 - Отметка строк с названиями
 - Нумерация строк внутри одной последовательности
 - Пометка нужных строк
 - Длины строк
- Разбиение по столбцам
- Сортировка
- Фильтрация
- Распространение формул
- Специальная вставка, формула в примечание к названию колонки
- Копирование и вставка