

Мини-обзор бактерии или археи, генома и протеома

Электронные таблицы для получения и
представления результатов

1.Среднее и не только

Пр.7 Про коров и молоко

- Кое-кто из студентов заметил, что средние содержания % жирности и % белка в молоке в стаде не совпадают с приведёнными в исходной таблице. Им премиальный балл!

КАК ТАК! Кто считать не умеет? Предлагайте объяснения.

Пр.7 Про коров и молоко

- Кое-кто из студентов заметил, что средние содержания % жирности и % белка в молоке в стаде не совпадают с приведёнными в исходной таблице. Им премиальный балл!

КАК ТАК! Кто считать не умеет? Предлагайте объяснения.

- кроме СРЕДНЕГО
- есть МЕДИАНА
- и другие интегральные характеристики числовых данных

Среднее реагирует на “выскочек”

Медиана не Врет устойчива к выбросам

2	2	4	4
4	6	8	10

медиана = 4 ✓
хорошо списывает
«типичного представителя»

2	2	4	4
4	6	8	1002

↑
Выброс!

медиана = 4 ✓
хорошо списывает
«типичного представителя»
чтобы получить медиану,
расставьте элементы
по возрастанию и возьмите
тот, что посередине

Медиана – это число,
которое является серединой
множества чисел.

Посчитайте среднее –
Это просто даже в
уме можно

Средняя зарплата - при
наличии миллиардеров ...

2. Мини-обзор

Зачёт его обязателен для получения зачёта за семестр

2.1. Мини-обзор это текстовый файл в формате .pdf с рисунками и таблицами

- Краткое описание выбранного вида бактерии или археи по литературным данным (Введение)
- Ваши результаты, полученные с помощью изучения генома и протеома выбранного вида (Результаты)
 - **Не менее ... результатов.** Каждый результат представлен в отдельном озаглавленном разделе
 - Обязательный раздел – гистограмма длин белков протеома
 - Результаты из списка предложенных тем
 - Самостоятельно придуманные исследование
- Краткое описание использованных в работе методов (Материалы и методы)
- Список использованной литературы

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. >ТРЕХ
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.

Разделы Мини-обзора

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

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.

Как искать литературу для Введения

- БД Pubmed (<https://pubmed.ncbi.nlm.nih.gov>)
 - Вводите слова для поиска
 - Уточнение где искать:
 - E.coli[ti] в названии статьи
 - Genome [ab] в аннотации
 - Можно и так [tiab] в аннотации и в названии
 - Livenskyi A [au] среди авторов статьи
 - 2023:2024[dp]
 - Пример комбинированного запроса
PARIS[tiab] Livenskyi A[au] 2023:2024[dp]
 - Есть фильтры списка находок. Для вас важен фильтр Free text

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>

Контрольные точки по мини-обзору

- 3 ноября 2024 – 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

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 длин 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 плазида 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”). ГЕН CDS не кодирующий белок!

- 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. Задание для выполнения в классе

- На добавление в таблицу данных из другой таблицы.
Функция vlookup (ВПР)

3.1. Данные для vlookup(ВПР)

- ДНК метилтрансферазы прокариот
- Функция
- REBASE

3.2. Разбить feature tables на три таблицы на НОВЫХ листах

- CDS кодирующие белки
- CDS псевдогенов
- Гены РНК

Демонстрация

Конец презентации