

ОБЗОР ГЕНОМА И ПРОТЕОМА STAPHYLOCOCCUS EQUORUM

1.

Конаныхин Максим Владимирович

2. Резюме

Цель работы изучить нуклеотидный состав цепочки ДНК *Staphylococcus equorum*, проверить работу второго правила Чаргаффа, наличие нуклеотидов отличных от аденина, цитозина, гуанина, тимина. Изучить процентное соотношение трех стоп-кодонов и рассмотреть соотношение белков по длинам (количеству аминокислот). Исследование статистики длин белков.

3. Ключевые слова

Нуклеотидный состав, длины белка, стоп-кодоны.

4. Введение

Staphylococcus equorum — неподвижные грамположительные кокки, диаметр клетки которых составляет от 0,6 до 1,2 мкм. Для представителей вида характерно деление в нескольких плоскостях, результатом чего есть расположение микробных клеток «виноградными гроздьями» в чистой культуре. Стафилококки — факультативные анаэробы, хемоорганотрофы с окислительным и ферментативным типом метаболизма, каталанзопозитивные и оксидонегативные. Формы колоний на плотных питательных средах — округлые, выпуклые, пигментированные (белые, желтые, золотистые). На жидких — равномерное помутнение.

Патогенные стафилококки продуцируют эндо- и экзотоксины, ферменты, нарушающие жизнедеятельность клеток.

Известна достаточно высокая чувствительность стафилококков к водным растворам солей серебра и его электролитическим растворам

Классификация:

Domain: Bacteria

Phylum: Bacillota

Class: Bacilli

Order: Bacillales

Family: Staphylococcaceae

Genus: *Staphylococcus*

Species: *S. equorum*

В этом обзоре мы изучим эту бактерию с помощью электронных таблиц, языка программирования Питон 3, и данных о геноме этой бактерии.

5. Материалы и методы

Данные о геноме который исследовался в этой работе с сайта ncbi с сайта Национального Центра Биотехнологической информации (NCBI).

Перейдем к “инструментам” с помощью которых обрабатывали данные из упомянутых выше таблиц.

Для получения таблиц 1 и 2 использовались программы 1 и 2 соответственно написанные на языке python 3

Для получения гистограммы длин белка использовались “Гугл таблицы”. В них применялась функция СЧЁТЕСЛИМН () для подсчета количества строк из столбца “длина белков” ячейки которых содержали подходящее для диапазона данного кармана число.

Для описательной статистики (таблица 3)

СРЗНАЧ()- рассчитывает среднее арифметическое из чисел входящих в диапазон ячеек

STDEV.P() – рассчитывает среднее квадратичное отклонение из чисел в выбранном диапазоне ячеек

МЕДИАНА() – выдает медианное значение из чисел в выбранном диапазоне ячеек

МАКС () – выдает максимальное число из чисел в выбранном диапазоне ячеек

МИН() – выдает минимум из чисел в выбранном диапазоне ячеек

Гистограмма в “Гугл таблице”. На первом листе из таблицы особенностей скопирован столбик с длиной генов. На втором листе в первом столбике указана длина белка (она указывалась арифметической прогрессией с шагом 29). В ячейках второго столбца количество белком длина которых лежит в диапазоне значений от значения в левой ячейке до значений в нижней от левой ячейки. (например в ячейки B6 указано количество белков длина которых лежит в диапазоне значений ячеек от A6 до A7). В последнюю ячейку второго столбца (B30) была “искусственно” вписано значение ≥ 813 .

6. Результаты и обсуждение

6.1 нуклеотиды

В цепи ДНК бактерии следующие содержание нуклеотидов:

В данном исследовании я попробовал найти необычный, отличный от стандартных А,Т,Г,С нуклеотид. Но к сожалению (а может и к счастью) он не был обнаружен. Также здесь я посчитал количество стандартных нуклеотидов и вот что получилось:

Таблица 1

Нуклеотид	Количество
Аденин	942 396
Гуанин	465 500
Цитозин	467 900
Тимин	946 397

Из таблицы 1 видно что правило Чаргаффа соблюдается как для аденина и тимина, так и для гуанина и цитозина. Модуль разности количеств аденина и тимина составляет 4001. Для цитозина и гуанина это значение составляет 2400.

6.2 стоп-кодона

Здесь будем считать стоп-кодона.

В стандартном генетическом коде существуют три стоп-кодона: TAG, TAA, TGA. Частота использования стоп-кодонов отличается для разных видов бактерий.

Частота использования трех стоп-кодонов для *Staphylococcus equorum*:

Таблица 2

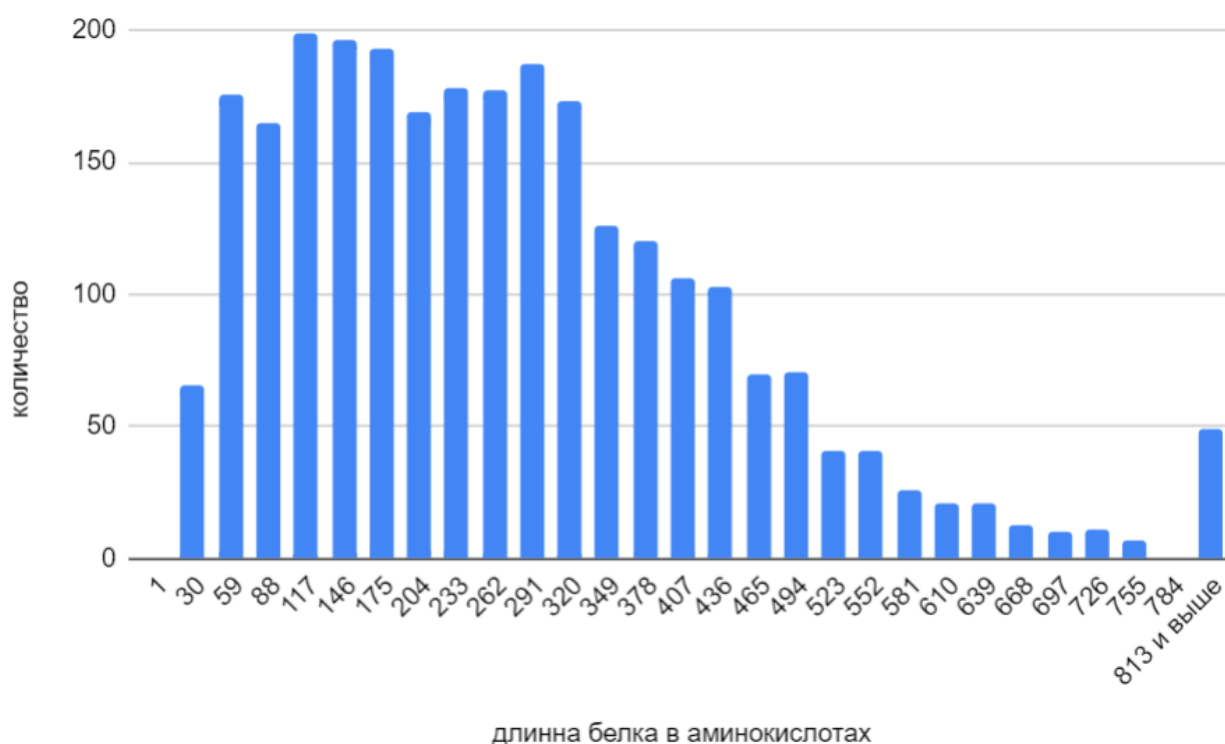
Стоп кодон	количество	частота %
TAA	32 789	49,82
TAG	13 563	20,61
TGA	19 466	29,58

Наиболее часто встречается TAA. Затем идет TGA , после TAG .

6.3

С помощью электронных таблиц и функции COUNTIFS можно построить гистограмму длин белков, кодируемых геномом бактерии:

Гистограмма распределения длин белка



Из диаграммы видно что для *Staphylococcus equorum* наиболее характерны белки длиной приблизительно от 45 до 320 аминокислот. Далее с уменьшением длины белков идет постепенное уменьшение их количества.

6.4 Описательная статистика белков

С помощью тех же электронных таблиц найдена некоторая статистическая информация о длине белка в аминокислотах.

Таблица 3

Среднее арифметическое	295
Среднее квадратичное отклонение	188
Медиана	267
Минимум	31
Максимум	1681

Из этой таблицы видно что среднее арифметическое близко к медиане(их модуль разности составляет 28 аминокислотных остатка).

Модуль разности между максимумом и минимумом составляет 1650 аминокислотный остаток.

7. Сопроводительные материалы

Геном:

https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/432/245/GCF_001432245.1_ASM143224v1/GCF_001432245.1_ASM143224v1_genomic.fna.gz

Протеом:

https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/432/245/GCF_001432245.1_ASM143224v1/GCF_001432245.1_ASM143224v1_feature_table.txt.gz

Код программы 1:

https://docs.google.com/document/d/1q1FDgNZpxhWU9IQ3mXgiC5zP_marOz3mqgFHegC7h-Y/edit?usp=sharing

Код программы 2:

https://drive.google.com/file/d/1C9FJYMMKFTFSFZVaiAEzba2tq5S_0NQr/view?usp=sharing

Гистограмма и описательная статистика в “Гугл таблицах”:

https://docs.google.com/spreadsheets/d/13fhyVf08Ua_PAih42EJRrW0VHfmJwIF8yfaGU2YzrR0/edit?usp=sharing

8. Список литературы

Информация о втором правиле Чаргаффа взята из интернет-ресурса livejournal из статьи

<https://galicarnax.livejournal.com/17558.html>

Ссылка на википедию

https://en.wikipedia.org/wiki/Staphylococcus_equorum