

Мини-обзор протеома бактерии *Aneurinibacillus*.

Барбашова София Валентиновна

Факультет биоинженерии и биоинформатики МГУ им. М. В. Ломоносова, ул. Ленинские Горы, 1с73, Москва, 119192

E-mail: barbashova@fbb.msu.ru

РЕЗЮМЕ

Работа посвящена анализу протеома бактерии *Aneurinibacillus* на основании файлов из базы данных NCBI [1].

Все расчеты и графики выполнены в Microsoft Excel 2016. С помощью данной программы я получила статистические данные о геноме бактерии: максимальное, минимальное, среднее значение длин белков, их медиану и стандартное отклонение; гистограмму длин белков; таблицу числа генов белков, псевдогенов и генов РНК на прямой и комплементарной цепи ДНК; таблицу числа генов по категориям.

Ключевые слова: Aneurinibacillus, bacterium, genome

1 ВВЕДЕНИЕ

1.1 Краткий обзор

Бациллы *Aneurinibacillus* были выделены из рода

Bacillus в отдельный род в 1996 году. Они являются аэробами или факультативными анаэробами, растут на простых питательных средах, таких как агар или соевый агат типтиказы. Способны к образованию эндоспор, устойчивых к неблагоприятным воздействиям (экстремальным температурам, высушиванию, химическим агентам)[2]. Рост происходит при рН от 5,5 до 9,0, а также в присутствии 2-5% NaCl. [3]

1.2 Роль в биотехнологии

Используются в генной инженерии в качестве хозяев для клонирования ДНК. Преимуществами, как хозяев для клонирования ДНК, являются способность секретировать интактные белки в окружающую среду, высокая технологичность, а также возможность длительного хранения промышленных штаммов в виде высушенных спор. [2]

2 МАТЕРИАЛЫ И МЕТОДЫ

2.1 Для анализа были использованы файлы GenBank сборки генома ASM143958v2 из базы данных NCBI [1].

Обработка данных осуществлялась с помощью функций MS Excel 2016.

2.2 При работе в Microsoft Excel 2016 применялись следующие встроенные функции:

=СЧЁТЕСЛИМН()

=МИН()

=МАКС()

=СРЗНАЧ()

=МЕДИАНА()

=СТАНДОТКЛОН()

3 РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

3.1 Распределение длин белков бактерии

Распределение длин белков бактерии *Aneurinibacillus* отражено на соответствующей гистограмме (Рис.1). Согласно данным, отраженным в гистограмме, справедливо утверждение о том, что чаще всего в геноме бактерии встречаются белки, длины которых принадлежат интервалу от 100 до 200 а.о. (точное количество – 809 белков) и интервалу от 200 до 300 а.о. (точное количество – 812). Данные о средней, максимальной и минимальной длине, а также о стандартном отклонении и медиане длин приведены в Таблице 1.

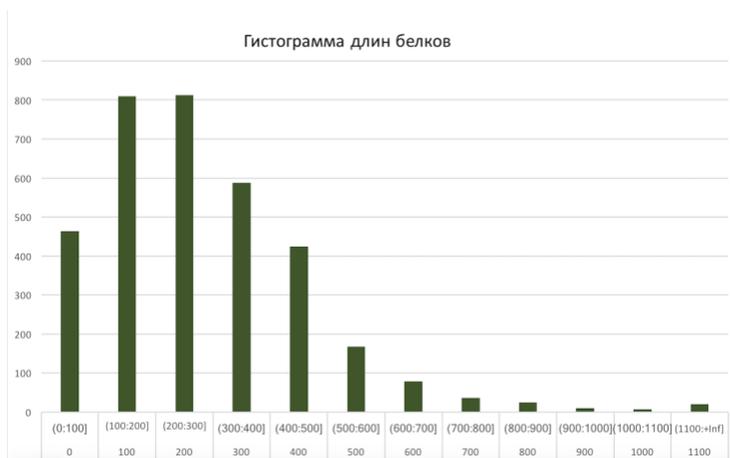


Рис.1 Гистограмма длин белков из протеома *Aneurinibacillus*. По вертикали – частота встречаемости белков, по горизонтали – интервалы длин белков (в аминокислотных остатках)

Средняя длина	286,9616391
Стандартное отклонение	185,0594758
Медиана	254
Максимальная длина	1913
Минимальная длина	37

Таблица 1. Данные о средней, максимальной и минимальной длине, стандартном отклонении и медиане длин белков.

3.2 Распределение генов на цепях ДНК

Результаты исследования распределения генов по разным цепям ДНК указаны в Таблице 2.

Цепь	Гены белков	Псевдогены	Гены РНК
Прямая	1856	45	24
Обратная	1585	40	131

Таблица 2. Распределение генов по разным цепям ДНК.

3.2 Число генов белков и генов РНК по категориям

Результаты исследования количества различных белков и генов РНК в геноме представлены в Таблице 3 и Таблице 4. Окончательная длина последовательности для расчета числа генов на 1 млн пар нуклеотидов (п.н.) найдена на странице бактерии на сайте NCBI [1].

Класс белка	Число	Оценка числа генов на 1 млн п.н.
Рибосомальный	89	24,28485866
Транспортный	170	46,38680868
Гипотетический	1652	450,770635
Другой	1530	417,4812781

Таблица 3. Распределение белков по классам

Класс РНК	Число	Оценка числа генов на 1 млн п.н.
тРНК	114	31,10644818
рРНК	34	9,277361737
Другой	9	2,455772224

Таблица 2. Распределение РНК по классам

4 ЗАКЛЮЧЕНИЕ

В результате анализа протеома *Aneurinibacillus* было обнаружено, что гены по комплементарным цепям распределены не равномерно. Число генов белков на прямой цепи превосходит число генов на обратной цепи, однако численность генов РНК на обратной цепи заметно больше числа генов РНК на прямой цепи.

Большая часть белков в геноме бактерии имеет длины, принадлежащие интервалу от 100 до 300 а.о. По гисторамме нетрудно заметить, что, начиная с максимального значения, количество более длинных белков уменьшается. Можно предположить, что геному «невыгодны» слишком длинные белки, а наиболее часто встречаемая длина – самая оптимальная для представленного организма.

5 СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

http://kodomofbb.msu.ru/~barbashova/term1/Supplementary_materials.xlsx

6 БЛАГОДАРНОСТИ

Выражаю благодарность преподавательскому составу курса практической информатики факультета биоинженерии и биоинформатики МГУ им. М. В. Ломоносова.

7 СПИСОК ЛИТЕРАТУРЫ

[1] NCBI:

https://www.ncbi.nlm.nih.gov/assembly/GCF_001439585.2

[2] Часичное цитирование статьи

[https://ru.wikipedia.org/wiki/Бациллы_\(род\)](https://ru.wikipedia.org/wiki/Бациллы_(род))

[3] Logan, NA и Vos, PD 2015. *Aneurinibacillus*. Bergey's Manual of Systematics of Archaea and Bacteria, 1-13.