

Обзор протеома бактерии *Sorangium cellulosum* So ce 56

Шугаева Т.Е.

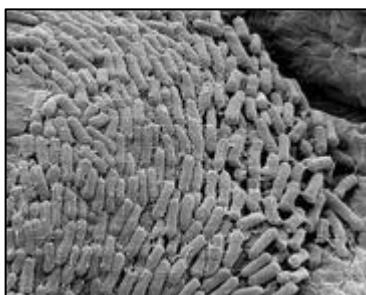
МГУ, 2014 год

Резюме.

Был проведен анализ протеома бактерии *Sorangium cellulosum* So ce 56: изучено распределение длин белков и расположение генов на цепях ДНК. Для получения статистических данных была использована программа Microsoft Excel (версия 2010 года). Полученные результаты оформлены в виде таблиц (см. Сопроводительные материалы), данные по распределению длин белков представлены в виде гистограммы (см. рисунок 1. Распределение длин белков.).

1. ВВЕДЕНИЕ

Бактерия *Sorangium cellulosum* относится к группе миксобактерий. Для данной группы характерно наличие относительно больших для прокариот геномов (9-10 миллионов пар нуклеотидов). В 2007 году геном бактерии *Sorangium cellulosum* был самым большим известным бактериальным геномом.



2. Фото бактерии *Sorangium cellulosum*
(электронная микроскопия)

Генетическая информация в клетке бактерии представлена кольцевой ДНК, содержащей 13033779 пар нуклеотидов. Всего в геноме закодировано 7426 белков и 319 РНК.

Был проведен анализ генома, в результате которого были получены статистические данные по распределению длин белков данной бактерии и расположению генов на прямой и обратной цепях ДНК. Цель работы заключалась в исследовании протеома *Sorangium cellulosum* So ce 56.

2. МЕТОДЫ

Общая информация о бактерии *Sorangium cellulosum* была получена из свободной универсальной интернет-энциклопедии Wikipedia [1].

Описание протеома бактерии было взято с сервера NCBI [2].

Фотография колонии бактерий была взята с сайта IDW [3].

Обработка данных производилась с помощью программы Microsoft Excel 2010. В процессе создания таблиц были использованы простейшие формулы, функция "СЧЁТЕСЛИМН", позволяющая рассчитать количество ячеек, удовлетворяющих нескольким условиям, а также функция "СЧЕТЕСЛИ", сходная с предыдущей (считает количество непустых ячеек, удовлетворяющих единственному условию). Данные функции понадобились для расчёта количества белков в заданном диапазоне и для расчёта количества генов на разных цепях ДНК.

Была использована доступная в Excel функция создания гистограммы.

3. РЕЗУЛЬТАТЫ

3.1 Распределение длин белков

Распределение длин белков бактерии *Sorangium cellulosum* So ce 56 представлено в виде гистограммы (см. рисунок 2. Распределение длин белков).



3. Распределение длин белков.

На горизонтальной оси указаны диапазоны длин (в а.о.), по вертикали указано количество белков.

Наибольший процент белков попадает в диапазон 300-350 а.о. (10, 9% от общего числа белков). Близки к этому

значения для диапазонов 250-300, 200-250 и 150-200 (10.4%, 10% и 9.9% соответственно).

3.2 Распределение генов по прямой и обратной цепям

Распределение генов по прямой и обратной цепям представлено в таблице 1.

Таблица 1. Распределение генов по цепям

	Белки	РНК
Прямая цепь	3739	276
Обратная цепь	3687	43
Всего	7426	319

Отдельно рассматривались гены, кодирующие белки, и гены, кодирующие РНК. Для генов, кодирующих белки соотношение 1:1. Для РНК 6:1.

4. ОБСУЖДЕНИЕ

Полученное распределение длин белков можно считать нормой для бактериального организма. Более 55% входят в диапазон от 100 до 400 а.о. Чуть более одного процента имеют длину более 1500 а.о. Максимальная длина белка 8417 а.о.

Нетипично соотношение генов, кодирующих РНК, на прямой и обратной цепях. Видно, что почти все гены на обратной цепи действительно отвечают за синтез молекул РНК, в то время как большинство генов на прямой цепи продукта не имеет.

5. СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Данные по протеому, гистограмму распределения длин белков и таблицу с расположением генов на цепях можно найти в файле [Sorangium_cellulosum_proteome.xlsx](#).

6. БЛАГОДАРНОСТИ

Выражаю огромную благодарность всем преподавателям факультета биоинженерии и биоинформатики МГУ.

7. СПИСОК ЛИТЕРАТУРЫ

- [1] http://en.wikipedia.org/wiki/Sorangium_cellulosum
- [2] ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/Sorangium_cellulosum_So_ce_56_uid61629
- [3] <https://idw-online.de/pages/en/news490849>