

Обзор протеома бактерии *Burkholderia ceracia* JBK9

Власов Антон¹

¹Факультет биоинженерии и биоинформатики МГУ им. М. В. Ломоносова.

АННОТАЦИЯ

В данной работе был исследован протеом бактерии *Burkholderia ceracia* JBK9. Были изучены особенности распределения длин белков, закономерности распределения белок-кодирующих и РНК-кодирующих генов в зависимости от направленности цепочки ДНК, на которой они находятся, а также пересечения белков.

Ключевые слова: *Burkholderia ceracia*, протеом, бактерия, Excel.

Дата публикации: 14 декабря 2017.

1 ВВЕДЕНИЕ

Классификация

Домен	Prokaryota
Царство	Bacteria
Тип	Proteobacteria
Класс	Betaproteobacteria
Порядок	Burkholderiales
Семейство	Burkholderiaceae
Род	<i>Burkholderia</i>
Вид	<i>B. ceracia</i>

Бактерия *Burkholderia ceracia* является условно-патогенным организмом, который в основном поражает лёгкие больных кистозным фиброзом, повышая смертность, также она может быть найдена повсеместно в окружающей среде, в том числе является патогенной для растений (поражает растения лука и табака). Длина генома (в парах нуклеотидов) – 8 481 212. Количество генов – 7 822.

Геном был секвенирован в 2015 году на факультете прикладной биологии Национального университета Кёнгука, Южная Корея [1]. Сама бактерия была открыта в середине XX века (1949 - Walter H. Burkholder) и уже тогда была признана патогенной для человека [2].

2 МАТЕРИАЛЫ И МЕТОДЫ

Данные о протеоме бактерии были взяты с сайта NCBI [3]. Для анализа генетической информации бактерии использовался файл *GCF_000701165.1_ASM70116v2_feature_table.txt*, содержащий информацию о белках, генах и РНК организма.

Данные были обработаны с помощью электронной таблицы Microsoft Excel 2016. Количество генов белков и РНК подсчитывалось с помощью функций СЧЁТЕСЛИМН(). Для подсчета белков разных типов использовался фильтр по содержанию в колонке `name`. Транспортные белки выбирались по ключевому слову “transport” (все белки, судя по их именам, выполняют транспортную функцию, дополнительный отбор не потребовался), рибосомальные – “S ribosomal”,

поскольку не все белки, содержащие “ribosomal” в названии, непосредственно входят в состав рибосом.

Подсчет длин пересечений велся независимо для каждой из 3 хромосом, прямой и обратной цепей. Разделение производилось с помощью фильтра по колонкам `chromosome` и `strand`.

Сводная таблица была создана автоматически посредством встроенного функционала Excel. Среднее значение, медиана и стандартное отклонение вычислялись встроенными функциями редактора.

3 РЕЗУЛЬТАТЫ

3.1 Белки организма

В протеоме бактерии *B. ceracia* JBK9 всего закодировано 7476 белков, статистические данные которых представлены в таблице 1.

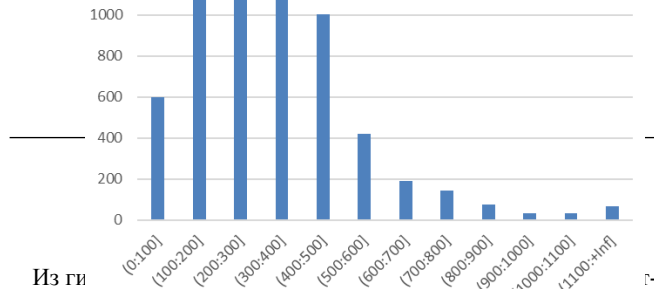
Таблица 1. Данные о длинах белков бактерии *B. ceracia* JBK9

Характеристика	Значение
Минимальная длина	26
Максимальная длина	5030
Средняя длина	326
Медиана длин	294
Стандартное отклонение	244,6

Информация о длинах белков представлена в числе аминокислот.

Распределение длин белков можно наблюдать на Диаграмме 1. На горизонтальной оси отложены длины белков с шагом 100 аминокислот, по вертикальной оси – количество белков с такой длиной в числе аминокислот.

Диаграмма 1. Гистограмма длин белков *B. ceracia*



Из ги
 ся белки длиной от 201 до 400 аминокислотных остатков, их количество колеблется от 1600 до 1781 а.к. Далее с увеличением длины количество белков уменьшается. Из таблицы 1 следует, что самый большой белок бактерии состоит из 5030 аминокислот.

В зависимости от функций закодированных белков, их можно разделить на 4 группы: рибосомальные, транспортные, гипотетические и остальные. Количества белков каждого типа приведены в Таблице 2.

Таблица 2. Группы белков бактерии *V.cerasia* JVK9

Тип	Число
Рибосомальные	78
Транспортные	829
Гипотетические	1381
Остальные	5188
Всего	7476

Из таблицы следует, что гипотетические белки составляют 18% от общего числа белков. Число рибосомальных белков почти в 100 раз меньше, чем число всех белков организма.

3.2 РНК бактерии

Всего в бактерии *B.cerasia* JVK9 71 ген кодирует различные РНК. В Таблице 3 приведены количества каждого вида РНК, закодированного в геноме.

Таблица 3. Количество генов, кодирующих разные виды РНК.

Обозначение	Расшифровка	Количество
tRNA	Транспортная	69
tmRNA	Транспортно-матричная	1
rRNA	Рибосомальная	18
ncRNA	Некодирующая	3

Из приведенной выше таблицы видно, что наиболее встречаемыми РНК являются транспортная и рибосомальная РНК. Также в протеоме содержатся гены трёх некодирующих РНК и один ген транспортно-матричной РНК.

3.3 Анализ протеома

Длина генома составляет 8 481 212 п.н. Всего в бактерии *B.cerasia* 7 822 гена. На 1 млн пар нуклеотидов приходится примерно 922 гена. Геном бактерии представлен тремя кольцевыми хромосомами. В Таблице 4 представлено количественное распределение генов по прямым и обратным цепям ДНК.

Таблица 4. Распределение генов по прямой и обратной цепи ДНК.

Цепь \ Тип гена	Гены белков	Псевдогены	Гены РНК
	Прямая	3579	125
Обратная	3897	130	47

Из таблицы следует, что немного больше генов закодировано на обратной цепи ДНК. Однако, в целом, число генов на прямой и обратной цепочках ДНК приблизительно равно, поэтому вывод о преобладании кодирования на одной из цепей сделать нельзя.

3.4 Пересечения белков

В ходе работы было подсчитано число пересечений белков на прямой и обратной цепях ДНК бактерии. Число пересечений на прямой цепи составило 361 а.к., а на обратной – 424.

Таблица 5. Данные о пересечениях белков бактерии *V.cerasia*

Характеристика	Прямая цепь	Обратная цепь
Минимальная длина*	3	3
Максимальная длина*	115	234
Средняя длина*	7,3	7,17
Медианная длина*	3	3
Число пересечений	361	424

*Информация о длинах пересечений белков представлена в числе аминокислот.

Количество пересечений различной длины с шагом 5 представлено на Гистограмме 2. На горизонтальной оси указаны интервалы длин, для которых велся подсчет количества белков, на вертикальной – количество белков в конкретном интервале.

Диаграмма 2. Гистограмма длин пересечений белков



Нетрудно заметить, что наиболее число пересечений на обратной цепи выше, чем на прямой. При этом большинство пересечений сосредоточено на интервале от 3 до 5 аминокислот. Об этом же свидетельствует медиана длин пересечений.

4 ОБСУЖДЕНИЕ И ЗАКЛЮЧЕНИЕ

1. Наиболее распространёнными белками *B.cepacia* являются белки с длиной от 200 до 400 а.к. Вероятно, именно такое преобладание длин белков характерно и для остальных бактерий.
2. Наличие почти 20% от общего числа белков показывает необходимость дальнейших исследований генома бактерии.
3. Функции некодирующих РНК в *B.cepacia* нуждаются в дальнейшем рассмотрении и изучении. Также вызывает интерес транспортно-матричная РНК бактерий.
4. При поиске рибосомальных белков было замечено, что у *B.cepacia* встречаются рибосомы с сегментацией 30S и 50S.
5. Вероятно, гены распределяются по прямой и обратной цепочкам случайно.
6. Пересечения белков в целом короткие, однако выделяются пересечения с длиной более 100 аминокислот. Интересны причины пересечения белков, а также механизм репликации на таких участках.

Работа над обзором протеома *B.cepacia* выявила интересные особенности устройства генома бактерий, которые сильно отличают их от эукариотических клеток.

5 СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Таблицы и расчёты представлены в файле: http://kodomofbb.msu.ru/~anton.vlasov/term1/pr13_vlasov.xlsx

БЛАГОДАРНОСТИ

Мне бы хотелось поблагодарить преподавателей факультета биоинженерии и биоинформатики за увлекательную возможность изучить протеом бактерии и сразу получить маленький опыт в практической биоинформатике. Также я благодарен одногруппникам, которые отвечали на мои глупые вопросы о бактериях, особенно Антону Иззи. Наконец, спасибо разработчикам Microsoft Excel за очень удобный инструмент.

6 СПИСОК ЛИТЕРАТУРЫ

- [1] <https://www.ncbi.nlm.nih.gov/genome/?term=Burkholderia+cepacia>
- [2] https://en.wikipedia.org/wiki/Burkholderia_cepacia_complex
- [3] ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/701/165/GCF_000701165.1_ASM70116v2