

Обзор генома бактерии *Clostridium beijerinckii* NCIMB 8052

Кравченко П.А.

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

20.12.2016

РЕЗЮМЕ

Данная работа была выполнена с целью изучения генома и протеома бактерии *Clostridium beijerinckii* NCIMB 8052.

Работа включает в себя общие закономерности состава генома *Clostridium beijerinckii*.

Было изучено распределение длин белков *Clostridium beijerinckii*, определено число генов, расположенных на прямой и обратной цепи ДНК. Проверена гипотеза случайного распределения генов на прямой и обратной цепи ДНК. В обзоре рассмотрено расположение квазиоперонов, тенденции их объединения, а также выявлена группа белковых продуктов, экспрессия которых ведётся особенно активно.

Для получения более подробной информации о GC составе был анализирован полный геном бактерии.

1. ВВЕДЕНИЕ

Clostridium beijerinckii - грамположительная, облигатно анаэробная бактерия, способная продуцировать эндоспоры. Геном *Clostridium beijerinckii* был секвенирован 27 июня 2007 года и содержит 6000632 пары нуклеотидов[4]. Весь генетический материал находится в одной хромосоме.

Таблица 1. Систематическое положение *Clostridium beijerinckii*

Domain:	Bacteria
Phylum:	Firmicutes
Class:	Clostridia
Order:	Clostridiales
Family:	Clostridiaceae
Genus:	<i>Clostridium</i>

У данной бактерии не обнаружено патогенных свойств, но некоторые представители класса Clostridia могут быть патогенны. Клостридии входят в состав нормофлоры желудочно-кишечного тракта. Иногда их обнаруживают в полости рта и на коже[4]. *Clostridium beijerinckii*, используя большое количество субстратов, способна жить на разнообразных средах. Она может быть использована в промышленности для получения бутанола, ацетона и изопропанола из пентоз, гексоз и крахмала[5].

Данный обзор написан с целью изучения генома и протеома бактерии *Clostridium beijerinckii*. Целью работы является выявления общих закономерностей её строения, таких как длины транскриптов, количество белков и РНК на прямой и комплементарной цепи ДНК, процентное содержание GC пар в генетическом материале. Поставлена задача проверки гипотезы равновероятного распределения генов по прямой и комплементарной цепям ДНК.

2. МЕТОДЫ

При написании обзора были использованы данные, полученные из базы данных NCBI*[2]. Для обработки и анализа информации была использована операционная система Windows и программы Microsoft Office Excel 2007 и FAR. В ходе работы были выполнены следующие операции с программным обеспечением:

Импортирование данных в электронные таблицы. Создание электронных таблиц. Автоматическое заполнение по образцу. Построение гистограммы распределения длин белков.

Функции: «СЧЁТЕСЛИ» для подсчета частоты встречаемости белков в данном диапазоне длин, «СЧЁТЕСЛИМН» для подсчета количества генов на прямой и обратной цепях, «СРЗНАЧ» для определения средней длины последовательностей, «МОДА» для определения моды длин последовательностей, «МИН» и «МАКС» для определения максимального и минимального значений соответственно, «МЕДИАНА» для определения медианы длин последовательностей, «СРОТКЛ» для определения отклонения от заданного параметра, «СЛУЧМЕЖДУ» для создания массива случайных данных.

Программа на языке Python для подсчёта количества пуриновых и пиримидиновых нуклеотидов.

3. РЕЗУЛЬТАТЫ

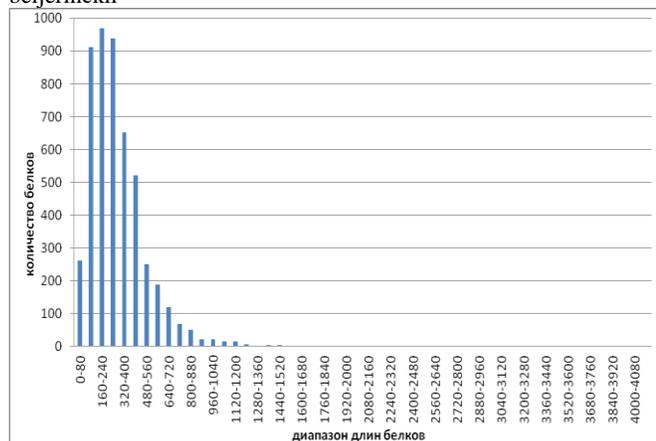
Работа с информацией, полученной из базы данных, показала, что в геноме *Clostridium beijerinckii* содержится 5231 значащих гена, из которых 5090 кодируют белки, а 141 кодируют РНК.

1. Распределение длин белков

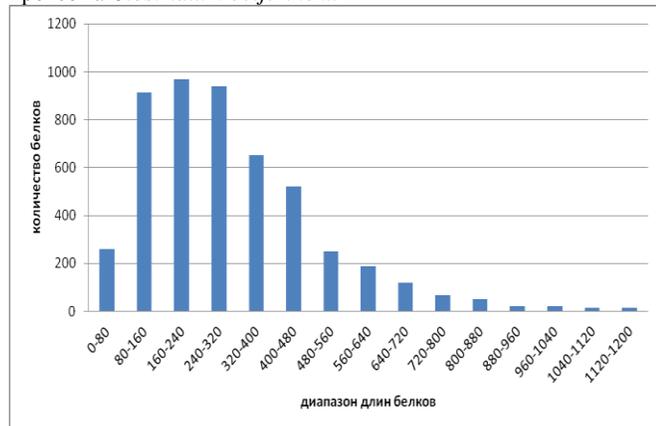
Гистограмма 1 иллюстрирует распределение длин белков протеома *Clostridium beijerinckii*, а гистограмма 2 иллюстрирует распределение длин наиболее многочисленных белков из протеома *Clostridium beijerinckii*.

Из гистограмм видно, что основная часть (84.45%) белков бактерии имеют длину в пределах от 80-и до 560-и аминокислотных остатков. Средняя длина белков равна 313-и, медиана длины 270-и, мода длины белков равна 150-и амк. остаткам. Самым коротким является белок, имеющий 37 амк. остатков. Длина самого длинного белка- 4034 амк. остатков.

Гистограмма 1. Длины белков из протеома *Clostridium beijerinckii*



Гистограмма 2. Длины наиболее многочисленных белков из протеома *Clostridium beijerinckii*



2. GC состав

Используя информацию из банка данных и написанную мной программу, удалось рассчитать процентное содержание GC пар в геноме *Clostridium beijerinckii*. Оно составило 29,856%. Выходные данные представлены в файле Excel(1) на листе "GC".

3. Распределение генов, кодирующих РНК и белки, по прямой и комплементарной цепи ДНК

Полученные мной данные о распределении белков и РНК по прямой и комплементарной цепи ДНК представлены в таблице 2. Таблица с ранжированным по распространённости списком белков находится на листе "Распространённость транскриптов" файла Excel(1). В геноме содержится несколько категорий РНК: rRNA (рибосомальная), tRNA (транспортная), ncRNA (некодирующая), misc_RNA(псевдогены). Таблица 3 иллюстрирует разбиение генов, кодирующих РНК (141 гена), на категории.

Таблица 2. Число генов белков и РНК на прямой и комплементарной цепи ДНК *Clostridium beijerinckii* NCIMB 8052

//	Прямая цепь	Обратная цепь
Белки	2650	2440
РНК	87	54

Таблица 3. Число генов РНК по категориям

Категория РНК	Число генов
rRNA	43
tRNA	94
ncRNA	1
misc_RNA	3

4. Проверка гипотезы равномерного распределения генов

На листе "Отклонение" файла Excel(1) проведены расчёты и найдена вероятность того, что в случайном эксперименте будет достигнуто отклонение, схожее с отклонением распределения генов на одну из цепей ДНК. Гипотеза равномерного распределения генов по прямой и комплементарной цепям ДНК подтвердилась.

5. Приблизительная оценка числа генов на 1 млн. пар нуклеотидов

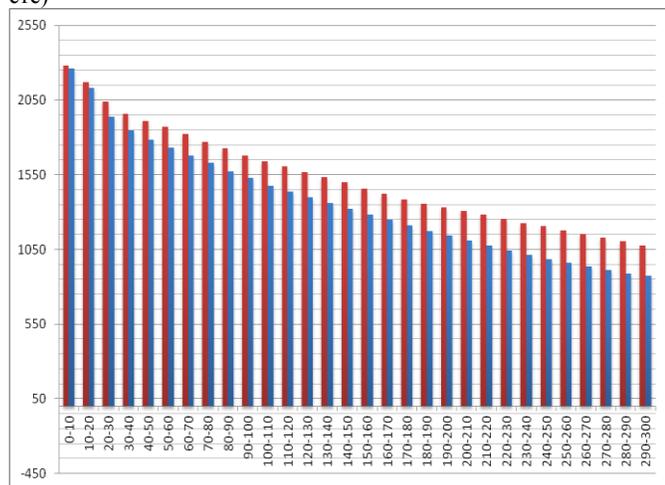
Приблизительная оценка числа генов на 1 млн. пар нуклеотидов даёт 872 нуклеотида на 1 ген.

6. Распределение квазиоперонов по прямой и обратной цепи ДНК

На гистограмме 3 показано распределение квазиоперонов на прямой и комплементарной цепи ДНК. Данные получены путём подсчёта количества генов, находя-

щихся на расстоянии N друг от друга, где N – максимальное количество нуклеотидов между генами, при котором оба входят в один квазиоперон. Синим цветом выделено количество квазиоперонов на прямой, а красным на комплементарной цепи. На оси абсцисс нанесена ранжированная шкала длин квазиоперонов, на оси ординат отмечено количество квазиоперонов, попадающих в данный промежуток. Очевидно, что на комплементарной цепи гены расположены ближе друг к другу и имеют более высокую тенденцию к образованию квазиоперонов.

Гистограмма 3. Квазиопероны на прямой и комплементарной цепи ДНК *Clostridium beijerinckii*. (Объяснения в тексте)



Clostridium beijerinckii имеет две кассеты генов - квазиоперонов, ответственных за сборку и работу жгутиков.

4. ОБСУЖДЕНИЕ

Полученные данные о распределении белков согласуются с общепринятыми представлениями о бактериальных протеинах, возможно, потому, что размеры бактерий не позволяют им иметь много особенно массивных белков. Однако в геноме представлен chromosomal replication initiator protein DnaA, имеющий длину 4034 амк. остатка.

Данный организм имеет 104 гена, кодирующих белки транскрипционной регуляции и 53- ДНК, связывающие регуляторные белки, что может свидетельствовать о высокой производительности механизмов репликации и транскрипции.

По наличию тРНК можно сделать вывод о том, что в белках данной бактерии не содержится таких аминокислотных остатков, как изолейцин, пирролизин, селеноцистеин, аспарагиновая к-та, глутаминовая к-та. В то же время, в геноме не представлено генов, кодирующих

тРНК с неопознанными амк. остатками. Однако в протеоме имеется белок Хаа-Pro aminopeptidase, обеспечивающий, по-видимому, посттрансляционные изменения белков *Clostridium beijerinckii*.

Бактерия имеет группу белков хемотаксиса и мембранных транспортёров, ответственных, скорее всего, за распознавание и улавливание пищи.

На листе “РНК” файла Excel(1) находится фрагмент таблицы из банка данных, по информации из которого можно предположить, какие функции выполняют в бактерии ncRNA и misc_RNA. Для работы bacterial RNase P необходим белок C5. Судя по данным из таблицы, C5 белок изменил свою функцию и стал метилтрансферазой, тогда как два продукта генов, получившихся при дупликации (misc_RNA), возможно, стали выполнять его функцию. Возможно, это связано с изменением субстрата, на котором живёт и питается *Clostridium beijerinckii*.

Низкий GC состав ДНК может говорить об относительно умеренности условий обитания бактерии.

Подтверждение гипотезы равновероятностного распределения генов по прямой и комплементарной цепям ДНК, возможно, говорит об отсутствии процессов, смещающих распределение.

Сложно сказать, почему гены на комплементарной цепи имеют большую, по сравнению с прямой цепью, тенденцию к объединению в квазиопероны. Возможно, это связано с работой транспозонов. В протеоме организма содержится 34 гена, кодирующего транспозазы.

5. СПИСОК ЛИТЕРАТУРЫ

- [1] <https://www.ncbi.nlm.nih.gov>
- [2] ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/016/965/GCF_000016965.1_ASM1696v1/GCF_000016965.1_ASM1696v1_feature_table.txt.gz
- [3] https://www.ncbi.nlm.nih.gov/assembly/GCF_000016965.1
- [4] <https://microbewiki.kenyon.edu/index.php/Clostridium>
- [5] Sun-Mi Lee, Min Ok Cho, Cheol Hee Park, Yun-Chul Chung, Ji Hyeon Kim, Byoung-In Sang and Youngsoon Um “Continuous Butanol Production Using Suspended and Immobilized *Clostridium beijerinckii* NCIMB 8052 with Supplementary Butyrate” *Energy Fuels*, 2008, 22 (5), pp 3459–3464 July 29, 2008

6. БЛАГОДАРНОСТИ

Автор выражает свою благодарность А.В.Алексеевскому за обучение основам биоинформатики и предоставление шаблона для написания обзора.

7. СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

- (1) Excel файл с перечисленными в тексте листами
http://kodomo.fbb.msu.ru/~pavel-kravchenko/prs/Kravchenko_supporting_materials.xlsx

8. Дополнения

*National Center for Biotechnology Information [1]