
Обзор протеома бактерии *Vibrio cholerae* 0395

Цветков Роман

2015 год

Факультет биоинженерии и биоинформатики МГУ

Аннотация: Холерный вибрион-одна из важнейших бактерий, представляющих угрозу здоровью людей. Описание ее генома и протеома - это задача, необходимая для четкого понимания механизма ее жизнедеятельности и дальнейшей разработки лекарств против холеры. Работа по исследованию протеома этой бактерии, выполнена с использованием базы данных NCBI. В результате получился анализ особенностей плазмид холерного вибриона, его белок-кодирующих и РНК-кодирующих последовательностей ДНК в 2 хромосомах бактерии. Было показано, что холерный вибрион - недавно возникшая и весьма прогрессивная бактерия, достигшая вершины идиоадаптации.

1 ВВЕДЕНИЕ

Сейчас холерный вибрион представляет большую угрозу здоровью людей на нашей планете. Понимание механизмов его жизнедеятельности-ключ к изобретению средств борьбы с ним. Для понимания, как функционирует *Vibrio cholerae* необходимо знать как устроены кодирующие последовательности вибриона, какие белки и РНК в ней закодированы. Для этого и необходимы анализ и исследование генома и протеома этой бактерии. Протеом-это совокупность всех белков, которые производит эта бактерия. Белки бактерии закодированы в ее ДНК, с которой транскрибируется РНК, с которой происходит трансляция белков. Такая РНК называется РНК-кодирующей. С транскрибированной РНК происходит трансляция белка. В работе представлен анализ последовательностей ДНК холерного вибриона.

2 МАТЕРИАЛЫ И МЕТОДЫ

Для выполнения этой работы я воспользовался программой Excel 2007, языком программирования Python 2.7.11, базой данных NCBI, программой Far Manager. Были скачаны через Net Vox Far Manager файлы со списком плазмид холерного вибриона и файл с закодированными в геноме белками. Далее я, используя Excel и Python построил гистограмму длин белков бактерии и составил таблицу числа генов белков и генов РНК на прямой и комплементарной цепочках ДНК. Я проверил гипотезу, что гены распределены по цепочкам случайно с вероятностью 0,5. Я разбил цепочку на 21 часть и для каждой части посчитал количество генов, далее вычислил среднюю для всех частей вероятность попадания гена на прямую или комплементарную цепь, посчитал среднее отклонение и дисперсию, которые оказались малыми. Гипотеза подтвердилась. Затем я

посчитал количество квазиоперонов(групп генов, каждый из которых удален от другого на расстоянии не более 100 нуклеотидов), составил таблицу генов имеющих пересечения. При этом я использовал Python 2.7.11. Также я составил таблицу плазмид известных прокариотических организмов с помощью Excel, используя функции MID и FIND, а также Python 2.7.11. В Excel я посчитал, используя функции COUNT, COUNTIF, MEDIAN, AVERAGE, MAX, MIN, среднюю длину всех известных плазмид, медиану их длин, длину самой длинной и самой короткой плазмиды. Род из названия организма я извлекал с помощью функции MID, записывая выражения вида =MID(H25;1;FIND(" ";H25)-1), выдававшие первое слово в ячейке(до первого пробела). Присваивая всем ячейкам одного столбца значения такого вида, я составил столбец с названиями родов организмов-хозяинов плазмид.

3 РЕЗУЛЬТАТЫ

В результате у меня получилась таблица генов белков и РНК в геноме холерного вибриона и таблица плазмид всех известных прокариотических организмов. На основе полученных данных я составил таблицы с анализом полученных данных.

3.1 Особенности плазмид *Vibrio cholerae* по сравнению с плазмидами других прокариотов

3.1.1 У меня получилась таблица генов всех известных плазмид прокариотических организмов, для каждой из которых указан род и вид организма-хозяина (Файл Plasmids.xlsx страница plasmids). Используя информацию, представленную в списке плазмид, я составил таблицу с числом плазмид для каждого рода и вида прокариотического организма(страница genera файла Plasmids.xlsx см. Сопроводительный материал 1). На странице info представлены выводы, из таблицы. Представлены данные о средней длине плазмид(58,26 тыс. нуклеотидов), медианной длине(34 тысяч нуклеотидов), максимальной(2351 тыс. нуклеотидов) и минимальной (1 тыс. нуклеотидов). Можно предположить, что преобладают плазмиды, длиной в несколько десятков нуклеотидов; короткие(меньше средней длины) встречаются чаще. Род *Vibrio* имеет как короткие плазмиды(длина не превышает 10 тыс. нуклеотидов), так и длинные плазмиды с длиной свыше 200 тыс. нуклеотидов. Максимальная длина 481, минимальная 3, средняя 74,8, медианная 55,5тыс. нуклеотидов. Таким образом, у *Vibrio* плазмиды в среднем длиннее, чем у остальных прокариотов, но короткие(меньше средней длины) как и у всех встречаются

чаще. Это также может говорить, что род *Vibrio* более прогрессивен и устроен более сложно, чем остальные прокариоты, что подтверждает гипотезу о недавнем происхождении рода *Vibrio* и вида *cholerae*.

3.1.2 Всего известно 6353 плазмиды. Плазмида представляет собой кольцевую ДНК. Гены, кодирующие рибосомальную РНК встречается редко, кодирующие транспортную практически никогда не встречаются. В частности, для рода *Vibrio* известно 57 плазмид, что намного больше медианного числа плазмид и больше среднего числа плазмид. Из этого можно сделать вывод, что род *Vibrio* находится в состоянии биологического прогресса и достиг успеха в эволюционной борьбе и, что, возможно, обусловлено тем, что он обладает эволюционными преимуществами. Также можно заметить, что среднее число плазмид у рода почти в 5 раз больше медианного, следовательно большинство родов прокариотов обладают 1-2 плазмидами, а меньшая часть родов имеют несколько десятков или несколько сотен плазмид, из этого следует вывод о нахождении большей части родов в состоянии биологического регресса, что неудивительно, учитывая древность группы прокариотов. Род *Vibrio* относится к последней группе. Вид *cholerae*, имеющий 10 плазмид, также можно назвать процветающим, так как все остальные виды рода *Vibrio* имеют меньшее число плазмид. Можно предположить, что род *Vibrio* возник позднее большинства других прокариот. Внешнее строение *Vibrio cholerae* можно увидеть на Рисунке1. [1]

Рисунок1. Внешнее строение холерного вибриона.



3.2 Особенности генома и протеома *Vibrio cholerae*

Также я составил таблицу генов белков и генов РНК в геноме холерного вибриона (лист Таблица генов в файле Results for proteins and RNAs of *Vibrio cholerae*). Ней представлены все гены в 2 хромосомах бактерии, кодирующие РНК или белки, с указанием хромосомы, к которой эти гены относятся, их длины, цепи ДНК (прямая или комплементарная) и продукта транскрипции с них. Помимо этого я посчитал число генов на прямой и комплементарной цепях (лист Число генов белков и генов РНК). С помощью Excel я построил гистограмму длин белков (лист Гистограмма длин белков и Картинка 1). Как видно из гистограммы и Таблицы1 белки в протеоме бактерии имеют длину чаще всего от 100 до 400 аминокислотных

остатков. Количество белков в зависимости от их длин можно узнать в Таблице1 и на Рисунке2.

Рисунок2. Гистограмма распределения длин белков.

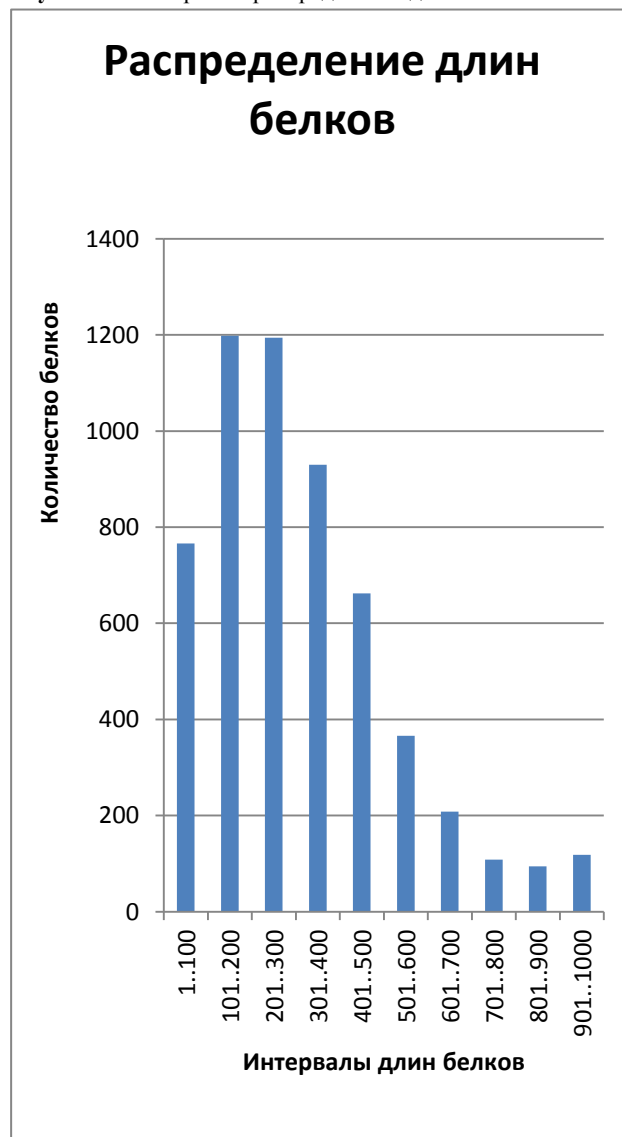


Таблица1. Количество белков определенной длины

Длина белков	1..100	101..200	201..300	301..400	401..500	Свыше 500
Количество белков	766	1198	1194	930	662	884

Картинка1. Гистограмма распределения длин белков

Я посчитал количество квазиоперонов в зависимости от порога нахождения пары генов в квазиопероне (Таблица 2 и лист Данные об особенностях генов). Как видно из Таблицы2, количество квазиоперонов при уменьшении порога вхождения гена в квазиоперон (то есть максимального расстояния в

нуклеотидах, при котором гены считаются в составе одного квазиоперона) число квазиоперонов растет, что неудивительно, так как чем большее расстояние может быть между генами, которые мы считаем, тем больше генов мы посчитаем [2].

Таблица2. Количество квазиоперонов в зависимости от порога

Порог	50	100	200
Количество квазиоперонов	1919	3033	4837

Также я с помощью Python 2.7.11 составил таблицу пересечений генов (лист Данные об особенностях генов) и обнаружил пересечений генов очень много, намного больше чем самих генов. Помимо этого я провел проверку гипотезы о том, что гены распределены между двумя цепями ДНК случайно с вероятностью 0.5 Для этого я разбил цепочку на 21 часть и для каждой части посчитал количество генов, далее вычислил среднюю для всех частей вероятность попадания гена на прямую или комплементарную цепь, посчитал среднее отклонение и дисперсию, которые оказались малыми (Таблица3). Гипотеза подтвердилась.

Таблица3. Проверка гипотезы случайного попадания гена на цепь.

Средняя вероятность попадания гена на определенную цепь	Дисперсия вероятности
50,88%	1,1%

Также я проверил с помощью Python 2.7.11 наличие генов длина которых не делится на 3. Таких генов не оказалось, следовательно можно предположить, что все нуклеотиды генома образуют триплеты генов (см. Сопроводит. Материал 2).

4 ОБСУЖДЕНИЕ

Поскольку род *Vibrio* имеет в среднем плазмиды большей длины, чем остальные прокариоты и большее число плазмид, то можно сделать следующие предположения:

- Род *Vibrio* находится в состоянии биологического прогресса.
- Род *Vibrio* возник недавно по эволюционным меркам и позже большинства прокариотов.
- Род *Vibrio* более прогрессивен и сложнее устроен, чем большинство остальных прокариотов.

Из таблицы генов РНК и белков *Vibrio cholerae* можно сделать следующие выводы:

- (1) Генов, кодирующих белки намного больше, чем генов, кодирующих РНК.
- (2) Генов, кодирующих РНК больше на прямой цепи, чем на комплементарной (76 на прямой, 46 на обратной), в то время как в среднем генов на прямой

цепи тоже немного больше (2072 гена на прямой цепи, 1984 на обратной).

- (3) Гены распределены между цепями ДНК случайно и практически равномерно.
- (4) В геноме очень много пересечений генов.
- (5) Генов, длина которых не делится на 3, нет.
- (6) Общее число генов около 4000.
- (7) У плазмид практически всегда отсутствует транспортная РНК.
- (8) Как можно заметить не вся ДНК входит в состав генов, есть участки, не кодирующие белок или РНК-интроны

В целом это характеризует *Vibrio cholerae*, как прогрессивный прокариотный организм, достигший вершины идиоадаптации и предельно специализированный и приспособленный к паразитизму. Все данные вписываются в современную научную картину мира. Очень большое число пересечений генов, наряду с наличием интронов иллюстрирует:

- Долгую эволюционную историю *Vibrio cholerae*, связанную с постоянными изменениями генома и генов, которое не всегда сопровождалось заменой или дополнением последовательности ДНК, а часто ограничивалось сдвигом рамок считывания генов, что часто приводило к сдвигу рамки считывания одного гена дальше рамки другого гена, то есть к формированию пересечения генов.
- Большую роль фактора случайности в эволюции: в геноме есть много пересечений генов, которые в случае повреждения выключают сразу несколько генов, что значительно опаснее для бактерии, чем выключение одного гена. Таким образом в ходе эволюции число пересечений генов должно было сократиться (возможно за счет замены интронов частью последовательности гена, входящего в состав пересекающихся генов и ликвидации пересечения).
- Подтверждение гипотезы об отсутствии "мусорной ДНК" в геноме и особой роли интронов, связанной с неизвестными сейчас функциями ДНК, не кодирующей белки: если бы интроны не выполняли никакой функции, то с течением времени мутантные особи, у которых произошла делеция интронов, получили бы эволюционное преимущество и вытеснили бы немутантных особей. Тогда количество интронов в геноме *Vibrio cholerae* было бы незначительным, а это не так.

БЛАГОДАРНОСТИ

Выражаю благодарность преподавателям биоинформатики на ФББ МГУ Алексею Владимировичу Алексеевскому, Сергею Александровичу Спирину, Дарье Владимировне Дибровой, Артуру Залевскому.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

1. [Таблица генов белков и РНК бактерии *Vibrio cholerae* 0395](#)

2. [Таблица плазмид](#)

ССЫЛКИ

1. **Википедия**
(https://en.wikipedia.org/wiki/vibrio_cholerae)
2. **Описание холерного вибриона на учебном сайте**
(http://kodomo.fbb.msu.ru/~tsvetcovroman/Vibrio_cholerae_0395)