
Мини-обзор протеома бактерии *Bacillus endophyticus*

Автор: Болихова Анастасия Кирилловна¹

¹Факультет биоинженерии и биоинформатики Московского государственного университета имени М.В.Ломоносова (1 курс)

РЕЗЮМЕ

В данном мини-обзоре приведены и проанализированы общие сведения о протеоме бактерии *Bacillus endophyticus*. Изучена закономерность распределения генов по прямой и обратной цепям ДНК, приведена гистограмма длин белков, дано количественное распределение генов по категориям (гены, кодирующие белки и гены, кодирующие РНК).

Ключевые слова: *Bacillus endophyticus*, протеом, РНК, псевдогены.

1 ВВЕДЕНИЕ

Bacillus endophyticus- штамм аэробных, эндоспорообразующих бактерий. Найден впервые в тканях здоровых хлопковых растений (*Gossypium sp.*, Dushanbe, Tajikistan). Выделен и классифицирован группой исследователей: Oleg N. Reva, Valerie V. Smirnov, Bertil Pettersson and Fergus G. Priest.

В ходе исследования было показано что данный штамм, а также еще три выделенных вместе с ним тесно связаны между собой и являются новыми видами рода *Bacillus*.

Статья о данном исследовании опубликована в International Journal of Systematic and Evolutionary Microbiology в 2002 году.

2 МАТЕРИАЛЫ И МЕТОДЫ

Для исследования протеома бактерии *Bacillus endophyticus* я пользовалась программой Excel 2013.

1) Для построения таблицы числа генов по категориям я:

- Для генов рибосомальных белков: посчитала все продукты трансляции ДНК, содержащие в названии “ribosomal protein”.
- Для генов транспортных белков: посчитала все продукты трансляции ДНК, содержащие в названии “transport”.
- Для генов гипотетических белков: посчитала все продукты трансляции ДНК, содержащие в названии “hypothetical”.
- Для генов остальных белков: посчитала все содержащиеся в имени “protein”, “transporter” или “transferase” и вычла сумму количеств рибосомальных, транспортных и гипотетических белков.
- Для генов транспортных РНК: посчитала все продукты транскрипции ДНК, содержащие в названии “ribosomal RNA”.
- Для генов рибосомальных РНК: посчитала все продукты транскрипции ДНК вида “tRNA-” + название нуклеотидного основания в трехбуквенном коде (выделила все содержащиеся “tRNA” и исключила из них

все ферменты, которые, например, можно узнать по окончанию “ase”).

2) Чтобы построить гистограмму белков бактерии я:

- Определила оптимальный размер карманов (интервалов значений) на гистограмме, для этого я нашла разность между наибольшей и наименьшей длиной белка в протеоме бактерии и разделила ее на выбранное мной количество интервалов (при разности между минимальной и максимальной длиной равной 4049 аминокислотам я посчитала наиболее удобным разбить эту величину на 100 интервалов).
- Затем я построила шкалу значений гистограммы по оси x- это возрастающий ряд чисел минимальное из которых равно минимальной длине белка, максимальное больше или равно максимальной длине белка, разность между каждыми двумя соседними числами равна размеру кармана, а общее количество чисел равно 101.
- В соответствие каждому значению шкалы (кроме последнего) я поставила количество белков в протеоме бактерии имеющих длину большую или равную данному значению и меньшую, чем следующее значение.
- По получившейся таблице из 2 столбцов и 100 строк (не берем последнюю строку так как она имеет только вспомогательное значение) построила гистограмму.

3) Для построения таблицы распределения генов по цепочкам я пользовалась функцией Excel СЧЁТЕСЛИМН, отбирая необходимые гены по двум критериям.

Первый это значение в столбце class (соответственно “protein_coding” для кодирующих белки, “RNase_P_RNA”, “tRNA” или “rRNA” для кодирующих РНК и “pseudogene” для псевдогенов).

Второй критерий это значение в столбце strand (“+” для прямой цепи и “-“ для обратной цепи).

3 РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

3.1 Количество генов в геноме по категориям (в зависимости от продукта их транскрипции или трансляции)

Геном — совокупность наследственного материала, заключенного в клетке организма. Геном содержит биологическую информацию, необходимую для построения и поддержания

организма, в нем закодирована структура белков и молекул РНК, необходимых для синтеза протеинов.

Белки в зависимости от выполняемых ими функций можно разделить на несколько основных категорий, это, например, рибосомальные или транспортные. Также по группам функциональным группам разбиваются и РНК, рРНК или рибосомальные РНК входят в состав рибосом вместе с рибосомальными белками. Транспортные РНК(тРНК) осуществляют транспорт определенных аминокислот к месту синтеза белка, то есть к рибосомам.

В таблице 1 приведено количество генов, кодирующих тот или иной тип РНК или белков.

В зависимости от того какое количество генов относится к определенной категории можно предположить сколько продуктов генов этой категории необходимо организму.

Таблица 1. Число генов в геноме бактерии *Bacillus endophyticus* по категориям.

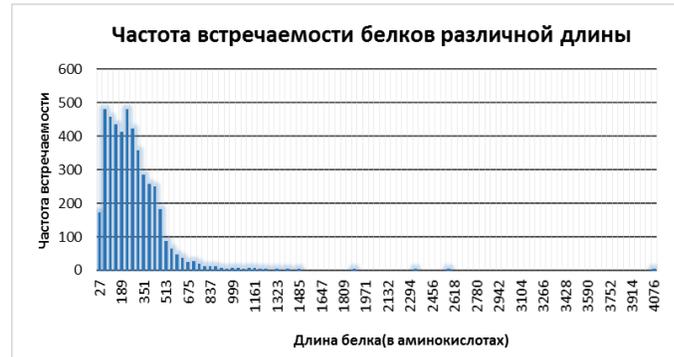
Категория	Число генов
Белки	
Рибосомальные белки	60
Транспортные белки	384
Гипотетические белки	1523
Все остальные	2269
РНК	
Транспортные	2328
Рибосомальные	385

Из таблицы видно, что число генов, кодирующих транспортные белки значительно превышает количество генов, кодирующих белки рибосомальные. Это возможно объясняется тем, что в клетке бактерии осуществляется множество видов транспорта, требующих большое количество транспортных белков разных типов. Рибосомы в клетке не несут большого различия между собой, а значит не требуют такого количества различных белков. Большое количество генов кодируют так называемые гипотетические белки, то есть белки с не определенной функцией из чего можно предположить, что молекулярное строение *Bacillus endophyticus* еще не очень хорошо изучено. Также значительно больше генов, кодирующих тРНК, чем генов, кодирующих рРНК. Это объясняется тем, что для синтеза обного белка необходима одна рибосома, состоящая из нескольких рРНК, и большое количество тРНК, а значит для функционирования клетке требуется большее количество тРНК, чем рРНК.

3.2 Нахождение белков различной длины в протеоме

Белки — высокомолекулярные органические вещества, состоящие из альфа-аминокислот, соединённых в цепочку пептидной связью. Аминокислотный состав белков определяется генетическим кодом. Длина белка чаще всего измеряется в аминокислотных остатках, гистограмм 1 наглядно де-

монстрирует количество белков той или иной длины в протеоме бактерии *Bacillus endophyticus*.



Гистограмма 1. Количество белков различной длины в протеоме бактерии *Bacillus endophyticus*.

Из гистограммы видно, что наибольшее количество белков в протеоме бактерии имеет длину примерно от 100 до 500 аминокислотных остатков. Хотя есть несколько “белков великанов”, имеющих длину более 1000 и даже более 4000 аминокислотных остатков.

Данные гистограммы обобщены в таблице 2, из которой мы видим, что точная длина самого большого белка равна 4076 аминокислотным остаткам, а длина самого маленького всего 27.

Длина белка в основном определяется его функциями. Можно предположить, что белки с наибольшей длиной выполняют более сложные функции. Это предположение частично подтверждается информацией, извлеченной из протеома бактерии. Наиболее длинные белки являются ферментами, обеспечивающими сложные процессы, например, белок с наибольшей длиной 4076 относится к классу редуктазы. Белки с наименьшей длиной в основном являются строительными белками, например, большинство белков мембраны имеет не большую длину. Белки же с самой маленькой длиной 27 аминокислотных остатков принимают участие в споруляции бактерии.

Таблица 2. Обобщенные данные о длине белков в протеоме бактерии *Bacillus endophyticus* (сопроводительная таблица к гистограмме 1).

Parameter	Value
min length	27
MAX length	4076
average length	286
standard deviation	192
median	257

3.3 Распределение генов в геноме по прямой и обратной цепям ДНК

ДНК у большинства организмов (кроме некоторых вирусов, содержащих одноцепочечную ДНК) это молекула состоящая из двух цепей нуклеотидов, соединенных между собой водородными связями. Цепи ДНК комплементарны между собой, но транскрипция каждого конкретного гена происходит только с одной из двух цепей.

В таблице 3 приведено распределение различных типов генов по прямой и обратной цепям ДНК.

Таблица 3. Распределение генов в геноме по прямой и обратной цепям ДНК.

цепь	protein_coding	RNA	pseudogene
Прямая цепь	2177	82	97
Обратная цепь	2355	34	82

Из таблицы данных видно, что между количеством генов кодирующих белки на прямой цепи и на обратной нет значительной разницы (различие составляет около 10%). Это также верно и для псевдогенов. Из чего можно заключить, что данные гены расположены на прямой и обратной цепи “случайным образом”, то есть в их распределении нет закономерности.

Значительная разница в количестве обнаруживается только между генами, кодирующими молекулы РНК. Число таких генов на прямой цепи более чем в два раза превосходит их число на обратной.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

http://kodomomo.fbb.msu.ru/~anastasia_b/book2.xlsx

БЛАГОДАРНОСТИ

Спасибо Порошину Владимиру Дмитриевичу за то, что поставил зачет по БЖД в начале декабря и освободил дополнительные полчаса времени для работы над этой статьей)

СПИСОК ЛИТЕРАТУРЫ

International Journal of Systematic and Evolutionary Microbiology (2002), 52, 101–107 (<http://www.microbiologyresearch.org/docserver/fulltext/ijsem/52/1/0520101a.pdf?expires=1512922004&id=id&accname=guest&checksum=AB2DC4C0FE3D6371DE31BCE3E9B496C5>)

https://ru.wikipedia.org/wiki/%D0%94%D0%B5%D0%B7%D0%BE%D0%BA%D1%81%D0%B8%D1%80%D0%B8%D0%B1%D0%BE%D0%BD%D1%83%D0%BA%D0%BB%D0%B5%D0%B8%D0%BD%D0%BE%D0%B2%D0%B0%D1%8F_%D0%BA%D0%B8%D1%81%D0%BB%D0%BE%D1%82%D0%B0
<https://ru.wikipedia.org/wiki/%D0%91%D0%B5%D0%BB%D0%BA%D0%B8>