

Обзор протеома археи *Fervidobacterium islandicum* (мини-обзор)

Автор: Антон Иззи

_Резюме

В данной работе мною был исследован геном бактерии *Fervidobacterium islandicum* по данным таблиц генов этого микроорганизма из сайта NCBI. Далее будут представлены данные, полученные в ходе работы, а именно число белков и генов РНК по категориям, гистограмма длин белков и таблица числа генов, псевдогенов и генов РНК на прямой и комплементарной цепочках ДНК. Все расчёты проводились в таблицах Excel

_Ключевые слова

Бактерия, протеом, статистика генов, таблица, гистограмма.

1. Введение

Fervidobacterium islandicum – термофильный анаэроб, который был добыт в Индонезии из горячего источника [1].

Оптимальная температура роста и жизни этой бактерии – приблизительно между 50 и 80 °С [2]. Является Грам-отрицательным микроорганизмом; также известно, что обладает подвижностью. Длина клетки - около 1,8 мкм, ширина - 0,6. Одиночные, иногда попадаются парами. Эта бактерия интересна тем, что может разложить птичьи перья при температуре 70 °С и pH 7 [1]. Размер генома составляет 2237377 пар оснований; генов всего 2089, из них 1928 кодирующих белки, 54 гена кодируют РНК, а остальные 107 являются псевдогенами.

В моей работе была составлена классификация генов по функциям, показывающая, сколько генов что кодируют. Всё это дело было оформлено в удобную таблицу. Далее мной была проведена статистика длин белков в виде гистограммы. Также был произведён подсчёт генов на прямой и комплементарной цепях ДНК бактерии.

2. Материалы и методы

Информация по геному моей бактерии была взята с сайта NCBI [3] (спасибо ему за свободный доступ). Для начала мне пришлось скачать файл, содержащий информацию о геноме моего микроорганизма (например данные о белках, РНК и т.д.), и затем оформить эти данные в виде таблицы (как никак удобно). Я пользовался таблицами Microsoft Excel 2016 для составления этой таблицы, а также некоторых других таблиц и гистограммы, характеризующих главную информацию, добытую мной в течении моего исследования. Для составления этих таблиц были использованы встроенные функции электронной таблицы Excel (СЧЁТЕСЛИМН, СРЗНАЧ и некоторые другие) что собственно и придаёт удобство работе в этой программе

3. Результаты

В геноме моей бактерии закодировано 1928 генов белков и 54 гена РНК. Данные подсчётов приведены в Таблице.1 и Таблице.2

Таблица.1. Классы белков и количество генов, кодирующие их

Класс белка	Количество
Транспортный	198
Рибосомальный	57
Гипотетический	549
Остальные	1124

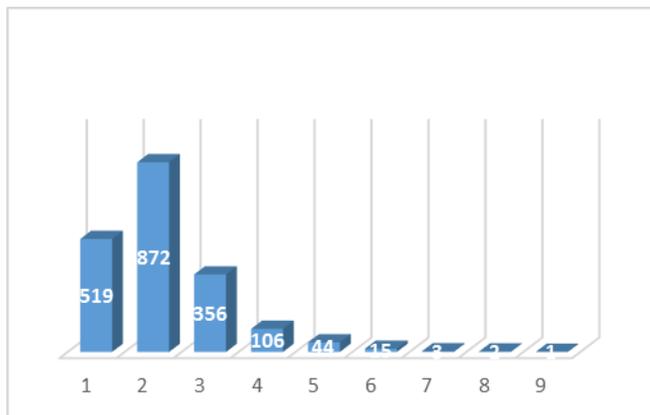
Таблица.2. Классы РНК и количество генов, кодирующие их

Класс РНК	Количество
тРНК	47
рРНК	6

Остальные	1
-----------	---

Однако это не вся работа. Также мною была составлена статистика длин белков бактерии, оформленная в виде гистограммы.

Диаграмма.1. Гистограмма длин белков



Данная гистограмма показывает распределение длин белков в соответствии с выбранным мною интервалом длин белков (я взял интервал длины 200). Наиболее полное представление о данных этой гистограммы даёт следующая таблица:

Таблица.3. Распределение длин белков в соответствии с диапазонами длины

Диапазоны	Размер диапазона	Количество белков
0	0-200	519
200	200-400	872
400	400-600	356
600	600-800	106
800	800-1000	44
1000	1000-1200	15
1200	1200-1400	3
1400	1400-1600	2
1600	1600<	1

Помимо прочего, мною также была составлена другая статистика, так сказать общая, показывающая минимальную и максимальную длину белка, среднюю длину белков и т.д. Результаты представлены в Таблице.4

Таблица.4. Общая статистика длин белков

Минимум	38
Средняя длина	330,155
Медиана	299,5
Максимум	1650
Стандартное отклонение	199,211

(если возникнут удивления по поводу не близкого расположения минимума и максимума – не обращайтесь внимание, small glitch).

Да, и я конечно же не забыл про распределение генов на прямой и комплементарной цепочках ДНК. Следующая таблица показывает подсчитанное мной распределение генов на цепочках ДНК (расшифровка: «+» - прямая цепь, «-» - комплементарная, или обратная):

Таблица.5. Распределение генов белков, РНК и псевдогенов на прямой и комплементарной цепочках ДНК *Fervidobacterium islandicum*

цепь	Кодирующая белок	псевдогены	гены РНК
+	979	47	33
-	949	60	21

И ещё я не удержался не выложить сюда некий забавный факт, обнаруженный в скачанной мною таблице генома археи с сайта NCBI, который связан с CDS (внимание на третью строчку!). В общем смотрите сами и улыбните свои глаза:

Таблица.6. Странный CDS

CDS	with_protein
Gene	pseudogene
CDS	without_protein
gene	pseudogene

4. Обсуждение и заключение

1. Транспортные РНК занимают лидирующую позицию по числу кодирующих их генов, и это не удивительно. Что касается белков, группа “остальные”, как оказалось, кодируется наибольшим количеством генов, но как никак не стоит забывать, что в эту группу может входить довольно много классов белков, в том числе и жизненно важных (таких как ферменты, мембранные белки и т.д.)

2. По данным гистограммы, самыми распространёнными белками являются белки с длиной от 200 до 400 аминокислотных остатков. Как вариант будущего исследования, представляется возможным провести статистику длин белков по геномам других бактерий, что позволит выявить наиболее распространённый интервал длин белков среди прокариот.

3. Что касается распределения белков по прямой и комплементарной цепям ДНК, что-либо сказать со 100% - ой уверенностью нельзя; гены распределяются практически одинаково, лишь с небольшим разрывом, будь то гены белков, псевдогены или гены РНК

4. Факт о том, что данная бактерия может перерабатывать перья при определённом рН, заинтересовал меня, и я конечно же начал думать, куда бы такую сверхспособность можно было бы применить во благо рода людского, но ничего адекватнее бойни на птицефабрике я не придумал.

5. И конечно же этот забавный момент с CDS... Я бы его прокомментировал конечно же, но пусть это останется на совести тех, кто заполнял эти данные и выкладывал их на сайт

5. Сопроводительные материалы

Данные по таблицам и геноме археи представлены в файле:

<http://kodomofbb.msu.ru/~skorpion7/%D0%B7%D0%B0%D0%B%D1%8F%D1%82%D0%B8%D0%B5%202.xlsx>

6. Благодарности

В первую очередь мне хотелось бы выразить благодарность преподавателям биоинформатики с моего факультета, благодаря которым я смог научиться пользоваться компьютером в действительно нужной области, а не попусту просаживать батарею. Также я хотел бы сказать спасибо моим корешам-однокурсникам, таким как Антон Власов, Стас Тихонов, Даниил Игумнов, которые в нужное время давали полезные советы и попросту выдерживали мою тормознутость и совершенно плоские вопросы.

7. Список литературы

[1] <https://www.ncbi.nlm.nih.gov/pubmed/12420177>

[2] <https://link.springer.com/article/10.1007/BF00423318>

[3] <ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/767/275/>