

Обзор протеома бактерии *Fervidobacterium islandicum* AW-1

Игумнов Д.С.

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

13.12.2017

РЕЗЮМЕ

Данная работа была выполнена в целях изучения генома и протеома бактерии *Fervidobacterium islandicum* штамма AW-1. Были изучены составы генома и протеома бактерии, выявлены и рассмотрены некоторые закономерности: длины белков *F. islandicum*, число генов, расположенных на прямой и обратной цепях ДНК соответственно. Также была проверена гипотеза равновероятного распределения генов на прямой и обратной цепях ДНК.

Ключевые слова: Fervidobacterium; Геном; Протеом; Обзор.

1. ВВЕДЕНИЕ

Fervidobacterium islandicum – гипертермофильная грамотрицательная бактерия. Систематическое положение представлено в *таблице 1*. Несмотря на то, что вид имеет отрицательную окраску по Граму, бактерии окружены одной липидной мембраной, то есть являются монодермными^{[1][2]}. Обитает в горячих источниках в условиях с температурным диапазоном от 50 до 80°C. Клетки окружены структурой в форме футляра - “тогой” – состоящей из пориноподобных белков^[3]. Способы усваивать высокомолекулярные углеводороды и в результате метаболизма выделять газообразный водород. Таким образом, вкуче с высокой термофильностью, бактерия может выступать в роли биотехнологического источника энергии, альтернативного ископаемому топливу^[4].

В процессе работы были изучены геном и протеом бактерии, выявлены закономерности расположения генов на цепях ДНК, длины белков. Также гены были посчитаны и распределены на соответствующие категории.

Таблица 1. Систематическое положение *Fervidobacterium islandicum*.

Kingdom:	Bacteria
Phylum:	Thermotogae
Order:	Thermotogales
Family:	Thermotogaceae
Genus:	Fervidobacterium
Species:	F. islandicum

2. МАТЕРИАЛЫ И МЕТОДЫ

Для работы были использованы данные, полученные из базы данных NCBI^[6]. Для обработки и анализа информации была использована операционная система Windows и программы Microsoft Office Excel 2010, FAR Manager. В ходе работы были выполнены следующие операции: скачанный с NCBI файл был интегрирован в Excel и отформатирован с автоматическим распределением по столбцам как тексто-

вая таблица, построена гистограмма распределения длин белков, написан скрипт на Python для проверки гипотезы равновероятного расположения генов на цепях ДНК. Использованные в Excel функции представлены в *таблице 2*. Для подсчета генов, кодирующих белки или РНК, по категориям использовались текстовые фильтры, представленные в *таблице 3*.

Таблица 2. Функции, использованные для работы с данными в Excel.

Подсчет	Функция
Средняя длина белка	СРЗНАЧ
Стандартное отклонение	СТАНДОТКЛОН
Медианная длина белка	МЕДИАНА
Минимальная длина белка	МИН
Максимальная длина белка	МАКС
Число белков в бине длины	СЧЁТЕСЛИМН

Таблица 3. Текстовые фильтры для подсчета генов по категориям.

Название столбца	Фильтр генов РНК	Фильтр генов белков
# feature class	gene tRNA/rRNA/ RNase_P_RNA	CDS/gene with_protein/ hypothetical
name	-	Ribosomal/transporter

3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Работа с информацией, полученной из базы данных, показала, что в геноме F. Islandicum AW-1 содержится 2237377 пар нуклеотидов, 2089 генов, из которых 1928 кодируют белки, 54 – РНК и 107 являются псевдогенами. Таким образом, на 1 миллион пар нуклеотидов статистически приходится ~934 гена.

Распределение генов по категориям

Гены были распределены на 2 основных класса: гены белков и гены РНК. Далее было произведено распределение и подсчет по соответствующим категориям, которые, как и результат данного подсчета, представлены в *таблицах 4 и 5*.

Как видно из полученных данных, количество генов белков, отнесенных к категории «остальные», превосходит количество распределенных по именованным категориям, и составляет 57,7% от общего числа. Скорее всего, это связано с тем, что данная категория по сути является «сборной» некоторого количества подкатегорий, значимостью которых по отдельности в данной работе было решено пренебречь. Также стоит заметить, что

гены гипотетических белков составляют более четверти от общего числа, что может быть связано с недостаточной изученностью данного штамма.

Таблица 4. Гены белков, распределенные по категориям.

Категория	Количество
Транспортные	198
Рибосомальные	57
Гипотетические	561
Остальные	1112

По распределению генов, кодируемых РНК-геномом, можно сказать, что большую часть из них составляют гены тРНК: 87%. Интересно заметить, что в отличие от генов белков, к категории «остальные» был отнесен всего лишь 1 ген, который кодирует рибонуклеазу P^[2], или РНКазу Р, являющуюся рибозимом, принимающим участие в метаболизме тРНК: её роль состоит в отсечении лишних нуклеотидов от тРНК^[5].

Таблица 5. Гены РНК, распределенные по категориям.

Категория	Количество
Транспортные	47
Рибосомальные	6
Остальные	1

Расположение генов на цепях ДНК

Результат подсчета генов на разных цепях ДНК представлен в *таблице 6*. Можно предположить, что мощность РНК-генома одной цепи связана обратно с количеством псевдогенов. Так, на кодирующей цепочке расположено больше псевдогенов, чем на комплементарной, но меньше генов, служащих матрицей для транскрибирования некодирующих РНК.

Таблица 6. Расположение генов на цепях ДНК.

Цепь	Гены белков	Псевдогены	Гены РНК
+	979	47	33
-	949	60	21

Опираясь на результаты работы скрипта, показанные в *таблице 7*, и считая вероятность такого отклонения достаточно высокой, можно сделать вывод, что расположение генов на цепях ДНК *F.islandicum* не противоречит гипотезе о равновероятном расположении генов.

Таблица 7. Расположение генов на цепях ДНК.

Выборка (тыс.)	1	10	50	100
Вероятность (%)	53,8	54,57	54	53,99

Статистика длин белков

Протеом данной бактерии был частично изучен: белки были распределены на бины длиной в 200 единиц (в данном случае – аминокислотных остатков). Как видно из *рисунка 1*, большая часть белков сосредоточена в интервале от 200 до 400. Белки, находящиеся в диапазоне от 0 до 600, составляют более 91% протеома бактерии.

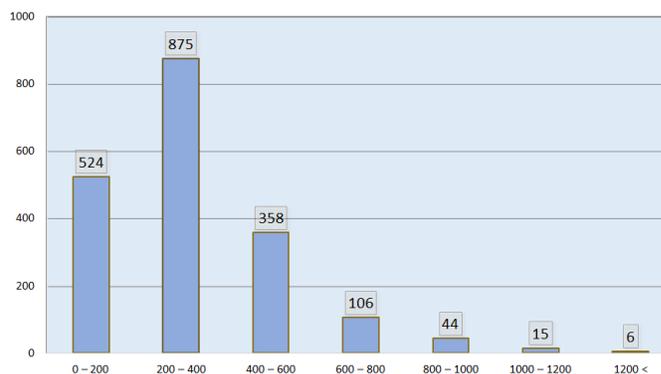


Рисунок 1. Гистограмма длин белков.

Данное распределение обусловлено тем, что средняя длина белка бактерий составляет 300-400 аминокислотных остатков^[8]. Этот факт подтверждается и результатами статистических подсчетов, которые представлены в *таблице 8*.

Таблица 8. Общая статистика длин белков.

Показатель	Значение
Средняя длина	330
Стандартное отклонение	199
Медиана	300
Минимальная длина	38
Максимальная длина	1650

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Данные бактерии и таблиц:

http://www.kodomo.fbb.msu.ru/~igumnov.daniel/term1/excel/Igumnov_review.xlsx

Скрипт на Python:

http://www.kodomo.fbb.msu.ru/~igumnov.daniel/term1/excel/probability_of_rejection.py

БЛАГОДАРНОСТИ

Благодарность выражается Алексеескому А.В. за помощь в освоении Microsoft Excel; Спирину С.А. и Дмитрию Пензару за ознакомление с азами Python; Дибровой Д.В. за вводный курс в HTML и CSS; всему коллективу специалистов, делившихся знаниями, за популяризацию FAR и написанные под-

сказки, невероятно облегчающие жизнь и поясняющие формулировки.

СПИСОК ЛИТЕРАТУРЫ

- [1] (1998) «Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaeobacteria, eubacteria, and eukaryotes». *Microbiol Mol Biol Rev* 62: 1435–1491.
- [2] (2011) «Origin of diderm (Gram-negative) bacteria: antibiotic selection pressure rather than endosymbiosis likely led to the evolution of bacterial cells with two membranes». *Antonie van Leeuwenhoek* 100: 171–182.
- [3] Huber, R., Woese, C.R., Langworthy, T.A., Kristjansson, J.K., and Stetter, K.O. "Fervidobacterium islandicum sp. nov., a new extremely thermophilic eubacterium belonging to the "Thermotogales". *Arch. Microbiol.* (1990) 154:105-111
- [4] (2006) «Microbial biochemistry, physiology, and biotechnology of hyperthermophilic Thermotoga species». *FEMS Microbiol Rev* 30: 872–905.
- [5] Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S (1983). "The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme". *Cell*. 35 (3 Pt 2): 849–57.
- [6] Assembly: GCA_000767275.3 ASM76727v3 [Электронный ресурс]// URL: ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/767/275/GCA_000767275.3_ASM76727v3/GCA_000767275.3_ASM76727v3_feature_table.txt.gz
- [7] Controlled vocabulary for ncRNA classes [Электронный ресурс]// URL: http://www.insdc.org/rna_vocab.html
- [8] Структура белка: введение для айтишников [Электронный ресурс]// URL: <https://geektimes.ru/post/181850/>