

Обзор генома бактерии *Ruminiclostridium thermocellum* AD2

Сафронов Вячеслав Викторович¹

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

РЕЗЮМЕ

В данном обзоре я исследовал геном бактерии *Ruminiclostridium thermocellum*, а именно: распределил белки и РНК по категориям и расположению в геноме и построил гистограмму распределения белков по их длине. Вся информация обрабатывалась с помощью Google Sheets.

1 ВСТУПЛЕНИЕ

R. Thermocellum (рис.1) это анаэробная термофильная бактерия способная перерабатывать целлюлозу в этанол.

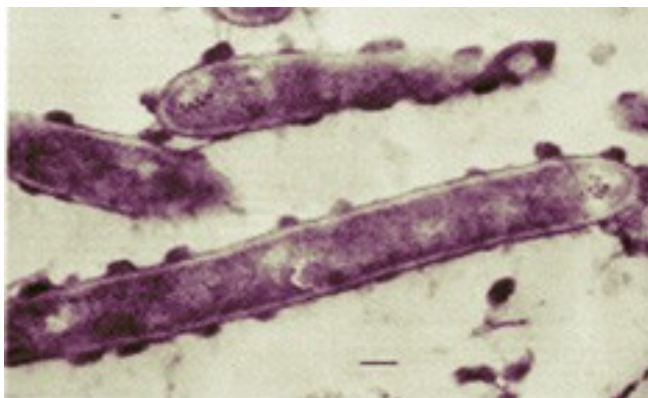


Рисунок 1. Микрофотография *R. Thermocellum*

R. Thermocellum живет при температуре 28-69°C, температурный оптимум 60°C.[1] Размер генома 3554854 пар оснований[2]

2 МАТЕРИАЛЫ И МЕТОДЫ

2.1 Используемые ресурсы и программы

Таблица с данными о геноме бактерии получена с сайта NCBI[3].

Для обработки данных использовался сервис Google Sheets.

2.2 Используемые функции Google Sheets

=МИН()

=МАКС()

=СЧЁТЕСЛИМН()

=СРЗНАЧ()

=СТАНДОТКЛОН()

=МЕДИАНА()

Фильтр

2.3 Определение числа генов белков и РНК в категории

Для определения числа генов белков в каждой категории был использован фильтр по значению "CDS" в диапазоне A:A. Получившаяся таблица скопирована на другой лист, после чего был

включен фильтр по значению "hypothetical protein" и подсчитано количество строк в получившейся таблице. Для подсчета числа генов транспортных белков фильтр был поставлен на «transport» и выбраны подходящие пункты. Аналогично и для рибосомальных белков. Количество остальных определено вычитанием найденных из общего количества. Для подсчета генов кодирующих тРНК и рРНК были использованы формулы =СЧЁТЕСЛИ(table1!\$A:\$A; "=tRNA") и =СЧЁТЕСЛИ(table1!\$A:\$A; "=rRNA") соответственно.

2.4 Распределение длин белков

Для построения гистограммы была использована таблица из практикума 13.

2.5 Распределение числа генов белков и РНК на прямой и комплементарной цепи

Для построения таблицы была использована функция =СЧЁТЕСЛИМН() с условием наличия + или — в столбце J:J в таблице

3 РЕЗУЛЬТАТЫ И МЕТОДЫ

3.1 Число генов белков и РНК по категориям

Длина генома *R. Thermocellum* составляет 3554854 пар нуклеотидов, на это количество приходится 3103 гена, таким образом геном *R. Thermocellum* содержит в среднем 886 ген на 1 млн пар оснований. Из Таблиц 1 и 2 можно увидеть какое количество генов относится к каждой из категорий.

Таблица 1. Число различных генов белков

Гены белков	
Транспортные	16
Рибосомальные	59
Гипотетические	671
Остальные	2289
Сумма	3035

Таблица 2. Число различных генов РНК

Гены РНК	
тРНК	56
рРНК	12
Сумма	68

3.2 Статистика длин белков

Как видно из гистограммы (рис.2), большинство белков бактерии имеют длину от 200 до 400 аминокислотных остатков. Также заметно экспоненциальное уменьшение количества белков по мере возрастания их длины



Рисунок 2. Распределение количества генов по их длине

3.3 Распределение генов по прямой и обратной цепочкам ДНК

В Таблице 3 представлены сведения о количестве генов на прямой и обратной цепочках ДНК

Таблица 3. Распределение генов по цепочкам

	Белок	РНК	Псевдоген
Прямая	1476	37	5
Обратная	1559	31	7

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Ссылка на таблицу Google Sheets:

https://docs.google.com/spreadsheets/d/1OX4gu8Q7H-8Xm1bDe4aPiJOWynrAE90mqfQ3htK_gtU/edit?usp=sharing

ССЫЛКИ

- [1] DORIS FREIER, CHERYLE P. MOTHERSHED AND JUERGEN WIEGEL. Characterization of Clostridium thermocellum JW20. APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Jan. 1988, p. 204-211. <http://aem.asm.org/content/54/1/204.full.pdf>
- [2] Ruminiclostridium thermocellum AD2, complete genome https://www.ncbi.nlm.nih.gov/nucleotide/NZ_CP013828.1
- [3] Таблица с данными о геноме на сайте NCBI. ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002/564/085/GCA_002564085.1_ASM256408v1/GCA_002564085.1_ASM256408v1_feature_table.txt.gz