

Мини-обзор генома бактерии *Methylocystis heyeri*

Автор: Лиференко Анастасия

Факультет Биоинженерии и Биоинформатики, Московский Государственный университет, Москва, Россия

Контактные данные: a.liferenko@fbb.msu.ru

Supplementary information:

Резюме: В данной работе рассматриваются стандартные данные о геноме бактерии *Methylocystis heyeri*, а также статистические данные о белках протеома и генах РНК данного штамма

Ключевые слова: *Methylocystis heyeri*, геном, протеом

Введение

Аэробные метанотрофные бактерии способны использовать метан в качестве единственного источника углерода и энергии. Они были разделены на два типа,, относящиеся к гамма- и альфапротеобактериям.

Мной выбранный бактерия- *Methylocystis heyeri*- типа II, обладает "фирменными" жирными кислотами метанотрофов типа I. Штаммы бактерии : H2(T), выделенный из кислого (pH 4,3) сфагнового торфяного озера (Teufelssee, Германия, 2001). Другой штамм Sakb1, выделенный из кислой (pH 4,2) почвы тропического леса (Таиланд,2005)

Описание клеток штаммов:аэробными, грамотрицательными, неподвижные палочки,покрытыми большими полисахаридными капсулами и содержащими внутрицитоплазматическую мембранную систему, типичную для метанотрофов II типа. Они обладают как частичной, так и растворимой монооксигеназой метана и используют сериновый путь усвоения углерода. Наиболее уникальной характеристикой этих штаммов является профиль жирных кислот фосфолипидов. Помимо характерной жирной кислоты метанотрофов.

Последовательности генов 16S рРНК имели 96-98% сходства с последовательностями других метанотрофов II типа в родах *Methylosinus* и *Methylocystis*. Филогения последовательности гена 16S рРНК и *rpoA* показала, что штаммы образуют новую линию внутри рода *Methylocystis*. Таким образом, предполагается, что эти два штамма представляют новый вид, *Methylocystis heyeri* sp. nov.

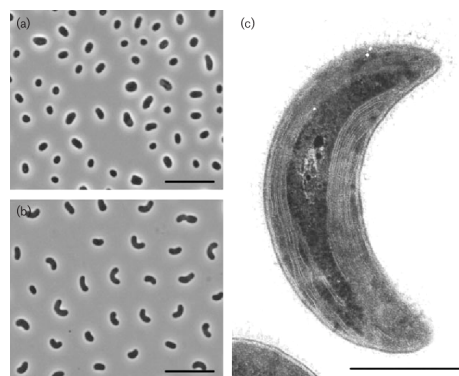


Рис.1. Фазово-контрастные микрофотографии клеток штаммов H2T (а) и Sakb1 (б), выращенных на твердой минеральной среде под метаном в течение 7 дней. (в) Электронная микрофотография ультратонкого среза клетки штамма H2T, выращенной на метане. второго типа, клетки также содержали большое количество жирной кислоты, которая ранее считалась характерной жирной кислотой метанотрофов первого типа

Domain	Bacteria
Phylum	Pseudomonadota
Class	Alphaproteobacteria
Order	Hyphomicrobiales
Family	Methylocystaceae
Genus	Methylocystis
Species	<i>M. heyeri</i>

Таблица 1. Таксономическое положение данной бактерии

2. Материалы и методы

Для проведения вычислительных исследований из электронной библиотеки NCBI были взяты: нуклеотидная последовательность генома, аннотация протеома *Methylocystis heyeri*.

Вычислительные исследования на основе этих данных были проведены с помощью программы для работы с электронными таблицами Google Sheets и Python.

Количество белок-кодирующих генов, генов РНК, медианная длина белка и некоторые другие статистические данные были получены с помощью программы Google Sheets. Поиск встречаемости нуклеотидов и подсчет стоп-кодонов был реализован при помощи скриптов Python. Для поиска участка *oriC* и *ter*, использовала сервис Webskew

3. Результаты

3.1 Стандартные данные о геноме *Methylocystis heyeri*.

Геном *Methylocystis heyeri* состоит из одной кольцевой хромосомы и двух кольцевых плазмид. Название, размер, GC-состав этих молекул (Таблице 2.)

Тип	Название	Размер (п.н.)	GC-состав (%)
Хромосома	NZ_CP046052	45,51947	63
Плазмида 1	unnamed 1	145,219	61
Плазмида 2	unnamed 2	30,622	60

Таблица 2. Состав и размер генома

3.2 Зависимость количества белковых продуктов от их длины

Чтобы получить картину распределения белковых продуктов по длинам, была составлена гистограмма (рис. 2).

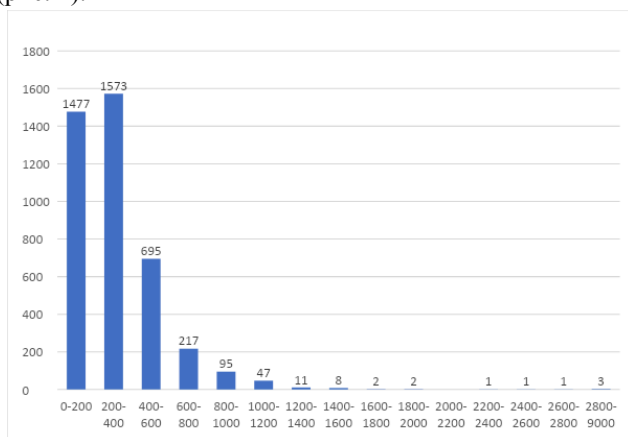


Рисунок 2 – Распределение белков по длинам аминокислотных цепей

Из этой гистограммы видно, что разброс протеома значителен, однако большая его часть приходится на белки длиной до 1400 аминокислотных остатков (а.о.). Наиболее многочисленными являются белки длиной от 200 до 400 а.о. – 1573 белка. Второе место занимают еще более мелкие белки длиной до 200 а.о. – 1477 представителей. Белков длиной более 1400 а.о. в сумме насчитывается 18 единиц, причем их длина не превышает 9000 а.о.

3.2. Распределение генов на прямой и обратной цепи генома

Всего у бактерии 4165 генов, кодирующих какой-либо белок. Из них на прямой(+) цепи находится 2046 генов, а на комплементарной (-) - 2119.

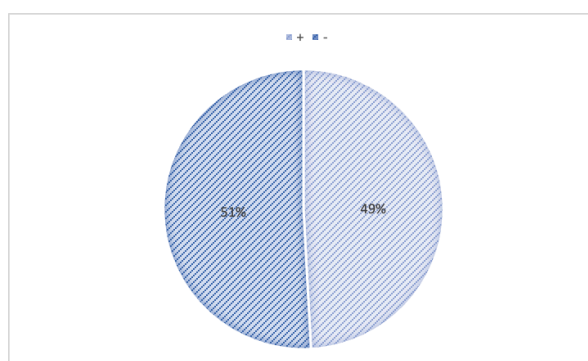


Рис 3. Распределение генов на прямой и обратной цепи генома бактерии

На рисунке выше (рис.3) представлено распределение относительно всего генома (на цепи прямого направления находится 49.2% генов, а на обратной - оставшиеся 50.9%). Можно сказать, что гены распределены примерно поровну на обеих цепях. В следующей таблице(Таблица 3) показано более точное распределение в плазмидах и хромосоме.

ДНК	Прямое направление цепи	Обратное направление цепи	%содержание от всех белков
хромосома	1965	2048	96,4 %
Плазмида 1	66	61	3,05 %
Плазмида 2	15	10	0,6 %
Всего	2046	2119	100 %

Таблица 3. Распределение генов, кодирующих белки на прямой и обратной цепи молекулы ДНК.

Для того, чтобы проверить случайность распределения на прямой и обратной цепи хромосомы и плазмиды, было рассчитано % содержание от всех белков, указанных в (Таблица 3)

3.2.3 Соотношение различных классов белков в геноме бактерии

С помощью таблицы геномных особенностей были получены данные о количестве разных классов белков в геноме *Methylocystis heyeri*

класс белков	кол-во	общая доля
рибосомальный	41	1,9%
гипотетический	822	19,6%
транспортный	202	4,8%

Таблица 4. Соотношение разных классов белков в геноме

3.3 Статистические данные о генах РНК

В ходе рассмотрения распределения генов различных функциональных РНК было выяснено:

	количество
Всего РНК-кодирующих генов	62
рРНК-кодирующих генов	11
тРНК-кодирующих генов	51

Таблица 5. РНК-кодирующие гены

3.5 Нуклеотидный состав геномных ДНК

Для того, чтобы проверить второе правило Чаргаффа, была написана программа, подсчитывающая число и частоту встреч каждой из букв А, Т, G, С. Используя написанную программу, считающую встречаемость каждого из стоп кодонов кодирующих белок последовательностей в долях (до 4 знака после запятой). Были получены данные, указанные в Таблице 6.

Основание	Доля встречаемости	Содержание(%)
A	0.4889	19.32%
T	0.4691	18.54%
C	0.7935	31.36%
G	0.7789	30.88%

Табл 6. Нуклеотидный состав геномной ДНК

Опираясь на результаты таблица, можно подтвердить второе правило Чаргаффа: содержание Аденина 19,31% приблизительно равно содержанию Тимина 18.54%, а

содержание Цитозина 31.36% приблизительно равно содержанию Гуанина 30.88%

3.4.Проверьте гипотезу о том, что гены распределены по двум цепочкам ДНК случайно с вероятностями 0,5

В геноме 4165 генов. На одной цепи 2046, на другой – 2119. Отклонение от ожидаемого числа 2082,5 – 36,5. Из 100 экспериментов 82 раза отклонение было больше или равно наблюдаемому 36,5. $82/100=0,82$. Вывод: отклонение 36,5 не противоречит гипотезе о независимом случайном равновероятном распределении генов на цепочках

3.7 Найдите в кольцевой хромосомной ДНК из выбранного вами генома участок *oriC*, в котором начинается репликация и участок *ter* в котором происходит терминация репликации

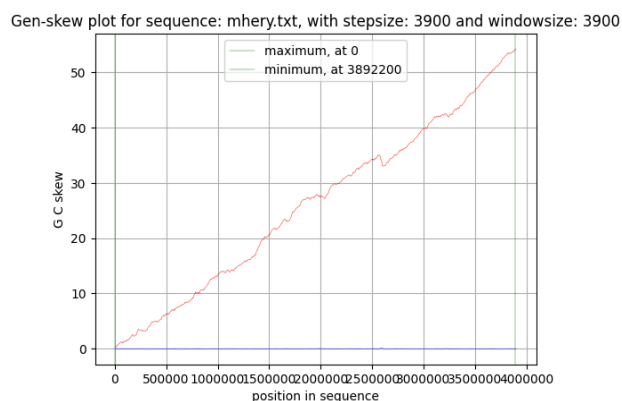


Рис 4.. Кумулятивный GC-skew график для кольцевой хромосомы NZ_CP046052

Если предположение о том, что точка начала репликации соответствует точке минимума (*oriC*), а точке максимума (*ter*) - точка конца, верно, то *oriC* - 3892200 позиции, а *ter* - 0 позиции (рис.4)

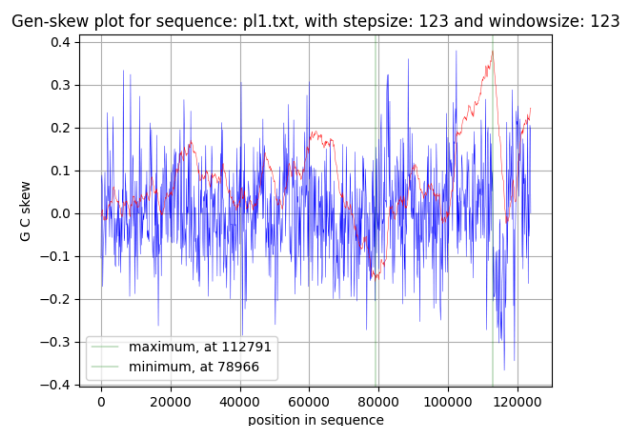


Рис 5.Кумулятивный GC-skew график для кольцевой плазмиды unnamed 1

Если предположение о том, что точка начала репликации соответствует точке минимума (*oriC*), а точке максимума (*ter*) - точка конца, верно, то *oriC* - 78966 позиции, а *ter* - 112791 позиции (рис.5)

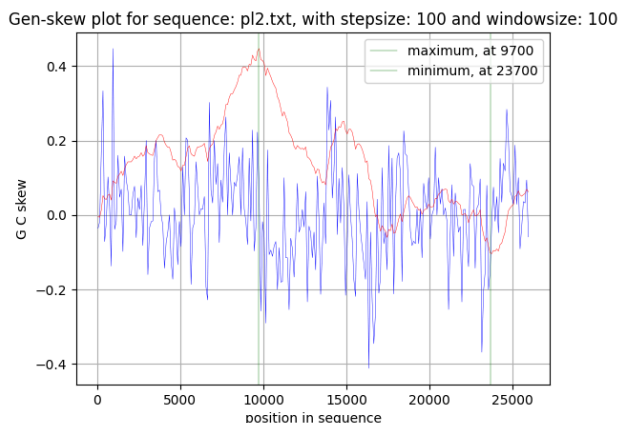


Рис 5. Кумулятивный GC-skew график для кольцевой плазмиды unnamed 2

Если предположение о том, что точка начала репликации соответствует точке минимума (oriC), а точке максимума (ter) - точка конца, верно, то oriC-9700 позиции, а ter - 23700 позиции (рис.5)

3.9 Частота встречаемости стоп-кодонов

Используя программу, считающую количество встреч каждого из стоп кодона, я получила данные, представленные в (Таблице 7)

Кольцевая хромосома 1	TGA	2486
	TAA	747
	TAG	800
plasmid unnamed1	TGA	86
	TAA	23
	TAG	22
plasmid unnamed2	TGA	17
	TAA	5
	TAG	4

Таблица 7. Частота встречаемости стоп кодонов

4. СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

-Файлы Excel(anastasia.l@pr8),

-Нуклеотидная последовательность хромосомы и плазмид из NCBI в формате FASTA

(https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP046052.1
https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP046053.1
https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP046054.1),

-скриптыPython

(<https://docs.google.com/document/d/1tVJyQc5IyrRZRcBF6-hPn4PcT7LwrDjHqA-GCuRD56w/edit?usp=sharing>)

- геном в формате fasta

([GCF_004802635.2_ASM480263v2_cds_from_genomic.fna.gz](https://www.ncbi.nlm.nih.gov/genbank/GCF_004802635.2_ASM480263v2_cds_from_genomic.fna.gz))

-таблица особенностей генома

([GCF_004802635.2_ASM480263v2_feature_table.txt.gz](https://www.ncbi.nlm.nih.gov/genbank/GCF_004802635.2_ASM480263v2_feature_table.txt.gz))

-сайт сервиса Webskew

(https://genskew.csb.univie.ac.at/webskew#usage_webskew)

БЛАГОДАРНОСТИ

Я выражаю благодарность своим преподавателям информатики за то, что за прошедшие занятия научили показали методы , которые помогли анализировать геном и протеом бактерии Methylocystis heyeri.

Список Литературы

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC441497/>
<https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ij.s.0.64623-0>
<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=391905>
<https://pubmed.ncbi.nlm.nih.gov/17329771/>
https://www.ncbi.nlm.nih.gov/assembly/GCF_004802635.2/#/ga
https://biocyc.org/organism-summary?object=GCF_004802635
https://www.kegg.jp/kegg-bin/show_organism?org=mhey

