

Обзор генома бактерии *Methylovorus glucosetrophus* SIP3-4

Клещенко М.Д.¹

¹ Факультет биоинженерии и биоинформатики, Московский Государственный Университет имени М. В. Ломоносова

КЛЮЧЕВЫЕ СЛОВА:

Methylovorus glucosetrophus SIP3-4, геном, метилотрофные бактерии

1 ВВЕДЕНИЕ

Methylovorus glucosetrophus SIP3-4 – бактерия, относящаяся к семейству Methylophilaceae. Оно включает в себя 4 рода, все представители которых – повсеместно распространенные облигатные или факультативные метилотрофы из наземных или пресноводных сред. Данная бактерия использует метиламин. Её геном кодирует ферменты: метанолдегидрогеназа, малатдегидрогеназа. Было установлено, что эти они являются отличительными для первичного окисления субстрата у этого вида^[2]. Вероятное практическое применение:

1. Метилотрофы выделяют ауксин и цитокинин, которые влияют на прорастание семян и рост корней растений, а также помогают растениям переносить водный стресс^[3]. Также такие бактерии выделяют осмопротекторы, как сахара и спирты, которые помогают защитить растения от высыхания и чрезмерного излучения^[3], фиксируют азот, соллюбилируют Р, К, Zn. Использование метилотрофных штаммов, в качестве биоудобрений может способствовать надлежащему производству сельскохозяйственных культур^{[3][4]}.

2. Возможно использование в почвах, покрывающих свалки (полигоны ТБО)^[5], так как аэробные метаноокисляющие бактерии преобразуют метан в углекислый газ, тем самым снижая выбросы парниковых газов в атмосферу^{[1][5]}.

2 МАТЕРИАЛЫ И МЕТОДЫ

Для анализа бактерии использовалась

1. таблица особенностей из файла GCF_000023745.1_ASM2374v1_feature_table.txt
2. геном бактерии из файла GCF_000023745.1_ASM2374v1_genomic.fna

Расчеты проводились в программах Excel и Python:

- Для подсчета длин последовательностей и их GC-состава была написана программа в Python [CM 2, программа 1]

- Для анализа распределения длин белков и построения диаграммы использовался Excel [CM 1, лист «Гистограмма»]
- Для анализа распределения генов по функциям использовался Excel [CM 1, лист «Статистика»]
- Для определения содержания нуклеотидов в молекулах ДНК использовалась программа в Python [CM 2, программа 2]

3 РЕЗУЛЬТАТЫ

Общие сведения о геноме

Из таблицы особенностей мы узнали, что геном *Methylovorus glucosetrophus* SIP3-4 состоит из одной хромосомы и двух плазмид (pMsip01, pMsip02). Из данных о сборке генома можно узнать GC-состав, однако точность его определения мала, поэтому посчитаем значения с помощью программы, написанной в Python [CM 2, программа 1]. Сведения о длине и GC-составе ДНК можно узнать из Таблицы 1.

Таблица 1. Общие сведения о геноме

ДНК	Длина, п.н.	GC-состав, %
Хромосома	2995511	0.54860
pMsip01	76680	0.45713
pMsip02	9816	0.47407

GC-состав генома свидетельствует о том, что бактерия проживает в умеренных условиях.

Статистические данные о белках протеома

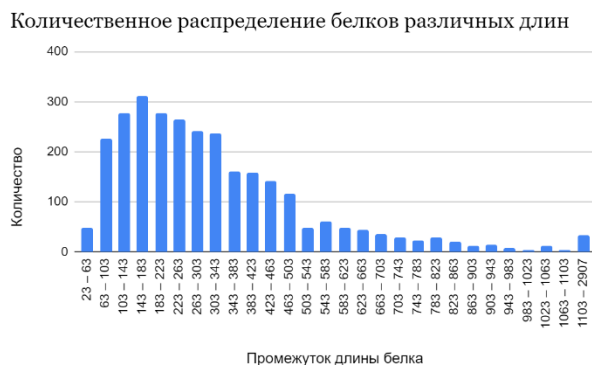
На основании таблицы особенностей был проведен анализ данных о белках протеома, а также построена гистограмма, отражающая количественное распределение белков разной длины. В Таблице 2 приведены данные о диапазоне длин белков, а также среднем и медианном значениях.

Таблица 2. Статистические данные о длинах белков протеома

Показатель	Число аминокислотных остатков
Минимальная длина	23
Максимальная длина	2907
Средняя длина	322,32
Медиана	268

Как видно из Рисунка 1, в диапазоне 143-183 аминокислотных остатков наблюдается пик. Распределение напоминает положительно скошенное распределение.

Рисунок 1. Гистограмма распределения длин белков. На горизонтальной оси располагаются интервалы, соответствующие количеству аминокислот, а на вертикальной оси – количество белков, длины которых попадают в данный диапазон.



Далее мною был проведен анализ белков по их функциям, результаты приведены в Таблице 3. Исходя из приведенных данных о том, что гипотетические белки составляют целых 14,16%, можем сделать вывод что геном изучен не до конца, ведь средняя длина таких белков составляет 189 аминокислотных остатков, медианное значение чуть меньше – 156, а самый длинный белок и вовсе 1452. Исходя из гистограммы, обсужденной выше, можно сделать вывод, что неизученными остаются белки, которые могут выполнять действительно важные функции, так как основную массу белков составляют протеины именно такой длины.

Таблица 3.

Тип белков	Количество белков	Доля белков, %
Все белки	2881	100,00
Рибосомальные	59	2,05
Транспортные	174	6,04
Гипотетические	408	14,16

Также был проведен анализ расположения генов, кодирующих белки, на прямой и комплементарной цепочках ДНК для каждой из молекул и рассчитана вероятность того, что при случайном распределении генов по двум цепочкам мы получим такую же или большую разницу между количествами генов на двух цепочках.

Таблица 4.

Молекула ДНК	Цепь	Количество генов	Вероятность случайного распределения
Хромосома	прямая	1342	0,0184
	обратная	1468	
pMsip01	прямая	20	0,00096
	обратная	54	
pMsip02	прямая	-	-
	обратная	12	

В случае, если вероятность больше 0,05 – можно говорить о том, что гены распределены случайно. В нашем случае значения как для хромосомы, так и для первой плазмиды значительно отличаются от 0,05 в меньшую сторону. О случайности распределения для второй плазмиды говорить не приходится, так как ни один ген не расположен на комплементарной цепочке.

Статистические данные о генах РНК

Из таблицы особенностей получены данные о различных закодированных РНК, результаты представлены в Таблице 5.

Таблица 5.

Гены РНК	Количество генов
Все гены РНК	60
Гены тРНК	49
Гены рРНК	6
tmRNA	1
SRP_RNA	2
RNase_P_RNA	2

Также я провела сравнение количества генов, кодирующих белки и различные РНК, получилось, что в 48,02 раза больше генов, несущих информацию про белки.

Описание нуклеотидного состава ДНК

Для этого исследования был написан код в Python [СМ 2, программа 2], который считает для каждой ДНК (только для одной цепи, которая имеется в файле с геномом) количество каждого нуклеотида в последовательности, а также частоту его встречаемости. Результаты приведены в Таблице 6.

Таблица 6.

Молекула ДНК	Нуклеотид	Количество, шт	Доля, %	Вероятность случайного распределения
Хромосома	A	672713	22,457	0,0699 * 10 ⁻⁷
	T	679448	22,682	
	G	822254	27,450	0,3668
	C	821096	27,411	
pMsip01	A	693122	22,561	0,01658 * 10 ⁻⁸
	T	700667	22,807	
	G	839131	27,314	0,9146
	C	839271	27,318	
pMsip02	A	695609	22,570	0
	T	703343	22,821	
	G	841432	27,301	0,8836
	C	841623	27,308	

В данном разделе исследования мы хотели узнать случайное ли распределение нуклеотидов внутри пар A-T G-C, для этого мы снова использовали формулу для определения случайности распределения [СМ 1, лист «Статистика»].

При случайном распределении мы бы получили максимально близкие числа нуклеотидов в паре, в данном случае те пары, у которых значение в 5 столбце выше, чем 0,05, имеют

практически равные количества нуклеотидов. К таким относятся G-C пары в каждой молекуле ДНК. То есть второе правило Чаргаффа работает для G-C пар у данной бактерии.

Наоборот, количества аденинов и тиминов сильно различаются во всех молекулах ДНК, что можно заметить из таблицы, ведь при малой вероятности, появляются серьёзные основания полагать, что эффект отражает некоторое природное явление.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

1. Ссылка на таблицу, в которой проводились статистические расчеты
<https://docs.google.com/spreadsheets/d/1gHnywndRxNWU3B3AUuKRnNmZM4jduROl/edit?usp=sharing&ouid=117704077599522511880&rtpof=true&sd=true>
2. Ссылка на Google Collab, в котором написаны программы Python
<https://colab.research.google.com/drive/13yaG6E2tsubAZzQWADgCqK6HXXVJDuxF?usp=sharing>
3. Ссылка для скачивания таблицы особенностей и генома бактерии
https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/023/745/GCF_000023745.1_ASM2374v1/

БЛАГОДАРНОСТИ

Я очень благодарна всем преподавателям биоинформатики за то, что давали нам полезную информацию в течение этого семестра. Результат не заставил долго себя ждать: при исследовании бактерии в данном мини-обзоре я использовала навыки, полученные на занятиях и в ходе выполнения домашних работ.

СПИСОК ЛИТЕРАТУРЫ

1. [HTTP://BACMAP.WISHARTLAB.COM/ORGANISMS/926](http://BACMAP.WISHARTLAB.COM/ORGANISMS/926)
2. LAPIDUS A, CLUM A, LABUTTI K, KALUZHNYA MG, LIM S, BECK DA, GLAVINA DEL RIO T, NOLAN M, MAVROMATIS K, HUNTEMANN M, LUCAS S, LIDSTROM ME, IVANOVA N, CHISTOSERDOVA L. GENOMES OF THREE METHYLOTROPHS FROM A SINGLE NICHE REVEAL THE GENETIC AND METABOLIC DIVERGENCE OF THE METHYLOPHILACEAE. J BACTERIOL. 2011 AUG;193(15):3757-64. doi: 10.1128/JB.00404-11. Epub 2011 May 27. PMID: 21622745; PMCID: PMC3147524.
3. KUMAR, M., KOUR, D., YADAV, A.N. ET AL. BIODIVERSITY OF METHYLOTROPHIC MICROBIAL COMMUNITIES AND THEIR POTENTIAL ROLE IN MITIGATION OF ABIOTIC STRESSES IN PLANTS. BIOLOGIA 74, 287–308 (2019).
4. YADAV, NEELAM & YADAV, AJAR NATH. (2018). BIODIVERSITY AND BIOTECHNOLOGICAL APPLICATIONS OF NOVEL PLANT GROWTH PROMOTING METHYLOTROPHS. 5. 342-344. 10.15406/JABB.2018.05.00162.
5. RAI, R.K., CHETRI, J.K., GREEN, S.J., REDDY, K.R. (2019). IDENTIFYING ACTIVE METHANOTROPHS AND MITIGATION OF CH₄ EMISSIONS IN LANDFILL COVER SOIL. IN: ZHAN, L., CHEN, Y., BOUAZZA, A. (EDS) PROCEEDINGS OF THE 8TH INTERNATIONAL CONGRESS ON ENVIRONMENTAL GEOTECHNICS VOLUME 2. ICEG 2018. ENVIRONMENTAL SCIENCE AND ENGINEERING(). SPRINGER, SINGAPORE.