

Обзор транскриптома бактерии *Lactobacillus amylovorus* Nakamura 1981

Никитин Павел^{1*}

¹Факультет биоинженерии и биоинформатики, Московский государственный университет имени М. В. Ломоносова, Ленинские горы д. 1 стр. 73, 119234, Москва, Россия

Получено 19.11.2019; отредактировано 04.12.2019; принято 13.12.2019

Ключевые слова: *Lactobacillus amylovorus*, прокариоты, протеом. Key words: *Lactobacillus amylovorus*, procariotes, proteomics

РЕЗЮМЕ

В настоящее время в открытом доступе появляется всё больше отсеквенированных и аннотированных геномов. Получить содержательные результаты из этих данных можно сравнительно простыми способами. В частности, при обработке таблиц удобна программа Microsoft Excel. В этой работе на основе открытых данных NCBI проанализированы особенности кодирования белков в геноме бактерии *Lactobacillus amylovorus* с использованием программного обеспечения MS Excel.

1 ВВЕДЕНИЕ

Лактобациллы – род грамположительных факультативно анаэробных бактерий. *Lactobacillus amylovorus* была описана L.K. Nakamura в 1981 году. Был впервые выделен из отходов ферментации крупного рогатого скота. Описана в качестве отдельного вида из-за обширных возможностей внеклеточного гидролиза сахаров и отличных от других представителей рода способов культивации (Nakamura, 1981). Геном состоит из одной хромосомы, его длина составляет 2.09 Mbp, также в клетке обнаруживаются две плазмиды. GC-пары составляют 38,08% генома [4]. Своими биохимическими особенностями может быть интересна для исследования с целью дальнейшего промышленного использования [2]. В данной работе проанализирована таблица, описывающая транскриптом *L. amylovorus*, отмечены некоторые особенности распределения генов белков и РНК.

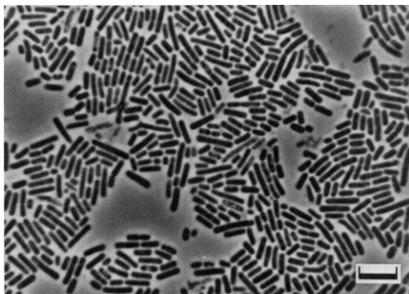


Рис. 1. Фазово-контрастная микрофотография живых клеток крахмал-гидролизующих бактерий *L. amylovorus*, Nakamura, 1981.

2 МАТЕРИАЛЫ И МЕТОДЫ

Для проверки гипотезы и составления сопроводительных таблиц и гистограмм была использована программа MS Excel. Все таблицы можно найти в сопроводительных материалах. Для анализа данных использовались различные функции MS Excel: СРЗНАЧ, СТАНДОТКЛОН, МЕДИАНА, МИН, МАКС, ВПР (VLOOKUP), а также сделанная плоская таблица. Использованы методы фильтрации, сортировки, относительной и абсолютной адресации, связи таблиц, специальной вставке, условного форматирования, распространения формул, построения гистограмм. Для проверки гипотезы случайного распределения генов был использован статистический критерий p-value. Информация о белках генома была получена из базы данных Genome NCBI и банка данных GenBank [4]. Информация о геноме:

Assembly: GCA_000191545.1

GenBank assembly accession: GCA_000191545.1 (latest)

3 РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

3.1 Распределение белков протеома по длинам

В рамках работы был проведен анализ длин белковых последовательностей. Как показано на гистограмме (Рис. 2), наиболее часты белки длиной от 50 до 350 аминокислотных остатков (составляют 67% всех белков). Около 3% всех белков имеют длину менее 50 аминокислот. 29% всех белков бактерии имеет длину от 350 до 900 аминокислотных остатков, и лишь 23 белка (1%) имеет длину более 950.

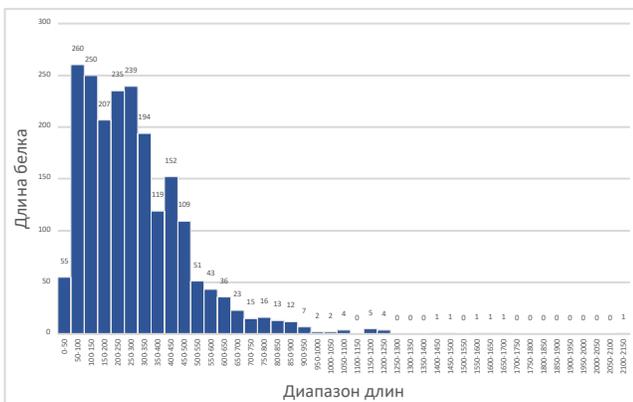
В Таблице 1 отражены минимальные, максимальные и медианные длины. Минимальная составляет 27 остатков, максимальная – 2124, а медианная – 254. Из этих данных можно сделать вывод, что большинство белков данной бактерии имеют относительно небольшую длину. Из среднего отклонения можно сказать, что подавляющее большинство белков имеет длину от 103 до 405 остатков.

* nikitin@fbb.msu.ru

Таблица 1. Статистические данные по длинам белков

Минимум	27
Максимум	2124
Среднее отклонение	151
Средняя длина	291
Медиана	254

Такое распределение можно объяснить следующим: белки прокариот, в отличие от эукариот часто одно- или двудоменные. Третьичная глобула, приходящаяся на один домен, имеет более-менее фиксированную длину. В распределении видно два максимума, которые, по всей видимости, получаются путем сочетания нормального распределения отдельно для одно- и для двудоменных белков. Медианная длина белка в 254 остатка является обычной для белков домашнего хозяйства и ферментов основных биохимических путей, а также может быть связана с отбором на длину генома в мире прокариот [3].

Рис. 2. Распределение белков по длине

3.2 Число генов по категориям, их распределение по прямой и обратной цепи DNA

Был проведен анализ распределения различных функциональных классов генов по прямой и обратной цепи DNA. Из данных таблицы 2 видно, что гены белков распределены по цепям в целом равномерно. Это подтверждает теорию о равнозначности обеих цепей DNA в клетках любого организма. О псевдогенах бактерии ничего неизвестно, что говорит о её плохой изученности, что связано с низкой распространенностью и особенностями культивации. А вот число генов RNA распределено по цепочкам явно неравномерно. Была проверена гипотеза о случайном распределении генов по цепям с помощью p-value. Критическим значением будем считать 0,001. В случае RNA имеем значение сильно меньше критического. Это нетривиальный факт, который требует дальнейших исследований. Это может быть также связано с особенностями поиска таких RNA, с регуляцией транскрипции таких генов, с тем, что гены tRNA транскрибируются другой полимеразой, нежели белки, и может быть, ей удобнее к ним так подступать, или, что возможно, с небольшой выборкой.

Таблица 2. Распределение генов разных функциональных классов по двум цепям DNA

Число генов	На прямой цепи DNA	На обратной цепи DNA	Всего	p-value
Белков	1068	991	2059	0,09
Псевдогенов	0	0	0	-
RNA	64	11	75	$3,1 \times 10^{-10}$

3.3 Анализ классов белков и RNA

Был проведен анализ классов белков, из которого мы выяснили, что гипотетические белки занимают в транскриптоме бактерии почти треть (Таблица 3).

Таблица 3. Гипотетические белки

Число гипотетических белков	Процент от всех
588	29%

Такую долю можно объяснить недостаточной изученностью бактерии по причинам, оговоренным выше. Исследование количества белков и RNA, составляющих рибосому (Таблица 4) показывает классическую картину строения рибосом прокариот. К ним относятся 5S, 16S, 23S rRNA, белки, составляющие 50S, 30S субъединицу, различные регуляторные белки (метильтрансферазы, ацетилтрансферазы, псевдоурдинсинтаза).

Таблица 4. Рибосомальные белки и RNA

Число рибосомальных белков	Число rRNA
65	12

Затем были более подробно рассмотрены распределения генов различных функциональных RNA по двум цепям в геноме (Таблица 5). Здесь также наблюдается неравномерность распределения по цепочкам. Число транспортных RNA укладывается в наши представления, поскольку триплетом нуклеотидов можно закодировать $4^3 = 64$ аминокислоты, то есть в клетке может быть до 64 различных транспортных RNA, просто в этом случае некоторые обслуживают принцип избыточности генетического кода, а одна замещена стоп-кодоном.

Таблица 5. Распределение функциональных классов RNA по цепям DNA

Число генов	На прямой цепи DNA	На обратной цепи DNA	Всего	p-value
Число генов rRNA	9	3	12	0,14
Число генов tRNA	55	8	63	$9,7 \times 10^{-10}$

Также было выяснено количество белков различных жизненно-важных функциональных классов (Таблица 6). В целом картина является довольно типичной для прокариот, однако обращает на себя внимание очень низкое количество трансмембранных белков. Это может быть

связано со сложностью в выделении таких белков или их транскриптомного анализа.

Таблица 6. Количество белков некоторых функциональных классов

ATP synthases	10
ATPases	29
Transporters	122
Transmembrane proteins	2
Regulator	86
RNA polymerases	7
DNA polymerases	13

ЗАКЛЮЧЕНИЕ

В ходе исследования было выявлено довольно несколько необычных закономерностей, которые требуют дальнейшего изучения. Необходимо также проверить корректность исходных данных, найти данные о геноме и протеоме из других источников и перепроверить результаты.

БЛАГОДАРНОСТИ

Хотелось бы выразить благодарность преподавателям факультета биоинженерии и биоинформатики МГУ: Алексеевскому А.В., Залевскому А.О. Спирину С.А. за предоставленные знания, а также Салимгарееву Р.С. за помощь в постижении биоинформатики.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Сопроводительные материалы доступны по ссылке: <https://docs.google.com/spreadsheets/d/1BBiIT3pEsJt4NIhD911AQVH-kemgfcXT2gtaFoMVV2Hs/edit?usp=sharing>

СПИСОК ЛИТЕРАТУРЫ

- [1] Nakamura L.K. (1981) *Lactobacillus amylovorus*, a new starch hydrolyzing species from swine waste-corn fermentation. *Int. J. Sys. Bacteriol.* 31: 56–63
- [2] Zhang D. X., Cheryan M. Direct fermentation of starch to lactic acid by *Lactobacillus amylovorus* // *Biotechnology Letters*. – 1991. – Т. 13. – №. 10. – С. 733-738.
- [3] Tiessen A., Pérez-Rodríguez P., Delaye-Arredondo L. J. Mathematical modeling and comparison of protein size distribution in different plant, animal, fungal and microbial species reveals a negative correlation between protein size and protein number, thus providing insight into the evolution of proteomes // *BMC research notes*. – 2012. – Т. 5. – №. 1. – С. 85.
- [4] <https://www.ncbi.nlm.nih.gov/genome/genomes/2539> (Была доступна 10.12.2019)