

Обзор протеома *Francisella philomiragia* subsp. *philomiragia* ATCC 25017

Денисова Александра^{1*}

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

РЕЗЮМЕ

В наше время для того, чтобы познакомиться с геномом интересующего организма часто достаточно просто знать его название. В современных базах данных хранится огромное количество геномов разных организмов. При этом, для первичного анализа генома порой достаточно умения пользоваться электронными таблицами. В этой учебной работе на основе данных NCBI составлен анализ особенностей генома и протеома *Francisella philomiragia* subsp. *philomiragia* ATCC 25017 с помощью Google Sheets.

Ключевые слова: *Francisella philomiragia* subsp. *philomiragia*, Google Sheets, genom

1 ВВЕДЕНИЕ

Francisella philomiragia subsp. *philomiragia* ATCC 25017 - штамм аэробной гамма-протеобактерии. Это практически неподвижные широко распространенные мезофильные кокки, являющиеся возбудителями пневмонии и бактериемии, которые обычно проявляются у животных с ослабленным иммунитетом, хроническими гранулематозами. Геном *Francisella philomiragia* subsp. *philomiragia* ATCC 25017 представлен одной хромосомой, длиной 2.05 Мб, кодирующей 1887 белков и все нематричные РНК бактерии, и плазмидой rPHI01, кодирующей 4 белка. (Zeytun et al., 2012). В данной работе проводится анализ генома *Francisella philomiragia* subsp. *philomiragia* ATCC 25017 с помощью электронных таблиц, целью исследования является получение представления о характеристиках бактерии, с помощью данных из ее 'feature table'.

2 METHODS

Работа была выполнена с помощью Google Sheets, потребовались следующие навыки использования данной программы:

Простая таблица

- Фильтр
- Распространение формул
- Сортировка

Оформление таблицы

- Ширина столбцов
- Примечания
- Замена формулы на значение
- Перенос строк

Адресация и распространение индексов

- адресация с использованием \$

Копирование и вставка

- CTRL-A
- CTRL-X
- CTRL-V

Связь таблиц

- VLOOKUP

Статистические функции

- СРЗНАЧ
- СТАНДОТКЛОН
- МИН
- МАКС
- BINOMDIST (случайное распределение генов по цепям)

Сделанные таблицы можно найти в сопроводительных материалах. Информация о геноме *Francisella philomiragia* subsp. *philomiragia* ATCC 25017 была получена с сайта National Center for Biotechnology Information Genom.

3 РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

3.1 Длины белков

Длины белков *Francisella philomiragia* subsp. *philomiragia* ATCC 25017 находятся в интервале от 27 до 2473 аминокислотных остатков (а.о.). Из гистограммы видно, что в геноме бактерии преобладают белки длины от 300 до 330 а.о., а среднее значение длины белка равно 323 а.о..

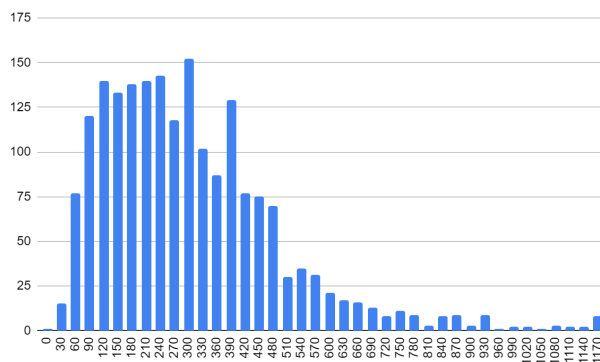


Рис.1 Гистограмма длин белков из протеома *Francisella philomiragia* subsp. *philomiragia* ATCC 25017, по оси Ох число аминокислотных остатков

3.2 Распределение генов по прямой и обратной цепи

По таблице распределения видно, что гены чаще расположены как на прямой, так и на комплементарной цепи в примерно равных соотношениях. На +-цепи бактерии находится 994 гена, а на другой - 976 генов. Вероятность такого распределения 0,702, что говорит о том, что мы можем считать его случайным. В случае генов тРНК, однако, видно, что большая часть последовательностей расположена на прямой цепи, оставим подробное изучение причины такого распределения.

	+	-	Всего
число генов белков	961	954	1915
п.о.	48,78%	48,43%	97,21%
число тРНК	25	14	39
п.о.	1,27%	0,71%	1,98%
число рРНК	4	5	9
п.о.	0,20%	0,25%	0,46%
число псевдогенов	4	3	7
п.о.	0,20%	0,15%	0,36%
общее число генов	994	976	1970
п.о.	50,46%	49,54%	100,00%

Таблица 1. Таблица числа генов белков, псевдогенов и генов РНК на прямой(+) и комплементарной(-) цепочках ДНК.

3.3 Гипотетические белки

Гипотетические белки - это такие белки, существование которых предсказано, но имеющихся данных недостаточно, чтобы определить, происходит ли их экспрессия в организме. Согласно определениям (Galperin et al., 2004) консервативные гипотетические белки, белки, обнаруженные в нескольких организмах, но функционально не охарактеризованные. В таблице признаков *Francisella philomiragia* subsp. *philomiragia* ATCC 25017 обозначено 513 гипотетических белков шести различных категорий.

По данным Таблицы 2 видно, что гены гипотетических белков занимают более 26% генома бактерии, при этом для 6% неизвестно, синтезируются ли они бактерией.

Название гена	Число генов	п.о.
conserved hypothetical protein	345	17,58%
hypothetical protein	117	5,96%
conserved hypothetical membrane protein	45	2,29%
conserved hypothetical lipoprotein	3	0,15%
conserved hypothetical exporter	1	0,05%
conserved hypothetical 4-hydroxybenzoyl-CoA thioesterase	1	0,05%
Общее число гипотетических генов	513	26,13%

Таблица 2. Число гипотетических белков в геноме и процент ото всех белков.

3.4 Рибосомальные белки, рРНК и их расположение

Бактериальная рибосома состоит из 30S и 50S субъединиц. Каждая из субъединиц, соответственно состоит из 21 и 34 белков, общее число рибосомальных белков 55.. Общее число генов рибосомальных белков *Francisella philomiragia* subsp. *philomiragia* ATCC 25017 согласно Таблице 3 составляет 56, при этом ген ribosomal protein S21 встречается три раза на комплементарной цепи. Ген 50S ribosomal protein L7/L12 воспринимается как один, хотя в ходе процессинга должно получаться два белка. Кроме того в таблице был анонсирован ribosomal protein, который был определен как мнимый и не учитывался при расчетах.

Общее название	Число генов	Общее число генов
30S ribosomal protein S	19	23
ribosomal protein S	4	
50S ribosomal protein L	26	33
ribosomal protein L	7	
ribosomal protein		56

Таблица 3. Число генов малой и большой субъединицы рибосомы

Бактериальная рибосома содержит 3 РНК, а именно 16S, 5S и 23S. Генома *Francisella philomiragia* subsp. *philomiragia* ATCC

25017 содержит 9 рРНК, по три каждого класса, которые расположены как на комплементарной, так и на прямой цепи ДНК.

	+	-
16S ribosomal RNA	2	1
5S ribosomal RNA	1	2
23S ribosomal RNA	1	2

Таблица 4. Распределение генов рРНК по прямой (+) и обратной (-) цепям (п.о. - процент от общего числа генов)

ЗАКЛЮЧЕНИЕ

Геном *Francisella philomiragia* subsp. *philomiragia* ATCC 25017 оказался интересным объектом для практики навыков и может оказаться интересным для дальнейшего изучения. В ходе же этой работы были выявлены интересные закономерности.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Сопроводительные материалы доступны на wiki-странице: <https://kodom.fbb.msu.ru/wiki/Users/savouriess2112>

БЛАГОДАРНОСТИ

Хотелось бы поблагодарить Анастасию Сухинину за помощь в открытии таблиц из Excel и за совместное обсуждение результатов.

СПИСОК ЛИТЕРАТУРЫ

Ahmet Zeytun, Stephanie A. Malfatti et al. (2012) Complete genome sequence of *Francisella philomiragia* ATCC 25017
Journal of Bacteriology. 194, 12, 6

Michael Y. Galperin, Eugene V. Koonin(2004) Conserved hypothetical' proteins: prioritization of targets for experimental study.
Published online 2004; 12,2.doi: 10.1093/nar/gkh885