

Мини-обзор генома и протеома бактерии *Saccharomonospora viridis* DSM 43017

Матвейшина Елена

Факультет Биоинженерии и биоинформатики МГУ им. М.В. Ломоносова

РЕЗЮМЕ

Данное исследование не представляет собой особой научной ценности. Оно проводилось как образовательное с целью научиться анализировать генетическую базу данных по генам и белкам, а также фиксировать результаты анализа в форме статьи. По итогам данной работы было рассчитано количество генов у бактерии *Saccharomonospora viridis* DSM 4301, кодирующих белки и РНК, в прямой и обратной цепи, а также выявлена встречаемость белков различной длины.

1 ВВЕДЕНИЕ

Систематическое положение:

- Domain: Bacteria
- Phylum: Actinobacteria
- Class: Actinobacteridae
- Order: Pseudonocardiales
- Family: Pseudonocardiaceae
- Genus: *Saccharomonospora*
- Species: *Saccharomonospora viridis*
- Strain: *Saccharomonospora viridis* DSM 43017

Общие сведения. Патоген человека. Обитает в теплом сене или компосте. И его споры легко распространяются в воздухе и вызывают инфекционное заболевание легких у фермеров, багассоз (заболевание легких, возникающее в результате систематического вдыхания пыли багассы (отход сахарного тростника); могут заболеть рабочие, занятые в производстве изоляционных строительных материалов, картона и др., на изготовление которых идет багасса) и кондиционную лихорадку.

Общие сведения о геноме. Хромосома кольцевая. Длина ДНК 4308349 пар нуклеотидов. Количество генов 3964. Количество белков 3828. Дата секвенирования генома: 2009 год.

2 МАТЕРИАЛЫ И МЕТОДЫ

Файлы с генетическими данными по бактерии *Saccharomonospora viridis* DSM 43017, а также общая информация о ней были взяты с сайта NCBI.

Обработка данных производилась в программе Excel 2013. Т.к. требовалось подсчитывать ячейки, удовлетворяющие определенному набору условий, то в основном использовались функции СЧЕТЕСЛИМН и СЧЕТЕСЛИ (countifs b countif в англоязычных версиях Excel).

3 РЕЗУЛЬТАТЫ

В результате анализа данных о плазидах с сайта NCBI было выяснено, что в геном бактерии *Saccharomonospora viridis* DSM 43017 не входит ни одна плазида, и получается, что представлен только одной хромосомой. (Плазмид нет у бактерий ни этого же вида, 1 плазида представлена у бактерии этого же рода *Saccharomonospora glauca*. Однако найти информацию по геному этой бактерии не удалось.) Для этой хромосомы было посчитано количество генов, кодирующих белки и кодирующих различные виды РНК, на прямой и комплементарной цепях ДНК. Были исследованы различные особенности расположения генов на хромосоме. Кроме того, был исследован длины белков в протеоме бактерии для выявления тенденций во встречаемости белков разной длины.

3.1 Распределение генов, кодирующих белки и РНК, по прямой и комплементарной цепи

По результатам подсчета генов составлена таблица 1. Видно, что количество генов больше в прямой цепи чем в комплементарной и для генов, кодирующих белок, и для генов, кодирующих РНК. Однако, эта разница для генов белков сравнительно незначительная – 10,6% Тогда как для генов кодирующих РНК, их количество на прямой цепи превышает их количество на комплементарной на 90%

Таблица 1. Количество генов, кодирующих белки и РНК в прямой и комплементарной цепи ДНК

Хромосома	Прямая цепь	Комплементарная цепь
Гены белков	2010	1818
Гены РНК	38	20

Была проверена гипотеза о том, что гены распределяют на прямой и комплементарной цепи с вероятностью 0,5. Гипотеза не подтвердилась, т.к. р значение биномиального коэффициента для распределения генов белков составило чуть более 0,001, а для генов РНК чуть более 0,012. Это меньше коэффициента доверия, равного 0,05. Чтобы можно было считать, что гипотеза подтвердилась, наши коэффициенты должны были оказаться больше.

3.2 Квазиопероны

Если гены расположены в цепочке ДНК достаточно близко (расстояние не превышает 100 пар нуклеотидов), то они образуют один квазиоперон. Если же считать, что расстояние, в рамках которого гены образуют один квазиоперон, меньше (к примеру 50 п.н.), только количество генов в одном квазиопероне уменьшится, а значится увеличится количество квазиоперонов. А если рассмотреть большее расстояние (200 п.н.), то количество квазиоперонов уменьшится, т.к. больше генов будут попадать в один квазиоперон. Данные по количеству квазиоперонов для *Saccharomonospora viridis* DSM 43017 представлены в таблице 2.

Таблица 2. Количество квазиоперонов

Длина квазиоперона	Прямая цепь	Комплементарная цепь
50 п.н.	1621	1484
100 п.н.	1443	1309
200 п.н.	1223	1120

3.3 Пересечения генов

Для *Saccharomonospora viridis* DSM 43017 весьма характерны пересечения генов. На прямой цепи пересекаются 387 генов. А на комплементарной 319 генов.

Делимость длины генов на 3

У *Saccharomonospora viridis* DSM 43017 не все гены делятся на 3. Среди генов, кодирующих РНК 44 гена из 58 на 3 не делятся. Что и не удивительно, ведь они кодируют мРНК и рРНК, а при их синтезе не происходит процесс трансляции, а значит наличие триплетов не имеет такого смысла, как при синтезе белка. Но кроме этого есть и 1 ген среди генов, кодирующих белки, который не делится на 3.

3.4 Распределение белков по длинам

Распределение белков по различным длинам представлено на рисунке 1. Видно, что очень маленькие длины (0-50 аминокислот) не характерны для данной бактерии. А уже в следующем диапазоне (50-75 аминокислот) происходит резкий скачок – 111 белков. Средние диапазоны длин белков (от 125 до 350 аминокислот) представлены довольно большим количеством

Рисунок 1. Длины белков бактерии *Saccharomonospora viridis* DSM 43017

белков, порядка 200 каждый. В этих диапазонах находится 2058 из 3828 белков, 53,8%. Максимальное количество белков (263) находится в диапазоне 250-275 аминокислот. В основном все белки сосредоточены в диапазонах до 500. Только 475 белков имеет длину от 500 до 1000 аминокислот, и только 49

свыше 1000. Максимальная встречающаяся длина белка у бактерии: *Saccharomonospora viridis* DSM 43017 1818 аминокислот, а минимальная 37 аминокислот.



ОБСУЖДЕНИЕ

В целом для бактерий характерно распределение длин белков в большинстве в рамках диапазона до 500 аминокислот. Так же характерно небольшое количество длинных (свыше 1000) и коротких (менее 50) белков.

Распределение генов по прямой и комплементарной цепям отличается для генов белков и для генов РНК. Гены белков распределены более-менее равномерно, тогда как генов РНК на прямой цепи почти в 2 раза больше, чем на комплементарной. Возможно такой сильный разброс генов РНК связан с тем, что самих генов РНК сильно меньше чем генов белка (несколько десятков против нескольких тысяч), и для них статистический метод анализа плохо применим, т.к. генов мало, а для бактерий характерно большое количество различных мутаций и видоизменений в геноме.

К тому же надо добавить, что геномы бактерий очень разнообразны, можно встретить множество видоизменений и в принципе говорить о типичности генома или протеома для бактерий довольно сложно.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

[Файл с данными по анализу длин белков и распределения генов по цепям ДНК](#)

[Файл с данными по анализу плазмид](#)

БЛАГОДАРНОСТИ

Выражаю благодарность своим преподавателям по информатике, которые обучали меня основам Excel. А также благодаря своим однокурсникам, которые помогли мне разбираться в том, чего я не понимала.

СПИСОК ЛИТЕРАТУРЫ

Сайт NCBI <http://www.ncbi.nlm.nih.gov/>