

Мини-обзор кодирующих последовательностей генома бактерии *Nitrosomonas europaea* ATCC 19718

Фокина Анастасия Станиславовна

Факультет биоинженерии и биоинформатики МГУ им. М.В. Ломоносова
Ленинские горы МГУ 1 стр. 73, г. Москва, 119234, Российская федерация

a.fokina@kodomo.fbb.msu.ru

Ключевые слова: Nitrosomonas europaea, анализ генома, бактерия, длина белков

Резюме

Мною был проведён статистический анализ генома бактерии *Nitrosomonas europaea* и исследованы его отдельные характеристики, такие как длины белков и их распределение по комплементарным цепям. Было установлено, что большинство белков протеома имеют длину до 1000 аминокислотных остатков, наиболее часто встречающееся значение длины полипептидной цепи (мода) – 347 аминокислотных остатков, а в самом протеоме много белков с неустановленными функциями.

Введение

Nitrosomonas europaea - грам-негативная бактерия семейства Nitrosomonadaceae, облигатный аэроб и хемолитоавтотроф, использующий окисление ионов аммония до нитрит-ионов для получения энергии. Она населяет почву, водоёмы и сточные воды, предпочитает загрязнённые районы с высоким содержанием соединений азота в воздухе.

Оптимальный pH для этого вида 6.0-9.0 почти нейтральный, а оптимальная температура 20-30°C. Она неподвижна и не имеет жгутика, хотя некоторые родственные виды способны передвигаться.

Бактерия катализирует первый этап окисления ионов аммония до нитрат-ионов, участвует в разложении промышленных и бытовых отходов, а её способность разрушать токсичные вещества (в частности, органические галогениды) позволяет использовать её для очистки сточных вод и почвы.¹ Этот организм определяет доступность азотистых соединений для растений, а значит, ограничивает интенсивность фиксации CO₂. Азотистая кислота, выделяемая *Nitrosomonas europaea*, может разрушать камень и другие строительные материалы, ведь бактерия часто поселяется на поверхности зданий, стенах и монументах. Геном этой нитрифицирующей бактерии был секвенирован The Joint Genome Institute в рамках проекта по изучению роли микроорганизмов в круговороте углерода.³

На рисунке 1 изображена микрофотография бактерии (Stan Watson, Woods Hole Oceanographic Institute).

Изучение *Nitrosomonas* затруднено тем, что бактерии крайне медленно делятся (на каждое деление уходит несколько дней), им требуется переработать огромное количество ионов аммония, чтобы накопить достаточно энергии для деления.⁵



Рисунок 1

Транскриптом данного штамма *Nitrosomonas europaea* представлен в виде таблицы, анализ которой позволяет заметить некоторые особенности распределения генов между цепями и по длинам продуктов.

Материалы и методы

Транскриптом бактерии был скачан из базы данных NCBI (National Center for Biotechnology Information) из банка RefSeq⁴.

Данные обрабатывались при помощи программы MS Excel 2007, она применялась для построения гистограммы длин белков и получения статистических данных, при этом использовались встроенные функции МИН, МАКС, МЕДИАНА, СТАНДОТКЛОН, СРЗНАЧ, МОДА, ОКРУГЛЕНИЕ, ДЛСТР, СЧЁТЕСЛИМН и некоторые другие. Плоские и сводную таблицы можно найти в сопроводительных материалах.

Сводная таблица применялась для анализа данных. Генами белков считались строки таблицы со значением 'CDS' в колонке '#feature' и значением 'with_protein' в колонке 'class', а генами РНК строки таблицы, где в колонке '#feature' был указан вид РНК (miscRNA, rRNA, tRNA), псевдогенами – со значением 'pseudogene' в колонке 'class'. Функции белков определялись по значению в столбце 'name', число белков 'other' получено вычитанием белков других видов из общего числа белков.

Оценка справедливости гипотезы о случайном распределении генов по цепям проводилась с использованием функции БИНОМРАСП, если она выдавала значение сильно меньше 1, значит, распределение случайно, если значение близко к 1 - неслучайно.

Результаты и обсуждение

1. Распределение генов по категориям.

Всего геном изучаемого штамма содержит 2623 гена, кодирующих белки и РНК, их распределение по категориям продуктов представлено в таблице 1. Число генов на 1 млн пар оснований было получено делением общего числа генов на длину генома (эти данные мне не удалось найти, поэтому я приняла за длину генома в парах оснований координату конца последнего гена – 2811935 пар оснований) и умножением на миллион пар оснований. Число генов на млн пар оснований – 933.

Таблица 1

Количество по полю ID	Названия столбцов								
Названия строк	misc_RNA	protein_coding	pseudogene	rRNA	tRNA	with_protein	without_protein	(пусто)	Общий итог
CDS						2462	113		2575
gene	3	2462	113	4	41				2623
misc_RNA								3	3
rRNA								4	4
tRNA								41	41
(пусто)									
Общий итог	3	2462	113	4	41	2462	113	48	5246

Белок кодируют 2462 генов, а РНК – 48 генов. Небольшое относительно общего числа генов количество псевдогенов свидетельствует о том, что, несмотря на затруднения, возникающие при культивировании данного штамма, его геном неплохо изучен. В геноме также присутствуют гены РНК с невыясненной функцией (misc-RNA), но больше всего транспортных (41).

В таблице 3 приведено распределение белков по их функциям, в колонку others выделены белки, функции которых не были отдельно рассмотрены в рамках данного мини-обзора. Функции этих белков установлены и играют важную роль в процессах жизнедеятельности клеток. Из белков, функции которых я рассматривала, больше всего протеинов с невыясненным значением (hypothetical protein, 616), а значит, изучение протеома этой бактерии ещё не закончено и функции многих белков предстоит установить. Белков, выполняющих транспортные функции, 96, участвующих в формировании рибосом, 61, мембранных белков 131, то есть это самый большой класс полипептидов после гипотетических.

2. Распределение белков по длинам.

Для изучения распределения белков по длине полипептидной цепи была составлена гистограмма, изображённая на Рисунке 2. Карманы гистограммы подбирались вручную, они разной длины, чтобы в каждом из карманов набралось статистически значимое число продуктов транскрипции генов. Больше всего оказалось коротких полипептидов, особенно много белков длины 101-150 аминокислот. После достижения длины в 500 аминокислотных остатков число белков определённой длины резко снижается, и в последних диапазонах (1501-1900 и 1901-3100), несмотря на их величину, очень мало белков.

В таблице 2 отражены статистические данные, полученные при анализе длин белков. Минимальная длина составляет 42 аминокислотных остатка, а максимальная – 3064. Среднее значение длины полипептидной цепи – 325 аминокислот, стандартное отклонение позволяет нам утверждать, что большинство белков имеют длину от 93 до 557 аминокислотных остатков, эти значения относительно невелики, значит исследуемая бактерия не нуждается в белках значительной длины. Интересно было бы узнать, какие функции выполняют самые короткие и самые длинные белки, и каково происхождение белков, имеющих самую большую длину среди всех белков протеома – 3064 аминокислотных остатка, потому что длину больше 1000 аминокислот имеют лишь 45 белков, а значит, столь длинные белки необычны для протеома данной бактерии. В таблице можно увидеть и другие статистические показатели: моду и медиану, которые, возможно, дают лучшее представление о длине всех белков протеома.

Таблица 2. Статистические показатели.

статистический показатель	значение
<i>минимум</i>	42
<i>максимум</i>	3064
<i>среднее значение</i>	324,9679
<i>стандартное отклонение</i>	231,8515
<i>медиана</i>	274,5
<i>мода</i>	347

Рисунок 2. Гистограмма распределения белков по числу аминокислотных остатков в полипептидной



3. Распределение генов по комплементарным цепям ДНК.

В таблице, которая представляет собой изначальные данные, в колонке strand указана цепь ДНК ('+' – прямая, '-' – обратная), на которой расположена рассматриваемая кодирующая последовательность. В таблице 3 проиллюстрировано распределение генов по комплементарным цепям, и видно, что число генов на прямой и обратной цепи примерно одинаково для генов белков, псевдогенов и генов РНК.

Таблица 3. Распределение генов по комплементарным цепям ДНК.

число генов	гены белков	псевдогены	гены рнк	тРНК	рРНК	misc-RNA	hypothetical protein	transport protein	ribosomal protein	membrane protein	other protein
прямая цепь	1148	54	27	23	4	2	282	48	37	49	732
обратная цепь	1314	59	18	18	0	1	334	48	24	82	826
всего	2462	113	45	41	4	3	616	96	61	131	1558

Сопроводительные материалы:

Файл Excel, на отдельных листах результаты выполнения каждого из заданий 13 практикума, описание которых представлено в данном отчёте.

Сопроводительные таблицы: https://kodomofbb.msu.ru/~a.fokina/term1/pr_14.xlsx

Список литературы:

1. Caranto, Jonathan D.; Lancaster, Kyle M. (2017-08-01). "Nitric oxide is an obligate bacterial nitrification intermediate produced by hydroxylamine oxidoreductase". *Proceedings of the National Academy of Sciences*. 114 (31): 8217–8222
2. Patrick Chain; et al. (May 2003). "Complete Genome Sequence of the Ammonia-Oxidizing Bacterium and Obligate Chemolithoautotroph *Nitrosomonas europaea*". *Journal of Bacteriology*. 185 (9): 2759–2773
3. http://genome.jgi-psf.org/80/finished_microbes/niteu/niteu.home.html
4. https://www.ncbi.nlm.nih.gov/genome/1013?genome_assembly_id=300478
5. <http://commtechlab.msu.edu/sites/dlc-me/zoo/zap0534.html>
6. <http://www.reed.edu/~davidd/thesis.html>
7. http://test.bsi.vt.edu/facultyfiles/biol_4684/Microbes/nitro.html