

Московский государственный университет имени М. В. Ломоносова

Мини-обзор генома бактерии *Martelella endophytica*

Автор: Мурзин Владислав Андреевич, студент 1 курса Факультета биоинженерии и биоинформатики

Резюме

Целью работы явилось исследование генома бактерии *Martelella endophytica* с получением количественных характеристик состава генов. Геном бактерии был проанализирован в данном мини-обзоре с выявлением ряда особенностей, таких как нуклеотидный состав, соотношение количеств генов разной длины и на разных цепях ДНК, степень покрытия генома генами.

Ключевые слова

Martelella endophytica, геном, протеом, нуклеотиды

Введение

Martelella endophytica – грамотрицательная аэробная бактерия класса Альфа-протеобактерий, обнаруженная в корнях галофитного растения Шиповник морщинистый (*Rosa rugosa*). Оптимальные параметры среды для роста бактерии – температура 28-30 градусов по Цельсию, рН в пределах от 7,0 до 8,5. Для осуществления дыхания использует убихинон Q-10. Для штамма YC6887T было показано ингибирование роста патогенных для растения оомицетов [1].

Размер генома – 4,8 Мб, представлен одной кольцевой хромосомой, без плазмид. Содержание GC пар в ДНК бактерии составляет 62,1% [2]. Хромосома содержит 4028 генов, с которых может быть получено 4090 различных РНК-транскриптов, и 62 псевдогена [3].

Материалы и методы

В ходе анализа генома бактерии применялись следующие методы работы с электронными таблицами:

- Вырезание, вставка, копирование и удаление содержимого столбцов и отдельных ячеек;
- Перенос данных между таблицами;
- Написание, комбинирование и распространение формул;
- Сортировка данных в пределах одного столбца;
- Избирательное отображение строк при помощи инструмента «Фильтр»;
- Визуализация данных при помощи построения гистограмм.

Результаты и обсуждение

Размер генома и его покрытие генами

Длина плазмидной ДНК бактерии составляет 4817335 нуклеотидов, что подтверждается как подсчётом нуклеотидов в файле GCF_000960975.1_ASM96097v1_genomic.fna.gz из страницы о геноме бактерии на NCBI [4], так и интернет-источником [3]. В состав генов включено 4315154 нуклеотида, покрытие генами составляет 89,58%. Такой высокий процент кодирующей ДНК свойственен бактериям, в частности, у *E. coli* покрытие генами составляет около 88%, а у *Methylophaga nitratireducenticrescens*, геном которой стал объектом изучения моего коллеги, - 89% [5]. Отчасти это обеспечивается меньшим количеством регуляторных элементов и перекрыванием некоторых кодирующих последовательностей.

Нуклеотидный состав генома

Количественный состав нуклеотидов в геноме *Martelella endophytica*:

- A: 911830;

- T: 912102;

- G: 1495946;

- C: 1497292;

- A/G: 6;
- T/C: 7;
- G/T: 3;
- A/C: 6;
- G/C: 4;
- A/T: 2;
- 137 невыясненных нуклеотидов.

Доля GC в геноме бактерии составляет примерно 62,1%, при этом количества А и Т примерно совпадают, так же, как и количества G и C.

Состав генов, кодирующих продукты разных классов

Всего в геноме *Martelella endophytica* 4537 участков, кодирующих различные транскрипты. Они примерно равномерно распределены между цепями ДНК. В таблице 1 приведено распределение генов продуктов разных классов по цепям ДНК.

Класс транскрипта	Количество	+ цепь	- цепь
некодирующие РНК (с неустановленной функцией)	1	1	0
белки	4400	2101	2299
псевдогены	71	39	32
РНК из РНКазы Р	1	0	1
рРНК	9	0	9
7S-SRP-РНК	1	1	0
транспортно-матричные РНК	1	0	1
тРНК	53	23	30
Всего	4537	2165	2372

Таблица 1: распределение генов по прямой (+) и обратной (-) цепям ДНК.

Длины белков

В протеоме *Martelella endophytica* 4400 белковых продуктов, из которых у 479 не выяснены функции. Распределение белков по числу аминокислотных остатков и цепи, на которой расположены их гены, приведено в таблице 2. Гистограмма длин белков приведена на рисунке 1, гистограмма распределения продуктов по цепям ДНК - на рисунке 2. Максимальные количества белков соответствуют длине в 200-300 аминокислот,

максимальная длина белка – 2829 аминокислотных остатков (гликозилтрансфераза),
 минимальная – 29 аминокислотных остатков (F-субъединица калиевой АТФазы).

Число аминокислотных остатков	Количество	+ цепь	- цепь
1-101	312	156	156
101-201	875	398	477
201-301	1169	601	568
301-401	993	463	530
401-501	500	230	270
501-601	229	110	119
601-701	120	54	66
701-801	66	23	43
801-901	56	29	27
901-1001	25	12	13
1001-1501	43	21	22
1501-2001	4	0	4
2001-2501	7	3	4

Таблица 2: распределение белков по длине и цепи гена

Распределение продуктов по положению их генов

На рисунке 2 приведена гистограмма, показывающая распределение генов различных продуктов по цепям ДНК. Видно, что это распределение достаточно случайное, за исключением генов рибосомальных РНК, которые все (в количестве 9 штук) собраны на обратной цепи.

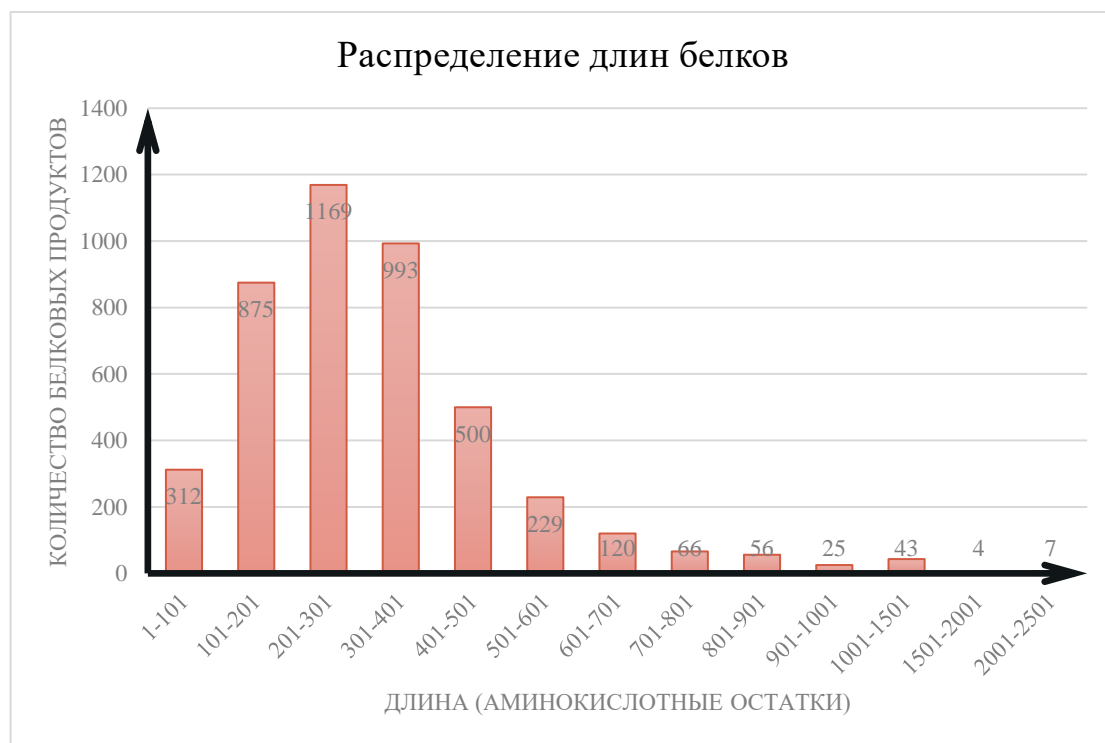


Рисунок 1: распределение длин белков

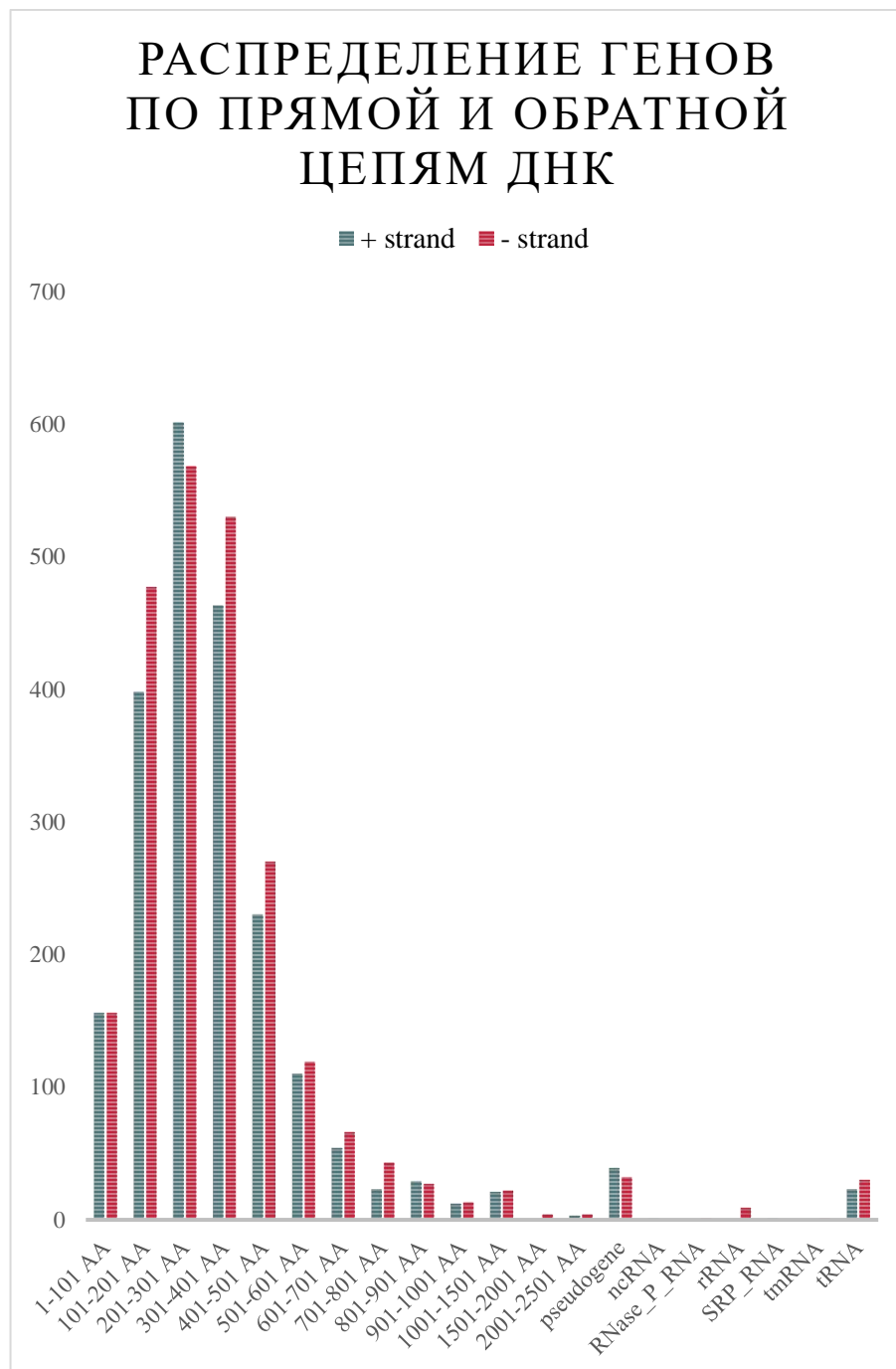


Рисунок 2: распределение генов различных продуктов по цепям ДНК. AA – аминокислотные остатки, синие столбцы – прямая цепь, красные столбцы – обратная цепь. Pseudogene – псевдогены, ncRNA – некодирующие РНК с невыясненной функцией, RNase_P_RNA – РНК из РНКазы P, rRNA – рибосомальные РНК, SRP_RNA – 7S-РНК из сигнал-распознающей частицы, tm_RNA – транспортно-матричные РНК, tRNA – транспортные РНК.

Анализ количества некоторых k-меров

В таблице 3 приведены встречаемости некоторых k-меров в геноме бактерии *Martelella endophytica*. Для их подсчёта использовалась программа, написанная мной на языке программирования python 3, ссылка на которую приведена в разделе «Сопроводительные материалы». Ожидаемое значение рассчитывается по формуле $\frac{4817335}{4^n}$, где 4817335 пар нуклеотидов – размер генома, а n – длина k-мера. Можно заметить, что, например, сайт метилирования встречается в геноме намного чаще, чем ожидалось бы для последовательности длины 4, что объясняется его важностью для защиты бактерии от чужеродной ДНК и различения старой и новосинтезированной цепей ДНК при репарации. Прибнов бокс, напротив, встречается намного реже, чем ожидается, вероятно, из-за способа организации бактериальных генов в опероны [7]. Количества сайтов рестрикции не дают чётких закономерностей, тем не менее, можно заметить, что сайт рестриктазы EcoRI, выделенной из кишечной палочки, встречается намного реже ожидаемого, в то время как сайт рестриктазы RtrI, выделенной из почвенной бактерии, близкой по систематике к *Martelella endophytica*, – намного чаще ожидаемого. Возможно, это можно объяснить степенью конкуренции между видами.

Название последовательности	Последовательность	Ожидаемое количество (для 1 цепи)	Реальное количество на прямой цепи	Реальное количество на обратной цепи
Шайна-Дальгарно [7]	AGGAGG	1176	1099	1107
Сайт метилирования	GATC	18818	41372	41372
Прибнов бокс [7]	TATAAT	1176	158	139
Сайт EcoRI [6]	GAATTC	1176	477	524
Сайт BamHI [6]	GGATCC	1176	793	793
Сайт AgeI [6]	ACCGGT	1176	1468	1468
Сайт PstI [6]	CTGCAG	1176	2698	2698
Сайт RtrI [6]	GTCGAC	1176	3115	3115

Таблица 3: встречаемость различных k-меров в геноме *Martelella endophytica*

Заключение

Данное исследование позволило выявить ряд закономерностей, свойственных для генома и протеома бактерии *Martelella endophytica*, а именно: процент GC в плазмидной ДНК, распределение белков по числу аминокислотных остатков, примерно случайное распределение генов по прямой и обратной цепям ДНК. Возможно, эти скромные данные помогут чуть лучше узнать малоизвестную бактерию *Martelella endophytica*.

Сопроводительные материалы

Электронная таблица с сопроводительными материалами доступна по ссылке

https://kodomo.fbb.msu.ru/~vladislove2020/term1/Murzin_supple.xlsx

Программа для подсчёта встречаемости k-меров доступна по ссылке:

<https://kodomo.fbb.msu.ru/~vladislove2020/term1/analysis.py>

Список литературы

- 1) Fehmida Bibi, “Martellella endophytica sp. nov., an antifungal bacterium associated with a halophyte”, *Int J Syst Evol Microbiol.* 2013 Aug;63(Pt 8):2914-2919, Gyeongsang National University
- 2) Ajmal Khan, Haji Khan, Eu Jin Chung, Mohammad T. Hossain, “Complete Genome Sequence of Martellella endophytica YC6887”, *Genome Announc.*, 2015 May 7;3(3):e00366-15 , Gyeongsang National University
- 3) Ensembl Bacteria release 48 - August 2020
- 4) Директория с данными о геноме бактерии *M. endophytica* на сайте NCBI ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/960/975/GCF_000960975.1_ASM96097v1
- 5) Nagimov R D, «Обзор генома и протеома бактерии *Methylophaga nitratireducentescens*», Moscow State University <https://kodomo.fbb.msu.ru/wiki/Users/nagrusbb/block3>
- 6) Список рестриктаз с сайтами разрезания и бактериями, из которых они были получены) https://ru.qaz.wiki/wiki/List_of_restriction_enzyme_cutting_sites
- 7) Д. Нельсон, М. Кокс, «Основы биохимии Ленинджера», том 3, изд. «Лаборатория знаний», ISBN: 978-5-00101-864-3, 978-5-00101-863-6, 2008