
Обзор протеома бактерии *Roseobacter denitrificans*

Автор: Янкович Иван Даркович

РЕЗЮМЕ

В этой работе описывается исследование генома и протеома бактерии *roseobacter denitrificans*, в котором были изучены соотношения различных нуклеотидов в последовательности ДНК, проанализировано распределение длин белков бактерии и расположение на прямой и обратной цепи ДНК рибосомальных белков и РНК. Кроме этого были определены функциональные классы закодированных в геноме РНК и соотношение количеств различных тРНК.

ВВЕДЕНИЕ

Roseobacter denitrificans это свободноживущая аэробная бактерия с розовой окраской. Клетки имеют яйцевидную или палочковидную форму и двигаются за счет жгутиков. Бактерия обитает в почве, на поверхности океана и вблизи гидротермальных источников. Бактерия не производит кислород, но он ей требуется для роста. Причем необходим низкий уровень кислорода. При чрезмерном количестве фотосинтетический аппарат клетки угнетается и бактерия растет хемотрофно с использованием органического углерода.¹ Этот вид уникален тем, что может двигаться без образования розеток. Первичный фотосинтетический аппарат – бактериохлорофилл а. Бактерия может перерабатывать органические фосфаты. Так как у нее апоксигенный фотосинтез, она окисляет монооксид углерода парникового газа и производит диметилсульфид. За счет разложения водорослей она влияет на климат.¹ Для человека безвредна и не вызывает никаких патологий. Бактерию применяют для изучения воздействия кислорода на фотосинтетические и метаболические процессы, так как она способна к росту без кислорода, используя нитрат в качестве акцептора электронов.

МЕТОДЫ

Информация о геноме бактерии получена с использованием базы данных NCBI.⁴

Функции Excel, использованные в работе:

- Фильтр (выбрать определенные столбцы)
- Сортировка(упорядочить строки)
- ВПР(связь таблиц,принимает на вход 4 аргумента)
- Функции СЧЁТЕСЛИ и СЧЁТЕСЛИМН (подсчет количества значений столбца данных)
- Функция ЕСЛИ и ЕСЛИМН (оценить значения в столбцах данных по одному или нескольким критериям)
- Адресация с использованием \$ (чтобы данные не менялись и сдвигались Знак \$ перед обозначением столбца или строки сохраняет соответствующие)
- Распространение формул осуществляется путем выделения ячейки с формулой и растягивания выделенной прямоугольной области

Программы Linux, использованные в работе:

- wordcount -wordsize n : подсчет всех возможных последовательностей длины n.

РЕЗУЛЬТАТЫ

1. Размер генома

Согласно NCBI, размер генома *roseobacter denitrificans* составляет 2909567 нуклеотида, Это относительно немного, но соответствует среднему диапазону для геномов Bacteria.

1.1 Встречаемость азотистых оснований

Анализ встречаемости каждого нуклеотида с помощью программы wordcount -wordsize 1 приводит к результату, отраженному в таблице 1.

Nucleotide	Amount	% content
T	619085	21,27756
A	622404	21,39164
G	834016	28,66461

С	834062	28,66619
---	--------	----------

Таблица 1

В соответствии с полученными данными, в геноме *roseobacter denitrificans* встречаются азотистые основания: аденин, тимин, гуанин и цитозин, причем чаще всего встречается цитозин (834062 нуклеотидов), а реже всего – тимин (619085 нуклеотида). Кроме того, количества нуклеотидов, содержащих аденин и тимин, а также гуанин и цитозин приблизительно равны, что подтверждает верность первого правила Чаргаффа.

2. Типы генов

Результат исследования типов генов, находящихся в геноме *roseobacter denitrificans*, позволяет получить следующую таблицу (таблица 2) с их встречаемостью.

Таблица 2

protein_coding	3947
Pseudogene	37
rRNA	3
tRNA	38

По полученным данным, можно отметить, что для данной бактерии характерно наличие большого количества генов, кодирующих белки (3947) для генома такого размера.

3. Длины белков



На гистограмме показана частота встречаемости белков различной длины в протеоме *roseobacter denitrificans*. Согласно исследованию, длины большей части генов лежат в пределах от 1000 до 3000 аминокислотных остатков, а пик гистограммы соответствует интервалу 2000-2500 остатков (839 белков).

4. Анализ РНК генома

В результате анализа генов, кодирующих РНК, было выяснено, что в геноме *roseobacter denitrificans*

закодировано 44 РНК, в таблице 3 представлено распределение РНК по функциям.

Было сделано исследование видов тРНК, присутствующих в геноме, в результате чего было выяснено, что в больше всего в геноме бактерии представлены гены тРНК, присоединяющей лейцин (5 генов), а также аргинин и серин (по 4 гена). Это распределение коррелирует с распределением количеств триплетов, кодирующих каждую аминокислоту (лейцин, аргинин, серин – 6; аспарагин, аспартат, цистеин, фенилаланин, тирозин – 2, триптофан – 1), что может говорить о том, какие аминокислоты чаще встречаются в белках бактерии.

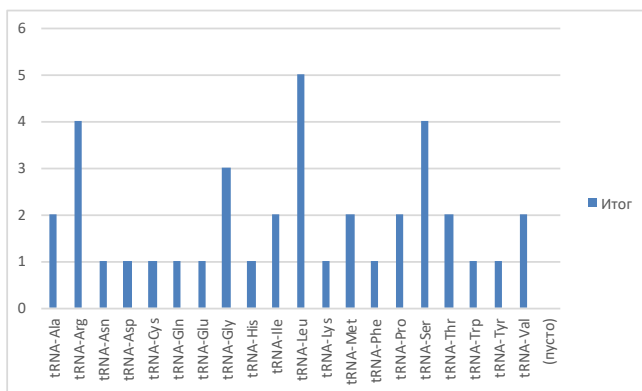


Таблица 3

Тип РНК	Количество
рибосомальная РНК	2
некодирующая РНК	3
транспортно-матричная РНК	1
транспортная РНК	38

5. Расположение и количество рибосомальных генов и РНК

Согласно проведенному исследованию, в геноме *roseobacter denitrificans* содержится по одному гену 16S рРНК, 23S рРНК и 5S рРНК, а также 60 рибосомальных белков, из которых 25 составляют малую субъединицу РНК (белки S), а 35 – большую (белки L). Молекул 23S и 5S РНК входят в состав большой субъединицы рибосомы, а 16S РНК – в состав малой. Большая часть генов рибосомальных РНК и белков расположена на прямой цепи ДНК. Данные о координатах генов РНК в таблицах 4 и 5.

Таблица 4

Name	start	End	strand
------	-------	-----	--------

16S ribosomal RNA	2859452	2860915	+
23S ribosomal RNA	2861750	2864580	+
5S ribosomal RNA	2864666	2864780	+

Таблица 5

Name	Start	End	strand
30S ribosomal protein S6	93557	93643	-
50S ribosomal protein L11 methyltransferase	15998	16870	+
30S ribosomal protein S9	518306	518797	-
50S ribosomal protein L13	518800	519261	-
30S ribosomal protein S2	555313	556095	-
50S ribosomal protein L27	564824	565093	-
50S ribosomal protein L21	565105	565755	-
30S ribosomal protein S6--L-glutamate ligase	858852	860231	+
30S ribosomal protein S6	936761	937114	+
30S ribosomal protein S18	937141	937368	+
50S ribosomal protein L9	937379	938014	+
50S ribosomal protein L32	1051224	1051430	+
30S ribosomal protein S4	1281747	1282367	-
50S ribosomal protein L11 methyltransferase	1588808	1590052	+
50S ribosomal protein L25/general stress protein Ctc	1789972	1790610	+
30S ribosomal protein S1	1798174	1799853	-
30S ribosomal protein S7	1892719	1893189	-
30S ribosomal protein S12	1893202	1893573	-
50S ribosomal protein L7/L12	1904001	1904372	-

50S ribosomal protein L10	1904447	1904962	-
50S ribosomal protein L1	1905202	1905900	-
50S ribosomal protein L11	1905902	1906327	-
30S ribosomal protein S20	2360396	2360659	+
30S ribosomal protein S15	2780746	2781015	-
50S ribosomal protein L36	2848592	2848717	-
50S ribosomal protein L35	3127565	3127765	+
50S ribosomal protein L20	3127778	3128143	+
30S ribosomal protein S21	3158854	3159060	-
50S ribosomal protein L31	3419502	3419723	-
50S ribosomal protein L19	3419735	3420118	-
30S ribosomal protein S16	3422562	3422930	-
30S ribosomal protein S10	3498417	3498734	+
50S ribosomal protein L3	3498755	3499615	+
50S ribosomal protein L4	3499612	3500229	+
50S ribosomal protein L23	3500226	3500522	+
50S ribosomal protein L2	3503098	3503940	+
30S ribosomal protein S19	3503944	3504222	+
50S ribosomal protein L22	3504219	3504599	+
30S ribosomal protein S3	3504599	3505306	+
50S ribosomal protein L16	3505320	3505733	+
50S ribosomal protein L29	3511745	3511951	+
30S ribosomal protein S17	3511963	3512202	+
50S ribosomal protein L14	3512272	3512640	+
50S ribosomal protein L24	3512643	3512954	+
50S ribosomal protein L5	3512947	3513510	+

30S ribosomal protein S14	3513531	3513836	+
30S ribosomal protein S8	3513849	3514241	+
50S ribosomal protein L6	3514252	3514785	+
50S ribosomal protein L18	3514797	3515156	+
30S ribosomal protein S5	3515263	3515838	+
50S ribosomal protein L30	3515851	3516039	+
50S ribosomal protein L15	3517604	3518086	+
30S ribosomal protein S13	3521745	3522113	+
30S ribosomal protein S11	3522129	3522518	+
50S ribosomal protein L17	3523779	3524201	+
50S ribosomal protein L34	3542914	3543048	-
30S ribosomal protein S12 methylthiotransferase RimO	3749890	3751269	-
50S ribosomal protein L28	3773699	3773998	+
50S ribosomal protein L33	4062215	4062382	-

СПИСОК ЛИТЕРАТУРЫ

1. Microbes wiki roseobacter denitrificans
https://microbewiki.kenyon.edu/index.php/Roseobacter_denitrificans
2. NCBI – assembly Roseobacter denitrificans.
https://www.ncbi.nlm.nih.gov/assembly/GCF_002983865.1
3. BioCyc database 01-June-2020 version 24.5
Summary of Roseobacter denitrificans, Strain OCh 114
Authors: P.S, P.M, I.K, A.K, R.C, P.K
https://biocyc.org/GCF_000014045/organism-summary
4. Директория с данными о геноме Roseobacter denitrificans на сайте NCBI
ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/983/865/GCF_002983865.1_ASM298386v1

Заключение:

В результате анализа свойств генома и протеома бактерии были определены некоторые важные для последующих исследований параметры, такие как распределение генов на прямой и обратной цепи ДНК, вероятно является случайным. Данные признаки могут быть важны для дальнейшей работы с культурой бактерии и модифицирования её генома с теми или иными целями. Кроме того определены максимальные, минимальные и наиболее часто встречающиеся длины белков изучаемого организма и встречаемость в геноме тРНК связывающихся с различными аминокислотами.

Сопроводительные материалы:

Хромосомная таблица и материалы исследования:
https://kodomofbb.msu.ru/~ivan.yankovich/term1/ivan.yankovich_suppl.xlsx