

Обзор протеома бактерии *Roseateles depolymerans*.

Бартыш Екатерина¹

¹Факультет биоинженерии и биоинформатики МГУ им. М. В. Ломоносова.

АННОТАЦИЯ

В данной работе я исследовала протеом бактерии *Roseateles depolymerans*. Были изучены особенности распределения длин белков, закономерности распределения белок-кодирующих и РНК-кодирующих генов в зависимости от направленности цепочки ДНК, на которой они находятся, а также исследованы пересечения генов и изучены категории достоверности существования белков.

1 ВВЕДЕНИЕ

Классификация	
Домен	Prokaryota
Царство	Bacteria
Тип	Proteobacteria
Класс	Beta Proteobacteria
Порядок	Burkholderiales
Семейство	Comamonadaceae
Род	<i>Roseateles</i>
Вид	<i>R. depolymerans</i>

Бактерия *Roseateles depolymerans* обитает в водной среде. Она является анаэробом, мезофилом (предпочитает температуру около 35 градусов по Цельсию) и миксотрофом. Болезни, вызываемые данной бактерией, еще не определены. Длина генома – 5 681 722. Количество генов - 4837. Геном был секвенирован 30 декабря 2015 года в Исследовательском Институте Биологии и Биотехнологии в Корее [1]. Сама бактерия была обнаружена еще в 1999 году японскими учеными [2].

2 МАТЕРИАЛЫ И МЕТОДЫ

Данные о протеоме бактерии были взяты с сайта NCBI [3]. Файл GCF_001483865.1_ASM148386v1_feature_table.txt.gz, содержащий информацию о белках и РНК организма, использовался для анализа информации.

Данные были упорядочены и изучены с помощью электронной таблицы OpenOffice 4.1.1. . Мною было подсчитано количество генов белков и генов РНК, была построена гистограмма длин белков и таблица числа генов белков и генов РНК на прямой и комплементарной цепочках ДНК, а также были изучены количественные характеристики пересечения генов и составлена статистика белков по категориям достоверности их существования.

Для подсчета были использованы такие функции как: «IF()», «COUNTIFS()», «MAX()», «MIN()» и т. д.

3 РЕЗУЛЬТАТЫ

3.1 Белки организма

В бактерии *Roseateles depolymerans* всего закодировано 4773 белков, статистические данные которых представлены в Таблице 1.

Характеристика	Значение
Минимальная длина, а.к.	29
Максимальная длина, а.к.	4828
Медиана длин, а.к.	296
Средняя длина, а.к.	347,05

Таблица 1. Данные о белках бактерии *Roseateles depolymerans*

Распределение длин белков можно пронаблюдать на Диаграмме 1. По оси x – длина белков, по оси y – количество белков с такой длиной.

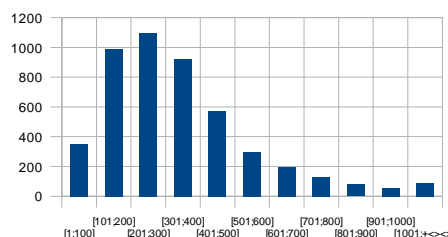


Диаграмма 1. Гистограмма распределения длин белков бактерии *Roseateles depolymerans*

Из гистограммы видно, что наиболее встречаемыми являются белки длиной от 101 до 400 аминокислотных остатков, их количество колеблется от 920 до 1096. Далее с увеличением длины количество белков уменьшается.

3.2 РНК организма

Всего в бактерии *Roseateles depolymerans* 71 ген кодирует различные РНК. В Таблице 2 приведены количественные характеристики каждого встречающегося вида РНК.

tRNA	ncRNA	tmRNA	rRNA
57	1	1	12

Таблица 2. Количество генов, кодирующих разные виды РНК.

Из приведенной выше таблицы видно, что наиболее встречаемыми РНК являются транспортная и рибосомальная РНК. Также в геноме содержится ген одной некодирующей РНК и транспортно-матричной РНК. Транспортно-матричная РНК участвует в терминеции трансляции по механизму, называемому транс-трансляцией, когда трансляция неполных, не содержащих стоп-кодонов мРНК приводит к тому, что рибосома задерживается на таких РНК, так как невозможна нормальная терминеция трансляции и диссоциация рибосомных субъединиц [4].

3.3 Анализ генома

Длина генома — 5 681 722 п.н. . Всего в бактерии *Roseateles depolymerans* 4844 гена. Получается, что на 1 млн п.н. приходится примерно 853 гена. Геном бактерии представлен кольцевой хромосомой [5]. В Таблице 3 представлено количественное распределение генов по прямым и обратным цепям ДНК.

	Прямая цепь	Комплементарная цепь
Число генов белков	2299	2474
Число генов РНК	41	30

Таблица 3. Распределение генов по прямой и обратной цепи ДНК

Как видно из таблицы, гены распределены достаточно равномерно с незначительными превышениями на прямой цепи как генов белков, так и генов РНК.

3.4 Пересечения генов

В Таблице 4 показано, как количественно распределяются пересечения генов в цепочках.

Количество пересечений на прямой цепи	349
Количество пересечений на обратной цепи	342

Таблица 4. Количественное распределение пересечений генов по цепям

Из таблицы видно, что в данном геноме довольно большое количество пересекающихся генов.

Я составила гистограмму длин участков пересекающихся генов на прямой и обратной цепи, которую можно увидеть на Диаграмме 2.

По оси x – длина общего участка, по оси y — количество таких участков.

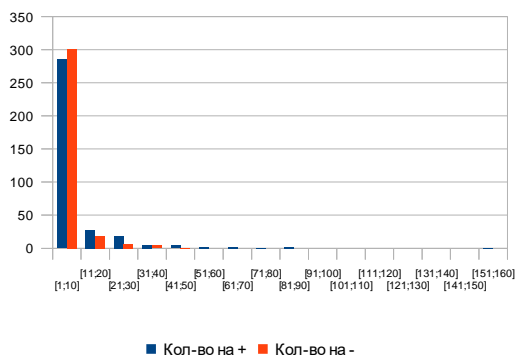


Диаграмма 2. Гистограмма длин участков пересекающихся генов на прямой и обратной цепи ДНК.

Как видно из диаграммы, наибольшая встречаемость принадлежит коротким участкам, причем их намного больше, чем участков других длин. Стоит отметить, что количество таких отрезков на прямой цепи больше, чем на обратной цепи (не считая коротких участков). Также интересным является появление участков большой длины, например, отрезки длиной в 160 или 85 аминокислотных остатков. Стоит также отметить, что такие длинные участки присутствуют только на прямой цепи ДНК.

3.5 Статистика белков по категориям достоверности их существования

В Таблице 5 я привела количественные характеристики для каждой категории достоверности существования белка для бактерии *Roseateles depolymerans*.

Evidence at transcript level	2
Inferred from homology	1248
Predicted	3518

Таблица 5. Количественные характеристики для каждой категории достоверности существования белка.

Evidence at transcript level — значит, что присутствие белка еще не доказано, однако была зафиксирована транскрипция.

Inferred from homology – указывает на то, что присутствие белка возможно, так как он есть у близких к данному виду видов.

Predicted – значит, что никаких четких доказательств нахождения данного белка в геноме нет. [6]

Перед статистическим подсчетом я хочу подметить, что по категориям распределены не все белки. Всего генов, кодирующих белки 4773, однако, если мы подсчитаем общее количество белков из Таблицы 5, то мы получим 4768 белков, то есть получается, что еще 5 генов не распределены по категориям. Я думаю, это связано с тем, что статус этих белков еще не определен точно, поэтому они нигде не отнесены.

Белки «Evidence at transcript level» — 0,04 %

Белки «Inferred from homology» — 26,15%

Белки «Predicted» — 73,71%

Остальные белки — 0,1%

4 ОБСУЖДЕНИЕ И ЗАКЛЮЧЕНИЕ

В этом сегменте я приведу общую сводку выводов, сделанных мною по ходу исследования.

1. Наиболее встречаемыми белками у бактерии *Roseateles depolymerans* являются белки длиной от 101 до 400 аминокислотных остатков, их количество колеблется от 920 до 1096. Далее с увеличением длины количество белков уменьшается. Это типично для бактерии, поэтому в этом нет ничего необычного.

2. Наиболее встречаемыми РНК у бактерии *Roseateles depolymerans* являются транспортная и рибосомальная РНК. Также в геноме содержится ген одной некодирующей РНК и транспортно-матричной РНК. По моему мнению, транспортно-матричная РНК является одной из интересных особенностей бактерии, и её роль в протеоме требует дальнейших исследований.

3. Длина генома бактерии *Roseateles depolymerans* — 5 681 722 п.н. . Всего в бактерии *Roseateles depolymerans* 4844 гена. Получается, что на 1 млн п.н. приходится примерно 853 гена. Также гены и белков, и РНК распределены на обратной и прямой цепи почти равномерно с незначительным превышением на прямой цепи. Это не является особенностью данной бактерии

4. Наибольшая встречаемость пересекающихся участков генов бактерии *Roseateles depolymerans* принадлежит коротким участкам. Еще одной интересной особенностью бактерии является наличие длинных пересекающихся участков, например, отрезки длиной в 160 или 85 аминокислотных остатков.

5. Статистика белков по категориям достоверности их существования показала, что белки данной бактерии нуждаются в дальнейшем изучении, так как большая часть из них еще достоверно не определена.

6. Хочу обратить внимание на то, что на сайте NCBI указано количество генов равно 4837, в то время как в плоской таблице их количество равно 4844. Вероятно, информация на сайте слегка устарела.

В итоге, по моему мнению, бактерия *Roseateles depolymerans* очень интересна для изучения, так как вместе с типичными свойствами обладает рядом примечательных особенностей.

5 СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Все расчеты и таблицы представлены в файле:

http://kodomo.fbb.msu.ru/~kamikki0/term_1/pr13.xls

БЛАГОДАРНОСТИ

Хочу поблагодарить всех преподавателей биоинформатики Факультета Биоинженерии и Бионформатики МГУ им. М. В. Ломоносова за помощь и предоставленную возможность исследовать протеом бактерии, а также своих одногруппников, помогавших мне в этом исследовании.

6 СПИСОК ЛИТЕРАТУРЫ

- [1] [NCBI, Roseateles depolymerans](#)
- [2] [Wikipedia, Roseateles depolymerans](#)
- [3] [NCBI, Roseateles depolymerans representative](#)
- [4] [Wikipedia, tmRNA](#)
- [5] [NCBI, Assembly\(Roseateles depolymerans\)](#)
- [6] [Uniprot, Protein existence](#)