

# Краткий обзор протеома бактерии *Corynebacterium pseudotuberculosis* 31

Солонович Вера

Факультет Биоинженерии & Биоинформатики МГУ им М.В.Ломоносова

## РЕЗЮМЕ

Данная работа посвящена исследованию протеома бактерии *Corynebacterium pseudotuberculosis* 31, направленному на установление распределения белков по их длине, расположению их на прямой и комплементарной цепях ДНК, установление точного числа генов белков и РНК.

## 1 ВВЕДЕНИЕ

Научная классификация организма:

Bacteria
Terrabacteria group
Actinobacteria
Actinobacteria
Corynebacteriales
Corynebacteriaceae
Corynebacterium
<i>Corynebacterium pseudotuberculosis</i>

Таблица 1. Научная классификация организма согласно Таксоному browser базы данных NCBI.[1]

### Общая информация:

*Corynebacterium* это род грам-положительных, факультативно аэробных палочкообразных бактерий, значительная часть которых является возбудителями опасных заболеваний.[1] Большинство из них устойчиво к высыханию, представители данного рода встречаются в норме в толстой кишке человека и потому оптимальными условиями для выживания тех является температура около 37 градусов и основной pH. [6,7]

Вид *C. pseudotuberculosis* впервые был описан Э.Нокаром в 1888 году. [1] Данный факультативный внутриклеточный анаэроб, не способен к спорообразованию и активному движению. Морфологически это полиморфные неподвижные неспорообразующие палочки, овоиды, реже кокки.[3] Небольшие белые колонии бактерий создают вокруг себя зону активного гемолиза.



Рис. 1, Рис. 2 Морфология колоний. 48-ми часовое культивирование в аэробных условиях с 5% содержанием CO<sub>2</sub>.

### Биохимия штамма:

Пептидогликан клеточной стенки в основном состоит из мезо-диаминопимелойкислоты (meso-DAP). Арабиноза и галактоза - важнейшие сахара клеточной стенки. Так же в ней представлены короткоцепочечные миколовые кислоты (22-36 атомов углерода). [1] Биохимия изолированных клеток существенно отличается от наблюдаемых в естественной среде колоний, в основном страдает бродильная способность. Штаммы продуцируют кислоты, но не газ, используя как источники углерода глюкозу, фруктозу, мальтозу, маннозу и сахарозу. [4]

*Corynebacterium pseudotuberculosis* является бета-гемолитической бактерией, штаммы, выделенные из обнаруженных жвачных животных, как правило, не занимаются редукцией нитрата. [1]

### Связь с заболеванием:

Животный патоген *C. pseudotuberculosis* является возбудителем казеозного лимфаденита, известного так же под названием псевдотуберкулёза, хронического кожного заболевания крупного рогатого скота, характеризующегося образованием гнойно-некротических очагов в лимфатических узлах, лёгких, печени и других органах и тканях. [5] Известны так же случаи заражения человека от домашнего скота, последние исследования клинической картины показали, что инфекция с лёгкостью распространяется на людей, и подобные заболевания среди людей не редки, хотя часто инфекция носит затяжной или рецидивный характер. Лечение требует хирургического вмешательства, лекарственная терапия антибиотиками не может являться основной мерой борьбы с патогеном. [11] Требуется активные меры, инфицированные животные должны быть подвергнуты лечению в течении нескольких недель.

Вспышки заболевания имеют цикличность в 5-10 лет, патоген способен сохраняться в почве в течении года. [4]



**Рис 3.** Вторичный абсцесс на правом бедре Арабского сернобыка (*Oryx leucoryx*), полученный в результате развития инфекции, при заражении *C. pseudotuberculosis*

### Краткая характеристика генома бактерии:

Впервые был секвенирован в 2012 году и насчитывает в себе 2.33 Мбр. Богат G-C связями (52,2% подобных связей на весь геном), что, в целом, характерно для геномов организмов из рода *Corynebacterium*. [1],[2]

## 2 МАТЕРИАЛЫ И МЕТОДЫ

Данные о геноме бактерии взяты из предоставленных материалов базы NCBI.

Анализ данных был произведён с помощью программы Microsoft Office Excel 2013, для построения диаграмм распределения и использования других методов обработки информации.

В данном случае для получения релевантной картины использовались следующие функции:

- **СЧЁТЕСЛИ** – для подсчёта общего количества генов белков и РНК, представленных в геноме
- **СЧЁТЕСЛИМН** – для ранжирования белков протеома по размеру в целях адекватного представления разброса размеров тех.
- **БИНОМРАСП** - для проверки гипотезы случайности распределения генов в последовательности.

Для анализа размеров белков бактерии были выбраны стандартные карманы длин размером в пятьдесят пар нуклеотидов каждый, с использованием данных о максимальной и минимальной возможных длинах белка, полученных с использованием функции «МАКС» и «МИН», а так же с опорой на значения, полученные при использовании к пулу длин белков функций «МЕДИАНА» и «СРЗНАЧ», была построена гистограмма длин белков протеома *Corynebacterium pseudotuberculosis* 31.

При оценке случайности распределения геномов было использовано биномиальное распределение, само по себе иллюстрирующее количество «попаданий» в серии независимых случайных экспериментов, таких, с заданной нашим экспериментом вероятностью успеха 0,5, то есть фактически воспроизводит эксперимент со случайными бросками монетки.

### 3 РЕЗУЛЬТАТЫ

#### 3.1 Количество групп генов бактерии и распределение их по цепям

В результате анализа данных *Corynebacterium pseudotuberculosis* 31 можно утверждать, что суммарно в геноме бактерии закодировано 2173 генов белков, тРНК и рРНК, с учётом 52 псевдогенов. Длина всего генома бактерии составляет 2,33 миллиона пар нуклеотидов, таким образом, на 1 миллион пар нуклеотидов приходится примерно 932 гена. Результаты детального анализа представлены в **Таблице 2**.

Категория:	Количество генов	Количество в расчёте на 1 Mbp:
CDS	2173	932,6
tRNA	47	20,2
rRNA	12	5,2

**Таблица 2.** Оценка числа групп генов бактерии *Corynebacterium pseudotuberculosis* 31 в расчёте на один миллион пар оснований.

Распределения генов по прямой и комплементарной цепям ДНК иллюстрирует нижеследующая **Таблица 3**.

Цепь:	CDS:	РНК:	рРНК:	тРНК:
+	1080	34	6	28
-	1093	25	6	19

**Таблица 3.** Распределение генов по различным цепям ДНК в геноме бактерии *Corynebacterium pseudotuberculosis* 31.

##### 3.1.1. Интерпретация данных.

Опираясь на данные, можно с лёгкостью заключить, что гены, кодирующие белки, преобладают на **комплементарной** или, так называемой, "-"-цепи. Разница незначительна и составляет 13 генов – это ~0,6% от общего количества генов в геноме.

Гены, кодирующие РНК, преобладают на **прямой**, или "+"-цепи. Разница составляет 0,4% (9 генов).

В частности, генов рибосомальных РНК на обеих цепях одинаковое количество, значит цепи различаются только по количеству генов транспортных РНК.

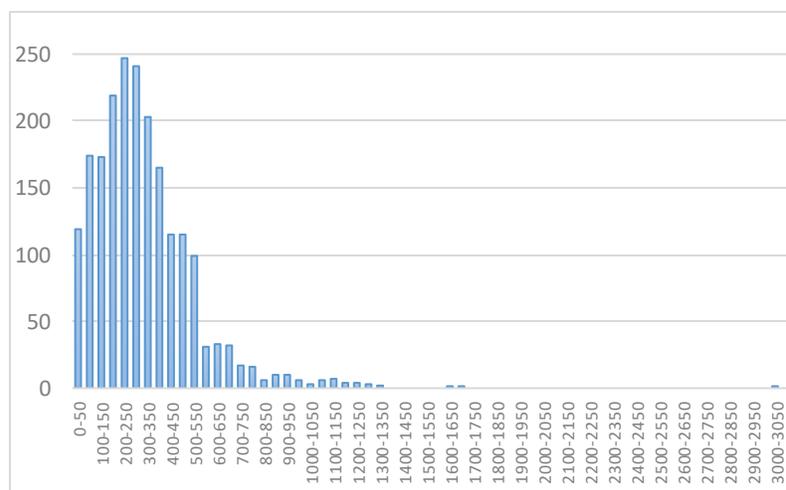
#### 3.2 Распределение длин белков

В геноме бактерии *Corynebacterium pseudotuberculosis* 31 закодировано 2173 белка. Некоторые данные о белках представлены в **Таблице 4**.

Характеристика:	Значение
Максимальная возможная длина белка:	3032
Минимальная возможная длина белка:	21
Медиана длин белков:	271
Среднее значение длин белков:	~309

**Таблица 4.** Статистические данные белков бактерии *Corynebacterium pseudotuberculosis* 31.

На основе полученных статистических данных были выбраны карманы длин шагом в 50 аминокислотных. Распределение белков *Corynebacterium pseudotuberculosis* 31 по длине представлено на **Рис.4** и **Рис.5**, гистограмма представлена в приближении для наглядности.



**Рис 4.** Распределение длин белков бактерии *Corynebacterium pseudotuberculosis* 31, полный масштаб.

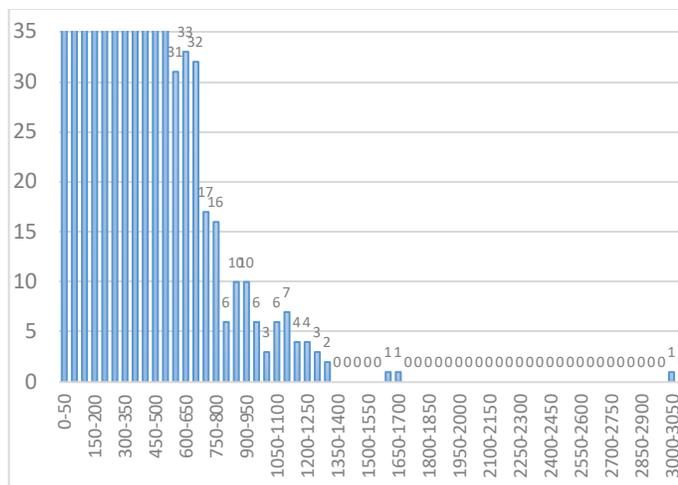


Рис 5. Распределение длин белков бактерии *Corynebacterium pseudotuberculosis* 31, в приближении.

### 3.3 Случайность распределения генов по цепям ДНК

Для проверки данной гипотезы была использована функция БИНОМ.РАСП, с интегральным аргументом «ИСТИНА». В результате была получена вероятность того, что число успешных испытаний будет не меньше значения аргумента "число\_успехов". В данном случае «успешными испытаниями» мы можем считать попадание на одну из цепей. В Таблице 5 представлены полученные результаты.

Биномиальное распределение:	Значение:
Для общего числа генов:	0,474685095
Для генов, кодирующих РНК:	0,903736771
Для генов, кодирующих белки:	0,398428762

Таблица 5. Оценка «случайности» распределения генов бактерии *Corynebacterium pseudotuberculosis* 31 по цепям ДНК

Если  $x$  = число\_успехов, в нашем случае – число попаданий на определённую цепь генов из всего генома *Corynebacterium pseudotuberculosis* 31,  $n$  = число\_испытаний, то есть общее число генов, и  $p$  = вероятность\_успеха, определённая в нашем случае, как 0.5, то интегральное биномиальное распределение выглядит следующим образом:

$$(1) \quad B(x; n, p) = \sum_{y=0}^x b(y; n, p)$$

Таким образом, разумным кажется вывод, что распределение генов по цепям является случайным, так как полученные вероятности попадания уже имеющегося на цепи в действительности или даже большего числа генов в общем случае составляет 47%. То есть практически в половине случаев мы бы наблюдали похожую картинку распределения генов по цепи.

### 3.4 Перекрывания генов по цепям

Цепь:	Количество перекрытий:
+	104
-	97
Суммарно:	201

Таблица 6 Перекрывания генов бактерии *Corynebacterium pseudotuberculosis* 31 по цепям ДНК

Из представленной таблицы следует, что микроорганизм *Corynebacterium pseudotuberculosis* 31 использует несколько рамок считывания (достоверно больше одной) при считывании генома. В данной таблице гены, кодирующие белки и гены, кодирующие РНК, не обособлены друг от друга по понятным причинам – их локализация может не коррелировать с функцией, а значит гены могут перекрываться независимо от своей природы.

### 3.5 Количество квазиперонов

Квазиперон – последовательность фиксированного размера, содержащая максимальное количество генов с установленными промежутками между теми. Очевидно, в зависимости от интервала между смысловыми участками и будет зависеть их количество. Для того, чтобы найти зависимость между количеством квазиперонов и размером интервала, были выбраны три значения данного интервала с шагом в 100 смысловых единиц (в данном случае – пар оснований): 50, 150 и 250.

Интервал:	50	150	250
+	213	243	237
-	202	251	241
Суммарно:	415	494	478

Таблица 6 Перекрывания генов бактерии *Corynebacterium pseudotuberculosis* 31 по цепям ДНК

## ОБСУЖДЕНИЕ И ЗАКЛЮЧЕНИЕ

Как можно наблюдать из представленных результатов, гены бактерии *C. pseudotuberculosis* 31 распределены по цепям случайной – их количество примерно совпадает, к тому же, научный метод проверки гипотезы случайного распределения генов по цепям аналитическими методами Excel показал, что наблюдаемая нами картина достаточно вероятна. К сожалению, из-за наличия псевдогенов в геноме, то есть генов, не имеющих конечного продукта, общее число представленных CDS последовательностей не соответствует числу конечных белковых продуктов. К счастью, количество таких псевдогенов сравнительно невелико.

**Рис.4** иллюстрирует преобладание в протеоме белков длиной от 200 до 250 аминокислотных остатков. После достижения пика значения идут на спад, белки большей длины не достигают таких впечатляющих значений. Абсолютное большинство белков протеома *Corynebacterium pseudotuberculosis* 31 имеют сравнительно небольшую длину, об это сразу же говорит смещённый пик в общем контуре гистограммы. Белков длиной более 1350 аминокислотных остатков почти не наблюдается.

Отдельные молекулы имеют, впрочем, значительную длину, но в каждом таком не пустом диапазоне количество «молекул-гигантов» не превышает одной. Функции этих белков исключений тоже разнообразны. Например, рекордсмен, имеющий 3032 аминокислотных остатков это синтаза жирных кислот, следующие достоверно изученные белки, это АТФ-зависимая хеликаза, синтаза птиоцерола и ДНК-зависимая РНК полимеразы, встречаются среди белков большой длины и «гипотетические» протеины (*Hypothetical protein*), чьё существование было предсказано, но в связи с недостатком экспериментальных свидетельств их наличия, существование оных ещё находится под вопросом.

Хотя в большинстве своём прокариотические организмы имеют довольно плотно посаженные гены в следствии отсутствия интронов в их ДНК, количество квазиперонов в геноме *Corynebacterium pseudotuberculosis* 31 сравнительно невелико и не коррелирует с размерами интервалов. Впрочем, возможно необходимо более детальное исследование для того, чтобы проследить закономерность, связывающую длину интервала и количе-

ство квазиперонов на цепи. В рамках данного исследования у нас не было достаточных средств для обнаружения оной.

## СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

[http://kodomom.fbb.msu.ru/~sol-era/term1/term1.html//GCA\\_feature\\_table\(1\).xlsx](http://kodomom.fbb.msu.ru/~sol-era/term1/term1.html//GCA_feature_table(1).xlsx)

## ВИЗУАЛЬНЫЕ МАТЕРИАЛЫ

1. <http://www.microbiologyinpictures.com/bacteria%20photos/corynebacterium%20pseudotuberculosis%20photos/corynebacterium%20pseudotuberculosis%2006.html>
2. <http://www.microbiologyinpictures.com/bacteria%20photos/corynebacterium%20pseudotuberculosis%20photos/corynebacterium%20pseudotuberculosis%2008.html>
3. <http://veterinaryrecord.bmj.com/content/162/26/862/F1.expansion.html>

## ЛИТЕРАТУРА

- I. Dorella, Fernanda Alves, et al. "Corynebacterium pseudotuberculosis: microbiology, biochemical properties, pathogenesis and molecular studies of virulence." *Veterinary research* 37.2 (2006): 201-218.
- II. Peel, Margaret M., et al. "Human lymphadenitis due to *Corynebacterium pseudotuberculosis*: report of ten cases from Australia and review." *Clinical Infectious Diseases* 24.2 (1997): 185-191.

## ДРУГИЕ ИСТОЧНИКИ

1. [https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=1087451&lvl=3&keep=1&srch-mode=1&unlock&mod=1&log\\_op=modifier\\_toggle#modif](https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=1087451&lvl=3&keep=1&srch-mode=1&unlock&mod=1&log_op=modifier_toggle#modif)
2. <https://www.ncbi.nlm.nih.gov/genome/genome-groups/2411>
3. <https://en.wikipedia.org/wiki/Corynebacterium>
4. <http://lavavet.com/corynebacterium-pseudotuberculosis/>
5. [https://web.mst.edu/~micro-bio/BIO221\\_2010/C\\_pseudotuberculosis.html](https://web.mst.edu/~micro-bio/BIO221_2010/C_pseudotuberculosis.html)
6. <http://zooresurs.ru/mrs/mrs-zb/42-pseudotuberculosis.html>
7. <http://www.gastroscan.ru/handbook/118/4456>