

# Исследование бактерии *Roseateles depolymerans*.

3. Раздобарин<sup>1</sup>

1) Факультет биоинженерии и биоинформатики МГУ имени М. В. Ломоносова

## Аннотация

Было проведено изучение и аналитическое исследование генома протеобактерии *Roseateles depolymerans*, в частности рассмотрены статистические данные о белках, псевдогенах и РНК бактерии. Для создания статистики использовались готовый расшифрованный геном бактерии и таблицы Excel. Результатом стали наглядные аналитические и статистические данные.

**Ключевые слова:** Бактерия, Протеобактерия, *Roseateles depolymerans*.

## 1. Введение

ТАБЛИЦА 1. КЛАССИФИКАЦИЯ.

Царство	Bacteria
Тип	Proteobacteria
Класс	Beta Proteobacteria
Порядок	Burkholderiales
Семейство	Comamonadaceae
Род	<i>Roseateles</i>
Вид	<i>Roseateles depolymerans</i>

*Roseateles depolymerans* — бета-протеобактерия открытая группой учёных из Национального института биологических наук и технологии человека, Ибараки, Япония, в 1999 году [1]. Согласно статье, образец был обнаружен в речной воде. Первые исследования показали, что бактерия способна расщеплять полигексаметилен карбонат и другие биоразлагаемые пластмассы. Также в ней содержатся каротиноиды, преимущественно спириллоксантин. Анализ последовательности генов показал, что бактерия является близким родственником родов *Rubrivivax*, *Ideonella* и *Leptothrix*, однако точного соответствия найдено не было. Согласно генетическим и фенотипическим признакам штамм принадлежит новому роду — *Roseateles*.

Задачей данной работы было изучение генома бактерии, предоставленного в геномной базе данных NCBI [2]. Для этого была проведена работа по изучению генома бактерии, получению понятных результатов и их оформлению.

## 2. Материалы и методы

Основным материалом работы послужил расшифрованный геном *Roseateles depolymerans* [3]. Его исследование проводилось при помощи Excel. В первую очередь был разобран состав генома. Для этого был проведён поиск по определённым маркерам в геноме, с помощью команды СЧЁТЕСЛИМН. «CDS – with\_protein» для кодирующих последовательностей, «gene – pseudogene» для псевдогенов и «ncRNA», «rRNA», «tmRNA», «tRNA» для РНК. Далее было рассмотрено расположение этих генов на цепях ДНК. Его можно определить по соответствующей отметке в графе «strand». Последнее и самое крупное исследование коснулось белков бактерии. Важно было посмотреть на длины этих белков — графа «product\_length», и отделить существующие белки от гипотетических. Соответствующие отметки можно найти в графе «name». Если белок гипотетический там будет стоять не название белка, а «hypothetical protein».

## 3. Результаты

### 3.1 Состав генома

Как уже было сказано, в первую очередь был определён состав генома для дальнейшего изучения и составлена таблица 2.

ТАБЛИЦА 2. СОСТАВ ГЕНОМА.

	class	protein_coding	pseudogene	RNase_P_RNA	rRNA	tmRNA	tRNA	with_protein	
# feature	1	0	0	0	0	0	0	0	0
CDS	0	0	0	0	0	0	0	4773	0
gene	0	4773	34	1	12	1	57	0	0
ncRNA	0	0	0	1	0	0	0	0	0
rRNA	0	0	0	0	0	0	0	0	12
tmRNA	0	0	0	0	0	0	0	0	1
tRNA	0	0	0	0	0	0	0	0	57

Из таблицы 2 видно, что геном бактерии содержит 4773 кодирующих последовательностей (белков), 34 псевдогена и 71 РНК различных типов. «ncRNA» — некодирующая РНК, «rRNA» — рибосомальная РНК, «tmRNA» — транспортно-матричная РНК и «tRNA» — транспортная РНК.

ТАБЛИЦА 3. РАСПРЕДЕЛЕНИЕ ГЕНОВ ПО ЦЕПЯМ ДНК.

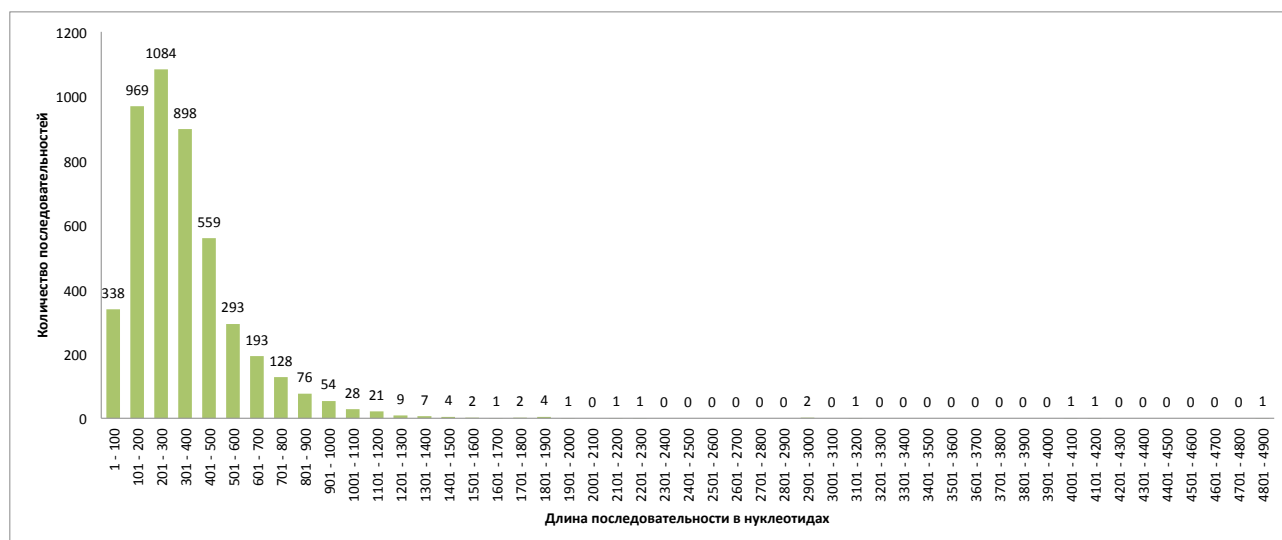
	белки	псевдогены	РНК
прямая	2299	19	41
комплементарная	2474	15	30

Также было интересно взглянуть на распределение генов по цепям ДНК. Из таблицы 3 видно, что отклонения от среднего довольно высоки.

### 3.2 Белки

Поскольку гены, кодирующие белки, составляют более 97% всех генов, они стали вторым этапом изучения.

ГИСТОГРАММА 1. СООТНОШЕНИЕ ДЛИН БЕЛКОВ



Из гистограммы 1 видно, что подавляющее большинство белков имеют длину от 201 до 300 нуклеотидов. Однако встречаются и белки длиной 29 нуклеотидов и длиной 4828. Это минимум и максимум соответственно. Остальные статистические данные видно из таблицы 4.

ТАБЛИЦА 4. СТАТИСТИКА ДЛИН БЕЛКОВ.

мин	макс	ср. длина	стд. откл.	медиана
29	4828	347,05	258,58	296

ТАБЛИЦА 5. ТИПЫ БЕЛКОВ.

тип белка	количество	кол-во на 1 млн п.н.
Рибосомальный	63	11.09
Транспортный	261	45.94
Гипотетический	1319	232.15
Остальные	3130	550.89

В таблице 5 представлено количество разных типов белков. Выделены рибосомальные и транспортные,

как выполняющие важнейшие функции, а также гипотетические белки. Последние названы так, поскольку не определены. При расшифровке генома были найдены старт-кодон и стоп-кодон. Логично предположить, что этот участок может что-то кодировать. Принято говорить, что такие участки кодируют гипотетический белок.

## 4. Обсуждение и заключение

### 4.1 Состав генома

Было упомянуто, что отклонение от среднего при распределении генов по цепям ДНК довольно высоки. Можно предположить, что такое распределение не случайно и несёт какую-то функцию. Определение этой функции требует большего исследования, чем простой анализ. Помимо этого, стоит определить функцию некодирующей РНК (ncRNA).

### 4.2 Белки

Самым длинным белком бактерии является «Hem-agglutinin-related protein». Определение функции этого белка в бактерии так же требует более детального изучения. Самым коротким же белком является гипотетический белок. Вполне возможно, что этот участок не несёт за собой никакой функции.

Более четверти белков — гипотетические. Такое количество гипотетических белков говорит о неизученности генома. Стоит понять, действительно ли эти участки что-то кодируют, и если да определить белок и его функцию.

## Сопроводительные материалы

[http://kodomo.fbb.msu.ru/~napster/term1/excel/supplementary\\_materials.xlsx](http://kodomo.fbb.msu.ru/~napster/term1/excel/supplementary_materials.xlsx)

В данном файле представлены все таблицы и гистограммы из статьи, а также оригинальный геном с NCBI [3], в формате Excel.

## Благодарности

Прежде всего я благодарен своей маме Раздобариной В.В. за помощь в оформлении и вёрстке статьи. Отдельная благодарность моим сокурсникам, которые всячески помогали и вдохновляли меня при написании статьи.

## Список литературы и ссылок

- [1] Статья группы учёных, обнаруживших бактерию [https://www.ncbi.nlm.nih.gov/pubmed?cmd=Retrieve&dopt=Abstract&list\\_uids=10319464](https://www.ncbi.nlm.nih.gov/pubmed?cmd=Retrieve&dopt=Abstract&list_uids=10319464)
- [2] Страница сайта NCBI, содержащая информацию о бактерии <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=76731&lvl=3&lin=f&keep=1&srchmode=1&unlock>
- [3] Ссылка на скачивание архива с файлом генома бактерии в текстовом формате [ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001/483/865/GCA\\_001483865.1\\_ASM148386v1/GCA\\_001483865.1\\_ASM148386v1\\_feature\\_table.txt.gz](ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001/483/865/GCA_001483865.1_ASM148386v1/GCA_001483865.1_ASM148386v1_feature_table.txt.gz)