

# Мини-обзор генома

## *Sphingobium japonicum* UT26S

Горбонос Александра

*Факультет Биоинженерии и биоинформатики, Московский Государственный Университет им. Ломоносова, Ленинские горы 1-73, Москва, Россия  
alexgorbonos@gmail.com*

### **Резюме:**

В ходе работы была проанализирована бактерия *Sphingobium japonicum*. С помощью открытых источников были получены необработанные данные о геноме бактерии. Исходные данные были проанализированы с помощью ЭВМ и компьютерной программы Microsoft Excel 2016.

**Ключевые слова:** *Sphingobium japonicum*, Excel, Гены

### **Введение:**

*Sphingobium japonicum* – желто-пигментированная, неподвижная, анаэробная грамотрицательная бактерия. В качестве единственного источника углерода и энергии она использует гамма-гексахлороциклогексан, созданный человеком хлор-содержащий пестицид, который оказывает негативное влияние на окружающую среду, из-за его токсичности и высокой устойчивости. В этой работе было исследовано расположение генов белков бактерии.

### **Материалы и методы:**

Файл с информацией о геноме был взят с сайта NCBI:

[https://www.ncbi.nlm.nih.gov/genome/2807?genome\\_assembly\\_id=172656](https://www.ncbi.nlm.nih.gov/genome/2807?genome_assembly_id=172656)

Файл с обработанными данными можно найти на моем сайте:

<https://kodomofbb.msu.ru/~alexgorbonos/term1/excel.html>

Данные были обработаны с помощью Microsoft Windows 10 October Update, Google Chrome, программы “Блокнот” и Microsoft Excel 2016.

Для сбора информации были построена гистограмма - количество белков от длины белка, с шагом в 200 аминокислот. Таблица с различными свойствами данного распределения. Таблица распределения генов по цепям ДНК.

## Результаты:

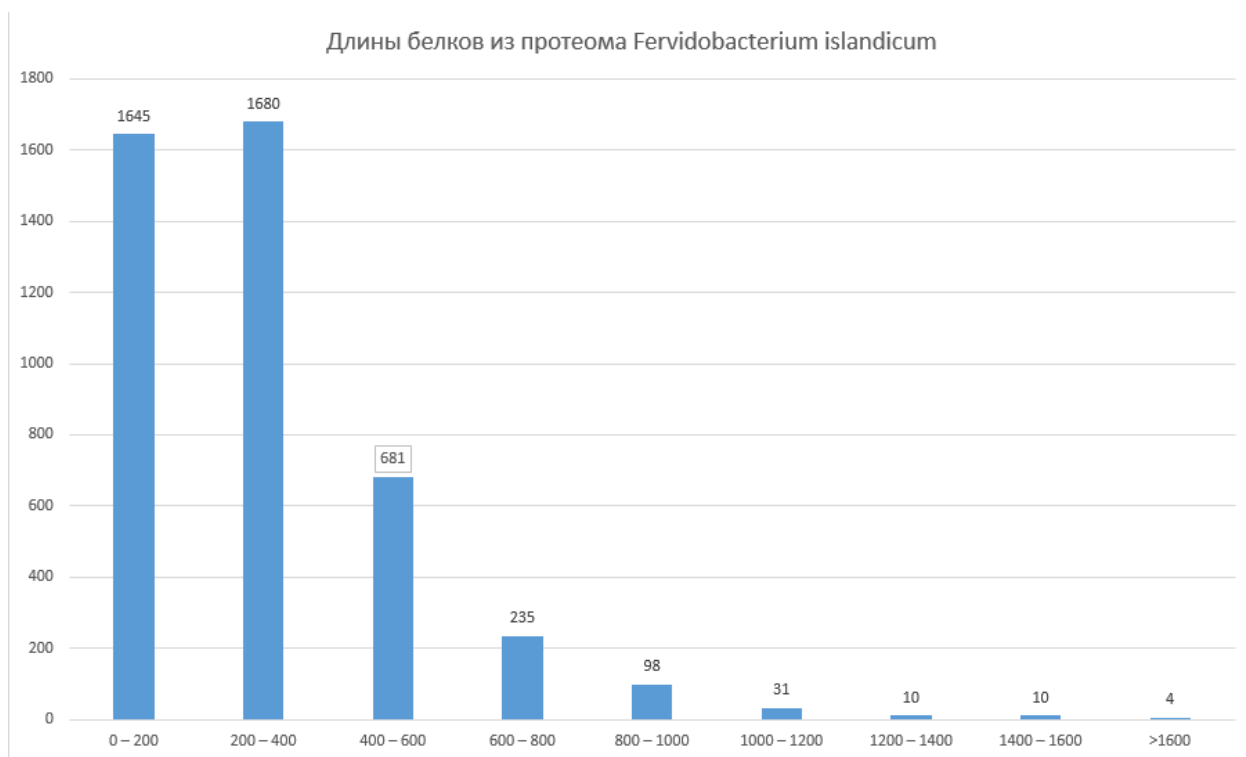


Рис. 1. Распределение длин белков. По вертикали – количество белков, по горизонтали их длина.

Средняя длина	Стандартное отклонение	Медиана	Минимальная длина	Максимальная длина
302,45	228,51	253	26	4159

Табл. 1. Свойства распределения длин белков.

strand	protein	pseudoge	RNA
+(main)	2093	0	37
-(complementary)	2301	0	30

Табл. 2. Распределение генов по цепям ДНК.

Больше всего белков длины 200-400 аминокислот, немного меньше длины 0-200. Таким образом “коротких” белков много больше чем всех остальных. (Здесь короткими называются длины сильно меньше среднего арифметического максимальной и минимальной длин). Что отражается значением медианы и средней длины.

Распределение длин белков не выглядит как нормальное гауссовское распределение. В дальнейшем, используя python и различные метрики, можно было бы понять, что это за распределение.

На комплементарной цепи ДНК-генов больше чем на прямой, однако РНК-генов больше на прямой, чем на обратной. Скорее всего они распределяются случайно, но возможно и

нет, ответа мне найти не удалось. Так же видно, что у бактерии отсутствуют псевдо-гены, что может означать, малую вероятность того, *Sphingobium japonicum* находится с кем-то в симбиотической связи.

Количество по полю assembly	Названия столбцов	gene	ncRNA	rRNA	tmRNA	tRNA	Общий итог
protein_coding	CDS	4394					4394
RNase_P_RNA		1	1				2
rRNA		9					9
SRP_RNA		1	1				2
tmRNA		1					1
tRNA		55					55
with_protein		4394					4394
(пусто)				9	1	55	65
<b>Общий итог</b>		<b>4394</b>	<b>4461</b>	<b>2</b>	<b>9</b>	<b>1</b>	<b>8922</b>

Табл. 3 Число генов белков по категориям.

## Заключение:

Данная работа показала некоторые факты о количестве различных белков, а так же их распределении. Для более детального исследования бактерии, следует посмотреть на взаимодействие генов друг на друга и другие не тронутые ранее параметры.

Анализировать эту информацию в excel проблематично, поэтому в дальнейшем нам придется прибегнуть к помощи различных языков программирования, а так же знаний статистики.

## Сопроводительные материалы

<https://kodomofbb.msu.ru/~alexgorbonos/term1/excel.html>

## Список литературы

<https://www.uniprot.org/proteomes/UP000007753>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2953701/>

[https://www.ncbi.nlm.nih.gov/genome/2807?genome\\_assembly\\_id=172656](https://www.ncbi.nlm.nih.gov/genome/2807?genome_assembly_id=172656)

[https://en.wikipedia.org/wiki/Pseudogene#Bacterial\\_pseudogenes](https://en.wikipedia.org/wiki/Pseudogene#Bacterial_pseudogenes)