

## Мини-обзор

# Мини-обзор протеома и генома бактерии *Pedobacter saltans* DSM 12145

Котюргин Александр<sup>1</sup>

<sup>1</sup>Факультет биоинженерии и биоинформатики МГУ 119234, г. Москва, ГСП-1, Ленинские горы МГУ 1, стр. 73

### АБСТРАКТ

Данные с сервера NCBI о бактерии *Pedobacter saltans* DSM 12145 были использованы для получения информации о распределении длин белков бактерии, а также для анализа распределения генов по прямой и обратной цепям, подсчета количества квазиоперонов и пересечений генов.

## 1 ВВЕДЕНИЕ

*Pedobacter saltans* DSM 12145 (эквивалентное название — *Pseudopedobacter saltans*) — граммотрицательная почвенная бактерия из семейства *Sphingobacteriaceae*. Является продуцентом гепариназы. Выделен из почвы в Исландии в 1990 году. Размер генома составляет 4635240 п.н..

В данной работе были построены гистограммы длин белков с шагом кармана 50, а также с шагом в 1/20 от разницы между длинами самого короткого и самого длинного белками. Была произведена проверка гипотезы, что гены распределены по прямой и обратной цепям с вероятностью 0,5, а также произведен подсчет числа квазиоперонов, пересечений генов и генов с длиной, не кратной трем.

## 2 МАТЕРИАЛЫ И МЕТОДЫ

В работе использовалась программа OpenOffice Calc. Обработка данных велась функциями BINOMDIST(), COUNTIF(), COUNTIFS(). Данные по геному бактерии были взяты с FTP-сервера NCBI.

## 3 РЕЗУЛЬТАТЫ

Были составлены две гистограммы длин белков по 20 карманов — одна с «естественным» делением, где за шаг гистограммы принята 1/20 разницы длины самого короткого и самого длинного белка, каковые составили 30 и 6409 аминокислотных остатков соответственно.

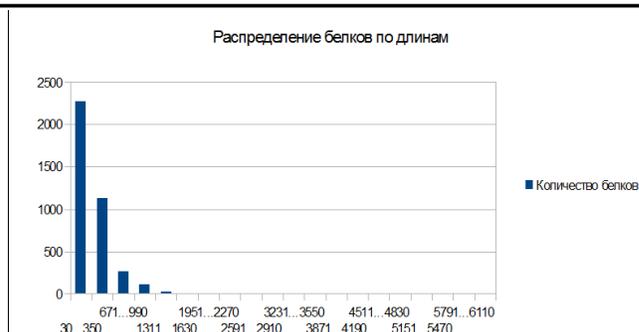


Рис. 1. Гистограмма длин белков с «естественным» шагом гистограммы.

В связи со слабой информативностью такой гистограммы, была также создана гистограмма с шагом 50 аминокислотных остатков:



Рис. 2. Гистограмма длин белков с шагом гистограммы 50.

Был произведен подсчет числа генов на прямой и обратной цепях ДНК для белок- и РНК-кодирующих генов (CDS и RNA соответственно)

**Таблица 1.** Распределение генов по цепям ДНК

Цепь	Тип гена		Всего по цепям
	CDS	RNA	
Прямая	1930	23	1953
Обратная	1862	44	1906
Всего по типам	3792	67	
		Всего генов	3859

После чего была проведена проверка гипотезы, что гены распределены по цепям с вероятностью 0,5 — с помощью функции BINOMDIST() было вычислено р-значение для нулевой гипотезы распределения с вероятностью 0,5:

**Таблица 2.** Р-значение для различных групп генов при нулевой гипотезе распределения с вероятностью 0.5

CDS	0,2766
RNA	0,0139
По всей совокупности генов	0,4590

Также был произведен подсчет числа квазиоперонов (т.е., групп генов, расстояние между которыми больше 100) и пересечений генов в совокупности CDS-генов:

**Таблица 3.** Подсчет числа особенностей генома

Число квазиоперонов	1437
Число пересечений генов	422

Генов, чьи длины не кратны трем, обнаружено не было.

## 4 ОБСУЖДЕНИЕ

Было получено, исходя из Рис.1., что подавляющее большинство белков имеет длину меньше 1000 и лишь единичные случаи имеют длину больше 1500 аминокислотных остатков. На более детальной гистограмме (Рис.2.) видно, что больше всего белков имеют длину от 101 до 450 аминокислотных остатков, а также то, что белков длиной от 1 до 50 резко меньше, чем прочих.

Исходя из р-значений, можно достоверно сказать, что CDS гены и вся совокупность генов распределена по прямой и обратной цепям случайно с вероятностью 0,5. Про RNA мож-

но предположить, что здесь влияние оказал малый размер выборки, который привел к тому, что р-значение меньше 0,05. Среди CDS-генов не было обнаружено генов с длинами, не кратным трем.

## БЛАГОДАРНОСТИ

Работа выполнялась на Факультете Биоинженерии и Биоинформатики МГУ.

## СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Ссылка на файл со всеми использованными в работе данными: [Kotyurgin\\_pr14.xls](#)

## СПИСОК ЛИТЕРАТУРЫ

<http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=762903>  
*Pedobacter saltans* DSM 12145, complete genome  
 Genome Assembly and Annotation report *Pseudopedobacter saltans* DSM 12145