

---

# Обзор протеома бактерии *Desulfovibrio vulgaris*, штамм 'Miyazaki F'

Маргасюк Сергей<sup>1</sup>

<sup>1</sup>МГУ, факультет биоинженерии и биоинформатики, 1 курс

---

## АННОТАЦИЯ

Данная работа посвящена исследованию протеома бактерии *Desulfovibrio vulgaris*, штамм 'Miyazaki F'. Рассмотрено распределение белков по длине и расположение генов на комплементарных цепях ДНК.

## 1. ВВЕДЕНИЕ

*Desulfovibrio vulgaris* – сульфатредуцирующая бактерия, принадлежащая к отряду Proteobacteria. [1]

Геном данного вида полностью секвенирован, он представлен единственной хромосомой. При этом среди других бактерий рода *Desulfovibrio* насчитывается 9 различных плазмид, среди других штаммов бактерии *Desulfovibrio vulgaris* – 3.[2] Всего в геноме 3260 генов, из них 3180 кодируют белки, 80 – РНК.

Цель работы – исследование протеома бактерии.

## 2. МЕТОДЫ

Данные о протеоме бактерии были взяты с сайта NCBI.[3] Для статистического анализа данных была использована программа LibreOffice Calc из пакета LibreOffice версии 5.0.4.2.

К данным из mt и pt файлов, содержащих информацию о белок-кодирующих и РНК-кодирующих генах, после импорта (на листы NC\_011769\_protein и NC\_011769\_rna) был добавлен столбец Gene type, показывающий, что именно кодирует данный ген: белок или РНК. С помощью сортировки список всех генов был разбит на два (в зависимости от расположения гена на прямой или обратной цепи). К полученным спискам были добавлены столбцы, содержащие координаты начала и конца генов, расстояние до предыдущего гена (на основании этих данных можно судить о конфигурации квазиоперонов и пересечений генов).

Далее были добавлены столбцы, содержащие информацию о номере гена в квазиопероне (определение квазиоперона см. далее): для первого гена этот номер полагался равным 1, для остальных с помощью функции IF рассматривалось два случая: если расстояние до предыдущего гена меньше константы L, то номер равен номеру предыдущего гена, увеличенному на 1, иначе номер гена в квазиопероне равен 1.

Для подсчета количества элементов, удовлетворяющих заданным условиям (в частности, при построении гистограммы, таблицы, описывающей расположение генов на прямой и обратных цепях и листа с дополнительными данными info) использовалась функция COUNTIFS.

Для проверки гипотезы о случайном распределении генов между прямой и обратной цепями с вероятностью 0.5 использовалась функция BINOM.DIST с параметром C=1 (подсчет вероятности данного или меньшего числа успехов в условиях гипотезы о случайном независимом распределении).

## 3. РЕЗУЛЬТАТЫ

### 3.1. Анализ распределения белков по длине

С помощью программы LibreOffice Calc построена гистограмма, отражающая распределение белков бактерии по длине. Для построения гистограммы взяты белки, кодированные генами, расположенными на единственной хромосоме. В качестве “карманов” гистограммы выбраны десять отрезков равной длины, покрывающие область от

минимальной длины белка (29 а.о.) до максимальной (2537 а.о.).

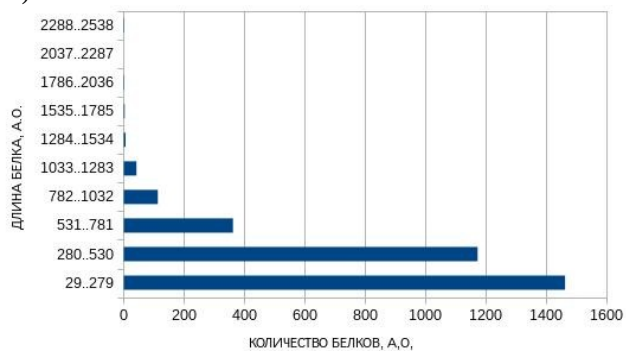


Рисунок 1. Распределение белков бактерии *Desulfovibrio vulgaris*, штамм 'Miyazaki F' по длине.

### 3.2. Анализ распределения генов по комплементарным цепям ДНК

С помощью программы LibreOffice Calc получены данные для таблицы, отражающей распределение генов, кодирующих белки и РНК, по комплементарным цепям.

цепь	белки	РНК
прямая	1598	39
обратная	1582	41
всего	3180	80

Таблица 1. Распределение генов бактерии *Desulfovibrio vulgaris*, штамм 'Miyazaki F' по комплементарным цепям.

Из приведенных данных видно, что гены распределены по цепям примерно поровну. Гипотеза о случайном распределении генов с вероятностью 0.5 по комплементарным цепям была проверена тестом на биномиальное распределение с помощью функции BINOM.DIST LibreOffice Calc. Вероятность того, что из 3180 генов на одной из цепочек обнаружено 1623 или меньше равна 0.82, что не противоречит предположению о случайном независимом распределении генов по цепочкам при уровне доверия 0.05. При проведении теста отдельно по белок-кодирующим генам и генам РНК получены вероятности 0.79 и 0.91 соответственно, что также не противоречит гипотезе.

### 3.3. Распределение генов по квазиоперонам; пересекающиеся гены

Пусть квазиоперон – множество генов, лежащих на одной цепи, при этом расстояние между соседними не превышает 100 п.н. (при выполнении работы были рассмотрены модификации данного определения с максимальным допустимым расстоянием L, равным 50 или 200). Предполагается, что принадлежность генов к одному квазиоперону является необходимым условием для

принадлежности их к оперону. Заметим, что конфигурация квазиоперонов (в отличие от конфигурации оперонов) может быть проанализирована программой на основе имеющихся данных. Далее приведены полученные результаты:

Показатель	Значение при L=50	Значение при L=100	Значение при L=200
Число квазиоперонов	2331	2074	1652
Число квазиоперонов на прямой цепи	1145	1027	827
Число квазиоперонов на обратной цепи	1186	1047	825
Максимальная длина квазиоперона	23	32	32

**Таблица 2.** Распределение генов бактерии *Desulfovibrio vulgaris*, штамм 'Miyazaki F' по квазиоперонам.

Кроме того, были получены данные о пересечениях соседних генов. Отдельно рассмотрены два случая: пересечение соседних генов, лежащих на одной и той же цепи ДНК, и пересечение соседних генов, лежащих на разных цепях ДНК. Далее приведены данные, характеризующие пересечения соседних генов, расположенных на одной цепи ДНК:

- число пересечений: 407
- число пересечений на прямой цепи: 199
- число пересечений на обратной цепи: 208
- максимальная длина пересечения: 65
- минимальная длина пересечения: 4

Следующие данные описывают пересечения генов, расположенных на комплементарных цепях ДНК:

- число пересечений: 11
- максимальная длина пересечения: 81
- минимальная длина пересечения: 4

Для анализа информации о пересечениях были получены статистические данные о длине генов:

Показатель	Значение для прямой цепи	Значение для обратной цепи	Значение для всего генома
Максимум	7613	7436	7613
Минимум	74	74	74
Среднее	1018	1040	1029
Стандартное отклонение	710	713	712

**Таблица 3.** Статистический анализ длин генов бактерии *Desulfovibrio vulgaris*, штамм 'Miyazaki F'

#### 4. ОБСУЖДЕНИЕ

Как результат рассмотрения таблицы 1 можно отметить неравномерность распределения длин белков бактерии: количество белков во второй половине рассмотренного диапазона крайне мало (13 белков из 3166).

Данные таблицы 2 свидетельствуют о примерно равном распределении квазиоперонов по цепям (здесь тест биномиального распределения не выполнялся). Кроме того, информация свидетельствует о том, что опероны часто не совпадают с квазиоперонами: максимальная длина квазиоперона в геноме равна 32, в то время как для оперона такое количество генов слишком велико.

Аналогично прочим параметром протеома, пересечения распределены по комплементарным цепям примерно поровну. Кроме того, при соотнесении данных таблиц 2 и 3 отметим, что пересечения имеют малую длину по сравнению с большинством генов бактерии. Можно попытаться объяснить это явление следующим образом: при мутации, затрагивающей область пересечения генов, оба гена оказываются изменены, поэтому вероятность получения жизнеспособных мутантов меньше (т. е. область пересечения генов меньше изменяется в ходе эволюционного процесса); таким образом, пересечения генов невыгодны для организма (особенно в ходе движущего отбора). Отдельно обратим внимание на то, что число пересечений генов, расположенных на комплементарных цепях ДНК, много меньше, чем число пересечений генов, расположенных на одной цепи. Приведенная гипотеза никак не объясняет это явление (при появлении мутации в одной цепи ДНК эта мутация, сохранившись, у потомства приведет к изменению в обеих цепях, т. е. затронет оба гена, вне зависимости от их распределения по цепям). Возможно, это явление может быть объяснено при более глубоком понимании причин появления пересечений генов.

#### 5. ЗАКЛЮЧЕНИЕ

В заключение можно отметить, что параметры протеома, рассмотренные в данной работе (количество белков, число пересечений генов, максимум, минимум, среднее и стандартное отклонение длины генов, число квазиоперонов), при подсчете их отдельно по множествам белков, гены которых расположены на прямой и обратной цепях соответственно, оказываются приблизительно равными; это позволяет выдвинуть гипотезу о независимом от этих параметров случайном распределении белков бактерии по комплементарным цепям.

#### 6. СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Далее представлена ссылка на файлы, содержащие таблицу с расчетами, использованными в данной работе: [review.xlsx](#), [review.ods](#)

#### СПИСОК ЛИТЕРАТУРЫ

- [1] John F Heidelberg и др. (2004) The genome sequence of the anaerobic, sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. *J. Nature Biotechnology* 22, 554 – 559

- [2] NCBI, информация о плазидах:  
[ftp://ftp.ncbi.nlm.nih.gov/genomes/GENOME\\_REPORTS/plasmids.txt](ftp://ftp.ncbi.nlm.nih.gov/genomes/GENOME_REPORTS/plasmids.txt)
- [3] NCBI, полный геном *D. vulgaris*, штамм 'Miyazaki F':  
[ftp://ftp.ncbi.nlm.nih.gov/genomes/archive/old\\_refseq/Bacteria/Desulfovibrio\\_vulgaris\\_\\_Miyazaki\\_F\\_\\_uid59089](ftp://ftp.ncbi.nlm.nih.gov/genomes/archive/old_refseq/Bacteria/Desulfovibrio_vulgaris__Miyazaki_F__uid59089)