

# МИНИ-ОБЗОР ГЕНОМА БАКТЕРИИ BORRELIELLA BURGENDORFERI (BORRELIA BURGENDORFERI)

Зубарева В.М.

*Факультет биоинженерии и биоинформатики, Московский государственный университет имени М.В.Ломоносова*

*Ленинские горы МГУ 1 стр. 73, г. Москва, 119234, Российская федерация*

*zubareva.valeria@kodomo.fbb.msu.ru*

*Ключевые слова: Borreliella burgdorferi B31, анализ генома, бактерия, протеом.*

## Резюме

Цель данной работы – исследование генома и протеома бактерии *Borreliella burgdorferi* B31, которая является возбудителем болезни Лайма.

В данной работе было определено общее число белков, число генов на различных цепях ДНК, различия в числе генов на плаزمиде и хромосоме, число квазиоперонов.

Также была выдвинута и проверена гипотеза о равномерном распределении генов по цепям ДНК. Дополнительно были составлены гистограммы длин белков и длин перекрываний генов.

## 1. Введение



*Borreliella burgdorferi* (*Borrelia Burgdorferi*) – наиболее распространенный в Северной Америке вид-возбудитель болезни Лайма<sup>1</sup>. Впервые была выделена в 1982 году Вилли Бургдофером<sup>2</sup>. Является Грамм-отрицательной спирохетой<sup>3</sup> и несет семь периплазматических флагелл и одну осевую нить на одну клетку; длина бактерии – 5-20 микрометров<sup>4</sup>. Двигается, волнообразно изгибая клетку. Это медленно растущий микроаэрофилл и мезофилл, чье деление происходит раз в 24-48 часов<sup>5</sup>. Способна жить в среде без железа, используя в активных центрах ферментов марганец<sup>6</sup>. Геном включает одну линейную хромосому и кольцевые плазмиды (до 21). Оптимум температур колеблется от 33° до 37°С.

**Рисунок 1: бактерия в среде**

Геном содержит 1522530 пар оснований на хромосоме, процент G-C пар – 28,6, число генов – 1531.

Болезнь Лайма – инфекционное трансмиссивное заболевание, которое передается иксодовыми клещами. Основные симптомы – жар, головные боли, кожная сыпь, в редких случаях наблюдаются поражения суставов, сердца и нервной системы. Несвоевременная и неадекватная терапия могут привести к хронической форме болезни, которая может закончиться инвалидностью или смертью<sup>7</sup>.

## 2. Материалы и методы

Для проверки гипотезы и составления сопроводительных таблиц и гистограмм была использована программа MS Excel. Все сделанные таблицы можно найти в сопроводительных материалах.

Информация о белках генома была получена из базы данных Genome NCBI<sup>1</sup> и банка данных GenBank.

Для анализа данных использовались различные статистические функции MS Excel: СРЗНАЧ, СТАНДОТКЛОН, МЕДИАНА, МИН, МАКС, а также сделанная плоская и сводная таблицы.

Для проверки гипотезы случайного распределения генов был использован критерий хи-квадрат (его расчёт можно найти в таблицах).

## 3. Результаты и обсуждение

### 3.1 Число генов по категориям

Таблица 1

# feature	class	Число
[-] CDS	with_protein	1339
Всего (CDS)		1339
[-] gene	protein_coding	1339
gene	pseudogene	192
Всего (gene)		1531
<b>Итого</b>		<b>2870</b>

В рамках данной работы все гены были подразделены на несколько категорий: гены, кодирующие белок, гены, кодирующие РНК и псевдогены.

Всего данный штамм несет 1531 ген, из которых белок кодируют 1339.

В ходе анализа выяснилось, что данный штамм не несет в своем

геноме генов, кодирующих РНК (из таблицы 1). Это может говорить о некорректности исходных данных.

Так же невелико число и псевдогенов, что говорит о достаточно неплохой изученности данной бактерии и данного штамма, что не удивительно, так как она является инфекционным агентом.

### 3.2 Распределение генов по прямой и обратной цепи, распределение генов между хромосомой и плазмидами

Так же в ходе данной работы мною был проведен анализ распределения генов различных классов по хромосомам и плазмидам.

Таблица 2

	Хромосома	Плазмиды
Белковые гены	753	586
Псевдогены	72	120

<sup>1</sup>[ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/008/685/GCF\\_000008685.2\\_ASM868v2/GCF\\_000008685.2\\_ASM868v2\\_feature\\_table.txt.gz](ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/008/685/GCF_000008685.2_ASM868v2/GCF_000008685.2_ASM868v2_feature_table.txt.gz)

Исходя из данных Таблицы 2 можно заметить, что белок-кодирующие гены распределены между хромосомой и 21 плазмидой в соотношении 1,285:1, то есть примерно поровну, а вот псевдогены в основном расположены на плазидах (соотношение 0,6:1). Всего на хромосоме находится 825 генов, а на плазидах – 706, что говорит о том, что гены расположены приблизительно равномерно между ними.

Затем был проведен анализ распределения генов по прямой и обратной цепям. Из данных Таблицы 3 можно выявить, что большая часть генов бактерии сосредоточена на прямых цепях хромосомы и плазмид (61,2% белок-кодирующих генов и 57% псевдогенов).

**Таблица 3**

Цепь	Гены	Псевдогены	Гены РНК
Прямая цепь	819	109	0
Обратная цепь	520	83	0

### 3.3 Проверка гипотезы о случайном распределении генов по прямой и обратной цепям

**Таблица 4**

Цепь	Количество генов	Ожидаемое распределение
Прямая	928	765
Обратная	603	766
Вероятность по критерию Пирсона:		0,0000000000000001

С помощью статистического критерия хи-квадрат была проверена гипотеза о равномерном распределении генов на прямой и обратной

цепи. Было проведено сравнение в целом по геному и отдельно по генам, кодирующим белок и псевдогенам по отдельности. Статистические расчеты, показанные в Таблице 4 и Таблице 5 показывают, что выдвинутая гипотеза подтверждается для всех исследованных типов генов.

Эти расчеты подтверждают теорию о равнозначности двух цепей ДНК для любого организма. Наиболее равномерно распределены по цепям псевдогены. Большой разрыв между количеством генов на цепях можно связать с некоторой некорректностью исходных данных.

**Таблица 5**

Цепь	Белковые	Псевдогены	Ожидаемое (белки)	Ожидаемое (псевдогены)
Прямая цепь	817	109	667	95
Обратная цепь	518	82	668	96
Вероятность о критерию Пирсона:			0,000000000000	0,04276102768

### 3.4 Распределение белков по длинам в геноме

В рамках работы так же был проведен анализ длин белковых последовательностей. Как

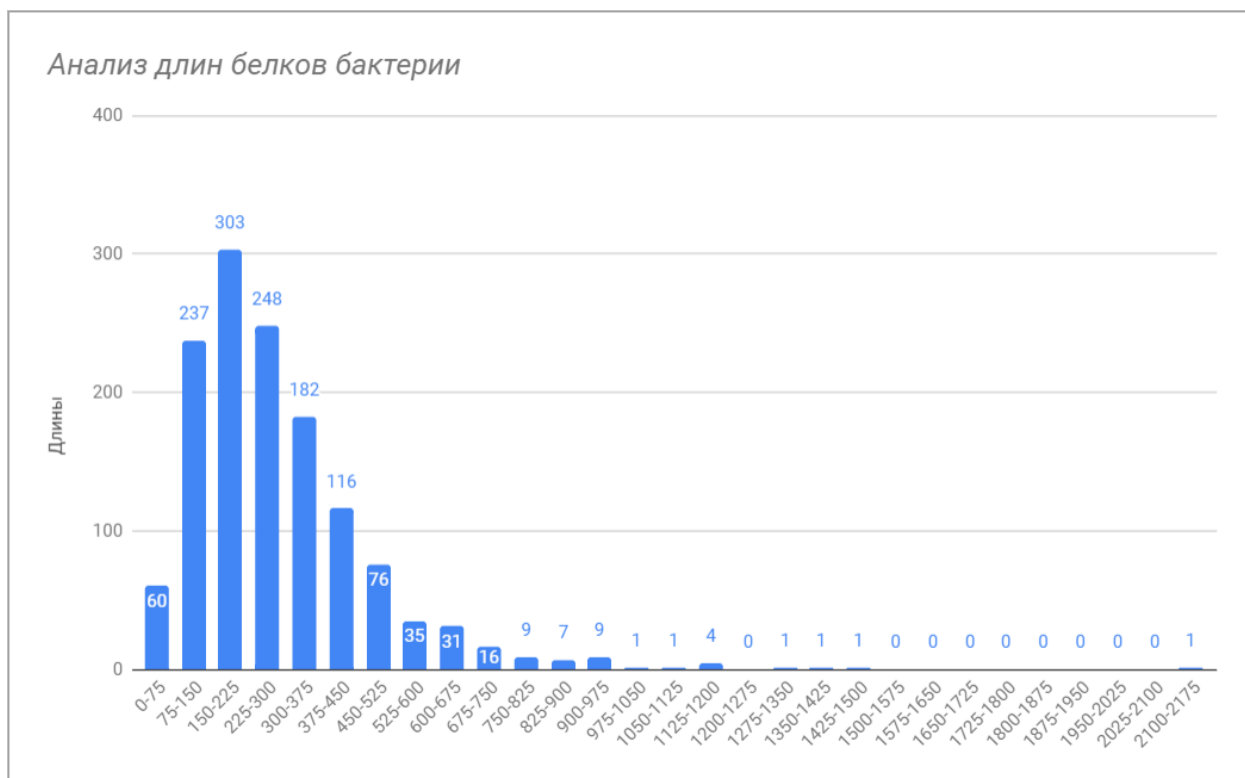
**Таблица 6**

Статистика длин белков	
Минимальная	40
Максимальная	2167
Медиана	249
Среднее отклонение	135,2472898

показано на гистограмме (Рисунок 2), наиболее часты белки длиной 75-375 аминокислотных остатков (составляют 72,4% всех белков). Это может быть связано с паразитическим образом жизни данной бактерии. Около 5% всех белков имеют длину менее 75 нуклеотидов. 20,5% всех белков бактерии имеет

длину от 375 до 750 аминокислотных остатков, и лишь 35 белков имеет длину более 750.

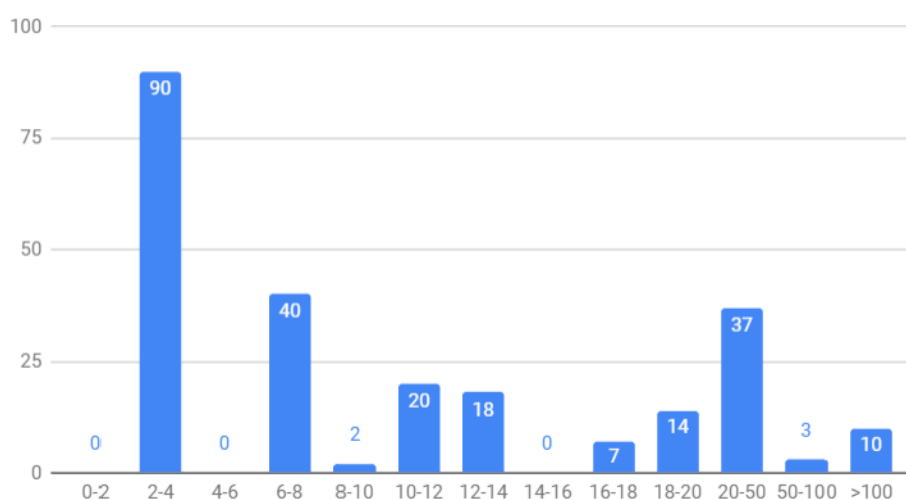
В Таблице 6 отражены минимальные, максимальные и медианные длины. Минимальная составляет 40 остатков, максимальная – 2167, а медианная – 249. Из этих данных можно сделать вывод, что большинство белков данной бактерии имеют небольшую длину, так как она не нуждается в длинных последовательностях. Из среднего отклонения можно сказать, что подавляющее большинство белков имеет длину от 114 до 384 остатков.



**Рисунок 2: Гистограмма длин белков**

### 3.5 Анализ квазиоперонов и перекрытий генов

#### *Длины перекрытий по 1 цепи*



**Рисунок 3: Гистограмма длин перекрытий**

В ходе работы так же было посчитано количество квазиоперонов<sup>2</sup>. В геноме их оказалось 647 (расчеты можно увидеть в сопроводительных таблицах). Также было посчитано количество перекрываний генов. Их оказалось 284, из которых 241 располагались на одной цепи. Из Рисунка 3 видно, что по одной цепи гены перекрываются преимущественно на 2-4 нуклеотида (37,3%), 6-8 нуклеотидов (16,6%) и 20-25 нуклеотидов (15,4%).

### 3.6 Анализ классов белков

В ходе работы так же был проведен анализ белков по классам. Как показано в таблице 7, максимальное число белков (56,3%) относятся к классу «другие», в который попали многие белки, которые играют огромную роль во внутриклеточном метаболизме. Также можно отметить огромное количество гипотетических белков, которых насчитывается 33,38%. В геноме примерно одинаковое число рибосомальных и мембранных белков, транспортных насчитывается несколько меньше.

На 1 м.п.н. в геноме бактерии насчитывается примерно 550 генов (для хромосомы).

Большое количество гипотетических белков можно объяснить недостаточной изученностью данного штамма бактерии.

Таблица 7

	Гипотетические	Рибосомальные	Транспортные	Мембранные	Другие
Число белков	447	54	36	48	754

## 4. Заключение

В ходе работы был проведен геномный и протеомный анализ бактерии-возбудителя болезни Лайма *Borrelia burgdorferi* B31. В процессе анализа было выявлено довольно много странных и интересных закономерностей, которые требуют дальнейшего изучения. Также необходимо проверить корректность исходных данных, и если они корректны, то факт отсутствия у бактерии генов РНК выглядит очень необычно и требует тщательного исследования.

## 5. Сопроводительные материалы

1. Сопроводительные таблицы:  
[https://kodomo.fbb.msu.ru/~zubareva.valeria/term1/zubareva\\_tables.xlsx](https://kodomo.fbb.msu.ru/~zubareva.valeria/term1/zubareva_tables.xlsx)

## 6. Благодарности

Хотелось бы выразить благодарность преподавателям биоинформатики факультета биоинженерии и биоинформатики МГУ: Алексеевскому А.В., Спиринову С.А за предоставленные знания и приобретенные навыки. Также выражаю благодарность Грызунову Никите за помощь при подготовке сопроводительных таблиц.

---

<sup>2</sup> максимальная последовательность генов, закодированных на одной цепочке с промежутками между генами не более порога 100 п.н.

## 7. Список литературы

1. <https://www.cmd-online.ru/vracham/spravochnik-vracha/iksodovye-kleshchevye-borreliozy/>
2. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2440571/>
3. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4688064/>
4. <http://adsabs.harvard.edu/abs/2000PNAS...9710899M>
5. <https://www.ncbi.nlm.nih.gov/genome/738>
6. <https://trove.nla.gov.au/work/8347273?selectedversion=NBD41788640>
7. <https://www.sciencedirect.com/science/article/pii/S0891552015000306?via%3Dihub>