

# Мини-обзор генома и протеома бактерии *Burkholderia* sp. штамма CCGE1001

Бусыгин С.А.<sup>1\*</sup>

<sup>1</sup>МГУ им. М.В.Ломоносова, факультет биоинженерии и биоинформатики, 1 курс

18.12.2016

## РЕЗЮМЕ

Данная работа посвящена анализу протеома и генома бактерии *Burkholderia* sp. штамма CCGE 1001. Расчитано количество генов по типам кодируемых продуктов, выявлены закономерности в распределении белков по длинам, найдено соотношение генов белков и РНК на прямой и комплементарной цепочках ДНК, проверена гипотеза о случайном распределении генов по цепочкам, установлено количество “квазиоперонов” в геноме бактерии и их величина, а также представлена статистика белков по критериям их существования.

## 1 ВВЕДЕНИЕ

Бактерии рода *Burkholderia* относятся к Грам-отрицательным бактериям, а точнее к бета-протеобактериям и не образуют спор. Их часто находят в почве и воде, особенно в местностях с тропическим климатом. В роду есть патогенные представители, вызывающие, в том числе, сап и мелиоидоз, из-за чего их считают потенциальными орудиями биотерроризма.[1] К тому же, бактерии характеризуются устойчивостью к разным видам антибиотиков, возможностью заражать различные виды животных, простотой получения монокультуры и отсутствием специфических симптомов. Вакцины против патогенных представителей рода пока не разработаны. С другой стороны, необходимо отметить высокую способность некоторых штаммов разлагать хлорорганические пестициды и полихлорированные дифенилы.[2] Геном состоит из двух хромосом длиной 4063449 и 2770302 пар оснований, а также плазмиды длиной 419743 пары оснований и был секвенирован 7 февраля 2011 года. У бактерии 6537 генов, из них 6453 кодируют белки. [3] В данной работе рассмотрены следующие вопросы: распределение белков по их длинам, количество генов белков и генов РНК на прямой и комплементарной цепях ДНК, а также распространённость квазиоперонов в геноме, роль длинных пересечений генов и критерии существования белков.

## 2 МЕТОДЫ И МАТЕРИАЛЫ

Данные о геноме и протеоме были взяты с сервера NCBI. Дополнительные данные о существовании белков были получены из базы данных Uniprot (1). Все данные были обработаны с помощью программы Microsoft Office Excel 2010 . (2)

## 3 РЕЗУЛЬТАТЫ

В ходе анализа генетических данных были получены и обобщены следующие результаты.

### 3.1 Количество генов в различных категориях

Все гены организма однозначно распределяются по группам на основании характера кодируемых ими продуктов.

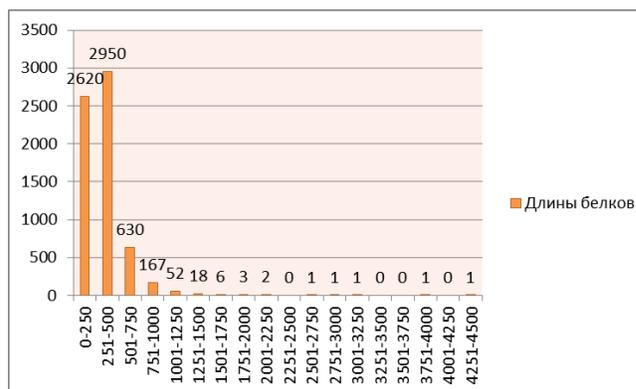
Таблица 1. Распределение генов по их продуктам.

Категория	Число генов в категории	В расчете на миллион нуклеотидов
CDS	6453	908,87
misc_RNA	1	0,14
rRNA	18	2,54
tRNA	62	8,73
ncRNA	2	0,28
tmRNA	1	0,14
Всего	6537	920,70

### 3.2 Распределение длин белков в протеоме

В протеоме бактерии подавляющее большинство белков не превышают 500 аминокислот в длину, хотя присутствуют белки, содержащие до 4347 аминокислот. Среди таких крупных белков несколько представителей семейства филаментозных гемоагглютининов, расположенных на поверхности клетки.[4] Бактериям эти белки помогают образовывать биопленки и удерживаться в организмах хозяев. На гистограмме представлено распределение белков по их длине.

Рисунок 1. Распределение белков по их длине.



\*sergebus@fbb.msu.ru

### 3.3 Распределение генов по цепям

Проведен расчет количества генов белков и РНК на разных цепях. Для этой бактерии характерно серьезное различие в количестве генов белка на прямой и обратной цепи.

Таблица 2. Распределение генов по цепям.

Цепь	Гены белка	Гены РНК
Прямая	3099	42
Обратная	3354	42

### 3.4 Проверка гипотезы случайного распределения генов по цепочкам.

Гипотезы случайности распределения генов по цепочкам ДНК проверялась с помощью биномиального распределения. Данные этого расчета приведены в таблице 2. В правом столбце таблицы представлены числовые значения вероятности получить такое же или меньшее количество генов на одной из цепей при условии случайного и равновероятного распределения генов по цепям. По вычисленным в таблице значениям можно предположить, что, из-за низкой вероятности сложившейся ситуации, на распределение генов по цепям воздействовали неизвестные факторы, вызвавшие значительные отклонения от случайного распределения генов.

Таблица 3. Оценка гипотезы о случайном распределении генов по цепям.

Биномиальное распределение	Значение
По всем генам	0,000839251
По генам белка	0,000782635
По генам РНК	0,543398819

### 3.5 Квазиопероны в геноме

Квазиоперон – некоторая совокупность генов, находящихся на одной цепи с промежутками между ними, меньшими порогового значения. Предполагается, что благодаря этому они транслируются с одной мРНК. При пороговой величине 100 нуклеотидов в геноме насчитывается 1329 квазиоперонов. Большая их часть состоит из двух или трех генов. Квазиоперонов, как и генов вообще, больше на обратных цепях ДНК. Самый крупный встреченный квазиоперон содержит 20 генов, больше половины из которых явного фагового происхождения. Возможно, этот квазиоперон возник при внедрении в геном бактерии плотно упакованной наследственной информации бактериофага. На рисунке 2 представлено распределение квазиоперонов по количеству генов, а таблица 4 позволяет оценить близость генов в геноме друг к другу через сопоставление данных о количестве оперонов при различных величинах порогового значения.

Рисунок 2. Распределение квазиоперонов по длине при пороге 100 п.н..



Таблица 4. Количество квазиоперонов в зависимости от порогового значения.

Порог	Количество квазиоперонов
15	869
25	939
50	1102
100	1329
150	1438

### 3.6 Длинные пересечения генов

В геноме бактерии 780 пересечений генов. Большинство из них происходят на одной цепи в одной рамке считывания. Самое крупное пересечение имеет длину 100 п. н. Так как в этом случае продукт одного из пересекающихся генов – гипотетический белок, можно предположить, что это вызвано ошибками в интерпретации результатов секвенирования.

Таблица 5. Информация о пересечениях генов.

Количество пар пересекающихся генов	На одной цепи	На разных
Всего	720	60
В одной рамке считывания	558	33
Со сдвигом на 1 нуклеотид	161	17
На два нуклеотида	1	10

### 3.7 Достоверность существования белков

В таблице представлена информация о количестве белков, существование которых было подтверждено различными методами. Подавляющее большинство было предсказано на основании анализа последовательностей ДНК и лишь четверть – из сопоставления с известными белками.

Таблица 6. Подтверждение существования белков

Способ	Количество белков
Свидетельства на уровне транскриптов	2
Предположение на основании гомологии	1647
Предсказание	4754

[\[3\]](#)

Информация из базы данных NCBI  
[https://www.ncbi.nlm.nih.gov/assembly/GCA\\_000176935.3](https://www.ncbi.nlm.nih.gov/assembly/GCA_000176935.3)

[\[4\]](#)

Молекулы клеточной адгезии: регуляция межклеточных взаимодействий.  
[http://medbiol.ru/medbiol/makr\\_cam/0000dc97.htm](http://medbiol.ru/medbiol/makr_cam/0000dc97.htm)

## 4 ОБСУЖДЕНИЕ

Гены не распределены случайно по двум цепочкам ДНК, а преобладают на обратной цепи (гипотеза подтверждена статистически). Количество квазиоперонов тоже зависит от цепи, что еще раз подтверждает гипотезу о приоритетном расположении на обратной цепи. Даже при небольшом пороговом значении около 20% генов находятся достаточно близко друг к другу, что может объясняться оперонной организацией генома, типичной для бактерий. Протеом типичен для прокариот, так как содержит, в основном, небольшие полипептидные цепи. Существование белков подтверждено, в основном, содержанием последовательностей ДНК.

## 5 СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

[\(1\)](#)

Файл с данными из базы Uniprot находится по адресу  
<http://kodomofbb.msu.ru/~sergebus/term1/uniprotdata.xlsx>

[\(2\)](#)

Файл со всеми расчетами и плоской таблицей, характеризующей геном, находится по адресу  
<http://kodomofbb.msu.ru/~sergebus/term1/Mybacteria.xlsx>

## 6 БЛАГОДАРНОСТИ

Автор выражает благодарность доценту Факультета биоинженерии и биоинформатики МГУ им. Ломоносова Алексеевскому А. В. за ценные советы и замечания, высказанные в процессе подготовки данной работы.

## 7 СПИСОК ЛИТЕРАТУРЫ И ДРУГИХ ИСТОЧНИКОВ

[\[1\]](#)

Alan Cann, Glanders and Melioidosis – coming soon to a crowded shopping centre near you?, MicrobiologyBytes 12.11.2007  
<https://microbiologybytes.wordpress.com/2007/11/12/glanders-and-melioidosis-coming-soon-to-a-crowded-shopping-centre-near-you/>

[\[2\]](#)

Burkholderia: Molecular Microbiology and Genomics; Horizon Bioscience Series; Tom Coenye, Peter Vandamme; Horizon Scientific Press, 2007  
<http://www.worldcat.org/title/burkholderia-molecular-microbiology-and-genomics/oclc/255160962>