

Мини-обзор протеома бактерии *Bordetella holmesii* H558

Миронова Екатерина

МГУ им. М. В. Ломоносова факультет биоинженерии и биоинформатики

e.mironova@fbb.msu.ru

РЕЗЮМЕ

В данном обзоре я продолжаю знакомиться с бактерией *Bordetella holmesii* H558. Я приведу мои выводы и результаты, связанные с геномом *B. holmesii*, полученные с помощью методов, доступных любому пользователю. В обзоре я рассматриваю распределение длин белков, число генов белков и РНК на цепях ДНК, опероны бактерии, сдвиг рамки считывания. Моя работа будет полезна людям, недавно заинтересовавшимся биоинформатикой и осваивающим простые методы обработки генетической информации. С результатами предыдущей работы можно ознакомиться на моей веб-странице [\[1\]](#).

1 ВВЕДЕНИЕ

Bordetella holmesii H558 - это штамм грамотрицательных (не окрашиваются кристаллическим фиолетовым при окрашивании по Граму) неподвижных бацилл, принадлежащих роду *Bordetella*. Бактерия не сахаролитическая (не расщепляет углеводы и высокоатомные спирты), продуцирует коричневый пигмент. Является человеческим патогеном: вызывает сепсис, эндокардит и респираторные заболевания, особенно сильно воздействует на пациентов с ослабленным иммунитетом (например, больных асплинией или СПИДом). Часто встречается в сочетании с *Bordetella pertussis* (возбудитель коклюша) [\[2\]](#). По данным NCBI, в геноме *Bordetella holmesii* H558 3696428 пар нуклеотидов и 3612 генов: 3396 кодируют полипептиды, 150 - псевдогены, рРНК 3 (5S), 3 (16S), 3 (23S), тРНК 56, 1 некодируемая РНК. Дата секвенирования генома: 8 февраля 2016 года [\[3\]](#).

Микрофотографии представителей рода *Bordetella*



Рис. 1. *Bordetella bronchiseptica* [\[4\]](#)

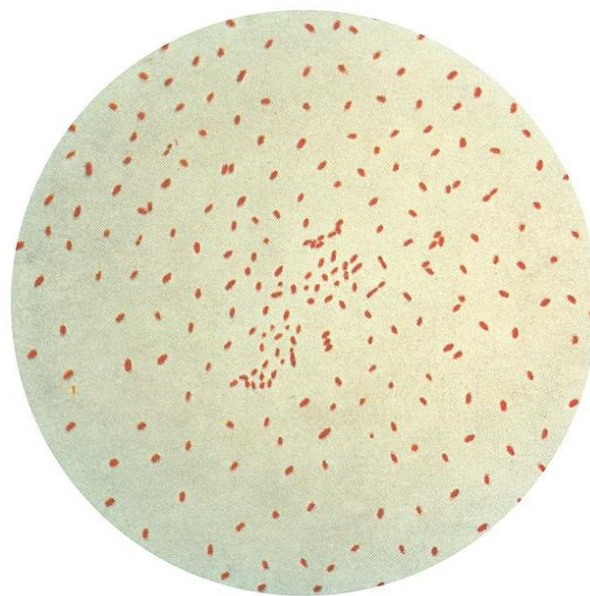


Рис. 2. *Bordetella pertussis* [\[5\]](#)

2 МАТЕРИАЛЫ И МЕТОДЫ

Использованы данные с сайта NCBI: *Bordetella holmesii* H558, complete genome, сайт NCBI [\[3\]](#), GCA_000341485.2_ASM34148v2_feature_table [\[6\]](#)

Данные обрабатывались в Microsoft Office Excel: построение диаграммы, таблиц, поиск необходимых данных при множественном условии (СЧЁТЕСЛИМН).

3 РЕЗУЛЬТАТЫ

В геноме *Bordetella holmesii* H558 3696428 пар нуклеотидов и 3483 генов: 3396 кодируют белки, 21 псевдоген, рРНК 3 (5S), 3 (16S), 3 (23S), тРНК 56, 1 некодируемая РНК. Как видно, не все результаты, полученные при работе с Excel, совпадают с данными на сайте NCBI (во введении, рис. 6, рис. 7). Количество генов на 1 млн нуклеотидных пар приблизительно равно 942,3 (расчёт по данным Excel).

3.1 Общее распределение длин белков протеома *Bordetella holmesii* H558

Большинство генов в длину находятся в диапазоне 250 – 350 аминокислотных остатков. Почти все белки длиной меньше 1600 а. о., но единично встречаются белки длиной в диапазоне 1900 – 2000 и 2800 – 2900 а. о. Обратим на них внимание. Самый длинный белок – adhesin (2885 а. о.). Это белок клеточной мембраны, участвующий в процессах межклеточной адгезии (слипания клеток). Самый короткий белок бактерии - ABC transporter (28 аминокислот).



Рис. 3. Гистограмма с распределением длин белков B. holmesii

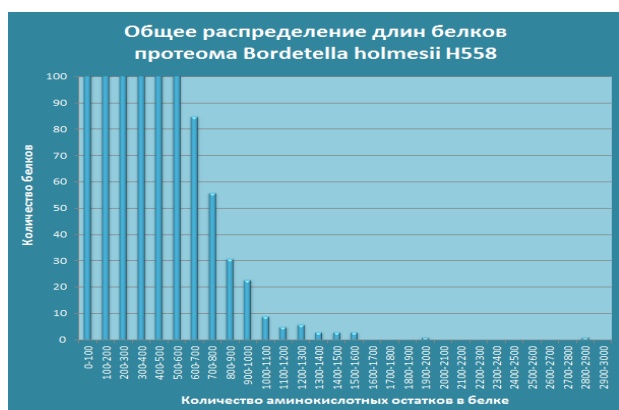


Рис. 4. Гистограмма, показывающая наличие белков большой длины

3.2 Распределение генов белков и РНК по цепям ДНК протеома Bordetella holmesii H558

В таблице распределения генов белков и РНК по прямой и комплементарной цепям были получены следующие результаты: генов РНК и на прямой, и на обратной цепях 33, а среди генов белков довольно большое отличие: 1866 на прямой, 1551 на комплементарной.

Табл. 1. Распределение генов по цепям ДНК

	Число генов белков	Число генов РНК
Прямая цепь	1866	33
Комплементарная	1551	33

3.3 Опероны Bordetella holmesii H558

У бактерий и архей оперон - участок ДНК с одним или несколькими генами белков, транскрибируемый в одну матричную ДНК. Таким образом, гены в одном опероне закодированы на одной цепочке ДНК. Как правило, расстояние между ними небольшое. Для начала проверим влияние величины диапазона на количество квазиоперонов: при увеличении порога на незначительную величину количество оперонов растет, но при значительном (около 50) заметно падает. При уменьшении порога сумма падает. Оставим оптимальную длину диапазона – 100. Квазиоперонов в прямой цепи ДНК 868, комплементарной 701. Но это кажущееся большое отличие, потому что соотношение с числом генов в цепи приблизительно одинаковое (46% и 44%)

3.4 Сдвиг рамки считывания

Пересекающихся генов на одной и той же цепи 423, а на разных цепях 92. Самый частый сдвиг рамки считывания – 3, а наибольший реальный – 92.

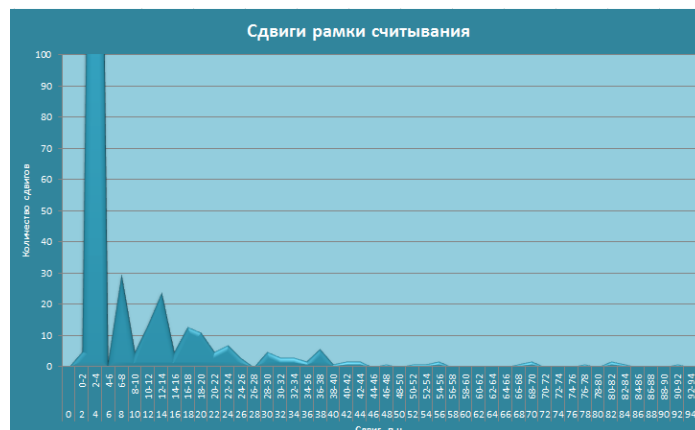


Рис. 5. Гистограмма встречаемости сдвигов рамки считывания определённой длины

4 ОБСУЖДЕНИЕ

Замечание при работе с ресурсом NCBI: Устройство сайта NCBI довольно сложно, поэтому нужно быть внимательным и брать данные из непротиворечащих источников.

```
##Genome-Annotation-Data-START##
Annotation Provider      :: NCBI
Annotation Date         :: 18/06/2015 13:35:50
Annotation Pipeline     :: NCBI Prokaryotic Genome Annotation
                        Pipeline
Annotation Method       :: Best-placed reference protein set;
                        GeneMarkS+
Annotation Software revision :: 3.0
Features Annotated      :: Gene; CDS; rRNA; tRNA; ncRNA;
                        repeat_region
Genes                   :: 3,612
CDS                     :: 3,396
Pseudo Genes           :: 150
rRNAs                   :: 3, 3, 3 (55, 165, 235)
complete rRNAs         :: 3, 3, 3 (55, 165, 235)
tRNAs                   :: 56
ncRNA                   :: 1
##Genome-Annotation-Data-END##
```

Рис. 6. Bordetella holmesii H558, complete genome [3]

Рис. 7. Bordetella holmesii list [7]

Также есть несоответствие в числе генов, выданным сайтом (приблизительно 3612-3614), и результатом, получающимся при работе в Excel (3483). Это происходит, потому что на шаге подготовки feature_table к работе в Excel были удалены строки gene, которые содержали дополнительные псевдогены.

4.1 Общее распределение длин белков протеома Bordetella holmesii H558

Белок адгезин – самый длинный белок моей бактерии. Говорит ли его большая длина об огромной важности этого белка для клетки? Сильно ли подвержены изменчивости аналогичные белки? Можно ли их использовать как критерии родов и штаммов бактерий?

ABC-транспортер в длину всего 28 аминокислот. Белок будет интересен для изучения пространственной структуры белков и их функций. Влияет ли его очень маленькая длина на распространение в живых существах?

4.2 Распределение генов белков и РНК по цепям ДНК протеома Bordetella holmesii H558

Предположение по формальности названий «прямая» и «комплементарная» цепь ДНК: прямой является цепь, где содержится больше генов.

Я считаю, что неравное распределение неслучайно, ведь функция второй цепи не только кодировка, но и обеспечение лучшей сохранности генетической информации, поэтому цепи не являются независимыми и одна из них преобладает в количестве генов.

4.3 Опероны *Bordetella holmesii* H558

Для начала, математическое утверждение: при увеличении порога на незначительную величину количество оперонов растёт, т. к. условия менее жёсткие, при значительном (около 50) падает, т. к. уменьшается само число возможных оперонов. При уменьшении порога сумма падает, т.к. условия стали более жёсткие.

Отношение количества квазиоперонов к количеству генов достаточно велико (приблизительно 50% в каждой цепи) и почти совпадает в цепях, в отличие от большой разности количества генов на двух цепях. Результат в 50% говорит о том, что в среднем в опероне 2 гена (по правилам статистики), но это утверждение нельзя считать абсолютно правильным, ведь в своих расчётах я учитывала и опероны длиной 1 ген.

4.4 Сдвиг рамки считывания

Самый частый сдвиг рамки считывания – 3, но тогда получается, что стоп-кодон одного гена совпадает со старт-кодоном другого. Получается противоречие, тогда пересечение генов – это, возможно, ошибка в предсказании самих кодирующих последовательностей.

5 ЗАКЛЮЧЕНИЕ

С помощью такого доступного инструмента как Excel можно прийти к следующим результатам: большинство длин белков бактерии *B. holmesii* H558 находится в диапазоне 50 – 400 аминокислотных остатков; распределение генов белков по обратной и прямой цепи неслучайно; по оперонам и сдвигу рамки считывания можно сказать, правильно ли предсказана кодирующая последовательность; квазиопероны примерно одинаково распространены в процентном соотношении по обеим цепям ДНК.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

[Файл Excel к мини-обзору](#)

БЛАГОДАРНОСТИ

Выражаю благодарность преподавательскому составу по практической информатике факультета биоинженерии и биоинформатики МГУ им. М. В. Ломоносова за помощь в оформлении данной статьи, определении методов исследования

СПИСОК ЛИТЕРАТУРЫ

- [1] Краткое описание белка бактерии *Bordetella holmesii* H558
- [2] *Bordetella holmesii*, Wikipedia
- [3] *Bordetella holmesii* H558, complete genome, сайт NCBI
- [4] Бордетеллы, Википедия
- [5] *Bordetella pertussis*, Wikipedia
- [6] Директория с файлами о *B. Holmessi* H558 на NCBI
- [7] *Bordetella holmesii* list