

Обзор протеома и транскриптома бактерии *Mycoplasma gallisepticum* S6

Рюмина Екатерина¹

¹Факультет биоинженерии и биоинформатики МГУ им. М. В. Ломоносова.

РЕЗЮМЕ

Данная работа направлена на изучение некоторых свойств транскриптома и протеома бактерии *Mycoplasma gallisepticum* S6, а именно: распределение белков по их длинам, распределение генов по прямой и комплементарной цепям (далее + и - цепи) и случайность этого распределения, число квазиоперонов и перекрываний генов (а также сдвиг рамки считывания при перекрывании генов). Результатом данной работы являются статистические данные, касающиеся свойств, указанных выше.

1 ВВЕДЕНИЕ

Систематическое положение:

Царство Bacteria
Terrabacteria
Тип Tenericutes
Класс Mollicutes
Отряд Mycoplasmatales
Семейство Mycoplasmataceae
Род Mycoplasma
Вид *Mycoplasma gallisepticum*
Штамм S6

Все представители рода *Mycoplasma* — облигатные паразиты. Кроме этого, их отличает крайне малый размер (0,3–0,8 микрон в диаметре), крайне малый геном и отсутствие клеточной стенки. Клеточная мембрана этих бактерий богата белками (они составляют до трети массы мембраны), которые помогают бактериям прикрепляться к клетке хозяина, а затем внедряться в нее. В клетке микоплазма может довольно долго персистировать, что может привести к смерти клетки. *M. gallisepticum* вызывает хронические респираторные заболевания у куриц и прочих птиц. Причем и лечение, и иммунный ответ хозяина осложнены тем, что бактерии живут в клетках хозяина. Также *M. gallisepticum* — один из немногих подвижных представителей рода *Mycoplasma*. Бактерия скользит по поверхности клетки. Механизм скольжения изучен до конца, но предположительно в нем задействованы белки

актин-миозинового цитоскелета, а также уникальные для данной группы микроорганизмов белки [2].

Краткое описание генома: геном состоит из одной хромосомы, длиной 985433 пар нуклеотидов [3]. Исследуемый геном был секвенирован 25.02.2016. Идентификатор генома: CP006916.3 [4].

2 МЕТОДЫ

Данные о геноме были взяты из базы данных NCBI [5]. Анализ данных производился с помощью программы OpenOffice 4.1.3

При построении таблиц и диаграмм были использованы встроенные функции по построению диаграмм, функции “IF”, “COUNTIF”, “ABS”, “MOD” и другие, и методы сортировки.

При проверке гипотезы о случайности распределения генов, была использована функция “RANDBETWEEN”. Сама проверка состояла в том, что в 100 столбцах длиной, равной числу генов, с помощью этой функции выпадал либо ноль, либо единица (которые означали + и - цепи). Далее считалось отклонение получившегося числа единиц от среднего (от половины числа генов). И сравнивалось с отклонением в реальном геноме. Количества перекрываний генов и квазиоперонов считались по столбцам, в которых указаны начала и концы генов, с использованием функции “IF”.

3 РЕЗУЛЬТАТЫ

3.1 Распределение числа белков в зависимости от их длины

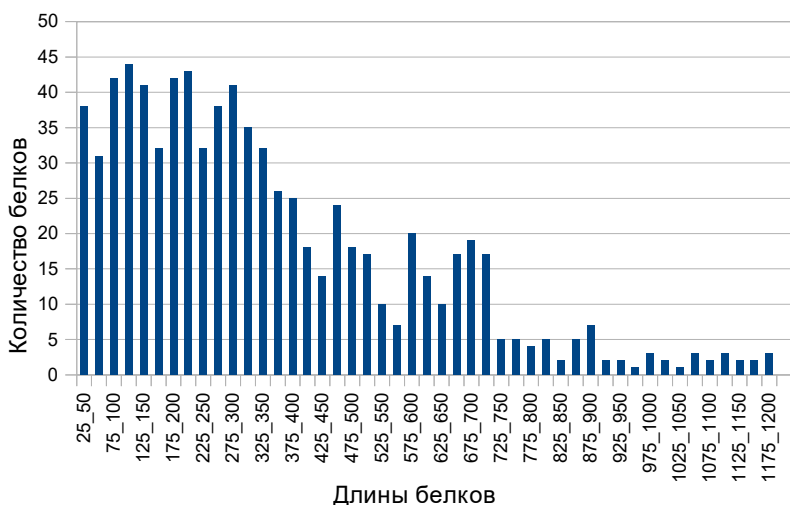
В геноме бактерии *M. gallisepticum* S6 закодировано 816 белков, данные о которых приведены ниже (в таблице 2 и гистограмме 1).

Таблица 1. Основные характеристики протеома бактерии *M. gallisepticum* S6.

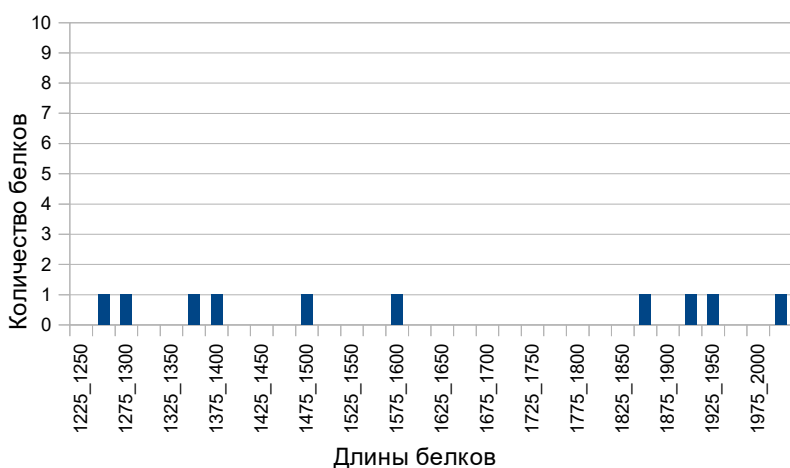
характеристика	значение
Максимальная длина белка	2019
Минимальная длина белка	30
Медиана длин белков	289
Среднее значение	360,7

Распределение белков по длинам показано в гистограммах 1 и 2.

Гистограмма 1. Белки до 1200 а. к.



Гистограмма 2. Белки от 1200 а. к.



3.2 Распределение генов по цепям

В геноме бактерии *M. gallisepticum S6* всего 879 генов: 816 из них кодируют белки, а 58 — разнообразные РНК. Также есть 5 псевдогенов. В таблице 2 приведено распределение этих генов по + и - цепям.

Таблица 2. Распределение генов по цепям

	Число генов	
	CDS	RNA
+ цепь	432	39
- цепь	384	19
всего	816	58

Из-за разнообразия РНК может быть интересно посмотреть на распределение генов РНК по цепям, которое представлено в таблице 3.

Таблица 3. Распределение генов РНК по цепям

	Число генов					
	antisense_RNA	rRNA	tRNA	tmRNA	ncRNA	RNase_P_RNA
+ цепь	3	6	28	0	1	1
- цепь	13	1	4	1	0	0
всего	16	7	32	1	1	1

Данные в таблицах позволяют нам сделать несколько выводов.

1. На + цепи закодировано больше генов (как суммарно, так и отдельно генов белков и РНК)
2. Разница между числом генов, кодирующих белки, составляет 5,9% (48 генов), а генов, кодирующих РНК — 51,3% (20 генов).
3. У бактерии *Mycoplasma gallisepticum S6* есть РНК, играющая роль фермента (рибозим), - Rnasa P (РНКаза П).

3.3 Гипотеза случайности распределения генов

Как описывалось выше, проверка проводилась путем «подбрасывания монетки». Из 100 симуляций, в каждой из которой 816 раз (сколько и CDS) выкидывалась единица или ноль, в 71 случае отклонение от среднего (408 генов) было больше, чем в геноме (24 гена). Из 100 симуляций, в каждой из которых выкидывалась единица или ноль ровно 58 раз (сколько генов РНК), только 2 раза отклонение было больше реального (того, которое наблюдается в геноме).

3.4 Число квазиоперонов

Для удобства, назовем "квазиопероном" максимальную последовательность генов, закодированных на одной цепочке с промежутками между генами не более порога 50, 100, 200 и 500 п.н. (причем квазиоперон может состоять и из одного гена). Результаты подсчета числа квазиоперонов в зависимости от максимального расстояния между генами, при котором они все еще считаются квазиопероном, приведены ниже в таблице 4.

Таблица 4. Число квазиоперонов

максимальное расстояние	50	100	200	500
+цепь	230	177	121	77
-цепь	207	177	127	81

3.5 Число перекрытий

В таблице 5 представлены данные о пересечении генов одной цепи (то есть в этой таблице не учтены случаи, когда перекрываются гены на + и - цепях).

Таблица 5. Пересечения генов

Число пересечений генов	
+цепь	60
-цепь	54
всего	114

При перекрытии генов важно, сдвигается ли при этом рамка считывания. Ниже приведены статистические данные, отражающие число разных сдвигов (сдвиг на 0 нуклеотидов — то есть отсутствие сдвига, на 1 и на 2 нуклеотида) рамки при перекрытии генов.

Таблица 6. Сдвиг рамки считывания

сдвиг рамки	0	1	2
+цепь	0	48	15
-цепь	1	38	12
всего	1	86	27

ОБСУЖДЕНИЕ

Гипотеза о случайном распределении генов по цепям не была подтверждена для генов, кодирующий РНК, но была подтверждена (в рамках того способа проверки, который применялся в данной работе) для генов, кодирующих белки. Но метод, используемый мной, не стопроцентно достоверен. И для большей уверенности стоило бы проверить другим методом, нежели просто сравнение отклонений.

Белки имеют длину от 30 до 2019 а. к.. Причем самые мелкие белки — это предполагаемые белки и рибосомальные белки. Среди самых крупных много ДНК-полимераз, а также других белков, связанных с ДНК (гираз, эндонуклеаз и т. д.). В целом, гистограмма позволяет зрительно разделить белки на три группы, которые представлены в таблице 7.

Таблица 7. Группы белков

длина, а.к.	число белков	процент
Не более 400	543	66,5
От 400 до 800	219	26,8
От 800 до 1200	54	6,6

Можно заметить, что уменьшение числа белков с соответствующей длиной зависит от этой длины нелинейно (число белков падает быстрее увеличения длины). Число квазиоперонов уменьшается с увеличением максимально допустимого расстояния. Что логично, поскольку гены все больше и больше «слипаются» - какие-то гены, которые при максимально допустимом расстоянии в 50 нуклеотидов были слишком далеко, чтобы считаться квазиопероном, при максимально допустимом расстоянии 500 нуклеотидов уже подходят под его определение. И из нескольких независимых генов (каждый из которых считался отдельным квазиопероном) получается один новый. То есть их число уменьшается. Но это непрерывное уменьшение так же означает, что нет очень хорошо выделенных кластеров генов (которые находятся на близком расстоянии между собой, но далеко от других). По крайней мере их не видно при том масштабе, который был мной использован. Число генов, которые пересекаются на + цепи, составляет 13,2% (62) от всех генов на + цепи, на - цепи — 13,4%, что, в общем-то, немало. При этом в подавляющем количестве случаев рамка сдвигается (из всего сдвигов всего один раз рамка не меняется — это меньше 1% случаев).

ИСТОЧНИКИ

[1] - *Mycoplasma gallisepticum* (ID 1113) - Genome – NCBI
ссылка: <https://www.ncbi.nlm.nih.gov/genome/?term=Mycoplasma%20gallisepticum%20S6%5Borgn%5D>

[2] - Taxonomy browser (*Mycoplasma gallisepticum*)
ссылка:

<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=1006581>

[3] – Содержание/genomes/all/GCA/000/211/545/
GCA_000211545.6_ASM21154v6

ссылка: ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/211/545/GCA_000211545.6_ASM21154v6/GCA_000211545.6_ASM21154v6_assembly_stats.txt

[4] – ASM21154v6 - Assembly - NCBI
ссылка:

https://www.ncbi.nlm.nih.gov/assembly/GCA_000211545.6

[5] - Genomes - Genome — NCBI

ссылка: <https://www.ncbi.nlm.nih.gov/genome/genomes/1113>

Данные, подтверждающие приведенные выше расчеты и таблицы содержатся в сопроводительных материалах.
ссылка: <http://kodomofbb.msu.ru/~e.caterina/term1/term1.html>