

Краткий обзор протеома и генома бактерии *Ureaplasma parvum serovar 3 str.* ATCC 700970

Эрмидис Александр-Павел

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

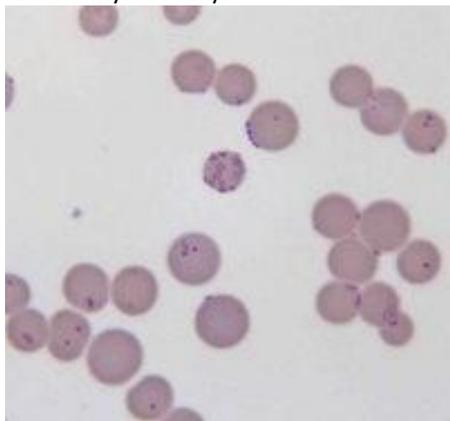
Аннотация

Представленная работа содержит в себе результаты исследования протеома и генома бактерии *Ureaplasma parvum serovar 3 str.* ATCC 700970. В ходе работы были рассмотрены особенности распределения белков по их длинам; закономерности распределения генов, кодирующих белки и РНК, учитывая направленность цепочки ДНК, на которой они располагаются.

1 ВВЕДЕНИЕ

Ureaplasma – род бактерий семейства *Mycoplasmataceae* порядка микоплазм, вызывающих воспаление слизистых оболочек половых органов и мочевых путей человека (уреаплазмоз); интересно отметить, что по своим свойствам *Ureaplasma* занимает промежуточное положение между вирусами и одноклеточными организмами; является специфичным паразитом, так как паразитирует микоплазма на мембранах клеток хозяина; являются грамотрицательными, неподвижными, аэрофильными организмами.

Непосредственно вид *parvum serovar 3 str.* ATCC 700970 круглой или коккобациллярной формы. Встречается преимущественно в мочеполовом тракте человека с менее частыми проявлениями в областях рта или дыхательных путей. Обычно это комменсальный организм, однако он может стать оппортунистическим патогеном. Может стать причиной неблагоприятных исходов беременности, неонатальных инфекций и мужского бесплодия. Как и любая микоплазма, не образует клеточную стенку.



НН

Рисунок 1. Колонии возбудителя уреаплазмоза – *Mycoplasma genitalium*.

Геном бактерии содержит 647 генов, которые кодируют 591 белок. Также встречаются последовательности, кодирующие 17 псевдогенов, 6 рРНК, 1 тмРНК, 30 тРНК и 2 нкРНК (тмРНК – транспортно-матричная РНК, нкРНК – некодирующая РНК, тРНК – транспортная РНК, рРНК – рибосомальная РНК).

2 МЕТОДЫ

Данные о протеоме бактерии были взяты с сайта-NCBI, а для анализа была использована программа Microsoft Office Excel 2013.

После скачивания архива GCF_000006625.1_ASM662v1_feature_table.txt.gz с протеомом, данные были импортированы и представлены в виде таблицы. После этого были созданы дополнительные листы:

1. Для подсчета числа белков и генов РНК по категориям. Примерная оценка количества генов на 1 млн. пар нуклеотидов.
2. Гистограмма длин белков из протеома бактерии.
3. Таблица, предоставляющая информацию о расположении числа генов белков, псевдогенов и РНК на прямой и комплементарной цепочках ДНК

При работе для каждой задачи были составлены отдельные таблицы с нужной информацией, которая с помощью фильтра данных была извлечена из основной таблицы. На протяжении работы для подсчета элементов были использованы функции: «СЧЁТЕСЛИМН()», «СЧЁТЕСЛИ()». Помимо этого использовались функции: «МАКС()», «МИН()», «СРЗНАЧ()», «МЕДИАНА()». Гистограммы были построены с карманами с шагом 1000, с помощью функции «СЧЁТЕСЛИМН()».

3 РЕЗУЛЬТАТЫ

3.1 АНАЛИЗ ПРОТЕОМА БАКТЕРИИ

В геноме *Ureaplasma parvum serovar 3 str.* ATCC 700970 насчитывается 647 генов, что можно увидеть в таблице 1, в которой показаны количества тех или иных видов последовательностей.

Таблица 1. Разновидности последовательностей в геноме изучаемой бактерии

	Белок	Псевдоген	нкРНК	тРНК	рРНК	тмРНК
Кол-во	591	17	2	30	6	1

Так же была построена таблица, отражающая основные статистические свойства протеома изучаемой бактерии.

Таблица 2. Основные статистические свойства протеома *Ureaplasma parvum serovar 3 str. ATCC 700970*.

Минимальная длина	105
Максимальная длина	15018
Средняя длина	1121,087
Стандартное отклонение	1124,388
Медиана	864

Распределение длин белков отражено на гистограмме (диаграмма 1).

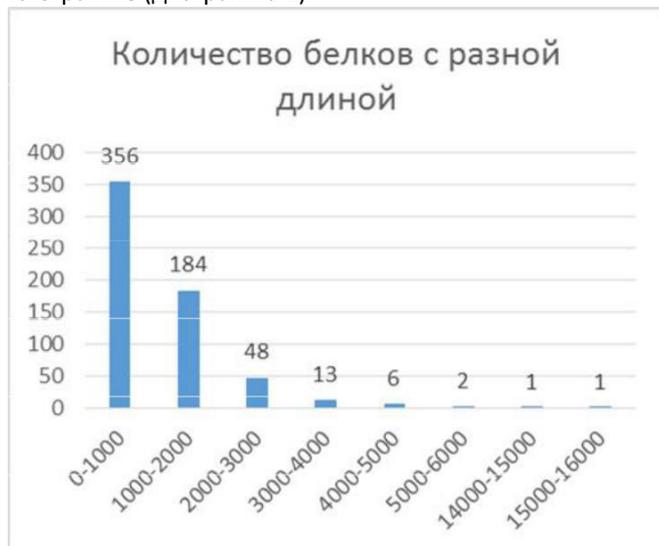


Диаграмма 1. Распределение белков бактерии по длине, кодирующихся на хромосоме 1.

По полученным данным можно заметить, что основное количество белков имеют длину до 3000 аминокислотных остатков при этом большую их часть составляют белки с длиной до 1000 аминокислот. Так же бросается в глаза отсутствие белков с длинами от 6000 до 14000 остатков.

Так же можно отметить довольно большую разницу между средними длинами и их медианами, что говорит о том, что содержание генов, кодирующих продуктов с маленькой длиной больше.

3.2 АНАЛИЗ ГЕНОМА БАКТЕРИИ

После анализа расположения генов по разным цепям ДНК была составлена таблица 2.

Таблица 2. Распределение генов бактерии по цепочкам ДНК в различных частях генома.

Цепь ДНК	Число генов	Белков	псевдогенов	Число генов	РНК
Прямая	322	6	16		
Обратная	269	11	22		

4 ВЫВОДЫ

При выполнении этой работы была проведена работа с данными генома *Ureaplasma parvum serovar 3 str. ATCC 700970* с помощью программы Excel 2013.

Были определены количество и виды последовательностей в геноме, посчитаны некоторые статистические величины для описания некоторых свойств генома, а также наглядно продемонстрировано различие в количестве белок-кодирующих последовательностей разной длины.

5 МЕСТО РАСПОЛОЖЕНИЯ ДОКУМЕНТА С РАСЧЕТНЫМИ ДАННЫМИ И ВСПОМОГАТЕЛЬНЫМИ ТАБЛИЦАМИ:

<http://kodomofbb.msu.ru/~aleksandr.pavele/term1/pr12.html>

6 Благодарности.

Выражаю благодарность преподавательскому составу кафедры биоинформатики факультета биоинженерии и биоинформатики МГУ имени М.В.Ломоносова, студентке 2 курса Софье Бруман и студенту 1 курса Ивану Поддъякову за оказанную поддержку.

7 Список литературы:

[1] Sample records for *ureaplasma parvum serovar*, <https://worldwidescience.org/topicpages/u/ureaplasma+parvum+serovars.html>

[2] Proteomes - *Ureaplasma parvum serovar 3* (strain ATCC 700970), <https://www.uniprot.org/proteomes/UP000000423>

