

Обзор протеома бактерии *Pasteurella multocida* subsp. *multocida* str. Pm70

Школиков А. С.

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

Ключевые слова:

Pasteurella multocida | Квазиопероны

В данной работе приведен статистический анализ данных по геному и протеому бактерии *Pasteurella multocida* subsp. *multocida* str. Pm70, проведенный с использованием ПО MS Excel. Было исследовано распределение генов по цепям хромосомы бактерии, произведен поиск потенциальных оперонов, рассмотрены пересечения генов и закономерности, связанные с ними, проведен анализ доли ферментов различных классов в протеоме *P. multocida*.

Введение

Pasteurella multocida – грамотрицательная бактерия, факультативно анаэробная, патогенная для многих птиц и млекопитающих^[1]. При контакте с животными способна передаваться и человеку, чаще всего через укусы^[2]. Попадая в кровь, вызывает обширное воспаление в районе места проникновения, в отсутствие лечения, приводящее к таким осложнениям, как остеомиелиты^[1]. В отдельных случаях способна вызывать менингиты^[3]. Чувствительна к большинству антибиотиков пенициллинового и тетрациклинового ряда^[1].

Однако наибольший вред деятельности человека бактерия приносит, вызывая болезни домашних животных. У птиц вызывает т.н. птичью холеру^[1], у свиней – атрофический ринит^[4]. К факторам вирулентности относятся гены, кодирующие различные белки наружных мембран, белки, регулирующие поглощение железа, метаболизм тиамина и клеточную

и ведутся активные работы по изучению механизмов, позволяющих бактерии поражать организм хозяина^[1].

Материалы и методы

Основными источниками материалов для работы послужили базы NCBI и Uniprot. Обработка данных проводилась с использованием Microsoft Excel 2010 и скриптов на Python 3.7.

Результаты и обсуждения:

Общие сведения о геноме

Исходя из материалов базы данных RefSeq^[2], геном данного штамма представлен одной кольцевой хромосомой и содержит 2113 гена, т.е. около 893 генов на 1 млн. п.н., 2017 из которых кодируют белок, 79 – некодирующие РНК, включая 56 транспортных, 19 рибосомальных 1 транспортно-матричную РНК и один рибозим (Рибонуклеаза Р). 17 участков являются псевдогенами. В протеоме бактерии представлено, как минимум, 110 белков, ассоциированных с мембраной, включая 82 пермеазы. Помимо пермеаз есть еще, по меньшей мере, 172 транспортных белка. 124 транспортных белка, в т.ч. 52 пермеазы, относятся к семейству ABC-транспортёров (*от англ. ATP-binding cassette*). Белков, входящих в состав рибосом либо регулирующих их функционирование не менее

83. Из белков бактериального иммунитета удалось обнаружить 9 CRISPR-ассоциированных белков и, как минимум, один (*CorA*), связь которого с CRISPR-системой является дискуссионной^[7].

Гипотетических белков в таблице обнаружено 207, то есть 9,8% от общего числа.

Для поиска белков по файлу, помимо встроенной в Excel функции поиска, был применен скрипт на Python 3.7, позволяющий искать комбинации слов, находящихся в ячейках таблицы на расстоянии друг от друга и в случайном порядке.

Также белки были рассмотрены по классам, основываясь на КФ шифре, взятом из таблицы по протеому исследуемого штамма из базы Uniprot.^[2] Если белок выполнял несколько функций, каждая считалась по отдельности. Для этой цели был использован скрипт на Python 3.7. По полученным данным преобладают трансферазы (более 37%), меньше всего транслоказ (менее 2%). Было проведено дополнительное сравнение с протеомами модельных бактерий – *E.coli* и *B.subtilis* (рис 2). Сходство с *E.coli* оказалось ожидаемо ближе, т.к. обе бактерии являются грамотрицательными и, к тому же, факультативными анаэробами и паразитами животных. У исследуемой бактерии процент транслоказ более чем в 2 раза превосходит таковой у *E.coli* и в 6 раз - *B.subtilis*. Вероятно, последнее можно объяснить более сложной организацией мембран и наличием периплазмы у грамотрицательных бактерий, тогда как различие с *E.coli* может быть как случайным при столь малой выборке, так и, например, свидетельствовать о более сложной организации переноса через мембрану у *P. multocida*. Данный вопрос может стать темой отдельного исследования, особенно, учитывая факт, что понимание строения мембраны бактерии и транспорта сквозь нее может сыграть роль в повышении эффективности лекарств от вызываемых ей болезней.

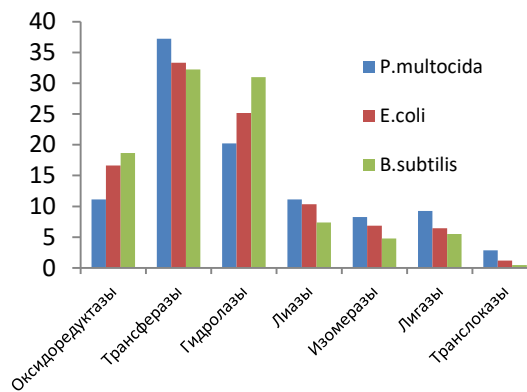


Рис.2 Сравнение процента различных классов белков в протеомах *P. multocida*, *E. coli* и *B. subtilis*

Стоит отметить, что КФ-шифр в использованных материалах был указан далеко не для всех белков, и нельзя утверждать, что соотношения в такой выборке полностью соответствуют таковым во всем протеоме. Возможно, например, что отдельные классы белков исследованы лучше других и потому больший их процент промаркирован.

Длины белков

Было изучено распределение длин генов у исследуемого штамма. Результаты приведены на гистограмме (рис. 1). Наиболее часто длина белка попадала в диапазон от 270 до 340. Среднее значение составило 328.93 аминокислотных остатков (а.о.), минимальное – 26 а.о., максимальное – 2712 а.о., стандартное отклонение – 221.5 а.о., медиана – 285 а.о. Близкие значения средней длины и медианы при достаточно большом стандартном отклонении говорят о разнообразии длин белков у данной бактерии.

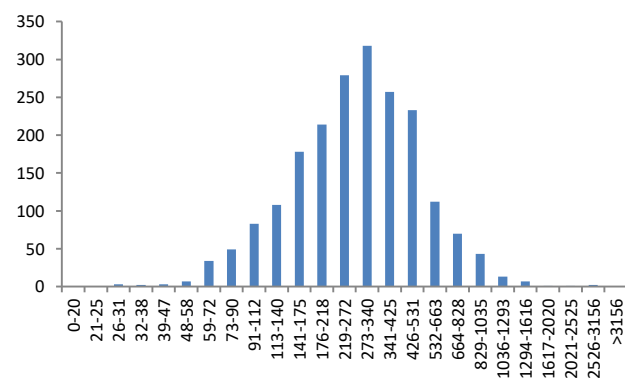


Рис.1 Распределение белков *P. multocida* по длине

Распределение по цепям

Было подсчитано число генов на прямой и комплементарной цепях ДНК (Табл. 1). Кроме того, проверялась гипотеза о случайности расположения генов на прямой и комплементарной цепях. Для этого при помощи функции БИНОМ.РАСП в MS Excel были рассчитаны вероятности ошибки первого рода в случае принятия гипотезы. Уровень значимости был принят равным 0,05. В результате гипотеза была отвергнута. Для отдельных групп генов случайность проверялась отдельно: гипотеза принята для РНК, для белок-кодирующих генов значение близко к уровню значимости, а для псевдогенов гипотеза отвергнута. В то же время, большинство генов тРНК и рРНК как раз собрано в группы, вероятно являющиеся оперонами (видно из таблицы). В данном случае, вероятно, имеет смысл проверка с использованием более сложного статистического аппарата.

Таблица 1. Число генов на цепях ДНК в хромосоме исследуемой бактерии.

	Белок-кодирующие	Псевдогены	РНК-кодирующие	Всего
+ цепь	1044	13	41	1098
- цепь	973	4	38	1015

Пересечения

Было посчитано количество пересечений генов различной длины. Как видно из полученных данных (Рис. 3.1), с ростом длины пересечения частота встречаемости уменьшается, однако для отдельных длин (Рис.3.2) частоты заметно отличаются от таковых для близких по длине перекрытий. Пики наблюдаются на значениях 1, 4, 8, 11, 14, в то время как пересечений длиной в 2, 3, 5, 6 пар оснований встречено не было.

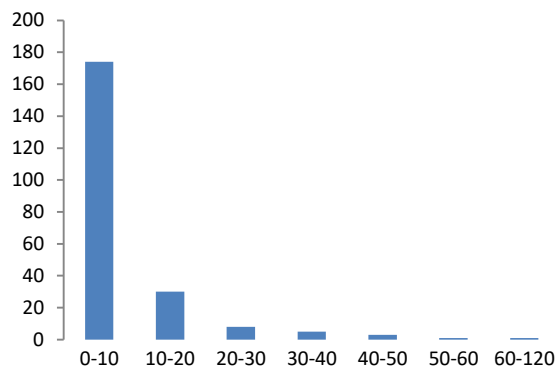


Рис.3.1 Число пересечений длины до 20 пар оснований

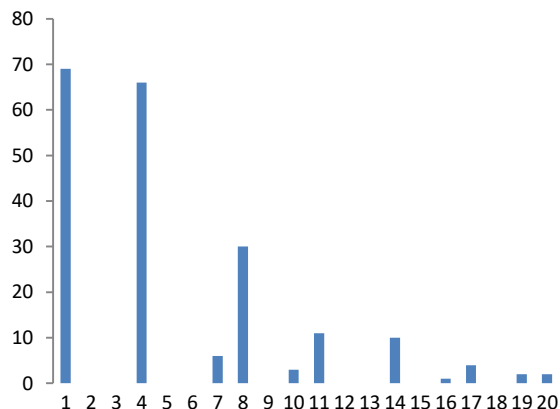


Рис.3.2 Число пересечений длины от 20 пар оснований

Для объяснения полученных результатов была выдвинута гипотеза относительно возможных сайтов перекрытия. Поскольку открытая рамка считывания гена должна оканчиваться стоп-кодом, которых для подавляющего большинства организмов существует три варианта (TAA, TGA, TAG), а начинается, как правило, с кодона ATG, то в случае пересечения в один нуклеотид возможны сочетания TAATG и TGATG, а в случае 4 нуклеотидов ATGA. Это вполне соответствует данным генома исследуемой бактерии, взятым из баз NCBI. Примечательно, что встретилось лишь три пересечения без сдвига рамки считывания, и во всех трех кодирующие последовательности лежат на разных цепях. Это вполне очевидно, ведь в случае пересечения без смещения рамки считывания генов, лежащих на одной цепи, стоп-кодон одного гена будет входить в состав второго, и транскрипция последнего будет обрываться в том же месте. Для пересечений на разных цепях характерны и другие особенности – они встречаются гораздо реже пересечений внутри одной цепи (13 против 209), но их средняя длина существенно больше (23,4 п.н. против

6,6 п.н.), причем максимальное обнаруженное пересечение имеет длину 108 п.н., тогда как из пересечений на одной цепи самое большое – 46 п.н. Объяснение этого факта может послужить темой для нового исследования.

Квазиопероны

В данной работе термином “квазиоперон” называются последовательности генов, закодированных на одной цепи с промежутками между генами не более 100 п.н. Поиск квазиоперонов был проведен с помощью MS Excel. Было обнаружено 485 квазиоперонов. Наибольшая длина – 12 (Рис. 4), квазиопероны такой длины встретились на обеих цепях. Последовательности из единственного гена в данной работе квазиоперонами не считались.

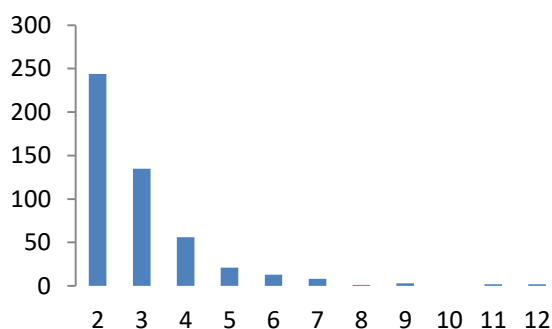


Рис.4 Распределение длин квазиоперонов

Список литературы:

1. Brenda A. Wilson and Mengfei Ho (2013) “Pasteurella multocida: from Zoonosis to Cellular Microbiology” [≥](#)
2. Arons M.S., Fernando L., Polayes I.M. (1982) “Pasteurella multocida – the major cause of hand infections following domestic animal bites.” [≥](#)
3. Eric D.R. Pond, Sameh El-Bailey and Duncan Webster (2015) “An unusual case of meningitis” [≥](#)
4. B. Martineau-Doizé, G. Dumas, R. Larochelle, J. C. Frantz, and G. P. Martineau (1991) “Atrophic rhinitis caused by Pasteurella multocida type D: morphometric analysis.” [≥](#)
5. Jatuponwiphat, Chumnanpuen, Othman, E-Kobon, Vongsangnak. (2018) “Iron-associated protein interaction networks reveal the key functional modules related to survival and virulence of Pasteurella multocida.”
6. Boyce JD1, Adler B. (2000) “The capsule is a virulence determinant in the pathogenesis of Pasteurella multocida”
7. Konstantin V. Severinov, Sergey A. Shmakov, Kira S. Makarova, Yuri I. Wolf and Eugene V. Koonin (2018) “Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis” [≥](#)

Заключение

В ходе исследования, основываясь лишь на материалах открытых баз данных, был проверен ряд гипотез, а также обнаружены явления, требующие дальнейшего изучения. В частности, на мой взгляд, перспективным может быть изучение структур, отвечающих за перенос через мембрану, а также причин большей длины перекрытий генов, лежащих на разных цепях в геноме исследуемой бактерии.

Благодарности

Автор признателен Спирину Сергею Александровичу за помощь в подборе статистических методов для проверки гипотез.

[Ссылка на сопроводительные материалы](#)