

Обзор протеома бактерии *Burkholderia cenocepacia* J2315

Иванова Антонина

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

РЕЗЮМЕ

Данная работа посвящена исследованию протеома бактерии *Burkholderia cenocepacia* J2315. В её ходе рассмотрены особенности распределения длин белков, закономерности распределения белок-кодирующих и РНК-кодирующих генов в зависимости от направленности той цепочки ДНК, на которой они находятся. Проверена гипотеза случайного распределения генов на прямой и обратной цепях.

1 ВВЕДЕНИЕ

Burkholderia cenocepacia — вид грамотрицательных аэробных бактерий, распространенных повсеместно, но чаще всего встречающихся в почве. Эти бактерии патогенны для людей, больных муковисцидозом и могут вызвать серьезные осложнения этой болезни. Инфицирование *Burkholderia cenocepacia* особенно проблематично, поскольку этот организм имеет высокий уровень устойчивости к антибиотикам, что делает его трудно искоренимым. В результате хронической инфекции может возникнуть тяжелое снижение функции легких. В связи с увеличением смертности, от больного муковисцидозом, являющегося членом эпидемии ET12 линии, которая возникла в Канаде или Великобритании и распространилась на Европу, был выделен штамм J2315 бактерии *Burkholderia cenocepacia*. Геном этого возбудителя инфекции полностью секвенирован и включает в себя три круговые хромосомы и плазмиду и кодирует широкий спектр функций, характерных для этого метаболически универсального рода. Хотя штаммы *B. cenocepacia* могут быть выделены из почвы и могут быть патогенными и для растений и для человека, J2315 является представителем рода *B. cenocepacia*, изолированным от окружающей среды. Бактерии из этого штамма распространяются между пациентами с муковисцидозом. Сравнительный анализ показал, что около 21% генома J2315 является уникальным по сравнению с другими штаммами *B. cenocepacia*, что говорит о геномной пластичности этого вида.

2 METHODS

В работе использованы данные, полученные с сервера NCBI.

Были взяты файлы NC_011000.ptt, NC_011000.rnt, NC_011001.ptt, NC_011001.rnt, NC_011002.ptt, NC_011002.rnt, NC_011003.ptt и NC_011003.rnt – по 2 файла для каждой хромосомы и плазмиды: .ptt файл с данными по белок-кодирующим генам и .rnt файл с данными по РНК-кодирующим генам. Для обработки информации применялась программа Microsoft Office Excel 2003. Гены, кодирующие РНК и белки, были разделены с помощью приема «Фильтр».

Для подсчета количества CDS и RNA генов на каждой из цепей применялась логическая функция «СЧЁТЕСЛИ».

Для сортировки белков по длинам и создания гистограммы были созданы карманы длин, принадлежность к которым определялась с помощью вышеупомянутой функции.

Для определения кратности генов трем был заведен столбец, в котором функцией «ОСТАТ» определялись остатки от деления соответствующих генов на 3, а затем подсчитывались ячейки, значение в которых не равнялось нулю.

Проверка гипотезы случайности распределения генов на цепях была проведена с помощью критерия согласия Пирсона по формуле:

$$\chi^2_{empir} = \sum_{i=1}^N \frac{(Empir_i - Teor_i)^2}{Teor_i}$$

где $Empir_i$ – предварительно вычисленное наблюдаемое отношение числа генов в цепи к общему числу генов; $Teor_i$ – предполагаемое отношение (0,5).

3 RESULTS

3.1 Распределение генов по цепям

Геном *Burkholderia cenocepacia* J2315. содержит 7116 генов, кодирующих белки, и 113 генов, кодирующих РНК. Их распределение по прямой и обратной цепям отражено в таблице 1.

Таблица 1. Распределение генов белков и РНК по прямой и обратной цепям в протеоме бактерии *Burkholderia cenocepacia* J2315.

| Цепь/ тип гена | CDS | RNA | Всего |
|----------------|------|-----|-------|
| (+) strand | 3590 | 55 | 3645 |
| (-) strand | 3526 | 58 | 3584 |
| Всего | 7116 | 113 | 7229 |

При проверке гипотезы о случайном распределении генов между цепями с вероятностью 0,5 было получено значение критерия $\chi^2=0,0001$. Критическое значение при уровне значимости 0,05 гораздо выше - 3,84. Следовательно, при уровне достоверности 0,05 гипотезу следует считать верной. Некоторые статистические данные для генома бактерии приведены в таблице 2.

Таблица 2. Статистические данные для генома бактерии *Burkholderia cenocepacia* J2315

| | |
|-------------------------|--------|
| Максимальная длина гена | 4557 |
| Минимальная длина гена | 24 |
| Медиана длин генов | 291 |
| Средняя длина гена | 322,85 |

Также было обнаружено, что в геноме бактерии встречаются 94 кодирующие последовательности, длины которых не делятся на три. Однако такие последовательности обнаружены только для генов, кодирующих РНК.

3.2 Распределение длин белков

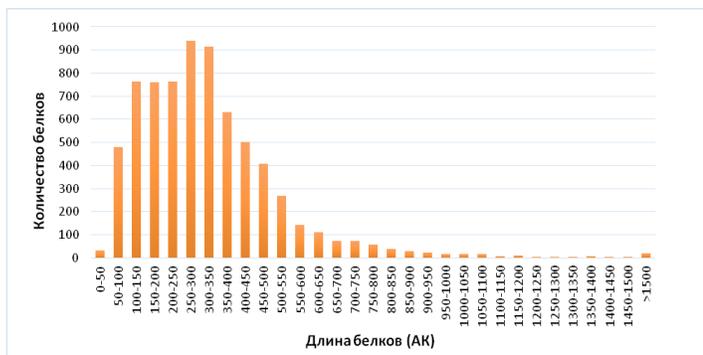
В геноме *Burkholderia cenocepacia* J2315 закодировано 7116 белков. Статистические данные о белках бактерии представлены в таблице 3.

Таблица 3. Статистические данные для белков бактерии *Burkholderia cenocepacia* J2315

| | |
|--------------------------|--------|
| Максимальная длина белка | 4557 |
| Минимальная длина белка | 24 |
| Медиана длин белков | 293 |
| Средняя длина белка | 322,84 |

Распределение белков по длине отражено на гистограмме (рис.1).

Рисунок 1. Распределение длин белков бактерии *Burkholderia cenocepacia* J2315



3.3 Плазмиды рода *Burkholderia*

К роду *Burkholderia* принадлежит 61 плазмида, 2 из которых относится к виду *Burkholderia cenocepacia* и лишь одна – к штамму J2315. Их длина лежит в пределах от 1641 п.н. до 2555070 п.н. Плазмида штамма J2315 содержит в себе 86 генов, кодирующих белки. Ее длина – 92661 п.н.

4 ОБСУЖДЕНИЕ

В целом у *Burkholderia cenocepacia* J2315 количество генов на прямой и обратной цепях различается незначительно. То же можно сказать и о количестве пересечений и квазиоперонов на соответствующих цепях.

Полученные в ходе работы данные позволяют с высокой вероятностью утверждать, что гены распределены по цепям случайно.

В процессе работы было обнаружено, что некоторые последовательности, кодирующие РНК, не кратны трем. Это можно объяснить особенностями синтеза РНК, для которого обязательна триплетность.

Длины белков на гистограмме не образуют симметричного купола, наблюдается перевес в сторону более коротких белков.

Наибольшее количество белков имеют длину в 250 - 300 а.к. Это достаточно хорошо согласуется с представлением о том, что для бактерий не характерны длинные полипептидные цепи.

Таким образом, *Burkholderia cenocepacia* J2315 проявляет типичные для прокариот признаки.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Таблицы и расчеты:

<http://kodom.fbb.msu.ru/~antonio/term1/term1.html>

ССЫЛКИ

1. Matthew T. G. Holden, Helena M. B. Seth-Smith, Lisa C. Crossman, Mohammed Sebahia, Stephen D. Bentley, Ana M. Cerdeño-Tárraga, Nicholas R. Thomson, Nathalie Bason, Michael A. Quail, Sarah Sharp, Inna Cherevach, Carol Churcher, Ian Goodhead, Heidi Hauser, Nancy Holroyd, Karen Munga, Paul Scott, Danielle Walker, Brian White, He-len Rose, Pernille Iversen, Dalila Mil-Homens, Eduardo P. C. Rocha, Arsenio M. Fialho, Adam Baldwin, Christopher Dowson, Bart G. Bar-rell, John R. Govan, Peter Vandamme, C. Anthony Hart, Eshwar Ma-henthiringam, and Julian Parkhill, The genome of *Burkholderia cenocepacia* J2315, an epidemic pathogen of cystic fibrosis patients, *J Bacteriol.* 2009 January; 191(1): 261–277.
2. NCBI, *Burkholderia cenocepacia* J2315, complete genome, organism overview <http://www.ncbi.nlm.nih.gov/genome/?term=Burkholderia+cenocepacia+J2315>