

Обзор протеома бактерии *Clostridium perfringens* ATCC 13124

Карпухина Анна

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

РЕЗЮМЕ

Данная работа посвящена исследованию протеома бактерии *Clostridium perfringens* ATCC 13124. Представлен анализ распределения генов на прямой и обратной цепях ДНК, изучены закономерности перекрывания генов и характер их объединения в квазиопероны. Проверена гипотеза случайного распределения генов на прямой и обратной цепях.

1 ВВЕДЕНИЕ

Clostridium perfringens ATCC 13124 - грамположительная, анаэробная, спорообразующая бактерия рода клостридий, хемоорганогетеротроф. Является возбудителем пищевых токсинфекций человека и одним из возбудителей газовой гангрены. В природе может быть обнаружена в разлагающейся растительности, почве, морских отложениях, желудочно-кишечном тракте позвоночных. *Clostridium perfringens* ATCC 13124 находит широкое применение в качестве удобного объекта контроля в различных исследованиях, используется для синтеза гемолизина, фосфолипаз С, некоторых токсинов.

Геном штамма ATCC 13124 представлен одной кольцевой молекулой ДНК, полностью секвенирован. Настоящая работа имела своей целью выявление особенностей распределения генов на прямой и обратной цепях, изучение степени перекрывания генов и их объединения в квазиопероны.

2 МЕТОДЫ

В работе использованы данные, полученные с сервера NCBI. Были взяты файлы CP000246.ptt и CP000246.rnt. Для обработки информации применялась программа Microsoft Office Excel 2013.

Гены (+) цепи и (-) цепи 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 выявления квазиоперонов применялась функция «ЕСЛИ», с помощью которой в специальном столбце получали значение «0», в случае, когда расстояние между генами было больше критического, или числовые значения, суммирующиеся,

начиная от единицы, в случае, когда расстояние между генами попадало в заданный диапазон. После этого функцией «СЧЁТЕСЛИ» проводился подсчет значений «0», число которых совпадает с количеством квазиоперонов.

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.

Поиск перекрываний генов осуществлялся с помощью функции «СЧЁТЕСЛИ», которая считала все ячейки с отрицательными значениями в столбце расстояний.

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

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

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

3 РЕЗУЛЬТАТЫ

Распределение генов по цепям

Геном *Clostridium perfringens* ATCC 13124 содержит 2876 генов, кодирующих белки, и 117 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 *Clostridium perfringens* ATCC 13124

Цепь/ тип гена	CDS	RNA	Всего
(+) strand	1326	35	1361
(-) strand	1550	82	1632
Всего	2876	117	2993

При проверке гипотезы о случайном распределении генов между цепями с вероятностью 0,5 было получено значение критерия $\chi^2=0,0083$. 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 *Clostridium perfringens* ATCC 13124

Максимальная длина гена, пн	7917
Минимальная длина гена, пн	74
Медиана длин генов	783
Средняя длина гена	919,5

В геноме *Clostridium perfringens* ATCC 13124 встречаются перекрытия генов. Общая частота перекрытий составляет приблизительно 0,0461. При этом частота перекрытий на прямой цепи незначительно превосходит частоту перекрытий на обратной (см. таблицу 3).

Таблица 3. Перекрытия генов в геноме бактерии *Clostridium perfringens* ATCC 13124

	Число перекрытий	Частота
(+) strand	71	0,0522
(-) strand	67	0,0411
Всего	138	0,0461

Также было обнаружено, что в геноме бактерии встречаются 83 кодирующие последовательности, длины которых не делятся на три. Однако такие последовательности обнаружены только для генов, кодирующих РНК.

Распределение длин белков

В геноме *Clostridium perfringens* ATCC 13124 закодировано 2876 белков. Статистические данные о белках бактерии представлены в таблице 4.

Таблица 4. Статистические данные для белков бактерии *Clostridium perfringens* ATCC 13124

Максимальная длина белка, ак	2638
Минимальная длина белка, ак	30
Медиана длин белков	265
Средняя длина белка	312,9

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

Рисунок 1. Распределение длин белков бактерии *Clostridium perfringens* ATCC 13124



Квазиопероны

Квазиоперон – совокупность генов одной цепи, расстояние между которыми не превышает некоторой пороговой величины. В зависимости от выбора данной величины количество квазиоперонов может меняться. В таблице 5 показано количество квазиоперонов *Clostridium perfringens* ATCC 13124 при разных критических значениях.

Таблица 5. Количество квазиоперонов *Clostridium perfringens* ATCC 13124 при разных пороговых расстояниях

Пороговое расстояние	50 bp	100 bp	200 bp
(+) strand	997	879	693
(-) strand	1091	934	671
Всего	2088	1813	1364

Видно, что при увеличении порогового расстояния количество квазиоперонов уменьшается. Распределение квазиоперонов между прямой и обратной цепями приблизительно одинаково.

4 ОБСУЖДЕНИЕ

В целом у *Clostridium perfringens* ATCC 13124 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.

Относительно большое количество квазиоперонов даёт возможность предположить, что многие биохимические процессы у *Clostridium perfringens* ATCC 13124 осуществляются группой белков. Перекрытия, возможно, объясняются тем, что соответствующие гены находятся в разных рамках считывания.

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

Длины белков на гистограмме не образуют симметричного купола, наблюдается перевес в сторону более коротких белков. Наибольшее количество белков имеют длину в 200-250 а.к. Это достаточно хорошо согласуется с представлением о том, что для бактерий не характерны длинные полипептидные цепи. Таким образом, *Clostridium perfringens* ATCC 13124 проявляет типичные для прокариот признаки.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Таблицы и расчёты:

http://kodomofbb.msu.ru/~ann_karpukhina/term1/excel.html

ССЫЛКИ

1. NCBI, Genomes
ftp://ftp.ncbi.nlm.nih.gov/genomes/archive/old_refseq/Bacteria/Clostridium_perfringens_ATCC_13124_uid57901/
2. NCBI, *Clostridium perfringens*
[http://www.ncbi.nlm.nih.gov/genome/?term=txid1502\[orgn\]](http://www.ncbi.nlm.nih.gov/genome/?term=txid1502[orgn])
3. Wikipedia, *Clostridium perfringens*
https://ru.wikipedia.org/wiki/Clostridium_perfringens