

Обзор генома и протеома бактерии *Neisseria meningitidis* MC58

Карань Анна¹

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

Резюме

Данная работа посвящена анализу протеома бактерии *Neisseria meningitidis* MC58, в частности, изучению распределения длин белков, характеру распределения генов между прямой и обратной цепями, степени объединения генов в квазиопероны и особенностям пересечения генов

1. Введение

Neisseria meningitidis MC58 – бактерия, относящаяся к роду *Neisseria*, который включает 25 видов. У всего рода известно 29 плазмид. Этот вид, один из наиболее известных патогенных представителей данного рода, является возбудителем менингококковой инфекции. В связи с этим собрано достаточно много информации о строении, биохимии, патогенезе, и в том числе об устройстве генома и генной регуляции, вида *Neisseria meningitidis*.

Классификация исследуемой бактерии:

Домен:	Bacteria
Филум:	Proteobacteria
Класс:	Betaproteobacteria
Порядок:	Neisseriales
Семейство:	Neisseriaceae
Род:	<i>Neisseria</i>
Вид:	<i>N. meningitidis</i>
Штамм:	<i>N. meningitidis</i> MC58

Данный вид содержит 1 хромосому и 4 плазмиды: pJS-B, pJS-A и две без названия, однако у штамма MC58 плазмид нет или они не секвенированы, так что, к сожалению, они не проанализированы в данной работе. В хромосоме 1953 гена, кодирующего белки, и 71 - тРНК и рРНК.

Целью данной работы является исследование особенностей распределения длин белков и описание особенностей локализации генов.

Также целью являлось овладение навыками работы с Microsoft Office Excel 2007 и обучение основным правилам написания статей и обзоров.

2. Методы

В работе были использованы файлы с последовательностями и таблицами локализации генов, скачанные из базы данных National Center for Biotechnology Information. NC_003112 - идентификатор единственной хромосомы. Обработка и анализа информации была использована программа Microsoft Office Excel 2007.

Из записей NC_003112.ptt, NC_003112.rnt данные были импортированы в файл Excel, в качестве разделителя был использован пробел, а последние ячейки потом объединялись. На новом листе были созданы карманы длин от 1 до 2800 с шагом 100, и с помощью функции СЧЕТЕСЛИМН было посчитано число белков в каждом диапазоне и составлена гистограмма.

С помощью функции СЧЕТЕСЛИ была составлена таблица числа генов тРНК/рРНК и белков на + и - цепях.

Функция ЕСЛИ использовалась для подсчета числа пересекающихся генов.

А с использованием БИНОМРАСП были определены Р-значения для подсчета статистической вероятности нахождения генов на + и - цепях. (уровень доверия был выбран 0,05)

Число квазиоперонов и пересечений тоже подсчитывалось с помощью функции ЕСЛИ.

3. Результаты

3.1 Распределение длин белков

Диаграмма 1. Распределение длин белков в протеоме *Neisseria meningitidis* MC58. По горизонтали указаны диапазоны длины белков, а по вертикали количество белков в соответствующем диапазоне.



Для исследуемой бактерии наиболее характерны белки длиной 100-200 нуклеотидов – 503 белка из 1953 (см. Диаграмму 1). В следующих диапазонах количество белков равномерно уменьшается, для бактерии характерны маленькие длины, так что белков до 500 87% (1696 из 1953). Максимальная длина белка – 2703 аминокислоты. Минимальная же – 22. (см. таблицу 1)

Таблица 1. Длина самого короткого и самого длинного белка

Локализация гена	Продукт	Длина белка	Категория
509391- 517502	Гемагглютинин. /гемолизин- подобный белок	2703	Длинный
1538830- 1538898	Гипотетический белок	22	Короткий

3.2 Распределение генов, кодирующих белки и РНК, между прямой и обратной цепью

Таблица 2. Распределение генов между прямой и обратной цепью

Конечный продукт	Прямая цепь	Обратная цепь
Белок	1024	929
РНК	40	31

Число генов на прямой цепи у данной бактерии больше, чем на обратной и для генов белков, и для генов РНК. (см. Таблицу 2) Однако, обычно предполагают распределение по цепям с вероятностью 0,5. Это гипотезу можно проверить на данных статистических данных, посчитав Р-значение биномиального распределения.

Таблица 3. Р-значения биномиального распределения между цепями для генов, кодирующих белки и РНК.

Конечный продукт	Р-значение
Белок	0,016695
РНК	0,171235

Из таблицы 3 видно, что гипотеза о случайном распределении генов подтверждается не полностью (если брать уровень доверия равный 0,05). Для белков она не верна, а как раз для них у нас больше данных, и можно было бы говорить о статистической достоверности, в отличие от РНК. Поэтому на этих данных гипотеза не подтверждается.

3.3 Число квазиоперонов

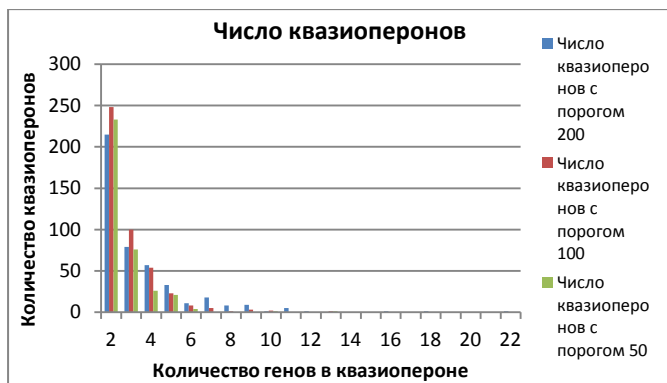
Таблица 4. Число квазиоперонов на прямой и обратной цепи, в зависимости от порогового расстояния

Цепь	200 п.н.	100 п.н.	50 п.н.
+ цепь	487	620	734
- цепь	452	556	702
всего	939	1186	1436

Под квазиопероном понимается такая совокупность генов, расположенных на одной цепи, что расстояние между любыми соседними генами внутри квазиопераона меньше некоторого порогового значения. В зависимости от выбора этого порогового расстояния получается различное число квазиоперонов. Из таблицы 4 видно, что с увеличением порогового расстояния число квазиоперонов уменьшается, то есть генов, объединенных в квазиоперон с какими-либо другими генами, становится больше.

Самый длинный квазиоперон с пороговым значением 50 и 100 данной бактерии состоит из 13 генов, а с 200 состоит из 22 генов.

Диаграмма 2. Количество квазиоперонов различной длины в зависимости от порогового значения.



Наибольшее число квазиоперонов маленьких длин, можно обнаружить, задавая пороговое значение 100, так как выявляется уже достаточно много близких генов (по сравнению с 50), но они еще не объединяются в крупные кластеры (как при 200). Наиболее крупные кластеры, как уже

было сказано, возникают при пороговом значении 200, однако они уже могут не иметь важного функционального значения.

Таблица 5. Р-значения для распределения квазиоперонов между цепями

	200 п.н.	100 п.н.	50 п.н.
Р-значение	0,133592	0,061886	0,206666

Из таблицы 5 видно, что квазиопероны распределены между цепями случайно.

3.4 Пересечения генов

Таблица 6. Число пересечений генов на прямой и обратной цепи

Цепь	Число пересечений	Р-значение	Частота пересечения генов
+ цепь	97	0,03871	0,091165
- цепь	73		0,076042
всего	170		0,083992

Таблица 6 еще раз показывает нам случайное распределение между цепями, в данном случае это пересечения генов.

4. Обсуждение

Для бактерий типично распределение белков по длинам, какое получилось в представленном исследовании. В основном представлены белки длины 100-200 аминокислот, и больше 75% белков длиной меньше 500. Также типичным признаком прокариот является высокая частота перекрытия генов, чуть меньше 10%.

Для данной бактерии характерны достаточно крупные квазиопероны, а так как обычно в одном опероне объединяются гены белков, участвующих в одном биохимическом процессе, то можно предположить наличие сложных биохимических путей у *Neisseria meningitidis* MC58. Также было обнаружено, что длины не всех генов нацело делятся на 3. Но все эти гены, конечным продуктом которых

является РНК, а значит у них не должно быть триплетной структуры.

На примере числа квазиоперонов, числа перекрываний, было показано случайное распределение между прямой и обратной цепью, однако на основе данных по распределению генов и характеру перекрываний подтвердить гипотезу не удалось.

5. Сопроводительные материалы

<http://kodomo.fbb.msu.ru/~annakaran/term1/block4/excel14.xlsx>

6. Ссылки

- ftp://ftp.ncbi.nlm.nih.gov/genomes/archive/old_refseq/Bacteria/Neisseria_meningitidis_MC58_uid57817/

