

Обзор протеома бактерии *Streptomyces avermitilis* MA-4680

Никишова А.В.

Факультет биоинженерии и биоинформатики МГУ им. Ломоносова

РЕЗЮМЕ

В данной работе представлен мини-обзор протеома бактерии *Streptomyces avermitilis* MA-4680 = NBRC 14893. Целью данного обзора было выявить частоту встречаемости белков различных длин, и определить наиболее распространённую длину. Также было исследовано распределение генов на прямой и комплементарной цепи.

1 ВВЕДЕНИЕ

Протеом – совокупность всех белков, производимых организмом. В данном исследовании я работала с бактерией, а протеом бактерий относительно невелик. Стрептомицеты примечательны тем, что они содержат не кольцевые, а линейные хромосомы (очень большие по размеру) с различными теломерами. Благодаря выработке авермектина Стрептомицеты активно используются в промышленности для производства антибиотиков, противогрибковых, противовирусных, и т. д.[1] Для понимания механизма выработки авермектина, было секвенировано несколько штаммов данной бактерии, протеом одной из которых и будет рассмотрен в данной работе. Исследование проводилось с помощью биоинформатических методов, что существенно упростило поставленную задачу.

2 МАТЕРИАЛЫ И МЕТОДЫ

В данной работе я использовала данные о протеоме бактерии *Streptomyces avermitilis* MA-4680 = NBRC 14893, полученные из базы данных National Center for Biotechnology Information [1]. Исследования проводились в программе Microsoft® Office Excel 2010.

Файл “GCF_000009765.2_ASM976v2_feature_table (1).txt.gz” был импортирован в файл Excel на лист “table”. Затем файл был превращен в плоскую таблицу (удалены строки «gene», поскольку они не информативны для прокариот, удалены пустые столбцы). Длины белков были скопированы на лист “bar chart” (для выделения длин белков использовался фильтр по столбцу “ # feature”). Далее были построены карманы от 0 до 2000 с шагом в 50 аминокислот и посчитаны частоты встречаемости длин белков (подсчет выполнялся при помощи функции СЧЁТЕСЛИМН) для каждого кармана. По частотам была построена гистограмма встречаемости различных длин белков. Также были посчитаны максимальные и минимальные значения длин, медиана и среднее.

На листе “numbers” мной была создана таблица, в которой подсчитано количество белков и РНК (сначала по категориям, потом суммированы) на прямой и комплементарной цепях. Для подсчета использовалась функция СЧЁТЕСЛИМН.

На листе “genes” создана таблица, в которой указана общая сумма всех белков и РНК по категориям. Также на этом листе выполнен подсчет количества генов на 1 млн. пар нуклеотидов.

Для проверки гипотезы о случайном распределении генов по цепочкам с вероятностью 0,5 была проведена «симуляция бросания монетки», для чего формула СЛУЧМЕЖДУ(0,1) была распространена на сто столбцов и на 7770 (по числу генов бактерии) строк, а затем подсчитаны отклонения для каждого случая и проведено сравнение с наблюдаемым отклонением.

3 РЕЗУЛЬТАТЫ

Данная бактерия содержит 7681 хромосомных генов, при этом на 1 млн п. н. приходится приблизительно 851 ген. В табл.1 можно увидеть количество генов белков и РНК по категориям и в сумме.

Табл.1 Количество генов белков и РНК

proteins	ncRNA	rRNA	tmRNA	tRNA	RNAs
7677	2	18	1	72	93

3.1 Распределение длин белков

В результате моего исследования было выявлено, что большее количество белков имеют длину от 100 до 500 аминокислотных остатков. При дальнейшем увеличении длины частота встречаемости белков постепенно уменьшается. Эти выводы наглядно иллюстрирует гистограмма (рис.1).

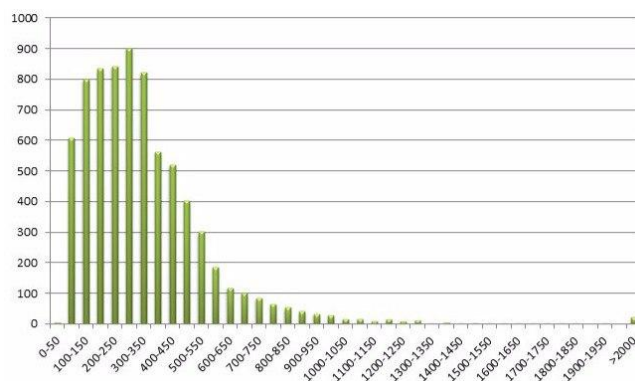


Рис.1 Гистограмма длин белков

Средняя длина белка составляет 336 АО, а медиана – 285АО. При этом длины белков имеют большой разброс: минимальное значение составляет 29 АО, а максимальное – 7746 АО.

3.2 Распределение генов по цепям

В ходе моего исследования распределения генов по цепям было выявлено, что количество генов на прямой и комплементарной цепи немного отличается. При этом количество РНК на прямой и обратной цепи практически равно, а количество белков на обратной цепи несколько превышает количество на прямой. Отклонение для белков составляет 64,5. Распределение генов белков и РНК по цепям приведено в табл. 1.

Табл.2 Число РНК и белков на прямой и комплементарной цепи

strand	proteins	ncRNA	rRNA	tmRNA	tRNA	RNAs
+	3774	1	12	1	33	47
-	3903	1	6	0	39	46

Проверка гипотезы о случайном распределении генов дала положительные результаты для генов РНК (вероятность того, что случайное отклонение будет превышать наблюдаемое, равна 1). Для белков вероятность невысокая (0,18), но вполне приемлемая. Результаты проверки говорят о том, что распределение генов по цепям случайное.

ОБСУЖДЕНИЕ

Гены организмов кодируют белки и РНК, при этом белков, как можно убедиться на примере нашей бактерии, гораздо больше. Среди РНК преобладает число транспортных РНК, а некодирующих и транспортно-матричных может быть всего одна-две.

Сравнивая количества генов белков и РНК на прямой и комплементарной цепи, можно сделать вывод, что, скорее всего, и белки, и РНК распределены по цепям случайно с вероятностью 0,5.

У бактерии данного штамма встречаются белки длиной от 29 АО до 7746 АО, при этом наибольшее количество белков имеют длину от 100 до 500 а. о., а белков с длиной больше 1000 а. о. очень мало. Возможно, это происходит из-за того, что для синтеза слишком длинных белков требуется много ресурсов, а функциональность таких белков ненамного превышает функциональность средних по длине. Таким образом, можно сделать вывод, что диапазон длин от 100 до 500 является наиболее выгодным.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

[Excel-файл](#), в котором можно найти гистограмму длин белков и таблицу распределения генов по цепям, а также сам протеом бактерии.

БЛАГОДАРНОСТИ

Выражаю благодарность преподавателю информатики А. В. Алексеевскому.

СПИСОК ЛИТЕРАТУРЫ

[1]NCBI:Streptomyces avermitilis <https://www.ncbi.nlm.nih.gov/genome/1040>