Мини - обзор

Исследование протеома бактерии *Deinococcus maricopensis*

Колупаева Анна

МГУ им.Ломоносова, Москва.

Завершён 19.02.15

РЕЗЮМЕ Исследовался геном бактерии *Deinococcus maricopensis* из рода *Deinococcus*. Были проанализированы таблица распределения длин белков, синтезируемых бактерией, в зависимости от их длины и таблица распределения генов по комплементарным цепям. Выяснилось, что наиболее часто встречаются белки длиной 200-300 аминокислот. Подтвердилась гипотеза о том, что гены распределяются по комплементарным цепям с равной вероятностью.

1 ВВЕДЕНИЕ

Бактерия *Deinococcus maricopensis*, представитель рода *Deinococcus*, является аэробом, питается гетеро- и хемотрофно. Обитает на простых питательных почвах, специфических условий местообитания не наблюдалось. Геном бактерии представлен одной кольцевой ДНК, общая длина генома 3 498 530 п.н.

Первая задача - определить, какая длина наиболее распространена среди генов, кодирующих белки. Для этого необходимо посчитать количество белков, входящих по длине в определённый диапазон, и найти наиболее часто встречающуюся длину. Затем умножить границы диапазона на 3, поскольку длина каждого гена в 3 раза больше длины кодируемого белка.

Вторая задача - наити количество генов, располагающихся на прамой и обратной цепях и проверить гипотезу, что они распределяются в соотношении 1:1 (т. е. вероятность нахождения гена на каждой цепи равна 0,5).

2 МАТЕРИАЛЫ И МЕТОДЫ

Я проводила исследование по данным о геноме бактерии Deinococcus maricopensis с сайта организации NCBI [1]. В частности я пользовалась таблицами с информацией о генах, кодирующих белки (иРНК) и генах, кодирующих тРНК и рРНК.

Все результаты были получены с помощью программы Microsoft Excel. Я использовала функцию CONCATENATE для создания диапазонов длин генов, функцию COUNTIFS – для подсчёта количества генов, обладающих заданным свойством (обладающий длиной из диапазона, находящийся на прамой или обратной цепи), а также функции MAX и MIN для вывода максимального и минимального значений длин белков.

3 РЕЗУЛЬТАТЫ

3.1 Максимальная длина белка, встречающаяся в геноме бактерии Deinococcus maricopensis – 3180 аминокислотных остатков (а.о.), а минимальная – 30 а.о. Изначально был задан интервалы длин белков от 0 до 3200 с шагом в 100 а.о. Но после получения гистограммы обнаружилось, что после отметки в 1500 а.о. количество белков, входящих в каждый интервал крайне мало. Поэтому все гены белков, имеющие длину более 1500 а.о. были объединены в одну колонку (Рис.1).

В итоге мы можем наблюдать следующую зависимость: найболее часто встречаются белки длины в интервалах 200-300 и 100-200 а.о. Далее линия, соединяющая вершины колонок, описывает убывающую гиперболу, т.е. количество белков длины, входящей в следующие интервалы, резко убывает. Отсюда можно сделать вывод, что наибольшую часть в геноме данной бактерии занимают гены длиной 600-900 п.н.

3.2 Было посчитано количество генов на прямой и на комплементарной ей цепи. Результаты отражены в таблице (Табл.1).

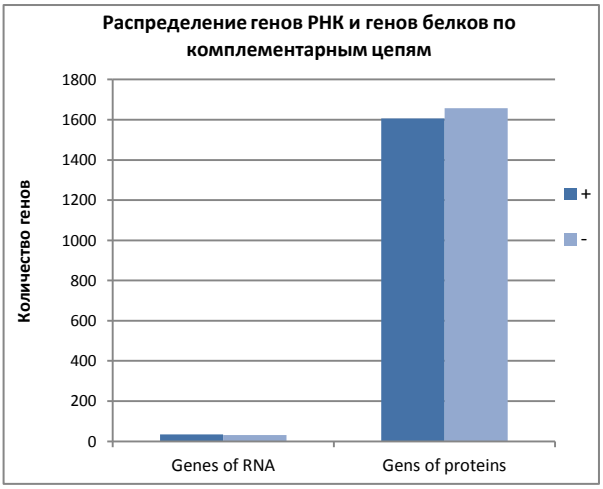
Рис.1 Гистограмма длин белков

В виде гистограммы представлено соотношение результатов отдельно для генов, кодирующих белки, и для кодирующих тРНК и рРНК (Рис.2).

Табл.1 Распределение генов по цепям ДНК

Направление	Гены РНК	Гены белков
+	35	1606
-	31	1658

Рис.2 Соотношение



4 ОБСУЖДЕНИЕ

Анализируя полученные закономерности, я сделала некоторые выводы.

В рамках первой задачи я предположила, что длина белка определяет сложность его функций. В таком случае логично, что в протеоме бактерии преобладают белки, не предназначеннные для выполнения функций, требующих большого количества выполняемых действий. Такие белки могут независимо друг от друга выполнять сравнительно несложные действия, а более сложноустроенные белки могут действовать комплексно для решения некоторых специализированных задач. Поэтому белков большой длины значительно меньше – для приспособления к меняющимся условиям среды гораздо

выгоднее использовать различные комбинации простых функций, чем создавать сложные белки, нацеленные на решение одной конкретной задачи.

После графического представления результатов второй задачи в виде гистограммы (Рис.2) становится очевидно, что гипотеза о равной вероятности нахождения гена на обеих цепях ДНК является верной. Но это можно доказать и с помощью несложных алгебраических вычислений. Если обозначить за *Q* отношение количества генов на прямой цепи к количеству на обратной и посчитать его для РНК и белков отдельно, получим следующие результаты:

$$Q_1 = 35/31 = 1{,}1290\ldots$$

$$Q_2 = 1606/1658 = 0{,}9686\ldots$$

Таким образом, мы наблюдаем, что при расширении области интереса (от РНК к белкам) соотношение генов, находящихся на разных цепях, близко к 1. А следовательно, вероятность расположения любого гена на одной из цепей равна 0,5.

5 СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Файл Excel: Review_of_the_proteom.xlsx

6 ЛИТЕРАТУРА

[1] Сайт NCBI (National Center for Biotechnology Information**):** ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/Deinococcus_maricopensis_DSM_21211_uid62225/