

Обзор протеома бактерии *Mesorhizobium opportunistum*, штамм WSM2075

Поляков Игорь

МГУ, факультет Биоинженерии и Биоинформатики, 1 курс

РЕЗЮМЕ

Данная работа посвящена исследованию протеома бактерии *Mesorhizobium opportunistum*, штамм WSM2075.

Количественные и статистические данные получены с помощью программы Microsoft Excel. Исследовалось распределение длин белков и расположение генов на разных цепях ДНК.

1. ВВЕДЕНИЕ

Бактерия *Mesorhizobium opportunistum* относится к классу Альфа-протеобактерии. К роду бактерии относят порядка 30 видов^[1]. В работе рассматривается штамм WSM2075, его геном полностью секвенирован. Геном этой бактерии состоит из одной кольцевой хромосомы. Всего в геноме 6,746 генов, из них 3337 кодируют белки, а 67 – тРНК и рРНК.

В ходе работы было изучено распределение длин белков, а также подсчитано количество генов на прямой и обратной цепях. Для ДНК в клетке бактерии (хромосома) подсчитано количество квазиоперонов (считалось, что соседние гены входят в один "квазиоперон", если они расположены на одной цепи и расстояние между ними меньше 100 п.н.) и перекрывающихся генов.

Основной целью работы является исследование протеома бактерии *Mesorhizobium opportunistum*, побочной – усовершенствование собственных навыков в программе Microsoft Excel.

2. МЕТОДЫ

Данные о протеоме бактерии были получены на сервере NCBI^[2]. Для их обработки использована программа Microsoft Excel 2007. При создании гистограммы использовался встроенный пакет анализа данных.

Для подсчета количества генов белков и РНК, расположенных на прямой и обратной цепях использовалась функция COUNTIF. Проверка гипотезы о случайном распределении генов на цепях осуществлялась с помощью функции БИНОМРАСП

3. РЕЗУЛЬТАТЫ

3.1. Распределение длин белков

Распределение длин белков представлено в виде гистограммы (рис. 1). Гены, кодирующие РНК, не учитывались.

По рисунку видно, что большинство белков имеет длину от 50 до 600 а.о. Больше всего белков попадает в диапазон 200-400 а.о.

3.2 Распределение генов по прямой и обратной цепям

Данные по количеству генов, расположенных на разных цепях ДНК, представлены в таблице 1.



Рисунок 1. Распределение длин белков в протеоме бактерии *Mesorhizobium opportunistum*. По горизонтали указаны диапазоны длин белков в аминокислотных остатках (а.о.), по вертикали – количество белков с такой длиной.

Таблица 1. Распределение генов, кодирующих белки и РНК, по прямой и обратной цепям

Цепь	Белки	РНК
Прямая	1750	41
Обратная	1587	26
Всего	3337	67

Как видно из таблицы, гены распределены по цепям примерно поровну. Для проверки гипотезы о том, что распределение генов по цепям случайно (с вероятностью 0,5), использовался тест биномиального распределения. По результатам этого теста гипотеза верна и для генов белков, и для генов РНК (p равно 0,998 и 0,975 соответственно).

3.3 Квазиопероны и перекрывающиеся гены

В таблице 2 приведены данные по количеству квазиоперонов и перекрывающихся генов.

Таблица 2. Квазиопероны и перекрывающиеся гены в геноме бактерии *Mesorhizobium opportunistum*.

ДНК	Квазиопероны	Перекрывающиеся гены
Кольцевая хромосома	1706	662

4. ОБСУЖДЕНИЕ

Более 75% белков имеют длину от 50 до 600 а.о. Для бактерий это значение можно считать нормальной средней длиной белков. Менее 1% белков имеет длину более 1300 а.о., из них только у шести длина превышает 2000 а.о.

Гипотеза о том, что распределение генов по цепям случайно, подтвердилась. В файле [Polyakov_pr15_task.xlsx](#) приведены данные о количестве квазиоперонов и перекрывающихся генов, а также их доля от общего количества генов.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

[Polyakov_pr15_task.xlsx](#) – файл с гистограммой, таблицей распределения генов по цепям, информацией по квазиоперонам и перекрывающимся генам

БЛАГОДАРНОСТИ

Я выражаю благодарность А.В. Алексеевскому и Ивану Русинову, которые ведут данный блок в курсе информатики.

СПИСОК ЛИТЕРАТУРЫ

- [1] List of genus *Mesorhizobium*:
<http://www.bacterio.net/mesorhizobium.html>
- [2] NCBI база геномов бактерий:
<ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/>