

Обзор протеома бактерии *Rhodopseudomonas palustris* CGA009

Португальская Дария

МГУ, факультет биоинженерии и биоинформатики, 1 курс

РЕЗЮМЕ

Был проведен анализ протеома бактерии *Rhodopseudomonas palustris*, штамм CGA009. Для получения статистических данных была использована программа Microsoft Excel (версия 2010 года).

Был проведен анализ генома, в результате которого были получены статистические данные по распределению длин белков данной бактерии и расположению генов на прямой и обратной цепях ДНК.

В ходе работы были установлены следующие закономерности:

- 1) Большинство белков имеют длину около 400 а.о.
- 2) Количество генов белков намного больше, чем генов РНК
- 3) Скорее всего гены распределены случайным образом по цепочкам ДНК

Полученные результаты оформлены в виде таблиц (см. Сопроводительные материалы), данные по распределению длин белков представлены в виде гистограммы (см. рисунок 2. Распределение длин белков.).

1 ВВЕДЕНИЕ

Бактерия *Rhodopseudomonas palustris* CGA009 относится к метаболической группе пурпурных фототрофных бактерий. *Rhodopseudomonas palustris* одна из самых известных метаболически универсальных бактерий, она для выживания использует четыре различные модели метаболизма: фотоавтотрофную, фотогетеротрофную, хемоавтотрофную и хемогетеротрофную модели.



1. Фотография колонии бактерии *Rhodopseudomonas palustris* CGA009

Генетическая информация в клетке бактерии представлена кольцевой ДНК, содержащей 4813 белков и 60 рнк, и плазмидой, содержащей 7 белков и не содержащей рнк.

Целью данного задания является освоение программы Microsoft Excel и приобретение навыков анализа данных. Также

это задание предполагает написание структурированного отчета.

2 МЕТОДЫ

Общая информация о бактерии была получена из сервера NCBI [1]. Также из этого сервера было взято описание протеома бактерии [2]. Были взяты файлы: NC_005296.rnt и NC_005296.ptt. Фотография колонии бактерий получена из электронной книги [3].

Обработка данных производилась с помощью программы Microsoft Excel 2010. В процессе создания таблиц были использованы простейшие формулы, функция “СЧЁТЕСЛИМН”, позволяющая рассчитать количество ячеек, удовлетворяющих нескольким условиям, а также функция “СЧЕТЕСЛИ”, сходная с предыдущей (считает количество непустых ячеек, удовлетворяющих единственному условию). Данные функции понадобились для расчёта количества белков в заданном диапазоне и для расчёта количества генов на разных цепях ДНК. Также была использована доступная в Excel функция создания гистограммы.

Все результаты доступны на моей странице на сервере kodoto [4].

3 РЕЗУЛЬТАТЫ

3.1 Распределение длин белков

Распределение длин белков бактерии представлено в виде гистограммы (см. рисунок 2. Распределение длин белков). На горизонтальной оси указаны диапазоны длин (в а.о.), по вертикали указано количество белков. Наибольший процент белков попадает в диапазон 200-400 а.о.



2. Распределение длин белков

Так же стоит отметить, что наблюдается скачкообразный рост в количестве белков на интервале 0 – 400 а.о. и стремительный спад на интервале 400-1200 а.о..

3.2 Распределение генов по прямой обратной цепям

Было подсчитано количество генов белков и генов РНК на прямой и комплементарных цепях.

Распределение генов по прямой и обратной цепям представлено в таблице 1.

	CDS	RNA	Всего
прямая цепь "+"	2400	24	2424
обратная цепь "-"	2413	36	2449
Всего	4813	60	4873

Прежде всего стоит отметить, что генов, кодирующих РНК, намного меньше, чем генов, кодирующих белки.

Количество генов, кодирующих белки, на прямой и на обратной цепи примерно одинаковое количество, а генов РНК на обратной цепи в 1.5 раз больше.

Была проверена гипотеза о том, что распределение генов по прямой и обратной цепям случайно.

Вероятность такого события достаточно высока и для РНК (0,9537695), и для белков (0,57996196).

ОБСУЖДЕНИЕ

Наибольшее количество белков имеет длину 400 а.о..

Практически нет белков длиной менее 200 а.о..

Распределение CDS и РНК по цепям может быть случайным поскольку вероятности случайного распределения велики.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Данные по протеому, гистограмму распределения длин белков и таблицу с расположением генов на цепях можно найти в файле [Portugalskaya_pr15.xlsx](#).

БЛАГОДАРНОСТИ

Выражаю огромную благодарность всем преподавателям факультета биоинженерии и биоинформатики МГУ.

ЛИТЕРАТУРА

- [1] Larimer FW et al (2004): Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodospseudomonas palustris*, *Nat Biotechnol* 22(1):55-61.
- [2]NCBI, полный геном *Rhodospseudomonas palustris*: ftp.ncbi.nlm.nih.gov/genomes/Bacteria/Rhodospseudomonas_palustris_CGA009_uid62901/
- [3] Anh LE-HOANG (2008): Study of *Rhodospseudomonas Palustris* toward Photoautotrophic Growth via the Media, Los Angeles, UCLA
- [4] Моя страница: kodomo.fbb.msu.ru/~dariya.portugalskaya/