

Обзор протеома бактерии *Nostoc sp.*, штамм PCC 7524

Кашко Наталия

МГУ, факультет Биоинженерии и Биоинформатики, 1 курс

РЕЗЮМЕ

Данная работа посвящена исследованию протеома бактерии *Nostoc sp.*, штамм PCC 7524. Количественные и статистические данные получены с помощью программы Microsoft Excel. Исследовалось распределение длин белков и расположение генов на разных цепях ДНК.

1 ВВЕДЕНИЕ

Бактерия *Nostoc sp.* относится к отряду Цианобактерии. К данному роду относят более 20 видов [1]. В работе рассматривается штамм PCC 7524, его геном полностью секвенирован.

Геном этой бактерии состоит из одной кольцевой хромосомы и двух плазмид (pNOS7524.01 и pNOS7524.02). Всего в геноме 5533 гена, из них 5449 кодируют белки, а 84 – тРНК и рРНК.

В ходе работы было изучено распределение длин белков, а также подсчитано количество генов на прямой и обратной цепях. Для трех молекул ДНК в клетке бактерии (хромосома и две плазмиды) 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 *Nostoc sp.*, усовершенствование собственных навыков в программе Microsoft Excel.

2 МЕТОДЫ

Данные о протеоме бактерии *Nostoc sp.* были получены на сервере NCBI [2]. Для их обработки использована программа Microsoft Excel 2013.

При создании гистограммы использовалась функция СЧЁТЕСЛИМН (COUNTIFS) – для подсчета количества белков, длина которых находится в указанном диапазоне.

Для подсчета количества генов белков и РНК, расположенных на прямой и обратной цепях, также использовалась эта функция. Проверка гипотезы о случайном распределении генов на цепях осуществлялась с помощью функции БИНОМ.РАСП (BINOM.DIST).

3 РЕЗУЛЬТАТЫ

3.1 Распределение длин белков

Распределение длин белков представлено в виде гистограммы (рис. 1). Для построения графика были объединены данные по хромосоме и двум плазмидам. Гены, кодирующие тРНК и рРНК, не учитывались.

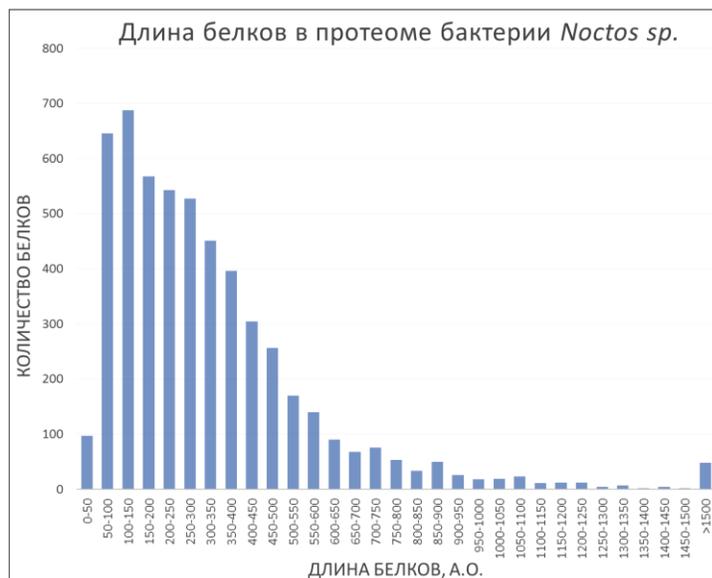


Рисунок 1. Распределение длин белков в протеоме бактерии *Nostoc sp.* По горизонтали указаны диапазоны длин белков в аминокислотных остатках (а.о.), по вертикали – количество белков с такой длиной.

По рисунку видно, что большинство белков имеет длину от 50 до примерно 500 а.о. Больше всего белков попадает в диапазон 100-150 а.о.

3.2 Распределение генов по прямой и обратной цепям

Данные по количеству генов, расположенных на разных цепях ДНК, представлены в таблице 1. Информация приводится для хромосомы и двух плазмид.

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

Цепь	Белки	рРНК, тРНК
Прямая цепь	2841	35
Обратная цепь	2608	49
Всего	5449	84

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 (р равно 0,99 и 0,95 соответственно).

3.3 Квазиопероны и перекрывающиеся гены

В таблице 2 приведены данные по количеству квазиоперонов и перекрывающихся генов для хромосомы, рNOS7524.01 и рNOS7524.02.

Таблица 2. Квазиопероны и перекрывающиеся гены в геноме бактерии *Nostoc sp.*

ДНК	Квазиопероны	Перекрывающиеся гены
Хромосома	1488	261
рNOS7524.01	48	6
рNOS7524.02	5	2
Всего	1541	269

4 ОБСУЖДЕНИЕ

Более 80% белков имеют длину от 50 до 500 а.о. Для бактерий это значение можно считать нормальной средней длиной белков. Менее 1% белков имеет длину более 1500 а.о., из них только у двух длина превышает 10000 а.о., причем эти белки содержат в своем составе повторы.

Гипотеза о том, что распределение генов по цепям случайно, подтвердилась.

В файле [Kashko_dop.xlsx](#) приведены данные о количестве квазиоперонов и перекрывающихся генов, а также их доля от общего количества генов. Для плазмид доля квазиоперонов составляет около 50%, а перекрывающихся генов - 5,77% для рNOS7524.01 и 20% рNOS7524.02. Это может быть объяснено относительно небольшими размерами плазмид, которые, однако, должны выполнять определенную функцию. Поэтому гены в плаزمиде должны располагаться компактно.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

[Nostoc_genes.xlsx](#) - основной файл с гистограммой и таблицей распределения генов по цепям

[Kashko_dop.xlsx](#) - файл с 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] <http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Undef&id=1177&lvl=3&keep=1&srchmode=1&unlock>
- [2] <ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/>