

Обзор протеома бактерии *Thermotoga thermarum* DSM 5069

Яковлева Александра

МГУ имени М.В. Ломоносова, факультет биоинженерии и биоинформатики, 1 курс 2014 года

РЕЗЮМЕ

Motivation: Интерес в исследовании представила из себя бактерия *Thermotoga thermarum* DSM 5069. Задавшись вопросом, какая должна быть длина белка для нормальной жизнедеятельности этой бактерии, и каково оптимальное расположение генов на прямых и комплементарных цепях белков и РНК, был получен ответ. Оптимальная длина белка – 250 аминокислотных остатков.

1 ВВЕДЕНИЕ

Организм *Thermotoga thermarum* DSM 5069 принадлежит к домену Bacteria, его полный геном насчитывает 2039943 п.н. Анализ, описание которого находится в данной статье, был проведен с целью разобраться в протеоме бактерии. Проведено исследование генов белков и генов РНК в организме.

2 МЕТОДЫ И МАТЕРИАЛЫ

Для написания анализа использовалась информация из банка геномов NCBI^[1], где были найдены все необходимые данные о генах белков и генах РНК исследуемой бактерии. Для построения гистограммы и вычисления табличных данных использовалась программа Microsoft Excel 2010, в частности функция COUNTIF и создание сводных таблиц.

3 РЕЗУЛЬТАТЫ

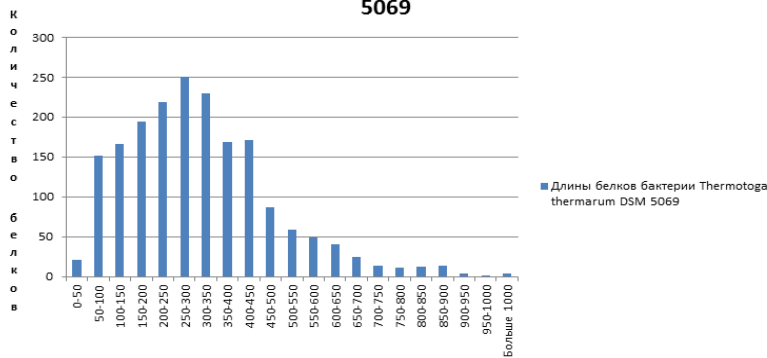
Оптимальная длина белка равняется 250 аминокислотным остаткам. Гены РНК и гены белков неравномерно располагаются на прямой и комплементарной цепи.

3.1 Длина белков *Thermotoga thermarum* DSM 5069

Для подсчета и выведения информации была построена гистограмма 1. На ее основе можно предположить, что оптимальной длиной генов белка такого организма является длина в 250-300 аминокислотных остатков. Их количество равно 250—ти различным последовательностям. При переходе от длины в 50 аминокислотных остатков к 50-100 наблюдается резкий скачок (от 21 до 151 последовательности), что наводит мысли о том, что короткие последовательности скорее всего не столь важны, ведь ими нельзя закодировать достаточное количество необходимой информации. Количество

белков с длиной в 250-400 и 400-450 аминокислотных остатков примерно одинаковы (169 и 171 соответственно), и от этих чисел начинается спад количества белков. Длина больше 900 уже является неудобной для кодировки информации, так как белков с длиной больше 900 аминокислотных остатков всего десять при общем количестве последовательностей в 1945 последовательностей. В среднем, оптимальная длина гена белка равна примерно от 100 до 450 аминокислотных остатков.

Длины белков бактерии *Thermotoga thermarum* DSM 5069



Гистограмма 1: По горизонтали длины белков в аминокислотных остатках, по вертикали – количество белков

3.2 Число генов белков и генов РНК на прямой и комплементарной цепях

На основе данных из таблицы №1 можно выдвинуть предположение, что для данного организма расположение генов на прямой цепи является оптимальным. Расположение генов белка на прямой цепи встречается примерно в 1,3 раза чаще, чем расположение генов на обратной цепи. Расположение генов на прямой цепи РНК встречается в 1,6 раза чаще, нежели на комплементарной.

(см. ниже)

Расположение гена на цепи	Прямая (+)	Комплементарная (-)
Гены белков	1095	850

Гены РНК	32	19
----------	----	----

ТАБЛИЦА №1

ОБСУЖДЕНИЕ

Таким образом, полученный результат наглядно демонстрирует, что оптимальная длина гена белка в рассматриваемом организме – 250-300 аминокислотных остатков. Очевидно, это число наиболее подходит для кодирования жизненно важной информации. Однако расположение генов белка и генов РНК на прямых и обратных цепях вводит в заблуждение и заставляет задуматься. Может ли быть такое, что прямая цепь больше подходит для кодирования информации у данной бактерии? Интуитивно кажется, что одинаковое разделение генов по двум цепям будет наиболее действенно ввиду равномерности и, как следствие, более удобного считывания информации, но, похоже, для данной бактерии это не столь важно.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

[Гистограмма в Excel](#)

ЛИТЕРАТУРА

[1] - <http://www.ncbi.nlm.nih.gov/genome/?term=thermotoga%20thermarum%20DSM%205069>
