

Обзор протеома археи *Thermococcus barophilus*

Худякова К. А.^{1,*}¹Факультет биоинженерии и биоинформатики МГУ, khudyakovaks@gmail.com.

РЕЗЮМЕ

Работа посвящена изучению расположения генов на цепях ДНК и распределения белков по длине у кольцевой хромосомы археи *Thermococcus barophilus* штамма MP. Было показано, что распределение генов по цепям описывается биномиальным распределением с вероятностью успеха 0,5. Длина 84% белков находится в пределах от 100 до 450 аминокислотных остатков.

1 ВВЕДЕНИЕ

Архея *Thermococcus barophilus* относится к классу Thermococci. Это гипертермофильный и барофильный организм, для которого оптимальными условиями являются температура 75°C и давление 40 Мпа. Эти археи строгие анаэробы и гетеротрофы, способные метаболизировать серу. Обнаружены представители этого вида были в гидротермальных источниках Атлантики на глубине 3550 метров в условиях высокой концентрации монооксида углерода.^[1]

Геном представлен кольцевой хромосомой размером 2,020,078 bp и кольцевой плазмидой размером 54,159 bp. В этой статье исследуются только РНК- и белок-кодирующие гены, расположенные на хромосоме.

Целью работы было исследование протеома археи. Была проверена и принята гипотеза о случайности расположения генов по цепям ДНК, а так же построено распределение белков по длине.

2 МЕТОДЫ

Использовались результаты секвенирования археи *Thermococcus barophilus* MP, найденные в базе данных NCBI^[2] по идентификатору NC_014804.

Данные из полученных файлов были импортированы в Microsoft Excel 2010. Лишняя информация была удалена, длины всех белков организма в аминокислотных остатках были собраны на отдельном листе (hist) и выбраны карманы для построения гистограммы (столбец Step, шаг 50). Для построения гистограммы (рис. 1) была подключена надстройка «Пакет анализа».

Для подсчета количества генов белков и РНК на «-» и «+» цепях использовалась функция СЧЁТЕСЛИ(COUNTIF), кроме того, была задействована функция БИНОМРАСП для проверки гипотезы о случайности распределения генов по цепям (таблица 1).

Вероятность того, что из 2258-и генов на одной цепочке обнаружено 1128 или меньше, равна 0.49, что не противоречит предположению о случайном независимом распределении генов по цепочкам при уровне доверия 0.05.

3 РЕЗУЛЬТАТЫ

3.1 Распределение длин белков

Распределение длин белков показано на гистограмме (рис. 1). 84% белков имеют длину от 100 до 450 аминокислотных остатков. Средняя длина белка равна 279 амк. Самый маленький белок состоит из 37 амк, самый большой из 2030. Больше всего белков попадает в диапазон 200-250 амк, их 287 штук.

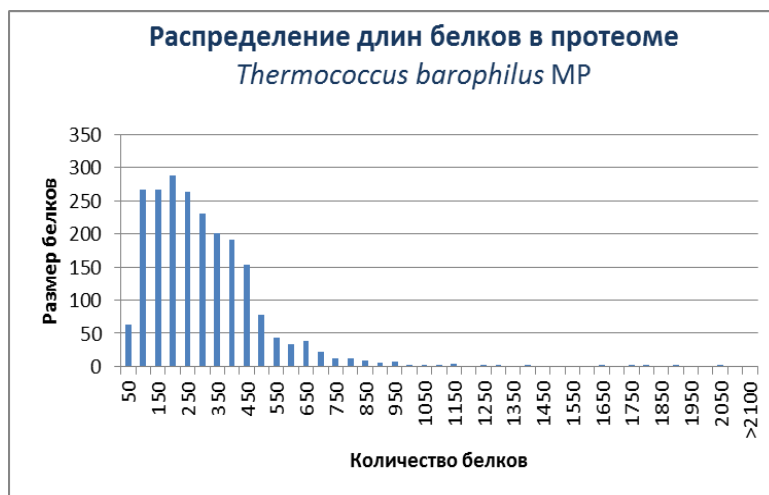


Рис. 1. Распределение длин белков в протеоме археи *Thermococcus barophilus*. На горизонтальной оси указаны длины белков (в аминокислотных остатках), на вертикальной – количество белков с такой длиной.

3.2 Распределение генов по цепям

Результаты по количеству генов на разных цепях ДНК представлены в таблице 1.

Значения для количества генов на прямой и обратной цепочках, приведенные в таблице, близки между собой, то есть гены распределены примерно поровну.

4 ОБСУЖДЕНИЕ

В диапазон длины 100 – 450 а.о. попало 84% всех белков организма, и это неудивительно, так как эта длина является наиболее оптимальной для существования и функционирования белка. Слишком короткие белки не были бы так высоко специализированы, как нужно для того, чтобы нормально функ-

*Кому следует адресовать корреспонденцию.

ционировать, а длинные полипептидные цепи имели бы громоздкие четвертичные структуры, из-за чего также не имели бы необходимых свойств.

Биномиальное распределение – это распределение числа успехов в n испытаниях Бернулли с вероятностью успеха в каждом испытании k . Мы проверили, распределено ли попадание гена на «+» или «-» цепь биномиально с вероятностью успеха 0,5. Вероятность, что данные распределены согласно гипотезе, 49%, что выше уровня значимости 5%, значит, оснований отвергнуть гипотезу нет. При интерпретации результата следует учитывать, что биномиальное распределение – это распределение независимых случайных величин, а попадание гена на ту или иную цепь может быть зависимым событием, потому что гены могут формировать опероны.

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

Цепь	Белки	РНК	Итог
Прямая цепь	1102	28	1130
Обратная цепь	1105	23	1128
Всего	2207	51	2258

5 СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

(1)<http://kodom.fbb.msu.ru/~khudyakova/term1/Thermococcus.xlsx> - Файл Excel с исходными данными, гистограммой и таблицей распределения генов по цепям.

БЛАГОДАРНОСТИ

Автор выражает благодарность факультету биоинженерии и биоинформатики МГУ за предоставленную возможность выполнить исследование

СПИСОК ЛИТЕРАТУРЫ И ССЫЛКИ

- [1] NCBI PubMed Vannier P., Marteinson VT, Fridjonsson OH, Oger P, Jebbar M. J Bacteriol. 2011 Jan 7;193(6):1481-2. <http://www.ncbi.nlm.nih.gov/pubmed/21217005>.
- [2] NCBI, полный геном T. barophilus: ftp.ncbi.nlm.nih.gov/genomes/archive/old_refseq/Bacteria/Thermococcus_barophilus