
Описание протеома бактерии *Archaeoglobus fulgidus*

Буянова М.Ф.¹

¹Факультет биоинженерии и биоинформатики МГУ имени М.В. Ломоносова.

РЕЗЮМЕ

В данном обзоре представлены результаты анализа протеома археи *Archaeoglobus fulgidus*. В частности, сделана попытка выявления закономерностей распределения длин белков по длинам. Рассмотрены и некоторые особенности генома: проверена случайность распределения генов, кодирующих белки и РНК, по прямой и обратной цепям ДНК, подсчитано количество квазиоперонов, перекрытий генов и показано наличие генов, длины которых не кратны трём. Все вычисления и анализ данных произведены при помощи программы Microsoft Office Excel 2007* в рамках учебного курса факультета биоинженерии и биоинформатики МГУ имени М.В. Ломоносова.

1 ВВЕДЕНИЕ

Archaeoglobus fulgidus – сульфатредуцирующая архея-термофил, которая была выделена из горячих морских осадков вблизи гидротермальных подводных источников, из чёрных курильщиков Срединно-Атлантического хребта, из нефтяных скважин Северного моря, из-под замёрзшей поверхности северного склона Аляски^[2]. Это первая архея с метаболизмом, основанным на серосодержащих соединениях, геном которой был секвенирован. Он состоит из одной хромосомы, плазмид не обнаружено.

В ходе работы был проведен анализ генома *Archaeoglobus fulgidus*, были рассмотрены как гены, кодирующие РНК (49 генов), так и гены, кодирующие белки (2420 генов). Это позволило сделать некоторые выводы о протеоме археи. В частности, по генам были рассчитаны длины белков, которые ими кодировались, и построена гистограмма распределения белков по длинам, после чего были выявлены самые часто встречающиеся длины белков в протеоме.

Попутно происходил и анализ собственно генома археи. Была проверена гипотеза о случайном распределении генов по обратной и прямой цепям ДНК при помощи функции биномиального распределения. Также были учтены такие особенности генома как перекрывание генов, квазиопероны, существование генов, длины которых в парах нуклеотидов не кратны трём.

Проведенная работа даёт как общие сведения о протеоме и геноме археи, так и сведения об интересных их особенностях.

2 МАТЕРИАЛЫ И МЕТОДЫ

Для проведения анализа использовались файлы базы данных NCBI^[1], содержащие данные о геноме археи. Для работы с

единственной хромосомой, имеющейся у *Archaeoglobus fulgidus*, были взяты два файла расширений .gnt и .ppt для получения данных по генам, кодирующим РНК и белки соответственно. Указанные файлы были импортированы в программу Microsoft Office Excel 2007 и представлены в удобном для чтения и анализа виде при помощи функции подразбиения текста на столбцы по разделителю (которым в данном случае являлся знак табуляции). Описанным способом была создана общая таблица, в ней добавлен столбец с типом гена (кодирующий белок или РНК), а исходный столбец Location заменён на два столбца с координатами начала и окончания гена еще одним применением разделителя ("."). Далее была создана таблица с указанием количества генов на прямой и обратной цепи ДНК (функция СЧЁТЕСЛИ) и была сделана проверка гипотезы о случайном распределении генов (функция БИНОМРАСП).

Для дополнительного исследования особенностей генома использовались функции ЕСЛИ (для выведения результатов в зависимости от выполнения заданных условий), ОСТАТ (для проверки кратности трём длины гена в п.н.). Также для комплексного отчёта по указанным особенностям создана сводная таблица (Вставка → Сводная таблица).

При рассмотрении генов, кодирующих белки, были выявлены максимальная и минимальная длины белков в протеоме. Были рассчитаны карманы (использовалась длина кармана, равная 100) и частота встреч длин белков в этих карманах (функция СЧЁТЕСЛИМН). На основе полученных данных была построена гистограмма (при помощи функционала надстройки MS Excel "Пакет анализа данных"), после чего происходил уже непосредственно визуальный анализ полученного распределения.

3 РЕЗУЛЬТАТЫ

Все результаты, как отмечалось ранее, получены в MS Excel 2007. Данные для создания этого мини-обзора содержатся в файле Buaynova_pr14.xlsx. (см. раздел "Сопроводительные материалы").

3.1 Распределение генов на прямой и обратной цепях ДНК

Результаты распределения генов (CDS и RNA) представлены в Таблице 1. Значение функции биномиального распределения для случая, когда 1206 из 2469 генов оказались на прямой цепи, равняется 0,13.

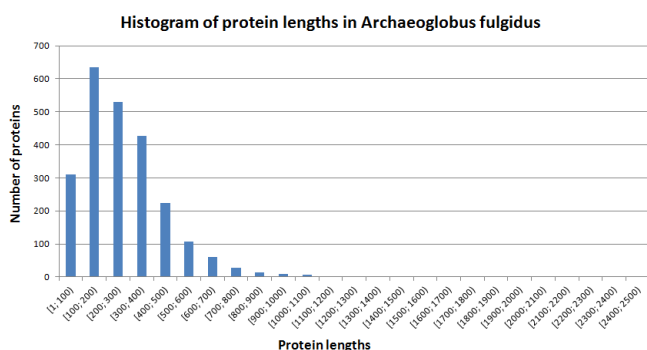
* версия Enterprise

Таблица 1. Распределение генов на прямой и обратной цепях ДНК.

Цепь / Тип гена	CDS	RNA
Прямая цепь	1187	19
Обратная цепь	1233	30

3.2 Распределение белков по длинам

Результаты распределения можно видеть на Рис. 1. Как видно из данной гистограммы, чаще всего у *Archaeoglobus fulgidus* встречаются белки с длинами в пределах от 100 до 200 аминокислотных остатков.



Максимальная длина – 2425 а.о. Минимальная – 25 а.о.

Рис. 1. Гистограмма длин белков археи *Archaeoglobus fulgidus*.

3.3 Квазиопероны и перекрытия между генами

Исходя из предположения, что каждый ген принадлежит какому-нибудь квазиоперону, а два соседних гена принадлежат одному квазиоперону, если они находятся на одной цепи и расстояние между ними не превышает 100 п.н., получаем количество квазиоперонов, которое для исследуемой археи равняется 998 (960 в генах, кодирующих белки, и 38 в генах, кодирующих РНК).

Перекрытием между генами считались участки цепи ДНК, которые принадлежат как минимум двум генам одновременно. Таких областей перекрытия получилось 613, причем из них только одна между генами, кодирующими РНК.

3.4 Гены, длина которых не кратна трём

Также анализ длин генов показал, что в геноме археи имеются гены, длины которых в п.н. не делятся на три. Таких генов всего 47 (из которых 34 кодируют РНК и всего 13 – белки).

4 ОБСУЖДЕНИЕ

4.1 Распределение генов по цепям

Вероятность того, что не более 1206 из 2469 генов обнаружено на одной цепочке, составляет 0.13, что не противоречит гипотезе о случайном независимом распределении генов между прямой и обратной цепями ДНК при уровне доверия 0.05.

4.2 Распределение длин белков

Самое частое значение для длины белка – от ста до двухсот а.о. Это относительно небольшая величина для белка, что легко объяснимо тем фактом, что у примитивных по строению и жизнедеятельности архей, древнейших организмов, просто отсутствует нужда в длинных и сложных белках. Однако, что интересно, несколько белков имеют достаточно большую длину. Например, белок из 2425 а.о. Это белок поверхностного слоя, кодирующийся геном *slgB-2. Archaeoglobus fulgidus*, как и большинство архей, обладает клеточной стенкой, представленной так называемым S-слоем – жёсткой сеткой белковых молекул^[3]. Этот слой защищает архею от физических и химических воздействий, препятствует контакту макромолекул с её мембраной. Можно сделать предположение о том, что для такого экстремального термофила, к коим относится исследуемая архея, наличие и особая прочность S-слоя жизненно важна. Поэтому именно поверхностный белок имеет такую большую длину по сравнению с остальными белками протеома.

4.3 Квазиопероны и перекрытия между генами

По количеству на каждые 2,5 гена приходится один квазиоперон. Это можно объяснить тем фактом, что многие белки в клетке работают в комплексе, и, соответственно, клетке выгодно проводить их синтез сцепленно.

Перекрытие же генов можно объяснить их нахождением в различных рамках считывания или же альтернативным сплайсингом. Пока что, однако, остается не вполне ясным, почему для генов, кодирующих РНК, встретилось лишь одно перекрытие.

4.4 Гены, длина которых не кратна трём.

Интересно отметить, что процент таких генов среди кодирующих белок составляет не более 0.5%, тогда как процент среди кодирующих РНК генов составляет 69%. Поэтому можно считать, что если ген кодирует белок, то в нем произошел сдвиг рамки считывания, вставка или делеция, которые привели к нарушению кратности.

Таким образом, для CDS это скорее исключение из правила. Для РНК триплетность не является обязательной, что и объясняет то, что так высок процент некратных трём длин генов.

5 СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Файл со всеми данными, использованными в обзоре:
http://kodomofbb.msu.ru/~emkeller/terms/term1/Buyanova_pr14.xlsx

6 БЛАГОДАРНОСТИ

Хочу выразить свою благодарность А. Алексеевскому, А. Жариковой и И. Русинову за проведение учебного курса по работе в Microsoft Office Excel и применению данной программы для решения биоинформатических задач.

7 СПИСОК ИСТОЧНИКОВ

- [1] NCBI, полный геном *A. fulgidus*:
ftp.ncbi.nlm.nih.gov/genomes/archive/old_refseq/Bacteria/Archaeoglobus_fulgidus_DSM_4304_uid57717
- [2] Wikipedia, Сульфатредуцирующие бактерии:
ru.wikipedia.org/wiki/Сульфатредуцирующие_бактерии
- [3] Wikipedia, Археи:
<ru.wikipedia.org/wiki/Археи>