МИНИ-ОБЗОР

Описание протеома цианобактерии Anabaena variabilis.

Выполнено Беловым Л. В., студентом 1-ого курса факультета биоинженерии и биоинформатики МГУ им. Ломоносова.

Резюме

В данном обзоре представлены результаты анализа протеома цианобактерии Anabaena variabilis. Была сделана попытка выявления закономерностей распределения длин белков по длинам. Рассмотрены 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

Anabaena variabilis это нитевидная цианобактерия. Этот вид способен к фотосинтезу. При этом Anabaena variabilis гетеротофна, так что может развиваться без света в присутствии фруктозы. Может превращать N2 в NH4+ фиксируя азот.

Anabaena variabilis филогенетически является двоюродным братом более известного вида Nostoc spirrilum. Оба этих вида вместе со многими другими цианобактериями , как известно, образуют симбиотические отношения с растениями. Другие цианобактерии, как известно, образуют симбиотические отношения с диатомовыми водорослями , хотя такие отношения не наблюдались с Anabaena variabilis.

Anabaena variabilis так же является моделью для изучения начал многоклеточной жизни из-за её нитевидности и потенциала к клеточной дифференциации (может формировать схожие со спорами клетки (акинеты), небольшие подвижные нити (гормонгии), и, самое главное, гетероцисты, являющиеся азотопродуцирующими клетками).

В ходе работы был проведен анализ генома Anabaena variabilis, были рассмотрены как гены, кодирующие РНК (62 гена), так и гены, кодирующие белки (5043 гена). Это позволило сделать некоторые выводы о протеоме цианобактерии. 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 цианобактерии. 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, содержащие данные о геноме археи. Для работы из 5 хромосом Anabaena variabilis были взяты два файла одной (NC_007413) с расширениями .rnt и .ptt 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).

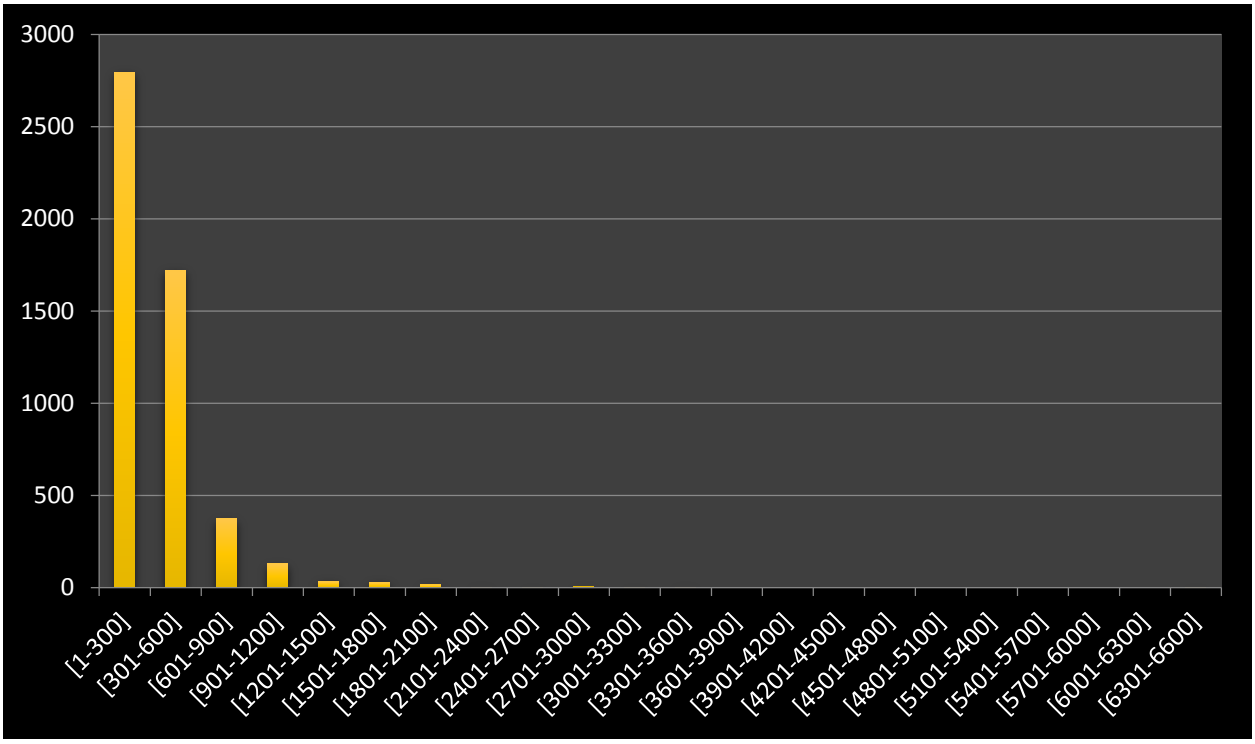
Для дополнительного исследования особенностей генома использовались функции ЕСЛИ (для выведения результатов в зависимости от выполнения заданных условий), ОСТАТ (для проверки кратности трём длины гена в п.н.). Также для комплексного отчёта по указанным особенностям создана сводная таблица (Вставка –>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

Гистограмма длин белков:

В ходе работы была составлена гистограмма длин белков цианобактерии Anabaena variabilis (смотри ниже). Можно увидеть, что данная цианобактерия имеет больше всего белков в диапазоне от 1 до 300 аминокислот.



(Рис. 1: «Гистограмма длин белков»)

Распределение белков по длинам:

Также было определено количество генов, кодирующих белки и РНК для прямой и обратной цепи. Все данные были занесены в таблицу 1. Можно увидеть, что всего бактерия имеет 5105 генов.

Таблица 1: «Распределение генов белков и РНК по цепям ДНК».

	Stand "+"	Stand "-"
CDS	2668	2375
RNA	28	35

(Stand “+” – прямая цепь Stand “-” – обратная)

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 цианобактерии равняется 3344 (3291 в генах, кодирующих белки, и 53 в генах, кодирующих РНК).

Перекрытием между генами считались участки цепи ДНК, которые принадлежат как минимум двум генам одновременно. Таких областей перекрытия получилось 228, причем из них только одна между генами, кодирующими РНК.

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 44 (из которых все 44 кодируют РНК).

Плазмиды:

В таблице 2 представлено количество плазмид для рода Anabaena и вида variabilis.

Таблица 2: ”Колличество плазмид рода Anabaena и вида variabilis.”

	Род (Genus) Anabaena	Вид (Species) Variabilis
Количество плазмид	14	4

(Немного о плазмидах)

Плазмиды — небольшие молекулы ДНК, физически отдельные от геномных хромосом и способные реплицироваться автономно. Как правило, плазмиды встречаются у бактерий и представляют собой двухцепочечные кольцевые молекулы, но изредка плазмиды встречаются также у архей и эукариот.

В природе плазмиды обычно содержат гены, повышающие устойчивость бактерии к неблагоприятным внешним факторам (в т. ч. устойчивость к антибиотикам), нередко они могут передаваться от одной бактерии к другой (иногда даже к бактерии другого вида) и, таким образом, служат средством горизонтального переноса генов.

Размер плазмид варьирует от 1 до свыше 1000 тысяч пар оснований. Количество идентичных плазмид в пределах одной клетки изменяется от одной до тысяч в зависимости от дополнительных обстоятельств. Плазмиды можно считать видом МГЭ (мобильных генетических элементов), поскольку они часто передаются при конъюгации — механизме горизонтального переноса генов.)

(4) Обсуждение и заключение

Вероятность того, что не более 2696 из 5105 генов обнаружено на одной цепочке, составляет 0.99, что не противоречит гипотезе о случайном независимом распределении генов между прямой и обратной цепями ДНК с вероятностью 0.5 .

Самое частое значение для длины белка – от 1 до 300 а.о. Это сравнительно небольшая величина для белка, что вполне объяснимо тем фактом, что примитивным по строению и жизнедеятельности цианобактериям, древнейшим организмам, нет необходимости в длинных и сложных белках.

Чуть более 65% генов имеют 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.

Количество генов, длина которых не кратна трём, среди генов, кодирующих белок , равно 0. Среди генов, кодирующих РНК, таких генов 83%. Впрочем для РНК триплетность не является обязательной, что и объясняет то, что так высок процент некратных трём длин генов. Для генов, кодирующих белок это означает, что нет никаких сдвигов рамки считывания, вставка или делеция, которые могли бы привести к нарушению кратности.

(5) Сопровождающие материалы

Файл .xlsx с данными:

http://kodomo.fbb.msu.ru/~volkiller/term1/Small_review/Belov_pr15.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) Источники

Wikipedia:

https://en.wikipedia.org/wiki/Anabaena_variabilis

NCBI:

http://www.ncbi.nlm.nih.gov/genome/?term=Anabaena+variabilis+ATCC+29413