

Обзор протеома бактерии *Thermus thermophilus* HB8

Корзина Анастасия

Факультет биоинженерии и биоинформатики МГУ им. М.В. Ломоносова

РЕЗЮМЕ

В ходе исследования протеома бактерии *Thermus thermophilus* HB8 с помощью Excel было выявлено, что протеом этой бактерии содержит более двух тысяч генов и наибольшее количество этих генов содержит около 150 пар оснований. Так же были изучены вопросы, связанные с распределением генов между прямой и обратной цепью и некратностью трем числа аминокислот в некоторых генах.

1 ВВЕДЕНИЕ

Выполняя этот обзор, автор преследовал несколько целей:

- Выявить особенности генов в протеоме бактерии *Thermus thermophilus* HB8;
- Описать эти особенности;
- Для большей наглядности снабдить обзор рисунками и таблицами;
- Овладеть методами анализа данных в программе Microsoft Office Excel 2007;*
- Научиться основным принципам написания обзоров на примере данного.*

2 МАТЕРИАЛЫ И МЕТОДЫ

Для написания обзора были использованы данные из базы NCBI: <ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/>. Анализ данных проведен в Microsoft Office Excel 2007. Были использованы следующие возможности этой программы:

- Функция «СЧЁТЕСЛИ» для подсчета количества генов на прямой и обратной цепях ДНК;
- Функция «ЦЕЛОЕ» для обнаружения подозрительных длин генов, не кратных 3;
- Функции «МИН» и «МАКС» соответственно для определения минимальной и максимальной длины генов;
- Функция, распределяющая значения по карманам гистограммы;
- Построение гистограммы и круговой диаграммы, по набору значений;
- Автозаполнение.

* Эти цели не относятся к теме самого обзора, но они не менее важны, чем первые три. Это связано с тем, что этот обзор первая работа автора и выполнена она с целью получения зачета по блоку, в котором изучается Excel.

**Все таблицы, рисунки и файлы представлены в разделе сопроводительные материалы.

3 РЕЗУЛЬТАТЫ

Обработка исходных данных и обобщение полученных результатов привели к следующим выводам:

- Генов, кодирующих белки, значительно больше, чем генов, кодирующих РНК (2173 и 53 пары оснований соответственно).
- Длины генов находятся в интервале от 22 до 2915 пар оснований. Среднее значение длины около 300 пар оснований. А с наибольшей частотой встречаются гены длиной в 150 пар оснований (в этом кармане их 64).
- На прямой и комплементарной цепи примерно одинаковое число генов (53.6% и 46.4% соответственно).
- Найдены такие гены, длины которых не кратны трем.

Результаты представлены на рисунках 1,2 и в таблице 1. Также в документе 1 представлены расчеты и таблицы с исходными данными.**

4 ОБСУЖДЕНИЕ

- Длины распределены неравномерно. Они даже не образуют симметричного колокола. Купол смещен сильно влево. Из-за такого сильного смещения купола среднее и самое частое значение так сильно различаются.
- Длина гена не кратна 3, значит, он считается другой рамкой (-1, -2, -3, +1, +2) и поэтому имеет еще продолжение.

5 СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ



Рис.1. Распределение длин генов из протеома термофильной бактерии. По горизонтальной оси отложены длины генов в парах оснований, по вертикальной отложена частота встречаемости этих длин в процентах.

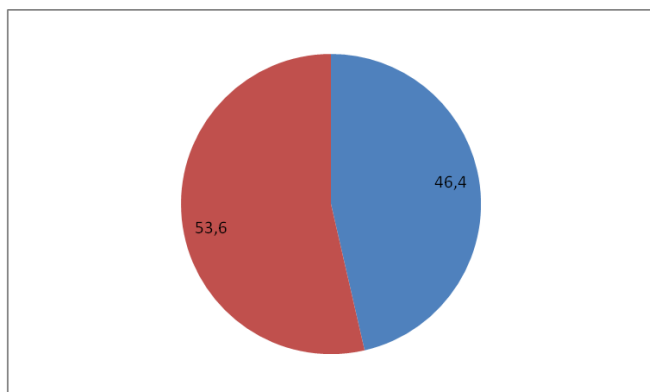


Рис.2. Распределение генов из протеома термофильной бактерии по прямой (красным) и комплементарной цепи (синим).

Табл.1. Распределение генов, кодирующих белки и кодирующих РНК, из протеома термофильной бактерии по прямой и комплементарной цепи.

	Прямая цепь	Обратная цепь	Всего
CDS	1009	1164	2173
RNA	24	29	53

Док.1. Исходные данные и расчеты среднего значения длины гена, самого частого значения длины гена, максимальной и минимальной длин генов, построение диаграмм, гистограмм и таблиц.

http://kodo.fbb.msu.ru/~askorzina/term1/Korzina_pr15_ex.xlsx

6 БЛАГОДАРНОСТИ

Автор выражает благодарность преподавателям информатики А.Алексеевскому и И.Русинову за предоставленные материалы по работе с Excel и шаблон оформления обзора.

7 СПИСОК ЛИТЕРАТУРЫ

<ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/>