

Обзор протеома бактерии *Bacteroides thetaiotaomicron* VPI-5482

Сигорских А. И.

Факультет биоинженерии и биоинформатики МГУ им. М.В.Ломоносова, 1 курс.

РЕЗЮМЕ

Данная работа посвящена исследованию протеома бактерии *Bacteroides thetaiotaomicron* VPI-5482 с целью изучения распределения длин белков и описания расположения их генов на прямой или комплементарной цепях ДНК. Данные были получены с сервера NCBI и обработаны с помощью программы Microsoft Excel.

1 ВВЕДЕНИЕ

Термин «протеом» обозначает всю совокупность белков организма и, в отличие от генома, определяемого последовательностью нуклеотидов, не сводится к сумме последовательностей аминокислот, так как включает в себя также пространственные структуры всех содержащихся в нём белков и функционального взаимодействия между ними. Протеомика, оперирующая большими массивами данных и потому тесно связанная с биоинформатикой, способна вывести на принципиально новый уровень изучение клеточной физиологии, раннюю диагностику различных заболеваний и уточнение аннотации генома, а благодаря сравнению протеомов близкородственных организмов позволяет выявить как общие для них белки, так и белки, которые обуславливают различия их фенотипов. Преимущество протеомики перед геномикой обуславливает тот факт, что наличие какого-то гена в геноме не позволяет однозначно судить о структуре кодируемого белка; более того, неясно даже, будет ли с этого гена производиться транскрипция. Таким образом, дальнейшее развитие протеомики способно сделать её одним из ключевых направлений биологии XXI века.

2 МАТЕРИАЛЫ И МЕТОДЫ

Данные о протеоме бактерии *Bacteroides thetaiotaomicron*, штамм VPI-5482 были получены с сервера NCBI [1], количественные и статистические данные дала обработка полученного описания протеома программой Microsoft Office Excel 2013.

Записи NC_004663.ppt (хромосома), NC_004663.mnt (РНК) и NC_004703.ptt (плазмида) были импортированы в файл Excel на соответствующие листы. Для установления частоты встречаемости белков разной длины была построена столбчатая гистограмма (лист Histogram в Excel-файле), для чего значения длины белков, кодируемых хромосомой и плазмидой были проанализированы функцией СЧЁТЕСЛИМН. «Шаг» разницы длин белков был выбран в 50 аминокислотных остатков, от 0 до 2200 АО (более длинных белков обнаружено не было). Также были посчитаны минимальное, максимальное и среднее значения длин.

На листе Strand Table были построены таблицы, указывающие количества генов на прямой (+) и комплементарной (-) цепях ДНК, и посчитанное функцией БИНОМРАСП Р-значение (вероятность статистической случайности исхода «+» при условии, что распределения является биномиальным).

3 РЕЗУЛЬТАТЫ

Таблица и гистограмма

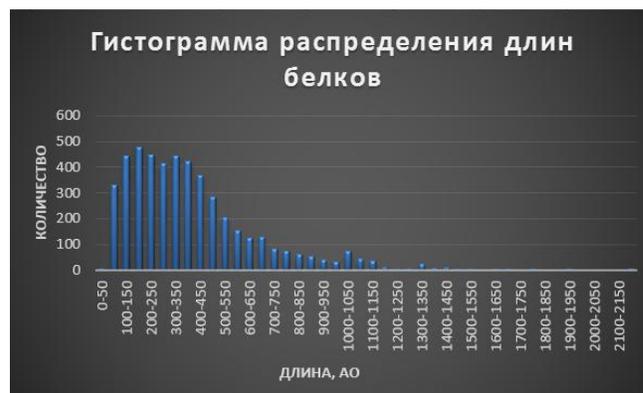


Таблица 1 – расположение генов на “+” и “-” цепях ДНК

Strand	Chromosome	Plasmid	RNA	Total sum
"+"	2365	29	40	2434
"-"	2413	9	46	2468
Total sum	4778	38	86	4902
Randomness check	Chromosome	Plasmid	RNA	Total
p-value	0,248271167	0,999764	0,295018	0,318704

На гистограмме наблюдается, что большинство белков имеет длину от 100 до 550 АО, минимальная длина белка 32 АО, максимальная 2183, среднее значение длины 389,0179.

Расположение генов белков и РНК указано в таблице 1; проверка гипотезы о том, что гены распределены по цепочкам случайно с вероятностью 0,5 дала чрезвычайно низкий результат в целом (0,318704), следовательно, гипотеза не подтверждена; однако, частный анализ распределения генов в плазмиде дал наоборот крайне высокий результат (0,999764), что подтверждает гипотезу о случайном распределении генов конкретно для плазмиды.

4 ОБСУЖДЕНИЕ

Подавляющее число белков имеет длину в пределах от 100 до 550 АО, что является нормой для бактерий; лишь малое число белков по длине превосходит 1000 АО; гипотеза о случайном распределении подтвердилась только в частном случае плазмиды.

5 СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

- Файл Excel *proteome_bacteriodes_thetaiotaomicron* – содержит все данные и расчёты.

6 СПИСОК ЛИТЕРАТУРЫ

[1]-

ftp://ftp.ncbi.nlm.nih.gov/genomes/bacteria/bacteroides_thetaiotaomicron_vpi_5482_uid62913/