

Описание протеома и генома бактерии *Streptomyces avermitilis* MA-4680=NBRC 14893

Веселова С.А.¹

¹Факультет биоинженерии и биоинформатики Московского Государственного Университета им. М.В. Ломоносова.

РЕЗЮМЕ

Данная работа посвящена исследованию протеома бактерии *Streptomyces avermitilis*. Было проанализировано количество генов белков и РНК на прямой и комплементарной цепях ДНК, распределение белков по их длинам, определено число генов белков и РНК.

1 ВВЕДЕНИЕ

Род *Streptomyces* широко распространен в природе, особенно известны его почвенные представители. Бактерии грамположительны, имеют гифы, воздушный мицелий и конидиальные споры - особенности, морфологически сходные с особенностями грибов [1].

Streptomyces avermitilis впервые был выделен в 1979 году из образца почвы в Японии, геном был секвенирован в 2003 году, представлен двуцепочечной ДНК. Интересен тот факт, что данная бактерия имеет линейную форму хромосом, а не привычную кольцевую, свойственную большинству представителей бактерий [1][2].

Streptomyces avermitilis MA-4680 известен как продуцент авермектина – вещества, выделяемого бактерией в качестве вторичного метаболита. Этот антигельминтный агент входит в состав различных лекарственных препаратов против онхоцеркоза и других заболеваний, вызываемых эндопаразитами [3][4].

2 МАТЕРИАЛЫ И МЕТОДЫ

Все необходимые данные для исследования были взяты с сервера NCBI, их обработка осуществлялась при помощи программы Microsoft Office Excel 2010.

В ходе работы были использованы функции:

- СЧЁТЕСЛИ – для подсчета количества генов CDS/RNA;
- СЧЁТЕСЛИМН – для подсчета количества значений длин белков в определенных карманах;
- БИНОМ.РАСП – для нахождения вероятности существования
- МЕДИАНА, МАКС, МИН, СРЗНАЧ – для получения информации о длинах белков

3 РЕЗУЛЬТАТЫ

Все результаты были получены с помощью вышеупомянутой программы Microsoft Office Excel 2010. Данные содержатся в файле pr13.xlsx (см. «Сопроводительные материалы»).

3.1 Количество генов белков и РНК, примерная оценка числа генов на 1Mbp

Как уже упоминалось выше, геном бактерии представлен двуцепочечной молекулой ДНК. По данным NCBI [5], геном имеет размер 9,025,608 миллионов пар нуклеотидов (Mbp).

В таблице ниже представлено распределение генов по категориям и их приблизительное количество на 1Mbp.

Таблица 1. Распределение генов по категориям.

Category	Amount	Per 1 Mbp
CDS	7588	840,31
rRNA	18	1,99
tRNA	72	7,97
tmRNA	1	0,11
ncRNA	2	0,22

3.2 Распределение длин белков в протеоме бактерии

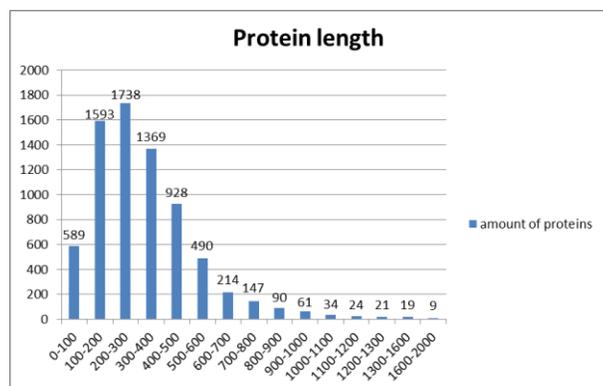
Для создания наглядной гистограммы потребовались предварительные расчеты, представленные в Таблице 2. Полученные значения помогли задать карманы длин таким образом, чтобы столбцы не состояли из одного элемента при каких-то значениях длин.

Таблица 2. Информация о длинах белков.

Category	Value
Min	29
Max	7746
Median	285
Average	337

В результате были выбраны карманы длин шагом в 100 аминокислотных остатков. Полученная гистограмма представлена ниже. Для наглядности над каждым столбцом сверху указано точное количество белков определенной длины. В процессе обработки данных было обнаружено, что 91.5% белков не превышают длину в 600 аминокислотных остатков.

Рисунок 1. Распределение белков по длине.



3.3 Распределение генов по прямой/комплементарной цепи

В таблице ниже представлено распределение белков (CDS) и РНК (RNA) по прямой/комплементарной цепи. В целом, оно достаточно равномерное.

Вероятность нахождения не более 3768 из 7681 генов на одной цепочке составляет 0,05 (это значение интегральной функции для биномиального распределения). Для подсчета данной вероятности была использована функция «БИНОМ.РАСП» (см. «Сопроводительные материалы»)

Таблица 3. Число генов белков и РНК на прямой и комплементарной цепи.

Число генов	Прямая	Комплементарная
Белки	3721	3867
РНК	47	46

3.4 Количество квазиоперонов и пересечения генов в геноме бактерии

Если каждый ген принадлежит какому-то квазиоперону, а два соседних гена принадлежат одному квазиоперону при условии расстояния между ними не более 100 п.н., то количество квазиоперонов – 3564 (3517 в генах, кодирующих белки).

Пересечением генов в геноме считались участки ДНК, принадлежащие сразу нескольким генам. В геноме *S. avermitilis* их 1036, и только 1 между генами, кодирующими РНК.

Таблица 4. Количество квазиоперонов и пересечений генов.

Категория	Квазиопероны	Пересечения
CDS	3517	1035
Total	3564	1036

4 ОБСУЖДЕНИЕ

4.1 Количество генов белков и РНК, примерная оценка числа генов на 1Mbp

Помимо генов, кодирующих рибосомальную и матричную РНК, бактерия имеет тмРНК (транспортно-матричная) и нкРНК (некодирующая), которые могут выполнять различные регуляторные функции.

4.2 Распределение длин белков в протеоме бактерии

В целом, можно сказать, что данный результат характерен для *S. avermitilis* как представителя бактерий. Однако, есть и очень длинные белки – например, состоящий из 7746 аминокислот-

ных остатков. Этот белок принадлежит к семейству поликетидсинтаз (ферментные комплексы, синтезирующие вторичные метаболиты)^[6]. Таким образом, такое значение длины доказывает, что выработка авермектина является одной из главных особенностей бактерии.

4.3 Распределение генов по прямой/комплементарной цепи

Вероятность распределения была рассчитана по формуле биномиального распределения (БИНОМ.РАСП). Полученное значение (0,05) не противоречит выдвинутой гипотезе о случайном независимом распределении генов между прямой и комплементарной цепями ДНК.

4.4 Количество квазиоперонов и пересечения генов в геноме бактерии

Соотнеся общее число генов бактерии и количество квазиоперонов, получаем, что на каждые 2.2 гена приходится 1 квазиоперон. Это можно объяснить тем, что белки зачастую работают в комплексе (например, вышеупомянутые поликетидсинтазы), и их удобнее, выгоднее синтезировать вместе.

Количество пересечений генов вписывается в теоретическую одну треть от общего количества генов (количество, характерное для бактерий). Эти пересечения позволяют обрабатывать несколько генов вместе, а также могло повлечь редукцию размера генома, что может быть эволюционным преимуществом^{[8][9]}.

Однако, мне осталось непонятным, почему для РНК-кодирующих генов пересечение только одно.

5 СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Использованные файлы Excel доступны по ссылке:
<http://kodom.fbb.msu.ru/~sophia.veselova/term1/pr13.xlsx>

6 БЛАГОДАРНОСТИ

Хочу выразить свою благодарность преподавателям дисциплины «Практическая информатика» за проведение учебного курса по работе в Microsoft Office Excel и применению данной программы для решения биоинформатических задач.

7 СПИСОК ЛИТЕРАТУРЫ

- [1] DOGAN *S. avermitilis* project.
- [2] Burg, R., Miller, B., Baker, E., Birnbaum, J. et. al.(1979) Avermectins, New Family of Potent Anthelmintic Agents: Producing Organism and Fermentation. *Antimicrobial Agents and Chemotherapy*. Vol 15, No 3, p. 361-367.
- [3] Demain, A. (1999) Pharmaceutically active secondary metabolites of microorganisms. *Appl. Microbiol. Biotechnol.* Vol 52, p. 455-463.
- [4] Onthocerciasis article on Wikipedia.
- [5] NCBI *Streptomyces avermitilis* page.
- [6] Polyketide synthase on Wikipedia.
- [7] Haruo I., Jun I., Akiharu H. et. al. (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nature Biotechnology* 21, 526 - 531 (2003)
- [8] Johnson Z, Chisholm S (2004). "Properties of overlapping genes are conserved across microbial genomes". *Genome Res.* 14 (11): 2268–72
- [9] Luo, Yingqin; Battistuzzi, Fabia; Lin, Kui; Gibas, Cynthia (29 November 2013). "Evolutionary Dynamics of Overlapped Genes in *Salmonella*". *PLoS ONE*. 8 (11): e81016.