
Обзор генома бактерии *Aneurinibacillus* sp. XH2

Безуглов Виталий

Факультет Биоинженерии и Биоинформатики МГУ им. М. В. Ломоносова

АННОТАЦИЯ

В данном обзоре представлено краткое описание генома, а также изучение протеома бактерии *Aneurinibacillus* sp. XH2.

КЛЮЧЕВЫЕ СЛОВА

Aneurinibacillus, геном, Microsoft Excel, белок, аминокислотные остатки, медиана.

1 ВВЕДЕНИЕ

Aneurinibacillus - род бактерий, выделенный в 1996 году из семейства *Bacillus*. Бациллы являются аэробами или факультативными анаэробами, многие представители - хемоорганогетеротрофы. Большинство бацилл - почвенные сапрофиты. Некоторые бациллы вызывают болезни животных и человека, например сибирскую язву, токсикоинфекции. Геном состоит из 3664835 нуклеотидов. Всего генов - 3680. Геном был секвенирован в Китае в Center for Bioengineering and Biotechnology, China University of Petroleum в 2015 году. Авторы: Zhang,Z., Qiao,N., Zhang,Y., Xiao,Z., Jing,L. and Zhao,J.-Y. В данном обзоре мной проанализированы данные по количеству белков, определенной длины, количеству генов как белков, так и РНК, закодированных на прямой и обратной цепях ДНК, а также по количеству генов белков и РНК различных типов.

2 МАТЕРИАЛЫ И МЕТОДЫ

2.1 Используемые программы и базы данных

Геном был скачан и обработан из базы данных сайта NCBI[1]. Для наглядного отображения длин белков использовалась программа Microsoft Excel 2013 (русскоязычная версия).

2.2 Используемые команды в Excel

При работе в MS Excel мною использовались следующие формулы:

- =СЧЁТЕСЛИМН()
- =МИН()
- =МАКС()
- =СРЗНАЧ()
- =СТАНДОТКЛОН()
- =МЕДИАНА()

2.3

Построение таблицы длин белков, числа генов на прямой и обратной цепи, а также количества белков и РНК по типам.

Была скачана хромосомная таблица бактерии[2]

Для анализа длин белков был скопирован столбец с длинами белков и скопирован в отдельный лист Excel, найдены максимальное, минимальное и среднее значение, а также медиана и стандартное отклонение.

Для построения плоской таблицы, необходимой для отбора строк, описывающих полноценные гены белков; строк, описывающих псевдогены, а также строк, описывающих гены РНК был проведён анализ первого и второго столбцов сводной таблицы с помощью команды =СЧЁТЕСЛИМН().

Аналогичным образом была составлена таблица числа генов на прямой и обратной цепи.

Для построения таблицы количества белков и РНК по типам был использован фильтр для поиска необходимых генов белков и РНК.

Среди белков с помощью фильтра по названию типов белков(ribosomal, transport, hypothetical) были выбраны белки данных типов(рибосомальные, транспортные, гипотетические).

2.4 Построение гистограммы длин белков

Для построения гистограммы на основе столбца с длинами белков была использована программа Excel, раздел ВСТАВКА - Вставить гистограмму. Перед этим длины белков были собраны по диапазонам в 100 единиц длины интервала объектов, а именно белков.

3 РЕЗУЛЬТАТЫ

3.1 Состав генома бактерии

Всего в геноме 3732 гена. По данным NCBI в геноме 3664835 пар нуклеотидов.

Гены:	Прямая цепь(+)	Обратная цепь(-)
белков	1871	1605
псевдогенов	53	50
РНК	24	129

Таблица 1. Таблица числа генов белков, псевдогенов, РНК на прямой и обратной цепи.

	protein_coding		pseudogene	SRP_RNA	tRNA	rRNA	tmRNA	ncRNA	other	RNase_P_RNA
gene	3476	0	103	1	114	34	1	2	0	1
CDS	0	3579	0	0	0	0	0	0	0	0
ncRNA	0	0	0	1	0	0	0	0	2	1
tRNA	0	114	0	0	0	0	0	0	0	0
rRNA	0	34	0	0	0	0	0	0	0	0
tmRNA	0	1	0	0	0	0	0	0	0	0

Таблица 2. Таблица, описывающая гены белков, гены РНК, псевдогены.

В геноме 3476 генов кодируют белки, при этом 1871 находятся на прямой цепи, а 1605 - на обратной, 103 псевдогена: 53 на прямой, 50 на обратной цепи, 1 ген SRP_RNA, 114 генов тРНК, 34 гена рРНК, 1 ген tmRNA.

3.2 Длины белков

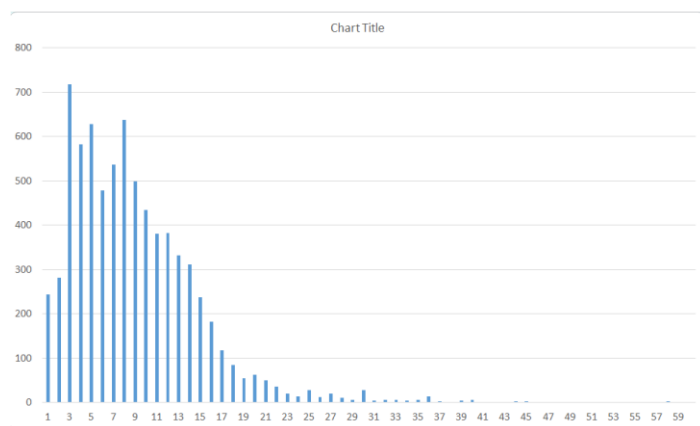


Рисунок 1. Гистограмма длин белков.

На рисунке горизонтальная ось обозначает сотни белков, то есть 1 – диапазон от 0 до 100, 2 – диапазон от 100 до 200 и т.д.
Минимальное значение - 0
Максимальное значение - 718
Среднее значение - 124,4
Стандартное отклонение - 203,39
Медиана - 8

3.3 Количество белков и РНК по типам

Белки	Кол-во
Рибосомальные	93
Транспортные	183
Гипотетические	1489

Таблица 3. Таблица числа генов белков и генов РНК по категориям.

В геноме большая часть генов белков - гипотетические. Количество транспортных и рибосомальных белков отличается приблизительно в 2 раза. Среди количества генов РНК большая часть - тРНК. Количество рРНК примерно в 3 раза меньше.

ЗАКЛЮЧЕНИЕ

Большее число генов белков по длине не превосходит 1700-1800 аминокислотных остатков. Гены белков закодированы как на прямой, так и на обратной цепи. Гистограмма длин белков показывает рост и быстрое падение длин. Поэтому медиана гораздо меньше среднего значения.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

http://kodom.fbb.msu.ru/~bezvita/term1/Практикум_13.xlsx

БЛАГОДАРНОСТИ

Выражаю благодарность преподавателям информатики ФББ МГУ за объяснение по написанию данного миниобзора.

СПИСОК ЛИТЕРАТУРЫ

[1]Геном:

<https://www.ncbi.nlm.nih.gov/nuccore/982535841>

[2]Хромосомная таблица:

ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001/439/585/GCA_001439585.2_ASM143958v2/