



Мини-обзор

Анализ протеома бактерии *Kosakonia arachidis* (*Enterobacter arachidis*)

Тедеев А. А.

Факультет биоинженерии и биоинформатики, Московский государственный университет имени М.В. Ломоносова, Ленинские горы д.1 стр.73, 119234, Москва, Российская Федерация

11.12.2020.

РЕЗЮМЕ

В данной работе представлен мини-обзор протеома бактерии *Kosakonia arachidis*, выполненный с помощью методов работы с электронными таблицами. В ходе работы был проанализирован нуклеотидный состав генома данной бактерии, рассчитано количество различных типов генов,

КЛЮЧЕВЫЕ СЛОВА: геном, протеом, бактерия, KOSAKONIA ARACHIDIS KACC 18508, MS Excel, электронные таблицы.

1 ВВЕДЕНИЕ

Большинство бактерий рода *Kosakonia* являются симбионтами и комменсалами растений, также встречаются виды, условно патогенные для человека. Геном штамма KACC 18508 вида *Kosakonia arachidis*, который был выделен из почвы 17 мая 2019 года на территории Южной Кореи, был зарегистрирован при поддержке *Kyungpook National University* 27 октября 2019 года. *Kosakonia arachidis* является аэробной мезофильной подвижной бактерией из семейства Enterobacteriaceae. Размеры бактерии находятся в пределах от 2-3.7 μm в длину и 0.5-0.7 μm в ширину. Ее ближайшими родственниками являются *Kosakonia radicincitans* и *Kosakonia cowanii* (*Madhaiyan et al., 2010*). Из *Kosakonia arachidis* KACC 18508 выделены 1 хромосома и 1 плазида. Геном содержит 4977 генов.

2 МАТЕРИАЛЫ И МЕТОДЫ

Хромосомная таблица извлечена из базы данных NCBI и обработана с помощью программы Microsoft Excel. Таблица с импортированными данными была отформатирована: были созданы листы *genes* со списком генов и информацией о них и *gene2* (прочее) (строки были отсортированы в лексикографическом порядке). Навигация по папкам осуществлялась с по-

мощью программы *Far Manager* (перенос файлов). Также в работе использовались горячие клавиши для копирования и вставки больших объемов данных и поиска-замены совпадений (Ctrl+F, Ctrl+H, Ctrl+C, Ctrl+V, Ctrl+A, Ctrl+X). Также в работе с геномом использовались сценарии, написанные на языке программирования *Python*.

В работе с Microsoft Excel использовались такие методы, как:

- создание плоских таблиц;
- сортировка строк и столбцов;
- создание примечания к ячейке;
- построение гистограмм по данным столбца;
- распространение формул;
- уничтожение формул;
- «фильтрация» значений;
- перенос данных между листами и т.д.

Использовались следующие формулы и функции:

- ВВП (для сопоставления данных из разных листов таблицы);
- СЧЁТЕСЛИ, СЧЁТЕСЛИМН;
- ОКРУГЛ и т.д.

Количество единичных нуклеотидов и k-меров подсчитано с помощью команды *wordcount*, GC-состав – с помощью команды *geecee* (применялся клиент *PuTTY* для протокола удаленного доступа *SSH*).

3 РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЯ

3.1 Общая характеристика генома

Геном *Kosakonia arachidis* KACC 18508 представлен одной кольцевой хромосомой и одной плазмидой. Размер хромосомы составляет 5096793 п.о., размер плазмиды – 79613 п.о. Это довольно крупный геном, так как среднее значение и медиана составляют ~3,65 Mb и ~3,46 Mb соответственно (*diCenzo et al., 2017*). По данным записи NZ_CP045299.1 базы данных *RefSeq NCBI* в геноме *Kosakonia arachidis* закодировано 4977 генов. Из них 4710 кодируют белки, 113 кодируют РНК, 154 – псевдогены.

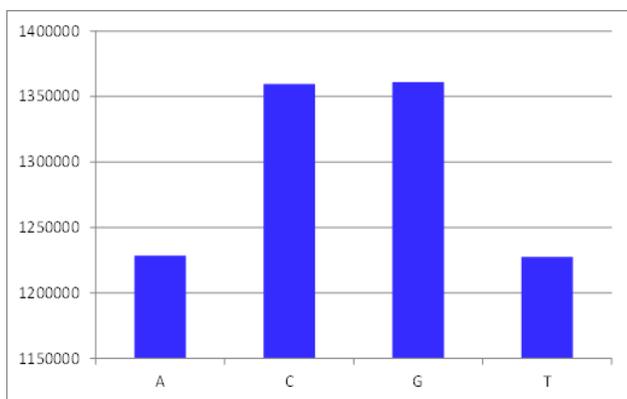
Таблица 1. Количество генов по типам.

Ген	Число
proteins	4710
pseudogenes	154
tRna	83
rRna	22
tmRNA	1
SRP_RNA	1
RNase_P_RNA	1
ncRNA	4
antisense_RNA	1

3.2 Соотношение нуклеотидов и GC-состав

В ходе анализа было установлено, что количества нуклеотидов А и Т, G и С равны, что подтверждает эмпирическое наблюдение Э. Чаргаффа.

Рисунок 1. Соотношение нуклеотидов в геноме.



Результаты исследования генома на GC-состав приведены в таблице 2. Относительно высокая доля гуанина и цитозина в ДНК согласуется со способностью *Kosakonia arachidis* усваивать молекулярный азот (согласно гипотезам некоторых исследователей, содержание GC у азотфиксирующих представителей выше [4]).

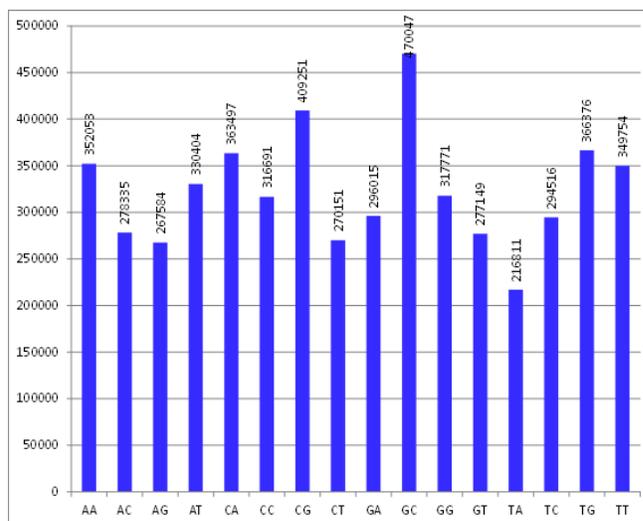
Таблица 2. GC-состав хромосомы и плазмиды.

Sequence	GC-content
NZ_CP045300.1	0.53
NZ_CP045299.1	0.48

3.3 Анализ k-меров

Количество каждого динуклеотидного k-мера представлено на рисунке. Самыми часто встречающимися «расстановками» оказались GC и CG, что согласуется с результатами расчета количества отдельных нуклеотидов [3.2].

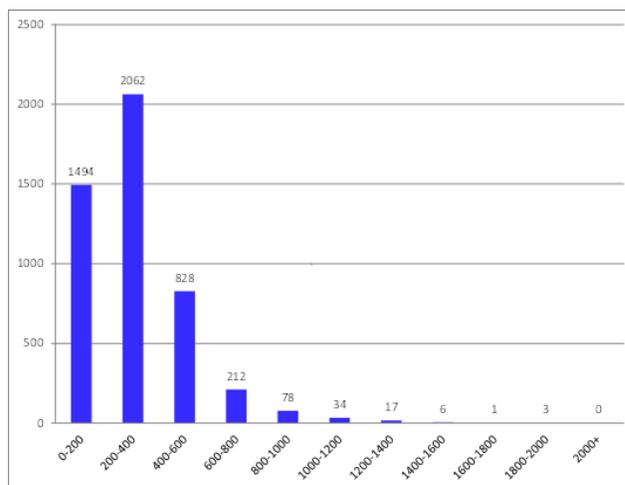
Рисунок 2. Соотношение k-меров длины 2.



3.4 Распределение белков по длинам

Наибольшее количество белков находится в пределах от 200 до 400 аминокислот.

Рисунок 3. Гистограмма длин белков



БЛАГОДАРНОСТЬ

Выражаю признательность преподавателям курса практической информатики факультета биоинженерии и биоинформатики МГУ за их роль в формировании студентов как высококлассных специалистов, а также за полученные знания и навыки.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

https://kodomofbb.msu.ru/~arsen/term1/Tedeev_supple.xlsx
https://kodomofbb.msu.ru/~arsen/term1/Data_table.xlsx

СПИСОК ЛИТЕРАТУРЫ

1. Munusamy Madhaiyan et al. (2010) *Enterobacter arachidis* sp. nov., a plant-growth-promoting diazotrophic bacterium isolated from rhizosphere soil of groundnut. *International Journal of Systematic and Evolutionary Microbiology*, 60, 1559–1564
2. Carrie Brady et al. (2013) Taxonomic evaluation of the genus *Enterobacter* based on multilocus sequence analysis (MLSA). *Systematic and Applied Microbiology*, 36(5):309-319
3. diCenzo et al. (2017) The divided bacterial genome: structure, function, and evolution. *Microbiol Mol Biol Rev*. 81: e00019–17–17. pmid:28794225
4. McEwan et al. (1998) Nitrogen-fixing aerobic bacteria have higher genomic GC content than non-fixing species within the same genus. *Hereditas*. 128: 173-178. Lund, Sweden. TSSN 0018-0661. Received July 7, 1997. Accepted March 4, 1998
5. [Страница генома *Kosakonia arachidis*](#)
6. [Сайт с характеристиками бактерии](#)