
Краткий обзор генома и протеома бактерии *Basilea psittacipulmonis* DSM 24701

Мухамедияров Рауль Маратович.

Факультет биоинформатики и биоинженерии, МГУ им. Ломоносова.

РЕЗЮМЕ

В настоящей работе было проанализировано соотношение нуклеотидов в геноме бактерии *Basilea psittacipulmonis*. Также рассмотрены типы генов и их соотношение. Были сделаны выводы о соблюдении второго правила Чаргаффа, GC-сдвиге, распределении длин белков и др.

КЛЮЧЕВЫЕ СЛОВА

Геном, протеом, белок, бактерия *Basilea psittacipulmonis* DSM 24701

ВВЕДЕНИЕ

Basilea psittacipulmonis была собрана с легкого мертвого попугая (*Melopsittacus undulates*) из зоомагазина города Базель, Швейцария в 2007 году[1].

Позже стало известно, что это грамотрицательная микроаэрофильная мезофильная (развивающаяся при 30-37 °C) бактерия[2]. По предварительной оценке не является патогенной[1].

В данном обзоре рассматриваются особенности генома и протеома *B. psittacipulmonis* с использованием простейших биоинформатических методов. Работа может положить начало другим исследованиям этой бактерии, которые позволят точно убедиться в том, что *B. psittacipulmonis* не наносит вреда попугаям.

Домен	Bacteria
Отдел	Proteobacteria
Класс	Betaproteobacteria
Порядок	Burkholderiales
Семейство	Alcaligenaceae
Род	Basilea
Вид	<i>Basilea psittacipulmonis</i> Whiteson et al. 2014

Таблица 1. Систематическое положение *Basilea psittacipulmonis*[2],[4].

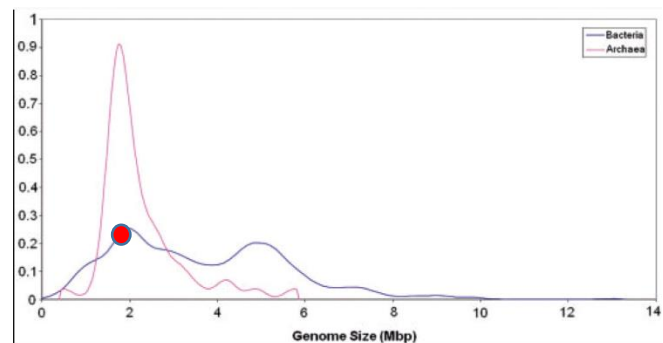
МАТЕРИАЛЫ И МЕТОДЫ

Геном (в формате .fasta) и таблица генов были скачаны с сайта NCBI[3]. При работе над обзором активно использовались электронные таблицы(ЭТ) Google sheets (функции: COUNTIFS, UNIQUE, ROUND, AVERAGE, STDEV, MEDIAN и др., диаграммы и некоторые таблицы). С помощью Python была написана программа для расчетов длины генома и соотношения нуклеотидов в нем (/home/students/y20/raul.mu-rov/term1/block3/overview/calc.py).

РЕЗУЛЬТАТЫ

1. Общие сведения

Согласно данным NCBI весь доступный геном *B. Psittacipulmonis* расположен только в хромосоме[3]. Размер генома составил 1 959 069 пар оснований(bp), что соответствует среднему размеру геномов бактерий – 2-5 Mbp(10^6 bp) [6].



2. Соотношение нуклеотидов.

Согласно полученным данным в исследуемом геноме количество нуклеотидов $A \approx T$, $G \approx C$. Это подтверждает второе правило Чаргаффа и принцип симметрии[7].

Основание	Содержание
A	29.7%
T	30.0%
G	20.1%
C	20.1%

Таблица 2. Относительное содержание нуклеотидов в геноме объекта.

Также по полученным данным доля GC-пар составляет **40.3%**. Как известно, повышенная доля GC-пар в ДНК увеличивает ее устойчивость к высоким температурам[5]. *B. Psittacipulmonis* – мезофил, поэтому GC-сдвиг мы не наблюдаем.

3. Типы генов и их соотношение

Электронную таблицу с обработкой генома можно найти по [ссылке](#).

Для подсчетов гены были разделены на 8 групп (таблица 3). Согласно полученным данным большинство генов бактерии кодируют белок. Также было подсчитано сколько генов расположено на «+»-цепи и сколько на «-»-цепи: 827 и 872, соответственно. Разница между полученными числами достаточно невелика, поэтому можно заключить, что гены расположены на двух цепях равномерно. Средняя длина гена у *B. Psittacipulmonis* составляет 1024 bp, медианное значение 840 bp. Максимальная длина у гена белка из семейства YadA, 13155 bp; а минимальная у гена цистеиновой tPHK, 74 bp.

4. Распределение белков по длинам

Тип гена	Число генов	%
protein_coding	1639	96,47
pseudogene	9	0,53
tRNA	39	2,3
rRNA	9	0,53
SRP_RNA	1	0,06
tmRNA	1	0,06
RNase_P_RNA	1	0,06
Total	1699	100

Таблица 3. Типы генов, их число и доля от общего числа.

Protein_coding – ген, кодирующий белок. Pseudogene – ген, переставший экспрессироваться. tRNA – транспортная РНК. rRNA – рибосомальная РНК. SPR_RNA – сигнальная РНК. tmRNA – транспортно-матричная РНК. RNase_P_RNA – рибонуклеаза Р РНК(рибозим). Total – общее число

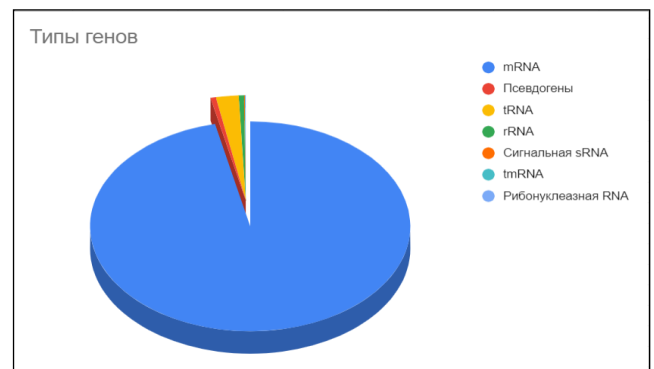
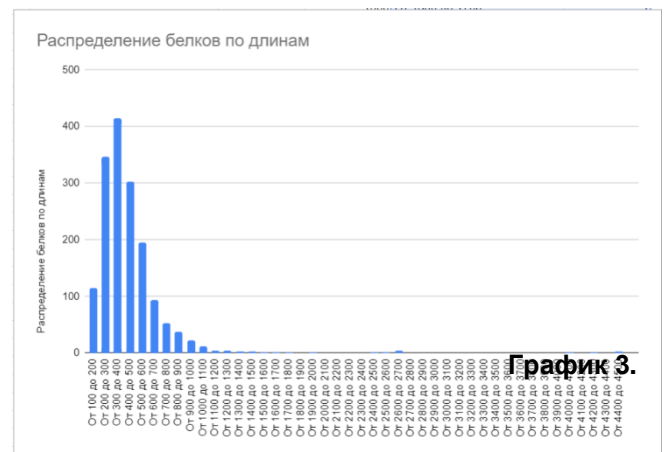


Диаграмма 1. Соотношение типов генов в геноме.



В ходе работы были рассмотрены белки бактерии. Среди 1639 молекул чаще всего встречаются белки с длиной цепи от 300 до 400 аминокислот (график 3). На графике наблюдается всего один крутой пик в указанной области. Подавляющее большинство генов имеет меньше 1000 аминокислотных остатков в своем составе.

Peter Yakovchuk, Ekaterina Protozanova and Maxim D. Frank-Kamenetskii, *Nucleic Acids Research*, 2006, Vol. 34, No. 2

6. Статья о размерах геномов прокариот <https://academic.oup.com/nar/article/36/21/6688/2410005>
7. Статья о принципе симметрии Prabhu VV (1993). "Symmetry observation in long nucleotide sequences". *Nucleic Acids Research*. **21** (12): 2797–2800. doi:10.1093/nar/21.12.2797. PMC 309655. PMID 8332488

ЗАКЛЮЧЕНИЕ

В ходе анализа генома были подтверждены некоторые закономерности его устройства.

Следует суммировать наиболее важные выводы:

- 1) Геном *B. Psittacipulmonis* имеет среднюю длину
- 2) Соотношение нуклеотидов подтверждает правило Чаргаффа и соответствует условиям обитания данной бактерии (мезофил)
- 3) Гены распределены практически равномерно по «+» и «-» цепям.

БЛАГОДАРНОСТИ

Автор выражает благодарность своей бабушке, которая приготовила ему чай во время работы над обзором.

ЛИТЕРАТУРА И ССЫЛКИ

1. *B. Psittacipulmonis* в немецкой коллекции микроорганизмов при Лейбницком институте: Leibniz Institute DSMZ, German Collection of Microorganisms and Cell Cultures GmbH - *Basilea psittacipulmonis* DSM 24701
2. Данные BacDrive о *B. psittacipulmonis*: <https://bacdrive.dsmz.de/search?search=Basilea+psittacipulmonis+&submit=>
3. Геном и протеом *B. psittacipulmonis* на NCBI: ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/743/945/GCF_000743945.1_ASM74394v1
4. Исследование генома *B. psittacipulmonis*: A genomic perspective on a new bacterial genus and species from the Alcaligenaceae family, *Basilea psittacipulmonis* (BMC Genomics. 2014 Mar 1;15:169. doi: 10.1186/1471-2164-15-169).
5. Исследование о влиянии температуры и солености на процент GC пар в геноме: Base-stacking and base-pairing contributions into thermal stability of the DNA double helix