

# Мини-обзор генома бактерии *Campylobacter cuniculorum*

Автор: Аркуша Вероника Евгеньевна<sup>1</sup>

<sup>1</sup>Факультет биоинженерии и биоинформатики, Московский Государственный Университет имени М.В.Ломоносова

Контактные данные: [arkushaver@mail.ru](mailto:arkushaver@mail.ru)

## Аннотация

Данный мини-обзор является анализом генома и протеома грамотрицательной бактерии *Campylobacter cuniculorum* с использованием функционала электронных таблиц и программирования на языке Python. Обзор носит описательный характер, результаты по возможности подкрепляются биоинформатическими закономерностями.

## 1 ВВЕДЕНИЕ

Таксономия: Bacteria; Proteobacteria; delta/epsilon subdivisions; Epsilonproteobacteria; Campylobacterales; Campylobacteraceae; *Campylobacter*; *Campylobacter cuniculorum*

Род *Campylobacter* был описан в ещё в 1963 году, однако подтвержденных данных о рассматриваемом нами виде не так много.

Представляет собой грамотрицательную бактерию спиралевидной формы, которая обычно обнаруживается в желудочно-кишечном тракте кроликов. Известно, что это зоонозный патоген, что означает, что он может передаваться от животных к человеку и был связан с гастроэнтеритом как у кроликов, так и у людей.

Этот штамм был изучен на предмет его устойчивости к антибиотикам, факторов вирулентности и генетических характеристик, чтобы лучше понять его потенциальное воздействие на общественное здравоохранение. Исследования показали, что *Campylobacter cuniculorum* DSM 23162 = LMG 24588 обладает способностью прилипать к эпителиальным клеткам кишечника и проникать в них, что может способствовать его патогенности.

Как и в случае с другими видами кампилобактерий, заражение *Campylobacter cuniculorum* DSM 23162 = LMG 24588 обычно связано с потреблением загрязненной пищи или воды, особенно недоваренной птицы или непастеризованного молока. Симптомы инфекции у людей могут включать диарею, боль в животе, лихорадку и рвоту.

В целом, изучение *Campylobacter cuniculorum* DSM 23162 = LMG 24588 важно для понимания потенциальных рисков, связанных с зоонозной передачей этой бактерии, и для разработки стратегий предотвращения и контроля ее распространения.

## 2 Материалы и методы

Данные по геному исследуемой бактерии были взяты с официального сайта NCBI (Национальный Центр

Биотехнологической информации). Для анализа полученных данных были использованы электронные таблицы Google Sheets, программа написанная на языке программирования Python.

## 3 РЕЗУЛЬТАТЫ

### 3.1 Описание стандартных данных о геноме

Геном *Campylobacter cuniculorum* включает одну кольцевую хромосому и две небольшие кольцевые плазмиды (Таблица 1) [2]. GC состав довольно низкий (31%) [2], поэтому нельзя говорить, что рассматриваемая нами бактерия сильно устойчива к денатурации в растворе (между А и Т двойная водородная связь, а между G и C - тройная), ей это и не нужно, так как средняя температура места ее обитания (ЖКТ) - ~37-38 °C

Таблица 1. Стандартные данные о геноме

ДНК	Длина (п.н.)	GC состав
Хромосома	1931011	31 %
Плазмида pCUN1	4923	28,5 %
Плазмида pCUN2	1834	35 %

С помощью fasta-файла[3] последовательности генома и программы, написанной на Python[5] можно определить нуклеотидный состав ДНК (Таблица 2). Данные показывают, что для исследуемой бактерии выполняется второе правило Чаргаффа – число букв А примерно равно числу букв Т (49.81% и 50.19% соответственно от суммы А+Т), а число букв G примерно равно числу букв C (49,9% и 50,1% соответственно от суммы G+C) в последовательности одной цепочки геномной ДНК.

\*файлы, которые использовались для написания программы были распакованы с помощью командной строки, они лежат в папке term1 на kodomo

Таблица 2. Нуклеотидный состав геномных ДНК

ДНК	A	T	G	C
Хромосома	661486	667731	300298	301496
Плазмида pCUN1	21215	1395	781	626
Плазмида pCUN2	610	586	335	303
Всего	664127	669172	301414	302425

\*Небольшое предположение: Плазмиды, как правило, синтезируют белки, антигены, расщепляют некоторые

соединения.. Мы видим, что плазида pCUN2 имеет не очень большой нуклеотидный состав, следовательно она выполняет свою основную задачу не столь часто, как pCUN1. Исходя из того, что основное место обитания нашей бактерии ЖКТ кролика, можно предположить, что ее функция - устойчивость к тяжелым металлам, антибиоткам, так как вряд ли животное каждый день употребляет пищу с содержанием вышеперечисленных веществ, однако риск попадания их в организм остается

### 3.2 Статистические данные о белках протеома

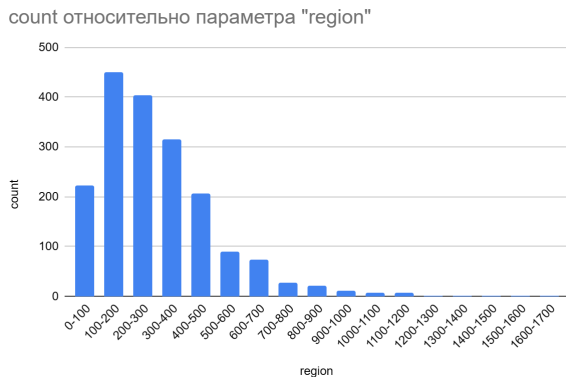
Были проанализированы количества и процентные содержания таких групп белков: рибосомальные, гипотетические и транспортные – по данным CDS [9, list "proteins\_CDS"] таблицы особенностей генома [4] (Таблица 3). Среди рибосомальных белков нет генов с одинаковыми названиями [6, list "ribosomal"]].

**Таблица 3.** Данные о белках избранных групп

Белки	Количество	Процент от всех белков
Рибосомальные белки	55	2,93 %
Гипотетические белки	383	20,4 %
Транспортные белки	104	5,55 %

\*Рассуждение: Изначально может показаться странным такое большое количество гипотетических белков. Однако это последовательности аминокислот, которые, как предполагается, кодируются геном, но не были охарактеризованы экспериментально. Эти белки могут быть идентифицированы с помощью секвенирования генома и биоинформатического анализа, но их функции и роль в биологических процессах неизвестны. Для определения функции и свойств гипотетического белка, подобного "pCUN2" и "pCUN1", потребуются дальнейшие исследования и эксперименты.

В электронных таблицах была построена гистограмма длин белков (Диаграмма 1) и подсчитаны некоторые статистические параметры для этого распределения (Таблица 4) [6, list "protein\_len-hist"].



**Диаграмма 1.** Гистограмма для белков

**Таблица 4.** Статистические параметры распределения длин белков

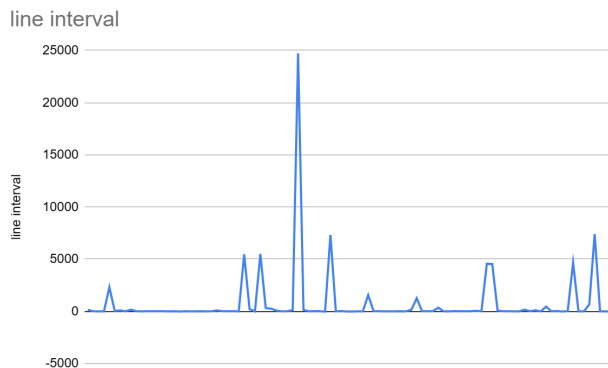
Средняя длина	108,5
Стандартное отклонение	152,5
Медиана	22
Минимальное значение	1
Максимальное значение	451

С помощью электронных таблиц [6, list "tables"] было установлено распределение белок-кодирующих генов по цепям ДНК (Таблица 5). По рассчитанной статистической значимости мы можем сделать вывод, что гены по цепям распределяются случайно (уровень значимости >0.01, для всех ДНК)

**Таблица 5.** Распределение белок-кодирующих генов

ДНК	«+» цепь	«-» цепь	Статистическая значимость
Хромосома	913	953	0,3666170
Плазида pCUN1	5	0	0,0625000
Плазида pCUN2	1	3	0,6250000
Всего	919	956	0,4057630

В электронных таблицах была построена гистограмма расстояний между генами белков на "+" цепи (Диаграмма 2) [6, list "protein\_len-hist"], т.е. из конца первого вычиталось начало второго и т.д. Это позволяет нам увидеть насколько далеко или близко друг от друга расположены гены на этой цепи.



**Диаграмма 2.** График расстояний между генами белков на "+" цепи

Примечание\* график выше не включает в себе все расстояния, так как если мы возьмем во внимание все гены белков (их 914), нам тяжело будет увидеть пики; поэтому график построен лишь для первых ста.

### 3.3 Статистические данные о генах РНК

Всего в геноме 49 генов РНК [6, list "per-replicon"], что составляет 2,61% от всех генов. Генов тРНК – 40, рРНК – 6 (Таблица 6). В геноме представлено по два гена каждого типа рРНК, образующих основу рибосом прокариот (16S, 23S). Вероятно, они представлены в одинаковой кратности для соотношения их продуктов 1:1, ведь в рибосоме присутствуют по одной молекуле каждого типа рРНК.

\*Рассуждения: Обратив внимание на таблицу ниже, можно заметить, что доля тРНК значительно больше остальных. Каждая клетка содержит множество различных транспортных

РНК, которые не только выполняют свою функцию транспорта аминокислот к рибосомам, но также участвуют в множестве других клеточных процессов. Транспортные РНК также имеют разнообразные структуры и специфические последовательности, что обуславливает их большое количество в геноме.

Хочется также обратить внимание на нкРНА: некодирующие РНК играют важную роль в бактериальных клетках. У бактерий существует несколько классов нкРНК каждый из которых выполняет различные функции.

Например, рибосомная РНК и транспортная РНК не являются кодирующими, но могут быть рассмотрены как нкРНК, которые присутствуют в клетках бактерий и играют ключевую роль в синтезе белка.

Есть также другие классы нкРНК, такие как микроРНК (miRNA) и сРНК (small RNA), которые участвуют в регуляции экспрессии генов, обеспечивая бактериям адаптацию к изменяющимся условиям окружающей среды.

Таким образом, нкРНК играют важную роль в бактериальной клетке, осуществляя различные функции, такие как участие в синтезе белков, регуляция экспрессии генов и адаптация к окружающей среде.

**Таблица 6.** Данные о РНК избранных групп

РНК	Количество	Процент от всех РНК
тРНК	40	81,6 %
рРНК	6	12,2 %
тмРНК	1	2,04%
нкРНК	2	4,08%

### 3.4 Дополнительные сведения и наблюдения

Гликолиз у *Campylobacter* spp. неполноный; следовательно, эти организмы не могут использовать сахара в качестве источника углерода, за исключением фукозы для некоторых штаммов. (в рамках обзора мы не будем углубляться в особенности этих штаммов-исключений, но в сопроводительных материалах я оставила на них ссылки [7-8]). При изучении последовательности генома можно выделить локус - *tsuABC*, который участвует в утилизации трикарбаллилатов.

### Сопроводительные материалы

1. Данные NCBI по геному бактерии  
[index of /genomes/all/GCF/002/104/335/GCF\\_002104335.1\\_ASM210433v1.nih.gov](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/104/335/GCF_002104335.1_ASM210433v1.nih.gov)
2. Описание генома с NCBI.  
[https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/104/335/GCF\\_002104335.1\\_ASM210433v1/GCF\\_002104335.1\\_ASM210433v1\\_assembly\\_stats.txt](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/104/335/GCF_002104335.1_ASM210433v1/GCF_002104335.1_ASM210433v1_assembly_stats.txt)
3. Fasta-файл генома с NCBI.  
[https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/104/335/GCF\\_002104335.1\\_ASM210433v1/GCF\\_002104335.1\\_ASM210433v1\\_genomic.fna.gz](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/104/335/GCF_002104335.1_ASM210433v1/GCF_002104335.1_ASM210433v1_genomic.fna.gz)
4. Таблица особенностей генома с NCBI.  
[https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/104/335/GCF\\_002104335.1\\_ASM210433v1/GCF\\_002104335.1\\_ASM210433v1\\_feature\\_table.txt.gz](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/104/335/GCF_002104335.1_ASM210433v1/GCF_002104335.1_ASM210433v1_feature_table.txt.gz)
5. Программа, написанная на Python, для определения нуклеотидного состава ДНК.

[https://colab.research.google.com/drive/1UDq389r\\_wfiZzpLXAbJfytemTz2uXq4V#scrollTo=zaHtYifMKz2](https://colab.research.google.com/drive/1UDq389r_wfiZzpLXAbJfytemTz2uXq4V#scrollTo=zaHtYifMKz2)

6. Таблицы Google Sheets. [bratzveron genome - Google Drive](#)
7. Muraoka WT, Zhang Q. 2011. Phenotypic and genotypic evidence for L-fucose utilization by *Campylobacter jejuni*. *J Bacteriol* 193:1065–1075. [Phenotypic and Genotypic Evidence for L-Fucose Utilization by Campylobacter jejuni | Journal of Bacteriology \(asm.org\)](#)
8. Stahl M, Friis LM, Nothhaft H, Liu X, Li J, Szymanski CM, Stintzi A. 2011. L-Fucose utilization provides *Campylobacter jejuni* with a competitive advantage. *Proc Natl Acad Sci USA* 108:7194–7199.

### Список литературы

1. Complete genome sequence of the *Campylobacter cuniculorum* type strain LMG 24588 Miller, W.G. 2017 Miller, W G, Yee, E, Revez, J, Bono, J L Rossi, M 2017, ' Complete genome sequence of the *Campylobacter cuniculorum* type strain LMG 24588 ', *Genome Announcements*, vol. 5, no. 24, e00543-17.
2. Zanoni RG, Debruyne L, Rossi M, Revez J, Vandamme P. 2009. *Campylobacter cuniculorum* sp. nov., from rabbits. *Int J Syst Evol Microbiol* 59:1666–1671
3. Lewis JA, Horswill AR, Schwem BE, Escalante-Semerena JC 2004. The tricarballylate utilization (*tsuRABC*) genes of *Salmonella enterica* serovar Typhimurium LT2. *J Bacteriol* 186:1629–1637.
4. Статья в Википедии "Кампилобактерии": [Кампилобактерии — Википедия \(wikipedia.org\)](#)
5. Н.Н. Скворцова "ОСНОВЫ МОЛЕКУЛЯРНОЙ БИОЛОГИИ" Учебное пособие