

Особенности генома *Corynebacterium deserti* GIMN1.010

Музалевская Юлия^{1,*}

¹ Факультет биоинженерии и биоинформатики Московского Государственного Университета *deserti*

* uliamuzalevskaya@gmail.com

Резюме: В данной работе представлен миниобзор генома *Corynebacterium deserti* GIMN1.010 – штамма бактерий из рода Коринебактерии, выделенного из песка пустыни в Китае. Количественный и качественный анализ генома и протеома был проведён с использованием методов электронных таблиц и языка программирования Python.

1. Введение

Штамм *Corynebacterium deserti* GIMN1.010 был выделен из пробы смешанного песка из пустыни западного Китая. Его выращивали в среде TGY (0,5% пептона, 0,5% дрожжевого экстракта, 0,1% глюкозы, pH доводили до 6,8) при 30 °C при встряхивании 200 об/мин в течение 2 дней с последующим выделением лимонно-жёлтых колониеобразующих бактерий на чашечках в среде TGY с 1,5% агара. Физиологическая характеристика и дополнительные биохимические тесты были выполнены в Институте микробиологии, провинция Гуандун, Китай.

Клетки штамма GIMN1.010 представляют собой положительные по Грамму, не образующие спор, короткие палочки неправильной формы. Полученные в среде TGY колонии были лимонно-жёлтыми, круглыми, непрозрачными, выпуклыми с правильными краями и диаметром примерно 0,5 мм. Изолят был положительным по каталазе и уреазе, но отрицательным по оксидазе и тирозиназе. Штамм GIMN1.010 был способен использовать ряд субстратов (таких как глюкоза, d-ксилоза, сахароза и салицин) в качестве единственных источников углерода для роста, но не желатин или казеин. Оптимальные pH и температура культивирования были pH 7 и 30 °C.[1]

2. Материалы и методы

Информация о геноме бактерии была взята из базы данных NCBI Genome. Хромосомная таблица находится в файле Genome features, последовательность – в файле Genome.txt. Таблица с результатами исследований находится в файле Minireview, использованные коды – в файле Python. В качестве методов использовались электронные таблицы Google Sheets:

- Различные функции, с помощью которых был проведён количественный и качественный анализ протеома и генома
- Диаграммы и гистограммы для наглядного изображения результатов анализа и язык программирования Python для разбора нуклеотидной последовательности.

3. Результаты

3.1. Стандартные данные о геноме *Corynebacterium deserti*

3.1.1. Число и названия ДНК, составляющих геном

У *Corynebacterium deserti* GIMN1.010 3 молекулы ДНК: основная кольцевая хромосома и две плазмиды – pCdes1 и pCdes2

- Общее количество пар нуклеотидов: 3 033 893;

- Основная кольцевая хромосома состоит из 2 972 149 п.н.;
- pCdes1 – 31 005 п.н.;
- pCdes2 – 30 739 п.н.

3.1.2. GC-состав каждой ДНК

Общий GC-состав генома 55,25%; на хромосоме 55,27%; на pCdes1 52,86%; на pCdes2 55,18%

3.2. Статистические данные о белках протеома

3.2.1. Гистограмма длин белков

Всего геном *Corynebacterium deserti* кодирует 2 697 белков, из них 50 синтезируются с плазмид, а остальные 2 647 – с хромосомы. Длины белков находятся в промежутке от 17 до 3 008 аминокислот, среднее значение составляет 318 аминокислот. Из гистограммы видно, что больше всего белков находится в диапазоне 210-240 аминокислот. Также наблюдается пик, соответствующий длинам 570-600 аминокислот. Индивидуальные распределения длин белков показывают пики при определённых длинах белков, которые, по-видимому, соответствуют семействам генов с высокой степенью дублирования в их соответствующих линиях.[2]



Рисунок 1. Гистограмма длин белков. По горизонтали отложены промежутки, соответствующие длинам белков; по вертикальной оси – количество белков длиной, находящейся в данном промежутке.

3.2.2. Число генов белков, закодированных на прямой и комплементарной цепочке

Всего у бактерии 2 697 генов, кодирующих какой-либо белок. Из них на прямой цепи находится 1 340 генов, а на комплементарной – 1 357. Расчёты находятся на листе Chain.

распределение генов генов по цепям

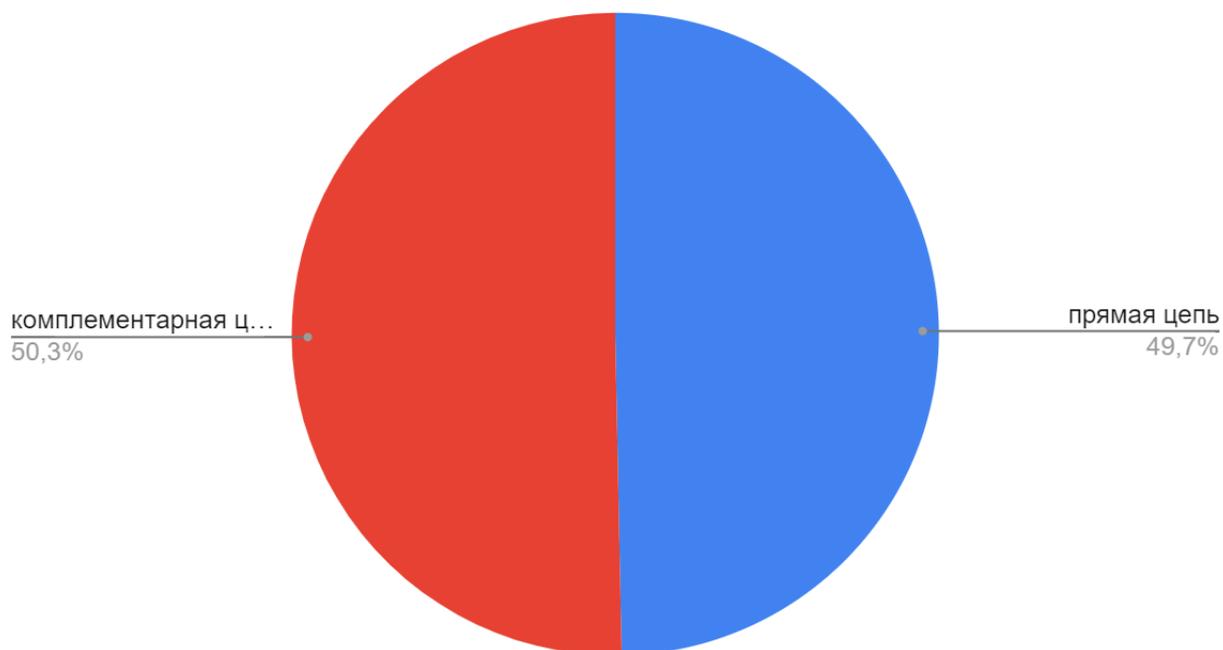


Рисунок 2. Диаграмма распределения генов по цепям. Синий – прямая цепь, красный – комплементарная цепь.

3.2.3. Число рибосомальных белков

Всего в протеоме бактерии 55 рибосомальных белков, то есть примерно 5%. Их перечень приведён на листе Ribosomal protein

3.2.4. Гипотетические белки

Количество гипотетических белков (hypothetical protein) – 461, что составляет 17,09% от числа всех белков бактерии. Перечень всех гипотетических белков их длиной, координатами начала и конца представлен на листе Hypothetical proteins.

3.2.5. Транспортные белки

Количество транспортных белков – 251, что составляет 9,3% от всех белков бактерии. Перечень всех транспортных белков с их характеристиками представлен на листе Transport proteins.

3.2.6. Обобщение

Далее приведены диаграммы распределения белков по типам на хромосоме и обеих плаزمидях. В сопроводительных материалах их можно найти на листе Protein diagram.

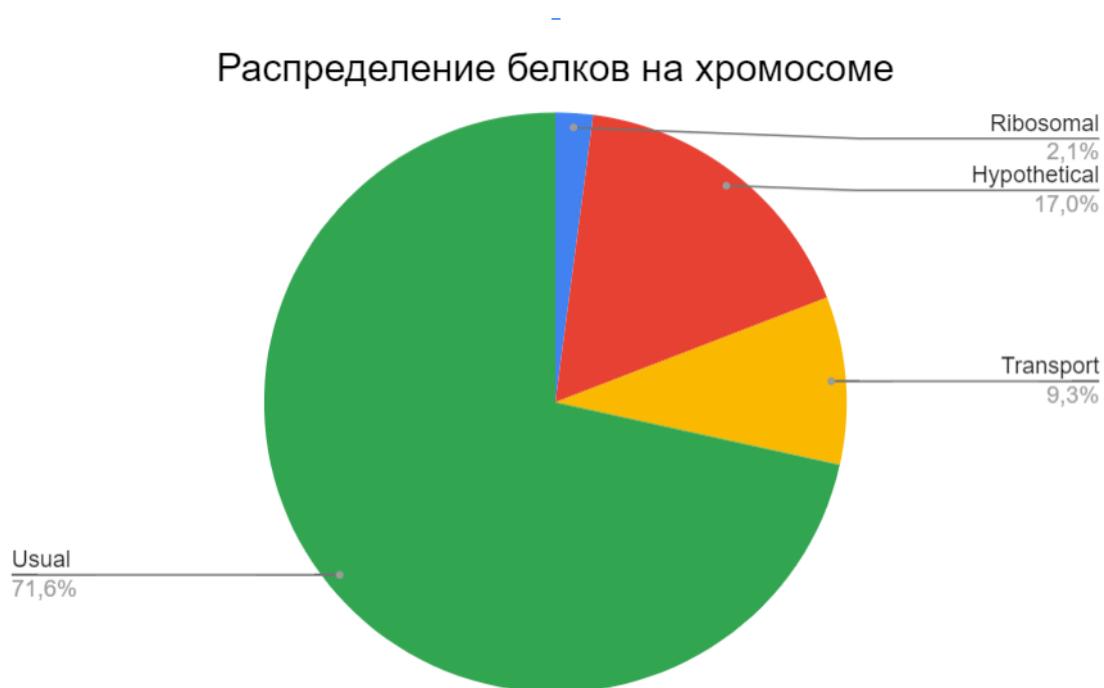


Рисунок 3. Диаграмма распределения белков по типам на хромосоме.

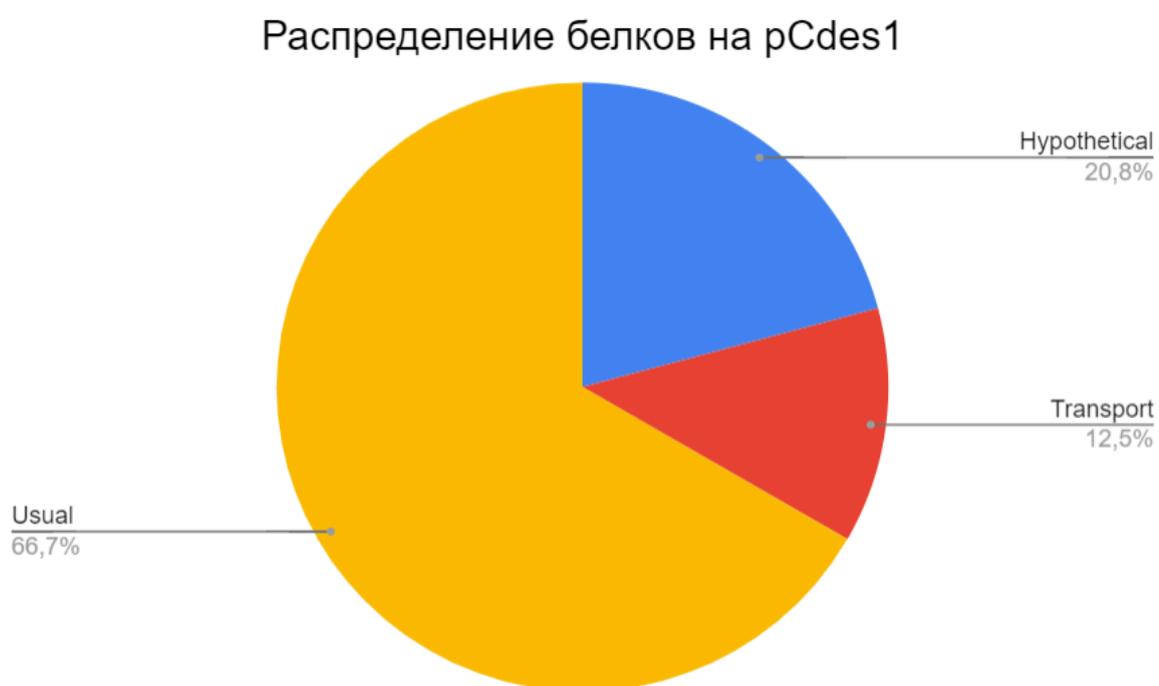


Рисунок 4. Диаграмма распределения белков по типам на pCdes1.

Распределение белков на pCdes2

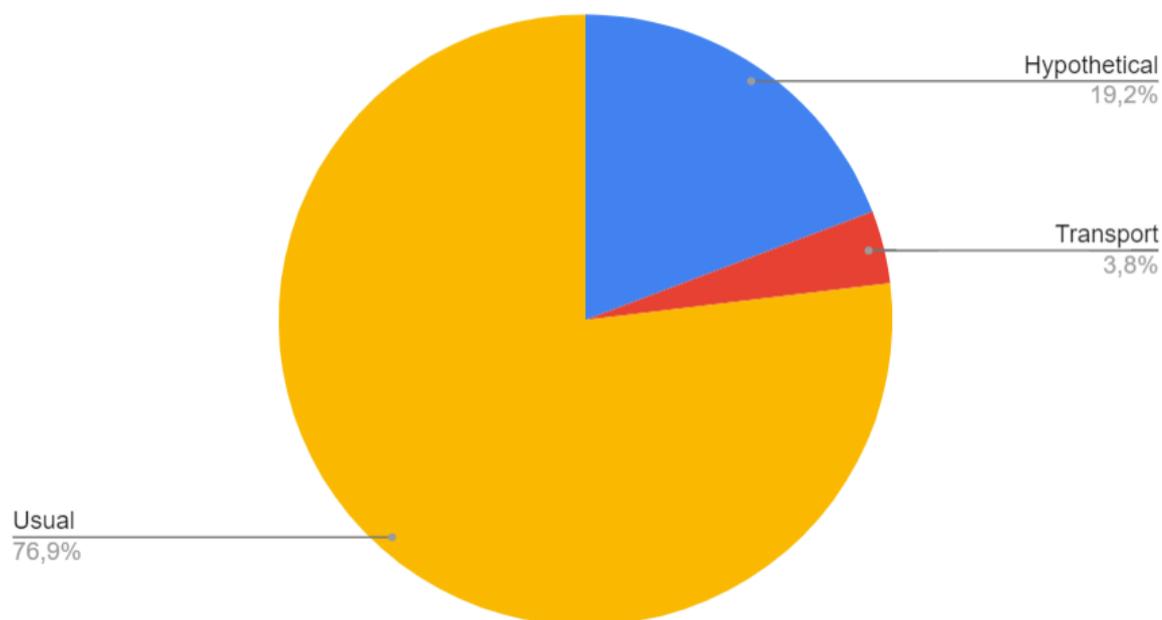


Рисунок 5. Диаграмма распределения белков по типам на pCdes2.

3.3. Статистические данные о генах РНК

3.3.1. Число генов РНК

В геноме *Corynebacterium deserti* 80 генов, кодирующих какую-либо РНК. Все эти гены сосредоточены на хромосоме. Генов, кодирующих РНК, значительно меньше, чем генов кодирующих белок (2 697).

3.3.2. Число рибосомальных РНК

Число рибосомальных РНК составляет 18, то есть 22,5% от всех генов, кодирующих РНК

3.3.3 Число транспортных РНК

Наибольшую часть всех РНК бактерии составляют транспортные РНК: их 59, что составляет 73,8%

3.3.4. Обобщение

Помимо приведённых рибосомальных и транспортных РНК у бактерии синтезируются:

1. 2 некодирующие РНК (ncRNA) – 2,5%.
2. 1 транспортно-матричная РНК (tmRNA) – 1,3%

Далее приведена диаграмма распределения всех РНК по типам, которую также можно найти в сопроводительных материалах на листе RNA.

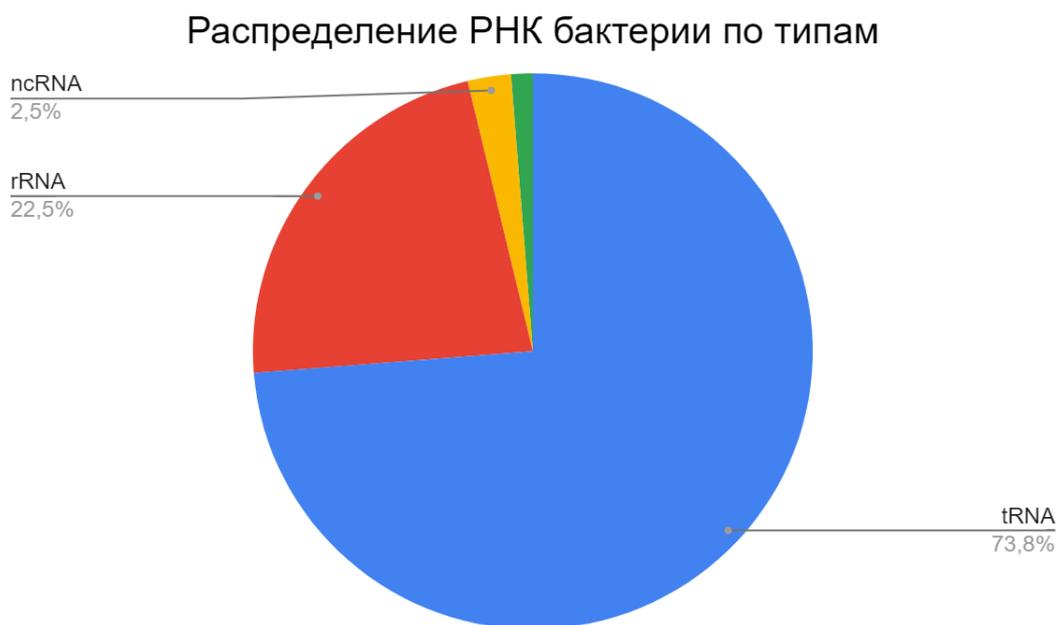


Рисунок 5. Диаграмма распределения РНК *Corynebacterium deserti* GIMN1.010 по типам.

3.4 Нуклеотидный состав геномных ДНК

Правило Чаргаффа гласит о том, что количество аденина на одной цепочке ДНК примерно равно количеству тимина, а гуанина – цитозину. То есть в общем количество пуринов примерно равно количеству пиримидинов ($A+G \approx T+C$). Проверим справедливость этого утверждения для изучаемой бактерии.

Таблица 1. Нуклеотидный состав прямой цепи хромосомы.

Нуклеотид	Количество	Процентное содержание
A	66 4502	22,36%
T	66 4740	22,37%
G	821 320	27,63%
C	821 587	27,64%

Исходя из таблицы, можно сделать вывод о том, что правило Чаргаффа выполняется для хромосомы *Corynebacterium deserti* GIMN1.010, так как и для пары А-Т, и для пары G-С разница в процентном соотношении составляет всего 0,01%.

Таблица 2. Нуклеотидный состав прямой цепи плазмиды pCdes1.

Нуклеотид	Количество	Процентное содержание
A	7 304	23,56%
T	7 311	23,58%
G	8 326	26,85%
C	8 064	26%

Исходя из таблицы, можно сделать вывод о том, что для плазмиды pCdes1 также выполняется правило Чаргаффа, но чуть менее точно, чем для хромосомы. Для пары А-Т разница в процентном соотношении составляет 0,02%, а вот для пары G-C более значительное число – 0,85%.

Таблица 3. Нуклеотидный состав прямой цепи плазмиды pCdes2.

Нуклеотид	Количество	Процентное содержание
A	6 995	22,76%
T	6 782	22,06%
G	8 288	26,96%
C	8 674	28,2%

Из третьей таблицы также видно менее строгое выполнение правила Чаргаффа. Однако для плазмиды pCdes2 разница в процентном соотношении заметна уже для обеих пар: А-Т – 0,7%, G-C – 1,24%

3.5. Оценка степени случайности распределения генов по двум цепочкам ДНК

На прямой цепи хромосомы находится 1 317 генов, на обратной – 1 330. Вероятность такого распределения или распределения с большей разностью составляет 40,78%. Это достаточно большая вероятность, а значит нет противоречия предположению о случайном распределении.

В отношении обеих плазмид можно сделать аналогичный вывод: для pCdes1 вероятность равна 27,06%, а для pCdes2 – 57,75%.

Сопроводительные материалы

1. Таблица Genome Features:
https://docs.google.com/spreadsheets/d/1d0KvoZbmRcD2qZu0XuJAcLrE_xLGvUcgZmCex2gtHA0/edit?usp=sharing
2. Genome.txt:
<https://drive.google.com/file/d/1B9R1zhIjkwSPBsJRGJaa523OtnWBDwO/view?usp=sharing>
3. Minireview:
<https://docs.google.com/spreadsheets/d/1hs8p7eAgi862TzuMSi0IY5kcxOLAU8c6nDEGJJd4uFs/edit?usp=sharing>
4. Python:
https://docs.google.com/document/d/1f6pNzOYTLGpCf6TxyF4p-MgzvxWRNoesjQHxYU6_3tk/edit?usp=sharing

Литература

1. Zhengfu Zhou, Menglong Yuan, Ran Tang, Ming Chen, Min Lin, Wei Zhang. *Corynebacterium deseri* sp. nov., isolated from desert sand. *International journal of systematic and evolutionary microbiology* **2012**.
2. Yannis Nevers, Natasha Glover, Christophe Dessimoz, Odile Lecompte. Protein length distribution is remarkably consistent across Life. *bioRxiv* **2021** doi: <https://doi.org/10.1101/2021.12.03.470944>