

Обзор генома бактерии *Acetobacterium woodii* DSM 1030

Додонова Анжелика¹

¹Факультет биоинженерии и биоинформатики Московского государственного университета

РЕЗЮМЕ

В данной работе был проведен анализ генома и протеома бактерии *Acetobacterium woodii* штамм DSM 1030. В ходе работы была подтверждена гипотеза случайного распределения генов по цепям ДНК, подсчитано количество квазиоперонов и пересечений генов, составлены гистограммы распределений длин белков, квазиоперонов и межгенных промежутков.

1 ВВЕДЕНИЕ

Acetobacterium woodii – анаэробная грамположительная бактерия, имеющая потенциально промышленное значение благодаря своему исключительному способу синтеза ацетата без побочных продуктов. Она может продуцировать на 50% больше ацетата, чем аэробные бактерии, используя только водород и углекислый газ. Поэтому увеличение экспрессии нужного фермента может превратить *A. Woodii* в более экономичную альтернативу бактериям, использующимся в современной биотехнологии. [1]

2 МЕТОДЫ

В работе использовалась программа Microsoft Excel 2016. Необходимые данные были получены из баз данных NCBI[2] и Uniprot[3].

3 РЕЗУЛЬТАТЫ

3.1 Распределение генов белков и РНК по категориям

В таблице 1 представлено распределение генов по категориям: гены рибосомальных, транспортных и гипотетических белков, всех остальных, а также гены тРНК и рРНК (другие РНК у бактерии не наблюдаются). В этой классификации псевдогены не учитывались. Примерное число генов на 1 млн. п.н. составило 880.

| категория | количество генов |
|----------------|------------------|
| рибосомальные | 58 |
| транспортные | 359 |
| гипотетические | 901 |
| другие | 2155 |
| тРНК | 59 |
| рРНК | 16 |

Таблица 1. Количественное распределение генов

¹To whom correspondence should be addressed.

3.2 Распределение длин белков в протеоме

После анализа протеома бактерии была построена гистограмма (рис. 1), отражающая распределение длин белков. Наибольшее количество белков лежит в диапазоне длины 101-400 а.к., но самый маленький белок имеет длину 22 а.к., а самый длинный – 2729 а.к. Средняя длина составила 330 а.к., медиана – 277 а.к., стандартное отклонение – 226 а.к., что говорит о высокой вариативности внутри протеома. Также следует отметить, что частота встречаемости белков длиннее 1300 а.к. чрезвычайно мала.



Рис. 1. Распределение длин белков в протеоме

3.3 Распределение длин белков в протеоме

Результаты анализа распределения генов белков и РНК, а также псевдогенов по прямой и обратной цепям ДНК представлены в таблице 2. Из таблицы можно предположить, что белок-кодирующие гены и псевдогены распределяются по цепочкам равномерно, тогда как генов РНК больше на комплементарной цепи, но причина этого феномена остается неясной. В работе была проверена гипотеза о том, что гены распределяются случайным образом, и она подтвердилась. Тем не менее, генов, как и квазиоперонов, незначительно больше на главной цепи.

| | белки | псевдогены | РНК |
|----------|-------|------------|-----|
| прямая | 1805 | 7 | 28 |
| обратная | 1668 | 6 | 47 |

Таблица 2. Распределение генов и псевдогенов по цепям ДНК

3.4 Поиск квазиоперонов

Последовательность генов считалась квазиопероном, если гены расположены друг на другом на одной цепи ДНК и расстояние между ними не превышает 100 п.н. Результаты приведены в таблице 3 и на рис. 2. Всего в геноме было обнаружено 1883 оперона. При увеличении длины квазиоперонов их встречаемость сначала резко снижается вплоть до достижения длины шести генов, а затем изменяется более равномерно.

| на прямой цепи | на обратной цепи | максимум | среднее | медиана |
|----------------|------------------|----------|---------|---------|
| 1011 | 872 | 21 | 1,84 | 1 |

Таблица 3. Квазиопероны



Рис. 2. Распределение длин квазиоперонов

3.5 Статистические данные о пересечениях генов и межгенных промежутках

Было подсчитано количество генов, имеющих пересечения. Статистические данные представлены в таблице 4. Всего было обнаружено 419 пересечений. В подавляющем большинстве случаев пересекаются гены, располагающиеся на одной цепи ДНК. Однако два исключительных пересечения, происходящие со сдвигом рамки считывания, относятся к генам с разных цепочек ДНК.

| максимум | среднее | медиана | на одной цепи | на разных цепях |
|----------|---------|---------|---------------|-----------------|
| 65 | 10 | 7 | 408 | 11 |

Таблица 4. Пересечения генов

Также была построена гистограмма (рис. 3) длин межгенных промежутков. Хотя подавляющее большинство имеют длину до 200 п.н., встречаются промежутки длиной до 3373 п.н., что больше, чем в 10 раз превышает среднюю длину белка из протеома данной бактерии.



Рис. 3. Распределение длин межгенных промежутков

3.6 Статистика белков по категориям достоверности их существования

Результаты представлены в таблице 5. Экспериментально было подтверждено существование 17 белков (для 3 из них данные по экспрессии генов подтверждают наличие транскриптов). Хотя ещё почти 900 белков были предсказаны на основе наличия ортологов у близких видов, большая часть не имеет и такого подтверждения существования.

| свидетельства на белковом уровне | свидетельства на уровне транскрипции | заключения на основе гомологии | остальные |
|----------------------------------|--------------------------------------|--------------------------------|-----------|
| 14 | 3 | 886 | 2542 |

Таблица 5. Данные о существовании белков

БЛАГОДАРНОСТИ

Автор выражает огромную признательность А.В. Алексеевскому за терпение и помощь в работе с базами данных.

ССЫЛКИ

- [1] https://microbewiki.kenyon.edu/index.php/Acetobacterium_woodii
- [2] <ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/247/605/>
- [3] <https://www.uniprot.org/uniprot/?query=yourlist:M201812206746803381A1F0E0DB47453E0216320D0CEB777&sort=yourlist:M201812206746803381A1F0E0DB47453E0216320D0CEB777>

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

<https://kodomu.fbb.msu.ru/~adodonova/term1/block4/pr13.xlsl>