

Мини-обзор генома бактерии *Ruminiclostridium thermocellum* AD2

Камышева Анна^{1*}

¹МГУ им. М. В. Ломоносова, факультет биоинженерии и биоинформатики, первый курс

РЕЗЮМЕ

Данная работа посвящена анализу генома бактерии *Ruminiclostridium thermocellum* AD2. Этот организм благодаря внеклеточному ферментативному комплексу способен расщеплять целлюлозу. Поэтому существует потенциал для применения *Ruminiclostridium thermocellum* AD2 в производстве биотоплива.

Я определила размер генома, число генов, кодирующих белки и различные типы РНК, а также расположение генов на прямой и комплементарных цепях. Изучила распределение длин белков бактерии.

Результаты представлены в виде таблиц и гистограммы, построенных при помощи программы Excel из пакета MsOffice.

1 ВВЕДЕНИЕ

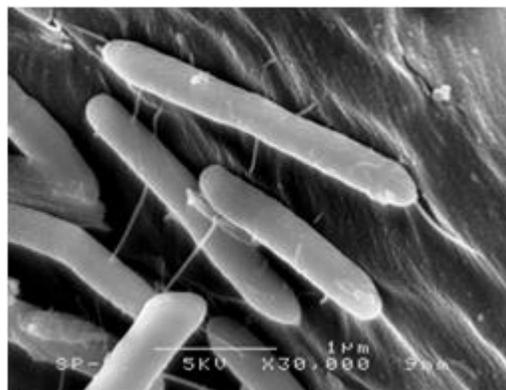
Ruminiclostridium thermocellum AD2 (*C. Thermocellum*) – анаэробная, термофильная, грамположительная бактерия. Как и большинство других бактерий рода *Clostridium*, *C. Thermocellum* имеет палочковидную форму тела. (Рис.1).^[1] Систематическое положение данного организма представлено в Таблице 1.

Таблица 1. Систематическое положение *Ruminiclostridium thermocellum* AD2.

Царство:	Bacteria
Тип:	Firmicutes
Класс:	Clostridia
Порядок:	Clostridiales
Семейство:	Clostridiaceae
Род:	<i>Clostridium</i>
Вид:	<i>C. thermocellum</i>

Ruminiclostridium thermocellum AD2 представляет интерес для науки своей уникальной способностью расщеплять кристаллическую целлюлозу, преобразуя ее в этанол. Этот процесс осуществляется благодаря особой внеклеточной ферментативной системе, называемой “cellulosome”. Эта система отличается от аналогичных целлюлаз грибов, возможностью ее использования для полного расщепления кристаллической целлюлозы (целлюлозы хлопка, к примеру).^[2]

Рисунок 1. Растровый электронный микроскоп, изображение изолированных бактерий *Ruminiclostridium thermocellum* AD2.^[3]



Из-за особенностей метаболизма данного организма можно предположить, что он способен обитать в любой среде, содержащей необходимую для его жизнедеятельности органику: в почве, на дне водоемов, в пищеварительном тракте животных.^[1]

Использование растительной биомассы в качестве источника топлива не возможно без расщепления целлюлозы. Именно этим обусловлен интерес к *Ruminiclostridium thermocellum* AD2. Лигноцеллюлозные биомассы являются крупнейшим источником органического вещества на Земле, что делает их перспективным возобновляемым сырьем для производства биотоплива и химических веществ.^[4] Тем не менее существуют некоторые ограничения к применению данного вида в производстве. В частности, из-за разветвленных путей ферментации, выход чистого этанола получается низким, вместе с ним производятся ацетат, формиат и лактат. Некоторые недавние исследования были направлены на оптимизацию этанол-продуцирующих метаболических путей, в надежде на создание более эффективного способа преобразования биомассы.^[2]

Целью данной работы является изучение генома бактерии *C.thermocellum*, в частности определение числа генов, кодирующих белки и РНК, определение размещения этих генов на прямой и обратной цепях ДНК, поиск закономерности в таком размещении и изучение распределения длин белков. Изучение генома *Ruminiclostridium thermocellum* AD2 может помочь в создании эффективных способов использования данного организма в биотехнологии.

*To whom correspondence should be addressed.

2 МАТЕРИАЛЫ И МЕТОДЫ

1) Информация о геноме бактерии была скачена с сайта “National Center for Biotechnology Information” (NCBI: Genome => Browse by organism => Search by organism). Скаченный файл содержит информацию о координатах генов, кодирующих различные типы РНК и белки, а также о расположении этих генов на прямой и обратной цепях ДНК (отредактированный файл можно найти в сопроводительных материалах).

2) Для обработки данных использовалось программное обеспечение MsOffice Excel. Файл был преобразован в формат Excel. При помощи фильтров из таблицы были удалены неинформативные строки, а также строки были отсортированы в порядке возрастания порядковых номеров старт-кодонов. Для получения результатов применялись следующие встроенные функции:

- СЧЁТЕСЛИМН – для построения гистограммы распределения длин белков
- СЧЁТЕСЛИМН, СУММ, ОКРУГЛ – для построения таблицы числа генов белков и генов РНК на прямой и комплементарной цепочках ДНК, определения общего числа белков, различных видов РНК.
- СЛУЧМЕЖДУ, СУММ, ABS – для проверки гипотезы о случайном и равновероятном распределении генов по цепям ДНК.

3 РЕЗУЛЬТАТЫ

Геном *Ruminiclostridium thermocellum* представлен одной кольцевой хромосомой, состоящей из 3554360 пар нуклеотидов. Всего в геноме присутствует 3096 генов. На 1 млн пар нуклеотидов приходится примерно 885 генов.

3.1 Число генов белков и генов РНК

Из 3096 генов *Ruminiclostridium thermocellum* AD2 3023 кодируют белки, и 73 гена – различные виды РНК. Количество генов, кодирующих различные виды РНК и белки, а так же процент их содержания в геноме бактерии продемонстрирован в Таблице 2.

Таблица 2. Число генов *Ruminiclostridium thermocellum* AD2 по типу продуктов их реализации.

	Protein	tRNA	rRNA	ncRNA	tmRNA
Amount	3023	56	12	4	1
%	97,642	1,809	0,388	0,129	0,032

Как видно из Таблицы 2, подавляющее большинство генов кодируют различные белки. Также в геноме данной бактерии присутствуют 4 некодирующих РНК, две из которых входят в

состав комплекса из белка и РНК, участвующего в распознавании специфических белков на плазматической мембране прокариот. Другая некодирующая РНК входит в состав рибонуклеазы Р. Четвертая является 6S РНК, участвующей в регуляции транскрипции.

Также в геноме бактерии присутствует одна транспортно-матричная РНК^[5], 56 транспортных РНК и 12 рибосомальных РНК.

3.2 Распределение генов по цепям

Молекула ДНК – двухцепочная, поэтому выделяют ее прямую цепь (+) и обратную цепь (-). Гены располагаются на обеих цепях молекулы, но распределены по ним не совсем равномерно. Распределение отражено в Таблице 3.

Таблица 3. Распределение генов по цепям ДНК (Прямой (+) и обратной (-)).

Strand	Protein	RNA
+	1468	39
-	1555	34

Из таблицы видно, что на обратной цепи закодировано на 87 белков больше, чем на прямой. На прямой же цепи, наоборот, на 4 больше генов РНК.

3.3 Гипотеза о распределении генов по цепям ДНК

Изначально было предположено, что вероятность нахождения гена на прямой цепи равна вероятности нахождения гена на обратной цепи. То есть вероятность найти данный ген на прямой цепи – 0.5. Для каждого гена из генома бактерии 100 раз проводился опыт с “подбрасыванием монетки”, где решка – 0, орел – 1 (с применением функций, описанных в методах для данного опыта). При выпадении 1, ген условно находился на прямой цепи, при выпадении 0 – на обратной. При суммировании результатов для всех генов по одному опыту, я получала модель ситуации, наблюдаемой в Таблице 3.

Повторив описанные действия 100 раз, я рассчитала сколько раз полученной отклонение от теоритического результата превышало или было равно рассчитанному мной для бактерии (3096 генов/2 -1468).

Проделав все описанное выше несколько раз, оказалось, что в 5-24 случаях из 100 возможное случайное распределение генов не меньше реального отклонения (Все расчеты можно найти в Сопроводительных материалах, лист 3).

3.4 Распределение длин белков

Распределение длин белков наглядно показано на Гистограмме 1.

Гистограмма 1. Распределение длин белков



Наибольшее число белков имеет длину в пределах 90-300 пар нуклеотидов (п. н.). По мере увеличения длины белка постепенно уменьшается число белков с данной длиной.

Пик длины приходится на белки с длиной 120-150 пар нуклеотидов (их 219). Число белков с длиной больше 1000 п. н. сильно уменьшается. Длина наибольшего белка достигает 8366 п. н..

ОБСУЖДЕНИЕ И ЗАКЛЮЧЕНИЕ

На основании приведенных результатов можно сделать следующие выводы:

- Гены в геноме бактерии *Ruminiclostridium thermocellum* AD2 расположены достаточно плотно (приблизительно 885 генов на 1 млн п.н.). Можно предположить, что такое размещение выгодно для бактерии, выигрывающей таким образом в стремлении уменьшить свой организм.
- 97,6% генов кодируют в организме бактерии различные белки. 2,2% приходится на транспортные и рибосомальные РНК. Остальные гены кодируют не совсем типичные элементы, встречающиеся не во всех организмах: 4 некодирующие РНК и одна транспортно-матричная РНК, участвующая в терминации трансляции. Возможно в дальнейшем, следовало бы обратить на них внимание и более подробно изучить их функции и роль в различных клеточных процессах данного организма.
- Гены распределены по цепочкам достаточно равномерно, с небольшим относительно общего числа генов отклонением. Интересным оказался тот факт, что на прямой цепи закодировано больше РНК, чем на обратной, а на обратной закодировано больше белков, чем на прямой. Возможно также такое равномерное распределение белков по цепям является еще одним механизмом уменьшения размера генома, которое может быть свойственно микроорганизмам.

- Было выяснено, что лишь в 5-24 случаях из 100, наблюдается похожее расположение генов по цепям (отклонение от ожидаемого расположения генов по цепям было равно или превысило реальное). Из такого результата можно сделать вывод, что гены распределены по цепочкам не равномерно и не случайно. Однако полностью возможность случайного распределения исключить в данном случае нельзя.
- Гистограмма длин белков *Ruminiclostridium thermocellum* AD2 достаточно типичная. Большинство белков находятся в диапазоне длины 90-300 п.н. Интересен самый большой по размеру белок в протеоме бактерии (8366 п.н.), отличающийся по размеру от других белков более чем на 6000 п.н.. Интересным могло бы быть более подробное изучение его функций.

Таким образом, благодаря использованию методов обработки данных при помощи программы Excel, удалось проанализировать геном бактерии *Ruminiclostridium thermocellum* AD2. На основе этих данных были выстроены гипотезы, которые могут быть подтверждены или опровергнуты в дальнейших исследованиях

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Отредактированная таблица, содержащая информацию о геноме бактерии, а также все точные формулы для расчета результатов, построения гистограммы и таблиц доступны по ссылке:

<http://kodomo.fbb.msu.ru/~annakamysheva/pr13.xlsx>

БЛАГОДАРНОСТИ

Хотелось бы выразить благодарность преподавателям курса практической информатики с факультета биоинженерии и биоинформатики МГУ им. М. В. Ломоносова за предоставление возможности технически претворить в жизнь этот мини-обзор.

СПИСОК ЛИТЕРАТУРЫ

- ^[1] https://microbewiki.kenyon.edu/index.php/Clostridium_thermocellum - характеристики *Ruminiclostridium thermocellum* AD2(форма, места обитания).
- ^[2] https://en.wikipedia.org/wiki/Clostridium_thermocellum - описание вида в википедии
- ^[3] https://www.jircas.affrc.go.jp/english/publication/highlights/2010/2010_13.html - фотография бактерии
- ^[4] <http://fimes.ru/tech/38669-mikrob-clostridium-thermocellum-pokazal-uchenym-kak-malo-oni-eshhe-znayut.html> - перспективы использования *Ruminiclostridium thermocellum* AD2 в биотехнологии.
- ^[5] <https://www.ncbi.nlm.nih.gov/> - таблица с данными