

Мини-обзор генома бактерии *Streptococcus mutans* UA159

Иван Черных

*Факультет биоинженерии и биоинформатики МГУ им. М.В.
Ломоносова*

*Ленинские горы МГУ 1 стр. 73, г. Москва, 119234, Российская
федерация*

ivan.chernykh@kodomo.fbb.msu.ru

Резюме

Streptococcus mutans — вид грамположительных, факультативно анаэробных бактерий рода *Streptococcus*, обычно обнаруживаемая в ротовой полости человека; вносит существенный вклад в возникновение кариеса. В данной работе было проведено исследование генома и протеома *Streptococcus mutans* UA159 с помощью программы Excel. В результате были получены распределения белков, закодированных в геноме, по длинам, по их функциям и по цепям ДНК. Также были получено распределение количества РНК-кодирующих генов по типу кодируемой РНК.

Ключевые слова: *Streptococcus mutans*, геном, протеом, белки, ДНК, РНК

Введение

Streptococcus mutans – это вид рода *Streptococcus*, который играет главную роль в образовании кариеса, переводя сахарозу в молочную кислоту¹. Кислая среда, создаваемая во рту этим процессом, является причиной того, что высоко минерализованная зубная эмаль становится уязвимой для разрушения. Сахароза используется *Streptococcus mutans*, чтобы произвести липкий внеклеточный полисахарид на основе декстрана, который позволяет им связываться между собой, формируя зубной налёт. *Streptococcus mutans* производит декстран при помощи фермента декстрансахаразы, используя сахарозу как субстрат. Сахароза — единственный сахар, который *Streptococcus mutans* может использовать, чтобы образовывать этот полисахарид^{2,3}.

Эти бактерии имеют большое медицинское значение, так как являются причиной возникновения кариеса, из-за чего их исследование является очень важным и актуальным вопросом. В данной работе исследовалось расположение генов

белков и РНК в геноме бактерии, а именно их расположение на цепях, длины кодируемых белков, их типы и функции.

Материалы и методы

Исходный файл с данными о геноме бактерии был взят с сайта NCBI⁴: <https://www.ncbi.nlm.nih.gov/genome/genomes/856>

Адрес исходного файла: <https://kodomofbb.msu.ru/~ivan.chernykh/term1/pr14.html>

Для анализа данных была использована программа Excel, в которой была построена гистограмма длин белков и две таблицы про распределение генов по цепям ДНК и по функциям белков и РНК.

При построении гистограммы был взят шаг в 100 аминокислот, так как этот формат был наиболее удобен для чтения, и при этом он показывал реальную тенденцию распределения белков по длинам. Наименьшим значением длины было выбрано 25, так как это даёт возможность быстро оценить наименьшую длину закодированного белка. Также было решено все белки длиннее 1525 аминокислот объединить в один столбец, так как разброс на таких длинах получался уже очень большой, что сильно снизило бы удобство восприятия диаграммы. Длина белков была получена из столбца “product_length”, где указана длина исключительно белков.

Далее была построена таблица, в которой показывается разброс генов по цепям ДНК. Блок-кодирующими генами были приняты гены, имеющие значение “protein_coding” в столбце “class”, псевдогенами – гены, имеющие значение “pseudogene”, а РНК-кодирующими – имеющие значение “RNA”.

Наконец, при построении последней таблицы были выделены классы белков: рибосомальные, транспортные и гипотетические, а также классы РНК: рибосомальные и транспортные. Рибосомальными белками считались те, что имели значение “ribosomal protein” в столбце “name”, транспортными – имеющие значение “transporter” или вообще часть “transport” в любом виде, а гипотетическими – имеющие значение “hypothetical protein”. В случае, если белок относился к гипотетическим и каким-то другим одновременно, он считался гипотетическим. Аналогично было сделано с РНК, “tRNA” – транспортные, “ribosomal RNA” – рибосомальные. После чего был произведён расчёт среднего количества генов на 1 миллион пар нуклеотидов.

Результаты и обсуждения

Длины белков

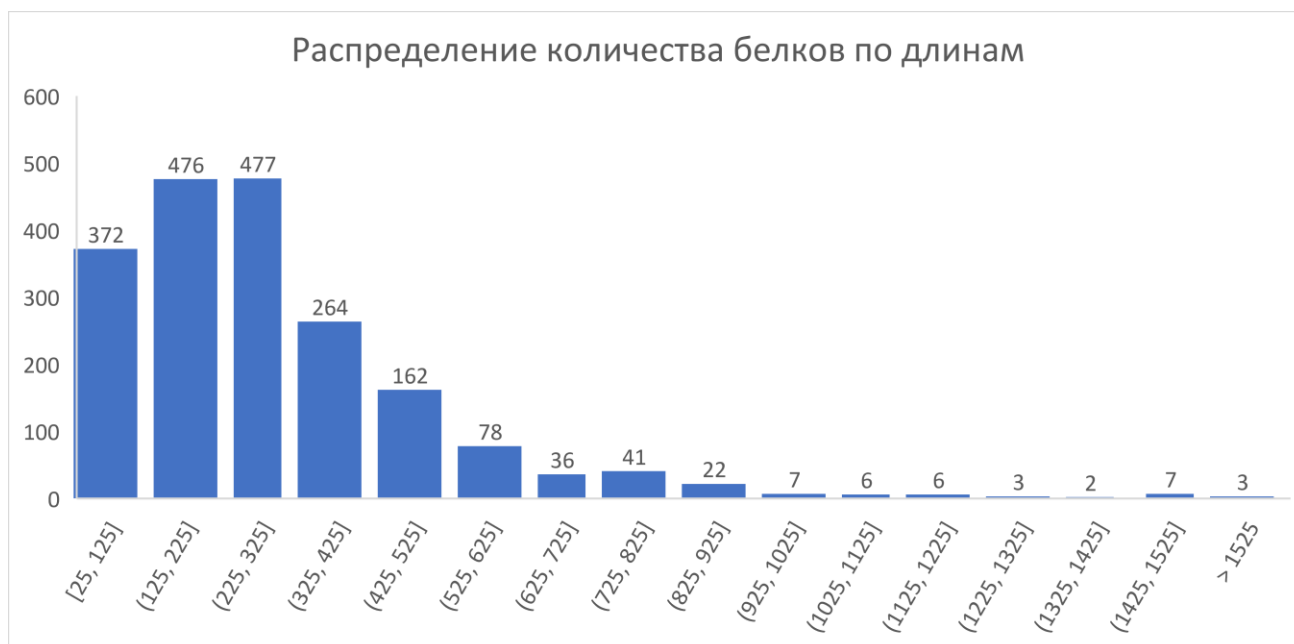


Рис. 1. Распределение количества белков по диапазонам длин. По горизонтали – диапазоны длин в аминокислотах, по вертикали – количество генов, кодирующих белки с данной длиной.

Параметр	Значение
Минимальная длина	25
Максимальная длина	2724
Средняя длина	295,7737
Стандартное отклонение	219,7168
Медиана	250

Табл. 1. Некоторые параметры длин белков.

Как видно из гистограммы (рис. 1), белки *Streptococcus mutans* очень разнообразны по длине, однако больше всего встречается в диапазоне 225 – 325. Это хорошо согласуется с данными из табл. 1, где приведены значения медианы и средней длины, которые как раз попадают в этот диапазон. Стандартное отклонение принимает довольно большое значение, что подтверждается большим разбросом по длине (например, самый длинный белок имеет 2724 аминокислоты в своём составе).

Разброс генов по цепям

Цепь	Число генов белков	Число псевдогенов	Число генов РНК
Прямая	943	0	44
Комплементарная	1019	1	35

Табл. 2. Распределение генов по цепям ДНК.

Как видно из табл. 2, в геноме *Streptococcus mutans* практически отсутствуют псевдогены. Также на комплементарной цепи ДНК генов белков больше, в то

время как РНК-кодирующих генов больше на прямой цепи, причём это отличие существенно. Вероятно, такое распределение нельзя назвать случайным, однако проверки на это не проводилось.

Распределение генов по типу закодированной РНК или белка

Ген	Количество	Приблизительное количество на 1 млн п. н.
Рибосомальные белки	53	26,0697834
Транспортные белки	156	76,73370208
Гипотетические белки	731	359,5662578
Остальные белки	1045	514,0174274
Транспортные РНК	65	31,97237587
Рибосомальные РНК	15	7,378240585
Остальные РНК	0	0

Табл. 3. Распределение генов по типу закодированной РНК или белка.

Как видно из табл. 3, транспортных белков в протеоме, примерно, в три раза больше, чем рибосомальных, что, вполне возможно, свидетельствует о том, что транспорт в клетке заметно более сложен и разнообразен, чем устройство рибосом. В то же время функции многих белков (почти половина от всего количества) остаются невыясненными.

Количество тРНК хорошо согласуется с теоретическими данными, так как всего присутствует по одному типу тРНК на каждый вариант кодона и ещё один дополнительный, функции которого не ясны. В геноме не присутствует генов остальных РНК, кроме транспортных и рибосомальных, чего тоже следовало ожидать.

Заключение

Исследование генома *Streptococcus mutans* дало множество интересных результатов, таких как неравномерное распределение генов, кодирующих белки и РНК, по цепям ДНК, а также малое количество псевдогенов в сравнении с другими бактериями. Однако большая часть генов не исследована до конца (731 гипотетический белок), что даёт возможность провести ещё множество исследований, которые могут привести к неожиданным результатам.

Сопроводительные материалы

<https://kodomofbb.msu.ru/~ivan.chernykh/term1/pr14.html>

Благодарности

Хотелось бы поблагодарить Алексея Дрождева и Александру Горбонос за моральную поддержку во время написания этого мини-обзора.

Список литературы

1. Loesche WJ. Microbiology of Dental Decay and Periodontal Disease. In: Barron's Medical Microbiology (Barron S et al, eds.). — 4th ed. — Univ of Texas Medical Branch, 1996. — ISBN
2. Ryan KJ; Ray CG (editors). Sherris Medical Microbiology. — 4th ed. — McGraw Hill, 2004. — ISBN
3. https://en.wikipedia.org/wiki/Streptococcus_mutans
4. <https://www.ncbi.nlm.nih.gov/>