

Краткий обзор генома и протеома *Lactobacillus dextrinicus* (Coster and White) Naakensen et al.

Литвинова А.В.

Московский государственный университет имени М. В. Ломоносова, факультет биоинженерии и биоинформатики,
1 курс

АННОТАЦИЯ

В данном обзоре приведен ограниченный анализ генома и протеома бактерии *Lactobacillus dextrinicus*. Нуклеотидный состав характеризуется 38% частотой GC-пар. Были определены точки начала и терминации репликации. В геноме было обнаружено 5 рРНК оперонов. Анализ протеома заключается в сравнении длин всех белков бактерии. Также исследовалась частота встречаемости кодонов.

1 ВВЕДЕНИЕ

Целью данного обзора является общий анализ генома и протеома *Lactobacillus dextrinicus* и представление их кратких характеристик.

Lactobacillus dextrinicus – бактерия, ранее описанная как *Pediococcus dextrinicus* Coster and White (Back W., 1978) и позже перенесенная в род *Lactobacillus* по ряду морфологических, физиологических признаков и на основании филогенетического анализа. Данная бактерия имеет несколько форм: кокк, диплококк и неправильные кластеры из кокков (Naakensen M. и др., 2009). *Lactobacillus dextrinicus* применяется в производстве пива и заготовке силоса. Длина генома бактерии – 1836976 п.н. В нем было идентифицировано 1812 генов, из них 1709 кодируют белки, 79 кодируют РНК и 24 представляют собой псевдогены (https://www.ncbi.nlm.nih.gov/nucore/NZ_CP044499.1).

2 МАТЕРИАЛЫ И МЕТОДЫ

Последовательность генома и данные о протеоме *Lactobacillus dextrinicus* были взяты с сайта NCBI (https://www.ncbi.nlm.nih.gov/nucore/NZ_CP044499.1).

Часть анализа проводилась в программе Excel, также был использован программный пакет EMBOSS для подсчета частот нуклеотидов и GC-состава в геноме (команды wordcount и geecee соответственно).

Сайт <http://genskew.csb.univie.ac.at/> был использован для поиска точек начала и окончания транскрипции.

В качестве статистического метода использовался критерий Хи-квадрат.

Визуализация расположения рРНК оперонов в геноме проводилась в программе SnapGene Viewer.

3 РЕЗУЛЬТАТЫ

3.1 Анализ генома

Нуклеотидный состав

В геноме бактерии *Lactobacillus dextrinicus* содержится 568336 (это 30,94% от общего количества нуклеотидов) адениновых нуклеотидов, 567632 (30,90%) тиминовых, 354460 (19,30%) гуаниновых и 346548 (18,87%) цитозиновых. Несмотря на различия в процентном содержании аденина и тимина, а также гуанина и цитозина, правило Чаргаффа соблюдается для этой последовательности, поскольку различия статистически незначимы по критерию Хи-квадрат ($p = 2,496E-19$). GC - состав этого генома примерно 38%.

Точка начала репликации (Ori)

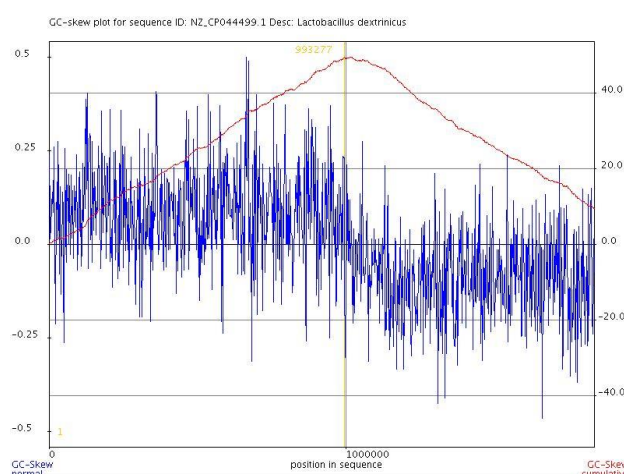


Рис. 1 График изменение GC-состава в геноме.

На рис.1 представлен график, на котором отмечены координаты Ori и точки терминации репликации (term).

Ori отмечен на 1 п.н, отсюда можно сделать вывод, что секвенирование проводилось в Ori, заранее зная последовательность прилежащего участка. Term имеет координату 993277 п.н., то есть плазида делится почти пополам: одна половина имеет длину 54% от общей длины, другая – 46%.

3.2 Анализ протеома

Длина белков

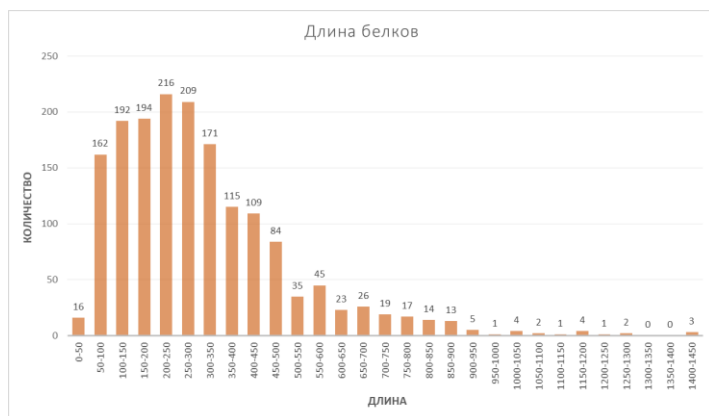


Рис. 2 График распределение белков в зависимости от длины

Для 1709 белков была построена гистограмма с их распределением по длине (рис.2). Самый короткий белок имеет длину 25 (белок с неизвестной функцией) аминокислотных остатков (а.о.), самый длинный – 1447. Примечательно, что последний является вирусным белком «рулеткой» (tare measure protein), который обуславливает проникновение бактериофага в клетку бактерии. Возможно, что данная бактерия была инфицирована. Инструментом BLAST NCBI идентифицировать вирус не удалось

Средняя длина составляет 305 а.о., медианная длина белка – 267 а.о. Из графика следует, что мода длины находится в пределах от 200 до 250 а.о.

Рибосомальные гены

Было выявлено 70 генов, кодирующих рРНК и рибосомальные белки. Полный список этих генов с указанием их координат представлен в таблице ribosomal в сопроводительных материалах.

Также было обнаружено, что гены рРНК имеют определенную организацию в геноме: они встречаются в виде оперонов (см. табл.1). Оперон состоит из подряд идущих генов 16S, 23S и 5S рРНК. У бактерий между генами 16S и 23S рРНК часто встречаются гены тРНК, такое строение имеют опероны 1 и 2, в спейсере находятся гены тРНК изолейцина и аланина. К оперону 4 прилегают гены тРНК серина и аспаргина, и вследствие их близкого расположения эти тРНК были включены в состав оперона. 3 и 5 опероны представлены только генами рРНК. На рис. 3 показано расположение оперонов в геноме и направление их транскрипции. Транскрипция всех рРНК оперонов осуществляется в направлении от Ori к term (см. рис 3)

Таблица 1 Координаты рибосомальных оперонов

Опероны	Координаты	Направление
Оперон 1		
16S rRNA	189065-190638	+
tRNA-Ile	190746-190819	+
tRNA-Ala	190845-190917	+
23S rRNA	191038-193958	+
5S rRNA	194034-194150	+
Оперон 2		
16S rRNA	599828-601401	+
tRNA-Ile	601509-601582	+
tRNA-Ala	601608-601680	+
23S rRNA	601801-604721	+
5S rRNA	604797-604913	+
Оперон 3		
5S rRNA	1412897-1413013	-
23S rRNA	1413089-1416009	-
16S rRNA	1416198-1417771	-
Оперон 4		
tRNA-Ser	1486783-1486873	-
tRNA-Asn	1486942-1487014	-
5S rRNA	1487022-1487138	-
23S rRNA	1487214-1490134	-
16S rRNA	1490323-1491896	-
Оперон 5		
5S rRNA	1754406-1754522	-
23S rRNA	1754598-1757518	-
16S rRNA	1757707-1759280	-

Частота встречаемости кодонов

В таблице codons в сопроводительных материалах представлены максимальные и минимальные частоты кодонов для всех аминокислот. Глутамин, лейцин, изолейцин, глутамат и аспарат имеют самые часто встречающиеся кодоны, доля которых варьируется от 4 до 5% от общего числа кодонов.

На месте 3 нуклеотида чаще всего стоит тимин (у 12 аминокислот), второй по встречаемости – аденин (у 6 аминокислот). Метионин и триптофан кодируются только одним кодоном, и на третьем месте всегда стоит гуанин. В остальных случаях на третьей позиции гуанин и цитозин встречаются гораздо реже, чем аденин и тимин.

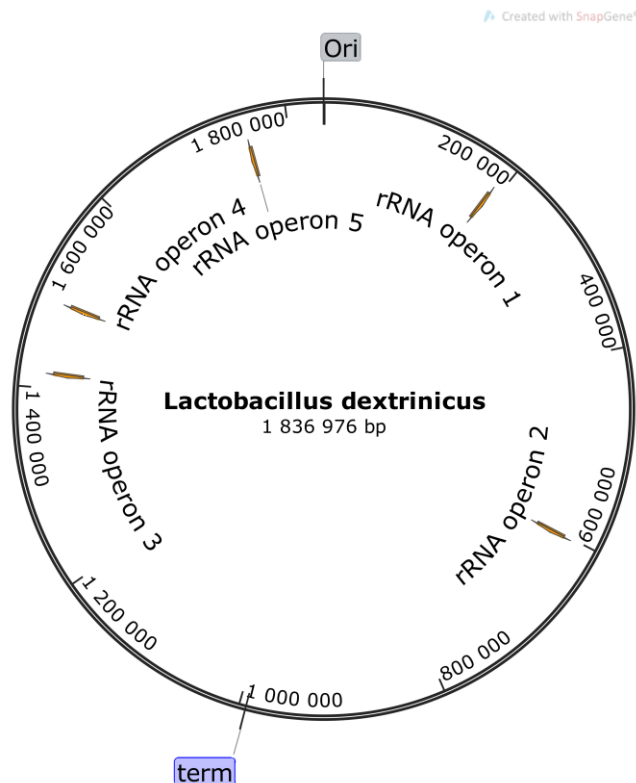


Рис. 3 Расположение рРНК оперонов в геноме. Стрелками указано направление их транскрипции.

4 ВЫВОДЫ

Дополнительный анализ генома *Lactobacillus dextrinicus* показал, что GC-состав равен 38%. Точка Ori расположен на 1 п.н, точка term – 993277 п.н.

Анализ рибосомальных генов показал наличие 5 рРНК оперонов в геноме. В анализе протеома было приведено сравнение всех белков по их длине, средняя длина составляет 305 а.о, наибольшее число белков имеет длину от 200 до 250 а.о. В протеоме бактерии был обнаружен бактериофаговый белок хвостовой «рулетки». Также были рассмотрены частоты встречаемости кодонов, и на 3-их позициях чаще всего находятся тимин и аденин.

ЛИТЕРАТУРА

- 1.Back, W. Elevation of *Pediococcus cerevisiae* subsp. *dextrinicus* Coster and White to Species Status [*Pediococcus dextrinicus* (Coster and White) comb. nov.]. INT. J. SYST. BACTERIOL. 28, 5 (1978).
- 2.GenSkew - visualization of nucleotide skew in genome sequences. <http://genskew.csb.univie.ac.at/>.
- 3.*Lactobacillus dextrinicus* strain LH506 chromosome, complete genome-Nucleotide-NCBI. https://www.ncbi.nlm.nih.gov/nucleotide/NZ_CP044499.1.
- 4.Haakensen, M., Dobson, C. M., Hill, J. E. & Ziola, B. Reclassification of *Pediococcus dextrinicus* (Coster and White

1964) Back 1978 (Approved Lists 1980) as *Lactobacillus dextrinicus* comb. nov., and emended description of the genus *Lactobacillus*. INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY 59, 615–621 (2009).

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Таблицы ribosomal и codons:

<https://kodomofbb.msu.ru/~ana-liva/term1/Litvinova-supple-fin.xlsx>