

Обзор протеома бактерии *Klebsiella oxytoca* E718

Посицельская Екатерина¹

¹Студентка факультета биоинженерии и биоинформатики МГУ им. М.В. Ломоносова

РЕЗЮМЕ

В данной работе был произведен анализ некоторых данных о геноме бактерии *Klebsiella oxytoca* E718, длин белков данной бактерии, а также количества плазмид, являющихся паразитами рода *Klebsiella*. Были выявлены некоторые закономерности и получены количественные оценки отдельных свойств генома.

ВВЕДЕНИЕ

Начиная с 1990-х годов происходит активная расшифровка геномов и сбор информации о живых организмах, в том числе и таких древнейших из них, как археи, бактерии, а также плазмид и других вирусов. Данные собраны в единую базу данных и доступны всем желающим.

Нас интересовала бактерия *Klebsiella oxytoca* E718, которая является факультативно-анаэробным организмом и обитает в различных средах (в качестве паразита, или свободно живущего в почве организма, устойчива во внешней среде благодаря образованию спор). Бактерия вызывает инфекцию человека, заражение которой может произойти при проведении операции на позвоночник. Поражает слизистые оболочки организма. Сопровождается болями в спине и другими проявлениями нарушения нормальной работы пищеварительной, выделительной или дыхательной систем - но может протекать и практически бессимптомно. Чаше встречается у детей, чем у взрослых, так как их иммунная система еще недостаточно развита.

Чтобы подвергнуть данные о геноме данной бактерии статистической обработке, мы пользовались программой эксель. В данном случае, нас интересовали три вопроса: какие плазмиды паразитируют на бактерии *Klebsiella oxytoca* E718, каков характер распределения длин белков, входящих в ее геном, и соотношение кодирующих белки и РНК-кодирующих участков ее генома.

МЕТОДЫ

Исходные данные о геноме бактерии *Klebsiella oxytoca* E718 мы получили с сервера NCBI. Эти данные представляли собой два файла - NC_018106.ptt, содержащий информацию об участках, кодирующих белки и NC_018106.rnt, содержащий информацию о РНК-кодирующих участках данного фрагмента генома. Доступ осуществлялся при помощи ftp-протокола, использовалась программа FAR-manager.

Нами были использованы программа Excel (версии 2003г и 2011г). В ней использовались сводные таблицы (для получения списков без повторов и частоты встречаемости каждого

элемента), а также функции: СЦЕПИТЬ: функция склеивает две текстовые строки; СРЕДНЕЕ; МЕДИАНА; ЧАСТНОЕ: возвращает целую часть от деления, использовалась для распределения значений по диапазонам при создании гистограммы.

Проблема разделителей между целой и десятичной частями числа решалась при помощи функции специальной вставки или замены и последующего разбиения на столбцы, если это основной случай встречаемости точки.

Для построения гистограммы использовалось меню “диаграмма”. Мы считали не количество каждого значения, а разбили весь интервал на диапазоны, чтобы получить более наглядную картину.

РЕЗУЛЬТАТЫ

Нами были получены следующие результаты:

1. Плазмиды, паразитирующие на бактериях рода *Klebsiella*

На представителях вида, к которому принадлежит *Klebsiella oxytoca* E718, могут паразитировать 296 различных плазмид.

2. Распределение длин белков в геноме *Klebsiella oxytoca* E718

Распределение длин белков близко к нормальному. Однако среднее сильно сдвинуто в сторону меньших значений. Большая часть белков имеет длины в диапазоне до 900 а/к, но в то же время максимальные значения длин доходят до 3150 а/к.

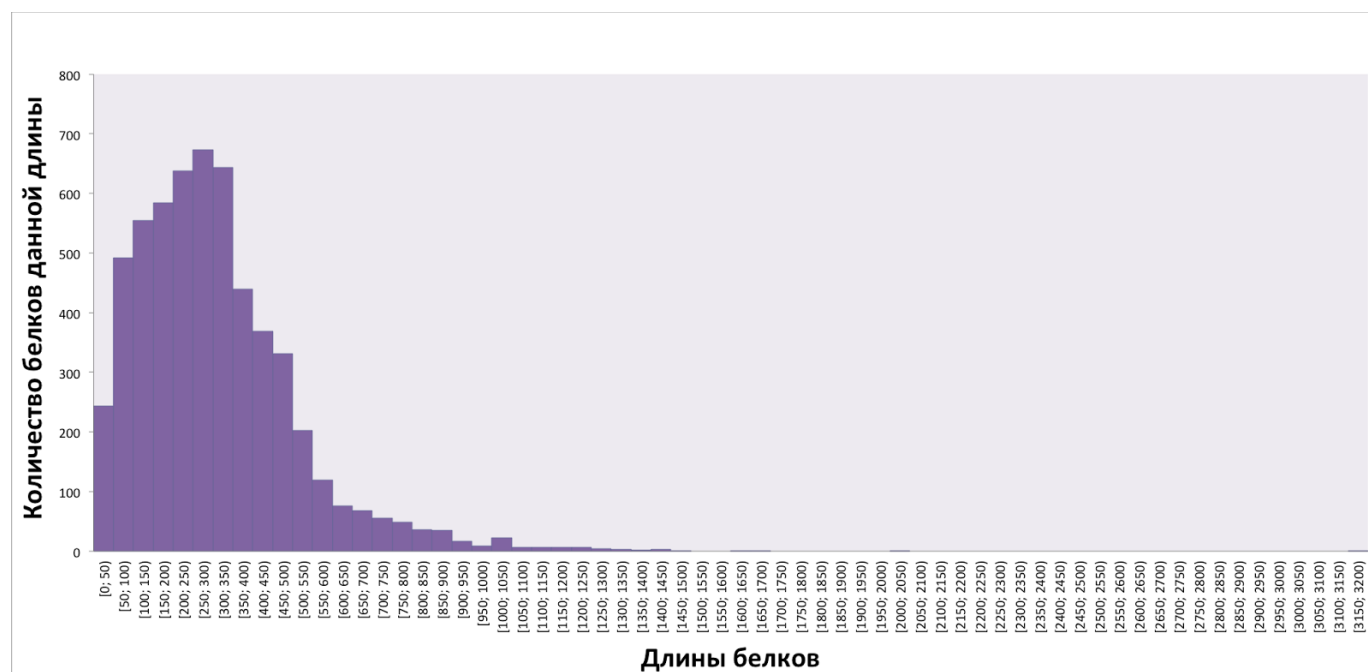


Рис. 1. Гистограмма длин белков из протеома бактерии *Klebsiella oxytoca* E718

3. Распределение кодирующих участков по цепям ДНК бактерии *Klebsiella oxytoca* E718

Кодирующие участки примерно равномерно распределены по цепям.

Таблица 1. Количество генов белков и генов РНК на прямой и комплементарной цепи ДНК

количес тво	прямая цепочка	комплементарная цепочка
генов	69	40
белков	2804	2897

ОБСУЖДЕНИЕ

Следует отметить, что белок-кодирующие участки значительно больше, чем РНК-кодирующих. Это очередной раз напоминает нам о загадочной природе кодирования – мы знаем, какой код соответствует какому белку, и знаем, как часть белков функционирует в организме. Однако вопрос о том, как именно линейная последовательность дает такое разнообразие так сложно пространственно и логически устроенных организмов остается открытым. Поэтому не удивительно, что белки – главный строительный материал этого разнообразия – является наиболее востребованным.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Полученные данные содержатся в файле

http://kodomo.fbb.msu.ru/~es_posicel/term1/Positselskaya_pr14_new.xls

На вкладке RNA отображаются данные о РНК-кодирующих участках, на вкладке CDS – о участках генома, кодирующих белки. На вкладке Bar Chart содержится гистограмма, изображенная на рис. 1. Вкладка work содержит расчеты и Таблицу 1.

На странице http://kodomo.fbb.msu.ru/~es_posicel/term1/Positselskaya_pr15.pdf содержится обсуждение результатов и отчет о ходе работы.

ИСТОЧНИКИ

1. <https://ru.wikipedia.org/wiki/%D0%9A%D0%BB%D0%B5%D0%B1%D1%81%D0%B8%D0%B5%D0%BB%D0%BB%D1%8B>
2. <http://www.ncbi.nlm.nih.gov/genome/?term=Klebsiella+oxytoca+E718>
3. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3353478/>