

Алгоритмы сравнительного анализа серий метагеномных образцов с использованием графов де Брейна для библиотек метагеномных чтений

Иванов А.Б.

научный руководитель — Ульянов В.И.

федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»

Санкт-Петербург

Метагеномные данные это совокупность информации о микроорганизмах (например бактериях, грибах), обитающих в образце, взятом из определенной среды (например почва, вода, организм человека). Трансплантация фекальной микробиоты является инновационным методом лечения рецидивирующих кишечных инфекций, вызванных устойчивыми к антибиотикам бактериями. Принцип действия основан на переносе бактерий из организма здорового человека (донора) в организм реципиента, подверженного заболеванию. Исследования показывают эффективность данного метода для лечения пациентов, равно как и приживаемость бактерий из организма донора в организме реципиента через три месяца после процедуры трансплантации [1]. Однако вопрос механизма приживания бактерий и определение влияния изменения микробиоты на реципиента остается открытым. Для ответа на эти вопросы необходимо получить как можно более полную информацию о метагеноме реципиента до и после процедуры трансплантации. Для этого производится метагеномная сборка — процесс соединения прочтений и построения длинной последовательности (контига), соответствующей отдельному геному или его части [2]. Однако данный метод строит консенсусные последовательности из коротких прочтений, тем самым теряя информацию о потенциальных изменениях небольшого числа нуклеотидов, типичных для близкородственных штаммов.

Существующие методы анализа метагеномов не используют всю имеющуюся информацию, а также не могут достоверно выявить все случаи штаммового замещения организмов, которые могут являться ключевыми для изменения состояния пациентов. Целью данной работы является разработка алгоритмов классификация бактерий в кишечнике пациента после трансплантации фекальной микробиоты по природе их появления с использованием максимального количества информации из метагеномов.

Для достижения поставленной цели был разработан алгоритм сравнения метагеномных образцов на основе графа де Брейна. Для одного метагенома строится граф де Брейна, на который затем выравниваются прочтения из другого метагенома. Критериями обнаружения прочтения в графе являются: соотношение глубины и ширины покрытия прочтения k -мерами, обнаружение парного прочтения в графе, обнаружение прочтения при различных значениях параметра k . В результате каждое прочтение классифицируется как найденное или не найденное в другом метагеноме. На основе данного алгоритма был построен классификатор, разбивающий прочтения реципиента после трансплантации на четыре категории: (1) было и у донора, и у реципиента до трансплантации, (2) прижилось от донора, (3) сохранилось у реципиента, (4) пришло в организм реципиента из внешней среды. Ключевым отличием алгоритма от существующих является классификация каждого прочтения в метагеноме (а не классификация контигов или собранных геномов из метагенома). Также точное выравнивание прочтений на граф и запрет перестроек, таких как замена, вставка и удаление нуклеотида, позволяет отслеживать изменения, присущие штаммам одного вида и выявлять случаи штаммового замещения бактерий.

Разработанный алгоритм был протестирован на искусственно сгенерированных наборах метагеномных данных с различными параметрами (распределение бактерий в метагеноме, наличие ошибок в прочтениях, видовой состав метагенома) и показал высокие результаты точности. Также было проведено сравнение разработанного алгоритма с известным алгоритмом для выравнивания прочтений на граф де Брейна — BrownieAligner [3].

Процент ошибок значительно снизился, что связано с использованием информации о парноконцевых прочтениях в реализованном алгоритме.

В результате был разработан алгоритм, классифицирующий прочтения из метагенома реципиента по природе их появления. При этом классифицируются все прочтения из метагенома, а также используется информация о парноконцевых прочтениях для повышения качества классификации. Полученные в результате классификации группы прочтений можно подвергать дальнейшему анализу: проводить геномную сборку отдельно для каждого класса прочтений, проводить таксономическую и функциональную аннотацию классов, исследовать их для выявления правил и зависимостей приживаемости бактерий у реципиента.

1. Li S. S. et al. Durable coexistence of donor and recipient strains after fecal microbiota transplantation //Science. – 2016. – Т. 352. – №. 6285. – С. 586-589.
2. Vollmers J., Wiegand S., Kaster A. K. Comparing and evaluating metagenome assembly tools from a microbiologist's perspective-not only size matters! //PloS one. – 2017. – Т. 12. – №. 1. – С. e0169662.
3. Heydari M. et al. BrownieAligner: accurate alignment of Illumina sequencing data to de Bruijn graphs //BMC bioinformatics. – 2018. – Т. 19. – №. 1. – С. 311.