

РЕЦЕНЗИЯ

На выпускную бакалаврскую работу
Бзикадзе Александра Важевича
“Применение Hi-C к метагеномике”

Данная работа посвящена актуальной на сегодняшний день проблеме в биоинформатике — de novo сборке метагеномных данных секвенирования. В рамках данного проекта был разработан алгоритм сборки для сравнительно недавно появившегося протокола секвенирования ДНК (именуемого Hi-C). Такого рода данные имеют принципиальные отличия от стандартных данных. Именно эти особенности позволяют использовать их для разрешения повторов, одной из ключевых фундаментальных проблем в сборке геномов и метагеномов. Несмотря на то, что уже существует достаточно большое количество различных геномных сборщиков, для данных Hi-C вычислительных методов по прежнему очень мало, особенно в области метагеномики.

Стоит отметить, что Александр хорошо разобрался в непростой области, провел полноценное исследование доступных данных и подробный анализ существующих методов. Помимо этого, важным моментом является и то, что разработанный алгоритм был реализован как часть существующего геномного сборщика SPAdes, который используется по всему миру, а значит предложенный метод имеет все шансы быть примененным в реальных биологических исследованиях.

Текст работы хорошо структурирован и не вызывает затруднений при чтении. Однако, в работе можно выявить ряд недостатков, в основном касающихся стилистики и содержания текста, нежели содержания самой работы:

- Очень кратко описаны непосредственно проблемы метагеномной сборки и почему Hi-C является хорошей технологией для их разрешения;
- Не отмечены недостатки программы `salsa2`;
- Несмотря на то, что программ, разработанных непосредственно для скаффолдинга метагеномных данных Hi-C не существует, было бы интересно увидеть в работе сравнение с существующими геномными скаффолдерами, например, той же `salsa2`;
- Введенный термин “сжатого графа” уже зарезервирован в рамках области сборки геномов и имеет вполне конкретный смысл (отличный от приведенного в тексте); термин стоит изменить, например, на “сокращенный граф” или “стяженный граф”;

- Некоторые процедуры предложенного вычислительного метода описаны расплывчато;
- В тексте присутствует умеренное количество научного жаргона, неформальных терминов и англицизмов (например пайплайн, датасет).

Перечисленные недостатки, однако, не умаляют актуальности, объема и качества проделанной работы. Данная бакалаврская работа заслуживает оценки отлично.

*Пржбельский Андрей Дмитриевич
н.с. лаборатории "Центр Алгоритмической Биотехнологии" СПбГУ
17 мая 2019 г.*