

ОТЗЫВ

Председателя диссертационного совета на диссертацию Антипова Дмитрия Юрьевича на тему: «Разработка алгоритмов для специальных задач сборки геномов», представленную на соискание ученой степени кандидата наук по специальности 1.5.8. —

Математическая биология, биоинформатика

. В современной молекулярной биологии подавляющее большинство статей содержит часть, сделанную методами биоинформатики. Либо исследование начинается с анализа баз данных (data mining), либо исследователь сравнивает свои находки с уже известными, которые также находятся в базах данных. Молекулярные биологи опираются на то, что находится в этих базах, и сильно зависят от их качества. Ошибки баз данных, ошибки сборок, приводят и к недостоверным результатам. Зависимость биологов от качества работы биоинформатиков очень велика. Поэтому работа диссертанта и всей группы профессора Певзнера и профессора Аллы Лapidус бесконечно ценна для мирового сообщества. О признании их результатов и, в частности, соискателя Д.Ю.Антипова, свидетельствуют цифры цитирования, т.е. практического использования работ.

В диссертации решаются две специальные задачи сборки. Это (1) гибридная сборка, основанная на разных типах секвенирования и (2) поиск плазмид в геномных и метагеномных исходных данных.

Гибридные сборки сейчас довольно распространены. В работе показано, что hybridSPAdes является одним из лучших существующих решений для бактериальных сборок. Создание алгоритма сборки – одна из заслуг соискателя.

Особенно интересен раздел, посвященный поиску плазмид в процессе сборки. С начала широкого применения антибиотиков идет непрерывное соревнование бактериальных сообществ, где устойчивость передается горизонтально, и человечества, занятого непрерывным созданием новых антибиотиков. Важной и актуальной задачей является создание инструментов для поиска плазмид, потому что именно ими передается устойчивость.

На базе инструмента SPAdes диссертант разработал специализированный сборщик plasmidSPAdes для сборки плазмид. Для полногеномных сборок plasmidSPAdes предсказывает возможные плазмиды с хорошей вероятностью.

Природные сообщества редко содержат единственный вид бактерий, а большинство видов микроорганизмов не поддается культивированию в лаборатории. Эти обстоятельства породили такой раздел биоинформатики как метагеномика, который посвящен одновременному анализу целых бактериальных сообществ. Разработанный диссертантом инструмент metaplasmidSPAdes предназначен для поиска плазмид в метагеномных данных. В пакете предусмотрена и дополнительная проверка принадлежности последовательности к плазмидам plasmidVerify. Несмотря на простую идею, лежащую в основе plasmidVerify, метод доказал свою работоспособность.

Текст диссертации написан ясно с исчерпывающими объяснениями, содержит незначительное количество опечаток. Большой объем проделанной работы привел к тому, что зачастую приходится обращаться к оригинальным статьям. Публикации и количество их цитирований производит впечатление. В тексте диссертации удалось обнаружить одну

опечатку (специального инструмента) и некорректное обозначение - баз, килобаз; следует нп или тпн или уж br или kb). Видно, что автор одинаково хорошо владеет как английским, так и русским; не понаслышке известны трудности перевода. Это большое достоинство, потому что биоинформатика существует на английском и рецензенту известны единичные случаи текстов на русском.

В порядке дискуссии хотелось бы задать соискателю ряд вопросов:

1. «Химерические ребра чаще всего возникают из-за специфики процесса секвенирования по одной клетке». Почему?
2. Стр 39: «за счет большего количества повторных участков малого размера...»; «длина максимального повторного участка в большинстве бактериальных геномов меньше 10 килобаз» - 'какова структура и происхождение больших повторных участков в бактериальных геномах'?
3. "Как сборка с меньшим значением K помогает закрыть разрывы в покрытии при сборке с большим K"?
4. Почему PacBio HiFi неправильно определяет кратность гомополимеров?
5. Можно ли с помощью Ваших методов для плазмид выявить митохондриальную ДНК из данных полногеномного секвенирования?

В Заключение главы 3 сказано, что описанные новые алгоритмы «могут позволить избежать использования сложных и затратных методов, связанных с лабораторным выделением плазмид». Хотелось бы уточнить – от выделения для секвенирования, но не для проверки. Результатом работы новых методов может быть предсказание последовательности плазмиды и возможность создания праймеров для ее клонирования. Но проверка предсказаний биоинформатики, т.е. обнаружение реальной плазмиды, остается за «мокрой» наукой. Необходимость проверки плазмиды следует из таблиц, предсказания нигде не 100%. Поэтому к достоинствам работы следует отнести раздел, где очерчены перспективы развития алгоритмов.

Диссертация Антипова Дмитрия Юрьевича на тему: «Разработка алгоритмов для специальных задач сборки геномов» соответствует основным требованиям, установленным Приказом от 19.11.2021 № 11181/1 «О порядке присуждения ученых степеней в Санкт-Петербургском государственном университете», соискатель Антипов Дмитрий Юрьевич заслуживает присуждения ученой степени кандидата наук по специальности 1.5.8. —Математическая биология, биоинформатика . Пункты 9 и 11 указанного Порядка диссертантом не нарушены.

Председатель диссертационного совета



ДБН, проф, проф

Подгорная О.И.

Дата 20.01.2022