

ОТЗЫВ

председателя диссертационного совета на диссертацию Гуревича Алексея Александровича на тему: «Вычислительные методы для анализа подверженных ошибкам метабологеномных данных», представленную на соискание ученой степени кандидата физико-математических наук по специальности 03.01.09 – Математическая биология, биоинформатика

С 2000 года, когда было объявлено об окончании программы «Геном человека» и создании чернового референса для человека, прочитаны уже десятки тысяч геномов различных организмов. Исходная идеология чтения геномов следовала основной догме молекулярной биологии, в соответствии с которой в геноме искали гены, все остальное только мешало и раздражало сборщиков. Обилие прочитанных эукариотических геномов позволяет видеть, что порядок количества генов не меняется от беспозвоночных до млекопитающих, тогда как величина генома возрастает на порядки. Именно количество повторяющейся ДНК коррелирует с увеличением сложности. Однако, идеология, заложенная исходно, приводит к тому, что качество получаемых геномных сборок различается очень сильно в зависимости от использованных методов и инструментов сборки. Необходимы методы сравнения сборок, что и определяет актуальность первой части работы.

Первыми прочитанными геномами были вирусы и «домашнее животное» биоинформатиков, так же как и геномных инженеров – кишечная палочка (*E.coli*). До сих пор именно на геноме *E.coli* зачастую проверяют новые алгоритмы сборки. Однако 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. Диссертант показал как определить достоверность метагеномных сборок, которые получаются после первичной сборки прочтений метагенома.

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 того, чтобы быстро обнаруживать известные антибиотики и их модификации в экспериментальных данных, что позволит сфокусировать максимальные усилия на обнаружении принципиально новых соединений и ускорить отбор новых антибиотиков; именно этому посвящена последняя часть диссертации.

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 известным, которое также находится 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 бесконечно ценна для биологов мирового сообщества.

bx 09/2 - 241 am 04.12.18

О степени обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации, их достоверности и новизне свидетельствует уникальный для диссертаций раздел «Распространение QUAST в научном сообществе». Из него следует, что представленные программы уже активно используются. Нашей группы еще нет в статистике ссылок, но она также там появится, потому что мы активно пользуемся предложенными программами и их модификациями, не приведенными в диссертации.

Надо отметить, что несмотря на, казалось бы, очень разные названия 4х глав диссертации, прослеживается очевидная преемственность программ – сначала приводится алгоритм сравнительно простой, потом показано как ее можно использовать для более сложных задач.

В тексте диссертации удалось обнаружить одну опечатку (Еще одна группа инструментов, например, REAPR [62] обнаруживаЮт) и одно неудачное выражение. Видно, что автор одинаково хорошо владеет как английским, так и русским, а диссертация подготовлена весьма тщательно (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 автору несколько вопросов

1. На странице 34 как основная метрика качества используется NGA50, значит ли это, что она самая важная/полезная из всей группы NGAх?
2. На странице 35 приведен Рисунок 1.7 — Выравнивание сборок данных одноклеточного секвенирования *S. aureus* на референсный геном». Что означает пробел наверху рисунка?
3. Стр. 64 «масса материнского иона». Это общепринятая терминология на русском языке?
4. Стр. 87. Одно из предсказанных программой PNP (пептидное природное соединение) массетолд-1252 уже было обнаружено экспериментально. Чем в эксперименте отличаются свойства нового варианта PNP от исходного соединения Massetolide A?

Возникшие вопросы ни в коей мере не умаляют высокого качества диссертации.

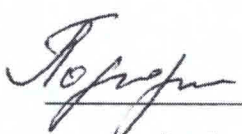
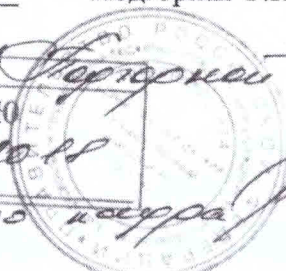
Диссертация Гуревича Алексея Александровича на тему: «Вычислительные методы для анализа подверженных ошибкам метабологеномных данных» соответствует основным требованиям, установленным Приказом от 01.09.2016 № 6821/1 «О порядке присуждения ученых степеней в Санкт-Петербургском государственном университете», соискатель Гуревич Алексей Александрович заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 03.01.09. – Математическая биология, биоинформатика. Пункт 11 указанного Порядка диссертантом не нарушен.

Председатель диссертационного совета

доктор биологических наук, профессор кафедры цитологии и гистологии, Биологического факультета, СПбГУ

26.11.2018

Результат проверки
по итогам обсуждения
работы

Подпись  Подгорная О.И.
ЗАВЕРЯЮ
26.11.2018

печать по кафедре