

ОТЗЫВ

члена диссертационного совета на диссертацию
Нурка Сергея Юрьевича на тему: «**СБОРКА ГЕНОМОВ
НЕКУЛЬТИВИРУЕМЫХ МИКРООРГАНИЗМОВ ПО ДАННЫМ
ВЫСОКОПРОИЗВОДИТЕЛЬНОГО СЕКВЕНИРОВАНИЯ**»,
представленную на соискание ученой степени кандидата физико-
математических наук по специальности
03.01.09 — «Математическая биология, биоинформатика»

Актуальность избранной темы

Развитие и активное внедрение методов секвенирования нового поколения открывает новые возможности для изучения разнообразных биологических объектов. В частности, в области микробиологии благодаря данным секвенирования нового поколения были кардинальным образом пересмотрены многие вопросы, касающиеся экологических отношений и филогенетических связей. Методы геномного секвенирования позволили выявить огромное количество так называемых некультивируемых микроорганизмов, о которых ранее и не подозревали. Это дало возможность описывать сообщества микроорганизмов более корректно и полно. И если идентификация микробов в составе сообществ методом секвенирования участков таксономически значимых районов ДНК сейчас не является большой проблемой, то реконструирование протяженных участков ДНК сопряжено с определёнными методическими сложностями.

На сегодняшний день, есть две группы подходов для изучения геномов некультивируемых микроорганизмов:

– секвенирование “из одной клетки”, осуществляемого с применением специализированных методов амплификации ДНК (например, MDA секвенирование),

– метагеномное секвенирование – секвенирование всей смеси ДНК из образца сообщества микроорганизмов. Особенности данных, получаемых с использованием каждой из этих методик, существенно усложняют реконструкцию продолжительных геномных фрагментов. В частности, применение к ним устоявшихся вычислительных подходов приводило к неоптимальным сборкам, содержащим заметное количество ошибок, по причине того, что многие вопросы, специфические для сборки данных MDA и метагеномного секвенирования, были разработаны недостаточно глубоко. Восполнению данных пробелов посвящена представленная работа. Исходя из чего ее актуальность не вызывает сомнения.

09/2 -124 от 26.04.2019

Научная новизна и практическая значимость работы

Работа Нурка С.Ю. внесла заметный вклад в разработку программного обеспечения для реконструкции прокариотических геномов на основании данных MDA и метагеномного секвенирования. Предложенные диссертантом решения были реализованы в геномном ассемблере SPAdes и метагеномном ассемблере metaSPAdes. SPAdes и metaSPAdes активно используются в ведущих лабораториях, занимающихся геномными и метагеномными проектами.

Основное содержание

Диссертационная работа изложена на 122 страницах и состоит из введения, трех глав и заключения, а также списка сокращений и условных обозначений, словаря терминов, списка литературы, списков рисунков и таблиц. Работа содержит 12 рисунков и 9 таблиц. Список литературы содержит 140 наименований.

В первой главе дается базовая информация о полногеномном секвенировании и сборке геномов на основе референсного генома, вводится понятие графов де Брюина и рассматривается возможность их применения для сборки геномов из наборов парных прочтений, полученных на секвенаторах Illumina, дается обзор подходов к секвенированию и анализу его результатов для некультивируемых микроорганизмов, обсуждаются основные проблемы реконструкции геномных фрагментов.

Во второй главе описываются алгоритмические и технические решения, лежащие в основе ассемблера SPAdes. С упором на анализ данных MDA секвенирования, дается сравнение SPAdes с другими ассемблерами, используемыми для сборки бактериальных геномов на основе данных MDA секвенирования.

Третья глава посвящена ассемблеру metaSPAdes и описанию результатов его сравнения с некоторыми другими метагеномными ассемблерами.

Можно отметить, что такое деление на главы не является традиционным для кандидатской диссертации. У математиков рекомендуется, чтобы первая глава была посвящена критическому анализу литературных источников по изучаемой проблеме. Во второй главе на базе выполненного анализа

приводятся новые решения автора исследования, в третьей главе описываются результаты реализации новых решений автора с их критической оценкой.

У биологов первым разделом также является обзор литературы, за которым следуют материалы и методы, а далее – результаты и обсуждение.

С одной стороны, решение Сергея Юрьевича видится вполне логичным, поскольку каждый из разделов посвящен одной глобальной идее. А в пределах каждого раздела можно найти элементы традиционной структуры работы.

С другой стороны, данный стиль изложения имеет свои недостатки. Они связаны с тем, что не всегда можно понять, где граница того, что сделано ранее, или сделано коллегами, а что непосредственно С.Ю. Нурком.

Вторая проблема, с которой пришлось столкнуться при прочтении – это описание материала. Я так и не поняла к чему относятся наборы данных ECOLI-SC и ECOLI-MC со страницы 65. Это результаты, полученные на одном и том же генотипе, или нет? На каком генотипе? В то же время для референсного генома дана информация до штамма. Аналогично на стр. 68 – SAUREUS. Мне представляется важным знать, что было подвергнуто сборке. Иначе, глядя на результаты сравнения ассемблеров, можно говорить только об общих тенденциях (которые весьма неплохи). В то же время в метагеномной части работы описание наборов данных присутствует. Если бы в работе был раздел или подраздел, посвященный материалам исследования, данной проблемы было бы можно избежать и видеть конкретные сведения для всех наборов данных.

При прочтении первой главы возникли вопросы

1. Что автор подразумевает под термином ГЕНОМ? Включает ли он только хромосому, или плазмиды тоже?
2. На странице 13 указано, что «в 1977 году был секвенирован первый вирусный геном, в 1995 – первый геном бактерии, в 1996 – археи, в 1998 – нематоды, в 2001 – геном человека.» С чем связано отсутствие в списке растений?
3. В этом разделе и далее: приставка «кило-» и кириллицей и латиницей обозначается маленькой буквой «к», а не большой.
4. В конце главы подробно описаны проблемы анализа метагеномных данных. Тут хотелось бы отметить один упущенный из обсуждения момент. Дело в том, что у микроорганизмов определенной «проблемой» являются не только горизонтально перенесенные гены, но и гомологи, которые у одних видов находятся на хромосоме, у других на плазмиде, у третьих и там и там в нескольких копиях. Учитывая размеры некоторых плазмид (Ti, Sym) и размеры

хромосом, вопрос об определении термина «геном» встает еще острее.

Главы 2 и 3 оставили благоприятное впечатление (за исключением одного недочета описанного выше), особенно в части весьма выигрышного сравнения результатов сборок при помощи SPAdes и metaSPAdes с результатами сборок другими ассемблерами.

Хотя работа написана достаточно понятным языком и построена логично, в ней имеются оформительские недочеты. Самым досадным из них является ошибка в названии собственного места работы на стр 8 (Центр Алгоритмической Юбиотехнологии).

Однако, сделанные замечания не умаляют достоинств работы. Результаты этой работы будут еще сотни, если не тысячи, раз использованы в геномных и метагеномных исследованиях.

Заключение

Представленное исследование выполнено на высоком научно-методическом уровне и **соответствует мировым стандартам** в области биоинформатики. Работа имеет огромное значение для фундаментальной и прикладной науки. Ее результаты по достоинству оценены научным сообществом. **Обоснованность научных положений и рекомендаций, сформулированных в диссертации, не вызывает сомнений.** Основные результаты работы представлены в виде докладов на международных конференциях и научных статей в престижных журналах, индексируемых Wos и Scopus, сотни раз процитированы. Актуальная версия пакета инструментов SPAdes (включая metaSPAdes) доступна по адресу <http://cab.spbu.ru/software/spades>

Диссертация Нурка Сергея Юрьевича на тему: «СБОРКА ГЕНОМОВ НЕКУЛЬТИВИРУЕМЫХ МИКРООРГАНИЗМОВ ПО ДАННЫМ ВЫСОКОПРОИЗВОДИТЕЛЬНОГО СЕКВЕНИРОВАНИЯ» соответствует основным требованиям, установленным Приказом от 01.09.2016 № 6821/1 «О порядке присуждения ученых степеней в Санкт-Петербургском государственном университете», соискатель Нурк Сергей Юрьевич заслуживает присуждения ученой степени кандидата физико-математических наук по специальности

03.01.09 — «Математическая биология, биоинформатика». Пункт 11
указанного Порядка диссертантом не нарушен.

Член диссертационного совета
Доктор биологических наук, доцент,
профессор кафедры генетики и
биотехнологии биологического
факультета Санкт-Петербургского
государственного университета



Татьяна Валерьевна Матвеева

24.04.2019