

## ОТЗЫВ

председателя 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

За последние 10 лет, достижения в области биотехнологий приоткрыли окно в удивительный, густонаселенный, не видимый нашему глазу, но имеющей огромное значение для человека мир микроорганизмов, обитающих внутри и вокруг нас. В то время как классические микробиологические (культуральные) методы оказались несостоятельны для подробного исследования микробиоты человека, для современных исследователей микробиома доступны высокопроизводительные методы метагеномного и метатранскриптомного секвенирования, а также метаболомики и single-cell геномики.

Диссертационная работа представляет результат исследований соискателя в области разработки методов восстановления (“сборки”) продолжительных геномных фрагментов из коротких прочтений. Актуальность работы обусловлена рассмотрением “неклассических” данных секвенирования, сборка которых оказывается сопряжена с уникальными дополнительными трудностями: секвенирования ДНК одиночных клеток (single-cell секвенирования) с использованием метода MDA, и секвенирования всей ДНК, выделенной из образца микробиоты (метагеномного секвенирования). Эти методики на текущий момент являются практически единственными способами анализа микроорганизмов, которые не удается культивировать в лабораторных условиях.

Беглое сравнение русского и английского вариантов текста свидетельствует о том, что перевод выполнен качественно. Дальнейшие комментарии относятся к русскому тексту диссертационной работы.

Во введении описывается актуальность исследования, научная новизна и практическая ценность диссертационной работы, сформированы цель и задачи исследования, а также положения диссертации, выносимые на защиту. Основной текст работы включает в себя три главы.

Первая глава представляет собой введение в полногеномное секвенирование с применением технологии Illumina, а также сборку из коротких прочтений и способы оценки ее качества. Стоит отметить, что раздел может быть интересен широкой аудитории, учитывая ограниченное количество подобных материалов на русском языке. Вторая глава описывает устройство геномного сборщика SPAdes и его особенностей, позволяющих эффективно работать с данными, полученными с применением метода MDA, который при всех недостатках до сих пор остается наиболее распространенным методом сборки single-cell секвенирования. Третья глава представляет сборщик metaSPAdes -- адаптацию SPAdes для сборки метагеномных данных. Вторая и третья главы включают подробные сравнения результатов сборки SPAdes и metaSPAdes с конкурирующими ассемблерами, обоснованно демонстрирующие эффективность разработанных подходов.

Основные методы и результаты, представленные в работе, опубликованы в международных рецензируемых научных журналах. Дополнительным подтверждением актуальности работы является широкое распространение разработанных программ в научном сообществе. Несмотря на то, что все работы выполнены в соавторстве, существенность вклада соискателя не вызывает сомнения.

Высокое качество представленной работы не исключает некоторые недочеты и опечатки:

09/2-126 от 26.04.2019

- Величину, названную в тексте “расстоянием вставки” принято называть “длиной вставки” (имеет отношение только к русскому варианту текста).
- Оказались пропущены некоторые ссылки, присутствие которых подразумевается в тексте (2 примера)
  - стр. 37 “При работе с данными секвенирования изолятов используется вероятностная модель схожая с использованной в работах cite.”
  - стр. 73 “При этом, недавние работы, осуществленные с использованием MDA секвенирования [70] и “синтетических длинных ридов” (TSLR cite),”
- Набор метрик, использованный при оценке качества метагеномных сборок в главе 3, несколько уже, чем использованный для оценки геномных сборок в главе 2. В первую очередь, отсутствует информация о количестве замен и инделов.
- Раздел 2.8 использует для обозначения графа нотацию, отличную от использованной во введении ( $DBG_k(R)$  вместо  $DBG(R, k)$ ).
- Обидная опечатка “Центр Алгоритмической Юиотехнологии”
- В подписи к таблице с результатами по набору данных ECOLI-MS он ошибочно назван ECOLI-SC.
- При перечислении свойств ошибок, допускаемых технологией Illumina вместо пунктов 3 и 4 оказалось два “третьих” пункта.

В порядке дискуссии также хотелось бы задать автору несколько вопросов:

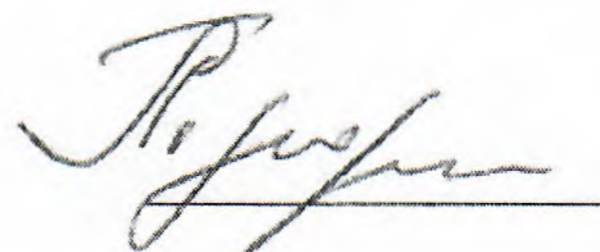
- В качестве метрики непрерывности сборок в таблицах результатов используется исключительно NGA50, почему именно она?
- Во введении неоднократно подчеркивалось значение выбора параметра длины k-мера, используемого при построении графа сборки. При этом стратегии этого выбора не освещались в данной работе. Как он осуществляется на практике?
- Метагеномные данные могут содержать и эукариотические геномы. Сталкивались ли вы с подобными случаями и эффективен ли для них metaSPAdes?
- Может ли использование наборов метагеномных данных, совмещенных с коллекциями изолятов, полученных из тех же образцов, помочь в решении проблемы тестирования метагеномных сборщиков?

Возникшие вопросы и указанные недочеты не носят принципиального характера и не умаляют достоинств представленной работы, а квалификация автора как исследователя не вызывает сомнения.

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.

Председатель диссертационного совета  
доктор биологических наук, профессор Кафедры цитологии и гистологии Биологического факультета, СПбГУ

26.04.2019



Подгорная О.И.