

ОТЗЫВ

**члена диссертационного совета на диссертацию
Пржибельского Андрея Дмитриевича РАЗРАБОТКА АЛГОРИТМОВ
ДЛЯ СБОРКИ ГЕНОМОВ И ТРАНСКРИПТОМОВ», представленную
на соискание ученой степени кандидата физико-математических наук по
специальности
03.01.09 — «Математическая биология, биоинформатика»**

Актуальность избранной темы

В последние годы методы секвенирования нового поколения развиваются очень бурно, поставляя массу новой информации о структуре и функционировании геномов живых организмов, их биоразнообразии, механизмах эволюции. Параллельно с совершенствованием молекулярно-биологических подходов к секвенированию геномов и транскриптомов, совершенствуются и биоинформатические методы, позволяя осуществлять сборку более точно, информативно. Однако, до сих пор остается много нерешенных вопросов. До сих пор вызывает трудности сборка геномов, содержащих большое количество повторов. Нередко повторяющиеся последовательности являются камнем преткновения, причиной возникновения брешей в данных геномного секвенирования. Еще сильнее усугубляется проблема при анализе данных секвенирования ДНК единичных клеток. В сборке транскриптомов тоже есть проблемы. Они часто связаны с наличием альтернативных транскриптов и сложностью их различения. Для решения этих проблем необходимо совершенствовать программы, осуществляющие сборку последовательностей (геномные ассемблеры), чему и посвящена представленная работа. Исходя из вышесказанного ее актуальность не вызывает сомнения.

Научная новизна и практическая значимость работы

Работа Пржибельского Андрея Дмитриевича внесла заметный вклад в разработку программного обеспечения для реконструкции

09/12-138 от 25.02.2010

прокариотических геномов. Предложенные диссертантом решения были реализованы в ассемблерах SPAdes и rnaSPAdes и модуле для разрешения повторов и скаффолдинга – exSPAnDer. Данные сборщики активно используются в ведущих лабораториях, занимающихся анализом данных NGS.

Основное содержание

Диссертационная работа изложена на 92 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 19 рисунков и 13 таблиц. Список литературы насчитывает 119 наименований.

Глава 1 «Задача сборки биологических последовательностей» посвящена обобщению имеющихся данных относительно общих алгоритмов сборки геномов и проблем, возникающих при работе с данными, полученными при секвенировании ДНК единичных клеток, при сборке повторяющихся последовательностей. Описаны трудности, которые возникают при сборке транскриптомов.

Вторая и третья главы посвящены сборке геномов и транскриптомов соответственно. Каждая из этих глав начинается с общего описания проблемы, за которым следует описание предлагаемых автором подходов для совершенствования процедуры сборки нуклеотидных последовательностей, а далее приводится сопоставление разработанного продукта с имеющимися аналогами по различным параметрам.

Хочется особо отметить очень хорошее: понятное, логичное, структурированное иллюстрированное удачными схемами и информативными таблицами изложение материала. После небольшой доработки многие разделы диссертации могут послужить основой для хорошего учебного пособия по биоинформатике.

Из работы понятно, какими способами можно решить проблему сборки повторяющихся последовательностей, начиная с выбора метода секвенирования и заканчивая усовершенствованиями алгоритма сборки.

В главе, посвященной сборке транскриптомов хорошо представлен вопрос выявления продуктов альтернативного сплайсинга, что является непростой задачей.

При сравнении своей разработки с другими сборщиками автор приводит таблицы с конкретными значениями метрик, смысл каждой из которых он предварительно расшифровывает.

Наилучшие показатели в таблицах выделены жирным шрифтом, показывая по каким параметрам какой из сборщиков превосходит другие.

Эти таблицы не только показывают, что разработанное программное обеспечение выгодно отличается от основных конкурентов, но и позволяет выбрать наилучшее решение для выполнения конкретных практических задач в будущем. Как известно: «Если рамки в одном становятся шире, то в другом непременно, увы, сужаются». По этой причине по некоторым параметрам другие сборщики могут оказаться более приемлемыми для решения конкретных задач.

В заключении к своей работе Андрей Дмитриевич не только подводит итог проделанного, но и говорит о перспективах. Видно, что его интерес направлен в сторону развития направления транскриптомики. Это очень приятно. В то же время вышли из поля зрения вопросы связанные с протяженными инвертированными повторами. А это тоже интересная область деятельности, поскольку более высокая вероятность ошибок на уровне секвенирования должна компенсироваться за счет алгоритмов сборки.

Замечания к данной работе у меня только редакторского плана. Они связаны с присутствием незначительного количества неудачных выражений. Несколько таких неудачных выражений сконцентрировались на стр 24. Я приведу их в качестве примера:

«На каждом шагу алгоритм принимает попытку выбрать лучшее возможное продолжение пути в соответствии с функцией оценки качества продолжений (англ. scoring function; описана далее). В случае если на вход алгоритму было передано несколько библиотек секвенирования, функции оценки строятся по каждой библиотеке независимо...»

В данном контексте выглядит не корректным выражение «на каждом шагу». Эта форма согласования используется в высказываниях типа: «Неприятности подстерегали на каждом шагу». В диссертации правильнее было бы сказать: на каждом этапе, на каждом шаге....

«Алгоритм принимает попытку» - тоже не совсем по-русски, хотя общий смысл ясен.

«На вход алгоритму передано несколько библиотек» - плохо согласованная фраза с элементами жаргонизмов.

Однако, сделанные замечания незначительны и не умоляют достоинств работы. Результаты этой работы будут еще сотни, если не тысячи, раз использованы в геномных и транскриптомных 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 опубликованы в виде пяти научных статей в журналах, индексируемых Wos и Scopus, в том числе три из них в журналах первой

квартили. Статьи сотни раз процитированы. Актуальная версия пакета инструментов SPAdes доступна по адресу [http:// cab.spbu.ru/software/spades](http://cab.spbu.ru/software/spades) rnaSPAdes – по адресу <http:// cab.spbu.ru/software/rnaspades>

Диссертация Пржибельского Андрея Дмитриевича «РАЗРАБОТКА АЛГОРИТМОВ ДЛЯ СБОРКИ ГЕНОМОВ И ТРАНСКРИПТОМОВ» соответствует основным требованиям, установленным Приказом от 01.09.2016 № 6821/1 «О порядке присуждения ученых степеней в Санкт-Петербургском государственном университете», соискатель Пржибельский Андрей Дмитриевич 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



21.02.2020