

## ОТЗЫВ

члена диссертационного совета на диссертацию Михеенко Аллы Александровны на тему: «РАЗРАБОТКА ВЫЧИСЛИТЕЛЬНЫХ МЕТОДОВ ДЛЯ АНАЛИЗА И ВИЗУАЛИЗАЦИИ ЭУКАРИОТИЧЕСКИХ ГЕНОМНЫХ СБОРОК», 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 структуры и функций геномов живых организмов. Хотя первые геномы были отсеквенированы десятки лет назад, до сих пор проблему сборки геномов нельзя считать решенной. Даже для модельных организмов остались вопросы, касающиеся качества сборки геномов, не говоря уже о сложных полиплоидных геномах растений, богатых разного рода повторяющимися последовательностями. Богатые повторами области, составляющие значительную часть многих эукариотических геномов, являются одной из самых больших проблем сборки геномных последовательностей, что требует разработки специальных программных подходов.

Развитие методов секвенирования и алгоритмов сборки привело к быстрому росту количества новых отсеквенированных геномов. При этом оценка качества сборки осталась актуальной задачей.

Накопление большого массива информации требует новых подходов для визуализации данных как для представления известной информации, так и для выявления скрытых закономерностей в данных, что может быть полезно и в различных биологических исследованиях, и для совершенствования алгоритмов разработчиками программного обеспечения.

Решению всех указанных выше вопросов и посвящена представленная работа, исходя из чего ее актуальность не вызывает сомнения.

## **Научная новизна и практическая значимость работы**

Работа Михеенко Аллы Александровны внесла заметный вклад в разработку программного обеспечения для оценки качества сборок больших эукариотических геномов, анализа длинных tandemных повторов, визуализации данных структуры и качества сборки геномов, что вылилось в создание следующих программных продуктов, активно используемых в ведущих мировых лабораториях:

-QUAST-LG: инструмент для оценки качества сборок больших эукариотических геномов (<http://cab.spbu.ru/software/quast-lg/>) и Icarus (<http://cab.spbu.ru/software/icarus/>): модуль визуализации для пакета QUAST (зарегистрирован в Роспатенте, регистрационный номер 2016619067);

– AGB: инструмент для интерактивной визуализации и анализа графов сборки (<http://cab.spbu.ru/software/agb/>).

– TandemTools: программный пакет для анализа длинных tandemных повторов (<http://cab.spbu.ru/software/tandemtools/>).

## **Основное содержание**

Диссертация представлена на 108 страницах, включает введение, три главы и заключение. Текст иллюстрируется 31 рисунком и 10 таблицами. Список литературы содержит 124 наименования. Такое деление на главы представляется удачным для логичной подачи материала исследования.

Первая глава посвящена проблемам сборки геномов и различным подходам к оценке качества сборки генома. Обсуждаются отличительные особенности геномов эукариот и прокариот, что влечет отличия в подходах к их сборке. Далее подробно описан инструмент QUAST-LG, разработанный для анализа и оценки качества больших эукариотических геномных сборок.

К этой главе у меня возникло два вопроса по сути излагаемого материала.

При описании подходов для оценки полноты сборки генома автор описывает как один из вариантов, поиск универсальных высококонсервативных генов и определяет их как «набор генов, которые, как ожидается, находятся в одной копии (т.е., не происходит их дублирования или потери) в любом секвенированном геноме из соответствующей филогенетической группы». Хочется узнать, какие ограничения у этого подхода видит автор с биологической точки зрения.

На стр. 24 автор пишет о возможности применения QUASt-LG для обнаружения мобильных элементов, а далее указывает, что альтернативным подходом была бы «классификация МГЭ на основе его геномной последовательности, но она выходит за рамки данной работы». Хотелось бы услышать, о каких работах за рамками данного исследования идет речь.

Вторая глава описывает новые алгоритмы обработки длинных для реконструкции tandemных повторов. Автор обсуждает как достижения, так и сохраняющиеся проблемы в области сборки и анализа длинных tandemных повторов. В этой главе представлен пакет TandemTools для выравнивания длинных ридов на сборки длинных tandemных повторов и анализа и улучшения их качества, представлено сравнение TandemTools с аналогами при анализе симулированных данных.

Третья глава посвящена проблеме визуализации данных в геномных исследованиях. Обсуждаются существующие подходы, сферы их применения и ограничения. Представлены инструменты Icarus для визуализации сборок генома и AGB для визуализации графов сборок. Данная глава написана понятно и вызвала единственный технический вопрос.

Чем обусловлены цветовые решения в Icarus при оценке подобия между сборками: синий цвет (для правильных контигов) или оранжевый (для неправильно собранных контигов). Для правильных контигов напрашивается зеленый цвет.

Вся работа написана достаточно понятным языком и построена логично, однако, в ней имеются незначительные оформительские недочеты, представленные, главным образом, неточными формулировками. Например: Стр 13. «У прокариот несколько генов могут быть организованы на одном опероне: они расположены рядом и управляют ферментами, которые осуществляют последовательные или близкие реакции синтеза». Следовало написать, что гены КОДИРУЮТ ферменты, а не управляют ими.

Стр. 16 – «Количество нуклеотидных оснований»

Стр. 22 - «качество человеческих сборок»

Однако, сделанные замечания не умоляют достоинств работы. Результаты этой работы будут еще сотни, если не тысячи, раз использованы в геномных 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. Актуальные версии инструментов доступны по адресам (<http://cab.spbu.ru/software/quast-lg/>), (<http://cab.spbu.ru/software/icarus/>), (<http://cab.spbu.ru/software/agb/>), – (<http://cab.spbu.ru/software/tandemtools/>).

Диссертация Михеенко Аллы Александровны на тему: «РАЗРАБОТКА ВЫЧИСЛИТЕЛЬНЫХ МЕТОДОВ ДЛЯ АНАЛИЗА И ВИЗУАЛИЗАЦИИ ЭУКАРИОТИЧЕСКИХ ГЕНОМНЫХ СБОРОК» соответствует основным требованиям, установленным Приказом от 01.09.2016 № 6821/1 «О порядке присуждения ученых степеней в Санкт-Петербургском государственном университете», соискатель Михеенко Алла Александровна 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

30.12.2020