

ОТЗЫВ

члена диссертационного совета на диссертацию Банкевича Антона Викторовича на тему: «Модели, численные методы и комплекс программ для сборки геномов из нестандартных данных секвенирования», представленную на соискание ученой степени кандидата физико-математических наук по специальности 05.13.18. – Математическое моделирование, численные методы и комплексы программ»

Диссертация посвящена решению задачи геномной сборки, то есть восстановлению геномной строки из данных секвенирования: набора коротких фрагментов геномной строки (прочтений), содержащих случайные ошибки. Задача геномной сборки из данных секвенирования является неотъемлемой частью современных биологических и медицинских исследований, что определяет ее актуальность и востребованность.

Различия в свойствах входных данных (распределение длин прочтений, модель ошибок в прочтениях, свойства искомой геномной строки) делают необходимым оптимизацию алгоритмов геномной сборки для различных типов данных. В диссертации рассматриваются два типа входных данных: одноклеточное секвенирование, для которого предложен алгоритм коррекции ошибок во входных данных, и метагеномное секвенирование, для которого предложено решение задачи оценки суммарной длины геномных строк на основе данных совместного секвенирования двумя технологиями.

Сборка данных одноклеточного секвенирования сложна из-за неравномерности распределения позиций прочтений в геномной строке и обилия ошибок специального вида: химерических прочтений. Для решения задачи сборки генома из данных одноклеточного секвенирования была использована модель, основанная на графе де Брюйна. В графе де Брюйна каждому ребру соответствует известная строка, геном соответствует некоторому (неизвестному) циклу в графе, а задача сборки генома сводится к поиску этого цикла. Несмотря на то, что графы де Брюйна являются уже классическим инструментом для решения задачи геномной сборки, в данной работе была предложена модификация графовой модели, специально адаптированная для применения к данным одноклеточного секвенирования. Важной задачей при использовании графовой модели секвенирования является поиск ошибок в графе де Брюйна, то есть рёбер, через которые не проходит геномный цикл. В работе предложен алгоритм поиска таких рёбер в идеальных данных, а также эвристические алгоритмы поиска ошибочных рёбер в случае неполноты входных данных.

Для оценки суммарной длины геномных строк в метагеноме вместо графовой модели была использована вероятностная модель секвенирования. Особенности свойств использовавшихся технологий секвенирования (длина прочтений, низкая вероятность ошибок) были применены для получения дополнительной информации: множества пар прочтений разных технологий, сгенерированных из близких участков геномных строк.

Разработанные модели и методы были реализованы в рамках программного обеспечения "SPAdes": Saint-Petersburg Assembler для решения задачи геномной сборки. Программное обеспечение SPAdes используется для решения таких важнейших биологических задач, как сборка данных одноклеточного секвенирования и сборка бактериальных геномов. Апробация результатов проводилась на представительных международных конференциях. Автор отмечает, что данное программное обеспечение

РК 000-21 м. 15.10.18

используется во многих лабораториях и институтах во всём мире, что подтверждается соответствующими публикациями.

Среди недостатков работы следует отметить, что не совсем четко представлены используемые численные методы, а для обоснования предложенных моделей был бы полезен более глубокий математический анализ рассматриваемых объектов исследования. В работе присутствуют опечатки и стилистические ошибки.

Несмотря на приведенные выше замечания, диссертация Банкевича Антона Викторовича на тему: «Модели, численные методы и комплекс программ для сборки геномов из нестандартных данных секвенирования» соответствует основным требованиям, установленным Приказом от 01.09.2016 № 6821/1 «О порядке присуждения ученых степеней в Санкт-Петербургском государственном университете», соискатель Банкевич Антон Викторович заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 05.13.18. – Математическое моделирование, численные методы и комплексы программ».

Член диссертационного совета
Доктор физико-математических наук,
доцент, профессор



Котина Е.Д.

29.05.2018