

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.

Актуальность

Появление в биологии технологии геномного секвенирования (метода обрыва цепи), предложенной в 1977 году Ф. Сенгером (Нобелевская премия 1980г.) ознаменовало наступление новой для биологии и ряда смежных наук эпохи получения новых биологических данных. Анализ данных секвенирования стал важнейшей задачей на стыке биологии, математики, статистики и компьютерных наук. Накопленные сведения о свойствах данных секвенирования позволили перенести их анализ в область компьютерного моделирования на основе проблемно-ориентированных программных комплексов. Одним из основных успехов в этой области стала расшифровка человеческого генома, что стало настоящим прорывом для медицины и биологии.

На расшифровку человеческого генома ушло более десяти лет и были затрачены усилия научных групп из десятков университетов во всём мире. Однако, технологии Секвенирования Нового Поколения (СНП), появившихся в начале XXI века позволили автоматизировать геномное секвенирование, в результате чего процесс генерации данных секвенирования значительно упростился и ускорился. В результате, многие биологические задачи, ранее решавшиеся на основе множественных биологических экспериментов оказалось возможным решить за счёт вычислительного анализа данных секвенирования нового поколения. Одной из важнейших таких задач является задача восстановления геномной строки (последовательности нуклеотидов в молекуле ДНК) из данных секвенирования, то есть набора её случайных подстрок, содержащих ошибки.

В свете сказанного тема диссертационной работы Банкевича А.В., в которой поставлена цель совершенствования существующих методов вычислительного анализа данных секвенирования и, в частности, задача восстановления геномной строки, является, несомненно, актуальной.

Структура работы

Наибольшее внимание в работе уделено решению двух частных, но исключительно важных проблем: проблема сборки данных одноклеточного секвенирования и задача оценки суммарной длины геномных строк в метагеноме.

В работе используется две математические модели геномного секвенирования: графовая модель, основанная на графе де Брюйна и вероятностная модель для описания совместного секвенирования метагенома двумя технологиями секвенирования и различными свойствами. В рамках предложенных моделей были сформулированы вычислительные задачи исправления ошибок в графе де Брюйна и оценки суммарной длины геномов в метагеноме, предложены численные методы для их решения. Также был предложен ряд эвристических модификаций предложенных численных методов для повышения качества результатов с учётом особенностей реальных данных.

PK 09/2-104 от 05.06.18

Достоинства работы

Несомненным достоинством рецензируемой работы является создание эффективных программных комплексов, способных качественно улучшить работу в сфере вычислительной геномики. Из других достоинств следует отметить:

- алгоритм и реализация поиска ошибок в графе де Брюйна на основе теории графов
- идею и разработку метода оценки суммарной длины геномов в метагеноме на основе данных совместного секвенирования
- полноту решения поставленных в диссертации задач

Замечания по работе.

В качестве недостатка можно отметить слабое соответствие специальности 05.13.18, поскольку, предложенные математические модели, по сути, не являются новыми, а ранее уже использовались для решения аналогичных задач. Однако, изменения, внесённые автором в классические графовые и вероятностные математические модели геномного секвенирования для отражения особенностей конкретного типа данных (данные одноклеточного секвенирования и данные совместного секвенирования метагенома двумя технологиями секвенирования), являются существенными для решения поставленных задач и могут считаться оригинальным результатом в области математического моделирования.

Заключение

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 технических наук,
профессор Кафедры компьютерного моделирования и многопроцессорных систем
Дегтярёв Александр Борисович

А.Б.Дегтярёв

