

## ОТЗЫВ

на диссертацию Гуревича Алексея Александровича на тему: «Вычислительные методы для анализа подверженных ошибкам метабологеномных данных», представленную на соискание ученой степени кандидата физико-математических наук по специальности 03.01.09 – Математическая биология, биоинформатика

Диссертационная работа А.А. Гуревича посвящена разработке новых методов анализа результатов геномных данных, полученных с использованием современных технологий секвенирования, и метаболомных данных, полученных с использованием масс-спектрометрии, с целью повышения точности метабологеномных исследований. Особенностью метабологеномных исследований является то, что с одной стороны они требуют анализа геномных последовательностей для выявления в ней фрагментов, потенциально кодирующих метаболиты, с другой стороны, необходимо выявить как можно больше метаболитов, присутствующих в исследуемом образце, на основе масс-спектрометрических данных, и выявить соответствие между геномами и метаболитами. Знание о таком соответствии может служить основой для открытия новых природных антибиотиков, что является актуальной задачей современной медицины для преодоления бактериальной резистентности в отношении известных лекарств.

Автор адекватно описал современное состояние и известные проблемы в области сборки геномов, оценки качества полученных геномов, а также оценки качества результатов метагеномных исследований и анализа масс-спектрометрических данных метаболитов. Текст диссертации показывает, что автор прекрасно разбирается в этих областях. В диссертационной работе было найдено решение двух схожих по постановке задач (геномной и метаболомной), суть которых основана на сравнение экспериментальных данных с референсным образцом, который может отличаться от исследуемого образца из-за мутаций или модификаций. Для обеих рассматриваемых задач были разработаны алгоритмические решения, реализованные в виде новых программных продуктов. При этом следует подчеркнуть, что разработанные программы оказались настолько качественными, что информация о них была опубликована в ведущих мировых научных изданиях, и их использование получило широкое распространение в мире.

Диссертационная работа Гуревича А.А. представлена на 118 страницах и состоит из введения, четырех глав и заключения. Работа включает 24 рисунка и 8 таблиц. Список литературы содержит 169 источников.

*Вх 09/2-261 от 30.11.12*

Во введении описывается актуальность исследования, научная новизна и практическая ценность диссертационной работы, сформированы цель и задачи исследования, а также положения диссертации, выносимые на защиту.

Каждая из четырех глав соответствует созданию определенного алгоритма и имеет схожую структуру, которая состоит из нескольких подразделов: (1) небольшой литературный обзор, связанный с областью исследований, на решение которых направлен создающийся алгоритм; (2) описание алгоритма; (3) результаты валидации и (4) апробация созданного алгоритма на конкретных экспериментальных данных.

Первая глава посвящена разработке и описанию алгоритма для разграничения ошибок сборки от естественных вариаций в секвенированном геноме. Разработанный алгоритм был реализован в программе QUASt, характеристики которой (метрики качества, визуализация результатов, веб-интерфейс) подробно описаны. В данной главе также приводится сравнение геномных сборок одноклеточного секвенирования *S. aureus* с использованием QUASt и оценивается его производительность.

Во второй главе описывается алгоритм для оценки метагеномных сборок, который был реализован в программе MetaQUAST. Дается небольшое описание метагеномных исследований. Приводятся описание дополнительных метрик и характеристик, использующихся в MetaQUAST. Оценивается чувствительность определения видового состава, эффективность уточнения мисассемблов и производительность MetaQUAST.

Третья глава посвящена дерепликации пептидных природных соединений посредством поиска их масс-спектров в соответствующей базе данных. Приводится описание исследований в области пептидных природных соединений, алгоритм Dereplicator, его валидация и применение для известных масс-спектрометрических данных.

В четвертой главе описывается алгоритм вариативной идентификации пептидных природных соединений VarQuest. Описывается проблема, на решение которой направлен данный алгоритм, варианты пептидных природных соединений и существующие стратегии вариативной идентификации пептидных природных соединений. Приводится схема и результаты тестирования VarQuest, а также описываются преимущества использования VarQuest для обнаружения модификаций и мутаций в пептидных природных соединениях, оценки их вариантов и разнообразия.

Диссертационная работа Гуревича А.А. актуальна, проведена на большом объеме данных и содержит статистически достоверные результаты. Достоверность и новизна результатов подтверждена статьями, опубликованными в международных рецензируемых научных журналах. Текст диссертации логически построен и написан хорошим

