

ОТЗЫВ

на диссертационную работу **Сафоновой Яны Юрьевны** на тему
**«ИСПОЛЬЗОВАНИЕ ГРАФОВЫХ МОДЕЛЕЙ ДЛЯ
БИОИНФОРМАТИЧЕСКОГО АНАЛИЗА ГИПЕРВАРИАБЕЛЬНЫХ
БИОЛОГИЧЕСКИХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ»**,
представленную на соискание ученой степени
кандидата физико-математических наук
по специальности 05.13.18 —
«Математическое моделирование, численные методы и
комплексы программ»

Появление технологий секвенирования и успех проекта “Геном человека” определили направление развития современной биологии. В частности, технологии секвенирования дали возможность представлять сложные биологические объекты в виде простых “цифровых” объектов и, таким образом, применять для их анализа вычислительные средства. Так, например, геном представляется в виде строки в четырехбуквенном алфавите и анализируется с помощью эффективных строчных алгоритмов и структур данных.

В результате, массовое использование технологий секвенирования в начале XX века положило начало сотрудничеству биологов с математиками, алгоритмистами и статистиками. В конечном счете, это сотрудничество сформировало биоинформатику, новое направление науки, посвященное разработке методов для обработки биологических данных и решения биологических задач. Биоинформатика является очень быстро развивающейся областью, и появление в ней новых технологий и экспериментальных средств влечет за собой развитие вычислительных методов.

В данной работе автор сосредоточился на решении двух вычислительных задач биоинформатики: сборка высокополиморфных диплоидных геномов и построение репертуаров иммунных рецепторов с помощью данных секвенирования. Как отмечает автор, обе эти задачи объединяет общая цель: устранение артефактов технологий секвенирования и сохранение естественного разнообразия изучаемых объектов. Несмотря на то, что такая цель присутствует в том или ином виде во многих вычислительных задачах биоинформатики, в случае диплоидных геномов и иммунных рецепторов, она становится особенно

важной. Информация о разнообразии диплоидного генома является путем к дальнейшему популяционному анализу, в то время как разнообразие иммунных рецепторов отражает состояние иммунной системы и используется в биомедицинских задачах.

Автор предложил использование двух математических моделей: графа де Брюйна (для решения задачи сборки генома) и графа Хэмминга (для решения задачи восстановления репертуара иммунных рецепторов). Несмотря на то, что граф де Брюйна является стандартной моделью для представления фрагментированных данных геномного секвенирования, автор проанализировал особенности такого представления для диплоидных геномов и изучил его свойства в случае высокого уровня полиморфизмов. Автор также впервые предложил использовать граф Хэмминга для представления данных иммуносеквенирования.

В качестве численных методов, автором предложено два эвристических алгоритма. Первый связан с модификацией графа де Брюйна, построенного для высокополиморфного диплоидного генома. Второй алгоритм предназначен для декомпозиции графа Хэмминга на плотные подграфы и последующего вычисления на их основе компонент репертуара иммунных рецепторов.

Предложенные алгоритмы были реализованы в виде программных комплексов `dipSPAdes` (алгоритм для сборки высоко полиморфных диплоидных геномов) и `IgRepertoireConstructor` (алгоритм для построения репертуара иммунных рецепторов из данных иммуносеквенирования). Оба программных комплекса являются зарегистрированными продуктами. Также автор предложил описание программного комплекса `IgSimulator`, предназначенного для моделирования иммунных рецепторов и оценки качества алгоритмов для их построения.

Среди недостатков рецензируемой работы хочу отметить то, что используемые математические модели не были описаны явно. В ряде случаев, эффективность предложенных эвристик обусловлена особенностью проанализированных данных: так, например, первый этап алгоритма `dipSPAdes` основан на предположении о том, что гаплоконтиги, покрывающие гомологичные позиции диплоидного генома, будут иметь далекие точки разрыва. В связи с этим, возникает вопрос, каков будет результат работы алгоритма в случае геномов с более высоким уровнем повторности. Отсутствие оценок эффективности работы алгоритмов и серьезного математического анализа рассматриваемых объектов

исследования в виде графов де Брюйна и Хэмминга, обедняет ценность диссертации, представленной к защите по физико-математическим наукам.

Несмотря на изложенные недостатки, результаты, полученные Сафоновой Я.Ю., востребованы биоинформатическим сообществом. Рецензируемая диссертация выполнена на стыке различных наук и соответствует специальности.

В соответствие с вышеизложенным, могу заключить, что по своему содержанию и полученным результатам, рассматриваемая диссертация соответствует требованиям, предъявляемым к диссертациям на соискание ученой степени кандидата наук, а ее соискатель, Сафонова Яна Юрьевна, заслуживает присуждения степени кандидата физико-математических наук.

Заслуженный работник высшей школы РФ,
доктор физико-математических наук,
профессор,
Овсянников Дмитрий Александрович

31.03.2017

