

ОТЗЫВ

**научного руководителя Певзнера Павла Аркадьевича
на диссертационную работу
Дворкиной Татьяны Евгеньевны
“РАЗРАБОТКА АЛГОРИТМОВ ДЛЯ АНАЛИЗА ГРАФОВ
ГЕНОМНОЙ СБОРКИ И ГЕНОМНЫХ СБОРОК”,
представленную на соискание ученой степени
кандидата физико-математических наук**

Специальность 1.5.8. Математическая биология, биоинформатика
Санкт-Петербург – 2023

Диссертация Дворкиной Т.Е. посвящена разработке эффективных алгоритмов и программных продуктов для анализа сложных участков геномных сборок. Задача сборки генома является одной из важнейших задач биоинформатики. Несмотря на существование целого ряда высокоэффективных геномных сборщиков, в геномной последовательности встречаются участки, которые регулярно представлены фрагментами в финальной сборке или вовсе отсутствуют. Первые две части диссертации Дворкиной Т.Е. посвящены разработке алгоритмов для прикладывания длинных прочтений и последовательностей белков к графам сборки из коротких прочтений. Представление коротких прочтений в виде графа сборки на промежуточном этапе является одним из основных подходов к сборке генома. Такое представление содержит больше связей между прочтениями, чем финальная сборка, и с учетом дополнительной информации, такой как прикладывания на граф длинных прочтений или последовательностей белков, позволяет получить больше информации о сложных участках. Основными результатами Дворкиной Т.Е. в рамках данного направления стали программные продукты SPAligner и ORFograph. Третья глава диссертации посвящена разработке алгоритмов и инструментов для анализа первыхборок центромерных участков генома человека, получение которых до недавнего времени было невозможным, так как

центромеры представляют собой длинные последовательности с высоким содержанием повторов. Как следствие, появилась потребность в автоматических методах для обработки и анализа центромерных последовательностей. В рамках данного направления были разработаны такие инструменты как StringDecomposer, CentromereArchitect, и HORmon. Инструменты, представленные в диссертации, находятся в открытом доступе и используются исследователями по всему миру, в том числе в рамках таких важных проектов как Консорциум Telomere-to-Telomere.

В процессе работы над диссертацией Дворкина Т. Е. продемонстрировала способности к самостоятельной исследовательской работе, грамотному изложению полученных результатов в научных текстах и в устных докладах.

Научные результаты, представленные в диссертации, были опубликованы в пяти статьях, из которых в четырёх Дворкина Т.Е. является первым автором. Все статьи были опубликованы в журналах, индексируемых базами Scopus и Web of Science. Также полученные результаты представлялись на международных конференциях в виде устных и постерных докладов.

Тема диссертационной работы актуальна, полученные научные результаты отличаются новизной и практической значимостью. Диссертация Дворкиной Т.Е. “Разработка алгоритмов для анализа графов геномной сборки и геномных сборок”, представленная на соискание степени кандидата физико-математических наук, соответствует требованиям, предъявляемым к диссертациям на соискание степени кандидата наук, и может быть допущена к защите.

Научный руководитель:

Певзнер Павел Аркадьевич,

кандидат физико-математических наук, профессор Калифорнийского университета в Сан-Диего, США,

Директор Национального центра вычислительной масс-спектрометрии,
США.

ppezvner@ucsd.edu

21 Февраля 2023 года

Paul Pez