

**ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА 24.1.120.01,
созданного на базе
Федерального государственного бюджетного учреждения науки
Института системного программирования им. В.П. Иванникова
Российской академии наук
Министерства науки и высшего образования РФ
по диссертации на соискание ученой степени кандидата наук**

аттестационное дело № _____

решение диссертационного совета от 07 декабря 2023 года № 2023/15

О присуждении Карпулевичу Евгению Андреевичу, гражданину РФ, ученой степени кандидата физико-математических наук.

Диссертация «Построение программного конвейера для выравнивания последовательностей в приложениях биоинформатики» по специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей» принята к защите 06 октября 2023, протокол № 2023/10 диссертационным советом 24.1.120.01, созданным на базе Федерального государственного бюджетного учреждения науки Института системного программирования им. В.П. Иванникова Российской академии наук (ведомственная принадлежность: Министерство науки и высшего образования РФ; адрес: 109004, г. Москва, ул. А. Солженицына, дом 25), создан Приказом Минобрнауки России о советах по защите докторских и кандидатских диссертаций от 2 ноября 2012 г. № 714/нк.

Соискатель Карпулевич Евгений Андреевич, 1990 года рождения.

В 2012 году соискатель окончил факультет вычислительной математики и кибернетики Московского государственного университета имени М.В. Ломоносова. В 2016 году окончил аспирантуру Федерального государственного бюджетного образовательного учреждения высшего образования «Национальный исследовательский университет «МЭИ».

Работает в должности научного сотрудника в отделе Информационных систем в Федеральном государственном бюджетном учреждении науки Институт системного программирования им. В.П. Иванникова Российской академии наук (ведомственная принадлежность: Министерство науки и высшего образования РФ).

Диссертация выполнена в Федеральном государственном бюджетном учреждении науки Институте системного программирования им. В. П. Иванникова Российской Академии Наук (ведомственная принадлежность: Министерство науки и высшего образования РФ).

Научный руководитель – кандидат физико-математических наук Турдаков Денис Юрьевич, заведующий отделом Информационных систем ИСП РАН.

Официальные оппоненты:

1. Макеев Всеволод Юрьевич, доктор физико-математических наук, член-корр. РАН, г.н.с. Федерального государственного бюджетного учреждения науки Институт общей генетики им. Н.И.Вавилова Российской академии наук
2. Алимова Ильсеяр Салимовна, кандидат технических наук, старший преподаватель кафедры программной инженерии Института информационных технологий и интеллектуальных систем Федерального государственного автономного образовательного учреждения высшего образования «Казанский (Приволжский) федеральный университет»

дали положительные отзывы на диссертацию.

Ведущая организация Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского», город Нижний Новгород в своем положительном заключении подписанном Иванченко Михаилом Васильевичем, доктором физико-математических наук, заведующим кафедрой прикладной математики Института информационных технологий, математики и механики

Национального исследовательского Нижегородского государственного университета имени Н. И. Лобачевского и утвержденная Грязновым Михаилом Юрьевичем, кандидатом физико-математических наук, проректором по науке и инновациям Национального исследовательского Нижегородского государственного университета имени Н. И. Лобачевского, указала, что диссертационная работа содержит новые научные результаты, которые могут быть использованы в научных исследованиях и промышленных проектах.

Соискатель имеет 15 опубликованных работ, в том числе по теме диссертации опубликовано 3 работы, из них в рецензируемых научных изданиях опубликовано 3 работы, кроме того, получено свидетельство о государственной регистрации программы для ЭВМ:

1. Гугучкин Е.П., Карпулевич Е.А. Модификация алгоритма выравнивания коротких прочтений для повышения качества пайплайна обработки данных полногеномного секвенирования человека. Труды Института системного программирования РАН. 2023;35(2):235-248. [https://doi.org/10.15514/ISPRAS-2023-35\(2\)-17](https://doi.org/10.15514/ISPRAS-2023-35(2)-17)
2. Кондратьева О.А., Карпулевич Е.А. Модификация метода расчета полигенных рисков с использованием графа вариации. Труды Института системного программирования РАН. 2022;34(2):191-200. [https://doi.org/10.15514/ISPRAS-2022-34\(2\)-15](https://doi.org/10.15514/ISPRAS-2022-34(2)-15)
3. Albert, E. A., Kondratieva, O. A., Baranova, E. E., Sagaydak, O. V., Belenikin, M. S., Zobkova, G. Y., Kuznetsova, E. S., Deviatkin, A. A., Zhurov, A. A., Karpulevich, E. A., Volchkov, P. Y., & Vorontsova, M. V. (2023). Transferability of the PRS estimates for height and BMI obtained from the European ethnic groups to the Western Russian populations. In *Frontiers in Genetics* (Vol. 14). Frontiers Media SA. <https://doi.org/10.3389/fgene.2023.1086709>
4. Свидетельство о государственной регистрации программы для ЭВМ №2022614027 “Программный комплекс "EVOGEN WEB SYSTEM””

В статье [1] автором поставлена задача совместно с соавтором, автору принадлежит основная часть: разделы 2-4, реализация инструмента и финальное редактирование текста.

В статьях [2;3] вместе с соавторами поставлена задача и проводилась редакторская правка, разработка программных конвейеров выполнена автором.

На основе разработанного программного конвейера получено свидетельство о государственной регистрации программы для ЭВМ [4].

Выбор официальных оппонентов и ведущей организации обосновывается их компетентностью и достижениями в области соответствующей специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей», наличием публикаций по теме диссертации и способностью определить научную и практическую ценность диссертации.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

- разработан новый метод выравнивания генетических последовательностей, позволяющий повысить точность идентификации генетических вариантов при анализе данных полногеномного секвенирования человека;
- доказана перспективность использования разработанного метода для выравнивания последовательностей в приложениях биоинформатики;
- предложена архитектура системы анализа данных NGS позволяющая расширить границы применимости полученных результатов.

Теоретическая значимость исследования обоснована тем, что:

- Разработан метод выравнивания генетических последовательностей на референсный геном с использованием данных об известных генетических вариантах.
- Разработаны алгоритмы в составе метода выравнивания генетических последовательностей и аналитические оценки их вычислительной и пространственной сложности через доказательство соответствующих теорем.

- Оценки, полученные в результате доказательства теорем, показывают, что вычислительная сложность алгоритмов построения создания модифицированного индекса референсной генетической последовательности остается линейной, а вычислительная сложность алгоритмов выравнивания генетических последовательностей на модифицированный индекс не изменяется по сравнению с выравниванием на индекс референсного генома.

Значение полученных соискателем результатов исследования для практики подтверждается тем, что:

- Разработана архитектура системы анализа данных NGS на базе программного конвейера, реализующего предложенный метод выравнивания генетических последовательностей
- В рамках предложенной архитектуры реализован программный конвейер на основе популярного инструмента выравнивания ридов minimap2. Программный конвейер апробирован на данных проекта "The Genome in a Bottle".
- Разработанный программный конвейер для обработки данных секвенирования ДНК человека с использованием модифицированного индекса превосходит существующие реализации по полноте (Recall) идентификации однонуклеотидных полиморфизмов, кроме того разработанный программный конвейер с использованием модифицированного индекса превосходит аналогичный с использованием стандартного индекса по качеству (по полноте, Recall) и количеству качественно выровненных ридов, при этом значения остальных метрик (Precision и F-score) не уменьшаются.
- Результаты исследования, содержащиеся в диссертационной работе Карпулевича Евгения Андреевича на тему «Построение программного конвейера для выравнивания последовательностей в приложениях биоинформатики», используются в процессе обработки данных секвенирования в ООО "Эвоген" и дают положительный эффект, что подтверждено Актом о внедрении результатов диссертации.

Оценка достоверности результатов исследования выявила:

- **теория** построена на известных проверенных данных и согласуется с опубликованными данными ученых в области выравнивания последовательностей;
- **идея базируется** на анализе научных публикаций, опубликованных в ведущих российских и зарубежных изданиях, а также практике применения программных конвейеров в биоинформатике;
- **использовано** сравнение авторских результатов с полученными ранее результатами, представленными в независимых источниках по рассматриваемой тематике;
- **установлено** совпадение авторских результатов с результатами, представленными в независимых источниках по данной тематике, в тех случаях, когда такое сравнение является обоснованным;
- **использованы** современные методики обработки исходной информации.

Личный вклад соискателя состоит в участии на всех этапах работы над диссертацией: в непосредственном участии соискателя в получении исходных данных и в научных экспериментах, в личном участии в апробации результатов исследования, в обработке и интерпретации экспериментальных данных, выполненных лично автором, в подготовке основных публикаций по выполненной работе, в апробации результатов на научных конференциях и оформлении диссертационной работы.

В ходе защиты диссертации были высказаны следующие критические замечания:

1. В диссертации показана работа реализации метода только на одном наборе тестовых данных, было бы интересно исследовать результаты анализа на нескольких наборах данных
2. В главе 3 следовало бы подробнее описать реализацию метода в виде библиотеки `minimap2_index_modifier`, какие компоненты были реализованы и как они связаны друг с другом.

3. Валидация разработанных алгоритмов и реализации на базе программного конвейера, равно как и демонстрация преимуществ по сравнению с существующими инструментами, проведена на эталонных данных. Степень общности этого результата для иных данных в тексте диссертации не полностью разъяснена.

Соискатель Карпулевич Евгений Андреевич согласился с замечаниями, ответил на задаваемые ему в ходе заседания вопросы.

На заседании 07 декабря 2023 диссертационный совет принял решение присудить Карпулевичу Е.А. ученую степень кандидата физико-математических наук за решение научной задачи построения программного конвейера для выравнивания последовательностей в приложениях биоинформатики, имеющей значение для развития математических методов в приложениях биоинформатики .

При проведении тайного голосования диссертационный совет в количестве 15 человек, из них 7 докторов наук по специальности рассматриваемой диссертации, участвовавших в заседании, из 22 человек, входящих в состав совета, проголосовали: за – 15, против – 0.

Заместитель председателя диссертационного совета,
доктор физико-математических наук

Петренко А. К.

Ученый секретарь диссертационного совета,
кандидат физико-математических наук

Зеленов С. В.

07 декабря 2023