

ЗАКЛЮЧЕНИЕ

диссертационного совета Д 002.077.04

на базе федерального государственного бюджетного учреждения науки
Института проблем передачи информации им. А.А. Харкевича
Российской академии наук по диссертации на соискание ученой степени
кандидата физико-математических наук

аттестационное дело № _____

решение диссертационного совета

от «11» февраля 2019 года, протокол № 3

О присуждении Гершгорину Роману Александровичу, гражданину
Российской Федерации, ученой степени кандидата физико-
математических наук.

Диссертация «Кратчайшее преобразование и реконструкция хромосомных структур» по специальности 03.01.09 – математическая биология, биоинформатика, принята к защите 6 декабря 2018, протокол № 6, диссертационным советом Д002.077.04 на базе Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук (127051, Москва, Большой Каретный переулок, д. 19, стр. 1, приказ о создании диссертационного совета №978/нк от «16» декабря 2013 года).

Соискатель Гершгорин Роман Александрович, гражданин Российской Федерации, 1991 года рождения, в 2013 году окончил механико-математический факультет Федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный университет имени М.В.Ломоносова». В период с 30.08.2013 по 29.08.2016 проходил обучение в аспирантуре Федерального государственного бюджетного учреждения науки

Института проблем передачи информации им. А.А. Харкевича Российской академии наук по специальности 03.01.09 – «Математическая биология, биоинформатика». В настоящее время работает исполняющим обязанности младшего научного сотрудника лаборатории №6 «Лаборатория математических методов и моделей в биоинформатике» Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук.

Диссертация выполнена в Федеральном государственном бюджетном учреждении науки Институте проблем передачи информации им. А.А. Харкевича Российской академии наук.

Научный руководитель – Любецкий Василий Александрович, доктор физико-математических наук, профессор, заведующий лабораторией №6 «Лаборатория математических методов и моделей в биоинформатике» Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук.

Официальные оппоненты:

1. Райгородский Андрей Михайлович, гражданин Российской Федерации, доктор физико-математических наук, профессор, главный научный сотрудник – заведующий лабораторией продвинутой комбинаторики и сетевых приложений Федерального государственного автономного образовательного учреждения высшего образования Московского физико-технического института (государственного университета);
2. Спирин Сергей Александрович, гражданин Российской Федерации, кандидат физико-математических наук, ведущий научный сотрудник Научно-исследовательского института физико-химической биологии им.

А.Н. Белозерского Московского государственного университета им. М.В. Ломоносова;

дали **положительные** отзывы на диссертацию.

Ведущая организация – Федеральное государственное бюджетное учреждение науки Институт общей генетики им. Н.И. Вавилова Российской академии наук, **в своем положительном заключении**, подписанным кандидатом биологических наук Касьяновым Артемом Сергеевичем, научным сотрудником лаборатории системной биологии и вычислительной генетики, и утвержденном доктором биологических наук, профессором Кудрявцевым Александром Михайловичем, директором Федерального государственного бюджетного учреждения науки Института общей генетики им. Н.И. Вавилова Российской академии наук, **указала, что** разработанные Гершгориним алгоритмы реконструкции могут быть использованы для изучения эволюции хромосомных структур клеточных органелл, геномной синтении и, как следствие, эволюции видов. Результаты его работы могут быть использованы в научно-исследовательских организациях, занимающихся исследованиями в областях молекулярной биологии и биофизики, таких как Институт молекулярной биологии им. В.А. Энгельгардта РАН, Институт теоретической и экспериментальной биофизики РАН, Институт биофизики клетки РАН, Институт цитологии и генетики СО РАН, Институт химической биологии и фундаментальной медицины СО РАН, Институт биофизики СО РАН, Московский государственный университет имени М.В. Ломоносова. Отзыв на диссертационную работу Гершгорина рассмотрен и утверждён на расширенном семинаре лаборатории системной биологии и вычислительной генетики Федерального государственного бюджетного учреждения науки Института общей генетики им. Н.И. Вавилова Российской академии наук 21 января 2019 года, протокол №4.

Выбор официальных оппонентов и ведущей организации обоснован тем, что один из оппонентов (А.М. Райгородский) является ведущим в нашей стране специалистом по теории графов и её приложениям, а другой (С.А. Спирин) – один из ведущих специалистов по биоинформатике, за последние годы ими опубликованы десятки работ в ведущих отечественных и международных журналах по этим направлениям. Ведущая организация – Институт общей генетики им. Н.И. Вавилова РАН является международно признанным центром научных исследований в области молекулярной биологии и её приложений.

Соискатель имеет 4 работы по теме диссертации, опубликованные в 3 научных журналах, включённых в перечень российских рецензируемых научных журналов и изданий для опубликования основных научных результатов диссертаций, общим объёмом 64 страницы; две из них входят в первый квартиль базы данных WoS. Соискателем опубликовано 2 тезисов российских и международных конференций.

Основные публикации по теме диссертации в журналах, включенных в перечень российских рецензируемых научных журналов и изданий для опубликования основных научных результатов диссертаций:

1. Lyubetsky V.A., **Gershgorin R.A.**, Seliverstov A.V., Gorbunov K.Yu. Algorithms for Reconstruction of Chromosomal Structures // *BMC Bioinformatics*. 2016, Vol. 17, № 40, 23 pages. DOI: 10.1186/s12859-016-0878-z.
2. Горбунов К.Ю., **Гершгорин Р.А.**, Любецкий В.А. Перестройка и реконструкция хромосомных структур // *Молекулярная биология*. 2015, Т. 49, № 3, С. 372–383. DOI: 10.7868/S0026898415030076. Перевод: Gorbunov K.Yu., Gershgorin R.A., Lyubetsky V.A. Rearrangement and Inference of Chromosome Structures // *Molecular Biology*. 2015, Vol. 49, № 3, P. 327–338. DOI: 10.1134/S0026893315030073.

3. **Gershgorin R.A.**, Gorbunov K.Yu., Zverkov O.A., Rubanov L.I., Seliverstov A.V., Lyubetsky V.A. Highly Conserved Elements and Chromosome Structure Evolution in Mitochondrial Genomes in Ciliates // *Life*. 2017, Vol. 7(1). DOI: 10.3390/life7010009.
4. Lyubetsky V.A., **Gershgorin R.A.**, Gorbunov K.Yu. Chromosome structures: reduction of certain problems with unequal gene content and gene paralogs to Integer linear programming // *BMC Bioinformatics*. 2017, Vol. 18, № 537, 18 pages.

В одной из этих работ соискатель является первым автором. В совместных работах соискателю принадлежат результаты по теоретической разработке и практическому применению методов и алгоритмов, представленных в диссертации, которые могут быть использованы для изучения эволюции хромосомных структур клеточных органелл, геномной синтении и эволюции видов, а также – по анализу полученных им результатов.

На автореферат диссертации поступило 5 отзывов, **все отзывы положительные**. В отзывах указывается, что работа имеет большое научное и практическое значение и по своей новизне и актуальности соответствует требованиям Высшей аттестационной комиссии. Диссертация является законченным научно-исследовательским трудом, выполненным автором самостоятельно на высоком научном уровне. В рамках диссертационного исследования соискателем разработаны новые алгоритмы кратчайшего преобразования и реконструкции хромосомных структур, которые были применены для исследования хромосомных структур митохондрий инфузорий и споровиков из класса Aconoidasida, пластид родофитной ветви у водорослей и споровиков, бактерий рода *Rhizobium*. Исследования являлись новыми на момент публикации, результаты, полученные автором, являются новыми научными знаниями

по специальности 03.01.09 – «математическая биология, биоинформатика».

В отзыве оппонента Райгородского Андрея Михайловича указаны замечания:

1. Автор не всегда чётко разграничивает свои результаты от упоминаемых им результатов других авторов. Например, во Введении текста диссертации автор приводит условие точности основного из его алгоритмов: «Получено эффективное решение задачи преобразования структур общего вида без паралогов; кроме того, оно является точным, если» (стр. 9). Хотя из автореферата, дальнейшего текста диссертации и из соответствующих ссылок ясно, что этот результат получен не автором, было бы лучше подчеркнуть это и во Введении.

2. Автор не приводит никаких данных об эффективности применяемого им пакета целочисленного линейного программирования (ЦЛП). Очевидно, он не разобрался в деталях реализованного в пакете алгоритма.

3. Во Введении автор называет свои алгоритмы эффективными, относя эту характеристику к сложности алгоритмов сведения к ЦЛП и финального алгоритма из главы 1, который по подсказке ЦЛП строит решение исходной графовой задачи. Конечно, о самой задаче ЦЛП и, тем более, о конкретном пакете, который диссертант использовал для её решения, нет аналогичных результатов. Действительно, в теории алгоритмов с оракулами характеристики сложности и точности всегда не касаются самого оракула, для которого они обычно не известны или тривиально экспоненциальные. Однако в контексте прикладной работы это можно было бы объяснить явно.

4. Описывая применения своих алгоритмов, автор не указывает время решения задач. Не приводит автор и сравнений результатов вычислений

одной и той же задачи в разных пакетах ЦПП, что позволило бы увидеть, насколько полученные решения близки между собой.

В отзыве оппонента Спирина Сергея Александровича указаны замечания:

1. Слишком краткий литературный обзор. Хотя упомянуто 14 работ, но каждой из них уделено по одному абзацу. По мнению рецензента, в диссертации на подобную тему следует излагать работы предшественников подробнее. Например, стоило кратко изложить алгоритм из статьи Comreau 2013 (номер 13 в списке использованных источников), а не ограничиваться лишь его упоминанием. Из текста диссертации непросто понять, чем каждый из описанных в ней пяти новых алгоритмов превосходит существовавшие ранее аналоги.

2. Две из восьми публикаций автора, приведённых как публикации по теме диссертации, из них одна — в реферируемом журнале, слабо связаны с темой диссертации. Впрочем, остальные четыре публикации в реферируемых журналах и двое тезисов, несомненно, связаны с темой работы и достаточно полно описывают её основные результаты.

3. Для статьи, имеющей номер 15 в списке использованных источников, указан неверный DOI (от другой статьи).

4. К сожалению, автор не провёл тестирования различных наборов весов элементарных операций, это можно было бы сделать на имеющихся примерах сравнением межгеномных расстояний, посчитанных по хромосомным перестройкам, с одной стороны, и по локальным заменам в отдельных генах, с другой. При этом в диссертации (стр. 39) приводятся конкретные наборы весов как «удобные» для линейных или же циклических хромосом, но не указывается, каким образом подбирались эти веса. Нелишним было бы и сравнение «брейкпойнтового» расстояния с расстоянием, определяемым как суммарная цена операций, например, по

тому, какое из них и насколько лучше коррелирует с числом локальных мутаций.

5. Общее замечание ко всему тексту — известная сумбурность изложения. Например, во введении сказано, что описанный в главе 1 алгоритм имеет линейную по времени и памяти сложность, но в самой главе 1 про сложность алгоритма не сказано вообще ничего. Определения прописаны не всегда достаточно чётко, например в (ключевом) определении общего графа двух структур на стр. 20 написано: «Ребра общего графа соединяют следующие пары вершин: ... Если линейный блок не имеет соседних общих генов (т.е. является цепью), то в общем графе ему соответствует изолированная особая вершина». В математическом тексте предпочтительна большая структурированность и чёткость. Ссылка на программную реализацию алгоритмов, которой следовало бы посвятить отдельный раздел введения, приводится «между делом» в четвёртой главе.

В отзыве ведущей организации отмечено, что в тексте имеется небольшое число опечаток и ошибок пунктуации, и указаны замечания:

1. Во Введении текста диссертации автор описывает свои алгоритмы словом «эффективные». Данную характеристику для приводимых алгоритмов использовать не совсем правомерно. Статистически значимое исследование их эффективности требует исследования эффективности используемого диссертантом пакета ЦЛП; такое исследование не проведено и вряд ли возможно. Обращение к «черному ящику» (в данном случае – ЦЛП) не позволяет сделать теоретический вывод о сложности рассматриваемой цепочки алгоритмов вида «сведение, ЦЛП, алгоритм из главы 1». Для биологических данных, на которых цепочка алгоритмов тестировалась, время счёта оказалось приемлемым, что, конечно, не позволяет сделать вывод о его теоретической эффективности на произвольных данных.

2. Во Введении (стр. 9) в тексте диссертации используется выражение «Получено эффективное решение задачи реконструкции» (и аналогично для других задач). Эта фраза неточна: что понимать под «решением задачи реконструкции»? Можно понимать цепочку алгоритмов, которая включает не только оригинальные алгоритмы сведения и финальный, но и обращение к «черному ящику», что не может быть эффективным для произвольных данных.

3. Задачи ЦЛП, которые нужно решить ради решения поставленных задач, имеют специальный вид по сравнению с произвольной задачей целочисленного линейного программирования. Для задач ЦЛП такого вида, по крайней мере, не противоречит известным теоретическим результатам предположение о решении их за хорошее полиномиальное время. Хотелось бы, чтобы диссертант исследовал возникающий таким образом новый подкласс задач ЦЛП.

В отзыве на автореферат Пентуса Мати Рейновича, доктора физико-математических наук, профессора, заместителя заведующего кафедрой Математической логики и Теории алгоритмов механико-математического факультета Федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный университет имени М.В. Ломоносова», отмечена максимальная общность рассматриваемых задач. В качестве замечания отмечено наличие опечаток и неудачных языковых оборотов, и предложено исследовать класс задач целочисленного линейного программирования, возникающий в результате сведения рассматриваемых графовых задачах.

В отзыве на автореферат Зарецкого Алексея Петровича, кандидата технических наук, старшего научного сотрудника лаборатории мультимедийных систем и технологий Федерального государственного автономного образовательного учреждения высшего образования

«Московский физико-технический институт (государственный университет)» приведены замечания:

1. Не очевидно, насколько эффективно применение целочисленного линейного программирования на реальных биологических данных;
2. Не очевидно, каким методом автор пользовался для верификации/валидации алгоритмов для задач преобразования и реконструкции с паралогами и равными ценами.

В отзыве на автореферат Салихова Камиля Маратовича, PhD по биоинформатике, University Paris-Est Marne-la-Vallee, руководителя группы Отдела разработки программного обеспечения ООО «Эйм Хай Технолоджис», отмечена эффективность реализации полученных алгоритмов. В качестве замечания указаны опечатки, недостаточная иллюстрированность при описании хромосомных структур и отсутствие сравнительного анализа с результатами предшествующих работ.

В отзыве на автореферат Гусак Юлии Валерьевны, кандидата физико-математических наук, младшего научного сотрудника лаборатории «Тензорные сети и глубинное обучение для интеллектуального анализа данных» автономной некоммерческой образовательной организации высшего образования «Сколковский институт науки и технологий», подчеркнута актуальность и новизна работы, в частности, при решении задачи реконструкции для структур с паралогами. Отмечены недостатки:

1. В главе 2 диссертант вводит новый термин «кратчайшее расстояние» для понятия «брейкпойнтовое расстояние», не очень ясна мотивация такого наименования.
2. В автореферате не упоминается, был ли приведен сравнительный анализ алгоритмов реконструкции для специального и кратчайшего расстояний.

В отзыве на автореферат Гараниной Ирины Андреевны, младшего научного сотрудника лаборатории протеомного анализа Федерального государственного бюджетного учреждения «Федеральный научно-клинический центр физико-химической медицины Федерального медико-биологического агентства», отмечена практическая применимость разработанных и реализованных алгоритмов для биоинформатики, в частности, для задач реконструкции предкового состояния и сборки полных геномов.

Отмечено замечание: из текста автореферата не понятен смысл введения специального расстояния и его отличие от кратчайшего. В описании раздела 3.4 упоминается ещё одно расстояние, «которое отличается от кратчайшего». Неясно, это то же самое специальное расстояние или другое. Имеется вопрос к соискателю: что автор подразумевает под «высококонсервативными элементами», по которым строились деревья пластид родофитной ветви.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

разработаны точные алгоритмы решения для важных задач преобразования и реконструкции хромосомных структур, позволяющие строить сценарии эволюции хромосомных структур;

разработан линейной сложности точный алгоритм для решения задачи преобразования структур без паралогов;

разработан кубической сложности алгоритм для решения задачи реконструкции с паралогами, равными ценами и произвольными структурами сведением к целочисленному линейному программированию кубического размера;

для специального расстояния, отсутствия паралогов и любых цен операций **разработан** квадратичной сложности точный алгоритм решения задачи реконструкции. В случае того же расстояния и присутствия

паралогов **разработан** квадратичной сложности точный алгоритм для решения задачи реконструкции сведением к задаче квадратичного булева линейного программирования;

разработаны алгоритмы для решения задачи преобразования с паралогом и равными ценами сведением к целочисленному линейному программированию (ЦЛП): для циклических хромосомных структур – к ЦЛП линейного размера и для произвольных структур – к ЦЛП квадратичного размера. Сложность алгоритмов сведения соответственно линейная и квадратичная;

точность алгоритмов везде указана относительно точности ЦЛП и указанных условий на цены; сложность – для алгоритмов сведения и финального.

Теоретическая значимость результатов обоснована тем, что:

полученные алгоритмы позволяют реконструировать эволюцию хромосомных структур с помощью алгоритмов эффективных прямых или сведения к целочисленному линейному программированию. Эти алгоритмы существенно обобщают известные до сих пор результаты и могут быть использованы для изучения эволюции хромосомных структур клеточных органелл, геномной синтении и эволюции видов.

Значение полученных соискателем результатов исследования для практики подтверждается тем, что предложен оригинальный подход к практическому решению NP-трудной задачи, который основан на целочисленном линейном программировании и алгоритмах сведения. Финальный алгоритм использует новые представляющие научный интерес понятия, связанных с преобразованием графов. В целом работа носит теоретический характер, хотя алгоритмы реконструкции хромосомных структур могут быть использованы при изучении эволюции хромосомных структур клеточных органелл, геномной синтении и эволюции видов.

Оценка достоверности результатов исследования выявила: результаты получили достаточное теоретическое обоснование и тестированы на искусственных и биологических данных.

Личный вклад соискателя состоит в создании и программной реализации алгоритмов, которые решают важную задачу реконструкции хромосомных структур на оригинальной основе, а также – в анализе полученных сценариев эволюции и подготовке основных публикаций по выполненной работе.

Диссертация охватывает основные вопросы поставленной научной проблемы и соответствует критерию внутреннего единства, что подтверждается наличием последовательного плана исследования, непротиворечивой методологической платформы, основной идейной линии, концептуальности и взаимосвязи выводов. По своему содержанию диссертация отвечает паспорту специальности 03.01.09 – математическая биология, биоинформатика.

Диссертационный совет пришёл к выводу о том, что диссертация Р.А. Гершгорина представляет собой завершённую научно-квалификационную работу, в которой решена задача разработки эффективных алгоритмов для задачи реконструкции, которая имеет важное теоретическое и практическое значение в области исследования эволюции хромосомных структур. По актуальности, новизне, практической значимости диссертация соответствует требованиям, установленным «Положением о порядке присуждения учёных степеней», утверждённым постановлением Правительства Российской Федерации от 24 сентября 2014 года №842 (в редакции постановления Правительства Российской Федерации от 21 апреля 2016 года № 335), предъявляемым к диссертациям на соискание учёной степени кандидата наук.

На заседании 11 февраля 2019 года диссертационный совет принял решение присудить Гершгорину Роману Александровичу учёную степень

кандидата физико-математических наук по специальности 03.01.09 –
Математическая биология, биоинформатика.

При проведении тайного голосования диссертационный совет в
количестве 17 человек, из них 5 докторов наук по специальности и
отрасли наук рассматриваемой диссертации, участвовавших в заседании,
из 21 человека, входящих в состав совета, проголосовали: за присуждение
учёной степени – 16, против присуждения учёной степени 0,
недействительных бюллетеней – 1.

Председатель

диссертационного совета Д 002.077.04

д.б.н., профессор



Гельфанд М.С.

Ученый секретарь

диссертационного совета Д 002.077.04

д.б.н., профессор



Рожкова Г.И.

11 февраля 2019 г.