

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 НАУК

аттестационное дело № __

решение диссертационного совета

от 14 декабря 2015, протокол № 13

о присуждении Солдатову Руслану Андреевичу, Российская Федерация, ученой степени кандидата физико-математических наук.

Диссертация «Методы предсказания структурных элементов РНК» по специальности 03.01.09 - Математическая биология, биоинформатика, принята к защите 2 октября 2015, протокол № 10, диссертационным советом Д 002.077.04 на базе Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук, (127051, г. Москва, Большой Каретный переулок, д. 19, стр. 1., приказ о создании №978/нк от 16 декабря 2013 года).

Соискатель Солдатов Руслан Андреевич, гражданин Российской Федерации 1989 года рождения, в 2011 году окончил с отличием Федеральное государственное образовательное учреждение высшего профессионального образования «Московский государственный университет им. М.В. Ломоносова», с 10.10.2011 по 09.10.2014 обучался в аспирантуре Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук, работает и.о. младшего научного сотрудника в Федеральном государственном бюджетном учреждении науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук.

Диссертация выполнена в учебно-научном центре «Биоинформатика» Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН).

Научный руководитель - доктор биологических наук Миронов Андрей Александрович, профессор факультета биоинженерии и биоинформатики Федерального государственного бюджетного образовательного учреждения высшего образования Московский государственный университет имени М.В.Ломоносова.

Официальные оппоненты:

1. Галзитская Оксана Валериановна, гражданка РФ, доктор физико-математических наук, руководитель группы биоинформатики

Федерального государственного бюджетного учреждения науки Института белка Российской академии наук;

- Кулаковский Иван Владимирович, гражданин РФ, кандидат физико-математических наук, старший научный сотрудник лаборатории вычислительных методов системной биологии Федерального государственного бюджетного учреждения науки Института молекулярной биологии имени В.А. Энгельгардта Российской академии наук;

дали положительные отзывы на диссертацию.

Ведущая организация - Федеральное государственное бюджетное учреждение науки Институт математических проблем биологии Российской академии наук, г. Пущино, в своем *положительном* заключении, подписанном кандидатом физико-математических наук, Яковлевым Виктором Вадимовичем, научным сотрудником лаборатории прикладной математики и утвержденном доктором физико-математических наук, профессором Лахно Виктором Дмитриевичем, директором ФГБУН Институт математических проблем биологии РАН указала, что диссертационная работа Солдатова Руслана Андреевича полностью соответствует критериям "Положения о порядке присуждения ученых степеней" (отзыв заслушан и одобрен на расширенном семинаре лаборатории прикладной математики ИМПБ РАН «23» ноября 2015 года, протокол № 8).

Соискатель имеет 4 опубликованных работы, из них 2 по теме диссертации в научных журналах, включенных в перечень российских рецензируемых научных журналов и изданий для опубликования основных научных результатов диссертаций, общим объёмом 14 страниц.

Наиболее значимые научные работы по теме диссертации:

- Солдатов РА, Миронов АА. Статистические методы сравнительно-геномного анализа, основанные на использовании диффузионных процессов // Биофизика. - 2013. - Т. 58. - С. 142–147.
- Soldatov RA, Vinogradova SV and Mironov AA. RNASurface: fast and accurate detection of locally optimal potentially structured RNA segments // Bioinformatics - 2014. - Vol. 30. - P. 457-463

Вклад диссертанта в опубликованные работы по теме диссертации состоит в непосредственном планировании исследований, участии в формулировках задач, теоретической разработке и практической реализации методов, сборе данных, обработке и анализе информации. В основных публикациях диссертант является первым автором.

На диссертацию и автореферат поступило 5 отзывов, *все положительные*. В отзывах указывается, что работа имеет высокий научный уровень, находится на переднем крае современной геномики, выводы детально обоснованы и убедительны, а полученные результаты открывают перспективу построения комплексных эффективных программных продуктов для анализа данных высокопроизводительного секвенирования, а в целом работа полностью соответствует требованиям Высшей аттестационной комиссии, предъявляемым к диссертациям на соискание ученой степени кандидата физико-математических наук.

В отзыве кандидата биологических наук Храмеевой Екатерины Евгеньевны, научного сотрудника Сколковского института науки и технологий, в качестве замечания отмечено, что в автореферате "содержится достаточно большое количество ошибок и опечаток, которые, однако, не затрудняют понимание изложенного материала", в том числе "1. В подписи в рисунке 1Б отсутствует описание верхней панели рисунка", "2. В тексте упоминается программа RNASlider, однако ни ссылки, ни описания этой программы нет", "3. На стр. 16 вместо ссылки на рисунок 6 написано "Ошибка! Источник ссылки не найден.", "4. Из текста не до конца понятно, что понимается под фоновым распределением свободных энергий. Нуклеотидные последовательности перемешивались с сохранением динуклеотидного состава? Или случайным образом из генома выбирались последовательности с похожим динуклеотидным составом?".

В отзыве кандидата медицинских наук, доцента Московцева Алексея Александровича, ведущего научного сотрудника отдела молекулярной и клеточной патофизиологии Федерального государственного бюджетного учреждения "Научно-исследовательский институт общей патологии и патофизиологии" Российской академии медицинских наук, отмечено, что автореферат "содержит некоторое количество опечаток, которые, впрочем, существенно не влияют на научное качество работы".

В отзыве кандидата биологических наук Базыкина Георгия Александровича, заведующего сектором молекулярной эволюции (сектор №4) Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук, "замечаний к тексту немного; большинство из них может быть списано на ограничения, накладываемые

форматом автореферата. На рис. 4 приведены числа, по-видимому являющиеся длинами геномных участков различных типов; однако не указано, являются эти длины константными, средними или характерными. В разделе "Распределение структурированных сегментов по регионам" не ясно, учитываются ли в анализе различия в фоновом и нуклеотидном составе между кодирующими и некодирующими областями, которые могут быть достаточно сильными. В описании сравнительно-геномного метода не описано, каким образом вычисляются параметры $a(z)$ и $b(z)$ ".

В отзыве кандидата физико-математических наук Спирина Сергея Александровича, старшего научного сотрудника Научно-исследовательского института физико-химической биологии имени А.Н. Белозерского Федерального государственного образовательного учреждения высшего профессионального образования "Московский государственный университет имени М.В. Ломоносова", "имеется несколько замечаний к изложению материала", среди которых: "...структуры РНК располагаются в специфических регионах геномы..."; "на странице 16 вместо двух ссылок на рис. 6 имеется дважды повторённый текст "Ошибка! источник ссылки не найден"; "Таблица 1 (стр. 13) составлена и описана весьма небрежно: в подписи к ней не указано, что означают числа в скобках и вне скобок, и только путём подсчётов удаётся выяснить, что числа вне скобок означают проценты (хотя знак процента опущен, а в подписи говорится о "доле"), а в скобках - количества; кроме того, вместо -3 в первой ячейке четвёртой строки стоит -2. На той же странице 13 в нижнем абзаце сказано, что при пороге $Z=-3$ предсказывается 30% структурных ДНК, а при пороге $Z = -2$ - 62%, далее говорится, что тРНК и рРНК "предсказываются плохо, в среднем 16%", но не сказано, при каком пороге, и тем самым непонятно, следует сравнивать эти 16% с 30% или же с 62%. На странице 19 в выражении для μ входят не определенные величины x_i (видимо, имелись в виду введённые несколькими абзацами ранее z_i)."

В отзыве кандидата физико-математических наук Первушина Дмитрия Давыдовича, сотрудника лаборатории вычислительной биологии РНК Центра Геномной Регуляции, Барселона, в качестве замечания отмечено, что "на мой персональный вкус, в автореферат можно было включить хотя бы один биологический пример локально-оптимальной структуры из *Bacillus subtilis*. Название диссертации является слишком общим и должно было быть конкретизировано с учётом темы работы", а также отмечена "некоторая небрежность в оформлении автореферата": "на стр. 16" отсутствуют ссылки на рисунки, на стр. 18 матрица A называется матрицей ковариации, в то время как она является матрицей, обратной к матрице ковариации, и др."

Выбор официальных оппонентов и ведущей организации обосновывается тем, что за последние годы ими было опубликовано большое количество научных работ, посвященных изучению механизмов регуляции клеточных процессов на уровне молекул РНК, алгоритмическим и статистическим методам анализа вторичной структуры РНК, а также сравнительно-геномным и эволюционным исследованиям нуклеотидных последовательностей.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

введено понятие локально-оптимального структурированного сегмента РНК с использованием Z-значений для определения границ такого рода сегментов
предложен метод эффективного вычисления Z-значений с учетом статистических характеристик последовательностей

разработана программа RNASurface для поиска локально-оптимальных сегментов, которая вычисляет профили структурированности и тепловую карту структурированности; диффузионная модель эволюции количественных характеристик последовательностей, которая позволяет выявить давление отбора на исследуемую характеристику;

показано, что теоретическое и практическое время работы и занимаемая память RNASurface не уступает самым эффективным программам данного класса; RNASurface имеет лучшее качество предсказания среди программ по предсказанию структурных элементов РНК на основе их энергетических свойств при апробации на полном геноме *Bacillus subtilis*; нетраслируемые области мРНК имеют значительно более высокую плотность предсказанных структурированных сегментов по сравнению с кодирующими областями в геноме *Bacillus subtilis*; диффузионная модель Z-значений позволяет значительно улучшить надежность предсказания аннотированных некодирующих РНК в геномах *Drosophila* по сравнению с Z-значениями в геноме *Drosophila melanogaster*.

Теоретическая значимость исследования обоснована тем, что:

предложены: новая постановка задачи о поиске структурированных сегментов без заранее заданной длины окна; новый подход по поиску локальных структурированных участков, реализованный в программе RNASurface; оригинальный метод оценки статистической значимости структурированности набора ортологичных последовательностей на основе диффузионных процессов.

Показано,

1. что переход от сканирования окном к обнаружению локально-оптимальных структурированных сегментов значительно улучшает качество предсказания известных РНК в геноме *Bacillus subtilis*.
2. высокая плотность структурированных сегментов в нетранслируемых областях матричных РНК по отношению к кодирующим областям в геноме *Bacillus subtilis*, что подтверждает гипотезу об активной регуляции экспрессии генов механизмами, основанными на вторичных структурах
3. преимущество сравнительно-геномного подхода по отношению к анализу структурированности в отдельных геномах.

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

разработан и внедрен новый метод поиска локально-оптимальных структурированных сегментов РНК, вычислительная эффективность которого позволяет осуществлять полногеномный анализ структурированных сегментов РНК, при этом метод может применяться для анализа потенциально важных структурированных участков РНК в геномах различных видов и транскриптах разных классов РНК.

Разработан новый метод сравнительно-геномного предсказания структурированных РНК на основе диффузионных процессов.

Представлены рекомендации для использования разработанных методов для анализа геномов и транскриптомов.

Оценка достоверности результатов исследования выявила:

– разработанные подходы опираются на современные алгоритмические, статистические и сравнительно-геномные биоинформатические методы и не вызывают сомнений.

– разработанные подходы были апробированы на геномах различных видов с использованием различных баз данных известных функциональных РНК.

– достоверность полученных результатов согласуются с представлениями о важности формирования вторичных структур РНК в нетранслируемых областях генома, и о наличие эволюционного отбора на структурные элементы известных функциональных РНК

Личный вклад соискателя состоит в непосредственном планировании исследований, участии в формулировках задач, теоретической разработке и практической реализации методов, сборе данных, обработке и анализе информации. Результаты получены автором самостоятельно.

Диссертация охватывает основные вопросы поставленной научной задачи и соответствует критерию внутреннего единства, что подтверждается наличием последовательного плана исследования, непротиворечивой методологической платформы, концептуальности и взаимосвязи выводов. По своему содержанию диссертация отвечает паспорту специальности 03.01.09 - математическая биология, биоинформатика. Описанные в диссертации алгоритмы, программы и результаты могут быть использованы в Федеральном государственном бюджетном учреждении науки Институт математических проблем биологии РАН, Федеральном государственном бюджетном учреждении науки Институт общей генетики им. Н.И. Вавилова РАН, Федеральном государственном бюджетном учреждении науки Институт молекулярной генетики РАН, Федеральном государственном бюджетном учреждении науки Институт биологии гена РАН, Федеральном государственном бюджетном учреждении науки Институт проблем передачи информации им. А.А. Харкевича РАН, Московском государственном университете имени М.В. Ломоносова, Новосибирском государственном университете и других учебных и научно-исследовательских организациях. Результаты, представленные в диссертации, могут быть использованы при чтении курсов по биоинформатике и структуре РНК, которые читаются в Московском государственном университете имени М.В. Ломоносова, Московском физико-техническом институте и других ВУЗах страны.

Диссертационный совет пришёл к выводу о том, что диссертация представляет собой завершённую научно-квалификационную работу, в которой решены задачи распознавания участков генома потенциально имеющих термодинамически стабильную вторичную структуру и оценки статистической значимости совместной структурированности набора исследуемых ортологичных последовательностей РНК, имеющих важное значение для аннотации структурированных участков РНК в геномных и транскрипционных данных. По актуальности, новизне, практической значимости диссертация соответствует требованиям, установленным «Положением о порядке присуждения ученых степеней», утвержденным постановлением Правительства Российской Федерации от 24 сентября 2014 года № 842, предъявляемым к диссертациям на соискание ученой степени кандидата физико-математических наук.

На заседании 14 декабря 2015 года диссертационный совет принял решение присудить Солдатову Руслану Андреевичу ученую степень кандидата физико-математических наук по специальности 03.01.09 - математическая биология, биоинформатика.

При проведении тайного голосования диссертационный совет в количестве 18 человек, из них 7 докторов наук по специальности рассматриваемой диссертации,

участвовавших в заседании, из 21 человека, входящих в состав совета, проголосовали за 18, против 0, недействительных бюллетеней 0.

Председатель
диссертационного совета Д 002.077.04
д.б.н.

Ученый секретарь
диссертационного совета Д 002.077.04
д.б.н., профессор

14 декабря 2015 г.

Мельничук
Гельфанд М.С.



Рожкова
Рожкова Г.И.