

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 биологических наук

аттестационное дело №
решение диссертационного совета
от 7 октября 2019, протокол № 19

о присуждении Казнадзей Анне Денисовне, гражданке Российской Федерации, учёной степени кандидата биологических наук.

Диссертация «Геномная ко-локализация генов углеводного метаболизма бактерий» по специальности 03.01.09 — Математическая биология, биоинформатика, принята к защите 22 июля 2019 года, протокол № 17, 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 диссертационного совета № 978/нк от 16 декабря 2013 года).

Соискатель Казнадзей Анна Денисовна, гражданка Российской Федерации, 1987 года рождения, в 2009 году окончила биологический факультет Московского государственного университета им. М. В. Ломоносова по специальности «Биохимия» В период подготовки диссертации являлась соискателем (с 2013 по 2016 год), а также работала в должности и.о. младшего научного сотрудника (с 2009 по 2010 год) и младшего научного сотрудника (с 2010 года по настоящее время) в ИППИ РАН.

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

Научный руководитель: Гельфанд Михаил Сергеевич, доктор биологических наук, профессор, заведующий учебно-научным центром «Биоинформатика», зам. директора по научной работе Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук.

Официальные оппоненты:

1. Озолин Ольга Николаевна, гражданка Российской Федерации, доктор биологических наук, профессор, зав. лабораторией функциональной геномики и клеточного стресса в Институте биофизики клетки Российской академии наук – обособленном подразделении Федерального государственного бюджетного учреждения науки «Федеральный исследовательский центр «Пущинский научный центр биологических исследований Российской академии наук».

2. Мошковский Сергей Александрович, гражданин Российской Федерации, доктор биологических наук, профессор, заведующий кафедрой биохимии медико-биологического факультета Федерального государственного бюджетного образовательного учреждения высшего образования "Российский национальный исследовательский медицинский университет имени Н.И. Пирогова" Министерства здравоохранения Российской Федерации

дали **положительные** отзывы на диссертацию.

Ведущая организация — Федеральное государственное бюджетное учреждение науки Институт общей генетики им. Н.И. Вавилова Российской академии наук дала **положительное** заключение, подписанное кандидатом биологических наук, старшим научным сотрудником, руководителем группы биоинформатики отдела вычислительной системной биологии Артамоновой Иреной Игоревной и утвержденное кандидатом биологических наук Брускиным С.А., заместителем директора по научной работе ИОГен РАН, в котором указала, что поставленная перед Анной Денисовной задача анализа организации генов

углеводного метаболизма у бактерий в контексте их эволюционной истории является, безусловно, актуальной. В диссертационной работе впервые описаны колокационные особенности более чем сотни тысяч генов углеводного метаболизма, а выявленные закономерности важны с точки зрения понимания общей организации и эволюции бактериальных хромосом, а также природы факторов, влияющих на взаимное расположение генов.

В работе также описан новый биоинформатический инструмент для сравнения нуклеотидных последовательностей, оптимизированный для пар достаточно удаленных гомологов. Ввиду огромного количества данных секвенирования, разработка инструментов, демонстрирующих преимущества в эффективности анализа такого рода данных весьма актуальна в современной вычислительной биологии. Инструмент успешно применяется в ходе анализа генов углеводного метаболизма. Также сообщается об успешном применении данного инструмента в опубликованных работах других исследовательских групп.

В заключительной части работы описан лабораторный эксперимент, проведенный для подтверждения гипотезы о функциях кассеты генов кишечной палочки, сделанный на основании анализа консервативных комбинаций генов углеводного метаболизма. Успешная проверка гипотез, сделанных в ходе биоинформатического анализа, с помощью лабораторных проектов, является важнейшим связующим звеном между теоретической и экспериментальной биологией и подтверждает актуальность соответствующих направлений.

Диссертационная работа хорошо оформлена. В ней достаточно мало опечаток и стилистических неточностей. Достоверность результатов никаких сомнений не вызывает, а выводы отражают суть полученных данных.

Уровень научных журналов и тезисов конференций, в которых результаты были опубликованы, также ее подтверждает. Результаты данной работы можно использовать во множестве типов биологических проектов, в частности, касающихся филогенетических исследований или функциональных предсказаний, а также в проектах, связанных с исследованиями генов углеводного метаболизма, например, в биотехнологических разработках. Результаты могут быть актуальны для научно-исследовательских организаций, занимающихся исследованиями в области сравнительной геномики, таких как Институт биологии гена РАН,

Институт Общей Генетики им. Н.И. Вавилова РАН, Институт молекулярной биологии им. В.А. Энгельгардта РАН, Сколковский институт науки и технологий, биологический факультет и факультет биоинформатики и биоинженерии Федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Московский государственный университет имени М. В. Ломоносова».

В отзыве содержится высокая оценка научного уровня и практической значимости выполненной работы и отмечено, что «работа удовлетворяет критериям пункта 9-14 «Положения о присуждении ученых степеней», утвержденного постановлением Правительства РФ от 24.09.2013 г. №842 в редакции с изменениями, утвержденными постановлением Правительства РФ от 28-го августа 2017 г. №1024, предъявляемым к диссертациям на соискание ученой степени кандидата наук, а ее автор заслуживает присуждения ученой степени кандидата биологических наук по специальности: 03.01.09 – математическая биология, биоинформатика».

Отзыв на диссертационную работу Анны Денисовны был заслушан, обсужден и одобрен на заседании научного семинара в отделе вычислительной системной биологии ИОГен РАН (Протокол № 3 от 21 мая 2019 г.).

Выбор официальных оппонентов и ведущей организации обоснован тем, что один из оппонентов (Озолинь О.Н.) является ведущим специалистом в области микробиологии, а другой — (Мошковский С.А.) является ведущим специалистом в области биохимии и обработки биоинформатических данных, за последние годы ими опубликованы десятки работ в ведущих международных журналах по этим темам. Ведущая организация - Институт Общей Генетики им. Н.И. Вавилова РАН является признанным центром научных исследований в области сравнительной геномики и эволюционной биологии.

Соискатель имеет 3 статьи по теме диссертации в международных рецензируемых журналах, включённых в перечень ВАК, общим объёмом 28 страниц. Кроме того, соискателем опубликовано 8 тезисов в материалах российских и международных конференций. Публикации по теме диссертации в

журналах, включенных в перечень российских рецензируемых научных журналов и изданий для опубликования основных научных результатов диссертации:

1. V. Novichkov, **A. Kaznadzey**, N. Alexandrova, D. Kaznadzey (2016) NSimScan: DNA comparison tool with increased speed, sensitivity and accuracy. *Bioinformatics* 32(15):2380-2381
2. **A. Kaznadzey**, P. Shelyakin, M. Gelfand (2017) Sugar Lego: gene composition of bacterial carbohydrate metabolism genomic loci. *Biology Direct* 12(1):28-42
3. **A. Kaznadzey**, P. Shelyakin, E. Belousova, A. Eremina, U. Shvyreva, D. Bykova, V. Emelianenko, A. Korosteleva, M. Tutukina, M. Gelfand (2018) The genes of the sulphoquinovose catabolism in *Escherichia coli* are also associated with a previously unknown pathway of lactose degradation. *Scientific Reports* 8(1):3177-3189.

В двух публикациях диссертант является первым автором. В рамках всех исследований вклад автора состоит в непосредственном планировании исследований, формулировке гипотез, теоретической разработке и практической реализации алгоритмов и методов, формулировании результатов и выводов, подготовке текста и иллюстративного материала, взаимодействии с научными журналами.

На автореферат диссертации поступило 5 отзывов, все отзывы положительные. В отзывах даётся высокая оценка работе и указывается, что диссертационная работа является законченным актуальным исследованием, которое вносит существенный вклад в понимание организации бактериальных геномов, в повышение эффективности методов сравнения нуклеотидных последовательностей и в соотнесение теоретических и практических знаний с помощью экспериментальных проверок гипотез, сделанных на стадии биоинформатического анализа. Работа выполнена автором самостоятельно, на высоком научном уровне с использованием современных биоинформатических методов и подходов. Выводы диссертации хорошо и четко сформулированы и соответствуют поставленным задачам. Объем проделанной работы, выдвинутые и проверенные гипотезы, разработанные методы и новые результаты не вызывают

сомнений в актуальности и ценности всей работы. Диссертация написана понятным языком и выдержана в едином стиле.

В отзыве ведущей организации указаны редакционные замечания, такие как ряд опечаток, неточности в подписях к рисункам и нечеткость графиков. Так, для иллюстрации эффективности инструмента NSimScan на графике приведено множество кривых, которые неудобно рассматривать, правильнее бы было сделать таблицу с числами, с выделенными местами, указывающими на преимущества инструмента по сравнению с его аналогами. Отмечена неточность пояснений к таблицам о склонности генов разных функциональных классов к образованию кассет, некоторые классы пропущены. При анализе вклада событий локальной дубликации в ко-локализацию генов сходных функций советуется помимо NSimScan использовать также надежно построенные филогенетические деревья для семейств гомологов.

В отзыве оппонента Сергея Александровича Мошковского отмечены следующие замечания:

1. Недостаточно ясно описана связь первого раздела результатов - создания метода NSimScan с последующими разделами, хотя она, безусловно, существует.
2. Успешно проверенная в работе гипотеза была сформулирована по итогам второй части результатов - углеводного "Лего" умозрительно, экспертным путем. Возникает вопрос создавался ли в процессе его сборки метод автоматического предсказания новых функций комбинаций генов?
3. Не вполне корректное описание чувствительности классификации на с. 19 литературного обзора. Чувствительность зависит от порога отсечения, поэтому рассматривать его без взаимозависимого понятия специфичности, упущенного в изложении, не имеет смысла.

В отзыве оппонента Озолинь Ольги Николаевны содержатся следующие замечания:

1. Есть некоторая неоднозначность в определении кассет. "Считалось, что гены формировали кассеты, если они располагались на хромосоме подряд, причем

расстояние между каждой парой не превышало 200 нуклеотидов". Но: "В кассете был разрешен один длинный интервал длиной 1500 нуклеотидов".

2. На Рис. 12, который отражает процент генов, формирующих кассеты у бактерий разных классов, надо было указать разбросы.

3. Некоторое удивление и поэтому желание получить комментарии вызывает отсутствие значимых ко-локализационных связей с генами категории "транскрипционные" (Рис. 13), т.к. известно, что гены локальных факторов транскрипции часто располагаются рядом с контролируемыми ими генами.

В отзыве на автореферат Григорян Ирины Аркадьевны, кандидата биологических наук, заведующей отделом микробиологии компании Hydromer, (Нью-Джерси, США) отмечена полнота проведенных исследований, а также высокий уровень текста самого и иллюстраций к нему, которые «будут понятны широкому кругу читателей». Отзыв не содержит замечаний, только пожелание сделать в будущем полный систематизированный каталог кассет генов для дальнейшей работы с ними другими исследователями.

В отзыве на автореферат Родионова Дмитрия Александровича, кандидата биологических наук, заведующего сектора №6 ИППИ РАН отмечено, что работа касается важных фундаментальных проблем современной геномики и системной биологии, оставляет очень хорошее впечатление и написана профессионально. Отмечен ряд редакторских замечаний, таких как отсутствие определения регулона и использование английских аббревиатур.

В отзыве на автореферат Калининой Ольги Вячеслововны, кандидата физико-математических наук, профессора Университета Земли Саар отмечено, что работа, по всей видимости, является самым масштабным исследованием подобного рода для генов углеводного метаболизма и выполнена на высоком методическом уровне. В качестве замечания указывается, что при сравнении работы инструмента NSimScan упомянуты не все опубликованные его аналоги для сравнения нуклеотидных последовательностей, например, не упомянут CD-HIT.

Кроме того, указан ряд опечаток и небрежностей в тексте, в том числе неправильно размеченная ось на рис. 9.

В отзыве на автореферат Цой Ольги Владиславовны, кандидата биологических наук, научного сотрудника Технического университета Мюнхена отмечено, что теоретическая и практическая значимость диссертационной работы не вызывает сомнений, особенным образом отмечена иллюстрация прикладной значимости предсказаний ранее неизвестных функций генов на основе анализа их консервативной ко-локализации. Приведены следующие вопросы: для определения кассет - сильно ли поменялись бы результаты без позволенного допуска в 1500 нуклеотидов? Можно ли при вычислении склонности к формированию кассет учитывать размер исследуемых групп? Насколько эволюционно далеко сохраняется ко-локализация генов одного функционального класса? Могут ли такие гены обладать разной специфичностью и регулироваться независимо? Обнаружен ли сайт связывания фактора YihW?

В отзыве на автореферат Мазина Павла Владимировича, кандидата биологических наук, научного сотрудника Сколковского технологического института сказано, что работа выполнена на высоком уровне, отмечено, что отдельный интерес представляет тенденция к ко-локализации у генов сходных функций. В отзыве приведены замечания по поводу неудобства интерпретации некоторых рисунков и опечаток, а также приведены вопросы: считалась ли кассета консервативной, если встречалась дважды, но в одном геноме? Насколько показанные распределения склонности генов к образованию кассет отличаются от случайных? Кроме того, отмечено, что из автореферата не совсем ясно, на чем основано утверждение о том, что значительная часть прокариотических генов не формирует устойчивых эволюционных комбинаций на хромосомах, а среди генов углеводного метаболизма эта доля еще меньше. Также приводится аргумент о том, что повышение экспрессии генов при росте культуры кишечной палочки на лактозе еще не доказывает наличие соответствующей активности кодируемых ими ферментов, а кроме того, подавление роста культуры на лактозе у мутанта по YihW не представляется очевидным из графика. Наконец, замечено, что в

автореферате не указано, как и где были найдены и описаны сайты связывания регулятора CRP.

На все вопросы и замечания диссертантом были даны исчерпывающие ответы.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

разработан и применен новый эффективный инструмент NSimScan для сравнения нуклеотидных последовательностей неблизкого родства;

описана геномная организация 148 тысяч генов углеводного метаболизма бактерий;

показано, что среди этих генов лишь около половины склонны к ко-локализации друг с другом; среди кассет, которые формирует другая половина, преобладают короткие кассеты;

показано, что функция и таксономическая принадлежность гена существенным образом влияет на его склонность к образованию кассет;

выявлены 45 эволюционно значимых ко-локационных связей между девятнадцатью функциональными классами генов углеводного метаболизма, из которых 12 связей оказались внутриклассовыми;

определен вклад событий локальной дупликации в ко-локализацию генов сходных функций, который оказался относительно небольшим;

подвергнуто успешной экспериментальной проверке предсказание функций генов кассеты *Escherichia coli*, которые кодируют, по всей видимости, мультифункциональные ферменты, вовлеченные в ранее неизвестный для кишечной палочки путь катаболизма лактозы.

Теоретическая значимость результатов обоснована тем, что: разработанный инструмент для анализа нуклеотидных последовательностей может быть успешно применен в рамках филогенетических исследований, а также для совершения функциональных и структурных предсказаний. Новые знания о ко-локационных тенденциях генов углеводного метаболизма вносят

существенный вклад в понимание организации бактериальных геномов и эволюционных механизмов, стоящих в ее основе в целом. Наконец, успешное подтверждение функциональных предсказаний с помощью надлежащих экспериментов является важным соотношением теоретических и практических знаний. Полученные результаты по экспрессии генов и регуляции *yih*-касеты кишечной палочки, помимо интереса с точки зрения впервые описанной альтернативы известному пути катаболизма лактозы, закодированного в опероне и Ф. Жакоба и Ж. Моно (лауреатов нобелевской премии, полученной за соответствующую работу), позволяют затронуть малоизученный, но крайне важный вопрос о мультифункциональных характеристиках бактериальных белков.

Практическая значимость результатов обоснована тем, что: полученные знания о геномной организации генов углеводного метаболизма могут быть использованы в биотехнологических проектах. Тщательное изучение механизмов, влияющих на ко-локализационные тенденции бактериальных генов позволяет улучшать методики получения промышленных штаммов. Инструмент для поиска нуклеотидных последовательностей, не уступающий аналогам по чувствительности, но отличающийся значительной скоростью работы и малой частотой ошибок может быть успешно применен для анализа данных секвенирования, количество которых с каждым днем стремительно увеличивается - в том числе, в медицинских и диагностических целях.

Свидетельства достоверности: результаты диссертации докладывались на российских и международных конференциях, а также опубликованы в ведущих рецензируемых международных журналах. Работа выполнена на большом объеме данных с использованием современных биоинформатических методов и подходов, применены корректные методы оценки статистических гипотез. Лабораторный эксперимент с кишечной палочкой подтвердил предсказания, сделанные с помощью анализа биоинформатических данных. Разработанный инструмент NSimScan для сравнения нуклеотидных последовательностей успешно применен в независимых опубликованных проектах, касающихся широкомасштабных исследований геномов бактерий, ретровирусов и других организмов.

Личный вклад соискателя состоит в планировании исследований, участии в постановке задач, теоретической разработке и практической реализации методов анализа данных и в написании статей. Результаты, приведенные в диссертации, получены автором самостоятельно.

По своему содержанию диссертация отвечает паспорту специальности «03.01.09 — математическая биология, биоинформатика».

Диссертационный совет пришел к выводу о том, что диссертация представляет собой завершённое научное исследование, в котором впервые описаны тенденции геномной ко-локализации аннотированных генов углеводного метаболизма бактерий в большой выборке геномов, разработан и успешно применен новый инструмент для сравнения нуклеотидных последовательностей, а также на основании исследования консервативных комбинаций функциональных классов выдвинута и экспериментально подтверждена гипотеза о неизвестных ранее функциях генов *Escherichia coli*. Выявлены как общие закономерности, вносящие существенный вклад в понимание эволюционных механизмов организации генов в геномах бактерий, так и конкретные новые свойства генов кишечной палочки, которая является одним из ключевых модельных организмов в биологических исследованиях. По актуальности, новизне и практической значимости диссертация соответствует требованиям установленным «Положением о порядке присуждения учёных степеней», утверждённым постановлением Правительства Российской Федерации от 24 сентября 2013 года № 842 (в редакции постановления Правительства Российской Федерации от 21 апреля 2016 г. № 335), предъявляемым к диссертациям на соискание ученой степени кандидата биологических наук.

На заседании 7 октября 2019 года диссертационный совет принял решение присудить Казнадзей Анне Денисовне учёную степень кандидата биологических наук по специальности 03.01.09 — математическая биология, биоинформатика.

При проведении тайного голосования диссертационный совет в количестве 21 человек, из них 9 докторов наук по специальности рассматриваемой

диссертации, участвовавших в заседании, из 21 человека, входящих в состава совета, проголосовали за - 14, против - 0, недействительных бюллетеней - 0.

Председатель

диссертационного совета Д 002.077.04

д.б.н., профессор



М.С. Гельфанд

Ученый секретарь

диссертационного совета Д 002.077.04

д.б.н., профессор

07 октября 2019 г.

Г.И. Рожкова