

"УТВЕРЖДАЮ"

Заместитель директора по научной работе
Федерального государственного бюджетного
учреждения науки
Институт общей генетики им. Н.И. Вавилова
Российской академии наук
к.б.н.



ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

на диссертационную работу

Казнадзей Анны Денисовны на тему **"Геномная ко-локализация генов углеводного метаболизма бактерий"**, представленную на соискание ученой степени кандидата биологических наук по специальности 03.01.09 - математическая биология, биоинформатика

Актуальность темы исследования. Бактерии – самые разнообразные организмы на нашей планете, в том числе, с точки зрения приспособленности к окружающей среде. В качестве источника энергии они способны использовать широкий спектр веществ, среди которых важнейшее место занимают углеводы. Кроме того, углеводы служат ключевым структурным элементом бактериальной клетки и участвуют во множестве внутренних ее процессов. Поэтому углеводный катаболизм и анаболизм относятся к одному из наиболее актуальных, с точки зрения научных исследований, сегментов бактериальной жизнедеятельности. Соответствующие проекты имеют как фундаментальное, так и прикладное значение, поскольку разработки, связанные с бактериями, широко применяются в сельском хозяйстве, биотехнологии и медицине. В диссертационной работе А.Д. Казнадзей впервые описаны ко-локационные особенности более чем сотни тысяч генов углеводного метаболизма. 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, появляющихся буквально каждый день, разработка инструментов, демонстрирующих преимущества в эффективности анализа такого рода данных, в том числе и с точки зрения скорости, весьма актуальна в современной вычислительной биологии. В рамках обсуждаемого исследования генов углеводного метаболизма инструмент успешно применяется для оценки вклада событий локальных дупликаций в ко-локацию генов со сходными функциями и принадлежащих к одним и тем же ортологическим группам. Автор также сообщает об успешном применении данного инструмента в опубликованных работах других исследовательских групп.

В заключительной части работы автор также описывает лабораторный эксперимент, проведенный для подтверждения предсказания функций генов в кассете кишечной палочки, сделанного на предыдущем этапе на основании анализа консервативных комбинаций генов углеводного метаболизма. Предсказанное участие генов кассеты, отвечающей за катаболизм серосодержащих сахаров, в лактозном метаболизме было успешно подтверждено, что, в свою очередь, позволило говорить о мультифункциональности соответствующих белков. Успешная проверка предсказаний функций белков, сделанных в ходе биоинформатического анализа, с помощью экспериментальных проектов, является важнейшим связующим звеном между теоретической и экспериментальной биологией и подтверждает актуальность соответствующих направлений.

Новизна исследования, степень обоснованности полученных результатов и выводов. Все три основные направления данной работы обладают высокой степенью новизны. Количество инструментов, используемых для анализа нуклеотидных последовательностей, достаточно велико, но постоянно увеличивающийся приток данных секвенирования требует все новых алгоритмов для оптимизации и ускорения соответствующих процессов. Автор показывает значительное усовершенствование обработки данных с помощью инструмента NSimScan с увеличением скорости, но без потери в чувствительности и без повышения частоты ошибок по сравнению с другими, широко используемыми инструментами в данной области.

Количество автоматически аннотированных генов всевозможных сегментов бактериального метаболизма также постоянно растет, однако существует не так много широкомасштабных исследований, касающихся их ко-локационных особенностей. С появлением все новых последовательностей ДНК в современных базах данных при этом растет потребность в их систематизации и анализе. В работе впервые описаны предпочтения более ста тысяч генов разных глобальных функций углеводного метаболизма к специфике своего геномного окружения, а также впервые описана значительная роль ко-локализации генов сходных функций.

Наконец, стоит отметить, что автоматическая аннотация порождает много ошибок, поэтому экспериментальная проверка теоретических гипотез такого рода исследований очень важна. Автор успешно предсказывает участие кассеты генов кишечной палочки *Escherichia coli* в лактозном метаболизме, впервые, таким образом, показывая альтернативный путь утилизации лактозы для данной бактерии, помимо известного пути, белки которого закодированы в лактозном опероне Жакоба и Моно.

Структура и объем диссертации. Структура диссертации стандартна, за исключением того, что методы не собраны в единую главу, а предваряют результаты в каждой главе. Это объясняется значительными различиями между типами исследований, описанными в каждой из этих глав, и, соответственно, использованными методами. Так, если вторая глава касается разработки биоинформатического алгоритма для поиска нуклеотидных последовательностей заданного сходства, то в четвертой подробно описано проведение лабораторного эксперимента по анализу экспрессии и регуляции генов кишечной палочки. Действительно, оказывается удобнее отдельно читать методы, непосредственно относящиеся к каждой из частей работы. Объем диссертации - 145 страниц, она состоит из 4 глав: "Литературный обзор", "Инструмент NSimScan для сравнения последовательностей ДНК удаленного сходства", "Организация генов углеводного метаболизма бактерий", и "Участие *uif*-кассеты

Escherichia coli в катаболизме лактозы". Диссертация содержит 21 рисунок и 3 таблицы. В конце приведен список литературы, содержащий 117 ссылок, и приложение, содержащее 4 таблицы. Текст написан хорошим языком, хоть и не лишен опечаток.

Объекты и методы исследования. Как уже отмечалось, методы исследования изложены не в отдельной главе, а приведены в начале каждой из основных глав (2-4) диссертационной работы. В главе "Инструмент NSimScan для сравнения последовательностей ДНК удаленного сходства" описаны биоинформатические методы сравнения эффективности работы инструментов для поиска нуклеотидных последовательностей, а также указаны контрольные выборки, на которых тестировался разработанный в работе инструмент: 53 рибосомальных семейства белков бактерий и метагеном корней огурца. Здесь также приведен подробный алгоритм работы самого NSimScan. В главе "Организация генов углеводного метаболизма бактерий" объектом исследования являлись 148 тысяч генов углеводного метаболизма, располагающиеся в бактериальных геномах 665 видов из разных семейств. В методах описаны способы создания двухуровневой классификации этих генов, анализа кассет, которые они составляют на хромосомах различных бактерий, а также статистического анализа, который проводился при поиске эволюционно значимых связей между разными функциональными классами. Наконец, в главе "Участие *yih*-кассеты *Escherichia coli* в катаболизме лактозы" описаны методы анализа экспрессии генов исследуемой кассеты кишечной палочки при росте культуры на разных средах, а также способы биоинформатического и экспериментального определения ее промоторов и сайтов связывания транскрипционных факторов.

Основные научные результаты и их значимость для науки и практики. Основные результаты изложены в главах 2-4, а также собраны в положениях, выносимых на защиту, выводах и заключении. Их новизна и фундаментальная значимость не вызывает сомнений. Практическая ценность их также очевидна. Она заключается, во-первых, в применении разработанного инструмента для поиска нуклеотидных последовательностей среднего сходства в широком спектре задач. Во-вторых, широкомасштабный анализ ко-локализационных особенностей генов углеводного метаболизма, с одной стороны, впервые позволил описать соответствующие тенденции у значительного количества генов разных функциональных классов и разных бактериальных таксонов, а с другой, позволил делать конкретные предположения о функциях конкретных генов бактерий, первое из которых удалось успешно подтвердить экспериментально уже в данной работе.

По теме диссертации опубликовано три статьи в рецензируемых международных научных журналах, входящих в основные библиометрические базы данных (PubMed, WoS и Scopus), все из которых входят в первую квартиль в своих областях.

Результаты работы также были доложены и опубликованы в сборниках тезисов восьми международных и российских конференций. Публикационная активность, таким образом, подтверждает научную значимость данной работы.

Достоверность полученных результатов. Степень обоснованности результатов и выводов. Конкретные рекомендации по использованию результатов и выводов. Никаких сомнения в достоверности полученных результатов нет. Уровень научных журналов и тезисов конференций, в которых они были опубликованы, также ее подтверждает. Результаты

данной работы могут быть использованы во множестве типов биологических проектов, в частности, фактически в любых проектах, касающихся филогенетических исследований или функциональных предсказаний (с помощью сравнения нуклеотидных последовательностей против баз данных разного размера), а также в проектах, связанных с исследованиями генов углеводного метаболизма и специфики их геномного окружения, которая важна, в частности, в ряде биотехнологических разработок.

Несмотря на общий высокий уровень выполнения и представления, к работе возник ряд замечаний, большинство из которых носит редакционный характер:

1) кроме ряда мелких опечаток, общее количество которых значительно меньше, чем в большинстве кандидатских диссертаций, стоит отметить ряд неудачных, неточных или даже неверных использованных выражений:

а) автор несколько раз использует форму «может позволять делать» (например, стр. 3 автореферата, второй абзац или стр. 6 текста диссертации, первый абзац): так как в большинстве случаев уже есть ряд соответствующих примеров, форма «позволяет делать» звучит лучше и точнее;

б) сравнение нуклеотидных последовательностей является не «одним из распространенных методов сравнительной геномики» (стр. 3 автореферата, третий абзац), а основой этого раздела науки;

в) методы сравнения нуклеотидных последовательностей необходимы для выявления не причин, а скорее механизмов ко-локализации генов (стр. 4 автореферата, первая строка)

г) в нескольких местах в тексте подчеркивается, что инструмент NSimScan «наилучшим образом подходит для поиска последовательностей, различающихся на 60-90%» - вероятно, для поиска последовательностей, совпадающих на 60-90%; и т. п.

2) Автор выбрал, по-видимому, не самую удачную форму для иллюстрации основных результатов в главах 2 и 3: на большинстве приведенных графиков представлено не менее дюжины однотипных кривых, что делает их практически нечитаемыми, особенно в формате автореферата. Например, для иллюстрации эффективности NSimScan приведен график зависимости логарифма частоты ошибок от чувствительности для семи программ, причем целевая программа запускалась с разными параметрами (рис. 5) - всего 12 кривых, которые на отдельных участках графика практически сливаются. Вероятно, эффективнее было бы проиллюстрировать параметры запуска целевой программы и аналогичных общеупотребительных программ, с которыми она сравнивается, на тестовых множествах с помощью таблицы, посвятив каждому прогону одной из программ строку и приведя все соответствующие параметры по столбцам, выделив жирным группу наилучших значений для каждого параметра.

3) В третьей главе диссертации в таблицах и на графиках или в их описаниях допущен ряд неточностей. Например, как функциональные классы, обладающие наименьшей склонностью к образованию кассет, упоминаются в тексте нуклеозидазы (23,28%), фосфотазы (37,77%) и мутаза (40,35%), но ничего не сказано про транскрипционные факторы, склонность к образованию кассет для которых составляет 35,29%, как следует из таблицы 2. Аналогично, среди наибольших по склонности к образованию кассет не упомянуты транспортные белки, имеющие в таблице третью по величине склонность, а

вместо них упомянуты трансальдолазы/транскетолазы с несколько меньшим значением склонности. В описании рисунка 9 автореферата (рис.10 в тексте диссертации) упомянуты кластеры COG, склонность для которых составляла более 90% и даже 100%, но на самом графике точка с самой высокой склонностью соответствует примерно 75-76%.

- 4) В конце третьей главы диссертации при прояснении вопроса, появляются ли гены одного и того же функционального класса, входящие в состав кассеты, в результате локальной дубликации, автор, судя по описанию метода, использует вес, определяемый программой NSimScan, для поиска ближайших эволюционных соседей. Это недостаточно корректно. Единственное свойство, прямо указывающее на ближайшее эволюционное родство, - это соседство на надежно построенном филогенетическом дереве для семейства гомологов.
- 5) Недостаточно аккуратны подписи к некоторым рисункам или на рисунках. Например, в автореферате присутствует два рисунка 12, в подписи к первому из них (рис. 15 в тексте диссертации) допущена ошибка, а на рисунке 13 автореферата (рисунок 20 в тексте диссертации) в пунктах а) и б) не подписаны вертикальные оси.

Однако все перечисленные замечания не подвергают сомнению достоверность полученных результатов и обоснованность сделанных в работе выводов и не снижают научную значимость работы.

Заключение

Диссертационная работа Казнадзей Анны Денисовны на тему "Геномная ко-локализация генов углеводного метаболизма бактерий", представленная на соискание ученой степени кандидата биологических наук, является законченной научно-квалификационной работой. По своей научной новизне и практической значимости работа соответствует указанной специальности. 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 и автореферат Казнадзей Анны Денисовны был заслушан, обсужден и одобрен на заседании научного семинара в отделе вычислительной системной биологии ИОГен РАН. Присутствовало на заседании 11 человек. Результаты голосования: «за» - 11 человек, «против» - нет, воздержались - нет (протокол заседания №3 от 21 мая 2019 г.). Основное направление научно-исследовательской работы отдела соответствует тематике диссертационной работы.

Отзыв подготовлен:

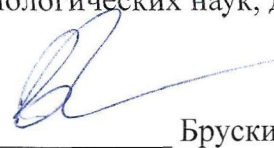
Руководитель группы биоинформатики
отдела вычислительной системной биологии,
кандидат биологических наук,
старший научный сотрудник ИОГен РАН

 Артамонова Ирена Игоревна

Подпись И.И. Артамоновой заверяю:

Заместитель директора по научной работе Федерального
государственного бюджетного учреждения науки
Институт общей генетики им. Н.И. Вавилова РАН,
Кандидат биологических наук, доцент



 Брускин С.А.

Адрес организации:

Федеральное государственное бюджетное учреждение науки Институт общей генетики им. Н.И. Вавилова Российской академии наук

119991, ГСП-1 Москва, ул. Губкина, д.3

Тел.: (499) 135-62-13

Email: iogen@vigg.ru

Электронный адрес и телефон составителя отзыва:

Артамонова И.И. irenart@vigg.ru, тел. +7 (916) 915 5809

19 сентября 2019 г.