

## ОТЗЫВ

Официального оппонента 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

Диссертационная работа Казнадзей А.Д. посвящена исследованию ко-локализации генов, кодирующих белки метаболизма углеводов у бактерий, в том числе, разработке нового алгоритма, позволяющего быстро и аккуратно сравнивать нуклеотидные последовательности среднего сходства.

Метаболизм - катаболизм и анаболизм - углеводов у бактерий является одним из наиболее актуальных для исследования сегментов бактериальной жизнедеятельности, поскольку с одной стороны, бактерии, являясь самыми распространенными и разнообразными организмами в мире, умеют очень быстро перестраивать свой метаболизм, приспосабливаясь к условиям среды, и исследование механизмов, лежащих в основе такого переключения, очень интересно с точки зрения фундаментальной науки. С другой стороны, полное понимание превращений углеводов, служащих ключевым структурным элементом бактериальной клетки и участвующих в большинстве внутренних её процессов - в особенности путей утилизации необычных субстратов - необходимо для успешного применения бактерий в сельском хозяйстве, биотехнологии и медицине. В диссертационной работе Казнадзей А.Д. впервые проведен масштабный анализ ко-локационных тенденций генов, кодирующих ферменты, транспортёры и регуляторы углеводного метаболизма: описана сеть эволюционных связей 148 тысяч генов для 665 видов бактерий и установлено, что 53% генов, кодирующих белки углеводного метаболизма, находятся в составе кассет. Оказалось, что склонность к формированию кассет у разных генов различается, а ключевыми факторами, влияющими на ко-локационные тенденции, являются функциональные и структурные характеристики гена и таксономическая принадлежность. Выявленные закономерности, безусловно, важны для понимания эволюции бактериальных хромосом в той части, которая определяет взаимное расположение генов на них.

Самостоятельную практическую значимость имеет разработанный в ходе выполнения работы алгоритм сравнения удаленных друг от друга (60-90% гомологии) нуклеотидных последовательностей NSimScan. В настоящее время появляется все больше данных полногеномного секвенирования, и разработка инструментов, позволяющих увеличить эффективность и скорость анализа больших массивов таких данных, а также снизить частоту ошибок, несомненно, актуальна. В рамках данной диссертационной работы NSimScan был успешно применён для поиска и масштабной характеристики генных кассет. Среди генов, кодирующих ферменты, транспортёры и регуляторы углеводного метаболизма автором, в частности, была обнаружена кассета генов утилизации сульфоквиновозы (сульфоглюкозы) кишечной палочки, которая по

составу очень похожа на кассету генов утилизации лактозы у бацилл. В последней части диссертации приведён объёмный экспериментальный материал, свидетельствующий об активации генов кассеты сульфоглюкозы при росте кишечной палочки на лактозе и о том, что эта активация опосредована закодированным в этой же кассете локальным регулятором YihW. Это наблюдение, безусловно, требует биохимической проверки способности соответствующих ферментов утилизировать галактозу, однако это очень интересно и может говорить о наличии у энтеробактерий альтернативных шунтов утилизации лактозы, которые могут помогать клетке метаболизировать этот субстрат, если основной *lac*-оперон по каким-то причинам не работает. Актуальность и практическая значимость диссертационной работы подтверждается также тем, алгоритм NSimScan был успешно использован и другими исследовательскими группами.

**Структура и объем диссертации.** Диссертационная работа Казнадзей А.Д. изложена на 145 страницах машинописного текста и состоит из 4 глав: "Литературный обзор", "Инструмент NSimScan для сравнения последовательностей ДНК удаленного сходства", "Организация генов углеводного метаболизма бактерий", и "Участие *yih*-кассеты *Escherichia coli* в катаболизме лактозы". Таким образом, результаты диссертационной работы описаны в трёх главах, и в каждой из них содержатся разделы «Материалы и методы», «Результаты и обсуждение» и «Заключение». Такая структура диссертации разумна, поскольку, несмотря на их очевидную логическую взаимосвязь – от разработки алгоритма к его практическому применению - типы исследований, описанные в каждой из глав, существенно различаются. Работа содержит 21 рисунок и 3 таблицы. В конце приведен список литературы, содержащий 117 ссылок, и приложение, содержащее 4 таблицы.

**Обзор литературы** изложен на 23 страницах и включает в себя три основных раздела, в которых описаны методы сравнения нуклеотидных последовательностей, и проведен сравнительный анализ основных инструментов, закономерности локализации генов, кодирующих белки метаболизма углеводов у бактерий, а также механизмы переключения метаболических путей при смене источника углерода. Большая часть последнего раздела посвящена способам утилизации лактозы у энтеробактерий и бацилл, а также возможным функциям *yih*-кассеты генов кишечной палочки, исследуемой в данной работе. В целом, обзор литературы отражает общую картину имеющихся к настоящему времени знаний.

**Вторая глава** посвящена разработке алгоритма для сравнения нуклеотидных последовательностей NSimScan. Она включает в себя описание области его применения, подробное описание работы самого алгоритма, а также методы оценки эффективности. Важно отметить, что NSimScan выложен в сеть для скачивания (<https://github.com/abadona/qsimsca>), как и подробное руководство по его использованию и примеры параметров для командной строки. Для оценки эффективности работы

разработанного инструмента были использованы гены рибосомных белков из 1244 бактериальных геномов разных видов. Из каждого из 53 семейств генов, содержащего более 600 аннотированных представителей, для перекрёстного сравнения случайным образом было выбрано по 200. На основании правильности/ошибочности полученных данных была вычислена чувствительность и селективность алгоритма. Эвристичность NSimScan была также сопоставлена с шестью другими поисковыми программами широкого пользования на примере решения задачи обнаружения репрезентативных последовательностей 16S РНК 749 таксонов в метагеноме корней огурца. Оказалось, что по уровню ошибок и чувствительности он работает на уровне самого чувствительного алгоритма SSearch, но при этом существенно превосходит его по скорости. И хотя при поиске последовательностей с уровнем сходства менее 70%, чувствительность NSimScan падает, при анализе больших массивов метагеномных данных он работает в 10 раз быстрее, чем MegaBLAST, обнаруживая более 95% таксонов, найденных MegaBLAST и 21,6% таксонов, которые MegaBLAST не находит. Имея чувствительность и селективность, сопоставимые с SSearch, NSimScan имеет большие преимущества при анализе последовательностей невысокого сходства (60-90%). Всё это совершенно адекватно отражено в первом выводе диссертации. Нет никаких сомнений в том, что NSimScan будет использован в самых разных проектах, требующих быстрого поиска гомологичных последовательностей. В данной работе он был использован для характеристики дублицированных генов углеводного метаболизма в бактериальных геномах и поиска их ортологов.

**Третья глава** диссертационной работы «Организация генов углеводного метаболизма бактерий» посвящена изучению кластеризации (ко-локализации) генов углеводного метаболизма бактерий. Она состоит из трех стандартных разделов «Материалы и методы», «Результаты и их обсуждение» и «Заключение».

Автором установлено, что ко-локализация не является строгой характеристикой генов, кодирующих белки с взаимосвязанными функциями. При этом в состав общих кассет часто входят два или несколько генов, кодирующих белки, принадлежащие к одному и тому же функциональному классу. Из 19 изученных функциональных классов почти две трети продемонстрировали склонность к подобной ко-локализации. Гены одного и того же класса, ко-локализованные в кассетах, делились на две группы – гены, кодирующие разные субъединицы белковых комплексов, и гены, кодирующие отдельные белки. Наиболее распространенным примером участников первой группы ожидаемо оказались гены, кодирующие субъединицы транспортных белков, при этом устойчивые ко-локационные связи формировали кассеты, в которых находилось не менее трёх генов, а события ко-локализации не более двух генов транспортеров не превышали порог случайности. Это наблюдение согласуется с субъединичным устройством ABC-транспортёров, которые, как правило, состоят не менее, чем из трёх субъединиц. Вторую

группу, в основном, составляли гены, белковые продукты которых участвуют в последовательных реакциях того или иного метаболического пути. Примерно в половине случаев белки, принадлежащие к одному функциональному классу и закодированные в ко-локализованных генах, принадлежали к одному кластеру COG и обладали значительной структурной гомологией. Поэтому у автора возникла гипотеза, что они могли появиться в результате локальной дупликации. Однако только в 3,6% случаев гены в паре продемонстрировали наибольшее сходство друг с другом, чем с ортологами в других геномах. Более того, оказалось, что менее 10% пар ко-локализованных генов из одного COG больше похожи на один и тот же ген в другом геноме, чем друг на друга, а в большинстве случаев они похожи на два разных гена. Автором предложено несколько вполне правдоподобных объяснений этому явлению, но на основании сделанных наблюдений сделан вывод о том, что ко-локализация генов, кодирующих белки с похожими функциями, в преобладающем большинстве случаев не является результатом событий локальной дупликации. Результаты этого раздела суммированы в виде второго, третьего и четвертого выводов.

На основании данных, полученных при анализе ко-локализаций/кластеризаций, было сделано предположение о том, что консервативные сочетания генов определённых функциональных классов могут указывать на сходные функции соответствующих кассет. Чтобы проверить это предположение, автором была выбрана консервативная кассета *yih*-генов энтеробактерий, белковые продукты которой, согласно относительно недавнему биохимическому исследованию, участвуют в процессе сульфогликолиза. По функциональному составу генов она была похожа на консервативную кассету генов утилизации лактозы у бацилл.

Экспериментальной проверке возможности участия кассеты *yih*-генов в метаболизме лактозы у кишечной палочки посвящена **четвёртая глава**. На первом этапе экспериментальной работы, были картированы стартовые точки транскрипции *yih*-генов, показано, что *yih*-кассета не является опероном (большинство генов может транскрибироваться с собственных промоторов) и установлено, что один из промоторов гена *yihT*, *yihTP1*, включается только при росте культуры на лактозе. Можно было более точно определить места инициации синтеза мРНК с помощью однократной инициации транскрипции *in vitro*, но для данной работы, в принципе, было достаточно и имеющихся данных 5'-конец-специфического секвенирования РНК и обратной транскрипции. Из данных обратной транскрипции следует, что синтез РНК с гена *yihV* активируется при росте на лактозе, что было подтверждено с помощью ОТ-ПЦР в реальном времени. Экспрессия *yihS* (изомераза), *yihT* (альдолаза) и *yihV* (киназа) в присутствии лактозы усиливалась, тогда как уровень *yihW*-мРНК снижался. Было сделано предположение, что *YihW* может являться репрессором для исследуемых генов, что было подтверждено с использованием штаммов, мутантных по кодирующему его гену. Было обнаружено, что,

как и в большинстве других кассет углеводного метаболизма кишечной палочки, локальный регулятор YihW работает координированно с глобальным регулятором cAMP-CRP, который эффективно связывался с регулярными областями всех исследуемых генов. Эти данные позволили автору сделать завершающий вывод, в котором содержится не только констатация факта участия *yih*-кассеты генов *Escherichia coli* в катаболизме лактозы, но и акцентируется обнаружение нового пути утилизации лактозы у кишечной палочки.

По результатам диссертационной работы опубликовано 11 научных работ - 8 тезисов международных и российских конференций и 3 статьи в ведущих научных журналах, все из которых включены в Перечень ВАК и относятся к Q1.

#### **Замечания:**

1. Есть некоторая неоднозначность в определении кассет (стр 51): «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. Автор уделил много внимания корректному определению статистической достоверности ко-локационных событий, что, действительно, очень важно для данного исследования. Тем более странно, что это не отражено на представленных графиках. В каких-то случаях это не существенно, но на Рис. 12, который отражает процент генов, формирующих кассеты у бактерий разных классов, разбросы надо было указать.
3. Некоторое удивление и поэтому желание получить комментарии вызывает отсутствие значимых ко-локаций с генами категории «транскрипционные» (Рис. 13, стр. 66), т.к. известно, что гены локальных факторов транскрипции часто располагаются рядом с контролируруемыми ими генами.

Указанные замечания не снижают прекрасного впечатления от работы Анны Денисовны, которая является продуманным и законченным в рамках поставленных задач исследованием. Диссертация хорошо написана и оформлена. Она содержит минимальное количество технических (отсутствие курсива на видовых названиях), орфографических (стр. 51, 80, 88) и стилистических (стр. 31, 47, 72, 91) погрешностей. Достоверность результатов не вызывает никаких сомнений, а выводы обоснованы и убедительны. Автореферат соответствует содержанию диссертации, а в 11 публикациях автора отражены все основные результаты.

По актуальности темы, объему проведенных исследований, практической значимости и новизне полученных результатов диссертационная работа Казнадзей Анны Денисовны на тему "Геномная ко-локализация генов углеводного метаболизма бактерий",

несомненно, соответствует 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.

20.09.19

О.Н. Озолин

Доктор биологических наук, профессор,  
заведующий лабораторией  
Функциональной геномики и клеточного стресса,  
Институт биофизики клетки Российской академии наук -  
обособленное подразделение Федерального государственного  
бюджетного учреждения науки  
«Федеральный исследовательский центр  
«Пущинский научный центр биологических исследований  
Российской академии наук»»  
142290, г. Пущино Московской области, Институтская, 3, ИБК РАН  
Тел.: +74967 739 140  
E-mail для связи: ozoline@rambler.ru

Подпись Озолин О.Н. Николаевич удостоверяю  
Учредитель секретари ИБК РАН  
(Шабунин К.С.)

