

**Отзыв на автореферат диссертации А.Д. Казнадзей «Геномная ко-локализация генов углеводного метаболизма бактерий» представленно к защите на соискание ученой степени кандидата биологических наук по специальности 03.01.09 «математическая биология, биоинформатика»**

Диссертация Анны Денисовны Казнадзей посвящена анализу взаимного расположения генов углеводного метаболизма на хромосомах бактерий. В работе сперва описывается разработанная при участии Анны новая программа NSimScan предназначенная для поиска гомологичных последовательностей — что было необходимо для дальнейшей работы. Автор проводит сравнение NsimScan с аналогичными программами, такими как blast и ssearch и утверждает, что NsimScan по ряду параметров превосходит их. Далее Анна определяет кассету генов углеводного метаболизма как группу генов относящихся к соответствующим функциональным классам и расположенных рядом на бактериальной хромосоме. Автор проводит статистический анализ состава таких кассет, зависимости его от типа бактерий, показывает, что склонность образовывать кассеты зависит от функционального класса к которому относится ген. На основании ко-локализации автор устанавливает связи между функциональными классами. Интересно, некоторые функциональные классы предпочитают иметь более одного представителя в кассете. В последней главе, на основании сходства кассеты *u1h* между представителями класса *Bacilli* (у которых она, как уже было известно ранее, участвует в катаболизме лактозы) и семейства *Enterobacteriaceae* (где было известно, что данная кассета участвует в метаболизме серосодержащих углеводных соединений) делается предсказание, что у *Enterobacteriaceae* данная кассета так же участвует в катаболизме лактозы. Проведённые эксперименты показывают, что экспрессия генов *u1h* кассеты кодирующих ферменты действительно увеличивается при росте *Escherichia coli* на лактозе, дополнительные эксперименты позволили автору установить факторы транскрипции и регуляторные связи ответственные за данные изменения экспрессии.

Хотя в целом работа выполнена на хорошем уровне, у меня возникло к ней несколько вопросов:

1. стр 15 «кассета считалась консервативной, если встречалась в исследованных геномах по крайней мере дважды» - даже если она встречалась несколько раз, но только в одном геноме?
2. Стр 15 «Такие наблюдения подтверждают гипотезу о том, что значительная часть прокариотических генов не формирует очевидных эволюционно- устойчивых комбинаций на бактериальных хромосомах, причем оказывается, что внутри сегмента углеводного метаболизма эта доля еще больше.» - не совсем ясно, на чем основано это утверждение, я не нашел в автореферате сравнения склонности различных генов углеводного метаболизма с другими функциональными группами образовывать «эволюционно- устойчивые комбинации»
3. Автор анализирует и обсуждает распределение кассет по размеру в зависимости от функций генов и видов бактерий, однако не проводит ни каких статистических тестов, чтобы показать, что найденные зависимости не случайны. Например рис. 6-7-8 — насколько показанные распределения случайны? Как выглядит аналогичное распределение для случайной выборки генов такого же размера? Автор указывает, что «Наибольшей склонностью к образованию кассет обладали представители классов *Dictyoglomi* и *Fusobacteria* (76%), *Thermotogae* (72%) и *Bacilli* (65%). Это соответствует опубликованным данным о том, что гены представителей этих классов (например, рода *Streptococcus*) часто лежат в составе длинных оперонов [Varghese Dehal et al, 2010; Gama-Castro et al, 2016]. Наименьшей склонностью к образованию кассет обладали представители классов *Planctomycetia* (37%) и *Chlamydiae* (37%) и *Chlorobia* (40%).» Однако не ясно, насколько указанные отличия

статистически значимы, может быть они объясняются просто размерами геномов и количеством генов метаболизма углеводов у этих бактерий.

4. В выводах указано, что *uif*-кассета участвует в катаболизме лактозы у *Escherichia coli*. Однако реально в автореферате показано, только что часть генов этой кассеты активируется при росте на лактозе. Однако это не доказывает ни того, что эти гены необходимы для катаболизма лактозы ни даже то, что они в нем участвуют. Для доказательства необходимо прямо показать, что данные ферменты катализируют реакции с участием лактозы, а так же, что нокаут хотя бы одного из этих ферментов приводит к замедлению (или прекращению) роста бактерий на лактозе.
5. Стр 24 «Наконец, общий рост культуры *E. coli* M182, мутантной по *uifW*, существенно снижен на лактозе относительно роста на глюкозе (Рис 13, с).» - насколько я могу видеть, все три культуры растут на лактозе существенно хуже, чем на глюкозе, никакой специфики для нокаута по *uifW* я не вижу
6. Так как CRP является глобальным регулятором и его нокаут приводит к заметному падению скорости роста клеток на обоих субстратах, кажется вероятным, что часть эффектов CRP на *uif*-кассету может быть непрямыми. Хотя в автореферате указано, что «с помощью метода филогенетического футпринта были предсказаны, а затем экспериментально подтверждены потенциальные сайты связывания CRP» однако я не нашел в тексте ни описания экспериментов доказывающих связывание CRP в соответствующих сайтах ни указания того, в промотерах каких именно генов обнаружены такие сайты.

Кроме того, в тексте попадает существенное количество ошибок или жаргонизмов, рисунки не всегда хорошо оформлены, что затрудняет чтение текста, например:

стр 5 «наилучшим образом он подходит для поиска последовательностей, различающихся на 60-90%» - вероятно не различающихся а сходных?

стр 5 «Функциональные классы могут формировать консервативные и, по всей видимости, эволюционно значимые ко-локализационные связи.» - если что-то консервативно, значит оно сохраняется эволюцией и значит оно зачем-то нужно, без всякой видимости

стр 8 рис 3 «Белым отмечены гены, кодирующие функции, не представленные в другой кассете.» - гены кодируют белки (или РНК), но не «функции»

стр 10 рис 5 – на графике указано, что по оси у отложен логарифм частоты ошибок, у некоторых методов эта величина превосходит  $10^1$ , значит ли это, что частота ошибок в этих случаях оказывается более  $e^{10}$ ? Кроме того рисунок 4 крайне трудно читается – подписи осей (значения) и размер символов на линиях очень мелкие. Чтобы понять с какими параметрами запускался NsimScan надо расшифровывать запись типа «sh3.k12.kt290.lthr45.sthr90». Было бы лучше показать линии для разных параметров NsimScan одним (или схожими) цветами, чтобы было легко отличить его от других методов.

стр 16 «При этом склонность к образованию кассет у некоторых кластеров среднего размера, включающих от двух до четырех тысяч генов, составляла более 90% (здесь представлены, в том числе, транспортеры систем ABC). Самые маленькие кластеры, включающих менее двух тысяч генов, с наиболее высокой склонностью к образованию кассет, принадлежали к классам дегидрогеназ, изомераз, киназ, эстераз и трансальдоз/транскетолаз.» - вопреки тексту я не обнаружил ни одного класса с склонностью более 90% на рис 9, более того, шкала на этом рисунке идет только до 80%... Дополнительно, самая высокая склонность к образованию кассет, которую я смог обнаружить на рисунке 9 для дегидрогеназ (желтые кружки) явно меньше 40%

стр 18 «Наиболее **распространенным** участником кассет среди функциональных классов оказались **гены**,» - число не согласовано

Стр 23 «мутантом по *yihW* и мутантом по *crp*» - под словом «мутант» имеется в виду нокаут по *yihW*?

рис 13А — очень неудобно организован, цвет служит одновременно для разделения и культур клеток и сред (глюкоза/лактоза) при этом среды так же разделены на рисунке пространственно, было бы гораздо проще сравнивать, если бы цвет обозначал только культуры.

Несмотря на указанные замечания и вопросы работа выполнена на высоком методологическом уровне, основные результаты не вызывают сомнений. По результатам работы опубликовано три статьи в международных научных журналах и сделано восемь докладов на конференциях. Работа «Геномная ко-локализация генов углеводного метаболизма бактерий» удовлетворяет требованиям ВАК, предъявляемым к диссертационным работам, а её автор Анна Денисовна Казнадзей заслуживает присуждения степени кандидата биологических наук.

к.б.н., н.с. Сколтех, CNBR  
Мазин Павел Владимирович

*sa.a.ka@gmail.com*



*13.09.2019*

Подпись *Мазин*  
РУКОВОДИТЕЛЬ ОТДЕЛА  
КАДРОВОГО АДМИНИСТРИРОВАНИЯ

*А.В. Мазин*  
*13.09.2019*