

ОТЗЫВ

на автореферат диссертации Семёна Александровича Лейна
«Эволюция транскрипционной регуляции метаболизма углеводов в бактериях»,
представленной на соискание учёной степени кандидата биологических наук по
специальности 03.01.09 - математическая биология, биоинформатика.

В пост-геномную эру одной из основных научных проблем, стоящих перед молекулярной биологией, является адекватная интерпретация результатов, полученных в результате секвенирования геномов. При этом особое значение приобретает задача аннотирования новых генов. В настоящее время лишь малое число становящихся доступными аминокислотных последовательностей принадлежит белкам с известной функцией, например, ферментам с экспериментально охарактеризованной энзиматической активностью. Огромное количество последовательностей в базах данных соответствует открытым рамкам считывания, обнаруженным при выполнении геномных проектов. Достигнутые к настоящему времени темпы накопления новых последовательностей делают невозможным экспериментальное изучение каждого белка. Возникает необходимость аннотирования генов на основании исключительно биоинформатических методов. Эти методы должны иметь максимально высокую предсказательную силу в отношении функций белков и быть способными использовать при этом исключительно данные о геномной последовательности организма. Одним из таких подходов является выявление и аннотирование регуляторных участков ДНК (промоторов, терминаторов, РНК-переключателей и сайтов связывания транскрипционных факторов), а также предсказание новых ДНК-связывающих белков для обнаруженных сайтов ДНК.

Диссертационная работа С.А. Лейна посвящена исследованию структуры целого ряда регулонов углеводного метаболизма у бактерий из отделов *Firmicutes* и *Proteobacteria* с помощью методов сравнительной геномики. Работа состоит из четырёх достаточно самостоятельных частей. Первая из них посвящена исследованию арабинозного регулона, контролируемого фактором транскрипции AraR, у бактерий отдела *Firmicutes*. Обнаружено, что ортологи белка AraR помимо фирмикут есть только у бактерий из порядка *Thermotogales*. Диссертантом были построены пять матриц позиционных весов для распознавания сайтов связывания AraR в бактериях разных таксонов и реконструированы их арабинозные регулоны. На основе филогении фактора транскрипции AraR сделаны выводы о возможных путях эволюции этих регулонов.

Вторая часть диссертационной работы посвящена исследованию путей утилизации галактозамина и N-ацетилгалактозамина, контролируемых фактором транскрипции AgaR, у бактерий отдела *Proteobacteria*. Филогенетический анализ позволил выделить пять

групп белков AgaR, для каждой из них был установлен распознаваемый мотив ДНК. Показано, что в разных видах протеобактерий существенно отличается набор генов, находящихся под контролем AgaR. Сделан вывод о ключевой роли горизонтальных переносов и дубликаций генов в эволюции AgaR-регулона.

Третья часть работы посвящена исследованию регуляции метаболизма углеводов фактором транскрипции HexR у бактерий отдела *Proteobacteria*. Проведённый анализ показал, что HexR-регулоны значительно различаются у разных бактерий как по размеру, так и по составу.

Четвёртая часть работы посвящена исследованию регуляции катаболизма сахаров в бактериях семейства *Bacillaceae*. Обнаружено, что в сеть регуляции метаболизма углеводов в бактериях этого семейства входит более 300 ортологичных рядов регулируемых генов. Среди 43 реконструированных регулонов лишь 14 полностью сохраняют свой состав во всех десяти исследованных геномах бактерий. Диссертант делает вывод о том, что в семействе *Bacillaceae* наблюдается большая пластичность транскрипционной регуляции, отвечающей за экспрессию генов метаболизма сахаров.

Содержание автореферата показывает, что диссертант в совершенстве овладел методами анализа нуклеотидных и аминокислотных последовательностей. Однако следует отметить, что он не продемонстрировал достаточного владения компьютерными базами данных со структурными и функциональными классификациями белков. Белки и гены в тексте называются по наиболее известному их представителю, т.е. тривиальными четырёхбуквенными именами. При этом не указываются названия соответствующих семейств в какой-либо глобальной классификации белков, например, PFAM или COG (во всём автореферате лишь однажды упомянут COG1820), что затрудняет использование читателем приводимых диссертантом результатов (в тексте самой диссертации есть только по одному упоминанию PF00248 и COG1820). Следует напомнить, что одни и те же четырёхбуквенные названия в разных организмах могут принадлежать негомологичным и даже функционально не связанным белкам. Например, диссертант использует названия *agaR* и *agaS* для обозначения генов белка-регулятора и фермента-изомеразы, в то время как в базе данных NCBI оба эти названия в случае бактерии *Pediococcus pentosaceus* использованы для обозначения генов фермента α -галактозидазы (GenBank, L32093). В тексте автореферата упоминаются представители двух принципиально разных групп гидролаз – амидогидролаз (стр. 12) и гликозил-гидролаз (стр. 13), но это не помешало диссертанту в других случаях говорить просто о гидролазах (стр. 9 и 17), что не позволяет читателю понять о ферментах какой именно группы идёт речь. Использование четырёхбуквенных названий белков сыграло злую шутку и с самим

диссертантом: в одном из случаев он перепутал названия двух разных регуляторов транскрипции – AgaR и AraR (стр. 11).

На мой взгляд, название диссертационной работы выбрано не очень удачно. Во-первых, три подряд идущих существительных в родительном падеже плохо выглядят с стилистической точки зрения. Во-вторых, в работе не исследовался как таковой процесс регуляции транскрипции в его динамике (этому вопросу посвящён только один из разделов обзора литературы), а диссертант занимался изучением белков-регуляторов и распознаваемых ими мотивов в промоторных областях генов. В-третьих, эволюция этого процесса и подавно не была предметом исследования рассматриваемой работы. Хорошо известно, что «*ничто в биологии не имеет смысла, кроме как в свете эволюции*», но это не является основанием для использования слова «эволюция» в названии каждой работы в области биологии. Конечно, эволюционный аспект присутствует в данной работе, но если мы обратимся даже к выводам, сделанным самим диссертантом в конце работы, то увидим, что эволюция упоминается лишь в двух из них.

Содержание автореферата не позволило мне понять логического основания сделанного диссертантом выбора набора изучаемых белков-регуляторов (несколько десятков групп ортологов) и набора исследуемых организмов (отдельные представители отделов *Firmicutes*, *Proteobacteria* и *Thermotogae*). Описание применённой методологии также представляется недостаточными: не ясно на основании какой формальной процедуры диссертант выбрал немногочисленные исследованные геномы из тысяч доступных в базах данных. Каким образом «была сделана репрезентативная выборка из 10 геномов семейства *Bacillaceae* исходя из филогенетического дерева видов»? Почему «ортологи фактора транскрипции AgaR (на самом деле AgaR – Д.Н.) были найдены в 19 геномах гамма-протеобактерий и двух геномах бета- и альфа-протеобактерий»? Ведь даже число доступных геномов *E. coli* исчисляется десятками. И, наконец, какое формальное определение ортолога было использовано в данной работе? Какими были пороги? Что делали с псевдогенами? Пытались ли понять причину существования в базах данных очень близких гомологов, аннотированных эукариотными белками (например, XP_004153705.1 и XP_004528836.1)?

В тексте автореферата есть предложение: «Геномы *Photothabdus luminescens* и *Aeromonas hydrophila* содержат ген гидролазы AgaH из семейства GH36, которая скорее всего также является альфа-N-ацетилгалактозаминидазой.» (стр. 13-14). Обратившись к базе данных CAZy (<http://www.cazy.org>), я обнаружил, что геном *Photothabdus luminescens* subsp. *laumondii* TTO1 действительно кодирует один белок семейства GH36 (NCBI, CAE13135), а бактерия *Aeromonas hydrophila* представлена там шестью штаммами. Чтобы выбрать из них тот, который изучался диссертантом пришлось воспользоваться текстом

диссертационной работы. Оказалось, что это – штамм *Aeromonas hydrophila* subsp. *hydrophila* ATCC 7966, однако его геном, согласно базе данных CAZy, кодирует сразу два белка семейства GH36 (NCBI, ABK36496 и ABK38110). На основании данных, содержащихся как в автореферате, так и в самой диссертации, понять о каком из них идёт речь мне не удалось. Ни один из них не проаннотирован как AgaH в базе данных NCBI. Однако уже несколько лет назад было показано, что семейство GH36 является гетерогенным и было предложено разбить его на 11 более мелких семейств с монофилетическим статусом (PMID:21639842). В состав одного из них (GH36D) попала единственная (в бывшем семействе GH36) экспериментально охарактеризованная α -N-ацетилгалактозаминидаза из *Clostridium perfringens* ATCC 10543 (PMID:12204375). В составе этого же семейства GH36D оказались также белки *P. luminescens* (NCBI, CAE13135) и *Aeromonas hydrophila* (NCBI, ABK38110). Таким образом, сделанный диссертантом вывод о том, что два обнаруженных им белка из *P. luminescens* и *A. hydrophila* являются α -N-ацетилгалактозаминидазами не обладает научной новизной. Вместе с тем, это утверждение не попало в список «основных результатов и выводов» и поэтому не может рассматриваться как серьёзная претензия к работе.

Хочется обратить внимание на не всегда удачные формулировки в тексте автореферата и диссертации. В автореферате несколько раз говорится про «методы сравнительной геномики», в т.ч. в целях и в задачах работы, однако в самой диссертации имеется утверждение «сравнительная геномика является эффективным методом для исследования...» (стр. 7), указывающее на то, что речь идёт лишь об одном методе. Некоторое противоречие между авторефератом и диссертацией наблюдается и при сравнении белков AgaB и AgaB-II. Так в автореферате говорится, что «AgaB-II на филогенетическом дереве FGGY семейства киназ стоит достаточно близко к AgaB» (стр. 11), в то время как в диссертации утверждается, что «ветви отвечающие AgaB, AgaB-II и AgaK находятся на большом филогенетическом расстоянии друг от друга» (стр. 61). В качестве цели работы назван «анализ эволюции» (стр. 4 автореферата), однако анализ является процессом, а не результатом и поэтому он едва ли может быть целью научной работы. Утверждение «структура оперонов и регулонов сильно различается даже внутри одного таксона» (стр. 9) не имеет особого смысла, пока не указано о таксоне какого уровня идёт речь. В автореферате многократно используется редко употребляемый в русскоязычной литературе термин «регулог» без объяснения его значения. К числу неудачных формулировок, на мой взгляд, следует отнести и одновременное использование терминов «метаболизм углерода» и «метаболизм углеводов» в списке «основных результатов и выводов» (выводы 3 и 4 соответственно).

Вывод диссертанта о существенной роли горизонтального переноса в эволюции генов метаболизма сахаров является довольно банальным, однако нигде в автореферате не сопровождается ссылками на более ранние работы (надеюсь, что в самой диссертации такие ссылки имеются).

Подписи к рисункам 2 и 5 в автореферате недостаточно подробны, чтобы разобраться в их содержании. Автореферат содержит довольно много опечаток. Некоторые из них существенно искажают смысл: диссертант очень самокритично именует свою работу «нестоящей» (стр. 15), а среди изученных им организмов оказались даже гномы (стр. 17). Некоторые опечатки носят систематический характер, что позволяет сделать вывод о том, что они не являются случайными, а отражают точку зрения автора. Так прилагательное «раннее» семь раз в тексте автореферата используется вместо наречия «ранее» (та же проблема была мной обнаружена и в тексте самой диссертации). В автореферате шесть раз использовано «et.al.» вместо «et al.», которое должно расшифровываться латинской фразой «et alii» (т.е. «и другие»). На стр. 6 автореферата утверждается, что список литературы в диссертации включает 247 наименований, в то время как на самом деле их 248, притом в этом списке фамилии авторов даже в тех случаях, когда они указаны на русском языке соединяются почему-то английским союзом «and». Список публикаций диссертанта в автореферате (стр. 22-23) и в диссертации (стр. 111-113) не совпадает и в обоих случаях не оформлен по единому стилю.

В целом, работа выполнена на современном уровне, а полученные результаты представляют большой теоретический и практический интерес. Считаю, что рецензируемая работа соответствует требованиям, предъявляемым к кандидатским диссертациям, а её автор, С.А. Лейн, заслуживает присуждения учёной степени кандидата биологических наук по специальности «математическая биология, биоинформатика».

2 октября 2014 года



Старший научный сотрудник
Института микробиологии им. С.Н. Виноградского РАН
(117312, г. Москва, Проспект 60-летия Октября, д. 7, корп. 2)
кандидат биологических наук
Даниил Геннадиевич Наумов

тел. (499) 135-05-91
эл. почта: daniil_naumoff@yahoo.com



СОБСТВЕННОРУЧНАЯ ПОДПИСЬ
ТОВ. Д. Г. Наумова
УДОСТВЕРЯЕТСЯ.

Подпись Даниил Геннадиевич Наумов