

ОТЗЫВ

официального оппонента на диссертационную работу Семёна Александровича Лейн "Эволюция транскрипционной регуляции метаболизма углеводов в бактериях", представленной на соискание ученой степени кандидата биологических наук по специальности 03.01.09 – «математическая биология, биоинформатика»

Особенностью современной биологии является появление огромного массива генетических данных, отражающих видовое многообразие планеты. Это облегчает аннотацию новых геномов, создаёт основу для их сравнительного изучения и реконструкции путей эволюции. Такого типа исследования методами сравнительной геномики имеют большое фундаментальное и практическое значение, т.к. позволяют обнаруживать действующие в биологических объектах метаболические системы и понять способы их регуляции. В настоящее время они особенно востребованы, т.к. в дополнение к данным экспрессионного анализа именно сейчас накапливаются данные полногеномного скрининга мест связывания факторов транскрипции, поэтому результаты информационного поиска регулонов могут предоставить важные ориентиры для систематизации и интерпретации результатов экспериментального тестирования. Актуальность рецензируемой работы определяется ещё и тем, что в качестве объекта исследования были выбраны системы метаболизма углеводов у бактерий. Т.к. углеводы являются основными источниками углерода и энергии, системы их утилизации эволюционно адаптированы для максимально эффективного использования всех возможных субстратов. При системном анализе это повышает вероятность обнаружения новых ферментов или новых регуляторных связей. Более того, именно бактериальное сообщество играет ведущую роль в ферментативном гидролизе углеводов у всех биологических объектов, поэтому первостепенное значение имеет информация о степени разнообразия используемых при этом метаболических путей.

Диссертационная работа изложена на 173 страницах и построена по стандартной схеме, включающей Введение, Обзор литературных источников, раздел «Материалы и методы», 4 главы с описанием проведённых исследований и Выводы. Список литературы включает 247 наименований. Работа проиллюстрирована 24 рисунками, а наиболее значимые результаты представлены в 9 таблицах и 4 приложениях. В первой главе, кроме короткого введения в суть проблемы, сформулирована цель диссертационной работы, конкретные задачи исследования, научная новизна и практическое значение работы, а также перечислены основные положения, выносимые на защиту.

В обзоре литературных данных рассмотрены основные принципы регуляции транскрипции у бактерий, а также методы сравнительной геномики, позволяющие аннотировать геномы и получать информацию об организации регуляторных сетей. Более обстоятельно изложена информация об изученных Семёном Александровичем регулонах AgaR, AgaR и HexR, а в последнем разделе представлены данные о механизмах регуляции путей утилизации сахаров у бацилл. Можно поспорить с некоторыми базовыми утверждениями автора, например с тем, что основной причиной способности бактерий быстро адаптироваться к разным условиям роста является «экономное использование своего генетического материала для экспрессии необходимых генов в правильное время» (стр. 11) или с определением оперона как набора «всех перекрывающихся транскрипционных единиц», уже хотя бы потому, что оперон это структурная единица генома, а транскрипционная единица – транскриптома (стр. 11), или с утверждением, что небольшие различия в мотивах сайтов связывания указывают на «отсутствие пересечения» регулонов (стр. 85). Есть фактические погрешности, например на стр. 14 написано, что σ -фактор отсоединяется от РНК-полимеразы «после формирования открытого комплекса». На самом деле, после синтеза 10-14 фосфодиэфирных связей и

иногда дольше остаётся в элонгирующем комплексе. На стр. 18 написано, что «активаторы класса I связываются с сайтом в 5' области -35 элемента промотора и взаимодействуют с С-концевым участком α субъединицы РНК полимеразы». На самом деле, выше элемента -35 (-45/-55) находится UP-элемент, т.е. сайты связывания С-концевых доменов α -субъединиц, а активаторы класса I присоединяются к промотору в соседних более удалённых участках, а центр сайта связывания для CRP в приведённом в Обзоре примере, находится в позиции -61.5. Тем не менее, тема и содержание Обзора точно соответствуют результативной части работы, что позволяет читателям составить представление о важности поставленных задач и спектре приемлемых для их решения подходов.

Методический раздел очень лаконичен. Он изложен всего на 7 страницах, из которых 3.5 занимает таблица со списком проанализированных геномов. Отчасти это обусловлено тем, что для решения поставленных задач было использовано уже готовое программное обеспечение, не требующее подробного описания. Тем не менее, некоторые методические приёмы и параметрические детали остались не ясными. Так, например, приходится только догадываться, какие критерии использовались для исключения близкородственных штаммов из списка геномов, содержащих гомологи анализируемых факторов транскрипции и о том, каким образом «белковые последовательности факторов транскрипции из оставшихся геномов проверялись на ортологичность».

Результаты исследовательской работы изложены в четырёх главах. Три из них посвящены регулонам факторов транскрипции AraR, AgaR и HexR, а четвёртая эволюции регуляторных систем катаболизма сахаров у бацилл. Особенностью AraR является его химерная структурная организация, в которой ДНК-связывающий домен принадлежит GntR семейству регуляторных белков, использующих POU домен для взаимодействия с ДНК, а субстрат-связывающий - к регуляторам семейства LacI. Поэтому гомологичные гены для изучения эволюции AraR искали для каждого из доменов в отдельности, а затем было собрано 28 геномов, попавших в обе выборки. Такая стратегия вполне приемлема, хотя закономерно возникает вопрос о том, что изменилось бы, если бы выборка была составлена при прямом поиске по двум доменам. Не понятно также, почему филогенетический анализ для гомологов этого белка осуществлялся только по всей первичной структуре (Рис. 3.1). Так как конечной целью работы был поиск регулируемых генов через поиск консервативных мотивов в их регуляторных участках, возникает вопрос, что изменилось бы, если бы филогенетические деревья были построены по ДНК-связывающим доменам. Семёном Александровичем были найдены потенциальные сайты связывания AraR в геномах: *Bacillales*, *Lactobacillales*, *Clostridiales* и *Thermotogales*, определена оперонная организация соответствующих генов и предложена функциональная аннотация для ряда неортологичных замещений. Наиболее ценным, вероятно, является сформулированная модель эволюции AraR регулона, включающая его таксон-специфическое расширение за счёт горизонтального переноса генов. Эти результаты представляют большую ценность и вполне адекватно сформулированы в первом выводе.

Второй вывод суммирует результаты, полученные при исследовании регулона AgaR. Этот регулятор по способности связываться с ДНК принадлежит к семейству DeoR и использует для этого домен, немного отличающийся от AraR по структурной организации (winged helix-turn-helix domain). Для характеристики его регулона был собран 21 геном, а филогенетический анализ ортологов позволил выявить 5 таксономических групп. Для поиска потенциальных сайтов связывания AgaR был использован метод филогенетического футпринтинга. Множественное выравнивание регуляторных областей AgaR-зависимых генов в каждой из групп позволило выявить консервативные мотивы,

имеющие общую составляющую (СТТТС или GAAAG) с разным взаимным расположением в разных группах. Эти мотивы были использованы для построения матриц позиционных весов, при помощи которых в дальнейшем было осуществлено полногеномное сканирование. Анализ оперонов и реконструкция регулонов утилизации N-ацетилгалактозамина (НАГА) позволила аннотировать ряд новых генов, так, например, у *Shewanella* был обнаружен паралог амидогидролазы NagA, а также гены, предположительно кодирующие пермеазу (agaP), киназу (agaK) и TonB-зависимый транспортёр внешней мембраны. Анализ выявленных оперонов свидетельствует о том, что движущей силой эволюции AgaR регулона был не только горизонтальный перенос, но и дупликация генов. В AgaR регулоне у *Shewanella*, например, таксон-специфические пермеаза AgaP и деацетилаза AgaA-II, могли быть образованы путем дупликации генов NagP и NagA с последующим приобретением новых функций. В рамках проведенного анализа в регулоне AgaR была обнаружена возможность двух любопытных эволюционных событий: недавнего горизонтального переноса из общего источника кластера aga генов *agaRZS-ptsIII-agaA* в 4 генома энтеробактерий и в геном *V. fisheri* и то, что один из двух *aga* кластеров у *S. proteamaculans* мог образоваться в результате дупликации предкового оперона с последующей потерей некоторых генов. Однако наиболее весомым результатом этой части работы, по-видимому, является установленный факт, что наиболее консервативным геном AgaR регулона является ген изомеразы *agaS*, которую, следовательно, можно считать основным ферментом, выполняющим функцию ГА-6-фосфат деаминазы/изомеразы, т.к. ген *agaI*, которому ранее приписывали эту функцию имеется только у двух организмов.

Особенностью регулятора HexR, исследованного в третьем разделе, является относительно большое разнообразие узнаваемых мотивов и наличие больших регулонов у *Vibrionales*, *Aeromonadales*, *Psychromonadales* и *Alteromonadales*, что позволяет квалифицировать HexR как глобальный регулятор у этих бактерий (третий вывод).

Самым масштабным является четвёртый раздел, посвящённый изучению эволюции регуляторных систем катаболизма сахаров у бацилл. Это исследование включало несколько последовательных этапов, начиная с поиска ортологов в 10 геномах для 43 факторов транскрипции из *B. subtilis*. Кроме хорошо изученных регуляторов, среди них были и совсем неисследованные факторы транскрипции. Используя три разных подхода, в зависимости от степени изученности конкретных регуляторных систем, были построены весовые матрицы для поиска сайтов связывания этих белков. Для каждого из всех 43 факторов удалось получить одну весовую матрицу, что указывает, с одной стороны, на консервативность соответствующих мотивов, а с другой, на адекватность полученных выборок. Контекст потенциально распознаваемых мотивов и размер вероятных регулонов представлены в чрезвычайно информативной Таблице 6.1, которая вполне обоснованно позволяет сделать основополагающий вывод о высокой степени эволюционной вариабельности регуляторных систем у прокариот.

Рецензируемая работа, таким образом, является цельным и законченным в рамках поставленных задач исследованием, в котором получен большой объём важного фактического материала. Весь он изложен четко и логично. Досадным является отсутствие информации о том, каким образом оценивалась предсказательная сила поисковых программ, т.е. каким образом составлялись тестовые компиляции (стр. 56), каков уровень случайного обнаружения целевых сайтов, как много потенциальных сайтов связывания было найдено в геномах помимо мест их ожидаемого присутствия. Смущают такие формулировки, как: «В результате анализа регулируемых генов оказалось...» (стр. 61) в тех случаях, когда речь идёт только об обнаружении потенциальных сайтов связывания факторов транскрипции, а регуляторное воздействие не установлено. Это тем

более удивительно, что автор удалил потенциальный сайт для связывания araR из обучающей компиляции на том основании, что экспериментально связь с этим фактором не была обнаружена. Кроме этого, в разделе «Научная новизна и практическое значение работы» написано, что: «...для каждого нового белка-регулятора были ... предсказаны ... молекулы-эффекторы». Это явное преувеличение.

Работа хорошо оформлена и проиллюстрирована, хотя на Рис. 1.1., Рис. 1.2. и Рис. 1.3. неправильно пронумерованы домены сигма-субъединицы РНК-полимеразы. Кроме того, не всегда имеются необходимые ссылки, например, на стр. 85 нет ссылки на экспериментальную работу. Обычных опечаток и синтаксических ошибок в работе мало (стр. 29, 34, 38, 40, 86), но почти на каждой странице есть погрешности печати (разрыв слов). Есть небольшое количество несогласованных предложений и стилистических погрешностей («...им были приписаны предполагаемые функциональные аннотации...», стр.73) и неудачных выражений («метод ... флуоресцентной поляризации», стр.90). Используется разное написание слова промотор, , на стр.21, 35.

Это, конечно, несколько не снижает фундаментальной значимости рецензируемой работы, которая является хорошо продуманным и полноценным исследованием. Достоверность результатов не вызывает никаких сомнений, а сформулированные выводы обоснованы и убедительны. Автореферат соответствует содержанию диссертации, а в 15 публикациях автора отражены практически все результаты.

По актуальности темы, объему и важности проведенных исследований и сделанных выводов рецензируемая работа, несомненно, соответствует требованиям, предъявляемым ВАК к диссертациям на соискание ученой степени кандидата биологических наук по специальности 03.01.09 – «математическая биология, биоинформатика», а автор заслуживает искомой степени.

Зав. лабораторией
функциональной геномики и клеточного стресса,
ФГБУН Институт биофизики клетки РАН
докт. биол. наук, профессор
142290, Московская область,
г.Пушино, ул.Институтская 3
ozoline@icb.psn.ru

О.Н.Озолин

31 августа 2014 г.

