

ОТЗЫВ

официального оппонента на диссертационную работу
Суворовой Инны Андреевны

на тему «Коэволюция транскрипционных факторов семейства GntR и их сайтов связывания», представленную на соискание ученой степени кандидата биологических наук по специальности 03.01.09 – математическая биология, биоинформатика.

Диссертация Инны Андреевны Суворовой посвящена масштабному изучению ДНК-паттернов, мотивов, распознаваемых факторами транскрипции семейства GntR, в контексте коэволюции мотивов сайтов связывания и соответствующих ДНК-связывающих доменов, а также локализации и представленности сайтов связывания в геномах различных бактерий. В ходе работы проанализировано более 1000 транскрипционных факторов и их сайтов связывания в более чем 300 бактериальных геномов. Были установлены как общие закономерности, характерные для мотивов семейства GntR, так и ряд интересных частных особенностей. В частности, для подсемейств FadR, HutC и YtrA получены корреляционные карты и идентифицированы конкретные соответствия между коэволюционирующими позициями сайтов связывания и позициями белковых выравниваний, соответствующими аминокислотам, узнающим эти нуклеотиды. Кроме того, для различных членов семейства GntR продемонстрирован крайне интересный феномен кластеризации хорошо выраженных боксов сайтов связывания. Это явление, судя по всему, схоже с гомотипической кластеризацией сайтов связывания, которая изучена у высших эукариот и является у них одним из важнейших механизмов, обеспечивающих устойчивость регуляторной сети. Также в работе исследованы особенности локализации одиночных и двойных сайтов связывания в дивергонах. В качестве конкретного исследования, с использованием

построенных мотивов проведен анализ регуляции метаболизма гексуронатов у Gammaproteobacteria, а также малоната и пропионата у Proteobacteria. Для идентификации и анализа мотивов использованы современные методы биоинформатики и сравнительной геномики, автор уверен в владении основными программными инструментами и методами статистической обработки результатов. Представленная работа предоставляет фундамент для дальнейшего анализа, например, интересен вопрос о согласованности иерархической кластеризации мотивов сайтов связывания и соответствующих им ДНК-связывающих доменов с учетом таксономической принадлежности.

Суммируя вышеизложенное, безусловно, высока актуальность темы диссертации, ее научная и практическая ценность. Не вызывает сомнений высокая квалификация автора как исследователя и способность автора эффективно использовать биоинформационный инструментарий. Научные статьи по теме работы опубликованы в достойных рецензируемых журналах.

К сожалению, работа не лишена определенных недостатков, как в части оформления, так и в части подачи материала; это, конечно, не снижает научной ценности работы, но затрудняет восприятие и интерпретацию материала.

Во-первых, ряд вопросов вызывает обзор литературы по теме диссертации, который по объему составляет более 30 страниц и, в основном, описывает хорошо изученные особенности генной регуляции у бактерий. На мой взгляд, ценность подробного изложения базовых принципов невысока, а ряд специфических подробностей (сигма-факторы, стадии транскрипции, регуляция экспрессии альтернативными структурами РНК) не имеет прямого отношения к теме диссертации, посвященной конкретному семейству факторов транскрипции GntR. Ряд сведений, приводимых в обзоре, относится к эукариотам: «Типичные для эукариот домены типа лейциновая застежка» (страница 18), «первым секвенированным эукариотическим геномом в 1996 году стал геном дрожжей *Saccharomyces cerevisiae*» (страница 24), «в базе данных KEGG ... насчитывается 3580 полных последовательностей геномов бактерий, 218 – архей, и 313 – эукариот» (там же). Есть и еще менее релевантная информация, например «существует целый

ряд программ поиска потенциальных генов» (Стр. 24). Раздел «Сравнительно-геномные методы исследования», страница 23, почему-то начинается с описания экспериментальных методов изучения ДНК-белкового взаимодействия. Суммируя замечания к тексту литературного обзора: при удивительной дотошности, изложение материала лишено связности в описании ряда методических и даже терминологических моментов. Например (стр. 28) «для каждого нуклеотида учитывается его частота и консервативность позиции»; необходимо явно различать эволюционную консервативность и стабильность нуклеотидов в конкретной позиции сайта связывания. Там же читаем: «Одной из главных проблем поиска сайтов связывания является выбор порога», но понятие порога распознавания не введено. Ключевой момент обзора, «Семейство транскрипционных факторов GNTR», который хотелось бы видеть в качестве самостоятельного раздела, является подпунктом третьего уровня в подразделе «1.4 ДНК-белковые взаимодействия»; а первый рисунок появляется только на странице 38.

Претензии к оформлению возникают и при знакомстве с рисунками и таблицами, например Рисунки 6,7,8,9,12 лишены подписей осей и размерностей; а в подписи таблице 3 не прокомментирован смысл значений в колонках, названных «Наблюдаемое» и «Ожидаемое».

Есть и содержательные вопросы к работе. Например, не ясно насколько проведенное исследование автоматизировано и опирается на формальные критерии, т.е. насколько оно может быть воспроизведено без участия эксперта. В частности, интересен выбор параметров: (стр. 41) «пороговое значение e-value» и (стр. 42) «сайты с более слабым весом (на 10% ниже порога)». Является ли выбор этих параметров критичным для стабильности получаемых мотивов связывания?

Не смотря на наличие объемного приложения ряд фактических данных в работе опущен. Так, на мой взгляд, критичной информацией является объем выборок идентифицированных сайтов связывания, которые использовались для обучения (построения) моделей мотивов в форме позиционно-весовых матриц. В частности, маленький объем выборок мог бы объяснить особенности Logo-визуализаций в Приложении Б, где наиболее стабильные (100% консервативные)

нуклеотиды в сайтах связывания имеют информационное содержание в районе 1 бита (ManR) или ровно 1 бит (DVU2802). Также, выравнивание мотивов друг относительно друга помогло бы оценке похожести между подсемействами.

Тем не менее, еще раз необходимо подчеркнуть, что указанные недостатки не являются принципиальными и не снижают высокого уровня представленной диссертационной работы.

Таким образом, в заключение, считаю необходимым отметить, что диссертационная работа Суворовой Инны Андреевны на тему «Коэволюция транскрипционных факторов семейства GntR и их сайтов связывания» является законченным научно-квалифицированным исследованием и полностью соответствует требованиям «Положения о порядке присуждения ученых степеней», утвержденного Постановлением Правительства РФ от 24 сентября 2013 г. №842. В свою очередь, Инна Андреевна определенно заслуживает присуждения учёной степени кандидата биологических наук по специальности 03.01.09 «математическая биология, биоинформатика».

кандидат физико-математических наук, старший научный сотрудник
Лаборатории вычислительных методов системной биологии
Федерального государственного бюджетного учреждения науки
Института молекулярной биологии имени В.А. Энгельгардта
Российской академии наук (ИМБ РАН)
ГСП-1, 119991, г. Москва, ул. Вавилова, д. 32
ivan.kulakovskiy@eimb.ru
+7 495 135 60 00

 / И.В. Кулаковский /
13 апреля 2016 года

