

ОТЗЫВ

официального оппонента на диссертационную работу

Зверкова Олега Анатольевича

**«Функции и эволюция РНК-полимераз в митохондриях и пластидах»,
представленную к защите на соискание ученой степени кандидата
физико-математических наук по специальности 03.01.09 –
«математическая биология, биоинформатика».**

В диссертационной работе О.А. Зверкова получены новые научные результаты о процессах транскрипции пластидных и митохондриальных геномов и о белках, закодированных в пластидах разных групп растений. Эти результаты получены с помощью разработанных соискателем оригинальных алгоритмов, которые были реализованы в компьютерных программах и успешно применены для решения возникающих биологических задач.

Тема диссертации, безусловно, является актуальной как в плане фундаментальной науки, так и в связи с изучением механизмов митохондриально обусловленных патологий человека.

Наиболее интересным результатом является предложенная О.А. Зверковым математическая модель транскрипции пластидных и митохондриальных геномов. В модели формализованы следующие процессы, влияющие на уровень транскрипции генов. Участие в транскрипции двух типов РНК-полимераз – фагового типа (работают в митохондриях и пластидах) и бактериального типа с разными сигма-субъединицами (работают в пластидах); процесс инициации транскрипции, включая abortивные попытки транскрипции (у бактериального типа РНК-полимераз), с учетом того, свободен ли промотор; процесс элонгации РНК-полимераз с учетом скорости транскрипции и разрешения конфликтов при столкновении двух РНК-полимераз, транскрибирующих противоположные цепочки ДНК или одну и ту же; процесс терминации транскрипции на терминаторах - шпильках ДНК, G-квадруплексе или при связывании специальной последовательности транскрипционным фактором (белком mTERF, обнаруженным у животных и растений); связывание транскрипционных факторов и взаимодействие их с РНК-полимеразами. Учтены также скорости распада мРНК.

Входными данными для модели является нуклеотидная последовательность митохондрии или пластиды с аннотированными генами, промоторами, терминаторами и сайтами связывания транскрипционных факторов. Модель является симуляционной, в ней моделируется связывание промоторов полимеразами и их понуклеотидное передвижение по ДНК в процессе транскрипции с течением времени. В приложениях моделируется транскрипция митохондрий целиком и транскрипция отдельных локусов пластид. Последнее ограничение вызвано ограничениями компьютерных

ресурсов. Выходными данными являются уровни транскрипции генов входного локуса. Также на выходе можно получить частоты определенных событий, например, столкновений полимераза.

Предложенная модель была реализована в компьютерной программе, доступной для потенциальных пользователей.

Следует отметить, что аналогов модели транскрипции подобной полноты в отношении числа моделируемых объектов в литературе мной не найдено (см. обзоры Vilar, Saiz 2013 года и Segal, Widom 2009 года). Ранее предлагались модели транскрипции только отдельных генов. О.Зверков выбрал удачные объекты для моделирования - пластиды и митохондрии, которые содержат относительно мало генов. К тому же в ряде случаев промоторы, терминаторы и регуляторные участки достаточно полно аннотированы. Кроме того, для нескольких организмов имеются экспериментальные данные об уровнях транскрипции генов, закодированных в геномах этих органелл.

Важными для оценки предложенной модели являются следующие обстоятельства. 1) Модель имеет сравнительно мало свободных параметров: в рассмотренных приложениях 4-6 или немногим больше. 2) Свободные параметры модели в конкретном случае определяются исходя из экспериментальных данных, прежде всего, об уровнях мРНК транскрибируемых генов. Эта обратная задача решается перебором допустимых наборов значений параметров с использованием разумных эвристик, сокращающих перебор. По-видимому, решение именно этой задачи является лимитирующим фактором по компьютерным ресурсам.

Подгонка параметров модели "под ответ" опасна в отношении переоптимизации и требует проверки. Вероятно, уникальность и ограниченность экспериментальных данных не позволила применить методы скользящего контроля. Тем не менее, косвенные подтверждения адекватности найденных параметров представлены соискателем. Наиболее красивое подтверждение получено при моделировании локуса 3 из генома пластиды *Arabidopsis thaliana*. В диссертации описано, что первоначально добиться соответствия уровней мРНК в модели и в эксперименте не удавалось. Исходя из этого, было сделано предположение о наличии в локусе неаннотированных терминаторов транскрипции. Путем перебора найдены положения двух терминаторов транскрипции, при введении которых модель давала хорошее соответствие экспериментальным данным. После этого на участках предсказанных таким способом положений терминаторов были обнаружены протяженные инвертированные повторы, которые могут образовывать крестообразные структуры ДНК или шпильки мРНК. Они могут быть терминаторами транскрипции. Оказалось, что на одном из участков для других растений ранее в литературе описан

консервативный инвертированный повтор и высказана гипотеза о возможной его роли как терминатора транскрипции.

Некоторые параметры модели оценивались Зверковым косвенными способами из-за отсутствия прямых экспериментальных данных. Так, скорость элонгации РНК-полимеразы бактериального типа и ее зависимость от температуры были взяты из работ о *E.coli*. Ограничение снизу скорости элонгации РНК-полимеразы фагового типа было рассчитано исходя из сравнения длин первого кодирующего экзона и следующего интрона в некоторых генах пластид и скорости передвижения рибосомы: транскрипция интрона и его сплайсинг должны завершиться до того, как рибосома дойдет до него.

Модель была применена О.А. Зверковым для шести локусов: полных митохондрий лягушки, крысы и человека, а также двух локусов генома пластиды резуховидки Таля (*Arabidopsis thaliana*) и одного локуса ячменя (*Hordeum vulgare*). Модель позволила сделать предсказания об интенсивностях связывания РНК-полимераз с разными промоторами во всех изученных локусах и о других параметрах. Например, о вероятностях “протекания” mTERF-терминатора по прямой и обратной цепи ДНК. Подробно обсуждаются биологически значимые выводы из полученных результатов, в частности, в отношении т.н. MELAS-мутации в митохондриях человека, приводящей к наследственным заболеваниям. Важно подчеркнуть, что многие предсказания, полученные из модели, могут быть проверены экспериментально (что конечно не входило в задачи представленной работы).

Перечисленные результаты составляют первую главу диссертации. Данная модель и результаты ее применения опубликованы в двух статьях в авторитетном журнале *Biology Direct*.

Во второй главе О.Зверковым описан оригинальный алгоритм кластеризации белков из разных геномов. Алгоритм подробно и хорошо описан. На мой взгляд, в нем удачно выделяются ситуации присутствия в геноме групп паралогов, дублицировавшихся в одном организме после расхождения сравниваемых видов (ортологичность “многие – ко многим”). Соискатель не ставил своей целью сравнение этого алгоритма с другими существующими алгоритмами кластеризации белков. Такое сравнение было бы интересным, но, возможно, отвлекло бы от изучения белков, закодированных в пластидах. Зверковым выполнена кластеризация белков, закодированных в пластидах четырех крупных таксономических групп растений и простейших. Этот результат является новым, из него соискатель делает значимые биологические выводы. Полученная классификация представлена в публично доступной базе данных. Эта база адекватно отображает кластеры, но имеет, пожалуй, слишком ограниченный набор

возможностей. Так, например, не представлены множественные выравнивания последовательностей из одного кластера, что является стандартной для биоинформатики характеристикой семейства белков.

В третьей короткой главе О.Зверков дополняет гипотезу о возможном участии консервативных сайтов, найденных в 5'-нетранслируемой области шести мРНК хлоропластных генов в задержке инициации трансляции до завершения сплайсинга этих мРНК. Эта гипотеза была выдвинута в работе Селиверстова и Любецкого, 2006 г. Усовершенствовав алгоритм, Зверков обнаружил те же консервативные сайты в хлоропластах еще трех растений. Ссылка на работу предшественников есть.

Кроме того, в хлоропластных геномах им обнаружены неконсервативные шпильки перед двумя другими генами, мРНК которых содержит интроны. Соискатель связывает эти шпильки с задержкой инициации трансляции. Однако в диссертации эта гипотеза не подтверждена дополнительными данными, тем самым, является плохо обоснованной.

Повторю вопросы по существу, которые возникли у меня после прочтения диссертации.

1) Возможна ли была верификация решения обратной задачи модели методом скользящего контроля (исключения данных о транскрипции одного гена, определение параметров модели, сравнения уровня транскрипции исключенного гена в модели с экспериментом)?

2) В работе рассчитаны интенсивности связывания промотора LSP1 РНК-полимеразой (и сайта связывания фактора mTERF) в икре лягушки в зависимости от времени с момента оплодотворения. Данные об уровне мРНК митохондриальных генов для трех лягушек взяты из работы Ammini, Hauswirth, 1999. Как можно прокомментировать сравнение результатов моделирования, основанных на трех данных одинаковых экспериментов?

3) Есть ли дополнительные данные, подтверждающие гипотезу о задержке инициации трансляции генов *accD* и *atpH*, кроме предсказания шпилек, продемонстрированных в тексте диссертации (глава 3)?

Кроме того, у меня много замечаний по структуре диссертации и по тексту. Некоторые из них приведены ниже.

Раздел 3.13. является последним во введении. Судя по его названию ("Заключение") в нем читатель ожидает подытоживание сказанного во введении. Почему же в нем описывается, что будет в следующей главе 1? (Заметим, не в главах 2 и 3.)

Названия разделов не всегда правильно отражают их содержание, что крайне неудобно для читателя. Так, например, раздел 4 главы 1 называется "Экспериментальные данные об уровнях транскрипции генов и временах

полураспада”. Уже первая фраза в разделе: “Решением называется набор неизвестных параметров в модели” - казалось бы, не о том. И действительно, в разделе приведено больше информации о результатах, полученных при моделировании, чем об экспериментальных данных из литературы. Так, в большой Табл. 1.2 приведены исключительно результаты моделирования. Экспериментальные данные также приведены в некоторых таблицах и сравнены с результатами моделирования. Однако раздел, описывающий только данные экспериментов, с критическим обсуждением этих данных был бы крайне полезен.

Важные для модели экспериментальные данные, взятые из публикаций, приведены частично в разделе “Введение”, частично - в разделе “Результаты”. Фраза “Другие опытные данные о митохондриях приведены в пункте 3 введения” (глава 1, стр. 33) отсылает к разделу, занимающему 10 страниц. В некоторых таблицах (например, Табл.1.1) отсутствуют ссылки на публикации, из которой взяты данные; необходимые ссылки приведены где-то в других разделах текста. В итоге, при чтении разделов о применении модели для конкретного локуса перед читателем возникает нетривиальная задача извлечь из текста такую важную информацию, как число параметров в модели и число использованных экспериментальных данных.

Временами Зверков излишне категоричен в утверждениях. Так, на стр. 28 утверждается “Скорости элонгации РЕР [РНК-полимераз бактериального типа] при разных температурах соответствуют скоростям РНК-полимеразы *E. coli*, так как соответствующие субъединицы этих полимераз – близкие гомологи [36].” В работе [36] ничего не говорится про скорости элонгации. Вероятно, соответствие скоростей - предположение О.Зверкова, позволяющее, за неимением других данных, хоть как-то оценить скорость элонгации.

В разделе 3.3 на стр. 30 автором дается оценка сверху скорости элонгации РНК-полимеразы фагового типа, исходя из длин первого кодирующего экзона и следующего за ним интрона в некоторых генах и скорости перемещения рибосомы. Поскольку про скорость рибосомы автор забыл написать, то результат расчетов повисает в воздухе. Не говоря уже о том, что последняя фраза 1го абзаца прямо противоречит дальнейшим рассуждениям.

В главе 2 при описании алгоритма кластеризации одно и то же обозначение s_0 используется для двух разных величин, что может привести к неправильному пониманию.

Приведенные недостатки текста диссертации требуют от читателя дополнительных усилий для понимания работы. Иногда удобнее обратиться к публикациям О.Зверкова и статьям, на которые он ссылается.

В заключение отмечу, что модель, разработанная соискателем, является первой и удачной попыткой моделирования в целом процесса транскрипции митохондрий и пластид. При всей ограниченности экспериментальных данных об этом процессе, вероятных неточностях в экспериментально измеренных величинах и возможной зависимости скоростей транскрипции и уровней мРНК от неизвестных пока факторов, примеры применения модели показывают потенциальную возможность получения разумных предсказаний уже при таком достаточно грубом описании процессов. В этом большая заслуга О.А.Зверкова. Алгоритм кластеризации последовательностей белков, предложенный О.А. Зверковым является новым и, по-видимому, перспективным для использования не только для пластидных белков. Гипотезы о задержке инициации трансляции ряда генов указывают на возможные объекты для экспериментального изучения данного феномена.

Недостатки работы, описанные выше, не влияют на мою положительную оценку основных результатов диссертации.

Таким образом, данная диссертационная работа является законченной научно-квалификационной работой, выполненной О.А. Зверковым. Основные результаты диссертации опубликованы в российских и международных журналах. Автореферат точно и полно отражает содержание диссертации.

Содержание представленной диссертации соответствует указанной специальности. Работа Зверкова Олега Анатольевича «Функции и эволюция РНК-полимераз в митохондриях и пластидах», представленная на соискание ученой степени кандидата физико-математических наук по специальности «математическая биология, биоинформатика» – 03.01.09, соответствует требованиям, предъявляемым к кандидатским диссертациям и п.9 «Положения о порядке присуждения ученых степеней», а сам диссертант достоин присуждения ученой степени кандидата физико-математических наук по специальности 03.01.09 – математическая биология, биоинформатика.

2 сентября 2014 г.

А.В. Алексеевский
к.ф.-м.н.,
в.н.с. НИИ ФХБ имени А.Н.Белозерского
МГУ

02.09.2014

Подпись А.В.Алексеевского подтверждаю.
Директор НИИ ФХБ имени А.Н.Белозерского
МГУ имени М.В.Ломоносова
Академик, профессор
В.П.Скулачев

