

## ОТЗЫВ

официального оппонента на диссертационную работу Соколова Владимира Сергеевича «КОМПЬЮТЕРНОЕ ИССЛЕДОВАНИЕ КОНТЕКСТНЫХ ХАРАКТЕРИСТИК ОТКРЫТЫХ РАМОК СЧИТЫВАНИЯ, СВЯЗАННЫХ С ЭФФЕКТИВНОСТЬЮ ЭЛОНГАЦИИ ТРАНСЛЯЦИИ, У ОДНОКЛЕТОЧНЫХ ОРГАНИЗМОВ», представленную на соискание ученой степени кандидата биологических наук по специальности 03.01.09 – математическая биология, биоинформатика

Предметом исследования представленной диссертационной работы является трансляция – один из фундаментальных процессов регуляции экспрессии генов. Показано, что контроль экспрессии генов на уровне трансляции осуществляется на различных стадиях: инициации, элонгации и терминации. Каждая из этих стадий контролируется определенным набором факторов. Взаимодействие и согласованное привлечение различных факторов обеспечивает координацию отдельных этапов в единый процесс и точный контроль активности генов. Несмотря на то, что в настоящее время накоплено достаточно много информации в области регуляции экспрессии генов, до сих пор не все участники этих процессов описаны, а также еще недостаточно изучены механизмы их действия и координации на уровне транскрипции и трансляции. Поэтому изучение различных характеристик мРНК, влияющих на эффективность трансляции остается одной из актуальных задач современной молекулярной биологии и биоинформатики.

В связи с этим диссертационная работа Соколова Владимира Сергеевича является актуальной и посвящена исследованию методами биоинформатики контекстных характеристик открытых рамок считывания, определяющих эффективность элонгации трансляции и оптимизацию нуклеотидных последовательностей в процессе эволюции. Это определило цель и задачи данной диссертационной работы, которые сформулированы корректно и соответствуют теме диссертационной работы.

В качестве целевых (модельных) объектов данного исследования автор выбрал одноклеточные организмы, поскольку все изменения, происходящие в геноме

одноклеточных, напрямую влияют на их приспособленность в силу чего эволюция их нуклеотидных последовательностей идет быстрее, чем у многоклеточных.

Диссертационная работа В.С. Соколова изложена на 163 страницах и состоит из введения, трёх глав основной части, заключения, выводов, одного приложения и списка литературы, включающего 196 источников.

Глава 1 содержит обзор литературы по теме диссертации, который посвящен анализу механизмов регуляции трансляции. Рассмотрены контекстные особенности открытых рамок считывания, связанные с эффективностью элонгации трансляции, в том числе: (а) влияние кодонного состава на эффективность элонгации трансляции; (б) молекулярные механизмы и математические модели, объясняющие неравномерность использования кодонов; (в) взаимосвязь функции гена с адаптацией его кодонного состава к пулу тРНК; (г) влияние скорости трансляции на точность синтеза и фолдинга белка; (д) влияние вторичных структур в мРНК на эффективность элонгации трансляции в иницирующей, кодирующей и терминирующей частях мРНК. Заключительный раздел обзора литературы посвящен описанию биоинформатических методов оценки эффективности экспрессии генов, в том числе индексам и программам анализа кодонного состава и вторичных структур нуклеотидных последовательностей.

В главе 2 изложены методы и алгоритмы. Приводится детальное обоснование преимущества разработанного в Институте цитологии и генетики СОРАН В.А. Лихошваем и Ю.Г. Матушкиным индекса EEI (Elongation Efficiency Index), который выбран для оценки эффективности элонгации трансляции. В отличие от большинства других индексов этот индекс позволяет рассчитывать не только кодонный состав генов, но и количество, и стабильность вторичных структур (локальных инвертированных повторов) в мРНК. В зависимости от комбинации этих показателей рассмотрены пять типов индекса EEI (EEI1-EEI5), которые позволяют оценить, какой из этих индексов (или показателей) оказывает основное влияние на эффективность трансляции. Для проведения исследований при непосредственном участии автора диссертационной работы было создано специальное программное обеспечение EloE (Elongation Efficiency), реализующее алгоритм расчета индекса EEI, на которое получено авторское свидетельство. Данная программа в виде веб-приложения доступна в сети Интернет (<http://www.bionet.sccc.ru:7780/EloE/>).



В третьей главе изложены результаты диссертационной работы и их обсуждение. С использованием разработанной программы EIoE проведен анализ 2771 геномов одноклеточных организмов, в том числе 2582 геномов бактерий, 165 – архей и 24 – одноклеточных эукариот. Для каждой группы организмов были рассчитаны их распределения по пяти индексам EEI1-EEI5 и определено, какой из пяти индексов лучше всего оценивает эффективность элонгации в каждой конкретной группе. Для этого были использованы гены рибосомных белков как маркеры высокоэкспрессирующихся генов. Для этих генов были рассчитаны ранг (M1-M5) и стандартное отклонение (R1-R5) и показана высокая корреляция индекса EEI с уровнем экспрессии этих генов.

Организмы, у которых ни один из индексов не позволил достоверно определить эффективность элонгации трансляции, были выделены в отдельную группу – группу «особых» организмов.

Для выяснения причин появления «особых» видов автор провел дополнительное исследование распределения геномов 62 особых штаммов *Mycoplasma* по индексам EEI и показал, что эволюционный отбор по эффективности трансляции генов у данных организмов шел на минимизацию количества потенциальных вторичных структур (совершенных локальных инвертированных повторов) в мРНК, а не на оптимизацию кодонного состава. На основе анализа профилей индекса локальной комплементарности (LCI) автор делает заключение, что сниженная стабильность потенциальных шпилек в 5'-районе, вероятнее всего, способствует сборке рибосомного комплекса и началу трансляции. Напротив повышенная стабильность потенциальных шпилек в 3'-районе может отвечать за терминацию трансляции или, возможно, транскрипции.

Анализ геномов Архей проведен в контексте исследования оптимизации первичной структуры их генов в процессе эволюции. Автору удалось показать, что при изменении условий обитания дрожжей наблюдаются наиболее значительные перестройки в первичной структуре их генов, что свидетельствует о приспособливании к новым условиям. В результате этого процесса изменяется профиль эффективности элонгации трансляции генов, что, вероятно, говорит об изменении профиля экспрессии генов. В новых условиях организм начинает более

эффективно экспрессировать гены, которые в прежних условиях обитания экспрессировались на низком уровне и наоборот.

Для двух видов дрожжей *Saccharomyces cerevisiae* и *Schizosaccharomyces Pombe* проанализирована связь между плотностью нуклеосомной упаковки и эффективностью элонгации генов. Показано, что отбор может идти как на уменьшение потенциала формирования нуклеосом (ПФН) для высокоэкспрессирующихся генов, так и на увеличение ПФН для низкоэкспрессирующихся. Для генов *S. pombe* с высоким уровнем трансляции наблюдаются районы с достоверной отрицательной корреляцией индекса эффективности элонгации трансляции и потенциала формирования нуклеосом как выше старта транскрипции, так и в кодирующей части генов. Тогда как для генов *S. cerevisiae* с низким уровнем трансляции наблюдаются районы с достоверной положительной корреляцией этих показателей. Полученные данные позволили автору предположить, что первичная структура низкоэкспрессирующихся генов у *S. cerevisiae* оптимизирована таким образом, чтобы при замедлении процесса транскрипции, вызванного образованием нуклеосомного комплекса, ускорялся процесс элонгации трансляции для компенсации данной задержки. Другими словами полученные данные подтверждают предложенную гипотезу о согласованности оптимизации процессов транскрипции и трансляции.

Замечания и вопросы, возникшие при оппонировании диссертации:

- 1) Список цитируемой литературы включает 196 источников, из них 80 статей (40 %) опубликованы до 2000 года (1959-2000 гг.) и только 23 работы (11 %) – за последние пять лет (2011-2015 гг.).
- 2) Для организмов, у которых значение параметра  $M < 30$ , используются разные термины – «особенные» и «особые» организмы.
- 3) На стр. 69 отмечается, что изучение особенностей трансляции у *Mycoplasma* может иметь практическое значение в медицине, т.к. некоторые виды микоплазм являются паразитами человека и домашних животных. Непонятно, как особенности трансляции у *Mycoplasma* связаны с их практическим использованием в медицине и ветеринарии. Здесь лучше говорить о связи особенностей трансляции с патогенезом *Mycoplasma*.



- 4) На стр. 83 отмечается, что среднее значение коэффициентов корреляции между GC-составом и EEI для микоплазм из 2-ой группы равно 0,44, а из 4-ой – 0,83. Из рис. 40 следует, что должно быть наоборот.
- 5) Из текста неясно, почему данные по коэффициентам корреляции между EEI и ПФН, представленные на рис. 55-60, приведены только для 15% последовательностей высокоэкспрессирующихся и низкоэкспрессирующихся генов. Чем обусловлен выбор порогового значения 15 %?
- 6) Из рис. 63-64 не ясно, почему автор делает заключение о том, что для  $\geq 90\%$  генов кодирующая часть находится в интервале  $[0; +400]$  нуклеотидов.

Высказанные замечания не снижают достоинств рецензируемой работы. Квалификация автора как специалиста в области математической биологии и биоинформатики, как и ценность представленных им результатов, не вызывает никаких сомнений.

Полученные автором диссертации результаты имеют теоретическое значение, так как позволяют понять закономерности оптимизации структуры геномов одноклеточных организмов в процессе их эволюции. В частности, филогенетическое исследование *Mycoplasma* позволило установить возможную связь оптимизации первичной структуры генов данных организмов с их средой обитания в процессе эволюции. Кроме того, анализ нуклеотидных последовательностей генов и их предковых форм у архей позволил установить, что наиболее сильные изменения в первичной структуре генов, связанные с оптимизацией элонгации трансляции, происходили при радикальной смене среды обитания данных организмов.

Созданное при выполнении диссертационной работы программное обеспечение EloE имеет практическое значение, так как может быть использовано для проектирования искусственных генно-инженерных конструкций – эффективных продуцентов биологически-активных веществ.

В целом, диссертационная работа Соколова В.С. является законченной научной квалификационной работой, новизна и значимость которой не вызывают сомнения. Полученные результаты являются новыми, выводы обоснованными. Содержание диссертации должным образом отражено в автореферате и в опубликованных работах. Материал диссертации соответствует указанной специальности.

## Заключение

Диссертация Соколова В.С. «Компьютерное исследование контекстных характеристик открытых рамок считывания, связанных с эффективностью элонгации трансляции, у одноклеточных организмов» удовлетворяет всем требованиям ВАК РФ, предъявляемым к диссертациям, и отвечает критериям п. 9 «Положения о порядке присуждения ученых степеней (Постановление от 24 сентября 2013 г. № 842), а ее автор, Соколов В.С. заслуживает присуждения ему искомой степени кандидата биологических наук по специальности 03.01.09 – математическая биология, биоинформатика.

Заведующий теоретическим отделом

ФБУН ГНЦ ВБ "Вектор"

д.б.н.



С.И.Бажан

Подпись заведующего отделом, д.б.н. С.И. Бажана заверяю

Ученый секретарь ФБУН ГНЦ ВБ «Вектор»

к.б.н., доцент



О.А. Плясунова



20 ноября 2015 г.