

УТВЕРЖДАЮ

Директор ИОГен РАН



член-корреспондент РАН

Н.К. Янковский

20 ноября 2015 г.

О Т З Ы В

Ведущей организации 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».

Актуальность. Диссертация Соколова Владимира Сергеевича посвящена анализу последовательностей матричной РНК одноклеточных организмов с целью определения факторов, влияющих на эффективность элонгации трансляции. Тема эта весьма актуальна и имеет большую практическую значимость, поскольку последовательности мРНК активно экспрессирующихся генов адаптированы под свойства организма-хозяина, из которых наиболее важны видоспецифические пулы аминок-ацил тРНК. Таким образом, при создании штаммов-продуцентов и гетерологической экспрессии в этих штаммах генов различных целевых белков необходима адаптация этих целевых генов под клеточное окружение нового хозяина. Понимание, какие

из свойств последовательностей мРНК существенны для эффективной трансляции, должно способствовать созданию синтетических генов с эффективной экспрессией в конкретном организме-хозяине. Кроме того, разработанные методы могут способствовать изучению самих организмов — источников генов. Действительно, получение последовательности генома одноклеточного организма в настоящее время является рутинной работой, а оценка фоновой экспрессии каждого конкретного гена позволяет дать грубое понятие о характерном метаболизме организма и до некоторой степени его биологии.

Несмотря на то, что оценки эффективности трансляции конкретного гена в конкретном организме-хозяине практикуются уже более 30 лет и вошли в круг стандартных методов синтетической биологии, существовал ряд неясных вопросов. В частности, были отмечены виды, в которых эффективность трансляции мРНК очень мало зависит от статистики используемых кодонов. Диссертация Соколова Владимира Сергеевича окончательно проясняет этот вопрос, ясно показывая, что у некоторых видов основным фактором, влияющим на эффективность трансляции, является не статистика используемых кодонов, а структура шпилек в мРНК, которые могут препятствовать движению рибосомы. Диссертант провел исчерпывающую классификацию различных видов бактерий и показал, в каких видах какой фактор является определяющим для эффективности трансляции, а в каких видах важна комбинация этих факторов. Таким образом, диссертация восполняет важный пробел в области важной для развития синтетической биологии, в чем и заключается ее безусловная актуальность.

Структура работы. Работа построена по классической схеме и состоит из трех глав: обзора литературы, методической главы, изложения результатов. Обзор литературы дает прекрасное введение в биологию элонгации трансляции и включает данные, полученные в самое последнее время с

использованием рибосомного профайлинга и параллельного анализа вторичных структур РНК. Кроме того, в обзоре литературы подробно излагается, какие именно характеристики мРНК, имеющие значение для эффективности элонгации трансляции, могут быть изучены методами биоинформатики. Очень благоприятное впечатление производит оригинальное изложение истории развития исследований в предметной области. Глава «Материалы и методы» содержит изложение процедуры формирования выборок последовательностей, что представляет собой важнейшую часть любого исследования в области биоинформатики. Также в этой главе дается подробное описание использованных критериев оценки эффективности элонгации. Диссертант вводит пять так называемых индексов, позволяющих как выделить основной фактор, влияющий на интенсивность элонгации (статистика использования кодонов или вторичная структура РНК), так и выделить гены с предположительно высокой экспрессией. К достоинствам работы относится то, что автор использует самообучающийся алгоритм, который строит положительную выборку генов с высокой экспрессией параллельно с построением меры эффективности элонгации. Алгоритм использует итеративную схему и в качестве затравки базируется на выборке рибосомальных белков, которые высоко экспрессируются в большинстве видов микроорганизмов.

В главе «Результаты» продемонстрированы данные, подтверждающие положения, выносимые на защиту. Очень хорошее впечатление производит иллюстративный материал: диссертация содержит более 50 рисунков и схем, ясно иллюстрирующих основные результаты работы. Специального раздела с обсуждением результатов не вводится, и результаты обсуждаются в нескольких кратких сводках по мере изложения.

Научная новизна. Диссертантом впервые был проведен систематический сравнительный анализ вкладов статистики использования синонимических кодонов в последовательности мРНК и формирования

вторичной структуры мРНК в эффективность элонгации трансляции. Проанализировано большое число видов бактерий и архей, а также дрожжей *Saccharomyces*. Построены индексы, позволяющие учитывать вклады этих двух факторов в определение скорости элонгации трансляции. На основании этих индексов построен метод выделения подгруппы генов с высокой экспрессией. Показано, что у бактерий основной фактор, определяющий скорость элонгации — это статистика использования кодонов, в то время как у архей большую роль играют инвертированные повторы в мРНК и таким образом ее вторичная структура.

Кроме того, автор проанализировал степень оптимизации последовательностей мРНК высокоэкспрессируемых генов в геномах с различным GC составом и показал, что по крайней мере у архей рост GC состава генома соответствует меньшей степени оптимизации последовательностей высокоэкспрессируемых генов.

Научная значимость результатов и их практическое значение.

Научная значимость результатов и практическое значение доказываются тем, что автором открыты новые явления в области строения участков мРНК, ответственных за элонгацию трансляции. С точки зрения фундаментальной науки, открытые диссертантом закономерности должны послужить дополнительным фактором для изучения биологии микроорганизмов, для которых недостаточно экспериментальных данных, но известна последовательность генома. Предсказание набора генов с высокой экспрессией может дать грубое понимание биологии изучаемого микроорганизма. С научно-практической точки зрения, открытые автором закономерности могут использоваться для повышения эффективности трансляции мРНК конкретных генов путем создания синтетических конструкторов, оптимально экспрессирующих гены, кодирующие конкретные белки в конкретных организмах — продуцентах. Существующие программы оптимизации синтетических последовательностей генов используют в

основном статистику использования кодонов или дикодонов и не используют потенциал формирования вторичных структур РНК. Диссертантом разработана программа EloE, доступная по сети Интернет, осуществляющая высокопроизводительный анализ эффективности элонгации генов в различных организмах.

Замечания к работе. Работа очень хорошо написана, и содержит сравнительно мало недостатков. По существу, я не согласен с единственным умозаключением диссертанта, достаточно категорично повторенным в нескольких местах работы. Например, на стр. 66 диссертант утверждает, что то, что среди архей нет организмов, принадлежащих к 3му и 5му типам индекса EEI (в этих типах индекса учитывается энергия шпилек мРНК, а не просто их количество), свидетельствует, что археи лишены энергозависимого механизма расплетания вторичных структур в мРНК. Я не могу согласиться с этим утверждением по крайней мере на материалах, доступных в тексте диссертации. При отнесении генома к конкретному типу индекса диссертант всегда брал тип индекса, вносящий максимальный вклад. Предположим, корреляция числа шпилек с интенсивностью экспрессии для некоторого генома всегда больше, чем корреляция суммарной оцененной энергии этих шпилек с той же интенсивностью экспрессии. Отсюда вовсе не следует, что корреляция суммарной оцененной энергии с интенсивностью экспрессии обязательно нулевая, она может быть также значимой, но меньшей по своей величине. Известно, что геномы многих архей АТ богаты, поэтому, для таких видов стандартная оценка энергии шпилек, основанная, в частности, на числе CG пар, может давать небольшие величины, и простой подсчет числа шпилек окажется лучше скоррелирован с эффективностью элонгации гена.

Есть некоторое количество мелких описок и опечаток, в том числе и в формулах, в частности в формуле (5) на стр. 37 у x нижние индексы должны быть j и k .

Сделанные замечания не умаляют общего высокого уровня работы.

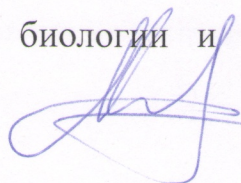
Рекомендации по практическому использованию. Результаты диссертации могут использоваться при проведении научных исследований в области биотехнологии и генетической биоинформатики. Описанные в диссертации методы и статистические закономерности могут быть использованы в Федеральном государственном бюджетном учреждении науки Институт общей генетики им. Н.И. Вавилова РАН, Федеральном государственном бюджетном учреждении науки Институт молекулярной генетики РАН, Федеральном государственном бюджетном учреждении науки Институт математических проблем биологии РАН, Федеральном государственном бюджетном учреждении науки Институт биологии гена РАН, Федеральном государственном бюджетном учреждении науки Институт проблем передачи информации им. А.А. Харкевича РАН, Московском государственном университете имени М.В. Ломоносова, Новосибирском государственном университете и других учебных и научно-исследовательских организациях.

Заключение. Диссертационная работа Соколова Владимира Сергеевича "Компьютерное исследование контекстных характеристик открытых рамок считывания, связанных с эффективностью элонгации трансляции, у одноклеточных организмов" является законченным научно-квалификационным исследованием, в котором решена задача анализа определения закономерностей в последовательностях мРНК различных биологических видов микроорганизмов, определяющих эффективность элонгации трансляции. Решение этой задачи имеет большое значение для задач моделирования живых систем, а именно субклеточных структур и клеток, математического моделирования эволюционных процессов в живой природе, а также задач компьютерной геномики и протеомики (пункты 1-3 паспорта специальности 03.01.09). По совокупности полученных результатов работа Соколова Владимира Сергеевича соответствует п.9 «Положения о присуждении ученых степеней», утвержденного Постановлением

Правительства РФ № 842 от 24 сентября 2013 года. Результаты достоверны и достаточно полно опубликованы, автореферат адекватно отражает содержание диссертации. Автор заслуживает присуждения ученой степени кандидата биологических наук по специальности 03.01.09 – Математическая биология, биоинформатика. Материалы диссертации были рассмотрены и обсуждены на Межлабораторном семинаре ИОГен РАН 17 ноября 2015 года (протокол № 1).

20 ноября 2015.

Зав. лаборатории системной биологии и вычислительной генетики
ИОГен РАН



д.ф.-м.н. Макеев В.Ю.

Подпись заведующего лабораторией системной биологии и вычислительной генетики ИОГен РАН, доктора физико-математических наук Макеева Всеволода Юрьевича удостоверяю

Ученый секретарь ИОГен РАН

д.б.н. Огаркова О.А.

