На правах рукописи

Зверков Олег Анатольевич

Функции и эволюция РНК-полимераз в митохондриях и пластидах

03.01.09 – Математическая биология, биоинформатика

Автореферат диссертации на соискание ученой степени кандидата физико-математических наук

Москва – 2014

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте проблем передачи информации им. А. А. Харкевича Российской академии наук (ИППИ РАН)

Научный руководитель – д.ф.-м.н. проф. Любецкий Василий Александрович.

Официальные оппоненты:

Туманян Владимир Гаевич, д.ф.-м.н., проф., Федеральное государственное бюджетное учреждение науки Институт молекулярной биологии им. В. А. Энгельгардта Российской академии наук, заведующий лабораторией;

Алексеевский Андрей Владимирович, к.ф.-м.н., Научно-исследовательский институт физико-химической биологии им. А. Н. Белозерского Московского государственного университета им. М. В. Ломоносова, ведущий научный сотрудник, и. о. заведующего отделом.

Ведущая организация: Федеральное государственное бюджетное учреждение науки Институт общей генетики им. Н. И. Вавилова Российской академии наук.

Защита состоится 18 сентября 2014 года в 16:00
на заседании диссертационного совета Д 002.077.04 на базе ИППИ РАН
Большой Каретный пер., д. 19, стр. 1, Москва, ГСП-4, 127994

С диссертацией можно ознакомиться в библиотеке ИППИ РАН и на сайте www.iitp.ru

Автореферат разослан 17 августа 2014 года

Ученый секретарь диссертационного совета,
д. б. н. профессор Рожкова Г. И.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. В биоинформатике велико значение быстрых и эффективных алгоритмов, поскольку зачастую возникают входные данные весьма большого объёма. Известные и новые методы вычислений требуют адаптации к работе на многопроцессорных вычислительных комплексах (суперкомпьютерах), которые стали в последнее время значительно доступнее.

К настоящему времени известны сотни полностью секвенированных геномов пластид, тысячи геномов митохондрий, скорость пополнения баз данных геномной информации растёт экспоненциальными темпами. Возникает такой объём информации, что доля геномов, доступных биохимическому исследованию, становится всё меньше. Поэтому возникает потребность в эффективных и быстрых алгоритмах компьютерного анализа данных, а также в создании специализированных баз данных. Существенно, чтобы алгоритмы опирались на «точные модели», т.е. было доказано, что они приводят к глобальным экстремумам соответствующих функционалов, имели низкую вычислительную сложность (полином 2–3 степени) и допускали эффективное распараллеливание.

Моделирование клеточных процессов требует нетривиальных алгоритмов и является важным инструментом биоинформатического исследования. Оно позволяет предсказать значения параметров биохимических процессов (например, инициации, элонгации и терминации транскрипции), которые трудно измерить непосредственно, а также – решить нетривиальную обратную задачу: выбрать значения параметров, которые соответствуют экспериментальным зависимостям.

Экспериментальные исследования, в том числе проведённые в Институте физиологии растений им. К. А. Тимирязева РАН (Зубо и др.), позволили предположить важную роль взаимодействия РНК-полимераз в процессе транскрипции пластомов растений и в ответе пластид на тепловой шок. Для проверки этого предположения и предсказания параметров, не определяемых в экспериментах, была поставлена задача моделирования процесса транскрипции в пластидах с одновременным участием многих РНК-полимераз, факторов и вторичных структур, взаимодействующих друг с другом. Затем задача была расширена на моделирование транскрипции в митохондриях.

Использование кластера MVS-100K в Межведомственном суперкомпьютерном центре РАН позволило впервые провести моделирование транскрипции для всей кольцевой ДНК митохондрий человека, крысы и лягушки, а также для существенных локусов пластид.

Построение близких по последовательности и минимальных по содержанию паралогов белковых семейств (кластеризация белков) позволяет уточнять аннотации белков, судить о работоспособности белковых комплексов, например РНК-полимераз бактериального типа. (В случае отсутствия последних транскрипция выполняется РНК-полимеразами фагового типа, что придаёт этому процессу другие черты.) Известно несколько баз данных ортологичных семейств белков. Однако большинство из них содержат небольшое *число* видов с пластидами или вовсе не содержат их. Например, (по состоянию на 1 июля 2013) OrthoDB не содержит растений и простейших, OrthoMCL включает только 11 водорослей и 14 споровиков; GeneDB – только 7 споровиков; в RoundUp и InParanoid таких видов ещё меньше; OMA и EggNOG почти не содержат видов с пластидами; в COG и KOG представлено два растения и ни одного споровика. Поэтому была поставлена задача: предложить эффективный алгоритм кластеризации белков и получить базы данных пластомных белков.

Изучение пластид споровиков (апикопластов) значимо, поскольку споровики вызывают опасные заболевания человека и животных, в том числе токсоплазмоз и малярию. Исследование регуляции экспрессии генов, кодируемых в апикопластах, важно для понимания роли апикопластов в передаче инфекции, а также в механизмах действия лекарственных средств на апикопласты, которые являются главной мишенью антибиотиков, не оказывающих прямого воздействия на экспрессию ядерных и митохондриальных генов хозяина. В частности, *Theileria* и *Babesia* переносятся иксодовыми клещами и вызывают заболевания крупного рогатого скота: *B. bigemina* и *B. bovis* – бабезиоз крупного рогатого скота, *Th. annulata* – тейлериоз крупного рогатого скота, *Th. parva* – лихорадку Восточного Берега; *Eimeria tenella* вызывает эймериоз кур; *Toxoplasma gondii* – токсоплазмоз, в том числе у человека; различные виды рода *Plasmodium* вызывают малярию у людей (*P. falciparum*, *P. vivax*) и других животных. Некоторые споровики, например *Cryptosporidium parvum*, не имеют пластид.

Исследование митохондрий человека, крысы и лягушки значимо для понимания молекулярных механизмов MELAS болезней человека (митохондриальная энцефаломиопатия, лактатацидоз, инсультоподобные эпизоды), болезней, связанных с недостаточностью гормона щитовидной железы, и т.д.

Цели работы

1. Разработать модель взаимодействия и конкуренции РНК-полимераз в митохондриях и пластидах, которая должна предсказывать уровни транскрипции всех генов. На её основе объяснить изменения уровней транскрипции генов: в митохондриях человека с MELAS-мутацией; в митохондриях крысы с эпигенетическими нарушениями, вызванными недостатком тиреоидного гормона; в пластидах растений после нокаутов минорных σ-субъединиц или теплового шока.

2. Разработать алгоритм построения сходных по последовательности и минимальных по содержанию паралогов семейств белков (кластеризации данного множества белков). Применить алгоритм к множествам белков, кодируемых в пластидах родофитной и хлорофитной ветвей и цветковых растений. На основе полученных семейств: рассмотреть вопрос о присутствии полноценной РНК-полимеразы бактериального типа у споровиков; указать белки, характерные для узких таксономических групп («филогенетические подписи»).

3. Предсказать белковые сайты и вторичные структуры мРНК, ответственные за задержку инициации трансляции до завершения процессинга мРНК в пластидах.

Методы исследования. В работе использованы методы теорий алгоритмов и массового обслуживания, методы моделирования и организации вычислительных экспериментов с использованием известных и оригинальных программ, в том числе для параллельных вычислений на суперкомпьютерах, методы математической биологии и биоинформатики.

Научная новизна. Моделирование взаимодействия РНК-полимераз, по крайней мере на длинных локусах ДНК, ранее не выполнялось. Моделирование основано на новом математическом и алгоритмическом подходе к изучению большой системы одновременно взаимодействующих объектов. Кластеризация получена на основе оригинального алгоритма в теории графов. Все полученные алгоритмы имеют низкую оценку вычислительной сложности, а биоинформатические результаты являются новыми.

Практическая значимость работы. Работа носит теоретический характер. В то же время, исследование может иметь прикладное значение.

Предложенные алгоритмы и их программные реализации могут применяться для исследования широкого класса задач. А именно, в медицинских исследованиях могут быть полезны разработанные методы количественной оценки влияния мутаций и эпигенетических нарушений на уровни транскрипции генов в митохондриях, предложенные нами объяснения механизма MELAS-синдрома у человека и нарушения метилирования мтДНК у крысы с недостатком гормона щитовидной железы.

Для создания новых видов растений, в том числе с ксенопластидами, могут быть полезны предложенные механизмы отклика на тепловой шок изолированных пластид и на нокауты транскрипционных факторов в пластидах.

Апробация работы. Компьютерные программы тестировались на биологических данных с экспериментально известными ответами, а также в процессе решения биологических задач. Результаты работы опубликованы и докладывались на следующих конференциях:

- Международная конференция "Moscow Conference on Computational Molecular Biology": MCCMB'07 (Москва, 27–31 июля 2007), MCCMB'13 (Москва, 25–28 июля 2013);
- 32-я, 33-я, 35-я, 37-я конференция «Информационные технологии и системы»: ИТиС'09 (Бекасово, 15–18 декабря 2009), ИТиС'10, (Геленджик, 20–24 сентября 2010), ИТиС'12 (Петрозаводск, 19–25 августа 2012), ИТиС'13 (Калининград, 1–6 сентября 2013);
- 7-я международная конференция "Bioinformatics of Genome Regulation and Structure\Systems Biology" BGRS\SB'10 (Новосибирск, 20–27 июня 2010);
- 51-я, 53-я, 54-я научная конференция МФТИ (Москва, 28–30 ноября 2008, 24–29 ноября 2010, 25–26 ноября 2011);
- 3-я Московская международная конференция "Molecular Phylogenetics" (Москва, 31 июля – 4 августа 2012).
- 8-я Международная научно-практическая конференция «Современные информационные технологии и ИТ-образование» (Москва, МГУ им. М. В. Ломоносова, 8–10 ноября 2013).

Работа также докладывалась на научных семинарах механико-математического факультета Московского государственного университета им. М. В. Ломоносова и на семинаре по Математической биологии и биоинформатике Института проблем передачи информации им. А. А. Харкевича РАН.

Публикации. По теме диссертации опубликовано 9 статей и 13 тезисов докладов на конференциях (см. список в конце автореферата). Все результаты, включённые в диссертацию, получены лично автором.

Структура и объём работы. Работа состоит из введения, трёх глав и списка литературы. Список литературы содержит 127 наименований. Объём работы составляет 112 страниц, включая 21 таблицу и 29 рисунков.

СОДЕРЖАНИЕ РАБОТЫ

Введение содержит общие сведения и вспомогательный материал к главам. Пункты 1–2 введения содержат общую характеристику работы и списки основных результатов и публикаций. В пункте 3 введения приводятся сведения об РНК-полимеразах в митохондриях хордовых животных (лягушки, человека и крысы) и в пластидах растений. Для митохондрий описывается структура и взаимное расположение промоторов, влияние белковых факторов на уровни транскрипции генов, описывается mTERF-зависимая и белок-независимая регуляция транскрипции, MELAS-мутация (MELAS-синдром – митохондриальная энцефаломиопатия, лактатацидоз, инсультоподобные эпизоды), времена полураспада РНК и т.д. Приводятся аналогичные сведения о пластидах высших растений и водорослей. Вводится понятие конкуренции РНК-полимераз. Описываются опыты с нокаутом генов пластид и с тепловым шоком изолированных хлоропластов.

В главе 1 изучается взаимодействие РНК-полимераз в митохондриях и пластидах. Глава начинается с описания анализируемых в дальнейшем локусов митохондриальной ДНК хордовых животных и пластидной ДНК растений. Подробно описывается предложенная нами модель взаимодействия и конкуренции РНК-полимераз. Описываются параметры РНК-полимераз бактериального (PEP) и фагового (NEP) типов, PEP- и NEP-промоторов, абортивного процесса для PEP.

Приводятся экспериментальные данные об инициации и элонгации транскрипции, терминации РНК-полимераз, поляризации терминаторов, формировании/связывании факторов, участвующих в этих процессах, как белковых, так и вторичных структур РНК. Описываются методики моделирования и оценки согласия результатов моделирования с опытными данными. Подробно сравниваются результаты моделирования с опытными данными для митохондрий и пластид.

В модели описывается следующая ситуация. В транскрипции локуса ДНК одновременно участвует множество РНК-полимераз, которые связываются с промоторами своего типа и затем движутся вдоль своей цепи, возможно навстречу друг другу. Это приводит к взаимодействию РНК-полимераз как между собой, так и с различными белковыми и структурными факторами на ДНК и РНК.

Описание модели. Задан локус (последовательность в четырёхбуквенном алфавите), на котором размечены участки: промоторы, сайты связывания белкового репрессора, сайты формирования терминаторов элонгации, кодирующие участки и т.д. Для каждого промотора задаётся *интенсивность* λ попыток связывания РНК-полимеразы с этим промотором; если её значение не известно из экспериментов, то оно вычисляется в модели как решение обратной задачи. А именно, интервалы времени между попытками связывания описываются пуассоновским процессом с параметром λ. Попытка считается *успешной*, если в момент её совершения промотор не занят другой РНК-полимеразой или каким-то фактором: регуляторным белком, вторичной структурой и т.д. Здесь возникает трудная математическая задача, так как модель представляет собой не просто систему пуассоновских процессов, что само по себе нетривиально, а систему «условных» пуассоновских процессов. Каждое условие задаётся крайне нетривиально – расположением многих ранее связавшихся полимераз и факторов, которые находятся в процессах движения, связывания и формирования. А именно, полимеразы перемещаются по локусу, а факторы возникают и исчезают на нём, каждая/ый по своему закону.

Итак, каждому NEP-промотору и каждому PEP-промотору (причём последний берётся в паре с фиксированной *группой* σ-субъединиц) сопоставляется пуассоновский процесс со своим параметром λ. В работе используются следующие группы: все σ-субъединицы и все σ-субъединицы кроме одной, нокаутирус-

мой (фактически это минорные σ-субъединицы Sig3 и Sig4). В опытах, не связанных с нокаутом, рассматривается группа, состоящая из всех σ-субъединиц.

Таким образом, каждому NEP-промотору соответствует свой стохастический процесс, определяющий промежутки времени между попытками связывания NEP. Это время равно $-(\ln\xi)/\lambda_N$, где ξ – равномерно распределённая случайная величина, заданная на интервале от 0 до 1. Параметр λ_N – искомое значение для этого промотора. Аналогично определяются стохастические процессы для каждого из PEP-промоторов. Промежутки времени снова вычисляются как $-(\ln\xi)/\lambda$, где λ равно λ_P для PEP в паре с группой всех σ-субъединиц и λ равно λ_4 для PEP в паре с группой всех σ-субъединиц кроме нокаутируемой Sig4 (случай локуса 1 из этой главы). Аналогично рассматривается локус 3 из этой главы, связанный с опытами по нокауту Sig3 или Sig4, для него определяются λ_P и λ_3 (когда нокаутируется Sig3) или λ_P и λ_4 (когда нокаутируется Sig4). Параметры λ называются *интенсивностями связывания* соответствующих промоторов и измеряются в обратных секундах. Определив интенсивности связывания в диком типе, мы используем их без изменения при описании нокаутов по разным σ-субъединицам и при описании теплового шока в том же или даже в близком виде.

Если передние края двух разнонаправленных полимераз занимают одну и ту же позицию, то в модели принимается, что элонгация обеих прекращается. Если на одной цепи ДНК полимераза *X* передним краем вплотную примыкает к полимеразе *Y*, то *X* не может обогнать *Y*. Для моделирования процесса элонгации нужно задать значения параметров v_N и v_P – скоростей элонгации NEP и PEP соответственно. Эти скорости зависят от температуры, нуклеотидного состава ДНК и вторичных структур, которые образуются на РНК в процессе транскрипции. Результаты получены в предположении постоянной скорости РНК-полимеразы (при фиксированной температуре) и без учёта вторичных структур РНК, так что элонгация моделируется как детерминированный процесс.

Каждому белковому фактору транскрипции *F* соответствует аналогичный стохастический процесс с параметром λ_F, который определяет промежутки времени между попытками связывания фактора со своим сайтом на ДНК. Как и выше, такая попытка считается успешной, если в момент её совершения сайт свя-

зывания свободен от всех РНК-полимераз и любых факторов. Терминация транскрипции на белковом факторе происходит, как описано ниже на примере фактора mTERF. Наконец, каждому терминатору транскрипции (крест-шпильке на ДНК) соответствует бернуллиевская случайная величина с параметром p, описывающая терминацию транскрипции на каком-либо нуклеотиде плеча шпильки.

Если РЕР связалась с РЕР-промотором, то сначала моделируется *абортивный* процесс, а затем – упомянутый выше процесс *элонгации* полимеразы. Для абортивного процесса следующим образом определяются число абортивных попыток и длины каждой из абортивных РНК. Длительность t абортивного процесса задаётся формулой $t = -(\ln \xi) \cdot t_0$, где t_0 – среднее время абортивного процесса (например, $t_0 = 0.4$ с). Общее число абортивных попыток k определяется как наибольшее число слагаемых в левой части неравенства $-(\ln \xi_1 + ... + \ln \xi_i + ... + \ln \xi_k) \le (t \cdot v_P / r_0)$, при котором оно остаётся верным. Параметр r_0 – средняя длина одной абортивной РНК (например, $r_0 = 4$). При каждой i-й абортивной попытке появляется РНК, длина которой равна целому числу, ближайшему к числу $-r_0 \cdot (\ln \xi_i)$. Таким образом, величина $-(\ln \xi_i)$ имеет смысл случайной поправки к среднему времени r_0 / v_P, уходящему на одну абортивную попытку, где v_P – скорость РЕР.

Для моделирования опытов по изменению уровня транскрипции в результате теплового шока[1] (локус 2 из этой главы) в модель введены следующие параметры: в течение времени t_1 растение находится при температуре T_1; затем в течение времени t_2 у одной массы хлоропластов температура повышается до T_2, а у другой такой же массы она остаётся равной T_1; затем в течение времени t_3 у обоих масс температура меняется на новое значение T_3, и сразу после этого у ряда генов хлоропластов измеряется отношение числа завершённых транскрипций в материале после шока к таковому числу в контрольном материале. В опыте эти параметры имели следующие значения: t_1 = 6–7 суток, $T_1 = 21$°C, $t_2 = 1.5$ часа, $T_2 = 40$°C, $t_3 = 15$ минут, $T_3 = 25$°C.

[1] Зубо Я.О., Лысенко Е.А. *и др.* Изменение транскрипционной активности генов пластома ячменя в условиях теплового шока // *Физиология растений*. 2008. Т. 55. С. 323–331.

В случае митохондрий фактором является ещё G-квадруплекс[2], который вовлекает короткие участки РНК, а также особую роль играет регуляторный белок mTERF[3]. Крест-шпильки на ДНК, характерные для пластид и бактерий[4], отсутствуют в рассмотренных митохондриях. В модели терминация транскрипции при столкновении РНК-полимеразы с белковым фактором mTERF описывается следующим образом. Как и выше, попытка связывания mTERF со своим сайтом считается успешной, если в момент её совершения сайт свободен от полимераз, ранее связавшихся молекул этого белка и других факторов. Если mTERF связался с сайтом и к нему приходит РНК-полимераза, то она проходит дальше и комплекс mTERF·ДНК диссоциирует («протекание терминатора»), или полимераза терминирует и комплекс сохраняется («непротекание терминатора»). В остальном дисциплина взаимодействия объектов остаётся прежней, как в пластидах.

Описанная выше модель используется следующим образом. Некоторые значения λ могут быть оценены из опыта, тогда оставшиеся (возможно, все) значения λ находятся как решение «обратной задачи». А именно, по известным из опыта значениям уровней транскрипции некоторых генов находятся λ, для которых в рамках модели имеет место наилучшее (в смысле фиксированной метрики) согласование с известными уровнями. Отметим, что в рассмотренных нами примерах эта обратная задача является сильно переопределённой и находится единственное решение или небольшое число решений, среди которых отбирается наилучшим образом согласующееся с дополнительными опытными данными.

В нашей предложена формула зависимости времени τ полураспада РНК от биологически значимых параметров: $\tau = \frac{1}{\mu}(1+d\lambda)\exp(w\lambda)\ln 2$, где $\lambda = \frac{\nu N}{1+\alpha N}$ – интенсивность попыток связывания рибосомы с её сайтом связывания[5], α – параметр в этой зависимости Михаэлиса – Ментен и ν – удельная интенсивность

[2] Pham X.H., Farge G. *et al.* Conserved sequence Box II directs transcription termination and primer formation in mitochondria // *J. Biol. Chem.* 2006. Vol. 281. P. 24647–24652; Wanrooij P.H., Uhler J.P. *et al.* G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation // *Proc. Nat. Acad. Sci. U.S.A.* 2010. Vol. 107. P. 16072–16077.

[3] Martin M., Cho J. *et al.* Termination factor-mediated DNA loop between termination and initiation sites drives mitochondrial rRNA synthesis // *Cell.* 2005. Vol. 123. P. 1227–1240.

[4] Любецкая Е.В., Селиверстов А.В., Любецкий В.А. У актинобактерий число длинных шпилек в межгенных трейлерных областях велико по сравнению с другими областями генома // *Молекулярная биология.* 2007. Т. 41, № 4. С. 739–742.

[5] Конечно, здесь λ отличается от также обозначаемой выше интенсивности попыток связывания РНК полимеразы с её сайтом; обозначение подчёркивает аналогичность этих параметров.

при малых N, где N – количество рибосом в митохондрии без MELAS-мутации. Далее, w – отношение линейного размера h РНКазы вдоль РНК к скорости V элонгации рибосомы (V=15 кодонов в секунду, $h = Vw = 15w$), d – отношение размера h_1 рибосомы вдоль РНК к той же скорости V элонгации рибосомы ($h_1 = 10$ кодонов, $h_1 = Vd$), μ – интенсивность взаимодействия РНКазы с определённым сайтом на мРНК, которая приводит к распаду РНК. Здесь в качестве причины распада РНК рассматривается только действие РНКазы, хотя аналогично можно рассмотреть действие и других факторов. Параметры ν, α и μ зависят от последовательности мРНК, N зависит от экспрессии многих генов и, в особенности, рибосомных генов.

У митохондрий с MELAS-мутацией время полураспада τ' аналогично выражается через N' – количество рибосом в митохондрии. Отсюда $\frac{1+d\lambda}{1+d\lambda'}\exp[w(\lambda-\lambda')]=\frac{\tau}{\tau'}$. В модели получена также зависимость интенсивности распада любой мРНК в результате взаимодействия с РНКазой $\frac{\mu}{1+d\lambda}\cdot\exp(-\lambda w)$.

Конец описания модели.

Значения N и N' определяются как абсолютные количества 12S или 16S рРНК, а w можно оценить в пределах от 2/15 до 4/3 секунд; ν и α не известны и зависят от сайта связывания рибосомы.

Полученные в модели зависимости для времени полураспада показывают, что малое уменьшение количества N рибосом резко уменьшает время полураспада некоторых РНК, а следовательно – количество соответствующего белка. Это может служить объяснением резкого изменения фенотипа при MELAS-мутации[6]. Можно предположить, что у больного человека время полураспада хотя бы одной (возможно, короткой) мРНК значительно уменьшается.

Модель применяется к полным кольцевым митохондриальным ДНК трёх хордовых животных: человека[7] *Homo sapiens*, крысы[8] *Rattus norvegicus* и шпор-

[6] Chomyn A., Martinuzzi A. *et al.* MELAS mutation in mtDNA binding site for transcription termination factor causes defects in protein synthesis and in respiration but no change in levels of upstream and downstream mature transcripts // *Proc. Nat. Acad. Sci. U.S.A.* 1992. Vol. 89. P. 4221–4225.

[7] Gelfand R., Attardi G. Synthesis and turnover of mitochondrial ribonucleic acid in HeLa cells: the mature ribosomal and messenger ribonucleic acid species are metabolically unstable // *Mol. Cell. Biol.* 1981. Vol. 1, № 6. P. 497–511. Piechota J., Tomecki R. *et al.* Differential stability of mitochondrial mRNA in HeLa cells // *Acta Biochim. Pol.* 2006. Vol. 3. P. 157–168.

цевой лягушки[9] *Xenopus laevis*, а также к трём нумерованным выше локусам ДНК пластид растений: резушки *Arabidopsis thaliana* и ячменя *Hordeum vulgare*.

Известны опыты[10], в которых у *Arabidopsis thaliana* и у других растений сравнивались уровни транскрипции генов в мутантных растениях (нокаут гена *sig4*) с соответствующими уровнями в диком типе, т.е. вычислялось отношение MT/WT уровня транскрипции гена до мутации к уровню транскрипции после неё. Проводились опыты[1] по тепловому шоку изолированных хлоропластов, в которых измерялось отношение HT/WT уровня транскрипции гена после теплового шока к уровня транскрипции до него. Подтверждением нашей модели является хорошее согласие её предсказаний с опытами, что видно из таблиц 1–5.

Таблица 1. Изменения уровней транскрипции генов в опыте (столбец 2) и в модели (столбец 3) для локусов 1 и 2. После знака ± указана среднеквадратичная погрешность.

Ген	Опыт	Модель
Локус 1 в *Arabidopsis thaliana*, **нокаут** *sig4*		
ycf1	0.73 ± 0.04	0.76 ± 0.01
ndhF	0.43 ± 0.10	0.47 ± 0.19
*rpl*32	1.52 ± 0.06	1.55 ± 0.02
Локус 2 в *Hordeum vulgare*, **тепловой шок**		
*rpl*23– *rpl*2	2.42 ± 0.27	2.64 ± 0.02
psbA	0.54 ± 0.01	0.54 ± 0.04

Глава 1 завершается следующим заключением. Предложено количественное описание (модель) взаимодействия РНК-полимераз в процессах инициации и элонгации транскрипции. Показано, что модель согласуется практически со всеми опытными данными, относящимися к пластидам растений, включая изменения уровней транскрипции генов после нокаутов σ-субъединиц РНК-полимераз и теплового шока изолированных пластид, относительные количества РНК и времена их полураспада в митохондриях лягушек, человека здорового и с MELAS-мутацией, крысы здоровой и с пониженным уровнем тиреоидного гормона.

---

[8] Enríquez J.A., Fernández-Silva P. *et al.* Direct regulation of mitochondrial RNA synthesis by thyroid hormone // *Mol. Cell. Biol.* 1999. Vol. 19. P. 657–670.

[9] Ammini C.V., Hauswirth W.W. Mitochondrial gene expression is regulated at the level of transcription during early embryogenesis of *Xenopus laevis* // *J. Biol. Chem.* 1999. Vol. 274. P. 6265–6271.

[10] Favory J.-J., Kobayshi M. *et al.* Specific function of a plastid sigma factor for *ndhF* gene transcription // *Nucleic Acids Res.* 2005. Vol. 33. P. 5991–5999. Zghidi W., Merendino L. *et al.* Nucleus-encoded plastid sigma factor SIG3 transcribes specifically the *psbN* gene in plastids // *Nucleic Acids Res.* 2007. Vol. 35. P. 455–464.

Таблица 2. Изменения уровней транскрипции генов в опытах и в модели для локуса 3, нокаут генов *sig3* и *sig4*. В столбцах 2 и 4 указаны значения, полученные в опыте. В остальном аналогично таблице 1.

Ген	**Нокаут *sig3***	**Модель *sig3***	**Нокаут *sig4***	**Модель *sig4***
psbB	1.02 ± 0.36	1.27 ± 0.12	0.69 ± 0.19	0.84 ± 0.11
psbT	0.98 ± 0.25	1.30 ± 0.12	0.96 ± 0.15	0.85 ± 0.11
psbN	0.49 ± 0.46	0.41 ± 0.12	1.03 ± 0.02	1.02 ± 0.19
psbH	1.31 ± 0.05	1.28 ± 0.12	1.01 ± 0.08	0.83 ± 0.11
petB	0.91 ± 0.15	1.09 ± 0.11	0.87 ± 0.29	0.83 ± 0.11
petD	0.92 ± 0.09	0.89 ± 0.10	0.81 ± 0.21	0.81 ± 0.11
rpoA	0.94 ± 0.14	0.82 ± 0.20	0.79 ± 0.11	1.01 ± 0.14
rps11	0.92 ± 0.33	0.90 ± 0.21	0.98 ± 0.31	1.01 ± 0.13
rpl36	0.88 ± 0.11	1.03 ± 0.21	1.54 ± 0.62	1.08 ± 0.18
rps8	1.11 ± 0.04	1.03 ± 0.21	0.83 ± 0.15	1.08 ± 0.18
rpl14	1.04 ± 0.15	1.03 ± 0.21	1.11 ± 0.02	1.08 ± 0.18
rpl16	1.09 ± 0.03	1.03 ± 0.21	1.18 ± 0.03	1.08 ± 0.18
rps3	1.24 ± 0.26	1.03 ± 0.21	1.25 ± 0.02	1.08 ± 0.18
rpl22	1.09 ± 0.13	1.03 ± 0.21	1.20 ± 0.12	1.08 ± 0.18
rps19	1.15 ± 0.50	1.03 ± 0.21	0.96 ± 0.07	1.08 ± 0.17
rpl2	0.94 ± 0.15	1.03 ± 0.21	0.95 ± 0.06	1.08 ± 0.17
rpl23	1.05 ± 0.04	1.06 ± 0.20	1.35 ± 0.33	1.10 ± 0.17

Таблица 3. Результаты для человека, полученные в опытах и в модели: человек здоровый и с MELAS-мутацией. Указаны интенсивности связывания с промоторами *LSP*, *HSP1*, *HSP2* и сайтом терминации *mTERF*; отношение *R* уровня транскрипции гена 12S рРНК к уровню гена COX2. Изменения уровня транскрипции в модели указывает, во сколько раз значение для здорового человека больше, чем для мутанта.

Параметры решения для здорового человека					Уровень транскрипции относительно гена ND1 в модели (вверху) и в опыте (внизу). Для ND1 в опыте 1.00±0.04.						
LSP	*HSP1*	*HSP2*	*mTERF*	*R*	ND2	COX1	COX2	ATP6/8	ND3	ND5	CYTB
0.0031	0.0031	0.0126	0.6456	23.955	1.00	1.00	1.00	0.96	0.96	0.96	0.96
В опыте для этих генов (вычислено):					1.40 ±0.34	1.04 ±1.23	1.72 ±1.23	0.91 ±0.78	1.04 ±0.16	1.86 ±1.09	2.31 ±1.06
Отклонение от опыта в процентах:					-29	-4	-42	+5	-4	-48	-58
Параметры решения при MELAS-синдроме					Изменение уровня транскрипции в модели						
					Phe	12S	Val	16S	Leu	Lys	CYTB
0.0031	0.0004	0.0126	0.5336	24.333	3.84	1.20	1.20	1.20	1.16	1.22	1.17

Таблица 4. Результаты для крыс, полученные в опытах и в модели: эутиреоида и гипотиреоида. Слева значения параметров у эутиреоида (вверху) и у гипотиреоида (внизу). Справа – сравнение результатов моделирования (вверху) и опытных данных (внизу). Здесь *HSP* = *HSP1*+ *HSP2*. Остальные обозначения, как в таблице 3.

LSP	*HSP*	*mTERF*	*R*	Отношение уровней транскрипции у гипотиреоида к эутиреоиду в модели (вверху) и в опыте (внизу)					
				COX1	ATP6/8	COX3	ND4	ND5	CYTB
0.1056	0.0721	0.9453	30.605	0.666	0.641	0.646	0.622	0.614	0.613
0.1056	0.0336	0.9453	30.637	0.61 ±1.02	0.33 ±0.42	0.33 ±0.42	0.61 ±1.02	0.78 ±0.96	0.35 ±0.39
Отклонение от опыта в процентах:				+9	+94	+96	+2	-21	+75

Таблица 5. Результаты для трёх лягушек в модели и в опыте. Данные приведены для части генов. Указаны два параметра: интенсивности связывания mTERF с сайтом терминации и РНК-полимераз с промотором LSP1. Затем – модельные (mod) и опытные (exp) уровни транскрипции генов (относительно нулевого момента Egg) вместе с их относительными отклонениями в процентах (dev).

время	*mTERF*	*LSP1*	ND1			COX2		
лягушка 1			mod	exp	dev,%	mod	exp	dev,%
Egg	0.0157	0.0034	1.0	1.0		1.0	1.0	
+5h	0.0448	0.0089	1.0	1.1	-12	0.9	0.8	+14
+10h	0.0872	0.0157	1.2	1.3	-5	1.1	1.1	+1
+14h	0.0793	0.0173	1.7	2.3	-26	1.6	1.6	-3
+16h	0.0960	0.0209	2.0	2.9	-31	1.7	1.4	+24
+18h	0.0542	0.0157	2.1	3.2	-34	1.9	1.7	+14
+20h	0.0655	0.0157	1.8	3.0	-41	1.6	1.4	+13
+23h	0.0721	0.0492	9.4	9.7	-4	7.6	5.1	+49
+48h	0.0542	0.0872	29.3	26.6	+10	26.2	13.4	+96
+96h	0.0407	0.0960	48.1	48.7	-1	45.3	20.9	+117
время	*mTERF*	*LSP1*	ND1			COX2		
лягушка 2			mod	exp	dev	mod	exp	dev
Egg	0.0089	0.0041	1.0	1.0		1.0	1.0	
+6h	0.0045	0.0023	1.2	1.3	-8	1.2	1.0	+22
+9h	0.0073	0.0045	1.3	1.5	-14	1.3	1.3	-1
+20h	0.0157	0.0157	3.8	4.6	-17	3.7	3.7	+1
+30h	0.0157	0.0230	7.2	7.2	0	7.1	6.8	+4
+48h	0.0407	0.1056	20.5	19.5	+5	19.7	19.7	0
+7days	0.0041	0.0073	6.5	6.1	+7	6.6	8.0	-18
время	*mTERF*	*LSP1*	16S			ND6		
лягушка 3			mod	exp	dev	mod	exp	dev
Egg	0.0960	0.0026	1.0	1.0		1.0	1.0	
+5h	0.0407	0.0050	2.2	2.2	+0.9	2.2	2.2	0.0
+14h	0.0230	0.0081	5.0	5.0	0.0	4.5	4.5	-0.2
+20h	0.0038	0.0028	5.9	6.0	-1.3	4.0	4.0	+0.5
+28h	0.0336	0.1056	92.2	92.0	+0.2	25.1	25.0	+0.4
+48h	0.0143	0.0306	44.1	44.0	+0.2	15.0	15.0	+0.3

Предсказаны характеристики транскрипции в митохондриях хордовых животных: доли РНК-полимераз, завершающих транскрипцию на mTERF-зависимом терминаторе в одном и другом направлениях (поляризация), интенсивность связывания регуляторного белка mTERF с сайтом терминации на ДНК, интенсивности инициации транскрипции на промоторах в пластидах растений и в митохондриях лягушки, человека, включая MELAS-мутацию, и крысы, включая гипотиреоида. Предсказаны значения уровней транскрипции всех генов, в то время как из опытов известны лишь их относительные количества и только для некоторых генов.

Предположен механизм влияния на фенотип MELAS-мутации: понижение количеств фенилаланиновой и валиновой тРНК, рРНК и, главное, резкое изменение времени полураспада некоторых мРНК.

Подтверждена корреляция между изменением метилирования сайта связывания mTERF и трёх промоторов, характерным для перехода от эутиреоида к гипотиреоиду с одной стороны, и изменением интенсивностей связывания белка mTERF и инициаций транскрипции с другой.

Глава 2 посвящена кластеризации пластомных белков, т.е. построению сходных по последовательности и минимальных по содержанию паралогов семейств таких белков. Описывается оригинальный алгоритм кластеризации, который применяется к белкам из трёх обширных групп пластид: родофитной и хлорофитной ветвей и цветковых растений. Результаты собраны в базе данных, доступной по адресу http://lab6.iitp.ru/ppc/. Среди её функций важен поиск белка по его филогенетическому профилю. На её основе рассматривается вопрос о присутствии полноценной РНК-полимеразы бактериального типа у споровиков, а также определяются белки, характерные для узких таксономических групп («филогенетические подписи»).

Результаты предложенного алгоритма хорошо согласуются с биологическими наблюдениями. Например, PsaA и PsaB имеют близкие последовательности и функционируют вместе в составе первой фотосистемы, но не заменяют друг друга и должны быть отнесены к разным кластерам, что и показывает алгоритм. Другой пример связан с регуляцией генов транспорта сульфатов в пласти-

ды: у Viridiplantae не ортологичные гены *cysA* и *cysT* образуют два кластера; в их 5'-лидерных областях найден общий регуляторный мотив.

Математически решается следующая задача. Дано множество последовательностей в фиксированном алфавите, разбитое на непересекающиеся подмножества (каждое подмножество состоит из белков, кодируемых в одном пластоме). Требуется по-другому разбить это множество на попарно непересекающиеся подмножества (кластеры), так чтобы в один кластер попали сходные по последовательности белки из разных пластомов, а белки из одного пластома как можно реже попадали в один кластер.

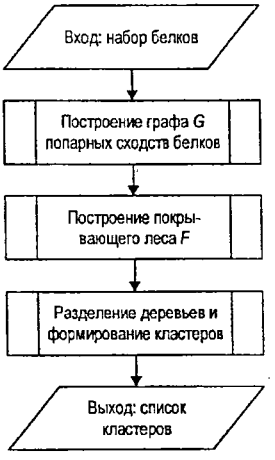

Рисунок 1. Общий план алгоритма кластеризации

Описание алгоритма (рис. 1). Пусть задан набор пластомов S_i и для каждого пластома перечислены его белки P_{ij}. Для всех пар белков (P_{ij}, P_{kl}) вычисляется характеристика сходства $s_0(P_{ij}, P_{kl})$, на основе которой определяется нормированное сходство $s(P_{ij}, P_{kl}) = 2s_0(P_{ij}, P_{kl})(s_0(P_{ij}, P_{ij}) + s_0(P_{kl}, P_{kl}))^{-1}$. Оно максимально (и равно единице), когда белки совпадают.

Рассматривается полный неориентированный граф G_0 с множеством вершин $\{P_{ij}\}$, в котором каждому ребру (P_{ij}, P_{kl}) приписано значение $s(P_{ij}, P_{kl})$ – *вес* этого ребра (петли отсутствуют). На основе G_0 строится разреженный граф G, включающий только рёбра (P_{ij}, P_{kl}), удовлетворяющие следующим условиям: $s(P_{ij}, P_{kl}) = \max_m s(P_{im}, P_{kl}) = \max_m s(P_{ij}, P_{km})$, $s(P_{ij}, P_{kl}) \geq L$, где максимумы берутся по всем белкам из соответствующих пластомов: i-го и k-го, L – параметр алгоритма. В случае $i = k$ предполагается ещё условие $m \neq l$.

Для графа G алгоритм процедурой Крускала строит лес F (ациклический подграф, компоненты связности которого – деревья), включающий все вершины из G. А именно, в G перебираются рёбра в порядке убывания их веса (при совпадении весов сначала выбираются рёбра, соединяющие белки одного пластома), которые объявляются рёбрами строящегося леса F, если добавление к F очеред-

ного ребра из G не приводит к появлению в F цикла. Сумма весов всех рёбер леса называется его весом. Вес полученного леса максимален по сравнению с любым другим лесом в G.

К лесу F применяется следующая процедура разделения деревьев, строящая набор C искомых белковых кластеров. Пусть T – дерево из F и e – ребро в T с минимальным по всем рёбрам в T весом s. Если $s < H$, где H – параметр алгоритма, и T не удовлетворяет сформулированному ниже *критерию сохранения дерева*, то T заменяется в F на два новых дерева путём удаления из T ребра e; в противном случае (т.е. когда критерий выполнен или $s \geq H$) дерево T перемещается из F в список C.

Критерий сохранения дерева T (рис. 2) состоит в выполнении трёх условий: $|T| \leq pn$, где $|T|$ – число вершин в дереве T, n – число всех пластомов в исходном наборе, p – параметр алгоритма; ребро (P_{ij}, P_{kl}) с минимальным в T весом соединяет белки P_{ij} и P_{kl}, у которых $i \neq k$; любая пара вершин P_{ij} и P_{il} дерева T, соответствующих белкам i-го пластома, соединена в T путём, состоящим из вершин, соответствующих белкам этого пластома (то есть, подграф в T, состоящий из вершин, относящихся к одному пластому, является связным). Конец критерия.

Если в F остались деревья, то рассматривается следующее дерево T из F, иначе алгоритм завершает работу. Полученный в результате набор деревьев C представляет кластеры исходных белков: один кластер состоит из последовательностей, приписанных вершинам одного дерева. Конец описания алгоритма.

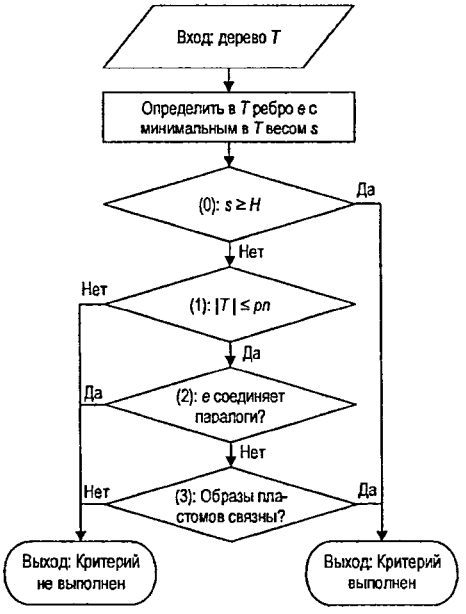

Рисунок 2. Схема проверки критерия сохранения дерева кластера

Предложение 1. Для любых белков P_0 и P_n, если в графе G существует путь от P_0 к P_n с весами рёбер не меньше H, то алгоритм помещает P_0 и P_n в один кластер. □

Предложение 2. Пусть C_1 и C_2 – две кластеризации одного множества белков при значениях H_1 и H_2 параметра H, соответственно. Если $H_1 > H_2$, то $C_1 = C_2$ или C_1 – измельчение C_2. □

Предложение 1 указывает ограничение снизу на размер кластера. Предложение 2 неформально означает, что при увеличении параметра H кластеры разделяются на части, но никогда не объединяются.

Следствие 1. Условие: указаны наборы белков, элементы которых должны находиться в разных кластерах. Существует не более одного максимального по включению интервала, для которого выполняется: при любом значении параметра H из интервала алгоритм выдаёт кластеризацию, удовлетворяющую условию, и никакие два её кластера нельзя объединить с сохранением условия. □

Следствие 2. Условие: указаны наборы белков, ни один набор не должен разделяться кластерами. Существует максимальный по включению интервал, для которого выполняется: при любом значении параметра H из интервала алгоритм выдаёт кластеризацию, удовлетворяющую условию, и ни один кластер нельзя разбить на меньшие с сохранением условия. □

В обоих следствиях границы интервалов – рациональные числа (или бесконечность), которые вычисляются алгоритмически. Число из пересечения этих интервалов бралось в качестве значения параметра H, своего для каждой филогенетической группы. Например, у цветковых растений это пересечение – узкая окрестность, включающая $H = 0.5$.

В результате применения описанного алгоритма получены семейства белков, кодируемых в пластомах багрянок и видов с пластидами, родственными пластидам багрянок, – родофитная ветвь; белков, кодируемых в пластомах рано отделившихся ветвей зелёных водорослей и видов с родственными пластидами (Viridiplantae, эвгленовые, *Bigelowiella natans*), – хлорофитная ветвь; белков, кодируемых в пластидах всех цветковых растений, и отдельно – однодольных растений. На этой основе получены распределения числа белковых семейств (кластеров) в зависимости от числа представленных в них видов для четырёх указанных

групп пластид. Найдены белки, специфичные для пластомов небольших таксономических групп водорослей и простейших; например, белки *ycf*88, *ycf*89 специфичны для диатомовых водорослей и их третичных эндосимбионтов. Показано, что у споровиков *Toxoplasma gondii* и *Plasmodium falceparum* присутствует полноценная РНК-полимераза бактериального типа. У *Neospora caninum* и *Plasmodium* spp. найдены α- и σ-субъединицы, кодируемые в ядре. Напротив, у споровиков таксономической группы Piroplasmida α- и σ-субъединицы РНК-полимеразы бактериального типа не найдены, а её субъединицы, обычно кодируемые в пластидах споровиков, значительно изменены или фрагментированы. Это позволяет предположить глубокое различие видов Piroplasmida с другими содержащими пластиды споровиками в части транскрипции в пластидах.

Короткая **глава 3** посвящена изучению сопряжения трансляции и транскрипции в пластидах с использованием оригинальной компьютерной программы поиска клики. А именно, изучению возможных механизмов задержки инициации трансляции до завершения процессинга транскрибированной мРНК. Элонгация РНК-полимеразы фагового типа существенно быстрее элонгации рибосомы, так что в этом случае остаётся достаточно времени для процессинга.

В результате предположен механизм задержки инициации трансляции до завершения редактирования мРНК генов *accD* и *atpH* у пластид растений видов *Adiantum capillus-veneris* и *Anthoceros formosae*, механизм вовлекает найденные неконсервативные длинные шпильки в 5'-нетранслируемой области около сайта связывания рибосомы. Эти шпильки имеют минимальные значения энергии. Найдены консервативные сайты перед шестью генами *atpF*, *clpP*, *petB*, *psaA*, *psbA*, *psbB* у трёх видов *Chara vulgaris*, *Zygnema circumcarinatum*, *Physcomitrella patens*. Получена корреляция между присутствием интронов в гене и наличием шпильки или консервативного сайта перед геном. У этих видов 5'-лидерные области значительно отличаются от аналогичных областей у сосудистых растений.

Для определения мотива в данном наборе нуклеотидных последовательностей использован алгоритм[11] поиска клики данного размера в многодольном графе. А именно, для заданного числа k формируется граф, в котором каждая доля

[11] Любецкий В.А., Селиверстов А.В. Некоторые алгоритмы, связанные с конечными группами // *Информационные процессы*. 2003. Т. 3, № 1. С. 39–46.

соответствует одной из последовательностей, и вершинам доли соответствуют все участки длиной k в этой последовательности, каждому ребру приписано сходство участков, соответствующих его концам. Сходство учитывает GC-состав участков, что является усовершенствованием алгоритма, ранее полученного в лаборатории. А именно, пусть среднее по всем геномам, из которых взяты последовательности, долей вхождений G и C равна p (и среднее вхождений A и T равна $1-p$), тогда сходство участков полагается равным сумме по позициям сходств нуклеотидов в них, последние вычисляются по таблице 6.

Таблица 6. Сходство нуклеотидов, используемое при вычислении сходства двух участков с одинаковой длиной

	A	C	G	T
A	1	½	½	p
C	½	1	$1-p$	½
G	½	$1-p$	1	½
T	p	½	½	1

Следующее простое предложение 3 относится к выбору значений в таблице 6. Если p маленькое, то появление на выравнивании двух участков A против T – малозначимое событие, а появление C против G – значимое; если p большое, то наоборот. Таким образом, большое сходство участков получается, если выравнивание содержит много редких событий, т.е. несёт много информации.

Предложение 3. Пусть $0 \le p < \frac{1}{2}$ и даны два случайных участка одинаковой длины в алфавите $\{A,C,G,T\}$, в которые буквы G и C входят с вероятностью $p/2$, а буквы A и T – с вероятностью $(1-p)/2$. Тогда в любой позиции выравнивания этих участков вероятность появления пары $\{A,T\}$ строго больше вероятности появления пары $\{G,C\}$. Если $\frac{1}{2} < p \le 1$, то вероятности связаны противоположным неравенством. □

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

Разработана математическая и компьютерная модель взаимодействия РНК-полимераз между собой, с вторичными структурами и белковыми факторами в процессах инициации и элонгации транскрипции. Модель применена к локусам пластид и митохондрий, и находится в согласии практически со всеми опытными данными, относящимися к пластидам растений и митохондриям, включая данные об изменениях уровней транскрипции генов после нокаутов σ-субъединиц РНК-полимераз и после теплового шока изолированных пластид, данные об относительных количествах РНК и временах их полураспада в митохондриях лягушек, человека здорового и с MELAS-мутацией, крысы здоровой и с пониженным уровнем тиреоидного гормона. (Глава 1)

На основе модели предсказаны характеристики транскрипции в митохондриях хордовых животных: доли РНК-полимераз, завершающих транскрипцию на mTERF-зависимом терминаторе в одном и другом направлениях (поляризация); интенсивность связывания регуляторного белка mTERF с сайтом терминации на ДНК; интенсивности инициации транскрипции на промоторах в пластидах растений и митохондриях лягушки, человека, включая случай MELAS-мутации, крысы, включая гипотиреоида. На основе модели предсказаны значения уровней транскрипции всех генов, в то время как в опытах известны лишь их относительные значения и только для некоторых генов. (Глава 1)

На основе модели предположен механизм влияния на фенотип MELAS-мутации: снижение концентраций как фенилаланиновой и валиновой тРНК, так и рРНК, а главное – резкое изменение времени полураспада определённых мРНК. (Глава 1)

На основе модели показана корреляция между изменениями метилирования сайта связывания mTERF и промоторов с интенсивностями связывания с ними mTERF и РНК-полимераз. (Глава 1)

Разработан алгоритм кластеризации множества белковых последовательностей. На его основе получены семейства сходных по последовательности и минимальных по содержанию паралогов белков, кодируемых в пластомах багрянок и видов с пластидами, родственными пластидам багрянок (родофитная

ветвь); белков, кодируемых в пластомах рано отделившихся ветвей зелёных водорослей и видов с родственными им пластидами: Viridiplantae, эвгленовые, *Bigelowiella natans* (хлорофитная ветвь); белков, кодируемых в пластомах цветковых и отдельно однодольных растений. На этой основе найдены белки, специфичные для пластомов небольших таксономических групп водорослей и простейших. (Глава 2)

Полученная кластеризация позволила заключить, что у споровиков *Toxoplasma gondii* и *Plasmodium falciparum* присутствует полноценная РНК-полимераза бактериального типа. У *Neospora caninum* и *Plasmodium* spp. найдены α- и σ-субъединицы, кодируемые в ядре. Напротив, у споровиков таксономической группы Piroplasmida α- и σ-субъединицы РНК-полимеразы бактериального типа не найдены, а её субъединицы, обычно кодируемые в пластидах, значительно изменены или фрагментированы. Это позволяет предположить глубокое различие видов Piroplasmida с другими содержащими пластиды споровиками в части транскрипции в пластидах. (Глава 2)

На основе оригинальной компьютерной программы (поиска мотива путём определения клики в многодольном графе с учётом GC-состава) предположен механизм задержки инициации трансляции до завершения редактирования транскриптов генов *accD* и *atpH* в пластидах растений видов *Adiantum capillus-veneris* и *Anthoceros formosae*. Механизм вовлекает длинные шпильки в 5'-лидерной области около сайта связывания рибосомы. Найдены консервативные сайты перед шестью генами *atpF*, *clpP*, *petB*, *psaA*, *psbA*, *psbB* у трёх видов *Chara vulgaris*, *Zygnema circumcarinatum*, *Physcomitrella patens*, которые в части случаев также участвуют в задержке инициации трансляции до завершения сплайсинга или редактирования. (Глава 3)

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи:

1. Lyubetsky V.A., Zverkov O.A., Pirogov S.A., Rubanov L.I., Seliverstov A.V. Modeling RNA polymerase interaction in mitochondria of chordates // *Biology Direct.* 2012. 7:26.
2. Lyubetsky V.A., Zverkov O.A., Rubanov L.I., Seliverstov A.V. Modeling RNA polymerase competition: the effect of σ-subunit knockout and heat shock on gene transcription level // *Biology Direct.* 2011. 6:3.
3. Любецкий В.А., Селиверстов А.В., Зверков О.А. Построение разделяющих паралоги семейств гомологичных белков, кодируемых в пластидах цветковых растений // *Мат. биол. и биоинф.* 2013. Т. 8, № 1. С. 225–233.
4. Зверков О.А., Селиверстов А.В., Любецкий В.А. Белковые семейства, специфичные для пластомов небольших таксономических групп водорослей и простейших // *Молекулярная биология.* 2012. Т. 46, № 5. С. 799–809.
5. Lyubetsky V.A., Seliverstov A.V., Zverkov O.A. Transcription regulation of plastid genes involved in sulfate transport in Viridiplantae // *BioMed Research International.* 2013. Vol. 2013. Article ID 413450, 6 pages.
6. Зверков О.А., Русин Л.Ю., Селиверстов А.В., Любецкий В.А. Изучение вставок прямых повторов в микроэволюции митохондрий и пластид растений на основе кластеризации белков // *Вестник Московского университета. Серия 16: Биология.* 2013. № 1. С. 8–13.
7. Зверков О.А., Селиверстов А.В., Любецкий В.А. Усредненная энтропия как характеристика консервативности участков генома // *Вестник Тамбовского университета. Серия: Естественные и технические науки.* 2013. Т. 18, Вып. 5. С. 2529–2531.
8. Lyubetsky V.A., Korolev S.A., Seliverstov A.V., Zverkov O.A., Rubanov L.I. Gene expression regulation of the PF00480 or PF14340 domain proteins suggests their involvement in sulfur metabolism // *Computational Biology and Chemistry.* 2014. Vol. 49. P. 7–13.
9. Seliverstov A.V., Zverkov O.A., Lyubetsky V.A. Translation of some chloroplast genes is checked to allow for splicing and editing // *Biophysics.* 2006. Vol. 51, S. 1. P. 18–22.

Тезисы докладов:

1. Lyubetsky V.A., Seliverstov A.V., Zverkov O.A. RNA Structures upstream *leuA* Genes in α-proteobacteria // *Proceedings of the International Moscow Conference on Computational Molecular Biology: MCCMB'07*. July 27–31 2007. P. 191–192.
2. Зверков О.А. Программный комплекс для согласования набора эволюционных деревьев и выявления эволюционных событий // *Труды 51-й научной конференции МФТИ*. Москва, 2008. С. 133–136.
3. Лопатовская К.В., Зверков О.А., Селиверстов А.В., Любецкий В.А. Транскрипция генов синтеза пролина у бактерий родов *Marinobacter*, *Pseudomonas* и *Shewanella* регулируется белком семейства tetR // *Труды 32-й конференции «Информационные технологии и системы»*. Бекасово, 15–18 декабря 2009. С. 278–281.
4. Зверков О.А., Селиверстов А.В., Рубанов Л.И., Любецкий В.А. Моделирование конкуренции РНК-полимераз: влияние нокаута сигма субъединицы и температуры на экспрессию генов // *Труды 32-й конференции «Информационные технологии и системы»*. Бекасово, 15–18 декабря 2009. С. 328–331.
5. Lyubetsky V.A., Zverkov O.A., Rubanov L.I., Seliverstov A.V. Interaction between nucleome and plastome: heat shock response regulation in plastids of plants // *Proceedings of the Seventh International Conference on Bioinformatics of Genome Regulation and Structure\Systems Biology*. Novosibirsk, June 20–27 2010. P. 161.
6. Зверков О.А., Селиверстов А.В., Любецкий В.А. Позиционная связь генов пластомов растений и водорослей // *Труды 33-й конференции «Информационные технологии и системы»*. г. Геленджик, 20–24 сентября 2010. С. 326–330.
7. Зверков О.А., Селиверстов А.В., Любецкий В.А. Об одном алгоритме кластеризации белков // *Труды 53-й научной конференции МФТИ*, Часть I. Радиотехника и кибернетика, Т. 1, М.: МФТИ, 2010. С. 118–119.
8. Зверков О.А., Горбунов К.Ю., Селиверстов А.В., Любецкий В.А. Кластеризация белков с учётом их доменной структуры // *Труды 54-й научной конференции МФТИ*. Т. 2. М.: МФТИ, 2011. С. 88–89.
9. Зверков О.А., Селиверстов А.В., Любецкий В.А. Семейства белков, кодируемых в пластомах Chlorophyta, Euglenozoa и Rhizaria // *Труды 35-й конференции «Информационные технологии и системы»*, 19–25 августа 2012. С. 298–302.

10. Zverkov O.A., Korolev S.A., Seliverstov A.V., Lyubetsky V.A. Transcription regulation of plastid genes *cysT* and *cysA* in Viridiplantae // *Contributions to the 3rd Moscow International Conference "Molecular Phylogenetics"*. July 31 – August 4, 2012. P. 85.
11. Зверков О.А. Использование быстрых алгоритмов в задаче кластеризации последовательностей // *Сборник избранных трудов VIII Международной научно-практической конференции «Современные информационные технологии и ИТ-образование»*. Москва, МГУ им. М.В.Ломоносова, 8–10 ноября 2013. С. 757–763.
12. Зверков О.А., Селиверстов А.В., Любецкий В.А. Построение разделяющих паралоги семейств гомологичных белков, кодируемых в пластидах цветковых растений // *Труды 37-й конференции «Информационные технологии и системы»*. Калининград, 1–6 сентября 2013. С. 172–177.
13. Kobets N.V., Goncharov D.B., Seliverstov A.V., Zverkov O.A., Lyubetsky V.A. Comparative analysis of apicoplast-targeted proteins in Toxoplasma gondii and other Apicomplexa species // *Proceedings of the International Moscow Conference on Computational Molecular Biology: MCCMB'13*, July 25–28, 2013.

Подписано в печать: 17.07.2014

Заказ № 10128 Тираж - 100 экз.
Печать трафаретная.
Типография «11-й ФОРМАТ»
ИНН 7726330900
115230, Москва, Варшавское ш., 36
(499) 788-78-56
www.autoreferat.ru