

Математическая модель поддержания тотипотентности и дифференцировки клеток при развитии меристемы побега *Arabidopsis thaliana*

Акбердин Илья Ринатович

Аспирант третьего года

Лаборатория теоретической генетики, ИЦиГ СОРАН, Новосибирск, Россия

akberdin@bionet.nsc.ru

Гормон ауксин (класс низкомолекулярных соединений, относящихся к производным индола) является одним из важнейших веществ, участвующих в регуляции различных биологических процессов растений, в особенности морфогенеза. Координация ауксином процессов роста и развития растений осуществляется через контролируемое формирование паттернов распределения ауксина и дозозависимых механизмов его действия. В течение последнего десятилетия наблюдается огромный интерес к применению технологий математического моделирования к задачам, связанным с развитием растений, опосредованным ауксином. Основной мотивацией теоретических исследований является следующее: а) предсказать движение и распределение паттернов ауксина в развивающейся ткани; б) показать, что некоторые гипотезы относительно роли ауксина, как одного из сигналов развития, проверяемые *in silico*, позволяют получить результаты, сравнимые с наблюдаемыми *in vivo*. Ранее нами была разработана математическая модель развития меристемы побега в терминах клеточного автомата. Анализ разработанной модели показывает, что динамика клеточного автомата не только качественно совпадает с морфодинамикой реального зародыша, но, при варьировании начальных параметров, также можно добиться количественного соответствия по числу клеток в зародыше на определённых этапах развития. Более того, было обнаружено, что при подборе параметров модели наблюдается поведение автомата, аналогичное морфодинамике мутантного фенотипа. Параллельно коллегами лаборатории теоретической генетики была предложена и разработана одномерная математическая модель (1D модель), описывающая динамику распределения ауксина в клетках, лежащих на центральной оси корня растения, с учетом регуляции ауксином собственного транспорта. Разработанная 1D модель воспроизводит некоторые важные особенности ауксин-зависимого развития корня. Анализ 1D модели позволяет предположить, что разные типы распределений концентрации ауксина в клетках могут реализоваться на разных этапах роста корня и задавать разные сценарии его развития. Из сравнительного анализа существующих моделей можно сделать вывод, что в корне и побеге преобладают разные типы транспорта ауксина. В корне ауксин более активно транспортируется из клетки, а в побеге – более активно транспортируется в клетку.

Однако остается неясным вопрос о критичности данных различий. В связи с этим мы построили серию математических моделей и проанализировали различные типы транспорта ауксина. Мы показываем, что преобладания различных типов транспорта является критичными для формирования наблюдаемых паттернов распределения ауксина в корне и меристеме побега.

Следующим иерархическим этапом развития моделирования в области поддержания тотипотентности и дифференцировки клеток при развитии меристемы побега *Arabidopsis thaliana* являлась разработка математической модели метаболизма ауксина в клетке побега. Для достижения данной цели мы реконструировали генную сеть метаболизма ауксина и, на основе разработанного нами же конвертора генных сетей в формат математических моделей, построили математическую модель метаболизма ауксина. Данная математическая модель воспроизводит экспериментальную динамику наработки ауксина и его конъюгатных форм в клетке *Arabidopsis thaliana*.

Способ вычисления ошибки дистанции МакАйнерни и его практическое применение для решения прикладных задач

Бутвиловский Александр Валерьевич

ассистент, к.м.н.

Одинцов Александр Олегович

студент

Белорусский государственный медицинский университет, Минск, Беларусь

E-mail: alexbutv@rambler.ru

Цель исследования: разработать и апробировать способ вычисления ошибки дистанции МакАйнерни.

Материалы и методы. Предлагаемый способ вычисления ошибки дистанции МакАйнерни ($SE_{Djk} = SD_{Djk} / \sqrt{n}$) предусматривает первоначальное определение отклонения (SD_{Djk}) следующим образом: $SD_{Djk} = \sqrt{(\sum d^2) / n}$, где $d = V_B - M_B$, $V_B = abs(RSCU_{ji} - RSCU_{ki}) / n$, n – число пар сравниваемых признаков. Для апробации предлагаемого способа проанализированы взятые с сервера NCBI (www.ncbi.nlm.nih.gov) последовательности мРНК, кодирующие ряд митохондриальных белков (субъединицы 1–6, 4L НАДН-дегидрогеназы (НАДН-ДГ), цитохром *b*, субъединицы 1–3 цитохром-*c*-оксидазы (ЦО), субъединица 6 АТФ-синтазы) человека и трихинеллы (*Trichinella spiralis*). В качестве контроля использованы мРНК, кодирующие аналогичные белки свободноживущего круглого червя цианорабдитис (*Caenorhabditis elegans*). Показатели RSCU рассчитаны при помощи программы MEGA 3 [1]. Сходство стратегий

кодирования изучаемых белков в мРНК определено по дистанции МакАйнерни [2] с вычислением ошибки предлагаемым способом.

Результаты и обсуждение. Дистанции МакАйнерни для мРНК, кодирующих субъединицу 1 НАДН-ДГ человека и трихинеллы (человека и цианорабдитис) равны $0,010 \pm 0,0015$ ($0,018 \pm 0,0020$), НАДН-ДГ 2 – $0,022 \pm 0,0016$ ($0,016 \pm 0,0016$), НАДН-ДГ 3 – $0,013 \pm 0,0022$ ($0,017 \pm 0,0021$), НАДН-ДГ 4 – $0,024 \pm 0,0017$ ($0,017 \pm 0,0018$), НАДН-ДГ 4L – $0,023 \pm 0,0023$ ($0,020 \pm 0,0024$), НАДН-ДГ 5 – $0,023 \pm 0,0013$ ($0,020 \pm 0,0021$), НАДН-ДГ 6 – $0,021 \pm 0,0021$ ($0,015 \pm 0,0019$), цитохром *b* – $0,009 \pm 0,0013$ ($0,018 \pm 0,0017$), ЦО 1 – $0,008 \pm 0,0009$ ($0,016 \pm 0,0016$), ЦО 2 – $0,010 \pm 0,0014$ ($0,017 \pm 0,0016$), ЦО 3 – $0,010 \pm 0,0014$ ($0,017 \pm 0,0016$), АТФ-синтазу 6 – $0,008 \pm 0,0013$ ($0,018 \pm 0,0019$). Предложенный способ является эффективным, так как он позволил вычислить ошибку дистанции во всех опытах. Отличия дистанций МакАйнерни недостоверны лишь для 3-х изученных мРНК, кодирующих ($25,0 \pm 13,06\%$) белки человека и трихинеллы (по сравнению с таковыми для мРНК человека и цианорабдитис) – субъединицы 3, 4L и 5 НАДН-дегидрогеназы. Для остальных изученных мРНК различия являются достоверными ($p < 0,05$), что свидетельствует о большем сходстве картины использования синонимичных кодонов в мРНК человека с таковой трихинеллы, а не цианорабдитис. Эти данные подтверждают возможность использования разработанного способа для решения прикладных задач.

Закключение. Предложенный способ вычисления ошибки дистанции МакАйнерни является эффективным. С помощью предлагаемого способа удалось установить, что картина использования синонимичных кодонов в мРНК человека более сходна с таковой трихинеллы (по сравнению с контролем).

Автор выражает благодарность доценту, к.м.н. Бутвиловскому В.Э. за помощь в подготовке тезисов.

Литература

1. Kumar, S. [et al.] (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment // Brief. Bioinform., № 5, p. 150–163.
2. McInerney, J.O. (1998) GCUA: General codon usage analysis // Bioinformatics, № 14(4), p. 372–373.

Модель классической аттенуаторной регуляции

Глотова Ирина Вячеславовна
студент

Московский государственный университет им. М.В. Ломоносова, Москва, Россия
E-mail: igbox@mail.ru

Моделирование процессов регуляции экспрессии генов у бактерий – актуальная задача. Массовый счет на основе таких моделей позволяет оценить сами модели,

сравнивая результаты счета с известными данными, и, с другой стороны, позволяет предсказать новые регуляторные сайты.

Предложенная в [1] модель классической аттенуаторной регуляции усовершенствована в следующих отношениях: вторичные регуляторные структуры в лидерной области допускают образование псевдоузлов (при этом используется метод подсчета энергии, предложенный в [2]) и учитывается влияние РНК-триплексов. Процедура вычисления энергии вторичной структуры подвергнута тщательному анализу.

На основе полученной модели при одних и тех же фиксированных параметрах проведен массовый счет для бактериальных оперонов, отвечающих за синтез Trp, Ile, Leu, Val, Thr, Phe и His. Полученные результаты согласуются с множественными выравниваниями лидерных областей и с экспериментальными данными, когда они известны. Для многих видов протеобактерий и актинобактерий модель предсказывает новые случаи классической аттенуаторной регуляции.

В таблице приведена небольшая часть полученных результатов: только для некоторых гамма-протеобактерий, которые относятся к числу наиболее изученных. Буква «А» указывает на случаи, в которых классическая аттенуаторная регуляция обнаружена биоинформатическими методами и/или экспериментально, а буква «Р» – на случаи, в которых модель предсказывает эту регуляцию.

Гамма-протеобактерии	<i>hisG</i>	<i>pheA</i>	<i>pheS</i>	<i>trpE</i>	<i>thrA</i>	<i>leuA</i>	<i>ilvA</i>	<i>ilvC</i>	<i>ilvB</i>	<i>ilvG</i>	<i>ilvI</i>
<i>Escherichia coli</i>	A, P	A	A, P	A, P	A, P	A			A, P	A	
<i>Salmonella typhi</i>	A, P	A	A, P	A	A, P	A			A	A, P	
<i>Klebsiella pneumoniae</i>	A, P	A	A, P	A, P	A, P	A			A, P	A	
<i>Erwinia carotovora</i>		A	A	A	A, P	A			A, P	A	
<i>Yersinia pestis</i>	A, P		A, P	A, P	A				A, P	A	
<i>Haemophilus influenzae</i>	A, P				A, P	A				A, P	
<i>Pasterella multocida</i>					A, P	A				A	
<i>Vibrio cholerae</i>	A	A, P		A, P	A	A, P				A, P	
<i>Vibrio vulnificus</i>				A, P							
<i>Vibrio parahaemolyticus</i>				A, P							
<i>Shewanella oneidensis</i>		A, P		A	A, P	A				A, P	
<i>Idiomarina loihiensis</i> L2TR					P						
<i>Pseudoalteromonas haloplanktis</i> TAC125					P						
<i>Pseudomonas putida</i>										A, P	
<i>Pseudomonas syringae</i>										A, P	

В дальнейшем планируется построить аналогичные модели и провести их сравнение с биоинформатическими и экспериментальными данными для других аттенуаторных регуляций и, в первую очередь, для Т-боксовой регуляции.

Литература

1. V. Lyubetsky, S. Pirogov, L. Rubanov, A. Seliverstov, "Modeling classic attenuation regulation of gene expression in bacteria", *Journal of Bioinformatics and Computational Biology*, vol.5, no.1, 2007, pp. 155-180.
2. H. Isambert, E. Siggia, "Modeling RNA folding paths with Pseudoknots: Application to hepatic delta virus ribozyme", *PNAS*, vol.97, no.12, 2007, pp. 6515-652

Улучшение сборки микробных геномов с использованием данных об их оперонной структуре

Диброва Дарья Владимировна
студентка

Московский государственный университет им. М.В. Ломоносова, Москва, Россия
E-mail: udavdasha@gmail.com

Сейчас секвенирование микробных геномов становится все более широко используемым, а задачи, связанные с ним, востребованными. Одна из них – задача контроля за ошибками.

В настоящей работе содержательно решается следующая задача. Дана пара бактериальных генов. Требуется на основе анализа банковских бактериальных геномов предсказать, могут ли составляющие эту пары гены лежать на геноме на значительном расстоянии друг от друга, или же на геноме эти гены должны располагаться рядом. Также предсказание требуется дополнить характеристикой достоверности.

Алгоритм, решающий сформулированную задачу, планируется применять при секвенировании бактериальных геномов:

- 1) для поиска возможных ошибок секвенирования;
- 2) для получения дополнительной информации о взаимном расположении контигов на геноме (другими словами, для формирования скаффолдов по набору контигов).

В алгоритме используется сходный с поиском оперонов подход (см. [1]).

При тестировании разработанного алгоритма на банковских геномах уже удалось выявить несколько вероятных ошибок секвенирования.

Литература

- [1] Prediction of operons in microbial genomes, *M.D. Ermolaeva et al.*, **1216-1221** *Nucleic Acids Research*, 2001, Vol. 29, No.5

Изучение последовательностей трипсиноподобных ферментов грибов

Дубовенко Алексей Жоржевич

студент

Московский государственный университет им. М.В. Ломоносова, Москва, Россия

E-mail: taurusmage@yandex.ru

Для патогенных организмов особенно важным инструментом являются пептидазы, поскольку с их помощью патоген проникает в организм хозяина и обеспечивает себя питательными веществами. Показано, что двумя важными группами секретируемых грибами пептидаз являются субтилазы и трипсины. Предварительные результаты позволили высказать предположение о существовании корреляции между наличием гена трипсина и фитопатогенностью грибов (Грубань, 2001; Hu, St.Leger, 2004). Был проведён анализ имеющихся на сегодняшний день в банке NCBI геномов грибов на наличие гена(ов) трипсина и его гомологов, а также были изучены последовательности трипсиноподобных белков из белковых банков.

Поиск с помощью программы BLAST в 54 секвенированных геномах грибов БД NCBI показал, что гены трипсина и его гомологов присутствуют у 16 представителей отдела Ascomycota, подотдел Pezizomycotina, а именно: у 11 представителей – патогенов растений, а также у 5 факультативных патогенов человека, представителей рода *Aspergillus*. Гены трипсина не обнаружены в оставшихся секвенированных геномах Ascomycota (20 Saccharomycotina, 10 Pezizonomycotina, 2 Schizosaccharomycetes), 5 Basidiomycota, а также единственного представителя группы Microsporidia. При дополнительном поиске в БД Swiss-Prot и TrEMBL с помощью поисковой системы SRS (<http://srs.ebi.ac.uk/>), а так же с помощью программ TBLASTN (среди Expressed Sequence Tags), PSI-BLAST (среди non-redundant sequences) найдено ещё 20 последовательностей трипсинов у 9 грибов-патогенов из отдела Ascomycota, 1 из Basidiomycota, 3 из Zygomycota.

Анализ первичной последовательности на предмет выделения аминокислот активного центра фермента показал, что из 43 трипсиноподобных последовательностей каталитическую функцию могут выполнять только 31, в активных центрах остальных последовательностей имеются замены, препятствующие выполнению каталитической функции трипсина. Анализ а.о. ответственных за субстратную специфичность показал, что одна последовательность предположительно является химотрипсином. Анализ а.о., по карбоксилу которых происходит отщепление пропептида фермента, показал, что

активация профермента трипсина у грибов осуществляется ферментом, отличающимся по субстратной специфичности от такового у животных.

Фрагментарность распределения трипсинов среди грибов говорит о том что их филогенетическое распространение у ранних грибов должно было быть шире, нежели у современных. Анализируя виды, в геномах которых не обнаружено генов трипсина, можно выделить пять групп: сапротрофы, патогены животных, дрожжи (в независимости от патогенности), два узкоспециализированных патогена растений, представитель группы *Microsporidia*.

На сегодняшний день можно судить об относительной справедливости гипотезы о том, что присутствие трипсина – это специфический маркёр фитопатогенности.

Литература

1. Грубань, Т.Н. (2001) Внеклеточные протеолитические ферменты несовершенных грибов. Автореферат на соискание учёной степени канд. биол. наук. Биологический факультет Московского государственного университета им М.В.Ломоносова. Москва.
2. Hu, G. and St.Leger, R.J. (2004) A phylogenetic approach to reconstructing the diversification of serine proteases in fungi. *Journal of Evolutional Biology* 17, 1204-1214.

Исследование корреляций между аминокислотными последовательностями транскрипционных регуляторов и нуклеотидными последовательностями их сайтов связывания с ДНК

Коростелев Юрий Дмитриевич

Студент 4 курса

Московский государственный университет им. М.В. Ломоносова, факультет биоинженерии и биоинформатики Москва, Россия.

Ранее нами был разработан метод поиска скоррелированных замен в последовательностях ДНК-связывающих белков и их сайтов связывания. При тестировании на семействе LACI-регуляторов были обнаружены статистически значимо скоррелированные пары позиций, хорошо соответствовавшие ДНК-белковым контактам в структуре комплекса PurR_Ecolі с его оператором purF (PDB: 1qpz) [1].

В настоящей работе исследована новая версия БД LACI (любезно предоставлена О.Лайковой), которая почти в два раза больше ранее изученной выборки. В программу введена процедура учета филогенетической близости регуляторов и операторов. Полученные скоррелированные пары были сопоставлены с белок-ДНКовыми контактами в трех известных структурах (1qpz, 1efa, 1rzt) и с экспериментальными данными по точечному мутагенезу LACI_Ecolі и его оператора [2]. Исследованы

перепредставленные пары слов (белковое слово – это аминокислотные остатки (а.о.) в заданных позициях, нуклеотидное слово — нуклеотиды в заданных позициях).

Полученный список скоррелированных пар позиций включает практически все полярные контакты боковых групп а.о. с азотистыми основаниями, существующие хотя бы в одной из структур. Две позиции регуляторов, 16 и 20 (нумерация по PurR_Ecoli) являются абсолютными лидерами по скоррелированности. У этой пары а.о. лежат на одной стороне распознающей альфа-спирали. Полученные таблицы сопряженности частот а.о. в позициях 16, 20 и нуклеотидов в позициях 5, 7 (нумерация позиций от 1 до 20, между 10 и 11 находится центр симметрии оператора) согласуются с экспериментальными данными [2].

Детальный анализ полученных предпочтительных пар позволил выявить несколько ярких зависимостей: T^{16} уменьшает вероятность пары $R^{20}-G_5$, что, по-видимому, обусловлено стерическими взаимодействиями, а комбинация $S^{16}R^{20}$ предпочитает G_5C_6 , что можно интерпретировать как возникновение «сложного» [3] контакта аргинина с 2-мя последовательными нуклеотидами. Существование таких зависимостей побудило нас исследовать корреляции между парами слов (белковое слово – а.о. в позициях 16,20, нуклеотидное – нуклеотиды в позициях 5,6,7). Было показано, что перепредставленные пары слов не объясняются ни индивидуальными корреляциями, ни филогенетическим следом, поскольку они встречаются на разных ветвях дерева регуляторов.

При исследовании семейства регуляторов NrtR в предположительном ДНК-связывающем домене были обнаружены две сильно скоррелированных с ДНК позиции, расположенные в одной спирали через один виток. Был отмечен случай скоррелированных замен в белке и операторе у двух очень близких регуляторов.

Таким образом, предложенный метод позволяет не только предсказывать специфические контакты с ДНК, но и исследовать более тонкие взаимодействия между факторами транскрипции и нуклеотидами, влияющие на белок-ДНКовое распознавание. Полученные результаты полезны для планирования генно-инженерных экспериментов.

Литература.

- 1) Y.D.Korostelev et al. (2007) Proc. MCCMB-2007, 158-159.
- 2) J. Sartorius, et al. (1989) *EMBO J.* 8: 1265–1270.
- 3) M. Nicolas et al. (2001) *Nucleic Acids Res.* 29: 2860-2874.

Изоферменты глицеральдегид-3-фосфатдегидрогеназы хордовых животных: выявление генов и исследование первичных структур

Куравский Михаил Львович
студент

Московский государственный университет им М.В. Ломоносова, Москва, Россия
e-mail: m_kuravskiy@yahoo.com

D-глицеральдегид-3-фосфатдегидрогеназа (ГАФД, КФ 1.2.1.12) является одним из гликолитических ферментов. В организме человека ГАФД представлена двумя изоферментами, которые экспрессируются в соматических клетках (ГАФД-1) и в сперматоидах (ГАФД-2). Высокая концентрация ГАФД-1 в клетке, достигающая 5–15% от общей концентрации цитоплазматических белков, дает основания предполагать, что она может выполнять иные функции, отличные от гликолитической. Имеются многочисленные данные, подтверждающие, что помимо гликолиза ГАФД-1 участвует в таких процессах, как эндоцитоз, слияние мембран, репарация, репликация и индукция апоптоза. ГАФД-2 включает в себя дополнительный N-концевой домен длиной 74 а.о., необходимый для фиксации фермента в фиброзном слое жгутика сперматозоида.

Целью данной работы является выявление генов, кодирующих изоферменты ГАФД 17-ти представителей хордовых (Chordata) и иглокожих (Echinodermata), и изучение их последовательностей. Проведенные исследования позволили выявить в геномах всех представителей классов млекопитающих (Mammalia) и лучеперых рыб (Actinopterygii) по два гена ГАФД, в геномах остальных организмов – по одному. Дополнительный N-концевой домен был найден только в составе одного из изоферментов млекопитающих. По результатам реконструкции филогенетических деревьев, последовательности ГАФД хордовых могут быть разделены на две группы. Первую группу образуют ГАФД-1 млекопитающих и лучеперых рыб, а также ГАФД *Gallus gallus* и *Xenopus laevis*, вторую – ГАФД-2 млекопитающих и лучеперых рыб. ГАФД *Strongylocentrotus purpuratus* и *Ciona savignyi* не могут быть надежно отнесены ни к одной из указанных групп ввиду низких соответствующих бутстреп-значений. Ветвь представителей членистоногих (Arthropoda) отделяется до расхождения ветвей ГАФД-1 и ГАФД-2. Следовательно, дупликация первоначального гена ГАФД произошла после расхождения первичноротых и вторичноротых и до формирования типа хордовых. Вывод согласуется с предположением, выдвинутым в работе [1].

Эволюционные расстояния между последовательностями ГАФД-2 каждой пары млекопитающих превышают эволюционные расстояния между последовательностями ГАФД-1 этих же пар организмов в среднем в 3,6 раза. Таким образом, давление

стабилизирующего отбора на ГАФД-1 существенно выше, чем на ГАФД-2. Это может быть связано с участием ГАФД-1 в разнообразных негликолитических процессах и подтверждает сделанный на основе изучения коротких функциональных мотивов изоферментов ГАФД человека вывод о том, что дополнительные функции ГАФД-1 не характерны для ГАФД-2 [2]. Изофермент ГАФД-2 лучеперых рыб менее специализирован, чем у млекопитающих: по данным EST он лишен N-концевого домена и экспрессируется во многих соматических тканях. Кроме того, данный изофермент является более консервативным, чем у млекопитающих (скорость накопления точечных мутаций в ГАФД-2 лучеперых рыб выше, чем в ГАФД-2 млекопитающих в среднем в 1,7 раза), что свидетельствует о том, что он может продолжать выполнять некоторые негликолитические функции, свойственные ГАФД-1 и утраченные ГАФД-2 млекопитающих.

Тезисы доклада основаны на материалах исследований, проведенных в рамках грантов РФФИ № 06-04-48240.

Автор выражает признательность профессору, д.б.н. Муронцу В.И. за помощь в подготовке тезисов.

Литература

1. Steinke, D., Hoegg, S., Brinkmann, H. and Meyer, A. (2006). *BMC Biol.*, **4**: 16.
2. Куравский М.Л. и Муронец В.И. (2007). *Биохимия*, **72(7)**, стр. 915-922.

Компьютерная система для моделирования и анализа растительных тканей

Лавреха Виктория Вадимовна
аспирант

E-mail: vvl@bionet.nsc.ru

Институт цитологии и генетики СО РАН, Новосибирск, Россия

Пененко Алексей Владимирович
аспирант

Институт математики им. Соболева СО РАН, Новосибирск, Россия

Николаев Сергей Васильевич
научный сотрудник

Институт цитологии и генетики СО РАН, Новосибирск, Россия

В данный момент разработка программ для моделирования роста и развития растительных тканей является актуальной задачей в биологии развития растений.

Представляемая система предназначена для поддержки процесса моделирования развития двумерных клеточных ансамблей, т.е. моделирования динамики их функционального состояния, биологических и биофизических характеристик, а также

геометрии на основе представлений о морфогенезе. Одна из моделей морфогенеза основана на том, что согласованное развитие клеток управляется полями концентраций специальных химических веществ, называемых морфогенами.

Предполагаемые направления использования этой системы включают, например:

- изучение влияния на морфогенез биофизических и молекулярно-биологических процессов как по отдельности, так и в совокупности;
- моделирование процессов роста и морфогенеза плоских растительных тканей.

Методы

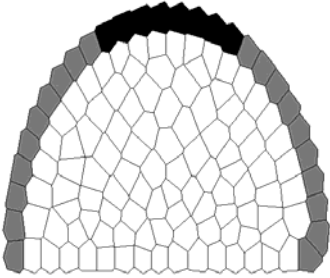
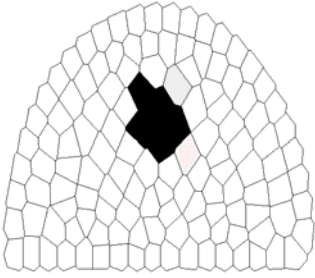
Данная компьютерная система имеет модульную архитектуру, что позволяет легко расширять её функциональность. Она включает в себя следующие модули:

- модуль моделирования процессов роста и деления клеток с возможностью регуляции экспрессии в них морфогенов и учетом влияния морфогенов на функционирование клеток, в том числе и на их дифференцировку;
- модуль расчета транспортных потоков воды, морфогенов и «питательных веществ» внутри ткани;
- модуль расчета деформации ткани на основе биомеханических моделей клетки.

«Привязка» моделей к реально наблюдаемым тканевым структурам является важным этапом верификации пригодности модели, и интерпретации экспериментальных данных. В связи с этим в систему планируется включить модуль обработки изображений, одним из результатов работы которого является построение геометрической модели реального объекта (клетки и/или ткани).

Результаты

На рисунках 1 и 2 показаны результаты работы компьютерной системы.

	
<p>Рисунок 1. Пример автоматического разбиения области на клетки.</p>	<p>Рисунок 2. Результат моделирования, качественно согласующийся с экспериментальными данными.</p>

Поиск и анализ антитерминаторных сайтов регуляции транскрипции фосфотрансферазных систем бактерий

Любецкая Анна Васильевна

студент

*Московский Государственный Университет им. М.В. Ломоносова, факультет
биоинженерии и биоинформатики, Москва, Россия*

E-mail: lyubetsky.anna@mail.ru

Один из основных механизмов импорта сахаров у бактерий являются фосфоенолпируват-зависимые фосфотрансферазные системы (ФТС), белки которых схожи у различных организмов и кодируются в одном опероне. В ряде случаев экспериментально была показана регуляция ФТС с помощью белков семейства BglG, гены которых обычно находятся непосредственно перед опероном ФТС. Регуляция обеспечивается наличием конкурирующих антитерминаторной и терминаторной шпилек в 5'-лидерной области оперона ФТС. В случае, когда в среде присутствует глюкоза, и потому бактерия не нуждается в дополнительных сахарах, ген белка-регулятора не транскрибируется, а транскрипция оперона ФТС ограничена терминаторной шпилькой. Если же клетка нуждается в импорте сахаров помимо глюкозы, то экспрессируется белок-регулятор, который взаимодействует с антитерминаторной шпилькой, располагающейся в области до терминаторной, тем самым разрушая последнюю и позволяя РНК-полимеразе синтезировать полноценный транскрипт ФТС. Такой механизм регуляции был экспериментально показан и изучен для ряда бактерий, но большее внимание уделялось изучению условий запуска системы ФТС, нежели структуре и работе антитерминаторного сайта.

В данной работе был проведен поиск и сравнительный анализ антитерминаторных сайтов у всех бактерий, в геноме которых были найдены (в том числе и автоматизированной аннотацией) опероны системы ФТС. Потенциальные сайты были найдены у ряда организмов, причем оказалось, что сайт во многих геномах повторяется два раза в изучаемых областях: структура из антитерминаторной и терминаторной шпилек обнаруживалась в лидерной области и оперона ФТС, и гена самого белка-регулятора. Структурный анализ областей выявил общее сходство шпилек, но их последовательности оказались малоконсервативными среди бактерий. Обнаружено несколько характерных структур для антитерминаторных шпилек у различных групп организмов.

Работа сделана под руководством профессора, д.б.н. Гельфанда М.С.

Литература

1. Гельфанд М.С. (2006) Цис-регуляторные структуры РНК бактерий // Молекулярная биология, 40(4), 609-619.
2. Lindner C., Hecker M., Le Coq D., Deutscher J. (2002) *Bacillus subtilis* Mutant LicT Antiterminators Exhibiting Enzyme I- and HPr Independent Antitermination Affect Catabolite Repression of the *bglPH* Operon // Journal of Bacteriology, 184(17), 4819-4828.
3. Gulati A., Mahadevan S. (2000) Mechanism of catabolite repression in the *bgl* operon of *Escherichia coli*: involvement of the anti-terminator BglG, CRP-cAMP and EIIA^{Glc} in mediating glucose effect downstream of transcription initiation // Genes to Cells, 5, 239-250.
4. Cote C.K., Honeyman A.L. (2003) The LicT protein acts as both a positive and a negative regulator of loci within the *bgl* regulon of *Streptococcus mutants* // Microbiology, 149, 1333-1340.
5. Kotrba P., Inui M., Yukawa H. (2003) A single V317A or V317M substitution in Enzyme II of a newly identified β -glucoside phosphotransferase and utilization system of *Corynebacterium glutamicum* R extends its specificity towards cellobiose // Microbiology, 149, 1569-1580.

SDPfox – Web-сервер для предсказания субстратной специфичности и поиска определяющих ее позиций белкового выравнивания

Мазин Павел Владимирович¹, Калинина Ольга Вячеславовна²

¹студент, ²(м.н.с.)

¹Московский государственный университет им. М.В. Ломоносова, факультет биоинженерии и биоинформатики, Москва, Россия

²УНЦ «Биоинформатика» ИППИ им. А.А. Харкевича РАН, Москва, Россия

EMBL-Heidelberg, Germany

E-mail: ¹iaa.aka@gmail.com, ²ok81@yandex.ru

Предсказание функции и специфичности белка по аминокислотной последовательности или по выравниванию семейства последовательностей является одной из основных задач биоинформатики, и приобрело в последние годы особую актуальность. Для выполнения белком своей биологической функции особенно важными являются остатки, которые можно условно разделить на два типа: одни отвечают за общую функцию белка и консервативны для всего семейства, другие - за специфическое распознавание взаимодействующей молекулы и, предположительно, консервативны для белков с одинаковой специфичностью (Specificity Determining Positions, SDP).

Ранее нашей группой была разработана группа методов SDPclust для поиска SDP на основании выравнивания [1]. Этот метод позволяет анализировать семейства белков, для которых отсутствует информация о специфичности отдельных представителей. SDPclust включает в себя 4 взаимосвязанных метода: SDPlight – метод для быстрого поиска SDP [2], SDPprofile – метод, осуществляющий перегруппировку последовательностей в соответствии с наилучшим сходством по предсказанным SDP [3],

SDPgroup – итеративная процедура для разбиения семейства на группы специфичности с использованием обучающей выборки (набора белков с известной специфичностью) [1], SDPtree – метод, строящий кластерное дерево предполагаемых специфичностей белков в случае, когда обучающей выборки нет, т.е. осуществляет деление семейства на группы специфичности.

В этой работе мы представляем Web-сервер SDPfox (<http://storage.bioinf.fbb.msu.ru/SDPfoxWeb/main.jsp>), реализующий методы SDPlight, SDPprofile и SDPgroup, и java-программу SDPproff реализующую все методы пакета SDPclust как в консольном, так и в визуальном режиме.

Мы провели сравнение SDPproff с другими программами и серверами осуществляющими предсказание детерминант специфичности и групп специфичности на реальных и генерированных данных. По результатам тестирования SDPproff показывает результаты по крайней мере не хуже, чем другие методы, а в ряде случаев существенно превосходит их.

Авторы выражают благодарность к.б.н. А.Б. Рахманиновой и проф. А.А. Миронову за ценное обсуждение.

Литература

1. P.V. Mazin, A.B.Rakhmaninova, O.V. Kalinina (2007) “SDPclust: a new tool for prediction protein specificity in MPA”. Moscow Conference on Computational Molecular Biology `07
2. П.В. Мазин и О.В. Калинина, “SDPlight – быстрый метод поиска позиций, определяющих специфичность в белковых выравниваниях” *Международная научная конференция студентов, аспирантов и молодых ученых “Ломоносов-2007”*, 2007
3. O.V. Kalinina A.A. Mironov, M.S. Gelfand and A.B. Rakhmaninova, “Automated selection of positions determining functional specificity of proteins by comparative analysis of orthologous groups in protein families”, *Protein Science*, 2004, pp. 13: 443-456

Роль нонсенс-кодонов в обеспечении оптимальности генетического кода

*Науменко Сергей Анатольевич
аспирант*

Институт прикладной математики им. М.В. Келдыша РАН, Москва, Россия

E-mail: sergey_clark@yahoo.co.uk

Исследование свойств генетического кода началось ещё до его окончательной расшифровки (ромбический код Гамова, код без запятых Крика). К настоящему моменту в рамках целого ряда исследований, использующих в качестве численной меры различия аминокислот показатели, связанные с гидрофобностью их боковых цепей, а в качестве

методов – генерацию множества случайных кодов со структурой вырождения, идентичной стандартной, удалось показать, что доля кодов, превосходящих стандартный по критерию минимизации расстояния между соседними кодонами по метрике кодируемых ими аминокислот, мала [1]. Обычно в подобных исследованиях положение нонсенов фиксируется, и их роль не учитывается.

В рамках данного исследования анализ оптимальности генетического кода строится на основе его простейшей модели, называемой *разметкой*, которая отражает деление множества кодонов на терминирующую и смысловую части [2]. Различие между смысловыми кодонами не вводится – существенно лишь расположение нонсенов на карте кода.

В качестве критериев оптимизации использованы вероятность появления нонсенса в результате мутации сдвига рамки считывания и количество соседних с нонсенсами смысловых кодонов. Указанные характеристики рассчитаны для всех возможных разметок с тремя нонсенсами.

Установлено, что разметка стандартного генетического кода обеспечивает максимально возможную вероятность прерывания трансляции с неправильной рамкой за счет появления нонсенса в этом случае. Кроме того, разметка обеспечивает минимально возможное количество уязвимых для точечных нонсенс-мутаций кодонов. Доля разметок с такими свойствами для случая трех нонсенов составляет всего 5,8%. По количеству возможных нонсенс-мутаций стандартная разметка не оптимальна, хотя и близка к оптимуму.

Произведено сравнение классов разметок с различным количеством нонсенс-кодонов. Класс разметок с тремя нонсенсами обладает по абсолютному значению максимально возможным количеством оптимальных по обоим критериям разметок.

Таким образом, в рамках простейшей модели и выдвинутых критериев оптимизации строго доказана оптимальность генетического кода, что служит дополнительным аргументом в пользу адапторной гипотезы его происхождения. Фиксация оптимальной разметки кода на самых ранних стадиях его эволюции, по-видимому, могла обеспечивать стабильность трансляции в условиях примитивного молекулярного аппарата.

Автор выражает признательность Фонду содействия отечественной науке за поддержку проводимых исследований.

Литература

1. S.J. Freeland, L.D. Hurst. The genetic code is one in a million. //J.Mol.Evol, 1998, 47: 238-248.
2. Г.Г. Малинецкий, С.А. Науменко, А.В. Подлазов. Об экстремальных свойствах разметки генетического кода. //Доклады академии наук (биохимия, биофизика, молекулярная биология) июнь 2007, том 414, № 6, стр. 831-835.

Изучение эволюции паралогичных семейств генов человека

Панчин Александр Юрьевич

Студент

*Московский Государственный Университет Им. М.В. Ломоносова, Факультет
Биоинженерии и Биоинформатики, Москва, Российская Федерация.*

E-mail: alexpanchin@yahoo.com

Изучение эволюции генетических последовательностей играет важную роль в современной биологии. При изучении эволюции участков генов, кодирующих белки, часто рассматриваются две величины: количество синонимичных (K_s) и количество несинонимичных (K_n) замен, приходящихся на сайт. Предполагается, что на синонимичные сайты, мутации в которых не влияют на аминокислотную последовательность кодируемого белка, отбор действует незначительно. Поэтому количество синонимичных замен принято считать своего рода “молекулярными часами”, по которым можно оценить время расхождения между двумя последовательностями. Мутации в несинонимичных сайтах могут приводить к изменениям кодируемого белка, поэтому они в большей степени подвержены естественному отбору. Соотношение K_n/K_s часто, используется для определения силы и направления отбора. Так $K_n/K_s = 1$, характеризует эволюционную нейтральность, $K_n/K_s \gg 1$ положительный отбор (направленный на сохранение новых полезных мутаций), а $K_n/K_s \ll 1$ отрицательный отбор (направленный на сохранение важных консервативных позиций). Известно, что темпы эволюции гена зависят от того, является ли он уникальным в геноме, или имеет паралоги. Оказалось, что при сравнении генов близких видов эукариот, значения K_n/K_s больше у уникальных генов, чем у генов, представленных в геноме несколькими копиями. Можно допустить, что дублируются главным образом гены, несущие важные функции, а так же гены, эффект которых зависит от их копияности, и поэтому, с самого начала подверженные сильному отрицательному отбору. Однако, K_n/K_s у генов дублировавшихся недавно, по-видимому, больше, чем у дублировавшихся давно – сам факт дубликации временно снижает силу отрицательного отбора на последовательности. В данной работе исследовалась эволюция семейств

дуплицированных генов человека, с целью выявить глобальные закономерности между временем, прошедшим с момента дупликации индивидуальных представителей генных семейств, и темпами их эволюции.

При выявлении недавно дуплицировавшихся генных семейств мы использовали геномное сходство генных последовательностей. Для построения выравненных кодирующих последовательностей использовалась программа ClustalW. Для получения попарных значений K_n и K_s между членами семейства использовался пакет программ Pam1. Была написана программа, считающая индивидуальные K_n и K_s для отдельных представителей семейств. Индивидуальные K_n и K_s — характеризуют число синонимичных и несинонимичных замен на сайт, происшедших в данной последовательности с момента последней дупликации.

В изученных семействах в целом же наблюдается тенденция уменьшения индивидуального K_n/K_s по мере увеличения времени от последней дупликации. Кроме того, получены данные, свидетельствующие о том, что некоторые паралоги внутри одного семейства эволюционируют асимметрично: ситуации, когда один из двух ближайших генов паралогов накапливает несинонимичные замены быстрее, чем другой, не редки. Интересным оказалось обнаружение консервативных синонимичных сайтов, в некоторых семействах. Это говорит о том, что отбор может вызывать предпочтение одного синонимичного кодона другому. Это наблюдение, а так же обнаруженная асимметричность индивидуальных K_s для многих ближайших паралогов заставляет пересмотреть представления о точности K_s , как правильно идущих «молекулярных часах».

Работа выполнялась совместно с научным руководителем Иреной Игоревной Артамоновой. Автор выражает благодарность М.С. Гельфанду и А.А. Миронову за обсуждение данной работы.

Изучение консервативности межгенных участков ДНК прокариот

Пятницкий Михаил Алексеевич
аспирант

ГУ НИИ биомедицинской химии им В.Н.Ореховича, РАМН, Москва, Россия

E-mail: mpyat@mail.ru

Хорошо известно, что даже некодирующие участки ДНК могут иметь высокие уровни консервативности из-за существования давления естественного отбора. Свойства консервативных межгенных участков широко исследованы у эукариот, однако для

прокариот эта область еще не достаточно изучена. В опубликованных работах анализируются повторы, периодичность, алгоритмы для поиска некодирующих РНК [1-3]. Целью настоящей работы было изучение характера эволюции и оценка доли консервативных позиций в некодирующих участках ДНК прокариот.

Нами разработано несколько вычислительных подпрограмм для изучения характера консервативности межгенных участков. На первом этапе ищутся все консервативные пары генов, т.е. пары соседних генов которые имеют ортологов во всех анализируемых геномах. Для поиска ортологичных генов нами был реализован наиболее распространенный алгоритм двунаправленного наилучшего совпадения. Результаты представляются в виде графа, где вершины (гены из всех изучаемых геномов) соединены ребром в том случае, если они являются наилучшим взаимным совпадением. Поиск клик в таком графе дает наборы взаимных ортологов, которые присутствуют во всех рассматриваемых геномах. В этом смысле полученные наборы генов аналогичны таковым из базы данных COG. Для каждой пары клик был проведен поиск генов которые имели одинаковую ориентацию. Участки ДНК между отобранными таким образом пар генов были выровнены с помощью программы MUSCLE.

Мы предложили простую математическую модель позволяющую оценить долю позиций, находящихся под действием отбора. В качестве исходных данных используется информация об энтропии каждой позиции выравнивания для ряда штаммов двух близкородственных видов. Подобный подход выявляет некоторые интересные паттерны эволюции межгенных участков.

Все программное обеспечение было реализовано как набор скриптов на языках Perl и R. Использовалась система управления базами данных MySQL. Все исходные коды доступны по запросу.

Эта работа была проведена совместно с М.С.Гельфандом.

Литература

1. N.Rajewsky, N.D. Succi, M.Zapotocky, and E.D. Siggia, "The Evolution of DNA Regulatory Regions for Proteo-Gamma Bacteria by Interspecies Comparisons", *Genome Res*, 12, 2002, pp. 298-308.
2. L.Wilson and P.Sharp, "Enterobacterial repetitive intergenic consensus (ERIC) sequences in Escherichia coli: Evolution and implications for ERIC-PCR", *Mol Biol Evol.*, 23(6), 2006, pp. 1156-1168.
3. S.Hosid, E.N.Trifonov and A.Bolshoy, "Sequence periodicity of E.coli is concentrated in intergenic regions", *BMC Mol Biol.*, 5, 2004, pp. 1-14.

Предсказание и анализ тонкой структуры сайтов с-белков бактериальных систем рестрикции-модификации

Сорокин Валерий Андреевич¹, Гельфанд Михаил Сергеевич²

аспирант¹, д.б.н., к.ф.-м.н., профессор²

*Московский Государственный Университет им. М.В. Ломоносова, Москва, Россия¹,
Институт Проблем Передачи Информации РАН, Москва, Россия²*

gtr-sorokin@mail.ru¹, gelfand@iitp.ru²

Типичная система рестрикции-модификации работает по принципу «токсин-антитоксин» и состоит как минимум из двух ферментов: метилазы и рестриктазы. Рестриктаза узнает специфичную для данной системы последовательность ДНК и вносит в неё разрыв. Метилаза призвана маскировать такие последовательности в хозяйской ДНК, предотвращая тем самым гибель клетки-носителя. Чужеродная ДНК содержит не маскированные специфичные последовательности, что вызывает её деградацию рестриктазой. Часто система рестрикции-модификации содержит третий элемент - С-белок (от англ. control protein), который является регулятором транскрипции рестриктазы и метилазы, а также регулирует собственную транскрипцию. Данная работа посвящена исследованию регуляторных сайтов в окрестности старта транскрипции С-белков.

Отправной точкой для работы послужили 46 известных С-белков, содержащихся в REBASE. Для каждого С-белка был выполнен поиск родственных последовательностей, которые рассматривались как потенциальные С-белки. В областях 5' от генов предполагаемых С-белков выделялись «островки консервативности», которые могли бы являться регуляторными сайтами. Для 70% предполагаемых генов С-белков удалось предсказать регуляторный сайт. В работе активно использовались вспомогательные автоматические процедуры (blast, hmmer, muscle), но в конечном итоге все результаты опираются на ручной анализ.

В результате проведённого исследования было обнаружено около 170 новых С-белков, что позволило существенным образом дополнить имеющуюся выборку из 46 С-белков, содержащихся в REBASE. Для каждого нового С-белка в окрестности старта транскрипции был предсказан регуляторный сайт. Многократное увеличение выборки предсказанных регуляторных сайтов позволило провести их анализ на качественно новом уровне. Имеющаяся на сегодняшний день классификация из трёх классов сайтов была дополнена ещё семью значимо отличающимися классами.

Наличие «слабых» копий сайтов в окрестности старта транскрипции С-белка подтверждает гипотезу о кооперативном связывании С-белков с ДНК. Исследование локусов, содержащих гены С-белков, позволило описать несколько новых систем рестрикции-модификации и показало, что системы рестрикции-модификации часто находятся в окружении из фаговых генов. В ряде систем в окрестности старта транскрипции метилазы были обнаружены последовательности, похожие на предсказанные сайты С-белков, что согласуется с представлениями о регуляции транскрипции метилазы С-белком.

Литература

1) Roberts, R.J. (2005) REBASE – restriction enzymes and DNA methyltransferases // Nucleic Acid Research, Vol.33, database issue, p. 230-232.

2) Blumenthal R.M. (2007) Regulatory circuit based on autogenous activation-repression: roles of C-boxes and spacer sequences in control of the PvuII restriction-modification system // Nucleic Acid Research, Vol.35, №20, p. 6935–6952

Влияние мобильных и экстрахромосомных элементов на эволюцию бактериального генома

*Сорокин Максим Игоревич
студент второго курса*

*Московский государственный университет им. М.В. Ломоносова, факультет
Биоинженерии и Биоинформатики, Москва, Россия*

E-mail: sorokin.maks@gmail.com

Большое количество полностью отсеквенированных бактериальных геномов (350 на данный момент) позволяет исследовать влияют ли мобильные элементы бактериального генома на его структуру. Было решено исследовать зависимость между частотой мобильных и экстрахромосомных элементов генома и его структурными характеристиками (размер, GC-состав, количество генов из разных функциональных категорий, общее количество предсказанных генов).

Для решения поставленной задачи была проанализирована база данных TIGR (The Institute for Genomic Research) с использованием языков программирования R и PERL.

Была обнаружена прямая зависимость размера генома от числа генов в нем, а также от его GC-состава, что согласуется с литературными источниками и, таким образом, показывает релевантность использованного нами подхода. Было показано, что увеличение размера генома обусловлено в основном быстрым возрастанием числа

неконсервативных генов, например, выполняющих функцию регуляции белковых взаимодействий или регуляции продукции токсинов, и в значительно меньшей степени - консервативных генов, например, ответственных за центральный метаболизм ДНК.

Было показано, что число горизонтально перенесенных генов не зависит от размера генома бактерии. Однако это число отрицательно коррелировало с количеством плазмид, свойственному этому виду бактерий. Возможно, это связано с тем, что плазмиды и хромосомальные горизонтально перенесенные гены – два альтернативных источника геномных инноваций. Тем не менее, была обнаружена обратная зависимость между числом плазмид и GC-составом.

Работа выполнена при поддержке гранта РФФИ № 08-04-01651-а.

Поиск вторичных структур РНК в интронах представителей рода *Drosophila*

Храмеева Екатерина Евгеньевна, Миронов Андрей Александрович, Гельфанд Михаил Сергеевич, Первушин Дмитрий Давидович
студентка, профессор, профессор, старший преподаватель
Московский государственный университет им. М.В. Ломоносова, Москва, Россия
E-mail: khrameeva@yandex.ru

Известно, что у многих организмов вторичная структура РНК оказывает значительное влияние на процессы транскрипции и трансляции, однако влияние таких структур на процесс сплайсинга до сих пор систематически не изучено. В настоящее время экспериментально показано наличие порядка 15 случаев участия вторичной структуры в регуляции сплайсинга у различных организмов (дрожжи, растения, насекомые и позвоночные). Также прослеживается связь вторичных структур с такими патологиями человека, как мышечная дистрофия, кистозный фиброз и паркинсонизм. Помимо экспериментальных методов, для изучения вторичных структур применяются различные биоинформатические подходы.

В данной работе был осуществлен сравнительный анализ интронов восьми представителей рода *Drosophila* на предмет наличия в них консервативных комплементарных мотивов, способных формировать стабильные РНК-структуры. При этом рассматривались окружения донорных и акцепторных сайтов сплайсинга, содержащие 200 соседних нуклеотидов интрона и 50 соседних нуклеотидов экзона. По последовательностям окружений соответствующих сайтов сплайсинга из разных организмов были построены множественные выравнивания. В каждом таком множественном выравнивании был осуществлен поиск консервативных участков длиной 9 нуклеотидов (с одной разрешенной неконсервативной позицией). Далее

осуществлялась проверка пар консервативных участков из донорного и акцепторного окружения на комплементарность, допуская GT пару на 9 нуклеотидов. Кроме того, требовали присутствия хотя бы двух GC пар на 9 нуклеотидов, чтобы избежать нахождения AT-богатых последовательностей.

Таким образом, получили набор из 52 интронов, для которых в окружении донорных и акцепторных сайтов найдены консервативные комплементарные участки длиной 9 нуклеотидов. Статистическая значимость этого набора интронов была оценена при помощи процедуры рандомизации и составила около 11%. Кроме того, было замечено, что для большинства интронов данные участки можно продлить, при этом наблюдается схожая картина: исходный 9-нуклеотидный участок служит «затравкой», от которой расходятся в различной степени протяженные участки вторичной структуры.

Для полученного набора из 52 интронов был осуществлен поиск альтернативных вторичных структур, т.е. таких, в которых кроме уже найденных комплементарных участков (А и В) имеется третий участок (С), также комплементарный А. При этом требовалось, чтобы альтернатива С имела длину не менее 7 нуклеотидов полностью перекрывалась с участком А по комплементарности (с одной разрешенной GT парой). Вместо консервативности было наложено требование, чтобы альтернативы присутствовали как минимум в 5 организмах. С такими ограничениями наличие альтернатив было показано для 35 из 52 интронов. Мы предполагаем, что в найденных генах регуляция сплайсинга осуществляется при помощи альтернативных вторичных структур. В качестве дальнейшей задачи стоит экспериментальная проверка данного утверждения.