

СЕКЦИЯ «БИОИНЖЕНЕРИЯ И БИОИНФОРМАТИКА»

ПОДСЕКЦИЯ «БИОИНФОРМАТИКА»

Захват и высвобождение кодирующей ДНК: эволюция бактериальных генов путем изменения положения стоп-кодонов¹

Вахрушева А.А.², Казанов М.Д.*

*студентка, к.б.н.**

Московский государственный университет им. М.В. Ломоносова, Факультет

Биоинженерии и Биоинформатики,

ИППИ РАН, Москва, Россия*

E-mail: avakhrusheva@gmail.com

Пути возникновения кодирующей нуклеотидной последовательности *de novo* остаются одним из туманных вопросов молекулярной эволюции. Один из возможных путей включения и исключения сегментов ДНК из состава гена – смещение положения стоп-кодона. Точечные нуклеотидные замены могут создавать преждевременный стоп-кодон (нонсенс-мутация) или уничтожать текущий стоп-кодон, что приведет к продолжению трансляции до следующего стоп-кодона в рамке считывания гена.

Цель работы – описание эволюции кодирующей последовательности генов бактерий, связанной с изменением положения сайта терминации трансляции. Мы исследовали методами сравнительной геномики семейства гомологичных генов, полученные из 623-х полных бактериальных геномов. Для каждого семейства были осуществлены выравнивания нуклеотидных последовательностей всех генов семейства. В этих выравниваниях мы проанализировали все случаи несовпадения положения стоп-кодона между разными генами семейства. Мы заключали, что положение стоп-кодона изменилось с захватом или высвобождением геном участка кодирующей ДНК, если кодирующая последовательность перед стоп-кодом одного гомолога очень хорошо выравнивалась с некодирующей последовательностью после стоп-кодона другого гомолога. Эти мутации поляризовались с использованием филогенетического анализа методом наибольшей экономии, что давало нам возможность отличать эволюционные приобретения от потерь С-концевого сегмента кодируемого белка.

В результате проведенного анализа были выявлены как случаи потери концевого участка белка и преобразования кодирующей последовательности в сегмент 3'UTR в результате зафиксированной нонсенс-мутации, так и случаи приобретения новой С-концевой аминокислотной последовательности в результате точечной мутации стоп-кодона. Часть наблюдаемых случаев не может объясняться ошибками секвенирования, поскольку и короткая, и длинная форма наблюдалась в нескольких вариантах в различных бактериальных геномах.

Полученные результаты свидетельствуют о том, что положение стоп-кодона эволюционно лабильно. Точечная мутация стоп-кодона – простой эволюционный путь к приобретению новой кодирующей последовательности. Наличие выравниваний большого количества ортологов из геномов разной степени филогенетического родства позволит нам исследовать закономерности дальнейшего эволюционного преобразования сегмента ДНК после его включения в состав гена.

1 Работа поддержана грантом РФФИ 08-04-01394.

2 Автор выражает признательность к.б.н. Базыкину Георгию Александровичу за помощь в подготовке тезисов.

Разработка микроматриц ДНК для генетического тестирования человека при проведении экстракорпорального оплодотворения

Виноградова С.В.

студентка

Московский государственный университет им. М.В. Ломоносова, Факультет Биоинженерии и Биоинформатики, и Центр «Биоинженерия» РАН, Москва, Россия

E-mail: kintany@gmail.com

В настоящее время экстракорпоральное оплодотворение (ЭКО) является одним из наиболее эффективных методов лечения бесплодия. Прогнозирование выживаемости эмбрионов является неотъемлемой частью данного подхода. Существующие методы анализа генома эмбриона (мультиплексный PCR, FISH) не отвечают полностью требованиям к диагностике. Применение многопараметрических методов с использованием микроматриц ДНК позволяет детектировать анеуплоидию, большие хромосомные делеции и дупликации, а также определять пол.

Целью работы была разработка микроматриц ДНК и подтверждение эффективности их использования.

Для создания микроматриц ДНК использовали ВАС-клоны из геномной библиотеки человека RP11. Библиотека содержит 543'797 клонов, средняя длина клона – 178'000 bp. Так как задачей было детектирование больших хромосомных перестроек, то достаточно разрешения 10^6 bp, т.е. необходимо было отобрать около 1000 клонов.

Отбор клонов производился с помощью Java-скриптов, так чтобы клоны были равномерно распределены по всем хромосомам. Для нанесения на микроматрицы использовали только клоны, не содержащие большого количества повторов – простых и инвертированных (для детекции повторов использовался сервер Genome Table Browser).

Для анализа образцов ДНК проводили гибридизацию. Для этого опытный и контрольный образцы ДНК методом ПЦР амплифицировали с вырожденными праймерами, с добавлением aa-dUTP. Далее продукты ПЦР в ходе химической реакции метили монореактивными эфирами Су3 и Су5 для получения флуоресцентно меченой ДНК. На одно гибридизационное поле микроматрицы наносили опытный и контрольный образцы, меченые двумя различными флуоресцентными красителями одновременно; далее использовали метод «переключения красок» (dye-swap).

После проведения гибридизации микроматрицы отмывали от неспецифического связывания и сканировали по двум каналам флуоресценции. Анализ интенсивности сигналов и изображения проводили в программе BlueFuse.

Для оценки эффективности данного подхода проводили гибридизацию с ДНК из клеточной линии НЕК293. Линия была заранее протестирована на микроматрицах высокой плотности, чтобы составить подробную карту хромосомальных перестроек. Это дало возможность оценки результатов, полученных с использованием микроматриц.

В результате гибридизации было получено, что микроматрицы верно воспроизводят картину хромосомальных перестроек с разрешением не менее 10^6 bp. Кроме того, они способны определять «пол» анализируемой ДНК.

Исследования конформационного пространства амилоидогенного пептида A β 1-42 с использованием программного пакета MaxFolder¹**Гармай Ю.П.¹, Захаров Г.А.²**¹аспирант, ²аспирант¹Санкт-Петербургский Институт Ядерной Физики РАН, Санкт-Петербург, Россия.²Институт Физиологии им. И.П. Павлова, РАН, Санкт-Петербург, Россия.

E-mail: yuri.garmay@mail.ru

В настоящий момент биоинформатика и компьютерная биология являются активно развивающимися разделами науки. Одной из важных задач компьютерной биологии является поиск оптимальной конформации различных биологических макромолекул. Теоретически, эта задача сводится к нахождению глобального минимума энергии молекулы, причём функция энергии зависит от огромного числа параметров и имеет множество локальных минимумов.

Большинство методов поиска конформаций используют либо алгоритм полного перебора значений варьируемых параметров, либо случайный выбор с сохранением оптимальных значений. Фактически оба метода основаны на грубой силе (brutforce-алгоритм) и не могут быть признаны оптимальными.

Методы поиска глобального минимума функции делятся на детерминированные и вероятностные. Реализовать детерминированный метод для поиска глобального минимума энергии большой молекулы сложно в силу большой вычислительной нагрузки. Метод генетических алгоритмов считается одним из оптимальных вероятностных методов поиска конформации.

В рамках дипломной работы одного из авторов был создан программный пакет MaxFolder, реализующий данный алгоритм для пептидов. Разработанный программный пакет был использован для исследования возможных конформаций амилоидогенного пептида A β 1-42 и его участка A β 12-28. Полученные конформации использовались для проведения молекулярно-механического моделирования.

Показано, что подавляющее большинство конформаций, полученных с помощью пакета MaxFolder близки к равновесной (релаксированной) конформации, найденной путем проведения расчета молекулярной динамики пептида в водном окружении. Большинство конформаций представляют собой разновидность β -поворота; чаще всего участком поворота служили а.к. Val²⁴ и Gly²⁵, что соответствует литературным данным о вероятностях образования β -поворотов. При использовании полученных с помощью MaxFolder конформаций в качестве стартовых, для достижения равновесной конформации требовалось значительно меньше времени, чем для получения равновесной конформации из линейного пептида. Таким образом, применение программного пакета MaxFolder позволяет существенно снизить объем вычислений, требуемый для получения релаксированной конформации пептида.

В некоторых случаях оптимальной найденной конформацией для A β 12-28 являлась развернутая цепь без образования определенной вторичной структуры. Однако по результатам молекулярно-механического моделирования стабильность такой структуры ниже, чем β -поворота, а время ее перехода в равновесную конформацию при молекулярно-механическом моделировании больше.

α -спирали в множестве полученных конформаций зафиксированы не были.

В дальнейшем авторами предполагается оптимизировать параметры генетического алгоритма для более точного нахождения оптимальных конформаций и исключения ложных результатов, а также адаптация разработанного пакета для работы с другими классами органических веществ.

1 Авторы выражают признательность к.ф.-м.н. Исаеву-Иванову В.В. и к.б.н. Щеголеву Б.Ф. за помощь в выполнении работы.

Эволюция генома бактерий родов *Yersinia* и *Erwinia***Глотова И.В.**

Студентка

Московский государственный университет имени М.В.Ломоносова, Факультет
биоинженерии и биоинформатики, Москва, Россия

E-mail: igbox@mail.ru

На сегодняшний день секвенировано большое число различных бактериальных геномов. Некоторые из этих геномов принадлежат близкородственным штаммам, как правило, патогенам. Периферийные гены в геномах близкородственных штаммов (т.е. гены, которые не присутствуют во всех штаммах) [1] являются интересным материалом для исследования эволюции геномов, функциональных подсистем и метаболических путей.

Мы провели исследование ортологических рядов близкородственных бактериальных штаммов рода *Yersinia* (виды *Y. pestis*, *Y. enterocolitica* и *Y. pseudotuberculosis*) и ортологических рядов множества этих штаммов и двух видов рода *Erwinia* (*E. carotovora* и *E. tasmaniensis*). Каждый ряд представляет собой набор ортологичных и/или паралогичных генов. В результате получены группы универсальных (присутствуют во всех рассматриваемых геномах) [1] и периферийных генов для бактериальных геномов рода *Yersinia* и *Erwinia*.

Анализ аннотированных ортологических рядов метаболической периферии показал, что в ней преобладают ряды, связанные с метаболизмом и транспортом сахаров, а также с симбиозом и путями переноса сигнала. Установлено, что приобретение генов транспорта и метаболизма сахаров в геномах штаммов рода *Yersinia* в основном происходило не на уровне предковой формы *Y. enterocolitica*, а в геномах более молодых видов *Y. pestis* и *Y. pseudotuberculosis*. Наибольшее соотношение потерянных к приобретенным генам наблюдается для вида *E. tasmaniensis*. Анализ потери/приобретения периферийных генов с различными функциями позволил определить межродовые и межвидовые различия в метаболических путях отдельных бактериальных геномов, в частности в путях сахарного метаболизма.

Литература

1. Duccio Medini *et al.* (2008) Microbiology in the post-genomic era. *Nature* 6: 419-430.

Анализ и аннотация аминокислотных остатков на интерфейсе димеризации мегануклеаз семейства LAGLIDADG.**Гришин А.В.**

Студент

ФГОУ ВПО РГАУ – МСХА им. К. А. Тимирязева, Агрономический факультет, Москва, Россия

E-mail: grishin-a1@yandex.ru

Мегануклеазами (хoming эндонуклеазами, homing endonucleases) называют эндонуклеазы, узнающие длинные последовательности ДНК (14–40 п.н.) и вносящие в найденном сайте двунитевой разрыв. Последовательности такой длины уникальны или встречаются в геноме малое число раз, в связи с чем представляется возможным использование подобных ферментов с целью создания высокоспецифичных эндонуклеаз с предопределенной специфичностью.

Все хoming эндонуклеазы могут быть поделены на 4 семейства (LAGLIDADG, HNH, His-Cys box и GIY-YIG); из них наиболее многочисленным и широко изученным является семейство LAGLIDADG. Последовательности, кодирующие хoming эндонуклеазы семейства LAGLIDADG встречаются в интронах I типа и интеинах бактерий, археобактерий и простейших эукариот. Нуклеазы этого семейства характеризуются наличием двух LAGLIDADG доменов, каждый из которых ответственен за узнавание половины рестрицируемого участка ДНК. Главную роль во взаимодействии двух LAGLIDADG доменов играет первая α -спираль каждого домена, содержащая единственный консервативный для этих доменов мотив LAGLIDADG (отсюда название семейства). Экспериментально была показана (Silva, 2004) возможность комбинирования отдельных LAGLIDADG доменов из различных белков и получения таким образом мегануклеаз с новой специфичностью.

При анализе структур 11 хoming эндонуклеаз семейства LAGLIDADG нами были выявлены наиболее важные для димеризации доменов аминокислотные остатки, являющиеся первоочередной мишенью для инжиниринга при комбинировании LAGLIDADG доменов различных белков. Также были проанализированы последовательности всех LAGLIDADG доменов, найденных в базе данных PFAM, построено множественное выравнивание консервативных участков этих последовательностей, а также дерево, основанное на этом выравнивании. Полученные данные были использованы в дальнейшем для поиска закономерностей распределения аминокислотных остатков в наиболее важных для димеризации доменов позициях.

Альтернативным подходом при создании высокоспецифичных мегануклеаз с предопределенной специфичностью является инжиниринг аминокислотных остатков, непосредственно участвующих в узнавании и связывании ДНК. При таком подходе подходящим объектом являются LAGLIDADG хoming эндонуклеазы, закодированные в интеинах, поскольку они имеют дополнительный ДНК-связывающий домен. Наиболее широко изученным подобным белком является нуклеаза PI-SceI. Для выявления наиболее важных для узнавания ДНК аминокислотных остатков было проведено сравнение структурных данных, доступных для этого белка, и экспериментальных данных по прямому мутагенезу как самого белка PI-SceI, так и сайта его связывания (Gimble, 1996; He, 1998).

Литература

1. Paques F. (2007) Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy // *Current Gene Therapy*, 7, 49-66.
2. Guhan N. (2003) Structural and functional characteristics of homing endonucleases // *Critical Reviews in Biochemistry and Molecular Biology*, 38(3):199–248.

Статистическое исследование водородных связей в структурах белков**Денисенко Е.В.¹***Студентка**Московский Государственный Университет имени М.В. Ломоносова, Факультет
биоинженерии и биоинформатики, Москва, Россия**E-mail: lesiad@ya.ru*

В работе при помощи компьютерных методов исследуются закономерности образования водородных связей между атомами белка. Работа преследовала две основных задачи: уточнение пространственных параметров взаимодействия и анализ предпочтений во взаимодействии разных типов атомов, входящих в состав белков.

Для решения первой из поставленных задач – определения параметров взаимодействий между атомами – была написана компьютерная программа, извлекающая необходимую информацию из заданных PDB-файлов.

Программа принимает на вход список исследуемых файлов формата PDB, названия типов атомов белка, взаимодействие между которыми требуется изучить, а также ряд технических параметров, в том числе максимальное расстояние между взаимодействующими атомами. Рассматриваются всевозможные пары атомов, удовлетворяющие заданным ограничениям, из всех заданных PDB-файлов; данные о взаимном расположении атомов выдаются в виде таблицы. Фактически для каждой пары атомов выдается пять чисел, являющихся координатами атомов друг относительно друга в полярных системах координат, связанных с данными атомами.

Программа также выдает данные о распределении количества пар взаимодействующих атомов в зависимости от значений полярных координат. В частности, для исследования распределения расстояний между взаимодействующими атомами рассматривались интервалы расстояний от 0 до 3,7 Å с шагом в 0,1 Å. Таким образом, для каждого из 37 интервалов программой выдается общее число пар атомов, удаленных друг от друга на соответствующее расстояние, а также относительное число таких пар. Относительное число рассчитывается как частное общего числа атомов и объема области пространства, ограниченной данным интервалом расстояния. Полученные данные о распределении значений параметров взаимного расположения атомов предполагается использовать для вывода эмпирического потенциала водородной связи.

Для решения второй задачи был проведён сбор статистики по различным типам атомов. Для этого из выборки структур негомологичных белковых цепей PDBselect25 (http://bioinfo.tg.fh-giessen.de/pdbselect/recent.pdb_select25) были взяты все 746 цепей, чьи структуры были получены методом рентгеноструктурного анализа с разрешением не хуже 1,5 Å.

Атомы кислорода и азота различных аминокислотных остатков были разбиты на 11 групп по химическим свойствам. Для каждой пары групп исследовалось распределение расстояний между атомами – их представителями. По такому распределению можно судить о наличии или отсутствии водородной связи между атомами данных групп. Была получена таблица, демонстрирующая, атомы каких групп могут образовывать водородную связь.

В перспективе предполагается провести аналогичный анализ взаимодействия атомов белка с атомами нуклеиновых кислот.

1 Автор выражает признательность руководителям работы к.ф.-м.н. Спирину С. А. и к.ф.-м.н. Алексеевскому А. В. за помощь в подготовке тезисов.

Предсказание гибких участков в белках**Дерюшева Е.И.**

Аспирант

Тульский государственный университет,
естественно-научный факультет, Тула, Россия

E-mail: janed1986@yandex.ru

В последнее время обнаружено, что многие белки не обладают уникальной третичной структурой, хотя и имеют четкую функцию при физиологических условиях. Такие белки принято называть белками с внутренней неупорядоченностью. Доля неупорядоченных областей в белках может быть разной, начиная от последовательности из нескольких аминокислот и заканчивая полностью неупорядоченной последовательностью длиной в десятки, а иногда и в сотни аминокислот. Поскольку развернутые участки белковой цепи играют важную роль в процессе функционирования белка, то их предсказанию уделяется большое внимание. На сегодня для этих целей разработаны специализированные программы, такие как PONDR, RONN, DisEMBL, PreLINK, IUPred, GlobPlot 2, FoldIndex.

В Институте белка РАН для предсказания гибких областей (петель) в белках была разработана программа FoldUnfold. Изначально программа была разработана для поиска полностью разупорядоченных белков или белков с большой долей неупорядоченности. Однако было показано, что использование дополнительной опции программы – выбор ширины окна, позволяет более детально исследовать белок по всей длине. Так ширина окна в 3 а.о. позволяет выявить почти все короткие участки (повороты, шпильки, петли), соединяющие элементы вторичной структуры.

Построение зависимости фактора Дебая – Валлера от номера остатка для группы белков показало, что участки, характеризующиеся высокой величиной фактора предсказываются при ширине окна в 11–13 а.о. и в основном являются функциональными. Другие, имеющие более низкие величины факторов Дебая – Валлера, являются элементами, соединяющими вторичную структуру. Такие петли имеют существенно более низкую подвижность и обеспечивают дополнительную жесткость трехмерной структуры. С помощью программы FoldUnfold была решена еще одна, трудная для других программ задача: определение границы между структурированной и неструктурированной областями в белках с большой долей неупорядоченности. Так, для убиквитин подобного домена определена граница между структурированной и неструктурированной областями в районе 30–31 а.о., тогда как каждая из других программ приводит к границе, колеблющейся от 28 до 70 а.о.

Кроме того, программа FoldUnfold была использована при теоретическом структурно-функциональном исследовании двух тканеспецифичных форм эукариотического фактора элонгации трансляции eEF1A. Было выявлено наличие дополнительной петли у одной из форм фактора, а также присутствие на С-конце полипептидной цепи длинной неструктурированной области. Для рибосомального белка S1 было выявлено присутствие устойчивого домена в многодоменных белках.

Литература

1. Дерюшева Е.И., Галзитская О.В., Сердюк И.Н. (2008) Предсказание коротких петель в белках с внутренней неупорядоченностью // Молекулярная биология, № 42, стр. 1067–1078.
2. Galzitskaya O.V., Garbuzynskiy S.O., Lobanov M.Yu. (2006) FoldUnfold: web server for the prediction of disordered regions in protein chain // Bioinformatics, №22(23), p. 2948-2949.
3. Galzitskaya O.V., Deryusheva E.I., Serdyuk I.N. (2008) Phylogenetic Analyses of the Loops in Elongation Factors EF1A: Stronger Support for the Grouping of Animal and Fungi // JCSB, №1, p. 73-80.

Построение нового профиля для поиска гистидиновых киназ, отвечающих за кворум-сенсинг у бактерий из филума *Firmicutes*

Диброва Д.В.

Студентка

Московский государственный университет имени М.В.Ломоносова, Факультет
Биоинженерии и Биоинформатики, Москва, Россия
e-mail: udavdasha@gmail.com

Среди белков, отвечающих у бактерий за прием сигнала из внешней среды и ответ на него, наиболее важными являются белки двухкомпонентных систем (ТСs – Two-Component Systems). Каждая такая система состоит из двух частей: сенсорного белка (которым является гистидиновая киназа, НК – Histidine Kinase) и белка, осуществляющего ответ клетки (его называют регулятором ответа, RR – Response Regulator). Принцип работы системы состоит в том, что в ответ на стимул НК автофосфорилируется по консервативному гистидиновому остатку, после чего передает фосфат на RR, а он в свою очередь активируется и осуществляет ответ. Часто RR является транскрипционным фактором, и ответ заключается в изменении экспрессии тех или иных генов. НК могут иметь различную доменную структуру, но всегда содержат два домена: АТФазный и содержащий фосфорилируемый гистидин (His-box).

Для поиска консервативных доменов обоих компонент ТСs существует ряд профилей в базе данных Pfam. Однако оказывается, что эти профили позволяют найти не все НК. Сравнение аннотаций белков двухкомпонентных систем в базе данных RefSeq с результатами поиска доменов Pfam показало наличие семейства белков, которые описаны как гистидиновые киназы, но не обнаруживаются при поиске. Среди этих белков был хорошо описанный в литературе как НК стафилококковый белок agrC, являющийся частью системы кворум-сенсинга [1].

По этим белкам был построен новый профиль. Детальный анализ показал, что хиты нового профиля обладают следующими свойствами, позволяющими утверждать, что они действительно являются НК:

- 1) У них имеется строго консервативный гистидин и характерная область вокруг него, включающая консервативные и вариабельные позиции;
- 2) На N-конце у них предсказываются несколько трансмембранных сегментов (чаще всего 7);
- 3) Их ближайшие соседи на геномах – гены RR;
- 4) Они содержатся в геномах бактерий филума *Firmicutes*, для которых и описана система кворум-сенсинга, включающая agrC-подобные белки;
- 5) Устройство их АТФазного домена на C-конце особое и полностью отвечающее описанию семейства гистидиновых киназ кворум-сенсинга НРК₁₀ [2].

Предполагается использовать построенный профиль при анализе ТСs бактерий из рода *Paenibacillus*, входящего в филум *Firmicutes*, геномы которых были недавно секвенированы. Известно, что эти бактерии могут в особых условиях образовывать колонии в виде разнообразных сложных узоров, что свидетельствует о развитой системе межклеточной коммуникации.

Литература

- [1] Quorum sensing in Staphylococci, Richard P. Novick and Edward Geisinger, Annu. Rev. Genet., 42: 541-64 (2008).
- [2] The histidine protein kinase superfamily, Grebe T.W. and Stock J.B., Adv. Microb. Physiol., 41: 139-227 (1999).

Поиск мотивов связывания белков-регуляторов семейства MerR**Жаров И.А.1**

Студент

Московский физико-технический институт, факультет радиотехники и кибернетики,
Москва, Россия

E-mail: peshwalk@mail.ru

Известно, что у *Bacillus subtilis* экспрессия генов *bmr* и *blt*, кодирующих белки-транспортеры, обеспечивающие множественную лекарственную устойчивость, регулируется белками BmrR и BltR соответственно. Эти регуляторы транскрипции относятся к семейству факторов транскрипции MerR. Механизм регуляции и сайты связывания этих регуляторных белков детально исследованы у *B. subtilis*. Однако регуляция систем множественной лекарственной устойчивости белками-регуляторами семейства MerR в других организмах до сих пор подробно не изучалась.

В данной работе был определен набор геномов, содержащих регуляторы, гомологичные белкам BmrR и BltR *Bacillus subtilis*, и построено общее филогенетическое дерево этих белков. Всего было исследовано 214 регуляторов в 87 геномах, идентифицировано 119 регуляторных сайтов. Сайты связывания данных белков с ДНК представляют собой нестрогие палиндромы. В процессе поиска в геномах было установлено, что центр сайта связывания находится на фиксированном расстоянии (13 нуклеотидных пар) от 5'-конца -35 промоторного бокса. При этом практически во всех случаях промотор имел длинный спейсер — 19 нуклеотидных пар, что является характерной особенностью промоторов, активируемых регуляторами семейства MerR. Таким образом, сайт связывания перекрывается с -35 промоторным боксом, что накладывает ограничение на структуру сайта. В работе показано, что для гомологов белка BltR характерно расположение гена регулятора и регулируемого гена в дивергоне. В случае гомологов белка BmrR, регулируемые гены могут находиться как в дивергоне, так и тандемно с геном регулятора. В ряде геномов было выявлено более одного регулируемого гена. Большинство генов, потенциально регулируемых гомологами белков BmrR и BltR, кодируют транспортные белки — Na⁺/H⁺-антипортеры (34 гена) и ABC-транспортеры (12 генов). Также потенциальные регуляторные сайты были найдены перед 42 генами, кодирующими ферменты (ацетилтрансферазы, алкогольдегидрогеназы) и другие белки.

Литература

1. Grkovic S., Brown M.H., Skurray R.A. (2002) Regulation of Bacterial Drug Export Systems // *Microbiology and Molecular Biology Reviews*, №66(4), p. 671–701.
2. Ahmed M., Lyass L., Markham P.N., Taylor S.S., Vázquez-Laslop N., Neyfakh A.A. (1995) Two Highly Similar Multidrug Transporters of *Bacillus subtilis* Whose Expression Is Differentially Regulated // *Journal of Bacteriology*, №177(14), p. 3904–3910.

1 Автор выражает благодарность д.б.н. Гельфанду М.С. и к.м.н. Казакову А.Е. за ценные советы и указания.

Изучение функции групп коэкспрессирующихся генов на примере анализа образцов из мозга больных глиобластомой

Ивлиев А.Е.

Аспирант

*Московский государственный университет имени М.В.Ломоносова, Факультет биоинженерии и биоинформатики, и НИИ ФХБ имени А.Н.Белозерского, Москва, Россия
E-mail: ivliev-alexa@yandex.ru*

Анализ коэкспрессии генов на основе данных, получаемых с помощью ДНК микрочипов, дает информацию о согласованных изменениях в уровнях экспрессии генов от образца к образцу в масштабе всего генома. Эту информацию используют при изучении механизмов развития заболеваний [1], а также путей регуляции экспрессии генов в клетке [2]. Цель данной работы - проверить, имеют ли группы коэкспрессирующихся генов человека связь с наборами мишеней факторов транскрипции – их потенциальных регуляторов. Были решены следующие задачи: поставлен метод полногеномного анализа коэкспрессии генов и функциональной аннотации модулей коэкспрессии; проведен анализ образцов из мозга больных глиобластомой. Данный вид опухоли был выбран как один из наиболее трудно диагностируемых и малоизлечимых видов опухолевых заболеваний. Экспериментальные данные, полученные с помощью ДНК микрочипов фирмы Affymetrix (модель U133A, 100 клинических образцов) были загружены из базы Gene Expression Omnibus. Нормализацию данных проводили с применением алгоритма MAS 5.0, считающегося оптимальным для анализа коэкспрессии генов [3], и квантильной нормализации. Для всех генов, имеющих детектируемый уровень экспрессии согласно алгоритму MAS 5.0, построена сеть коэкспрессии. Модули коэкспрессирующихся генов выделены методом WGCNA [4]. Проведена функциональная аннотация модулей с помощью анализа их обогащения категориями Gene Ontology. Для большинства модулей обнаружена связь с одним (реже несколькими) из следующих биологических процессов: синтез белка, репликация ДНК и клеточное деление, репарация, сплайсинг, синтез АТФ в митохондриях, иммунный и воспалительный ответ, взаимодействие клеток с внеклеточным матриксом. Также обнаружены модули, связанные с тканеспецифическими функциями: развитие нервной системы, синаптическая передача, миелиновая оболочка аксонов в белом мозговом веществе. С помощью точного теста Фишера проведено сравнение состава групп коэкспрессирующихся генов с наборами мишеней, предсказанных Xie и коллегами [5], а также создателями базы данных TransFac [6] для широкого круга факторов транскрипции. Поиск показал, что некоторые модули коэкспрессии значительно пересекаются с предсказанными списками факторов транскрипции, например, митотический модуль обогащен мишенями факторов E2F – известных регуляторов клеточного цикла [7]. Согласно полученным результатам, группы генов, коэкспрессирующихся в мозге больных глиобластомой, связаны с различными клеточными и тканевыми процессами. Показано значимое сходство модулей с наборами мишеней факторов транскрипции, которое указывает на возможную связь между действием этих регуляторов и наблюдаемыми изменениями экспрессии генов. Работа выполнена при поддержке РФФИ (07-04-01160).

Литература

1. Fuller TF *et al.* Mamm Genome. 2007, 18(6-7):463-72.
2. Allocco DJ *et al.* BMC Bioinformatics. 2004, 5:18.
3. Lim WK *et al.* Bioinformatics. 2007, 1;23(13):i282-8.
4. Zhang B and Horvath S. Stat Appl Genet Mol Biol. 2005, 4:Article17.
5. Xie X *et al.* Nature. 2005, 434(7031):338-45.
6. Matys V *et al.* Nucleic Acids Res. 2006, 34:D108-10.
7. Zhu W *et al.* EMBO J. 2004 Nov 24;23(23):4615-26.

Поиск локальных областей с аномальным GC-составом в масштабе полного бактериального генома
Илатовский А.В.

сотрудник

Петербургский Институт Ядерной Физики им. Б.П.Константинова, Гатчина, Россия
E-mail: andreyi@omrb.pnpi.spb.ru

Одним из важных механизмов эволюции прокариот является горизонтальная передача генетического материала. Знания о горизонтально переданных генах важны для понимания структуры и эволюции генома исследуемого микроорганизма. Часто нуклеотидные последовательности геномов близких организмов неизвестны, поэтому необходимо использовать методы идентификации горизонтально переданных участков, основанные на анализе только одного генома. Известно что каждый геном уникален по GC-составу и использованию кодонов, что позволяет произвести идентификацию, анализируя эти характеристики генома.

Обычно для определения GC-состава применяют метод рамки фиксированной длины. Однако, результаты применения этого метода значительно зависят от выбора длины рамки, что не позволяет рассматривать его как надежный инструмент анализа GC-состава. Целью настоящей работы являлось создание методов и компьютерных программ для статистического анализа распределения GC-состава в бактериальных геномах и выявление областей со значимым отклонением GC-состава от среднего по всему геному.

Для формирования глобального распределения GC-состава был разработан биномиальный тест для оценки случайности измеренного значения GC-состава. На его основе построен алгоритм разбиения генома на сегменты, позволяющий получить глобальное распределение GC-состава с заданным уровнем достоверности.

В качестве первого приложения разработанного метода был выбран статистический анализ распределения GC-состава в геноме бактерии *E. coli* K-12. На основе статистического анализа искусственных последовательностей ДНК с заданными свойствами была изучена способность разработанных методов распознавать статистически значимые отклонения GC-состава и определены минимальные размеры участков ДНК, поддающиеся такому анализу. В результате исследования глобального и локальных распределений GC-состава в полном геноме бактерии *E. coli* K-12 было показано, что глобальное распределение GC-состава в этом геноме значимо отличается от нормального распределения, определено местоположение областей с аномальным GC-составом в геноме бактерии *E. coli* K-12 и проанализирована их возможная биологическая роль.

Исследование кислород-зависимой регуляции азотфиксации

Климова Е.Ю.

Студентка

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

E-mail: jane1390@rambler.ru

Азотфиксация – важный биохимический процесс, представляющий собой превращение молекулярного азота в ионы аммония. Несмотря на широкую распространенность азотфиксации среди различных таксономических групп, регуляция данного процесса практически не исследовалась методами биоинформатики.

Центральный фермент азотфиксации, нитрогеназа, крайне чувствительна к присутствию молекулярного кислорода, что и обуславливает необходимость регуляции процесса фиксации азота в соответствии с содержанием в среде молекулярного кислорода. На уровне экспрессии генов ответ на присутствие кислорода осуществляется гомологичными факторами транскрипции FnrN и FixK и двухкомпонентной системой FixL-FixJ [1,2].

Целью данной работы стало исследование FixK- и FnrN-зависимой регуляции в геномах различных представителей порядка *Rhizobiales*: *Bradyrhizobium japonicum*, *Bradyrhizobium* sp. BTAi1, *Sinorhizobium meliloti*, *Nitrobacter winogradskyi*, *N. hamburgensis*, *Rhodopseudomonas palustris*, *Rhizobium etli*, *R. leguminosarum* и *Agrobacterium tumefaciens* методами сравнительной геномики.

Исследование эволюции регуляторных систем FnrN и FixK показало присутствие во всех исследованных геномах как минимум одной копии генов для данных регуляторных белков. Анализ аминокислотных последовательностей белков-регуляторов FnrN и FixK выявил консервативность цистеиновых остатков, необходимых для формирования железо-серного кластера, присутствующего только в FnrN [3].

Несмотря на различия в структуре эффекторного домена, последовательности ДНК-связывающих доменов FnrN и FixK крайне близки, что позволяет предположить, что оба белка связываются с идентичными сайтами. Для поиска сайтов связывания FnrN и FixK было построено распознающее правило. В работе была исследована эволюция регулируемых генов, показаны массовые случаи дупликаций, как с сохранением, так и с исчезновением регуляции. Кроме того, была предсказана регуляция генов соxВ и creВ, кодирующих белки системы защиты от молекулярного кислорода.

Работа выполнена под руководством Д.А. Равчеева.

Литература.

1. Fischer, H.M. (1994) Genetic regulation of nitrogen fixation in rhizobia // *Microbiol. Rev.*, v. 58, p. 352–386.
2. Dixon, R., Kahn, D. (2004) Genetic regulation of biological nitrogen fixation // *Nat. Rev. Microbiol.*, v. 2, p. 621–632.
3. Korner, H., Sofia, H.Z., Zumft, W.G. (2003) Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs // *FEMS Microbiol. Rev.*, v. 27, p. 559–592.

Анализ пространственной структуры данных магнитной энцефалографии**Корнилина Е.Д.***Студентка**Московский государственный университет имени М.В.Ломоносова, Факультет
вычислительной математики и кибернетики, Москва, Россия**E-mail: ekornilina@gmail.com*

Магнитная энцефалография (МЭГ) — быстро развивающаяся область экспериментального изучения высшей нервной деятельности человека, функциональных областей мозга и диагностики различных патологий. Причиной этого является возможность неинвазивного получения данных о процессах, происходящих как в коре головного мозга, так и в глубоких его отделах.

Магнитное поле на поверхности головы описывается в результате решения так называемой прямой задачи МЭГ. Согласно исходным данным о положении, ориентации и значении момента токового источника строится картина магнитной индукции на всей поверхности головы. Для локализации источников биомагнитной активности необходимо решать обратную задачу МЭГ. В качестве признакового пространства, при решении задач классификации типа магнитной активности в записи и распознавания участков с патологией использовался набор сферических гармоник разложения распределения магнитной индукции по поверхности головы (данные на сфере). Основная сложность в реализации данной процедуры связана с неполнотой данных – датчики покрывают не всю поверхность сферы, охватывающей голову, а примерно 2/3 ее части. Для решения этой проблемы разработана процедура экстраполяции данных с помощью полиномов Лежандра и присоединенных сферических функций.

В данной работе решалась задача изучения слуховой коры при помощи пространственной локализации источников вызванной активности и анализа стохастической динамики сигнала при звуковой стимуляции испытуемого. Измеряемый сигнал представляет собой пространственно-временную структуру: 148-мерный вектор измерений на поверхности головы, развернутый во временной ряд с частотой опроса датчиков 500 Гц. Производилась начальная фильтрация сигнала и выделение полезной составляющей. С помощью разложения в ряд по сферическим функциям описывалась пространственная структура поля, после чего осуществлялся выбор моментов времени для решения обратной задачи локализации токовых источников путем подгонки модели к экспериментальным данным в эти моменты времени. Далее решалась обратная задача локализации источников с учетом физиологических ограничений, получаемых с помощью ЯМР-томографии.

Заключительным этапом было проведение анализа стохастической динамики выделенного сигнала при отсутствии аудиостимуляции испытуемого и во время подачи звукового сигнала.

Литература

1. Дедус Ф.Ф., Махортых С.А., Устинин М.Н., Дедус А.Ф. Обобщенный спектрально – аналитический метод обработки информационных массивов. Задачи анализа изображений и распознавания образов. М.: Машиностроение, 1999, 357с.
2. Дергузов А.В., Махортых С.А. Спектральные разложения и классификация данных магнитной энцефалографии. Pattern Recognition and Image Analysis. Vol. 16. № 2. 2006.
3. Nikiforov A.F., Suslov S.K., Uvarov M.N. Classical orthogonal polynomials of a discrete variable, Berlin – Heidelberg – New York: Springer Verlag, 1991.
4. Sarvas J. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem, Phys. Biol., Vol. 32, N 1, pp. 11-22, 1987.

Восстановление спектра квазивидов вируса гепатита В по сиквенс-хроматограмме

Красникова А.Е.1

Студент

Московский государственный университет имени М.В.Ломоносова, Физический факультет, Москва, Россия
E-mail: kraska-kraska@list.ru

В ходе лечения инфекционного гепатита В вирус проявляет резистентность к некоторым лекарствам, что требует изменения хода лечения. Такое поведение HBV (*Hepatitis B virus*) объясняется высокой способностью вируса к мутациям. Генетические варианты вируса, сосуществующие в организме, называют «квазивидами». Своевременное определение изменения состава и/или соотношения встречаемости квазивидов вируса у больного позволяет назначать более гибкий и эффективный курс лечения.

Для определения состава квазивидов и доли каждого из них в смеси производится сиквенирование генома вируса. По заранее составленному словарю, т.е. набору всевозможных квазивидов вируса, отбираются наиболее вероятные квазивиды, составляющие данную смесь. Пусть:

$F = \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}$ – высоты пиков экспериментальной хроматограммы (f_j – нормированная

высота пика в j -й позиции хроматограммы),

$\gamma = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{bmatrix}$ – вектор частот квазивидов в смеси (γ_i - частота встречаемости i -го квазивида в смеси),

U – матрица размера $m \times n$, такая, что:

$U_{i,j} = \begin{cases} 1, & \text{если в } j\text{-м штамме в соответствующей позиции буква совпала с буквой пика } f_i, \\ 0, & \text{не совпала.} \end{cases}$

Для вычисления вектора частот квазивидов в смеси в работе использован метод наименьших квадратов, сводящийся к поиску таких коэффициентов γ_i , которые минимизировали бы следующую квадратичную форму (функцию невязки):

$$Q = (F - U\gamma)^2 \rightarrow \min$$

при дополнительных условиях:

$$\begin{cases} 0 < \gamma_i < 1, i = 1 \dots n \\ \sum_{i=1}^n \gamma_i = 1 \end{cases}$$

На выходе алгоритм дает вектор частот γ , несущий информацию о составе смеси квазивидов. Проверка работы алгоритма будет выполнена на модельных данных, имитирующих сиквенс-хроматограммы с заданными характеристиками.

Литература:

1. Kay, A. and Zoulim, F. Hepatitis B virus genetic variability and evolution. HAL author manuscript. *Virus Res.* (2007), 127164.
2. Pozhitkov, A., Stemshorn, K., and Tautz, D. An algorithm for the determination and quantification of components of nucleic acid mixtures based on single sequencing reactions. *BMC Bioinformatics* (2005), 6:281.
3. Лоусон Ч., Хенсон Р. Численное решение задач метода наименьших квадратов/пер.с англ. М., Наука, 1986.

1 Автор выражает признательность профессору, д.ф.-м.н. Иванову П.С. за помощь в подготовке тезисов.

Horizontal transfer of retrotransposons in eukaryotes*Novikova O.S.**PhD scientific researcher**Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia**E-mail: novikova@bionet.nsc.ru*

Horizontal transfer or horizontal transmission (HT) can be defined as the process by which genes move between reproductively isolated species. It is not surprising that many examples of HT of transposable elements (TEs) have been identified in eukaryotes. TEs have the capacity to insert themselves into the chromosomes of possible vectors and, subsequently, into host chromosomes. They were found virtually in all investigated eukaryotes and represent the ubiquitous components of eukaryotic genomes (Bannert, Kurth, 2004; Vitte, Panaud, 2005). Recent evidence suggests that TEs may provide the genome with potent agents to generate genetic and genomic plasticity (Kidwell, Lisch, 1997). Two major classes of TEs are recognized (Wicker et al., 2007). Class I elements use RNA-mediated mechanisms for their transposition and are called retrotransposons. Two major orders of retrotransposons are recognized: LTR retrotransposons, which have long terminal repeats (LTRs); and non-LTR retrotransposons, which lack LTRs (Wicker et al., 2007). Class II elements transpose through DNA-mediated mechanisms and are called DNA transposons (Feschotte, Pritham, 2007). HT is well known for DNA transposons and LTR retrotransposons (Robertson, 1993; Silva, Kidwell, 2004). However non-LTR retrotransposons rarely undergo HT, and their phylogenies are largely congruent to those of their hosts (Malik et al., 1999).

Three criteria can be used for HT event recognition. The first criterion is the inconsistency between the phylogenies of TEs and host species. There are potential problems with application of this criterion for HT detection. Multiple TE lineages can be present within genomes. Moreover, TEs are multicopy components of genomes. Comparisons of paralogous TE copies instead of orthologs along with varying rates of their sequence evolution are the main sources for incongruence in phylogenetic analysis, which could be misidentified as HT. The second criterion, which seems to offer the strongest evidence, is a higher degree of observed sequence similarity for TEs than for functional genes, so called 'slowdown effect on evolutionary rates'. Once inserted, a new copy of transposable element is presumed to evolve without functional constraints. Thus, all types of mutations should have an equal chance to be fixed (Volff et al., 2000; Bensasson et al., 2001). The third criterion of inferring HT is the discontinuous distribution of TEs among closely related taxa, i.e., presence of a TE in one lineage and its absence in a sister lineage (Kaplan et al., 1985; Lohe et al., 1995).

We are combining a search of retrotransposons in publicly available eukaryotic genomes using bioinformatic approaches and an experimental analysis of diverse eukaryotic species in order to identify putative HT events. The whole genome mining of retrotransposable elements sequences is performed using uGENE software (freely available at <http://ugene.unipro.ru/>). The uGENE was designed in collaboration with UniPro Company (<http://unipro.ru/>) as an interactive visual software system aimed to simplify work with DNA sequences, alignments and annotations. The standard molecular biology methods are used in experimental analysis: PCR and cloning followed by sequencing of individual clones.

The bioinformatic whole genomes analysis of number eukaryotic species (35 fungi, two plants and three insect species) allowed us to describe three putative HT so far: (i) for CR1B non-LTR retrotransposons which have been transmitted between Bombycidae moths and Lycaenidae butterflies; (ii) for Tad-like non-LTR retrotransposons which have been transferred between Eurotiomycetes and Sordariomycetes fungi; and (iii) for Ty3 LTR retrotransposons which were presumably transmitted between basidiomycetes (Fungi) and mosses (Bryophyta, Plantae). The actual mechanisms of horizontal transfer are still unknown for eukaryotic TEs since it is not possible to show experimentally how the HT can occur. Parasites, symbionts, bacteria, or viruses all could be suggested as potential vectors for horizontal transfer.

Консервативные гидрофобные ядра белковых доменов**Пеков Ю.А.¹***студент**Московский Государственный Университет имени М.В. Ломоносова, Факультет
биоинженерии и биоинформатики, Москва, Россия
yurapekov@gmail.com*

Гидрофобное взаимодействие является одним из основных факторов, определяющих пространственную структуру белковой молекулы. Совокупность неполярных атомов белка, занимающих некоторую область пространства внутри молекулы, будем называть гидрофобным кластером. Как правило, среди гидрофобных кластеров белкового домена выделяется один самый большой, который называется гидрофобным ядром данного домена.

Для автоматического выявления гидрофобных кластеров (в частности, гидрофобного ядра) в заданной пространственной структуре белка или комплекса биологических макромолекул был создан ряд компьютерных программ (Swindells, 1995, Zehfus, 1995 и др.). В данной работе использована программа CluD (Alexeevski et al., 2004). Эта программа принимает на вход файл с пространственной структурой белка в PDB-формате, а также ряд параметров, в том числе список атомных групп, которые считаются неполярными. На выходе выдаётся набор гидрофобных кластеров данной структуры, причем для каждого кластера приводится список атомов, образующих данный кластер.

Целью работы является создание метода (компьютерной программы) для выявления консервативных гидрофобных кластеров семейства родственных белковых доменов. Известно, что консервативные (то есть сохраняющиеся у всех или большинства белков семейства) особенности структуры, как правило, наиболее важны с биологической точки зрения. В частности, можно ожидать, что аминокислотные остатки, принимающие участие в формировании консервативных кластеров, и в первую очередь консервативного гидрофобного ядра, важны для укладки доменов семейства.

Программа принимает на вход выравнивание последовательностей белков семейства. Структуры белков, при их отсутствии в текущей директории, автоматически скачиваются с сайта банка PDB. Для каждой структуры запускается программа CluD, и полученные гидрофобные кластеры отдельных структур сравниваются между собой в соответствии со входным выравниванием. На выходе получаем список консервативных кластеров, а также средства визуализации: сценарий для визуализатора пространственных структур RasMol и HTML-файл, представляющий исходное выравнивание, в котором остатки, принимающие участие в образовании разных консервативных кластеров, выделены разными цветами. Сценарий RasMol выделяет кластеры теми же цветами на изображении пространственной структуры белка.

Программа была протестирована на нескольких семействах белковых доменов. При сравнении гидрофобного ядра индивидуального домена, полученного программой CluD, и консервативного гидрофобного ядра того же домена, как правило, можно отметить большую компактность последнего — меньшее количество выступающих частей. Это, на наш взгляд, свидетельствует о том, что консервативное ядро представляет собой ту "основу" гидрофобного ядра, которая освобождена от "случайностей", то есть деталей, не имеющих решающего значения для укладки белковой глобулы.

1 Автор выражает признательность к.ф.-м.н. Спиринову С. А. и к.ф.-м.н. Алексеевскому А. В. за помощь в подготовке тезисов.

Компьютерный анализ регуляции экспрессии генов рибосомных белков у бактерий.**Петрова С.А.¹**

Студент

Московский государственный университет имени М.В.Ломоносова, Факультет
биоинженерии и биоинформатики, Москва, Россия

E-mail: plazmoliz@yandex.ru

Бактерии используют различные регуляторные механизмы для контроля экспрессии генов. Один из важных механизмов регуляции генов связан с образованием вторичной структуры мРНК в регуляторной (или лидерной) области гена. Рибосомный белок связывается с регуляторным сайтом мРНК, имитирующим сайт связывания этого белка с рибосомальной РНК. При связывании запирается сайт инициации трансляции и рибосомальные гены не экспрессируются.

Используя известные регуляторные сайты, мы построили паттерн и использовали его для поиска потенциальных регуляторных структур в бактериальных геномах. Кроме того, регуляторные сайты были уточнены методами сравнительного анализа. Был проведен эволюционный анализ регуляции и показана вариабельность регуляции рибосомных белков.

Для анализа регуляции нами были использованы следующие программы: для построения рибосомальных оперонов и поиска ортологов GenomeExplorer (Mironov, 2000) и Operon (Витрешак, unpublished); для выравнивания лидерных областей генов ClustalW (Thompson, 1997); для предсказания вторичной структуры мРНК mfold (Zuker, 2003) и Multal (Mironov, unpublished); для поиска регуляторных сигналов по паттерну RNAPattern (Vitreschak, 2002).

Нами были рассмотрены лидерные области 6 оперонов: *rpsO* (*S15*), α , *spc* (*S8*), *S10* (*L4*), *L1-L11*, β (*L10*). На основании сравнения с известными структурами *E. coli* были найдены регуляторные сайты для 16 других гамма-протобактерий. В результате сравнения этих участков были уточнены регуляторные структуры и выявлены наиболее консервативные участки.

Для оперона *L1-L11* был проведен массовый поиск регуляторных структур. На основании выборки известных регуляторных структур мы построили паттерн и провели поиск во всех доступных геномах. Регуляторный сайт был найден в различных группах (Грам-положительных бактериях, протеобактериях, цианобактериях, бактериоидах и др.). При этом в некоторых группах бактерий регуляторные сайты найдены либо перед обоими генами оперона *rplK* и *rplA* (у бацилл, лактобацилл, и клостридий), либо только перед одним геном *rplA* (клостридии, актиномицеты, бактериоиды, цианобактерии), либо перед геном *rplK* (альфа-, бета-, гамма-протобактерии). Обнаружена корреляция между локализацией регуляторной структуры и расстоянием между генами оперона. Для различных бактерий предсказана либо ко-трансляция, либо независимая трансляция генов *rplK* и *rplA*.

Таким образом, эволюционный анализ оперона *L1-L11* показал вариабельность регуляции генов рибосомных белков: разные бактерии выбирают различную регуляторную стратегию: (1) регуляторный сайт расположен перед первым геном оперона (трансляция второго гена сопряжена с трансляцией первого гена, классический случай, экспериментально изученный у *E. coli*); (2) сайт расположен перед каждым из двух генов оперона (трансляция второго гена независима от трансляции первого и регулируется своим сайтом); (3) сайт расположен только перед вторым геном (в этом случае регулируется только трансляция белка L1). Было реконструировано дерево регуляторных событий для *L1-L11* оперона.

1 Автор выражает признательность к.ф.-м.н. Витрешаку А.Г. и профессору, д.б.н., к.ф.-м.н. Гельфанду М.С. за помощь в проведении исследования и подготовке тезисов.

Методы классификации гомологичных белков на примере геномов организмов рода *Mycoplasma* и практическое применение классификаций

Попенко А.С.

Студент

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

E-mail: kirika77@gmail.com

К настоящему времени были секвенированы 14 геномов микоплазм: *M. mobile*, *M. penetrans*, *M. mycoides*, *M. synoviae*, *M. pulmonis*, *M. pneumoniae*, *M. hyopneumoniae* (штаммы 232, 7448, J), *M. gallisepticum*, *M. agalactiae*, *M. arthritis*, *M. genitalium*, *M. capricolum*. Из них единственным свободноживущим организмом является *M. mobile*. Паразитами человека являются *M. pneumoniae*, *M. genitalium* и *M. penetrans*. Эти организмы поражают дыхательные пути, мочеполовой тракт, часто вызывают хронические заболевания. Существующие схемы антибактериальной терапии инфекций, вызванных представителями рода *Mycoplasma*, не во всех случаях эффективны, поэтому существует потребность в создании новых лекарственных средств.

Методы сравнительной геномики и протеомики позволяют определить круг белков микоплазм – потенциальных мишеней для разработки новых лекарств. Такими мишенями могут быть белки микоплазм, секретируемые в хозяйскую клетку и подавляющие ее защитные механизмы. Интерес представляет специфичность белков и отдельных белковых доменов для таксона, вида или штамма микоплазм.

В работе методами сравнительной протеомики на основе расшифрованных геномов были созданы две классификации белков по сходству последовательностей: кластеры на основе выдачи программы BLAST и доменные архитектуры по результату анализа доменных структур белков с использованием базы данных Pfam. Автоматически сгенерированы все множественные выравнивания последовательностей белков, предположительно, гомологичных по классификации на основе BLAST. Построенные 927 выравниваний проанализированы визуально для выделения участков достоверного выравнивания. Во многом классификации идентичны. Так, для 1626 из 1845 кластеров соответствует единственная доменная архитектура, а для 730 доменных архитектур из 817 – единственный кластер. Имеется 8 разобранных типичных примеров несовпадений, выявляющие границы применения той или иной классификации. Анализ выравниваний для ряда случаев расхождений в классификациях позволил высказать обоснованные гипотезы о гомологичности и ортологичности белков, найти противоречия в аннотациях исходных записей из банка белковых последовательностей Uniprot, предложить свои определения ортологичности и предсказания функций ряда белков. На основе классификаций удалось выделить несколько групп белков, участвующих в патогенезе, перспективных в качестве мишеней для разработки лекарств, как то: транспортные белки, специфичные адгезины и иммуноподавляющие белки.

Полученные классификации предполагается использовать для аннотирования геномов и создания единой базы данных по микоплазмам.

Выражаю благодарность своему научному руководителю Алексеевскому А.В. за руководство и значительную помощь в работе.

VSDocker1: утилита для высокопроизводительного виртуального скрининга на Windows-кластере с использованием AutoDock 4**Прахов Н.Д.²***Магистрант**Нижегородский государственный университет им. Н. И. Лобачевского, биологический факультет, Нижний Новгород, Россия**E-mail: nikita.prakhov@gmail.com*

Сегодня виртуальный скрининг, в основе которого лежит молекулярный докинг, играет важную роль в процессе компьютерного конструирования лекарств, а также в изучении структуры и взаимодействий молекул. Очевидно, что для решения серьезных исследовательских задач из этой сферы, необходимо анализировать большое количество лигандов за приемлемое время. Это стало возможным с появлением параллельных вычислительных систем и суперкомпьютеров. Проблемой на этом пути становится выбор программного обеспечения. Одной из наиболее широко используемых программ молекулярного докинга является AutoDock. Но AutoDock изначально не предназначен для организации параллельных вычислений. Однако существует пример адаптации AutoDock 4 для работы на кластере под управлением Linux (Zhang et al., 2008), что делает возможным проведение виртуального скрининга. Тем не менее для операционной системы Windows аналогичной утилиты не существовало.

Цель работы: создание утилиты для проведения виртуального скрининга с использованием AutoDock 4 на кластере в среде OS Windows. Работа стала возможной благодаря наличию в университете им. Н. И. Лобачевского высокопроизводительного вычислительного кластера. На кластере допустим запуск приложений, написанных на основе свободно распространяемой реализации стандарта MPI — MPICH2 (www.mcs.anl.gov). В настоящее время технология MPI является основным средством программирования для кластерных систем и компьютеров с распределенной памятью. В результате было разработано консольное приложение VSDocker, написанное на языке C++ с использованием библиотеки MPICH2. К основным возможностям VSDocker относятся следующие:

1. Автоматическое восстановление работы утилиты после непредвиденной остановки (реализуется без потери уже рассчитанных данных), например в результате сбоя в работе кластера.

2. Корректная работа в случае “зависания” отдельных копий AutoDock 4.

3. Отображение оценки времени оставшегося до завершения всей задачи.

4. Отсутствие большого числа параметров для запуска.

Результирующие файлы (*.dlg) с результатами расчетов сохраняются в архиве вместе с директориями, в которых они находились. Архивация делает возможным легкий перенос и распаковку файлов на настольном компьютере, а также их дальнейший анализ, например, с помощью скрипта “summarize_results4.py” (Lindstrom et al., 2007). Программа VSDocker свободно доступна для некоммерческого использования по адресу <http://bio.nnov.ru>.

Литература

1. Zhang, S., Kumar, K., Jiang, X., Wallqvist, A., and Reifman, J. (2008) DOVIS: an implementation for high-throughput virtual screening using AutoDock4 // BMC.Bioinformatics. Vol. 9, p. 126.
2. Lindstrom, W., Garrett, M., Weber, C. and Huey, R. (2007) Using AutoDock4 for Virtual Screening.
3. www.mcs.anl.gov/research/projects/mpich2 (MPICH2: High-performance and Widely Portable MPI).

1 Разработка программы VSDocker выполнена при поддержке РФФИ (проект № 08-04-01816)

2 Автор выражает признательность к.м.н. Гайнуллину М.Р. за помощь в выполнении работы и написании тезисов.

Совместная метаболическая активность тли *Acyrtosiphon pisum* и симбиотической бактерии *Buchnera aphidicola* str. APS

Ракитин Д.И.

Студент

Московский государственный университет имени М.В. Ломоносова, Биологический факультет, Москва, Россия
E-mail: callipso@yandex.ru

Симбиоз широко распространен в природе и присущ многих группам растений и животных. Самым тесным вариантом симбиотического взаимодействия организмов является эндоцитобиоз, при котором симбионт живет непосредственно в клетках хозяина. Примером такого типа взаимодействия является симбиоз тли *Acyrtosiphon pisum* и бактерии *Buchnera aphidicola* str. APS, продуцирующей для организма-хозяина ряд аминокислот и кофакторов.

В ходе исследования с помощью методов сравнительной геномики были выявлены гены *Buchnera aphidicola* str. APS, отвечающие за метаболические пути синтеза десяти незаменимых аминокислот. При этом было обнаружено отсутствие некоторых генов, кодирующих белки нескольких реакций путей синтеза валина, лейцина, изолейцина и аспартата, хотя данные химического анализа, заимствованные из литературных источников, показывают наличие конечного химического продукта данных метаболических путей. Данное разногласие попытались устранить утверждением о том, что недостающие гены находятся в организме насекомого хозяина. Для этого белки родственной бактерии *E. coli*, катализирующие недостающие в клетке *Buchnera* реакции, сравнивали при помощи программы BLAST с черновым вариантом генома тли *Acyrtosiphon pisum*. Для всех недостающих звеньев (*ilvE*, *ilvA*, *aspC*) были обнаружены кандидаты, функция которых была подтверждена сравнением с полной базой GenBank.

Ранее считалось, что все реакции синтеза аминокислот проходят в организме симбиотической бактерии, из наших наблюдений, однако следует, что одна часть пути синтеза проходит в организме насекомого-хозяина, а другая часть в организме симбиотической бактерии.

Составление информативных имен последовательностей с использованием аннотаций¹

Решетов ДА2

Студент

Московский государственный университет имени М.В. Ломоносова, Факультет биоинженерии и биоинформатики, Москва, Россия

E-mail: reshetov@genebee.msu.ru

Аннотации белковых и нуклеотидных последовательностей, а также сами последовательности хранятся в базах данных под уникальными идентификаторами. Эти идентификаторы, обычно мало информативны в человеческом общении, поэтому они заменяются в публикациях и презентациях на более содержательные имена. Такие замены часто совершаются вручную, что отнимает время и приводит к ошибкам и неточностям.

Нами разработан SNAD (Sequence Name Annotation-based Designer) – инструмент для автоматической замены идентификаторов последовательностей в списках, выравниваниях и филогенетических деревьях на информативные имена. Эта замена управляется шаблонами (уже готовыми или задающимися пользователем) с использованием аннотаций из внешних баз данных.

Мы надеемся, что разработанный инструмент повысит эффективность взаимодействия между исследователями.

SNAD находится на сайте <http://goudhaas.lumc.nl/SNAD/>

Конвертирование идентификаторов с использованием SNAD

Девять идентификаторов последовательностей из GenBank (поле “Before ID conversion”) конвертированы с использованием шаблона включающего в себя следующие характеристики из GenBank: “Primary ID” (gi); название организма в формате “G[enus].species”; название белка и локуса. Использовано три разделителя: пробел, “[“ и ”]”. Первые три характеристики были ограничены размерами в 9, 6, и 17 символов, соответственно. Результаты конвертации показаны в секции “After ID conversion”. Каждая строчка соответствует одному конвертированному идентификатору. {...} показывает, что более одной характеристики “G[enus].species” найдено для данного ID (см. GenBank:P03033, поле source/organism).

Before ID conversion (enter your data in the box below)

SUPPORTED FORMATS: ID list GenBank or UniProt, Alignment fasta, msf, dustalw, Tree newick, nexus

Example IDs: CAB04658, NF_248318, AAB994

or upload it from file: _____

Send results to e-mail: _____

Names/format designer (select name template, refine if needed, and the query)

Select name template: Complex protein name

Template-building facility

#	Delimiter	Characteristic	min	max	abbreviate	Location	Source
01		Primary ID	9	9	no	↑ ↓ ↕	GenBank
02	space	Organism=>G[enus] species	6	6	no	↑ ↓ ↕	GenBank
03		Protein, product	17	17	no	↑ ↓ ↕	GenBank
04		CDS, locus tag				↑ ↓ ↕	GenBank

Show * if characteristic is not available

OUTPUT:

Alignment/tree options: alignment format, tree format, show sequence sizes

Interactive display options: names only, annotation table, entry link

After ID conversion (see results below and in separate window)

Number of ID: submitted = 9, converted = 9. Time total: 4.45 sec (0.49 per ID).

3879212	C.eleg[C. elegans protei	R13H4.6
15669508	M.jann[hypothetical prot	MJ1318
1592066	M.jann[ATP-dependent pro	MJ_1417
1786643	E.coli[DNA-binding ATP-d	b0439
1651586	E.coli[UmuD protein.]*
2144964	E.coli[mucA protein]*
126224	{...} [LexA repressor]*
133353	E.phag[Repressor protein]*
763050	S.phag[repressor protein]*

Primary ID, 9 positions Organism=>G[enus]species, 6 positions Protein, product, 17 positions CDS, locus tag

- 1 Работа частично поддержана Коллективным соглашением в области биоинформатики между Медицинским центром Лейденского университета (Нидерланды) и МГУ (CRDF GAP1473), а также EU FP6 IP Vizier LSHG-ST-2004-511960.
- 2 Автор выражает благодарность к.ф.-м.н. Сидорову ИА и д.б.н. Горбалене АЕ за помощь при работе над проектом и в подготовке тезисов.

Исследование генов с высокой связностью в сети коэкспрессии как потенциальных мишеней лекарственных средств

Руднева В.А., Ивлиев А.Е.

Студент, аспирант

*Московский государственный университет имени М.В.Ломоносова, Факультет биоинженерии и биоинформатики и НИИ ФХБ имени А.Н.Белозерского, Москва, Россия
E-mail: rudnevav@gmail.com, ivliev-alexa@yandex.ru*

В настоящее время новые лекарства разрабатываются преимущественно для уже известных белков-мишеней [1] путем подбора химических веществ, специфически связывающихся с этими мишенями. Проблема поиска новых мишеней лекарственных средств - это актуальная задача современной фармакологии. В последние годы были предложены биоинформатические подходы к поиску мишеней. В частности, предложено использовать анализ сетей коэкспрессии, которые отражают динамическую структуру транскриптома тканей и клеток человека при патологии. Предполагается, что потенциальные мишени следует искать в модулях коэкспрессии, связанных с патологическими процессами, а точнее, среди хабов, т.е. генов с высокой связностью, в пределах этих модулей [2]. Целью данной работы являлось проверить справедливость этого предположения на примере уже известных мишеней лекарственных средств.

Анализ проводился на выборке противоопухолевых препаратов как одного из наиболее многочисленных классов лекарств. Мишени к препаратам этого класса были извлечены из базы данных DrugBank [3] – всего 147 мишеней. Для построения сетей коэкспрессии из базы Gene Expression Omnibus загружены данные по экспрессии генов в тканях больных раком груди, крови, простаты, прямой кишки и мозга - всего 577 клинических образцов. Данные были исходно получены с помощью ДНК микрочипов фирмы Affymetrix (модель U133A) и нормализованы методом MAS 5.0. Проведен предварительный скрининг данных с целью исключения потенциально некачественных образцов. Методом WGCNA [4] построены сети коэкспрессии для каждой группы больных. Выделены модули коэкспрессирующихся генов и проведена их функциональная аннотация с помощью анализа обогащения категориями Gene Ontology. С помощью точного теста Фишера проведена оценка обогащения модулей мишенями противоопухолевых лекарственных средств. Для проверки наличия взаимосвязи хабов с известными мишенями проведен тест Уилкоксона с поправкой на множественность тестирования Бенжамини-Хохберга и нулевой гипотезой об отсутствии различия между средней связностью мишеней и остальных генов в пределах каждого модуля. Обнаружено, что известные мишени противоопухолевых препаратов действительно ассоциированы с ограниченным набором модулей коэкспрессии, вовлеченных в процесс развития заболевания. Однако не было выявлено тенденции мишеней иметь высокую связность, т.е. быть хабами, в соответствующих модулях. Согласно полученным результатам, выделение групп коэкспрессирующихся генов дает дополнительную информацию при поиске потенциальных мишеней лекарственных средств, однако поиск потенциальных мишеней среди генов с высокой связностью представляется недостаточно обоснованным. Работа выполнена при поддержке РФФИ (07-04-01160).

Литература

1. Yildirim MA *et al.* Nat Biotechnol, 2007, 25(10):1119-26.
2. Horvath S *et al.* PNAS, 2006, 103(46):17402-7.
3. Wishart DS *et al.* Nucleic Acids Res, 2008, 36:D901-6.
4. Zhang B and Horvath S. Stat Appl Genet Mol Biol, 2005, 4:Article17.

Предсказание субстратной специфичности цистеиновых пептидаз семейства С1¹**Семашко Т.А.**Аспирант²*Московский государственный университет имени М.В.Ломоносова, Факультет биоинженерии и биоинформатики, Москва, Россия**E-mail: t.semashko@gmail.com*

Цистеиновые пептидазы семейства С1 (семейства папаина) — большая группа гидролаз различного происхождения, содержащих в активном центре остаток цистеина и обладающих сходством первичной и пространственной структуры [1].

Задачей работы являлся анализ первичной и пространственной структуры различных цистеиновых пептидаз семейства С1 с целью выявления возможных отличий в их субстратной специфичности. Особый интерес представлял анализ локальных изменений в структуре субстрат-связывающей области.

Объектами исследования являлись: растительные ферменты папаин и бромелаин, катепсины В и L млекопитающих, а также главная пищеварительная пептидаза из личинок жука-вредителя мучных запасов *Tenebrio molitor* (ДСР). Степень изученности этих ферментов различна. Для папаина и катепсинов имеются данные о первичной и третичной структуре. Бромелаин, напротив, слабо охарактеризован в структурном отношении. Для него и ДСР существуют данные о первичной структуре препрофермента.

Нами проведено сравнение аминокислотных последовательностей, а также третичных структур или их моделей для выбранных ферментов. Модели структур бромелаина и ДСР строились с использованием сервера SwissModel [2] в автоматическом режиме. Все изученные ферменты имеют консервативное строение активного центра, однако несколько различающуюся зону связывания.

Для характеристики субстратной специфичности вышеперечисленных пептидаз, исходя из литературных данных, был предложен гипотетический ряд субстратов общей формулы А-Хаа-Уаа-В, где А = Glp (пироглутамил), Abz (*o*-аминобензоил); В = рNA (*n*-нитроанилид), АМС (7-амидо-4-метилкумарин) и АФС (7-амидо-4-трифторметилкумарин); Хаа = Phe, Val; Уаа = Ala, Arg, Cys(SBz), Gln, Cit. С использованием метода молекулярного докинга было показано, что все вышеперечисленные вещества могут являться субстратами исследованных пептидаз, однако для некоторых комбинаций фермент-субстрат возможно связывание, которое в дальнейшем приведет к гидролизу по пептидной связи между аминокислотными остатками, а не к отщеплению детектируемой группы. Так как из литературных данных известно, что катепсин В проявляет при кислых рН дипептидил-пептидазную активность, а при нейтральных — эндопептидазную, протонирование фермента при определенном рН для подготовки к докингу осуществлялось с помощью программы H++ [3]. Докинг проводился с использованием программы Lead-Finder [4]. Расчетные характеристики коррелируют с данными о субстратной специфичности пептидаз, полученными экспериментально с использованием синтезированных нами субстратов, описанных выше.

Литература

1. <http://merops.sanger.ac.uk> (Rawlings et al. (2008), NAR 36, p. 320-325)
2. <http://swissmodel.expasy.org/workspace> (Schwede et al. (2003), NAR 31, p. 3381-3385)
3. <http://biophysics.cs.vt.edu/H++/index.php> (Gordon et al. (2005) NAR 33, p. 368-371)
4. <http://www.moltech.ru> (Stroganov et al.. (2008) J Chem Inf Model. 48(12), p. 2371-2385)

1 Тезисы доклада основаны на материалах исследований, проведенных в рамках гранта РФФИ (гранты № 09-04-91289-ИНИС_а и 09-04-01449-а).

2 Автор выражает признательность в.н.с., к.х.н. Филипповой И.Ю., с.н.с., к.б.н. Элпидиной Е.Н. за помощь в подготовке тезисов.

**Получение клеточных линий с суперпродукцией и нокдауном протимозина альфа;
исследование влияния протимозина альфа на чувствительность клеток к
окислительному стрессу**

Соколов В.В., Суворова А.А., Захарова Н.И.

Студент, студент, аспирант

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

zakharova_n.i@mail.ru

Протимозин α (ПроТа) – мультифункциональный ядерный белок, обладающий свойствами онкобелка. Для исследования механизмов его функционирования было решено получить клеточные линии с повышенным и пониженным уровнем ПроТа с целью анализа влияния этого белка на экспрессию генов, а также на чувствительность клеток к окислительному стрессу.

Последняя задача обусловлена тем, что в нашей лаборатории было обнаружено взаимодействие ПроТа с белком Keap1, ингибитором транскрипционного фактора Nrf2, регулирующего гены ответа клетки на окислительный стресс. Повышение концентрации ПроТа приводит к диссоциации комплекса Nrf2-Keap1 и увеличению концентрации свободного Nrf2, способного взаимодействовать с промоторными областями генов белков-антиоксидантов. Это позволило предположить, что изменение внутриклеточной концентрации ПроТа должно влиять на чувствительность клеток к окислительному стрессу. Для экспериментальной проверки этой гипотезы было решено исследовать чувствительность клеток с разным уровнем ПроТа к действию окислителей.

Для создания клеточной линии с суперпродукцией ПроТа белок-кодирующая область его кДНК была встроена в лентивирусный вектор под контролем цитомегаловирусного промотора и энхансера трансляции. Были получены псевдовирусные частицы, которыми инфицировали клетки линии HeLa. Клетки с интегрированной в хромосому провирусной ДНК селектировали на среде, содержащей антибиотик G-418. По данным электрофоретического анализа уровень ПроТа в полученных таким образом клетках оказался примерно в 3 раза выше, чем в контрольных клетках, инфицированных «пустым» лентивирусным вектором.

Клеточная линия с пониженным уровнем ПроТа была получена аналогичным образом с помощью сконструированного ранее лентивирусного вектора, кодирующего короткую интерферирующую РНК, специфичную к ПроТа. В этих клетках уровень ПроТа оказался примерно в 3 раза ниже по сравнению с контрольными клетками, инфицированными «пустым» лентивирусным вектором (по результатам микрочипового исследования).

С целью тестирования чувствительности полученных клеточных линий к окислительному стрессу их обрабатывали сублетальными дозами перекиси водорода и анализировали кривые роста клеток. Оказалось, что высокий внутриклеточный уровень ПроТа защищает клетки от гибели, вызванной окислительным стрессом, а снижение уровня ПроТа, наоборот, способствует гибели клеток.

Для тестирования роли системы Nrf2-Keap1 в защитном действии ПроТа сконструирован лентивирусный вектор, кодирующий мутант ПроТа E(44,50)G, потерявший способность взаимодействовать с белком Keap1. В настоящее время проводится получение клеточной линии с суперпродукцией мутантного ПроТа с целью анализа ее чувствительности к действию окислителей.

В заключение, в настоящей работе впервые показано, что внутриклеточный уровень протимозина альфа коррелирует с резистентностью клетки к окислительному стрессу. Предложен способ тестирования молекулярного механизма этого явления.

ViTaVi: a web-tool for Virus Taxonomy Visualization¹**Tukhtubaeva N.E.¹, Sedlyarov V.O.¹, Lauber C.²***Student, student, PhD student*¹ *Faculty of bioengineering and bioinformatics, Moscow State University, Moscow, Russia*² *Leiden University Medical Center, Leiden, The Netherlands**E-mail: nadkafbb@gmail.com*

Viruses are exceptionally diverse in morphology, replication strategies, genetic organization and other characteristics. In current RNA Virus Taxonomy multiple species demarcation criteria are applied to accommodate molecular and phenotypic diversity. These periodically revised criteria, which are developed by experts, hinder efficient and objective classification of both newly identified and already recognized RNA viruses. Virus species recognized by current taxonomy tend to form compact clusters in phylogenetic trees. Phylogenetically based pair-wise genetic distances offer a resort from drawbacks of existent Virus Taxonomy. These probability-based distances represent a robust and objective criterion for demarcation of species level to achieve a comprehensive classification of viruses from a particular family.

This project is aimed at developing a user friendly web application that connects parts of a software tool for pair-wise distance-based clustering of virus sequences. A second major aspect was the development of a tool for graphical representation of taxonomic structures

A web-tool was developed in Perl using the CGI (Common Gateway Interface) modules. The CGI-script operates on an Apache HTTP-server installed on a UNIX platform. This web-tool is intended for multi-user usage and includes three main steps. Firstly, all possible distance thresholds (where cliques form) are calculated based on a pair-wise distance matrix. Secondly, the user can choose thresholds which may correspond to taxonomic units (Species, Genus, Subfamily, Family and Order (bottom-up)). Finally, the resulting taxonomic structure is visualized.

The graphical platform was developed in Perl. It provides an opportunity for working with complex taxonomic structures. In general, three main stages of the program development can be defined:

1. Developing a procedure for parsing taxonomic structures in an XML-like format
2. Developing a data structure for storing taxonomic units
3. Coding a module for taxonomic structure visualization.

In summary, we developed a user-friendly web-tool for representing results of pair-wise distance-based clustering of sequences.

1 We thank Dr. Gorbalenya A.E. and Dr. Sidorov I.A. for supervising our work.

К вопросу о происхождении лингвистической симметрии цепей геномных ДНК**Усанов Н.Н.**

Студент

Московский государственный университет имени М.В.Ломоносова, Химический факультет, Москва, Россия
E-mail: nikolai@usanov.ru

Лингвистическая симметрия текстов хромосомных ДНК прокариот, архей и эукариот универсальна, ее легко обнаружить, сравнивая частотные словари первой и второй цепей, полученные методом скользящей рамки при ее длине $N = 3 \div 12$. Корреляции по Пирсону частот встречаемости строк в парах цепей, в подавляющем большинстве случаев можно расценивать как сильные или очень сильные.

Явление не объяснимо с точки зрения алгоритма построения текста второй цепи из текста первой: принципа комплементарности и инверсии направления текста. Любой тест геномной ДНК, созданный псевдослучайным способом, коренным образом отличается от «живого» текста именно отсутствием симметрии между первой и второй цепями даже на уровне три- и тетра-нуклеотидных сочетаний, что легко проверяется экспериментально *in silico*. Иными словами, корреляции в парных рядах частот встречаемости, полученных для строк искусственных ДНК большой длины, практически не обнаруживаются. Устойчивое существование симметрии, как и ее происхождение, также не имеет объяснений, если следовать логике большинства известных перестроек генома, включая крупноблочные делеции, транслокации и вставки чужеродной ДНК в процессе горизонтального переноса генов, наличие IS-элементов и мультипликации собственных участков. Более того, случайные мутации должны разрушать симметрию, в то время как она существует в геномах практически всех организмов.

В нашей модели, мы попытались создать алгоритм построения искусственных ДНК с высоким уровнем лингвистической симметрии путем эволюции хаотических текстов. С этой целью, в качестве исходных матриц использовались псевдослучайные последовательности, сгенерированные программно и имеющие значения корреляции Пирсона для частот встречаемости $r < 0.07$. Процесс непрерывного встраивания инвертированных копий фрагмента первой цепи во вторую был выбран в качестве гипотетического молекулярного механизма, позволяющего хаотической последовательности измениться до состояния явной лингвистической симметрии.

Для реализации этого алгоритма, в псевдослучайной строке ДНК выбирался фрагмент со случайной длиной ($N = 50 \div 5000$ п.о.) и положением в тексте, который встраивался на случайной позиции в комплементарной цепи. Из измененного текста второй цепи рассчитывалась и записывалась новая последовательность первой. Для поддержания постоянства длины искусственного «генома» во время его «эволюции», периодически и хаотически совершались делеции (утери фрагментов ДНК). Было обнаружено, что при генерации инвертированных повторов с сопутствующим удалением «избытков» текста, лингвистическая симметрия, сходная по значениям коэффициентов Пирсона с таковой природных организмов, возникала через $10^4 - 10^5$ циклов трансформации случайного текста. Важным и существенным представляется вопрос о том, с какой скоростью случайные мутации могут разрушать возникающую симметрию. В модельной системе, при внесении 30 случайных мутаций на один цикл инверсии с участием вставок размером 5000 нуклеотидов, уже через 256 тысяч итераций возникает и стабилизируется уровень разбалансированности, соответствующий геномам прокариот с самым маленьким уровнем лингвистической симметрии. Полученные данные позволяют по новому взглянуть на молекулярные механизмы процессов эволюции.

Литература

1. Усанов Н.Н., Усанов Н.Г. Симметричные структуры в комплементарных цепях геномных ДНК и возможные алгоритмы их организации // Препринт ИПМ № 43, Москва, 2008

Полиморфизм "анонимных" фрагментов ДНК и позиционирование инвертированных повторов в секвенированных последовательностях генома овцы
Феофилов А.В.¹

Аспирант

*Российский Государственный Аграрный Университет – МСХА им. К.А.Тимирязева,
Центр нанобиотехнологий, Москва, Россия*

E-mail: foton3@yandex.ru

Анализ полиморфизма ISSR-PCR маркеров (фрагментов ДНК, фланкированных инвертированными повторами микросателлитных локусов) широко применяется для решения задач общей и частной генетики. Их использование основано на предположении о том, что в геноме микросателлитные локусы распределены случайно и их полиморфизм сходен. В то же время накапливаются данные о том, что полиморфизм существенно отличается от одного микросателлита к другому. Это приводит к необходимости предварительного экспериментального исследования эффективности использования фрагментов разных микросателлитных локусов в качестве праймеров в полимеразной цепной реакции (PCR) для получения на геномной ДНК разных видов наиболее информативных (по количеству выявляемых локусов и их полиморфизму) спектров продуктов амплификации. Для того, чтобы упростить процедуру такого подбора, необходимо выяснить возможные причины отличий в информативности между такими локусами. Можно ожидать, что к ним относятся особенности корового мотива микросателлита, локализация их в кодирующих белки и некодирующих последовательностях геномной ДНК. В этой связи, в настоящей работе выполнено сопоставление спектров ISSR-PCR маркеров, полученных с использованием фрагментов двух микросателлитов (AG)₉C и (GA)₉C на геномной ДНК животных романовской и кулундинской пород овец и их позиционированием в секвенированных последовательностях ДНК домашней овцы, имеющих в международном банке данных (GenBank). В результате получены следующие данные. Использование праймера (AG)₉C в PCR на геномах двух пород овец позволило получить спектры продуктов амплификации, включающие 13 надежно выявляемых фрагментов ДНК, а для праймера (GA)₉C - только 8. Обращает на себя внимание точность отжига праймеров, поскольку сдвиг на один нуклеотид может привести к преобразованию одного праймера в другой. Для того, чтобы выяснить возможные причины наблюдаемых отличий в информативности этих микросателлитных локусов при их использовании в качестве праймеров в PCR, проведен анализ количества участков гомологии к прямым и инвертированным повторам каждого из праймеров на основе секвенированных последовательностей генома овцы, представленных в GenBank, с использованием стандартной компьютерной программы BLASTn (100% совпадений). Для (AG)₉ выявлено 19 участков гомологии, для (CT)₉ – 3 и только 1 – для (GA)₉ - (TC)₉. Интересно отметить, что последовательность (AG) относительно чаще локализуется в структурных генах, чем (GA), что соответствует выявленным различиям по индексу PIC (polymorphic information contents) между продуктами амплификации, полученным с использованием в PCR в качестве праймеров (AG)₉C и (GA)₉C на геномах исследованных пород овец. Наблюдаемые различия позволяют утверждать, что полиморфизм ISSR-PCR маркеров зависит от корового мотива микросателлитного локуса (AG или GA), используемого в качестве праймера, а также позиционирования в нуклеотидных последовательностях, принадлежащих структурным генам, или локализованных в других геномных участках. Из полученных данных следует, что предварительный поиск участков гомологии для микросателлитных локусов в секвенированных последовательностях GenBank может существенно упростить подбор праймеров для выявления наиболее информативных ISSR-PCR маркеров геномов разных видов.

¹ Автор выражает благодарность академику РАСХН (иностранный член), профессору, д.с.-х.н. Глазко В.И. за помощь в подготовке тезисов

Молекулярный механизм ферментативного гидролиза АТФ в активном центре белка**RecA из *E.coli*.**Швецов А.В.[#]

Студент

Санкт-Петербургский Государственный Политехнический Университет, Физико-Механический факультет, Санкт-Петербург, Россия

E-mail: alexxy@omrb.pnpi.spb.ru

Работа посвящена теоретическому исследованию механизма АТФазной активности белка RecA который является одним из центральных белков участвующих в реакции гомологической рекомбинации у прокариот. Аналоги этого белка существуют во всех организмах начиная от археобактерий (RadA) заканчивая эукариотическими организмами (Rad51).

С помощью пакета программ GAMESS для квантово-химических расчетов на многопроцессорном вычислительном комплексе была исследована структура и энергетика исходного, переходного и продуктивного состояния в реакции ферментативного гидролиза АТФ в активном центре белка RecA из *E.coli*. Работа состояла из нескольких этапов: 1) построение фермент-субстратного комплекса на базе литературных рентгеноструктурных данных; 2) построение упрощенной модели комплекса, пригодной для проведения квантово-механических расчетов; 3) квантово-химические расчеты энергии комплекса и ее минимизация с помощью пакета программ GAMESS.

Пространственные структуры фермент-субстратного комплекса и его упрощенных моделей строились с помощью ICM [1], пакета программ для молекулярного моделирования (Molsoft LLC, USA). На базе известных рентгеновских структур комплексов RecA с различными лигандами (АДФ, АТФ- γ -S и ADP-AlF₄) [2-5], а также литературных данных по механизмам гидролиза АТФ другими ферментами [6,7] были построены несколько вариантов активированного комплекса, которые могут корректно воспроизводить реакцию гидролиза АТФ. Расчеты показали, что природа энергетического барьера в этой химической реакции вероятно связана с энтальпийными и энтропийными потерями, необходимыми при сборке всех элементов в активированный фермент-субстратный комплекс.

Литература

1. Abagyan R, Totrov M. *J Mol Biol* 1994; **235**:983-1002.
2. Story RM, Steitz TA. *Nature* 1992; **355**:374-376.
3. Datta S, Prabu MM, Vaze MB, Ganesh N, Chandra NR, Muniyappa K, Vijayan M. *Nucleic Acids Res* 2000; **28**:4964-4973.
4. Datta S, Krishna R, Ganesh N, Chandra NR, Muniyappa K, Vijayan M. *J Bacteriol* 2003; **185**:4280-4284.
5. Xing X, Bell CE. *Biochemistry* 2004; **43**:16142-16152.
6. Topol IA, Cachau RE, Nemukhin AV, Grigorenko BL, Burt SK. *Biochim Biophys Acta* 2004; **1700**:125-136.
7. Grigorenko BL, Nemukhin AV, Topol IA, Cachau RE, Burt SK. *Proteins* 2005; **60**:495-503.
8. Parr RG, Yang W. *Density-functional theory of atoms and molecules*. New York, Oxford UK: Oxford University Press ; Clarendon Press; 1989.
9. Schmidt M.W. et al. *Journal of Computational Chemistry* 1993; **14**:1347-1363.

[#] Научный руководитель Петухов М. Г., доцент НОЦ «Биофизика» при СПбГПУ, с.н.с. ПИЯФ РАН