ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

ОСНОВАН В 1997 г.

Том 17
4/1
Октябрь 2013

VAVILOV JOURNAL OF GENETICS AND BREEDING

FOUNDED IN 1997

Vol. 17
4/1
October 2013

«Вавиловский журнал генетики и селекции» / «Vavilov Journal of Genetics and Breeding» до 2011 г. выходил под названием «Информационный вестник ВОГиС» / «The Herald of Vavilov Society for Geneticists and Breeding Scientists».

«Вавиловский журнал генетики и селекции» включен ВАК Минобрнауки России в Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени доктора и кандидата наук (по биологическим наукам).
(Редакция 17 июня 2011 г.: http://vak.ed.gov.ru)

«Вавиловский журнал генетики и селекции» включен в федеральный почтовый Объединенный каталог «ПРЕССА РОССИИ».
Персональный подписной индекс № 42153.

Адрес редакции:
«Вавиловский журнал генетики и селекции»,
ИЦиГ СО РАН,
Проспект Академика Лаврентьева, 10,
Новосибирск, 630090

Факс: (383) 3331278
e-mail: vavilov_journal@bionet.nsc.ru

Ответственный секретарь редакции:
С.В. Зубова,
тел. 363-4922 *1351

Регистрационное свидетельство ПИ № ФС77-45870 выдано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций 20 июля 2011 г.

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

ОСНОВАН В 1997 г.

Том 17
4/1
Октябрь 2013

Содержание

VAVILOV JOURNAL
OF GENETICS AND BREEDING
FOUNDED IN 1997

Vol. 17
4/1
October 2013

Content

ПРЕДИСЛОВИЕ

Главная особенность современной биоинформатики и системной биологии – глубокая интеграция с высокопроизводительными экспериментальными и компьютерными технологиями.

За последние 15 лет молекулярная биология и генетика вышли на качественно новый уровень исследований, основанных на использовании высокопроизводительных экспериментальных технологий, таких как скоростное секвенирование геномной ДНК, многолокусное генотипирование, многопараметрическое профилирование экспрессии генов с использованием ДНК-чипов, ChIP-on-chip технологий, MPSS, протеомных технологий, позволяющих анализировать протеомы органов, тканей и групп клеток с масс-спектрометрической расшифровкой аминокислотных последовательностей белков и т. д.

В связи с беспрецедентно огромными объемами данных, генерируемых современной экспериментальной биологией, критически возрастает роль таких научных направлений, как биоинформатика и системная компьютерная биология, обеспечивающих возможность автоматического конвейерного анализа и интерпретации получаемых экспериментальных данных, моделирования биологических систем и процессов.

Системная компьютерная биология имеет важнейшее значение для решения широкого круга прикладных задач в области биомедицины (моделирование механизмов возникновения патологий) и биотехнологии (моделирование и оптимизация метаболических путей при создании бактериальных штаммов-суперпродуцентов, планирование экспериментов по созданию генетически модифицированных организмов с заданными целевыми свойствами).

Наиболее перспективным путем развития исследований в области молекулярной генетики и биотехнологии является симбиоз теоретических, компьютерных и экспериментальных подходов.

В настоящем выпуске журнала представлены результаты исследований, проводимых в СО РАН по различным направлениям как биоинформатики и системной компьютерной биологии, так и генетической и метаболической инженерии и биотехнологии, включая разработку методов и программных систем оценки влияния однонуклеотидных полиморфизмов на развитие социально значимых заболеваний, генерации Web-сервисов и вычислительных конвейеров для унифицированного доступа к ресурсам в области биоинформатики; компьютерный анализ данных по экспрессии генов в клетках мозга; моделирование механики и морфодинамики растений; построение сегрегационных моделей сложных количественных признаков; комплексные экспериментально-биоинформатические исследования в области генетической и метаболической инженерии, ориентированные на получение этанола и органических кислот; анализ перспективных свойств бактерий, обладающих липолитической активностью, а также исследование свойств термостабильных липаз.

Приглашенные редакторы **Н.А. Колчанов**
Н.Л. Подколодный
С.Е. Пельтек
Ю.Л. Орлов

УДК 61:575; 658.011.56

ПРОГРАММНЫЙ КОМПЛЕКС SNP-MED ДЛЯ АНАЛИЗА ВЛИЯНИЯ ОДНОНУКЛЕОТИДНЫХ ПОЛИМОРФИЗМОВ НА ФУНКЦИЮ ГЕНОВ, СВЯЗАННЫХ С РАЗВИТИЕМ СОЦИАЛЬНО ЗНАЧИМЫХ ЗАБОЛЕВАНИЙ

© 2013 г. **Н.Л. Подколодный[1], Д.А. Афонников[1], Ю.Ю. Васькин[2], Л.О. Брызгалов[1], В.А. Иванисенко[1], П.С. Деменков[1], М.П. Пономаренко[1], Д.А. Рассказов[1], К.В. Гунбин[1], И.В. Процук[2], И.Ю. Шутов[2], П.Н. Леонтьев[2], М.Ю. Фурсов[2], Н.П. Бондарь[1], Е.В. Антонцева[1], Т.И. Меркулова[1], Н.А. Колчанов[1]**

[1] Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия, e-mail: pnl@bionet.nsc.ru;
[2] ООО Новосибирский центр информационных технологий «УНИПРО», Новосибирск, Россия

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

В данной работе описана модульная компьютерная информационная система SNP-MED, предназначенная для оценки влияния однонуклеотидных полиморфизмов (ОНП) на функцию генов, связанных с развитием социально значимых заболеваний, включающая программные компоненты «Геномика», «Протеомика», «Генные сети» и базу данных «Информационный ресурс» (БДИР).

Ключевые слова: биоинформатика, однонуклеотидные полиморфизмы, персонализированная медицина.

ВВЕДЕНИЕ

Для современной постгеномной биологии характерно бурное развитие высокопроизводительных экспериментальных методик, позволяющих в одном эксперименте проводить измерения параметров целого генома, транскриптома, протеома. Разработанные ДНК-чиповые и транскриптомные технологии позволяют изучать динамику экспрессии десятков тысяч генов одновременно. Новое поколение методов высокоразрешающей масс-спектрометрии позволяет наблюдать за динамикой изменения концентраций РНК, белков, изучать потоки низкомолекулярных соединений в применении к фундаментальным медицинским проблемам. Новые технологии секвенирования геномов организмов (так называемые технологии секвенирования нового поколения, СНП) обеспечивают недорогой и эффективный способ ресеквенирования геномной ДНК человека, стоимость которого постоянно уменьшается и в ближайшее время может составить менее 1000 USD. Эти достижения позволяют перейти в медицине к новой парадигме, так называемой «персонализированной медицине», современной концепции здравоохранения, сформировавшейся в постгеномную эпоху, которая предполагает при проведении лечения учет индивидуальных геномных особенностей каждого пациента (генетическую предрасположенность к разным заболеваниям, воздействиям различных лекарств и т. п.). В основе персонализированной медицины лежат информация о персональных геномах и современные технологии оценки рисков заболеваний с учетом геномных полиморфизмов.

Выяснение молекулярных механизмов генетической предрасположенности к различным

заболеваниям, таким как сердечно-сосудистые, психоневрологические и онкологические, является одной из основных проблем современной медицинской генетики, молекулярной физиологии и патологии. В рамках этой проблемы в настоящее время во всем мире проводятся широкомасштабные исследования, посвященные изучению связи вариаций геномной последовательности ДНК с различными патологиями.

В настоящее время интенсивно развиваются методы биоинформатики, направленные на оценку влияния полиморфизмов на различных уровнях описания молекулярно-генетических систем: генома, транскриптома, протеома и генных сетей.

Замены одного нуклеотида (однонуклеотидные полиморфизмы, ОНП) – наиболее распространенный и интенсивно изучаемый тип полиморфизма последовательностей ДНК. Данные по ОНП человека в настоящее время получаются в достаточно большом количестве, в частности в результате проектов по полногеномным исследованиям ассоциаций (Genome-Wide Association Studies, GWAS) (Torkamani *et al*., 2008). В связи с этим становится возможным детальное изучение влияния полиморфизмов на возникновение социально значимых заболеваний. Оно основывается преимущественно на двух методиках: а) анализе ассоциаций полиморфизмов с заболеваниями; б) системной биологии.

Первый подход основан на статистическом выявлении взаимосвязей между геномными вариациями и риском заболеваний (Psychiatric GWAS Consortium …, 2009). Например, исследования ассоциаций позволили установить взаимосвязь ряда полиморфизмов с предрасположенностью к раку кожи (Gerstenblith *et al*., 2010). Большое количество результатов таких исследований доступно в виде баз данных (Johnson, O'Donnell, 2009). Для реализации такого подхода необходима кропотливая работа по сбору данных и дальнейшей их верификации методами доказательной медицины. Данный подход имеет ряд недостатков. При статистическом анализе ОНП невозможно разделить функционально значимые полиморфизмы и полиморфизмы, сцепленные с признаком, что требует дополнительных исследований для применения результатов, полученных на одной популяции, перед использованием в клинике. Вторым недостатком является сложность изучения редких полиморфизмов в связи с трудностями создания достаточно большой выборки для статистической обработки.

Альтернативным подходом может служить метод предсказания эффекта единичных замен нуклеотидов *in silico*, основанный на информации о структурно-функциональной аннотации генома человека и особенностей его функционирования в рамках модели генных и метаболических сетей (Weston, 2004). Этот подход основан на том, что ОНП в генах могут влиять на человеческий фенотип на разных уровнях экспрессии гена: полиморфизмы в некодирующих регуляторных областях могут вносить повреждения в последовательности сайтов связывания транскрипционных факторов, сплайсинга, нарушая их функционирование на уровне транскрипции или трансляции. Полиморфизмы в кодирующих участках генов могут становиться причиной аминокислотных замен и приводить к изменениям функциональных или структурных свойств кодируемого белка. В совокупности такие повреждения на уровне отдельных генов могут влиять на функционирование генных сетей (Системная компьютерная биология ..., 2008) и приводить к фенотипическим нарушениям на уровне организма. Современные методы биоинформатики позволяют выявлять повреждающие эффекты мутаций как в некодирующих регуляторных участках генов (Савинкова и др., 2009), так и на уровне белков (Иванисенко и др., 2011).

Подход *in silico*, разумеется, не обладает такой степенью доказательности, как широкомасштабные исследования генетических ассоциаций, тем не менее, в последнее время он интенсивно развивается, так как в дополнение к результатам GWAS позволяет оценивать влияние мутаций на возникновение патологий, на основе современных знаний об их молекулярных механизмах (Na *et al*., 2013).

Необходимо отметить, что в настоящее время разработано большое количество информационных ресурсов (баз данных и компьютерных программ), которые нацелены на решение отдельных задач по оценке влияния ОНП на определенные функции генома (Mooney *et al*., 2010). Однако разнородность этих ресурсов,

огромный объем информации, необходимой для всесторонней оценки риска возникновения патологий, ее сложность не позволяют экспертам-медикам или биоинформатикам систематизировать и анализировать ее вручную. Эффективное решение этой задачи возможно только при использовании компьютерной системы, которая позволяет в автоматическом режиме на основе данных персонального генома проводить первичную оценку рисков возникновения, прежде всего, социально значимых заболеваний. В основе работы такой системы должен лежать принцип интеграции разнородных данных, полученных в результате работы большого числа компьютерных программ при обработке большого числа баз данных. Результаты работы подобной системы, полученные как с учетом известных ассоциаций полиморфизмов с заболеваниями, так и предсказанные на основе биоинформационного анализа предрасположенности к тому или иному заболеванию, могут использоваться медиком-экспертом для учета индивидуальных особенностей пациента.

В настоящей работе описана модульная компьютерная информационная система (МКИС) для оценки влияния полиморфизмов на возникновение социально значимых заболеваний, включающая базы данных, алгоритмы и математические модели влияния полиморфизмов на функцию генов и генных сетей, особенно при таких социально значимых заболеваниях, как рак и заболевания, связанные с нарушением метаболизма.

МАТЕМАТИЧЕСКИЕ МОДЕЛИ И МЕТОДЫ АНАЛИЗА ВЛИЯНИЯ ОНП НА ФУНКЦИЮ ГЕНОВ, СВЯЗАННЫХ С ПОЯВЛЕНИЕМ СОЦИАЛЬНО ЗНАЧИМЫХ ЗАБОЛЕВАНИЙ

Полиморфизм кодирующих районов генов. Достаточно хорошо обоснованной в настоящее время является модель влияния ОНП на структуру и функцию белков – принципиальных компонент генных сетей. Аминокислотные замены в белках могут существенно влиять на функционирование генных сетей и даже приводить к качественному изменению динамики их функционирования. Так, например, мутации в некоторых транскрипционных факторах способны изменять спектр целевых генов, регулируемых данными регуляторными белками (Farnebo *et al*., 2010).

В общем виде мутации по механизму их действия на белок можно разделить на два больших класса: 1) мутации, нарушающие функцию белка, но сохраняющие его пространственную укладку, и 2) мутации, нарушающие структуру (Sanchez-Ruiz *et al*., 2010).

Нарушение функции белка может быть вызвано мутациями, непосредственно расположенными в функциональных центрах белка (каталитических сайтах, сайтах связывания, сайтах посттрансляционных модификаций и т. д.), а также и удаленными от функциональных центров мутациями, действие которых может приводить к изменениям пространственной структуры активных центров белков.

Нарушение структуры белков может проявляться на уровне формирования белковой укладки, т. е. образования белковых форм, принципиально отличающихся по третичной структуре от нативного белка либо вообще не имеющих компактной структуры, либо характеризующихся снижением термодинамической стабильности белков. В последнем случае мутантные белки могут обладать правильной укладкой, однако время существования белка в этой нестабильной форме оказывается ниже по сравнению с нативным белком, и такие белки с более высокой скоростью подвержены протеолитической деградации.

Масштабные исследования полиморфизмов при помощи компьютерных моделей показали, что 90 % единичных мутаций, связанных с заболеваниями, так или иначе понижают стабильность белковой глобулы. В то же время около 70 % наблюдаемых полиморфизмов (из их общего пула) являются, по данным компьютерного моделирования, нейтральными. Примерно 30 % мутаций могут обуславливать заболевания полигенного характера. Использование новых методов компьютерного анализа и представления данных позволило создать информационные ресурсы, посвященные анализу и аннотации ОНП в белках, и привязать информацию об ОНП к структурной и функциональной аннотации белка (Ramensky *et al*., 2002; Cavallo *et al*., 2005; Karchin *et al*., 2005). Это позволяет предсказывать роль ОНП в возникновении за-

болеваний человека и тем самым планировать ассоциативные исследования по мутациям в геноме человека (Yue *et al*., 2006).

Полиморфизм некодирующих районов генов. Гораздо меньше информации имеется о регуляторных ОНП (рОНП), способных влиять на уровень экспрессии генов-кандидатов. Имеющиеся в литературе примеры рОНП показывают, что такие однонуклеотидные замены часто приводят или к разрушению сайтов связывания различных транскрипционных факторов (ТФ), или к образованию новых сайтов, или же изменяют сродство ТФ к их сайтам. Эти события могут не только менять уровень транскрипции генов, но также радикально влиять на характер их экспрессии вплоть до изменений ее тканеспецифичности и способности реагировать на внешние сигналы. С развитием высокопроизводительных экспериментальных подходов к выявлению участков связывания транскрипционных факторов в масштабе генома (ChIP-seq) и накоплением больших массивов этой информации в результате работы международного постгеномного проекта ENCODE (http://genome.ucsc.edu/) появилась возможность масштабной идентификации регуляторных районов в последовательностях генов на основании данных о скоплении мест связывания различных ТФ в определенных геномных локусах. Это, в свою очередь, открывает возможность отбора ОНП, попадающих в эти локусы и потенциально способных влиять на связывание ТФ с данным районом ДНК. Исследования показали высокую предсказательную способность такого подхода к выявлению рОНП. Так, более 70 % нуклеотидных замен, попадающих в такие районы, способны менять связывание белков с ДНК. Таким образом, становится возможным предсказывать участие некодирующих ОНП в развитии различных патологий.

Влияние полиморфизма на функцию генных сетей. Системный подход к реконструкции механизмов, обеспечивающих проявление полиморфизмов через патологические режимы функционирования генных сетей, приобретает все большее значение и демонстрирует свою эффективность. В качестве примера укажем на работу А. Торкамани с соавт. (Torkamani *et al*., 2008), которые провели анализ молекулярных путей в генных сетях, связанных с заболеваниями человека (биполярное аффективное расстройство, заболевание коронарной артерии, болезнь Крона, гипертензия, ревматоидный артрит, диабеты 1 и 2 типов). Анализировались гены, участвующие в генных сетях, а также влияние на их функцию полиморфизмов, данные по которым были получены в результате полногеномных ассоциативных исследований. Результаты анализа показали, что механизмы возникновения патогенных состояний являются общими для многих заболеваний и обусловлены множеством как общих, так специфических факторов риска. В частности, одни и те же сигнальные пути в генных сетях могут отвечать за возникновение сразу нескольких заболеваний. К таким важным путям были отнесены пути передачи сигналов, вовлекающие рецепторы, аденилатциклазы, протеинкиназы, системы передачи кальциевых сигналов.

В качестве еще одного примера укажем, что к настоящему времени уже известны многие десятки мутаций в генных сетях, контролирующих пищевое поведение, ассоциированные с одним и тем же признаком – избыточной массой тела. При этом примечательно, что механизмы действия этих мутаций основаны на нарушениях регуляторных процессов функционирования указанных выше генных сетей.

Складывающийся к настоящему времени биоинформационный подход к выявлению молекулярно-генетических механизмов мультифакториальных заболеваний включает следующие этапы (Moore *et al*., 2010): реконструкция генных сетей, нарушения которых ассоциированы с теми или иными заболеваниями; идентификация генов и белков, участвующих в функционировании этих генных сетей; функциональная оценка влияния мутаций (полиморфизмов) на уровень экспрессии генов либо структуры/функции/активности белков; оценка характера функциональных нарушений на уровне локальных и интегральных генных сетей организма.

РЕАЛИЗАЦИЯ ПРОГРАММНОГО КОМПЛЕКСА SNP-MED

Архитектура программного комплекса SNP-MED. МКИС SNP-MED включает три функциональных модуля, реализованных в виде

программных компонент (ПК) (рис. 1), и базу данных «Информационный ресурс», которая содержит всю необходимую информацию для работы этих компонент.

1. Программная компонента «Геномика» для оценки эффекта влияния ОНП на функционирование регуляторных районов генов. В основе работы этого модуля лежит широкомасштабный поиск совпадений полиморфизмов пациента с известными данными о полиморфизмах и их взаимосвязи с социально значимыми заболеваниями, представленными в общедоступных базах данных, а также предсказание влияния ОНП в регуляторных районах на функцию генов.

2. Программная компонента «Протеомика» для оценки эффекта влияния ОНП в участках генов, кодирующих белки. В основе работы этого модуля лежит широкомасштабный анализ влияния полиморфизмов пациента на нарушение структуры и функций белков, кодируемых генами человека.

3. Программная компонента «Генные сети» для оценки интегрального эффекта влияния ОНП на генные сети. В основе работы этого модуля лежит оценка влияния генетических рисков, найденных в результате работы программных компонент «Геномика» и «Протеомика» на уровне генов и регуляторных взаимодействий, на структуру и функцию генных сетей.

Входные данные в виде последовательности персонального генома или списка ОНП подаются на вход системы пользователем. Эти данные вначале поступают на обработку модулем «Геномика». В результате отбираются ОНП, аннотация которых уже содержится в базах данных, а также ОНП в регуляторных районах генов.

Затем производится обработка ОНП, локализованных в кодирующих участках генов (их влияние на термостабильность и активные центры). На последнем этапе анализа выявленные повреждающие мутации проецируются на структуру генных сетей, которые загружаются пользователем в систему в одном из стандартных форматов. В результате работы ПК пользователь получает аннотацию ОНП, их классификацию по степени повреждающего эффекта, оценку их влияния на функционирование генов, белков и структуру генных сетей.

Для создания МКИС SNP-MED была использована биоинформационная платформа UGENE, которая представляет собой один из широко используемых программных пакетов

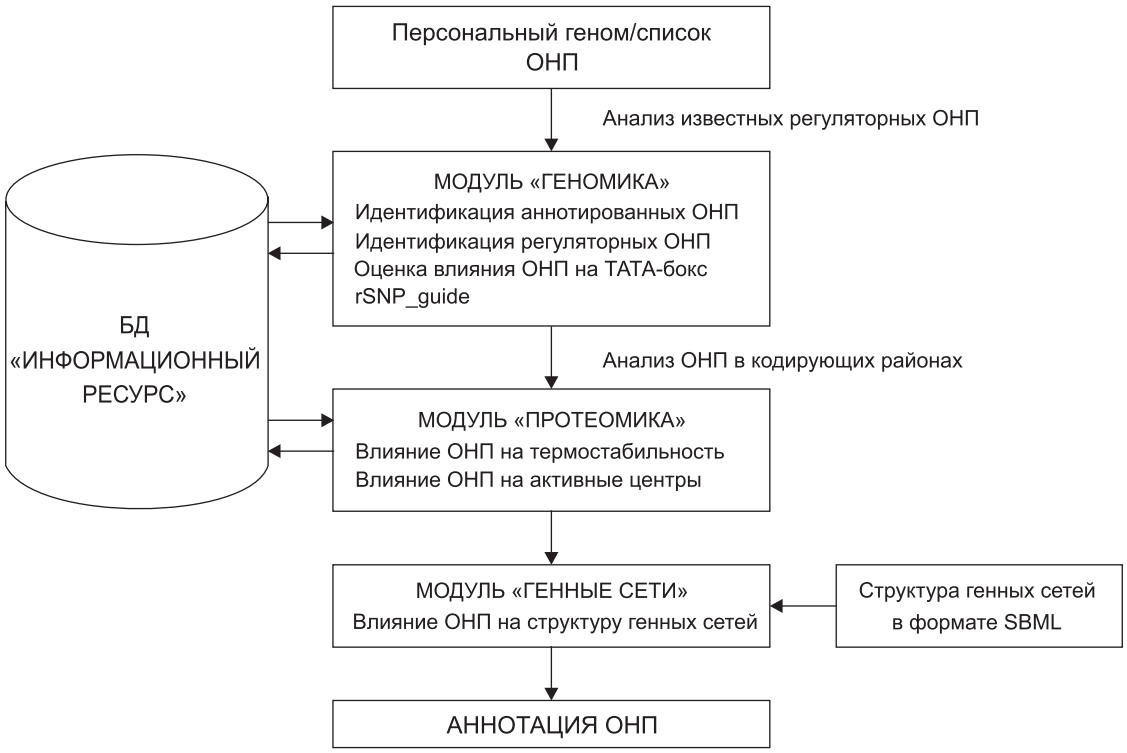


Рис. 1. Основные модули МКИС SNP-MED и их взаимосвязь.

для выполнения различных процедур анализа биологических данных (Okonechnikov, 2012). Высокая популярность UGENE обусловлена, прежде всего, широким спектром его возможностей, кроссплатформенностью, а также его свободным использованием, так как UGENE распространяется на условиях лицензии GNU General Public License, v 2.0.

Для аннотирования однонуклеотидных полиморфизмов используется конструктор вычислительных схем, входящий в состав UGENE, так называемый Workflow Designer (далее WD). Этот модуль позволяет создавать многоступенчатые конвейеры (вычислительные схемы) для обработки биологической информации. Каждая ступень такого конвейера – отдельный алгоритм. В процессе работы схемы они обмениваются между собой сообщениями, содержащими входные данные для одних алгоритмов и/или результаты работы других.

Внутренняя архитектура WD позволяет объединить компоненты МКИС в рамках общего интерфейса вычислительных схем за счет добавления новых вычислительных элементов, соответствующих отдельным алгоритмам. Такой подход позволит в дальнейшем использовать эти компоненты в любых вычислительных схемах WD.

Рассмотрим подробнее функционирование отдельных модулей МКИС SNP-MED.

Программная компонента «Геномика» включает следующие программные модули.

1. Программный модуль для поиска известных ОНП, ассоциированных с социально значимыми заболеваниями, представленных в общедоступных базах данных. К числу этих баз данных относятся такие, как Diseases, dbSNP (Sherry *et al*., 2001), Exome variant server, 1000 Genomes, HapMap. В работе этого модуля также используются последовательность ДНК генома человека и ее функциональная разметка: локализация генов, регуляторных районов, участков сегментных дупликаций и эволюционно-консервативных районов (http://genome.ucsc.edu/). Данный модуль на основе информации о локализации ОНП извлекает доступную аннотацию из упомянутых источников и подает ее на выход модуля.

2. Программный модуль для оценки вероятности нахождения ОНП в регуляторных районах генов. Этот модуль использует информацию по аннотации сайтов связывания десятков транскрипционных факторов с участками генома, полученного в ходе выполнения проекта ENCODE (Rosenbloom *et al*., 2013). Наличие полиморфизмов в этих участках потенциально способно влиять на связывание ТФ с данным районом ДНК. Так, по экспериментальным данным, более 70 % нуклеотидных замен, попадающих в такие районы, способны влиять на связывание белков с ДНК.

3. Программный модуль для оценки влияния ОНП на функционирование районов ТАТА-боксов. В основе анализа – модель взаимодействия белка TBP и фрагмента ДНК, содержащего ТАТА-бокс, описывающая связывание за четыре последовательных шага, отражающих критически значимые этапы функционирования ТАТА-бокса (Пономаренко и др., 2008). Эта модель с высокой точностью позволяет оценить значение аффинности, что было подтверждено экспериментальными исследованиями.

4. Программный модуль для оценки влияния ОНП на функционирование регуляторных районов на основе технологии rSNP_Guide. Данная технология позволяет определить тип транскрипционного фактора, связывание с которым наиболее сильно нарушается в результате мутации в регуляторном районе ДНК.

Программная компонента «Протеомика» включает:

1. Программный модуль для предсказания влияния ОНП на термодинамическую стабильность белков.

2. Программный модуль для идентификации ОНП в функциональных сайтах белков.

База данных «Информационный ресурс» (БДИР) включает информацию, необходимую для функционирования МКИС при анализе влияния ОНП на функцию генов, связанных с появлением социально значимых заболеваний. Содержит аннотацию человеческого генома и известных ОНП, которые использует модуль «Геномика», а также информацию, полученную в результате аннотации ОНП свободно распространяемыми программами:

1) результаты анализа белок-кодирующих последовательностей с использованием алгоритма SIFT (Ng, Henikoff, 2003), оценивающего влияние мутации на функцию белка;

2) результаты анализа белок-кодирующих последовательностей с использованием алгоритма PolyPhen (Adzhubei *et al*., 2010), позволяющего выявить повреждающие мутации в аминокислотных последовательностях;

3) результаты анализа белок-кодирующих последовательностей с использованием алгоритмов PhyloP, LRT и GERP (Pollard *et al*., 2010), оценивающих степень повреждающего эффекта влияния ОНП на основе анализа эволюционной консервативности последовательностей геномов.

Наличие БДИР не требует компьютерных расчетов для каждой ОНП, запрошенной пользователем; программа осуществляет поиск уже подготовленной информации, что обеспечивает быструю обработку данных.

При проектировании программного комплекса большая часть необходимых для аннотирования ОНП алгоритмов была разработана в виде отдельных Web-приложений. В локальном варианте остался доступен лишь анализ влияния полиморфизмов на функционирование TATA-боксов. Соответственно, при разработке функциональности новых вычислительных элементов WD было принято решение для удаленных алгоритмов использовать запросы на соответствующие сервисы, а локальные вызывать непосредственно.

В целом все алгоритмы с программной точки зрения имеют общий интерфейс – требуется передать некоторые атрибуты последовательности, содержащей нуклеотидную замену, а также идентификатор ОНП. На рис. 2 приведена общая схема взаимодействия вычислительных элементов, используемая в данном программном комплексе. Что касается упомянутой процедуры анализа, то в аналогичной диаграмме этапы формирования запроса и получения ответа отсутствуют, а вызов алгоритма производится вычислительным элементом.

Одним из атрибутов каждого вычислительного элемента (за исключением алгоритма «*SNP ChIP*») является локальный путь до базы данных БДИР, содержащей аннотацию генома человека. Обращения к ней позволяют получить параметры последовательности, содержащей полиморфизм, необходимые для вызова каждого из алгоритмов. Затем, используя эти данные, вычислительный элемент вызывает исполнение соответствующего сценария в среде Python, который, в свою очередь, производит обра-

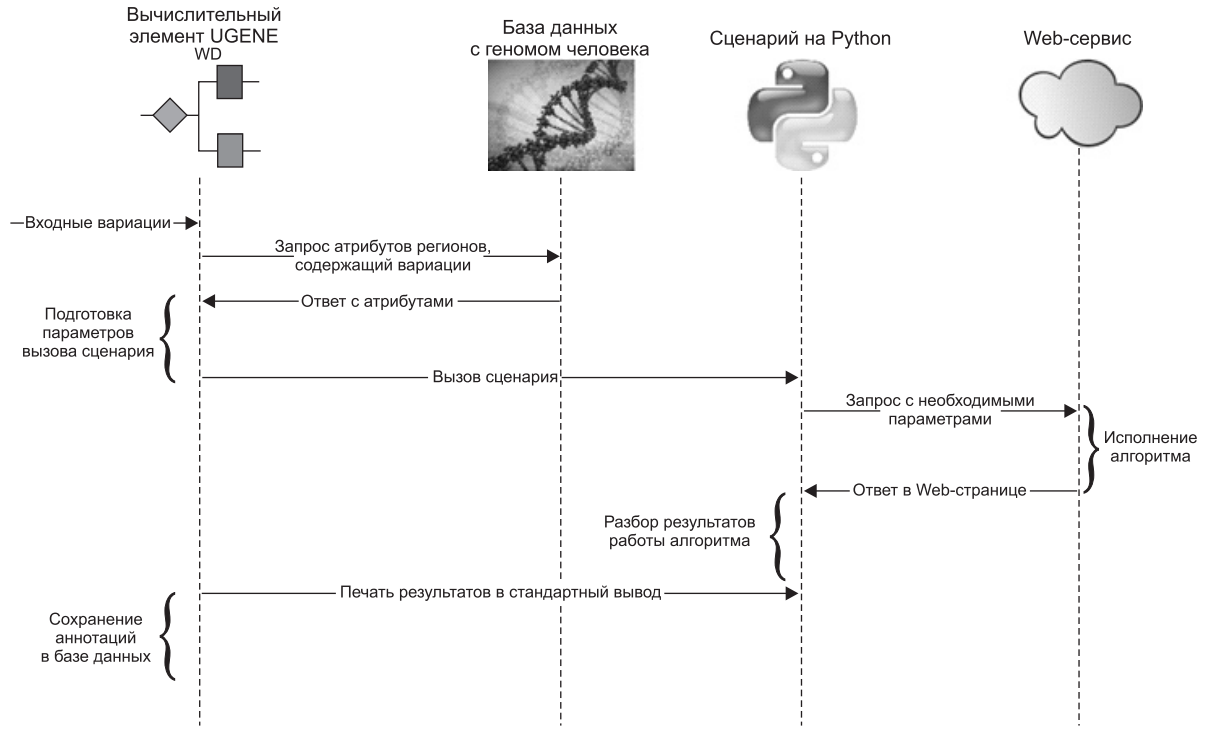


Рис. 2. Диаграмма взаимодействия вычислительного элемента МКИС SNP_MED и БДИР в процессе получения аннотации ОНП на основе использования технологии Workflow Designer UNIPRO.

щение к удаленному сервису, реализующему необходимый алгоритм. После завершения работы сервиса сценарий получает результаты обработки, аннотации ОНП в HTML-формате, которые впоследствии подвергаются разбору. Затем аннотации полиморфизма попадают в стандартный вывод сценария, откуда их считывание производит вычислительный элемент.

На последнем этапе работы того или иного алгоритма полученная информация заносится в общую для всех вычислительных элементов базу данных. Таким образом, последовательно проходя через различные ступени конвейера, аннотация каждой ОНП постепенно дополняется новыми атрибутами в этом хранилище. На основе его содержимого заключительный элемент вычислительной схемы («*Write SNP Report*» на рис. 3) формирует два текстовых отчета – о влиянии полиморфизмов на генную и регуляторную области соответственно (рис. 4, 5).

Данные выдаются в формате SNP-report. Это текстовый файл, в котором аннотация полиморфизма занимает одну или несколько строк. В них располагаются данные в виде столбцов, разделенных знаками табуляции. Число столбцов в каждом из отчетов зависит напрямую от числа задействованных в конвейере алгоритмов. К примеру, для его полной версии (рис. 3) каждой нуклеотидной замене в отчете будут представлены идентификаторы области генома, в котором она обнаружена, а также следующие параметры: идентификатор и местоположение поврежденного гена; идентификатор и кодон поврежденного белка; перечень связанных с поврежденной областью социально значимых заболеваний; оценка повреждающего эффекта с помощью алгоритма SIFT; оценка влияния полиморфизма на термодинамическую стабильность первичной и третичной структур белка; идентификация вариации в функциональных сайтах белка.

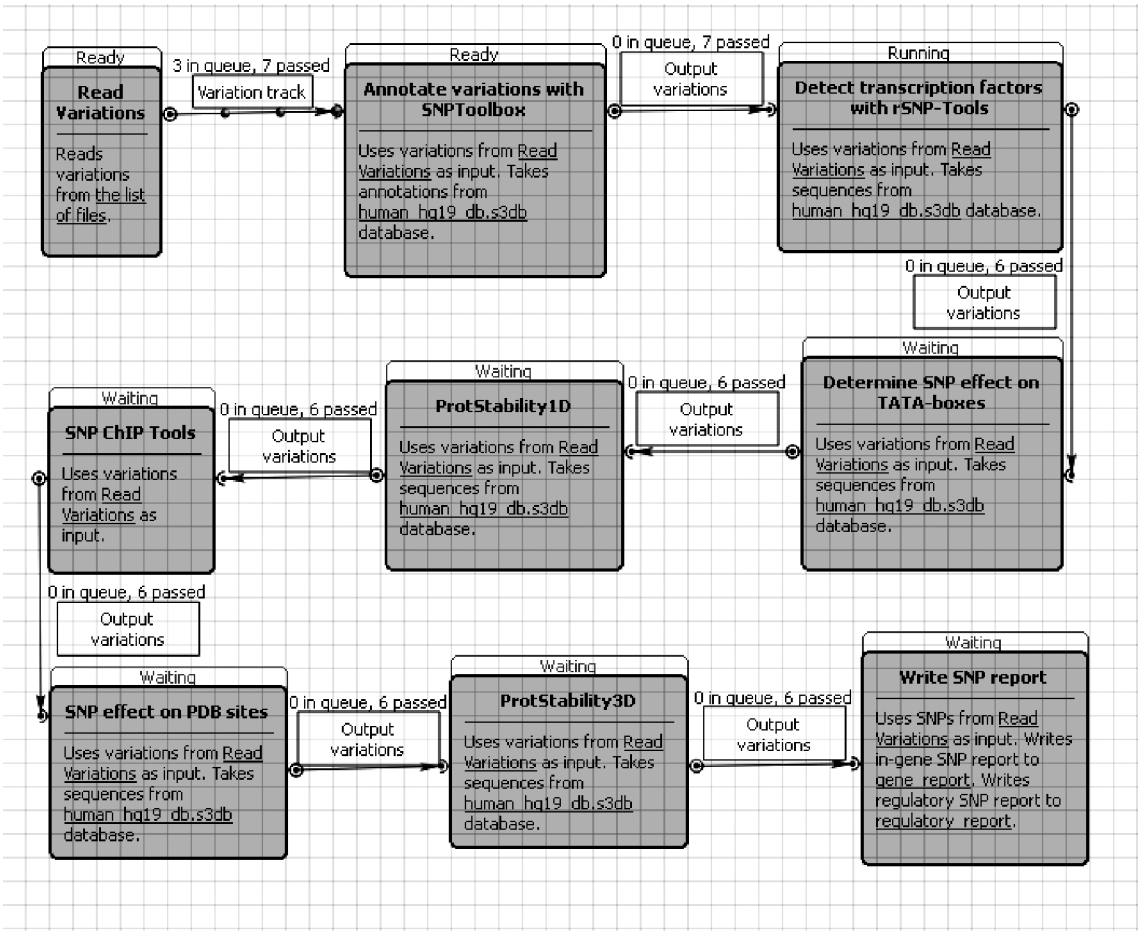


Рис. 3 Детальная последовательность обработки данных в процессе функционирования конвейера МКИС SNP_MED, реализованного в UGENE Workflow Designer в процессе вычислений.

#Chr	Position	Allele	dbSNP	Gene	Clinical_significance	Location	Protein	Codon	Substitution
chr11	94800448	A/G	-	B2R6B8	CDS.	Exon: 94800056..94800900	Q9BRL6	AAC->GAC	N20D
chr13	113634072	G/A	-	AB002360	Factor VII CDS.	Intron: 113623029..113669076	-	-	-
chr13	113634072	G/A	-	CCDS45070	Factor VII CDS.	Intron: 113623029..113669076	-	-	-
chr13	113634072	G/A	-	KIAA0362	Factor VII CDS.	Intron: 113623773..113669076	-	-	-
chr13	113634072	G/A	-	CCDS9527	Factor VII CDS.	Exon: 113633982..113634072. Donor splice-site.	E9PDN8	GAT->AAT	D31N
chr14	22356190	T/A	-	AJ004871	CDS.	Intron: 22192546..22447340	-	-	-
chr14	22356190	T/A	-	FR159098	CDS.	Exon: 22205173..23016490	-	-	-
chr14	22356190	T/A	-	TCRA	Cysticercosis(2.2), Taeniasis(2.2), Leukemia(1.4), Disease(1.2), Toxocariasis(1.2), Echinococcosis(1.1), Cancer(1.0), Lymphoma(0.8), Myoma(0.8), Pneumothorax(0.7), Arthritis(0.5)	Intron: 22294244..22994626	-	-	-
chr14	22356190	T/A	-	FR004500	CDS.	Exon: 22321073..22981890	-	-	-
chr14	22356190	T/A	-	M97714	CDS.	Intron: 22337544..22733720	-	-	-
chr14	22356190	T/A	-	AV2S1A1	CDS.	Exon: 22356037..22356190. Donor splice-site.	-	TGG->AGG	W16R
chr3	126386191	G/C	-	C3orf46	CDS.	Exon: 126385949..126386191. Donor splice-site.	Q8IVU5	AAG->AAC	K133N
chr4	265207	C/A	-	B4DXR9	CDS.	Exon: 264464..266419	B4DXR9	GGC->GTC	G448V
chr4	265207	C/A	-	B4DXR9	CDS.	Exon: 264464..266419	B4DXR9	GGC->GTC	G460V
chr4	152571673	C/G	-	FAM160A1CDS.	CDS.	Exon: 152570611..152571744	Q05DH4	CCA->CGA	P827R
chr5	33997584	G/C	-	AMACR	Cancer(2.0), Adenocarcinoma(2.0), Carcinoma(2.0), Disease(1.8), Adenoma(1.0), Prostatitis(1.0), Angiomyolipoma(0.8), Adrenoleukodystrophy(0.6)	Intron: 33989608..33998745	-	-	-
chr5	33997584	G/C	-	A5YM47	prostate cancer; effects in AMACR are the cause of alpha-methylacyl-CoAracemase deficiency (AMACRD) effects in AMACR are the cause of congenital bile acidsynthesis defect type 4 (CBAS4) CDS.	Intron: 33989608..33998745	-	-	-

Рис. 4. Пример отчета об известных ОНП, ассоциированных с заболеваниями, в формате SNP-report.

Верхняя строка содержит названия колонок отчета. Ниже представлены аннотации полиморфизмов, полученные при работе МКИС SNP-MED.

Пример такого отчета приведен на рис. 4.

Что касается отчета о влиянии ОНП на регуляторную область, то он включает следующую информацию о затронутом ОНП регионе: идентификатор промотора, соответствующего данной области; оценка повреждающего эффекта регуляторной области; перечень связанных с поврежденной областью социально значимых заболеваний; список поврежденных сайтов связывания транскрипционных факторов; оценка влияния вариации на функциональную активность ТАТА-боксов.

Пример аннотационного отчета о регуляторных ОНП приведен на рис. 5.

В рамках МКИС SNP-MED разработаны следующие сценарии биоинформационного анализа влияния ОНП на функции генов, связанных с появлением социально значимых заболеваний.

1. Сценарий биоинформационного анализа данных на геномном уровне для идентификации ОНП, ассоциированных с социально значимыми заболеваниями.

2. Сценарий оценки влияния ОНП на регуляторные районы генов. Пользователю выдается список выявленных ОНП, находящихся в регуляторных районах генов и известных в качестве ассоциированных с социально значимыми заболеваниями, с оценкой их функциональной значимости.

3. Сценарий влияния ОНП на функциональную активность ТАТА-боксов в регуляторных областях генов. Пользователю выдается список генов, ТАТА-боксы которых подвержены значимому изменению уровня активности в результате найденных ОНП, с описанием эффекта ОНП.

4. Сценарий оценки влияния ОНП на функциональную активность сайтов связывания транскрипционных факторов в регуляторных областях генов. Пользователю выдается список генов, у которых сайты связывания транскрип-

#Chr	Position	Allele	dbSNP	Promoter_of_Gene	Clinical_significance	From_transcroption_start	rSNPTools_factors	ChIPTools
chr17	39993435	T/C	-	C9JKC4	-911	-	-	-
chr17	39993435	T/C	-	5NT3L_HUMAN	-911	-	-	-
chr17	39993435	T/C	-	KLH10_HUMAN	-608	-	-	-
chr17	39993435	T/C	-	AK302141	-642	-	-	-
chr17	39993435	T/C	-	AK301797	-642	-	-	-
chr17	39993435	T/C	-	5NT3L_HUMAN	-946	-	-	-
chr17	39993435	T/C	-	C9JKC4	-911	-	-	-
chr17	39993435	T/C	-	5NT3L_HUMAN	-911	-	-	-
chr17	39993435	T/C	-	KLH10_HUMAN	-608	-	-	-
chr17	39993435	T/C	-	AK302141	-642	-	-	-
chr17	39993435	T/C	-	AK301797	-642	-	-	-
chr17	39993435	T/C	-	5NT3L_HUMAN	-946	-	-	-
chr17	39993435	T/C	-	C9JKC4	-911	-	-	-
chr17	39993435	T/C	-	5NT3L_HUMAN	-911	-	-	-
chr17	39993435	T/C	-	KLH10_HUMAN	-608	-	-	-
chr17	39993435	T/C	-	AK302141	-642	-	-	-
chr17	39993435	T/C	-	AK301797	-642	-	-	-
chr17	39993435	T/C	-	5NT3L_HUMAN	-946	-	-	-
chr17	39993435	T/C	-	C9JKC4	-911	-	-	-
chr17	39993435	T/C	-	5NT3L_HUMAN	-911	-	-	-
chr6	32635046	A/G	rs76356512	HLA-DQB1	-579	-	-	-
					Disease(2.0),			
					Lymphopenia(1.3),			
					Scleroderma(1.1),			
					Leukopenia(1.1),			
					Podoconiosis(1.0),			
					Sarcoidosis(1.0),			
					Narcolepsy(0.8),			
					Dermatitis(0.7),			
					Schizophrenia(0.7),			
					Elephantiasis(0.7),			
					Thyroiditis(0.7),			
					Glomerulonephritis(0.7),			
					Hyperthyroidism(0.6),			
					Arthritis(0.6),			
					Nephritis(0.6),			
					Cancer(0.6),			
					Agammaglobulinemia(0.6)			
chr6	32635046	A/G	rs76356512	A2AAZ0	-579	-	-	-
					Leprosy; diabetes,			
					type 1; pancreatitis,			
					autoimmune;			
					pancreatitis, chronic			
					calcifying;			
					periodontitis;			
					infertility, tubal			

Рис. 5. Пример отчета о влиянии полиморфизмов на регуляторную область в формате SNP-report.

Верхняя строка содержит названия колонок отчета. Ниже представлены аннотации полиморфизмов, полученные при работе МКИС SNP-MED.

ционных факторов подвержены изменению функциональной активности в результате найденных ОНП, с описанием эффекта ОНП.

5. Типовой сценарий биоинформационного анализа данных на протеомном уровне, включая предсказание влияния ОНП на термодинамическую стабильность белков, идентификацию ОНП в функциональных сайтах белков.

6. Типовой сценарий биоинформационного анализа влияния ОНП на генные сети.

ЗАКЛЮЧЕНИЕ

Разработана модульная компьютерная информационная система SNP-MED для анализа влияния ОНП на функцию генов, связанных с появлением социально значимых заболеваний, в состав которой входят:

1. Программная компонента «Геномика», включающая программные модули или интерфейсы к сервисам поиска известных ОНП, ассоциированных с социально значимыми заболеваниями, оценки вероятности нахождения ОНП в регуляторных районах генов, оценки влияния ОНП на функционирование районов ТАТА-боксов и регуляторных районов.

2. Программная компонента «Протеомика», включающая программные модули или интерфейсы к сервисам идентификации известных ОНП в кодирующих участках генов, ассоциированных с социально значимыми заболеваниями, предсказания влияния ОНП на термодинамическую стабильность белков и идентификации ОНП в функциональных сайтах белков.

3. Программная компонента «Генные сети», позволяющая оценивать эффект влияния ОНП на генные сети.

4. База данных «Информационный ресурс», включающая информацию, необходимую для функционирования МКИС SNP-MED при анализе влияния ОНП на функцию генов, связанных с появлением социально значимых заболеваний.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке Министерства образования и науки Российской Федерации (Госконтракт № 14.512.11.0094).

ЛИТЕРАТУРА

Иванисенко В.А., Деменков П.С., Иванисенко Т.В., Колчанов Н.А. Protein structure discovery: пакет программ для решения задач компьютерной протеомики // Биоорган. химия. 2011. Т. 37. № 1. С. 22–35.

Пономаренко П.М., Савинкова Л.К., Драчкова И.А. и др. Пошаговая модель связывания TBP/TATA-бокс позволяет предсказать наследственное заболевание человека по точечному полиморфизму // Докл. АН. 2008. Т. 419. С. 828–832.

Савинкова Л.К., Пономаренко М.П., Пономаренко П.М. и др. Полиморфизмы TATA-боксов промоторов генов человека и ассоциированные с ними наследственные патологии // Биохимия. 2009. Т. 4. № 4. С. 149–163.

Системная компьютерная биология // Под ред. Н.А. Колчанов, С.С. Гончаров, В.А. Лихошва, В.А. Иванисенко. Новосибирск: СО РАН, 2008.

Adzhubei I.A., Schmidt S., Peshkin L. *et al.* A method and server for predicting damaging missense mutations // Nature Meth. 2010. V. 7. No. 4. P. 248–249.

Cavallo A., Martin A.C. Mapping SNPs to protein sequence and structure data // Bioinformatics. 2005. V. 21. P. 1443–1450.

Farnebo M., Bykov V.J., Wiman K.G. The p53 tumor suppressor: a master regulator of diverse cellular processes and therapeutic target in cancer // Biochem. Biophys Res. Commun. 2010. P. 85–89.

Gerstenblith M.R., Shi J., Landi M.T. Genome-wide association studies of pigmentation and skin cancer: a review and meta-analysis // Pigment Cell Melanoma Res. 2010. V. 23. No. 5. P. 587–606.

Johnson A.D., O'Donnell C.J. An open access database of genome-wide association results // BMC Med. Genet. 2009. V. 10. No. 1. P. 6.

Karchin R., Diekhans M., Kelly L. *et al.* LS-SNP: large-scale annotation of coding non-synonymous SNPs based on multiple information sources // Bioinformatics. 2005. V. 21. P. 2814–2820.

Mooney S.D., Krishnan V.G., Evani U.S. Bioinformatic tools for identifying disease gene and SNP candidates // In Genetic Variation. 2010. P. 307–319.

Moore J.H., Asselbergs F.W., Williams S.M. Bioinformatics challenges for genome-wide association studies // Bioinformatics. 2010. V. 26. P. 445–455.

Na Y.J., Cho Y., Kim J.H. AnsNGS: An annotation system to sequence variations of next generation sequencing data for disease-related phenotypes // Healthcare Inform. Res. 2013. V. 19. No. 1. P. 50–55.

Ng P.C., Henikoff S. SIFT: Predicting amino acid changes that affect protein function // Nucl. Acids Res. 2003. V. 31. P. 3812–3814.

Okonechnikov K., Golosova O., Fursov M. *et al*. Unipro UGENE: a unified bioinformatics toolkit // Bioinformatics. 2012. V. 28. P. 1166–1167.

Pollard K.S., Hubisz M.J., Rosenbloom K.R., Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies // Genome Res. 2010. V. 20. No. 1. P. 110–121.

Psychiatric GWAS Consortium Steering Committee. A Framework for Interpreting Genome-Wide Association Studies of Psychiatric Disorders // Mol. Psychiatry. 2009. V. 14. No. 1. P. 10.

Ramensky V., Bork P., Sunyaev S. Human non-synonymous SNPs: server and survey // Nucl. Acids Res. 2002. V. 30. P. 3894–3900.

Rosenbloom K.R., Sloan C.A., Malladi V.S. *et al.* ENCODE Data in the UCSC Genome Browser: year 5 update // Nucl. Acids Res. 2013. P. D56–D63.

Sanchez-Ruiz J.M. Protein kinetic stability // Biophys. Chem. 2010. V. 148. P. 1–15.

Sherry S.T., Ward M.H., Kholodov M. *et al.* dbSNP: the NCBI database of genetic variation // Nucl. Acids Res. 2001. No. 29. P. 308–311.

Torkamani A., Topol E.J., Schork N.J. Pathway analysis of seven common diseases assessed by genome-wide association // Genomics. 2008. No. 92. P. 265–272.

Weston A.D., L H. Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine // J. Proteome Res. 2004. V. 3. No. 2. P. 179–196.

Yue P., Melamud E., Moult J. SNPs3D: Candidate gene and SNP selection for association studies // BMC Bioinformatics. 2006. No. 7. P. 166.

THE SNP-MED SYSTEM FOR ANALYSIS OF THE EFFECT OF SINGLE-NUCLEOTIDE POLYMORPHISMS ON THE FUNCTION OF GENES ASSOCIATED WITH SOCIALLY SIGNIFICANT DISEASES

N.L. Podkolodnyy[1], D.A. Afonnikov[1], Yu.Yu. Vaskin[2], L.O. Bryzgalov[1], V.A. Ivanisenko[1], P.S. Demenkov[1], M.P. Ponomarenko[1], D.A. Rasskazov[1], K.V. Gunbin[1], I.V. Protsyuk[2], I.Yu. Shutov[2], P.N. Leontyev[2], M.Yu. Fursov[2], N.P. Bondar[1], E.V. Antontseva[1], T.I. Merkulova[1], N.A. Kolchanov[1]

[1] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: pnl@bionet.nsc.ru;
[2] Novosibirsk Center of Information Technologies «UNIPRO», Novosibirsk, Russia

Summary

This paper describes the SNP-MED modular computer-based information system for estimation of the influence of single nucleotide polymorphisms (SNPs) on the function of genes associated with the risk of socially significant diseases. The system includes software components Genomics, Proteomics, Gene networks and the Information resource database (BDIR).

Key words: bioinformatics, SNP, personalized medicine.

УДК 577.113.3:57.087.1

ОЦЕНКА ПО ТЕХНОЛОГИИ rSNP_Guide SNPs ПРОМОТОРОВ ГЕНОВ *APC* И *MLH1* ЧЕЛОВЕКА, СВЯЗАННЫХ С РАКОМ ТОЛСТОГО КИШЕЧНИКА

© 2013 г. **Д.А. Рассказов[1], Е.В. Антонцева[1], Л.О. Брызгалов[1], М.Ю. Матвеева[1], Е.В. Кашина[1], П.М. Пономаренко[1], Г.В. Орлова[1], М.П. Пономаренко[1], Д.А. Афонников[1, 2], Т.И. Меркулова[1, 2]**

[1] Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия, e-mail: pon@bionet.nsc.ru;
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

Каждый из 6 регуляторных SNPs (Single nucleotide polymorphisms) генов *APC* и *MLH1* человека (rs75996864, rs76241113, rs78037487, rs80112297, rs80313086 и rs1800734) был оценен по созданной ранее технологии rSNP_Guide на значимость изменения связывания каждого из 40 факторов транскрипции с соответствующими районами ДНК. В результате для каждого SNP все анализируемые белки были ранжированы по убыванию уровня статистической значимости α (*t*-тест Стьюдента) изменения их сродства к аллельным вариантам указанной ДНК. Установлено, что самыми вероятными проявлениями SNPs rs75996864, rs76241113, rs78037487, rs80112297 и rs80313086 гена *APC*, а также SNP rs1800734 гена *MLH1* человека являются изменения в связывании именно тех транскрипционных факторов (NF-Y, NFkB, c-Myb, RAR, YY-1, Sp-1), для которых ранее было показано участие в развитии рака толстого кишечника. Полученные результаты служат новым основанием для исследований ассоциации SNPs rs75996864, rs76241113, rs78037487, rs80112297, rs80313086 гена *APC* с раком толстого кишечника общепринятыми медико-генетическими методами.

Ключевые слова: однонуклеотидный полиморфизм, регуляция экспрессии генов, рак толстого кишечника, *APC*, *MLH1*, комплекс ДНК с регуляторным белком, *t*-тест Стьюдента.

ВВЕДЕНИЕ

Начало III тысячелетия н. э. было ознаменовано эпохальным достижением науки в области молекулярной биологии – расшифровкой генома человека. В 2004 г. было завершено секвенирование так называемого «референсного» (т. е. общепринятого стандарта) генома человека (The International Human Genome Sequencing Consortium, 2004). Это событие считается началом новой постгеномной эры науки о жизни. Для нее характерна быстрая расшифровка индивидуальных геномов пациентов относительно референсного генома человека, что закладывает основы для развития персонализированной медицины с возможностью диагностики, терапии и мониторинга заболеваний с учетом генетической предрасположенности и индивидуальной чувствительности/устойчивости к лекарственным препаратам.

В этой связи интенсивно ведется статистическое выявление ассоциаций между геномными вариациями и риском заболеваний, NHGRI GWAS catalog (Hindorff *et al*., 2009), что позволило, в частности, связать наличие ряда однонуклеотидных замен (SNP, Single Nucleotide Polymorphism) с генетической предрасположенностью к раку кожи (Gerstenblith *et al*., 2010). Экспериментально установленные случаи полиморфизма генов человека докумен-

тируются в базе данных dbSNP (NCBI Resource Coordinators, 2013), ассоциации полиморфизма с патологиями – в базе данных OMIM (Hamosh *et al.*, 2005).

Особые успехи были достигнуты при определении молекулярных механизмов негативного действия SNPs, расположенных в белок-кодирующих районах геномов, вследствие относительной простоты выяснения возможных причин нарушения функции белка: с сохранением пространственной укладки либо с ее нарушением (Sanchez-Ruiz, 2010). Однако молекулярные механизмы влияния SNPs из некодирующих районов генов на возникновение патологий остаются в своем большинстве неясными. В настоящее время выявлено множество регуляторных SNPs, связанных с проявлением различных патологий. В частности, минорный аллель rs1800734 (идентификатор базы данных dbSNP) гена *MLH1* человека, кодирующего белок-гомолог гена *mutL* репарации *E. coli* (Win *et al.*, 2013), был ассоциирован с раком толстого кишечника (Hitchins *et al*., 2007).

В данной работе мы исследовали аллели rs75996864, rs76241113, rs78037487, rs80112297 и rs80313086 гена *APC* (adenomatous polyposis coli – аденоматозного полипоза толстого кишечника), кодирующего белок-супрессор опухоли (Polakis, 2011). Для этих SNPs ранее (Антонцева и др., 2011; Antontseva *et al*., 2012) впервые были получены данные по задержке в геле (EMSA, electrophoretic mobility shift assay) белками ядерного экстракта из раковой клеточной линии HCT-116 (рак толстого кишечника). С помощью технологии rSNP_Guide (Ponomarenko *et al*., 2001, 2002) было установлено, что самые статистически достоверные (*t*-тест Стьюдента) аллельные изменения при связывании с белком/белками присущи именно тем участкам ДНК гена *APC*, соответствующим потенциальным сайтам связывания транскрипционных факторов, участие которых в канцерогенезе толстого кишечника уже было экспериментально доказано. Полученные компьютерно-экспериментальные данные являются основанием для проведения дальнейшего стандартного медико-генетического исследования связи SNPs rs75996864, rs76241113, rs78037487, rs80112297 и rs80313086 гена *APC* с раком толстого кишечника человека.

МАТЕРИАЛЫ И МЕТОДЫ

Из базы данных dbSNP (NCBI Resource Coordinators, 2013) были извлечены последовательности нуклеотидов $S = \{s_{-25}\ldots s_{-1}(s_0^{WT} / s_0^{minor}) s_1\ldots s_{25}\}$ с аллельной вариацией в центре предковых и минорных аллелей rs1800734, ранее уже ассоциированных с раком толстого кишечника гена *MLH1* человека, и исследуемых 5 SNPs, rs75996864, rs76241113, rs78037487, rs80112297 и rs80313086 гена *APC* человека (табл. 1–3).

В табл. 1 и 3 представлены взятые из базы данных SNPChiPTools (Антонцева и др., 2011; Antontseva *et al*., 2012) результаты эксперимента по задержке в геле двуцепочечных олигонуклеотидов, несущих предковые (верхняя дорожка) или минорные (нижняя дорожка) аллели, с белками экстрактов ядер раковых клеток линий HCT-116 (рак толстого кишечника), HeLaS3 (рак шейки матки), K562 (эритролейкемия) и HepG2 (гепатома).

Эти экспериментальные данные были исследованы по технологии rSNP_Guide (Ponomarenko *et al*., 2001, 2002), как это показано на рис. 1 и в табл. 2 на примере аллелей rs1800734, ассоциированных с раком толстого кишечника, гена *MLH1* человека (табл. 1, колонка «HCT-116»).

После ввода «http://samurai.bionet.nsc.ru/cgi-bin/03/programs/rsnp_lin/rsnpd.pl» в Интернет браузер пользователь получает начальную форму (рис. 1, а), которую он должен заполнить исследуемыми им экспериментальными данными. Применяя команду «Calculate», пользователь автоматически получает результат анализа этих данных по технологии rSNP_Guide (рис. 1, б). Дополнительная команда «Intermediate Report» позволяет получить все данные промежуточных расчетов без исключения (табл. 2). В первой колонке этой таблицы приведены общепринятые обозначения 40 белков-факторов транскрипции, которые представлены в текущей версии rSNP_Guide. В следующих колонках ii–v даны оценки наибольшего сродства этих белков к анализируемой двунитевой ДНК, полученные на основе анализа контекстных мотивов сайтов связывания этих белков, найденные во «введенной» нити (+) ДНК (рис. 1) и в комплементарной к ней (−) нити ДНК предкового и минорного аллелей «введенного» SNP. В текущей версии rSNP_Guide (Ponomarenko *et al.*, 2001, 2002)

эти оценки уровней сродства ДНК/белок соответствуют теории статистической механики взаимодействия регуляторных белков с ДНК (Berg, von Hippel, 1987) на основе построения позиционно-весовых матриц олигонуклеотидов ДНК, как это было описано ранее (Ponomarenko *et al*., 1999). На основе вычисления среднеарифметического M_0 отрицательных

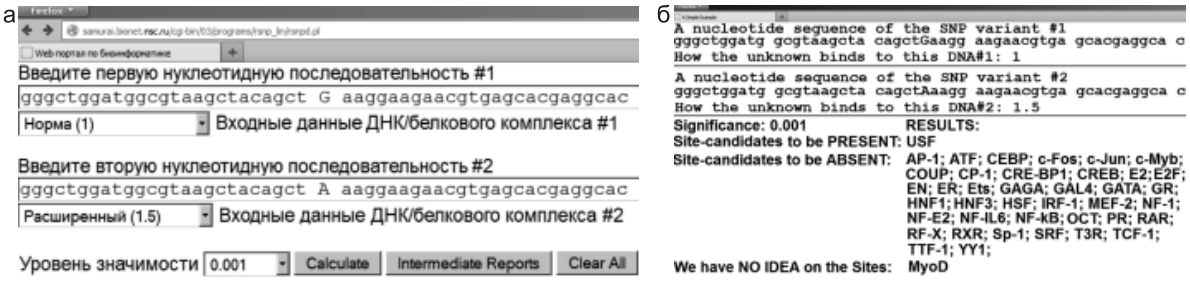
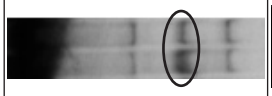
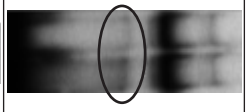
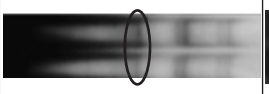


Рис. 1. Интерфейс ввода (а) и вывода (б) технологии rSNP_Guide на примере анализа экспериментальных данных о задержке в геле двуцепочечных олигонуклеотидов, несущих rs1800734, с белками ядерных экстрактов из раковой клеточной линии HCT-116 (рак толстого кишечника, табл. 1). Комплементарная нить к введенной нити ДНК строится автоматически для их совместного анализа.

Таблица 1

Оценка по технологии rSNP_Guide ассоциированных с раком толстого кишечника аллелей rs1800734 гена *MLH1* человека

Аллель	Последовательность ДНК, ген *MLH1*, rs1800734 ± 25 п.о.			
Предок, WT минорный	gggctggatggcgtaagctacagct**G**aaggaagaacgtgagcacgaggcac gggctggatggcgtaagctacagct**A**aaggaagaacgtgagcacgaggcac			
Клетки Рак	HCT-116 толстого кишечника	HeLaS3 шейки матки	HepG2 гепатома	K562 эритролейкемия
Задержка в геле				
Комплекс[#]	MINOR больше	WT меньше	WT больше	возник слабый MINOR
Данные для rSNP_Guide	X_{WT} = 1,0; X_{MINOR} = 1,5	X_{WT} = 0,5; X_{MINOR} = 1,0	X_{WT} = 1,0; X_{MINOR} = 0,5	X_{WT} = 0,0; X_{MINOR} = 0,5
Ранг (значимость)	rSNP_Guide: ранжирование транскрипционных факторов-кандидатов			
I ($\alpha < 0{,}001$)	USF			
II ($\alpha < 0{,}0025$)		USF	USF, MyoD	
III ($\alpha < 0{,}005$)	Ets, GR		GR	
IV ($\alpha < 0{,}01$)	ATF, MyoD, YY1	ATF, c-Jun, Ets, GR, YY1	ATF, Ets, YY1	c-Myb
V ($\alpha < 0{,}025$)	c-Jun, c-Myb, CREB; GAGA, GATA, NF-IL6, RF-X	c-Fos, c-Myb, CRE-BP1, CREB, GAGA, GATA, MyoD, NF-IL6, RF-X	c-Jun, NF-IL6, CREB, GAGA, RF-X	ATF, c-Fos, c-Jun, CRE-BP1, CREB, GATA, T3R
VI ($\alpha < 0{,}05$)	c-Fos, CEBP, PR, CRE-BP1, RAR, T3R, TCF-1	CEBP, PR, RAR, T3R, TCF-1	CEBP, c-Fos, RAR, c-Myb, PR, CRE-BP1	NF-IL6, RAR, RF-X

[#] Здесь и далее: анализируемый комплекс ДНК/белок (выделен овалом) был охарактеризован референтом базы данных SNPChiPTools (Антонцева и др., 2011; Antontseva *et al*., 2012).

Таблица 2

Пример обоснования полученной по технологии rSNP_Guide оценки ассоциированного с раком толстого кишечника полиморфизма rs1800734 промотора гена *MLH1* человека (транскрипционный фактор-кандидат USF, для MyoD мало данных)

Фактор транскрипции, ТФ	Сродство ТФ/ДНК, нити (+)/(–)				Сходство Декарта между ТФ и моделью SNP				*t*-тест Стьюдента ($\alpha < 0{,}001$) соответствия ТФ модели SNP			
	Предок, WT		Минорный аллель									
	(+)	(–)	(+)	(–)	X_{++}	X_{+-}	X_{-+}	X_{--}	X_{++}	X_{+-}	X_{-+}	X_{--}
i	ii	iii	iv	v	vi	vii	viii	ix	x	xi	xii	xiii
AP-1	–0,09	0,01	–0,09	0,01	2,62	1,98	1,80	0,52	–1,19	–1,11	–1,06	0,94
ATF	0,39	0,33	0,39	0,33	1,83	1,54	1,66	1,32	–0,40	–0,68	–0,91	0,14
CEBP	–0,51	0,38	–0,37	0,23	2,79	2,56	1,42	0,89	–1,35	–1,70	–0,68	0,57
c-Fos	0,25	0,20	0,25	0,20	2,10	1,61	1,71	1,05	–0,67	–0,75	–0,96	0,41
c-Jun	0,36	0,30	0,36	0,30	1,90	1,55	1,67	1,26	–0,46	–0,69	–0,92	0,20
c-Myb	–0,00	0,35	0,26	0,35	2,06	1,83	1,46	1,11	–0,62	–0,97	–0,71	0,35
COUP	–0,32	0,10	–0,32	0,08	2,80	2,32	1,67	0,56	–1,37	–1,46	–0,92	0,91
CP-1	–0,34	–0,12	–0,15	–0,12	2,90	2,14	1,97	0,29	–1,46	–1,28	–1,23	1,17
CRE-BP1	0,32	0,12	0,32	0,12	2,12	1,47	1,85	1,06	–0,68	–0,61	–1,11	0,40
CREB	0,19	0,29	0,19	0,28	2,08	1,74	1,56	1,08	–0,64	–0,88	–0,82	0,38
E2	–0,37	–0,12	–0,35	–0,13	3,05	2,32	1,99	0,25	–1,61	–1,46	–1,24	1,21
E2F	–0,16	–0,44	–0,16	–0,44	3,16	2,04	2,43	0,27	–1,72	–1,18	–1,68	1,19
EN	–0,20	–0,53	–0,22	–0,31	3,17	2,11	2,37	0,26	–1,73	–1,25	–1,63	1,20
ER	–0,21	–0,30	–0,29	–0,30	3,10	2,17	2,22	0,07	–1,67	–1,30	–1,48	1,39
Ets	–0,39	0,57	–0,46	0,62	2,60	2,72	0,99	1,28	–1,16	–1,86	–0,25	0,18
GAGA	0,46	–0,60	0,46	–0,60	2,89	1,24	2,86	1,16	–1,46	–0,38	–2,12	0,30
GATA	0,11	–0,01	0,57	–0,01	2,22	1,34	2,05	1,04	–0,78	–0,48	–1,31	0,42
GR	0,56	–0,12	0,56	–0,12	2,22	1,05	2,32	1,24	–0,79	–0,19	–1,57	0,22
HNF1	–0,35	–0,05	–0,28	–0,00	2,89	2,27	1,83	0,38	–1,46	–1,41	–1,09	1,08
HNF3	–0,11	–0,40	–0,07	–0,40	3,05	1,93	2,39	0,32	–1,62	–1,07	–1,64	1,14
HSF	–0,41	–0,37	–0,41	–0,37	3,32	2,38	2,32	0,19	–1,88	–1,51	–1,58	1,27
IRF-1	0,03	–0,47	0,03	–0,47	3,01	1,76	2,50	0,53	–1,58	–0,90	–1,76	0,93
MEF-2	–0,11	–0,37	0,08	–0,37	2,93	1,80	2,35	0,43	–1,49	–0,94	–1,61	1,03
MyoD	0,57	0,80	0,12	0,40	1,81	1,94	1,47	1,63	–0,37	–1,08	–0,72	–0,16
NF-1	–0,33	–0,41	–0,37	–0,04	3,11	2,32	2,09	0,29	–1,67	–1,45	–1,35	1,18
NF-E2	0,08	–0,20	–0,23	–0,20	2,86	1,96	2,13	0,41	–1,43	–1,10	–1,38	1,06
NF-IL6	0,15	0,48	0,15	0,48	1,95	1,93	1,30	1,27	–0,51	–1,07	–0,56	0,19
NF-kB	–0,22	–0,47	–0,22	–0,47	3,24	2,12	2,46	0,26	–1,80	–1,26	–1,71	1,20
OCT	–0,37	–0,31	–0,23	–0,33	3,16	2,21	2,26	0,10	–1,72	–1,35	–1,51	1,36
PR	–0,37	0,38	–0,37	0,26	2,70	2,49	1,38	0,89	–1,26	–1,62	–0,63	0,57
RAR	0,32	0,07	0,21	0,07	2,23	1,54	1,88	0,96	–0,79	–0,67	–1,14	0,50
RF-X	0,41	–0,10	0,33	–0,18	2,40	1,32	2,22	0,97	–0,96	–0,46	–1,48	0,49
RXR	–0,14	0,16	–0,14	0,11	2,56	2,09	1,63	0,66	–1,13	–1,23	–0,88	0,80
Sp-1	–0,31	–0,84	–0,31	–0,80	3,71	2,36	2,95	0,74	–2,27	–1,50	–2,21	0,72
SRF	–0,25	–0,56	–0,02	–0,51	3,22	2,00	2,56	0,44	–1,78	–1,13	–1,82	1,02
T3R	0,24	0,04	0,24	0,04	2,26	1,53	1,90	0,91	–0,83	–0,67	–1,15	0,55
TCF-1	0,16	–0,54	0,28	–0,54	2,95	1,51	2,67	0,82	–1,51	–0,64	–1,92	0,64
TTF-1	0,03	–0,26	0,03	–0,18	2,74	1,76	2,15	0,48	–1,30	–0,90	–1,41	0,99
USF	0,77	0,53	0,77	0,53	1,31	1,39	1,86	1,92	0,12	–0,53	–1,12	0,45
YY1	0,50	–0,24	0,50	–0,37	2,51	1,12	2,52	1,13	–1,08	–0,25	–1,77	0,33

оценок конструируются все четыре возможные модели влияния заданного SNP на комплексы ДНК/ белок: модель X_{--} оценивает неспецифическое связывание ДНК/белок для обеих нитей ДНК каждого из аллелей (колонки (ix и xiii)); модель X_{++} – специфическое связывание ДНК/белок введенными в rSNP_Guide (рис. 1) M_{WT} и M_{MINOR}; модели X_{+-} и X_{-+} – специфическое связывание одной из двух нитей ДНК. В колонках vi–ix для каждой из этих четырех моделей влияния «введенного» SNP на связывание ДНК/белок представлены оценки их сходства с изменениями сродства каждого из рассматриваемых 40 белков к регуляторной ДНК с этим SNP, полученные с помощью наиболее часто используемой Декартовой меры сходства.

Наконец, в колонках (x–xiii) для каждой модели представлены отклонения ее значений от 95 %-й доверительной границы *t*-теста Стьюдента для гипотезы «H_0: заданная SNP неразличимо одинаково меняет связывание ДНК с любым регуляторным белком». На этой основе rSNP_Guide автоматически генерирует выходные данные (рис. 1).

Положительные оценки указывают на те из 40 анализируемых белков, сродство которых к содержащей SNP последовательности ДНК изменяется в значимом соответствии с моделью патогенного изменения комплекса ДНК/белок во введенных данных эксперимента по задержке в геле для «введенного» SNP (рис. 1). Одновременное несоответствие модели X_{--} при соответствии любой другой модели SNP указывает те белки, для которых изменения их связывания с ДНК соответствуют «введенным» данным (рис. 1). При одновременном отсутствии достоверного сходства всех четырех возможных моделей SNP с каким-либо из анализируемых 40 белков (в качестве примера в табл. 2 рассмотрен транскрипционный фактор MyoD) по технологии rSNP_Guide принималось решение о невозможности оценить влияние «введенной SNP» на связывание этого белка с ДНК. Такое решение принимается и при одновременном достоверном сходстве какого-либо белка как с моделью X_{--}, так и с любой из X_{++}, X_{+-} или X_{-+} моделей. В остальных случаях rSNP_Guide принимает решение о недостоверном влиянии «введенного SNP» на связывание заданного белка с ДНК.

Результатом rSNP_Guide для ассоциированного с раком толстого кишечника SNP rs1800734 была оценка в качестве самых достоверных аллельных различий гена *MLH1* с теми транскрипционными факторами, участие которых в канцерогенезе толстого кишечника было ранее экспериментально доказано (табл. 1): USF (Bruno *et al*., 2004; Ansorge *et al*., 2007; Pare *et al*., 2008; Belanger *et al.*, 2010; Christensen *et al*., 2013), Ets (Wai *et al*., 2006) и GR (например, Byrne *et al*., 2010). Этот факт послужил отправной точкой для исследования 5 аллелей: rs75996864, rs76241113, rs78037487, rs80112297 и rs80313086 гена *APC* человека, которые не были еще ассоциированы с раком толстого кишечника и для которых в предыдущей статье (Антонцева и др., 2011; Antontseva *et al*., 2012) были представлены первые аргументы в пользу такой ассоциации.

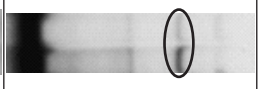
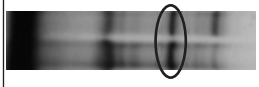
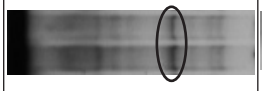
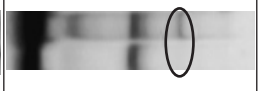
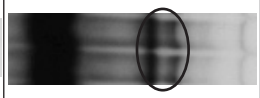
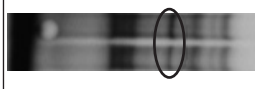
РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Результаты оценки по технологии rSNP_Guide пяти SNPs: rs75996864, rs76241113, rs78037487, rs80112297 и rs80313086 гена *APC* человека даны в табл. 3, Для rs75996864 технология rSNP_Guide предсказала самым вероятным белком-кандидатом, сайт связывания которого с ДНК был изменен, фактор транскрипции RAR. Это согласуется с независимыми экспериментальными данными об участии этого белка в канцерогенезе толстого кишечника (Wei *et al*., 2003; Groubet *et al*., 2003; Xu *et al*., 2004; Mulholland *et al*., 2005; Kameue *et al*., 2006). Следующими по значимости кандидатами были транскрипционные факторы YY1 и c-Myb, участие которых в развитии рака толстого кишечника было ранее независимо показано (Chinnappan *et al*., 2009; Ramsay *et al*., 2003).

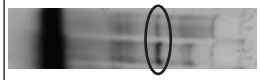
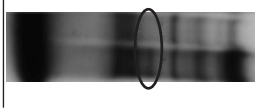
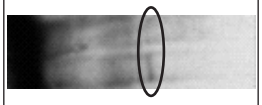
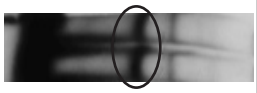
Для четырех других SNPs, rs76241113, rs80112297, rs80313086 и rs78037487, по оценке технологии rSNP_Guide самыми вероятными оказались вариации потенциальных сайтов связывания транскрипционных факторов, чье участие в развитии рака толстого кишечника было также доказано экспериментально: Sp-1 (см. обзор Allgayer, 2010); MyoD (Arasaradnam *et al*., 2012), NFκB (Andersen *et al*., 2010) YY1 (Chinnappan *et al*., 2009) и CP-1 (синоним NF-Y, см. например, Park *et al*., 2007).

Таблица 3

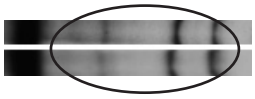
Оценка rSNP_Guide пяти аллелей rs75996864, rs76241113, rs78037487, rs80112297 и rs80313086 гена *APC* человека для ассоциации с раком толстого кишечника

Аллель	Последовательность ДНК, ген *APC*, rs75996864 ± 25 п.о.			
Предок, WT минорный	ggcgcacgtgaccgacatgtggctg**T**attggtgcagcccgccagggtgtca ggcgcacgtgaccgacatgtggctg**G**attggtgcagcccgccagggtgtca			
Клетки Рак	HCT-116 толстого кишечника	HeLaS3 шейки матки	HepG2 гепатома	K562 эритролейкемия
Задержка в геле				
Комплекс	WT больше	MINOR больше	возник слабый MINOR	без изменений
Данные для rSNP_Guide	X_{WT} = 1,0; X_{MINOR} = 0,5	X_{WT} = 1,0; X_{MINOR} = 1,5	X_{WT} = 0,0; X_{MINOR} = 0,5	X_{WT} = 1,0; X_{MINOR} = 1,0
Ранг (значимость)	rSNP_Guide: ранжирование транскрипционных факторов-кандидатов			
I (α < 0,0005)		USF		
II (α < 0,0025)	RAR	RAR		RAR
III (α < 0,005)				USF
IV (α < 0,01)	c-Myb, T3R, YY1	c-Myb		c-Myb, T3R
V (α < 0,025)	ATF, c-Jun, CP-1, CREB, MyoD	c-Jun, CP-1, CREB, MyoD, T3R, YY1		ATF, c-Jun, CP-1, CREB, MyoD, YY1
VI (α < 0,05)	AP-1, c-Fos, CRE-BP1, RXR, Sp-1, USF	AP-1, ATF, c-Fos, CRE-BP1, RXR, Sp-1	ATF, c-Fos, CREB, CRE-BP1, RXR, Sp-1, T3R	AP-1, c-Fos, Sp-1, CRE-BP1, RXR
Аллель	Последовательность ДНК, ген *APC*, rs76241113 ± 25 п.о.			
Предок, WT минорный	caggcttgctgcgggggagggggg**A**aggtggttttccctcgcactgtctt caggcttgctgcgggggagggggg**G**aggtggttttccctcgcactgtctt			
Клетки Рак органа	HCT-116 толстого кишечника	HeLaS3 шейки матки	HepG2 гепатома	K562 эритролейкемия
Задержка в геле				
Комплекс	MINOR больше	MINOR нет	без изменений	WT больше
Данные для rSNP_Guide	X_{WT} = 1,0; X_{MINOR} = 1,5	X_{WT} = 1,0; X_{MINOR} = 0,0	X_{WT} = 1,0; X_{MINOR} = 1,0	X_{WT} = 1,0; X_{MINOR} = 0,5
Ранг (значимость)	rSNP_Guide: ранжирование транскрипционных факторов-кандидатов			
I (α < 0,0005)	Sp-1			
II (α < 0,001)			Sp-1	
III (α < 0,0025)				
IV (α < 0,005)	MyoD		MyoD	MyoD
V (α < 0,01)	c-Myb, NF-kB, Ets		c-Myb, NF-kB, Ets	
VI (α < 0,025)	GAGA, E2F		E2F, GAGA	c-Myb, E2F, Sp-1, GAGA, NF-kB
VII (α < 0,05)	GR, PR, T3R	c-Myb, E2F, PR, GAGA, GR, MyoD, NF-kB, RAR, T3R, YY1	GR, PR, T3R	Ets, GR, NF-E2, PR, T3R, YY1

Продолжение таблицы 3

Аллель	Последовательность ДНК, ген *APC*, rs80112297 ± 25 п.о.			
Предок, WT минорный	ccatggccaggcttgctgcgggggg**A**gggggaaggtggttttccctcgca ccatggccaggcttgctgcgggggg**G**gggggaaggtggttttccctcgca			
Клетки Рак органа	HCT-116 толстого кишечника	HeLaS3 шейки матки	HepG2 гепатома	K562 эритролейкемия
Задержка в геле				
Комплекс	MINOR больше	MINOR нет	возник MINOR	
Данные для rSNP_Guide	X_{WT} = 1,0; X_{MINOR} = 1,5	X_{WT} = 1,0; X_{MINOR} = 0,0	X_{WT} = 0,0; X_{MINOR} = 1,0	
Ранг (значимость)	rSNP_Guide: ранжирование транскрипционных факторов-кандидатов			
I (α < 0,0005)	Sp-1			
II (α < 0,01)	Ets, MyoD, NF-kB		Ets	
III (α < 0,025)	c-Myb, E2F, YY1	c-Myb, Ets, YY1, NF-kB	NF-kB, YY1	
IV (α < 0,05)	GR, NF-E2, PR, T3R	NF-E2, PR, RAR, T3R	E2F, GR, MyoD, NF-E2, PR	
Аллель	Последовательность ДНК, ген *APC*, rs80313086 ± 25 п.о.			
Предок, WT минорный (редкий[$]:	cttgctgcggggggaggggggaagg**T**ggttttccctcgcactgtcttaaac cttgctgcggggggaggggggaagg**C**ggttttccctcgcactgtcttaaac cttgctgcggggggaggggggaagg**G**ggttttccctcgcactgtcttaaac)			
Клетки Рак органа	HCT-116 толстого кишечника	HeLaS3 шейки матки	HepG2 гепатома	K562 эритролейкемия
Задержка в геле				
Комплекс	WT больше	WT меньше	MINOR больше	возник слабый MINOR
Данные для rSNP_Guide	X_{WT} = 1,0; X_{MINOR} = 0,5	X_{WT} = 0,5; X_{MINOR} = 1,0	X_{WT} = 1,0; X_{MINOR} = 1,5	X_{WT} = 0,0; X_{MINOR} = 0,5
Ранг (значимость)	rSNP_Guide: ранжирование транскрипционных факторов-кандидатов			
I (α < 0,0005)			Sp-1	
II (α < 0,0025)				
III (α < 0,005)			c-Myb, GR	
IV (α < 0,01)	GR	NF-kB, GR	NF-kB	
V (α < 0,025)	c-Myb, E2F, Sp-1 MyoD, NF-kB	c-Myb, E2F, Sp-1	E2F	GAGA, NF-E2
VI (α < 0,05)	CEBP, COUP, Ets, GAGA, MEF-2, NF-E2, PR, T3R, YY1	CEBP, COUP, Ets, GAGA, MEF-2, NF-E2, PR, RAR, T3R	CEBP, COUP, GAGA, Ets, MyoD, MEF-2, NF-E2, PR, T3R	RAR, T3R, TCF-1

Окончание таблицы 3

Аллель	Последовательность ДНК, ген *APC*, rs78037487 ± 25 п.о.			
Предок, WT	agggcgctccccattcccgtcggga**G**cccgccgattggctgggtgtgggcg			
минорный	agggcgctccccattcccgtcggga**C**cccgccgattggctgggtgtgggcg			
Клетки	HCT-116	HeLaS3	HepG2	K562
Рак органа	толстого кишечника	шейки матки	гепатома	эритролейкемия
Задержка в геле				
Комплекс	без изменений	MINOR нет	возник слабый MINOR	
Данные для rSNP_Guide	X_{WT} = 1,0; X_{MINOR} = 1,0	X_{WT} = 1,0; X_{MINOR} = 0,0	X_{WT} = 0,0; X_{MINOR} = 0,5	
Ранг (значимость)	rSNP_Guide: ранжирование транскрипционных факторов-кандидатов			
I ($\alpha < 0{,}0025$)	c-Myb, CP-1			
II ($\alpha < 0{,}01$)	Ets, Sp-1			
III ($\alpha < 0{,}025$)	NF-E2	c-Myb, GAGA	T3R	
IV ($\alpha < 0{,}05$)	GAGA, RF-X, T3R	Ets, RF-X	NF-E2	

[$] Редкий аллель не был оценен из-за отсутствия других примеров встречаемости такого типа аллелей.

С целью верификации оценок rSNP_Guide в табл. 1 и 3 приведены результаты для исследуемых аллелей генов *APC* и *MLH1* человека в случае использования раковых клеток линий HCT-116 (рак толстого кишечника), HeLaS3 (рак шейки матки), K562 (эритролейкемия) и HepG2 (гепатома).

Можно видеть, что как для ранее уже ассоциированного SNP rs1800734 (табл. 1), так и для 5 остальных исследованных нами SNPs (табл. 3) наблюдается одна и та же закономерность: для раковых клеточных линий различных органов человека имеются и элементы сходства оценок rSNP_Guide, отражающие общие черты канцерогенеза в целом и существенные различия, соответствующие орган-специфическим особенностям патогенеза.

В заключение представляется важным обсудить также ограничения текущей версии технологии rSNP_Guide (Ponomarenko *et al*., 2001, 2002) и возможные пути усовершенствования технологии в следующих версиях.

Самым дискуссионным вопросом проведенного анализа, несомненно, является слишком большое количество транскрипционных факторов-кандидатов вблизи общепринятого порога статистической значимости ($\alpha < 0{,}05$). Это замечание типично для всех биоинформационных методов распознавания сайтов связывания факторов транскрипции и, в первую очередь, обусловлено высоким регуляторным потенциалом ДНК, лишь часть из которого реализуется в каждой конкретной клеточной ситуации (Kolchanov *et al*., 2007). В этой связи большую актуальность приобретает дополнительная независимая верификация «припороговой» части оценок rSNP_Guide, например, с помощью данных секвенирования футпринтов для транскрипционных факторов в масштабе генома человека в базе данных экспериментов ChIP-Seq с Энциклопедией ДНК-элементов ENCODE (Auerbach *et al*., 2013).

Очевидным недостатком текущей версии rSNP_Guide (Ponomarenko *et al*., 2001, 2002) является также ее ограниченность 40 регуляторными белками (табл. 2), тогда как, например, система SITECON (Oshchepkov *et al*., 2004) включает более 100 транскрипционных факторов. Следовательно, требуется обновление rSNP_Guide за счет существенного расширения количества анализируемых белков.

Наконец, среди транскрипционных факторов-кандидатов, сайты связывания которых изменяются в случае SNPs rs75996864, rs76241113, rs78037487, rs80112297, rs80313086 гена *APC* человека был предсказан фактор T3R, участие которого в развитии рака толстого кишечника до сих пор не было описано. Поскольку в более чем 30 экспериментальных статьях описано участие этого белка в развитии других типов рака у человека, например, при патогенезе рака молочной железы (Alvarado-Pisani *et al.*, 1986), представляется целесообразным экспериментально проверить влияние изученных SNPs в гене *APC* на связывание T3R.

Представленные нами результаты свидетельствуют о получении по технологии rSNP_Guide (Ponomarenko *et al.*, 2001, 2002) новых оснований для дальнейшего исследования ассоциации SNPs rs75996864, rs76241113, rs78037487, rs80112297, rs80313086 гена *APC* с раком толстого кишечника общепринятыми медико-генетическими методами.

Работа была поддержана Госконтрактом № 14.512.11.0094 и Соглашением № 8740 Минобрнауки РФ, а также частично – молодежным проектом поддержки ведущих научных школ НШ-5278.2012.4, программами Президиума РАН «Молекулярная и клеточная биология» № 6.6 и «Фундаментальные науки – медицине» № 23.

ЛИТЕРАТУРА

Антонцева Е.В., Брызгалов Л.О., Матвеева М.Ю. и др. Поиск регуляторных SNPs, связанных с развитием рака толстой кишки, в генах *APC* и *MLH1* // Вавилов. журн. генет. и селекции. 2011. Т. 15. Вып. 4. С. 644–652.

Allgayer H. Pdcd4, a colon cancer prognostic that is regulated by a microRNA // Crit. Rev. Oncol. Hematol. 2010. V. 73. No. 3. P. 185–191.

Alvarado-Pisani A.R., Chacon R.S., Betancourt L.J., Lopez-Herrera L. Thyroid hormone receptors in human breast cancer: effect of thyroxine administration // Anticancer Res. 1986. V. 6. No. 6. P. 1347–1351.

Andersen V., Christensen J., Overvad K. *et al*. Polymorphisms in NFkB, PXR, LXR and risk of colorectal cancer in a prospective study of Danes // BMC Cancer. 2010. V. 10. P. 484.

Ansorge N., Juttner S., Cramer T. *et al*. An upstream CRE-E-box element is essential for gastrin-dependent activation of the cyclooxygenase-2 gene in human colon cancer cells // Regul. Pep. 2007. V. 144. No. 1/3. P. 25–33.

Antontseva E.V., Bryzgalov L.O., Matveeva M.Yu. *et al*. Search for regulatory SNPs associated with colon cancer in the *APC* and *MLH1* genes // Russ. J. Genet: Appl. Res. 2012. V. 2. No. 3. P. 222–228.

Arasaradnam R.P., Quraishi M.N., Commane D. *et al*. MYOD-1 in normal colonic mucosa–role as a putative biomarker? // BMC Res. Notes. 2012. V. 5. P. 240.

Auerbach R.K., Chen B., Butte A.J. Relating genes to function: identifying enriched transcription factors using the ENCODE ChIP-Seq significance tool // Bioinformatics. 2013. V. 29. No. 15. P. 1922–1924.

Belanger A.S., Tojcic J., Harvey M., Guillemette C. Regulation of UGT1A1 and HNF1 transcription factor gene expression by DNA methylation in colon cancer cells // BMC Mol. Biol. 2010. V. 11. P. 9.

Berg O.G., von Hippel P.H. Selection of DNA binding sites by regulatory proteins, Statistical-mechanical theory and application to operators and promoters // J. Mol. Biol. 1987. V. 193. No. 4. P. 723–750.

Bruno M.E., West R.B., Schneeman T.A. *et al*. Upstream stimulatory factor but not c-Myc enhances transcription of the human polymeric immunoglobulin receptor gene // Mol. Immunol. 2004. V. 40. No. 10. P. 695–708.

Byrne A.M., Foran E., Sharma R. *et al*. Bile acids modulate the Golgi membrane fission process via a protein kinase Ceta and protein kinase D-dependent pathway in colonic epithelial cells // Carcinogenesis. 2010. V. 31. No. 4. P. 737–744.

Chinnappan D., Xiao D., Ratnasari A. *et al*. Transcription factor YY1 expression in human gastrointestinal cancer cells // Int. J. Oncol. 2009. V. 34. No. 5. P. 1417–1423.

Christensen L.L., Tobiasen H., Holm A. *et al*. MiRNA-362-3p induces cell cycle arrest through targeting of E2F1, USF2 and PTPN1 and is associated with recurrence of colorectal cancer // Int. J. Cancer. 2013. V. 133. No. 1. P. 67–78.

Gerstenblith M.R., Shi J., Landi M.T. Genome-wide association studies of pigmentation and skin cancer: a review and meta-analysis // Pigment Cell Melanoma Res. 2010. V. 23. No. 5. P. 587–606.

Groubet R., Pallet V., Delage B. *et al*. Hyperlipidic diets induce early alterations of the vitamin A signalling pathway in rat colonic mucosa // Endocr. Regul. 2003. V. 37. No. 3. P. 137–144.

Hamosh A., Scott A.F., Amberger J.S. *et al*. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders // Nucl. Acids Res. 2005. V. 33. P. D514–D517.

Hindorff L.A., Sethupathy P., Junkins H.A. *et al*. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits // Proc. Natl Acad. Sci. USA. 2009. V. 106. No. 23. P. 9362–9367.

Hitchins M.P., Wong J.J.L., Suthers G. *et al*. Inheritance of a cancer-associated MLH1 germ-line epimutation // New Eng. J. Med. 2007. V. 356. No. 7. P. 697–705.

Kameue C., Tsukahara T., Ushida K. Alteration of gene expression in the colon of colorectal cancer model rat by dietary sodium gluconate // Biosci. Biotechnol. Biochem. 2006. V. 70. No. 3. P. 606–614.

Kolchanov N.A., Merkulova T.I., Ignatieva E.V. *et al*. Combined experimental and computational approaches to study the regulatory elements in eukaryotic genes // Brief Bioinform. 2007. V. 8. No. 4. P. 266–274.

Mulholland D.J., Dedhar S., Coetzee G.A., Nelson C.C. Interaction of nuclear receptors with the Wnt/beta-catenin/Tcf signaling axis: Want you like to know? // Endocrinol. Rev. 2005. V. 26. No. 7. P. 898–915.

NCBI Resource Coordinators, Database resources of the National Center for Biotechnology Information // Nucl. Acids Res. 2013. V. 41. P. D8–D20.

Oshchepkov D.Y., Vityaev E.E., Grigorovich D.A. *et al*. SITECON: a tool for detecting conservative conformational and physicochemical properties in transcription factor binding site alignments and for site recognition // Nucl. Acids Res. 2004. V. 32. Web Server issue. P. W208–W212.

Pare L., Marcuello E., Altes A. *et al*. Transcription factor-binding sites in the thymidylate synthase gene: predictors of outcome in patients with metastatic colorectal cancer treated with 5-fluorouracil and oxaliplatin? // Pharmacogenomics J. 2008. V. 8. No. 5. P. 315–320.

Park S.H., Yu G.R., Kim W.H. et al. NF-Y-dependent cyclin B2 expression in colorectal adenocarcinoma // Clin. Cancer Res. 2007. V. 13. No. 3. P. 858–867.

Polakis P. An Introduction to Wnt Signaling // Targeting the Wnt Pathway in Cancer. N.Y.: Springer, 2011. P. 1–18.

Ponomarenko J.V., Merkulova T.I., Vasiliev G.V. *et al*. rSNP_Guide, a database system for analysis of transcription factor binding to target sequences: application to SNPs and site-directed mutations // Nucl. Acids Res. 2001. V. 29. No. 1. P. 312–316.

Ponomarenko J.V., Orlova G.V., Merkulova T.I. *et al*. rSNP_Guide: an integrated database-tools system for studying SNPs and site-directed mutations in transcription factor binding sites // Hum. Mutat. 2002. V. 20. No. 4. P. 239–248.

Ponomarenko M.P., Ponomarenko J.V., Frolov A.S. *et al*. Oligonucleotide frequency matrices addressed to recognizing functional DNA sites // Bioinformatics. 1999. V. 15. P. 631–643.

Ramsay R.G., Ciznadija D., Vanevski M., Mantamadiotis T. Transcriptional regulation of cyclo-oxygenase expression: three pillars of control // Int. J. Immunopathol. Pharmacol. 2003. V. 16. No. 2 (Suppl). P. 59–67.

Sanchez-Ruiz J.M. Protein kinetic stability // Biophys. Chem. 2010. V. 148. P. 1–15.

The International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome // Nature. 2004. V. 431. No. 7011. P. 931–945.

Wai P.Y., Mi Z., Gao C. *et al*. Ets-1 and runx2 regulate transcription of a metastatic gene, osteopontin, in murine colorectal cancer cells // J. Biol. Chem. 2006. V. 281. No. 28. P. 18973–18982.

Wei H.B., Han X.Y., Fan W. *et al*. Effect of retinoic acid on cell proliferation kinetics and retinoic acid receptor expression of colorectal mucosa // World J. Gastroenterol. 2003. V. 9. No. 8. P. 1725–1728.

Win A.K., Hopper J.L., Buchanan D.D. *et al*. Are the common genetic variants associated with colorectal cancer risk for DNA mismatch repair gene mutation carriers? // Eur. J. Cancer. 2013. V. 49. No. 7. P. 1578–1587.

Xu X.L., Yu J., Zhang H.Y. *et al*. Methylation profile of the promoter CpG islands of 31 genes that may contribute to colorectal carcinogenesis // World J. Gastroenterol. 2004. V. 10. No. 23. P. 3441–3454.

rSNP_Guide-BASED EVALUATION OF SNPs IN PROMOTERS OF THE HUMAN *APC* AND *MLH1* GENES ASSOCIATED WITH COLON CANCER

D.A. Rasskazov[1], E.V. Antontseva[1], L.O. Bryzgalov[1], M.Yu. Matveeva[1], E.V. Kashina[1], P.M. Ponomarenko[1], G.V. Orlova[1], M.P. Ponomarenko[1], D.A. Afonnikov[1, 2], T.I. Merkulova[1, 2]

[1] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia, e-mail: pon@bionet.nsc.ru;
[2] Novosibirsk National Research State University, Novosibirsk, Russia

Summary

Six regulatory SNPs (single nucleotide polymorphisms) of two human *APC* and *MLH1* genes (rs75996864, rs76241113, rs78037487, rs80112297, rs80313086 and rs1800734) were evaluated by the previously developed rSNP_Guide method to compute the significance of the changes for binding of the SNP region with 40 transcription factors. For each SNP, all analyzed proteins were ranged according to the significance of the changes for protein binding to the alleles evaluated by Student's *t*-test. We found that rs75996864, rs76241113, rs78037487, rs80112297, and rs80313086 of *APC*, as well as rs1800734 of *MLH1* in humans concerned to the greatest extent the binding of transcription factors NF-Y, NFkB, c-Myb, RAR, YY-1, and Sp-1, which are known to be involved in colon cancer development. Our results substantiate investigation of associations of rs75996864, rs76241113, rs78037487, rs80112297, and rs80313086 in the *APC* gene with colon cancer by using commonly accepted medical and genetic protocols.

Key words: single-nucleotide polymorphism, gene expression regulation, colon cancer, *APC*, *MLH1* genes, DNA–regulatory protein complex, Student's *t*-test.

УДК 577.133.3:57.087.1

SNP_TATA_COMPARATOR: WEB-СЕРВИС ПРИМЕНЕНИЯ УРАВНЕНИЯ РАВНОВЕСИЯ TBP/TATA-КОМПЛЕКСА В СРАВНИТЕЛЬНОЙ ОЦЕНКЕ SNPs ПРОМОТОРОВ ГЕНОВ, СВЯЗАННЫХ С БОЛЕЗНЯМИ ЧЕЛОВЕКА

© 2013 г. **Д.А. Рассказов[1], К.В. Гунбин[1], П.М. Пономаренко[1], О.В. Вишневский[1, 2], М.П. Пономаренко[1], Д.А. Афонников[1, 2]**

[1] Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия, e-mail: pon@bionet.nsc.ru;
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

Исследование проявления полиморфизма регуляторных районов генов на уровне их экспрессии имеет важное фундаментальное и прикладное значение. Создан Web-сервис SNP_TATA_Comparator для применения оценки *in silico* на основе уравнения равновесия TBP/TATA-комплекса (TBP, TATA-связывающий белок), экспериментально доказанного для биохимических проявлений *in vitro*, к связанным с болезнями SNP (Single nucleotide polymorphism) кор-промоторов генов человека.В результате обеспечен свободный доступ в режиме «реального времени» к анализу имеющейся персонифицированной информации об отклонениях индивидуальных геномов от так называемого референсного генома человека (Вып. 68, Flicek *et al*., 2011), т. е. варианта генома, который общепринят в качестве стандарта для сравнительного анализа). При этом автоматически учитываются все доказанные транскрипты гена (база данных GENCODE, Вып. 17). Web-сервис SNP_TATA_Comparator предназначен для диагностики, мониторинга, профилактики и лечения заболеваний с учетом индивидуальных геномов пациентов в рамках персонифицированной медицины.

Ключевые слова: SNP, TATA-бокс, TBP, экспрессия генов, болезни человека.

ВВЕДЕНИЕ

Голдберг и Хогнесс (Lifton *et al*., 1978) открыли (A+T)-богатый район длиной 8 п.о. в промоторах генов гистонов дрозофилы, который назвали «TATA бокс» по виду его консенсуса TATA(t/a)A(t/a)g (синонимы: Goldberg-Hogness box и Hogness box). Затем на TATA-боксе промотора гена кональбумина был открыт (Davison *et al*., 1983) стабильный комплекс ДНК/белок, формирование которого предшествовало связыванию РНК полимеразы II (RNAPII). Год спустя из всего ДНК/белкового комплекса-мишени для связывания RNAPII был выделен (Parker, Topol, 1984) инициирующий транскрипционный фактор TFIID, который связывал TATA-бокс промотора гена *HSP70* дрозофилы. Было замечено (Fire *et al*., 1984), что после связывания RNAPII с анкерным комплексом вокруг TATA-бокса на этой основе формируется иной стабильный ДНК/белковый-комплекс, названный преинициаторным из-за начала транскрипции только после его самосборки. Позднее (Schmidt *et al*., 1989) у дрожжей нашли ген TATA-связывающего полипептида ДНК-связывающей субъединицы TFIID (TATA-связывающий белок, TBP). Параллельно для TATA бокса рентгеноструктурным анализом были установлены трехмерные структуры свободной B-спирали ДНК (Drew *et al*., 1981) и комплекса TATA-бокса с

TBP (Kim J. *et al*., 1993; Kim Y. *et al*., 1993). Наконец, были измерены (Hahn *et al*., 1989) уровни 10^{-9} М специфического сродства TBP к TATA-боксу и 10^{-5} М неспецифического – к произвольной ДНК.

В настоящее время TATA-бокс – один из самых изученных регуляторных сигналов геномов эукариот (Ponomarenko *et al*., 2013). Хотя опыт с антителами против TBP (Wieczorek *et al*., 1998) продемонстрировал возможность инициации транскрипции *in vitro* без TBP, все еще не было найдено гена эукариот с такой инициацией транскрипции *in vivo*. Поэтому TBP/TATA-комплекс общепринято считать обязательным для преинициаторного комплекса RNAPII якорем на ДНК (Auble, 2009). С помощью микрочипов была построена полногеномная карта экспериментально доказанных TATA-боксов для 17181 генов человека (Yang *et al*., 2011). Описано более 50 связанных с заболеваниями SNPs (Single nucleotide polymorphisms) TATA-боксов промоторов генов человека (Савинкова и др., 2009). В этой связи исследование проявления полиморфизма регуляторных районов генов на уровне их экспрессии имеет важное фундаментальное и прикладное значение. С этой целью был создан Web-сервис SNP_TATA_Comparator для сравнения TATA-боксов индивидуальных геномов пациентов с референсным геномом человека – URL=http://beehive.bionet.nsc.ru/cgi-bin/mgs/tatascan/start.pl.

МАТЕРИАЛЫ И МЕТОДЫ

Web-сервис SNP_TATA_Comparator анализирует последовательность ДНК $\{s_{-140}\ldots s_{-1}\}$ нуклеотидов $s \in \{a, t, g, c\}$ длиной 140 п.о. непосредственно перед стартом транскрипции генов человека. Он автоматически берет ее из базы данных Ensembl (Flicek *et al*., 2011) референсного генома человека, т. е. варианта генома, общепринятого в качестве стандарта для сравнительного анализа с использованием разметки генов из базы данных GENCODE (Harrow *et al*., 2012), как показано на рисунке на примере связанного с β-талассемией человека SNP A-31g гена *HBB* β-гемоглобина.

Web-сервис SNP_TATA_Comparator был написан на языке программирования Perl. Он доступен по URL=http://beehive.bionet.nsc.ru/cgi-bin/mgs/tatascan/start.pl. Сначала пользователь получает пустую форму запроса «выбор гена» (рис., слева). После ввода пользователем (непрерывная стрелка) кода или ID интересующего его гена SNP_TATA_Comparator по команде «Search» находит (штриховые стрелки) в базе данных Ensembl (Flicek *et al*., 2011) все документы об этом гене и предлагает пользователю их список для выбора одного из них с целью дальнейшего анализа. Затем (точечные стрелки) по команде «Search» Web-сервис находит в базе данных GENCODE (Harrow *et al*., 2012) все старты транскрипции этого гена и дает пользователю их список для выбора одного из них для анализа его кор-промотора. Наконец, по команде «GetSeq» он находит (пунктирные стрелки) в базе данных Ensembl (Flicek *et al*., 2011) фрагмент длиной 140 п.о., локализованный в референсном геноме человека (Вып. 68) непосредственно перед выбранным стартом транскрипции.

Итак, пользователь автоматически получает две копии интересующего его фрагмента референсного генома человека: в качестве стандарта для предстоящего сравнения (окно «Base sequence») и (окно «Editable sequence») для внесения в копию этого стандарта интересующих пользователя индивидуальных отличий.

По команде «Calculate» с помощью формул (1–4) SNP_TATA_Comparator сравнивает (стрелки «штрих-точка-точка») введенный пользователем частный вариант генома с референсным геномом человека (Вып. 68, Flicek *et al*., 2011). Окно «Result» содержит результат работы Web-сервиса SNP_TATA_Comparator для интересующего пользователя SNP TATA-бокса человека. Этот результат интерпретируется согласно приведенному ниже описанию указанных выше формул (1–4).

Прежде всего, Web-сервис SNP_TATA_Comparator анализирует единообразно независимо один от другого оба варианта генома человека, «референсный» и «индивидуальный». Каждое положение скользящего окна длиной 26 п.о. $\{s_{m-13}\ldots s_m\ldots s_{m+12}\}$ в m-й позиции характеризуется оценкой равновесной константы диссоциации K_D комплекса между ДНК и TATA-связывающим белком (TBP), выраженной в натуральных логарифмических единицах (ln) так называемой «аффинности»:

$$
\begin{cases}
-\ln[K_D](\{\zeta_i\}) = 10{,}9 - 0{,}2\left(\ln[K_{D;dsDNA}](\{\zeta_i\}) + PWM(\{\zeta_i\}) - \ln[K_{D;ssDNA}](\{\zeta_i\})\right) \\
-\ln[K_{D;dsDNA}](\{\zeta_i\}) = \underset{1\le k\le 11}{\text{mean}}\left[35{,}1 + 3{,}4\sum_{k+5\le j\le k+8}\omega(\zeta_j\zeta_{j+1}) - 0{,}8\sum_{\substack{k\le j\le k+13\\ \zeta_j\zeta_{j+1}=TA}} F_{TA}(j)\right]; \\
PWM(\{\zeta_i\}) = \underset{1\le k\le 11}{\max}\left[\sum_{k+5\le j\le k+14} \text{weight}_{\zeta_j j}\right]; \\
-\ln[K_{D;TBP/ssDNA}](\{\zeta_i\}) = \underset{\substack{k=k_{TATA}\\ \gamma\in\{\zeta;-\zeta\}}}{\text{mean}}\left[14{,}5 + 0{,}9\sum_{\substack{k\le j\le k+13\\ \gamma_j\gamma_{j+1}\in WR}} F_{WR}(j) + 2{,}5\sum_{\substack{k\le j\le k+13\\ \gamma_j\gamma_{j+1}\in TV}} F_{TV}(j)\right];
\end{cases}
\tag{1}
$$

здесь: 10,9 ln – величина неспецифического сродства TBP/ДНК ($\approx 10^{-5}$ M), взятая из эксперимента (Hahn *et al*., 1989); 0,2 – коэффициент, взятый из работы Пономаренко с соавт. (2008) и численно равный нормированному на число этапов отношению длин TATA-бокса и «скользящего окна» (здесь: (15 п.о./26 п.о.)/3 = 0,2, стехиометрический коэффициент каждого из трех последовательных шагов образования TBP/ДНК-комплекса); $\text{weight}_{k\zeta}$ и k_{TATA} – матрица

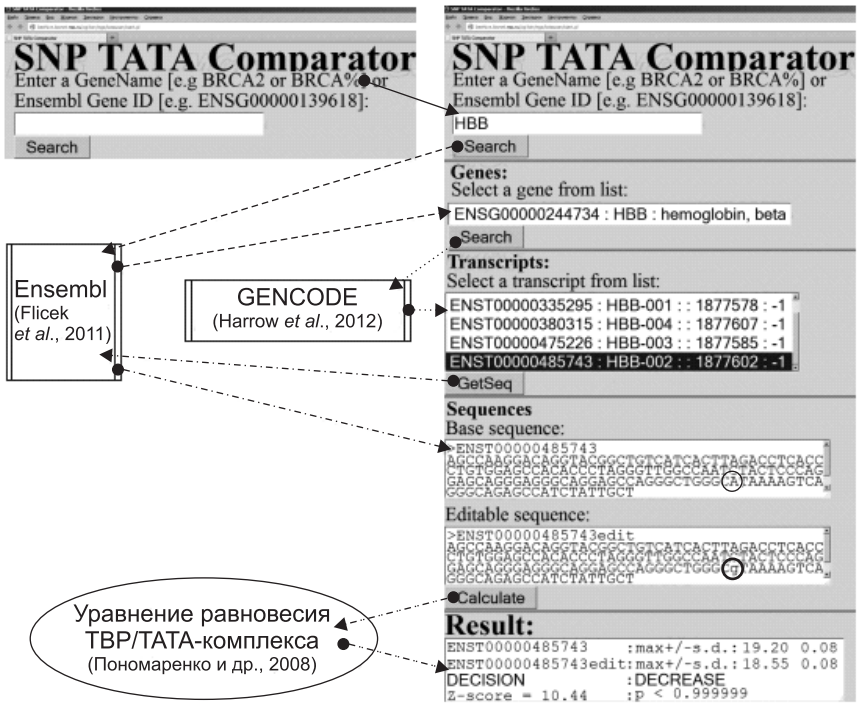


Рис. Применение Web-сервиса SNP_TATA_Comparator на иллюстративном примере сравнительной оценки для связанной с β-талассемией SNP A-31g (○) промотора гена *HBB*, кодирующего β-цепь гемоглобина человека.

Вверху слева – результат ввода в Интернет адреса http://beehive.bionet.nsc.ru/cgi-bin/mgs/tatascan/start.pl исходно пустой формы «входных данных» Web-сервиса SNP_TATA_Comparator. После ввода в нее (стрелка) информации о целевом гене SNP_TATA_Comparator по команде «Search» находит (штриховые стрелки) в базе данных Ensembl (Flicek *et al*., 2011) все ее документы и предлагает пользователю их список для выбора одного из них для анализа. Затем (точечные стрелки) по команде «Search» SNP_TATA_Comparator находит в базе данных GENCODE (Harrow *et al*., 2012) все старты транскрипции этого гена и предлагает пользователю их список для выбора одного из них для анализа его кор-промотора. Далее по команде «GetSeq» он берет (пунктирные стрелки) из базы данных Ensembl (Flicek *et al*., 2011) фрагмент длиной 140 п.о. референсного генома человека (вып. 68) и предлагает пользователю две его копии: в окне «Base sequence» в качестве стандарта предстоящего сравнения и в окне «Editable sequence» для внесения пользователем в эту копию анализируемых им отличий (○) от указанного стандарта. Наконец, SNP_TATA_Comparator по команде «Calculate» (стрелки «штрих-точка-точка») сравнивает введенный пользователем вариант геномной ДНК с референсным геномом человека с помощью формул (1–4) и представляет пользователю результат этого сравнительного анализа, интерпретируемого согласно разделу «Материалы и методы».

Бухера (Bucher, 1990), вес нуклеотида ζ в k-й позиции TATA-box ($-1 \le k \le 13$ относительно канонического варианта $T_0A_1T_2A_3A_4A_5G_6$ и позиция с максимумом PWM-скора этих весов в «окне сканирования»; $F_{TA}(i)$, $F_{WR}(i)$, и $F_{TV}(i)$ – веса динуклеотидов TA, WR, и TV в i-х позициях TATA-бокса, взятые из работ (Пономаренко и др., 1997; Ponomarenko *et al*., 1999); $\omega(\zeta_i\zeta_{i+1})$, ширина малой бороздки в ангстремах (Karas *et al*., 1996), взятая из базы данных (Колчанов и др., 1998); «–ζ» означает «комплементарный к ζ».

Формула (1) описывает следующие шаги формирования TBP/TATA-комплекса: (i) TBP скользит вдоль ДНК (Coleman, Pugh, 1995) в силу их неспецифического сродства (Hahn *et al*., 1989) → (ii) остановка скольжения на TATA-боксе (Berg, von Hippel, 1987; Bucher, 1990) → (iii) эндотермическая (Powell *et al*., 2002) стабилизация TBP/ДНК-комплекса увеличением изгиба оси двойной спирали ДНК от 19° (Drew *et al*., 1981) до 90° (Kim J. *et al*., 1993; Kim Y. *et al*., 1993).

Web-сервис SNP_TATA_Comparator находит максимальные оценки K_D^{REF} и K_D^{USER} для участка [–70; –20] локализации всех доказанных к настоящему времени TATA-боксов промоторов эукариот относительно старта транскрипции по отдельности для соответственно референсного генома человека (Вып. 68) и «введенного» пользователем, как это показано на рис. На этом рисунке можно видеть, что K_D^{REF} и K_D^{USER} имеют оценки «± s.d.» стандартного отклонения (± σ):

$$\pm\sigma_{\#} = \sqrt{\sum_{j=-13}^{12}\sum_{\xi=1}^{3}\ln\left[\frac{K^{\#}_{D;s_j\to\xi}}{K^{\#}_{D}}\right]\Big/(78\times 77)}\ ; \qquad (2)$$

здесь: $\#\in\{\text{REF, USER}\}$; $s_j\to\xi$, замена нуклеотида s_j заданной ДНК на другой ξ.

Формула (2) оценивает величину стандартного отклонения K_D^{REF} и K_D^{USER} с помощью всех $78 = 26 \times 3$ возможных одиночных замен каждого нуклеотида на три других варианта в каждой из 26 позиций «окна сканирования» при том его положении, когда были получены эти максимальные оценки формулы (1). Эти две независимые оценки, σ_{REF} и σ_{USER}, стандартным способом объединяются:

$$\pm\sigma = \sqrt{\sigma^2_{REF} + \sigma^2_{REF.}} \qquad (3)$$

На этой основе оценивается 95 %-й доверительный интервал, $\pm\Delta_{95\,\%} = \tau_{95\,\%;\upsilon=78}\sigma$, для t-критерия Стьюдента недостоверных различий двух сравниваемых вариантов генома, на основе которого SNP_TATA_Comparator принимает решение:

$$\begin{cases} \ln\left[\dfrac{K_D^{REF}}{K_D^{USER}}\right] \ge \Delta_{95\,\%} \xrightarrow{\text{yields}} \text{достоверный ИЗБЫТОК продукта гена;} \\ \ln\left[\dfrac{K_D^{REF}}{K_D^{USER}}\right] \le -\Delta_{95\,\%} \xrightarrow{\text{yields}} \text{достоверный ДЕФИЦИТ продуктов гена;} \\ \text{ИНАЧЕ} \xrightarrow{\text{yields}} \text{ПРОГНОЗ недостоверный.} \end{cases} \qquad (4)$$

Результат формулы (4) показан в строке «DECISION» окна «Results» на рис.

Наконец, в качестве дополнительной верификации результата формулы (4) Web-сервис SNP_TATA_Comparator оценивает часто используемую Z-статистику (отношение абсолютной разницы оценок к их стандартному отклонению $|K_D^{REF} - K_D^{USER}| / \sigma$) и уровень ее значимости *p*. Величины Z и p показаны в нижней строке окна «Results» на рис. и в табл.

РЕЗУЛЬТАТЫ

Результаты работы Web-сервиса SNP_TATA_Comparator по сравнительной оценке связанных с болезнями человека SNPs промоторов генов представлены на рис. и в табл.

На рис. детально шаг за шагом показан алгоритм применения пользователем Web-сервиса SNP_TATA_Comparator на иллюстративном примере A-31g, связанного с β-талассемией SNP промотора гена *HBB*, кодирующего β-цепь гемоглобина (Takihara *et al*., 1986) – одного из самых изученных наследственных заболеваний.

Как можно видеть, согласно решению SNP_TATA_Comparator с помощью формул (1–4), биохимической причиной β-талассемии был дефицит β-цепей гемоглобина, что соответствует данным эксперимента (Takihara *et al*., 1986). С помощью Web-сервиса SNP_TATA_Comparator

Таблица

Результаты SNP_TATA_Comparator по сравнительной оценке SNPs промоторов генов, связанных с болезнями человека

Геном человека					Результаты SNP_TATA_Comparator					Экспериментально подтверждено (ссылка)
Ген	мРНК	SNP	TATA,REF→USER	Заболевание	$K_D^{REF} \pm \sigma_{REF}$	$K_D^{USER} \pm \sigma_{USER}$	Решение	*Z*	*p*	
HBB	№ 2	A-31g	gC(A→g)TAAAAg	β-талассемия	19,20 ± 0,08	18,55 ± 0,08	дефицит	10,44	10^{-6}	(Takihara *et al*., 1986)
CYP2A6	№ 1	T-48g	agTA(T→g)AAAgg	рак легких	20,02 ± 0,10	18,52 ± 0,09	дефицит	20,70	10^{-6}	(Pitarque *et al*., 2001)
SOD1	№ 4	A-27g	ccT(A→g)TAAAgt	боковой амиотрофический склероз	19,83 ± 0,09	18,78 ± 0,08	дефицит	16,67	10^{-6}	(Niemann *et al*., 2007)
EDH17B2	№ 3(1)	A-27c	TGATATC(A→c)A	рак молочной железы	18,14 ± 0,08	17,85 ± 0,09	дефицит	4,52	10^{-3}	(Peltoketo *et al*., 1994)
DARC	№ 3	T-33c	tcTTA(T→c)CTTgg	анемия, но устойчивость к малярии	19,20 ± 0,08	18,55 ± 0,08	дефицит	10,44	10^{-6}	(Penner, Davie, 1994) (Tournamillt *et al*., 1995)
MBL2	№ 1	T-35c	tcTA(T→c)ATAgcc	риск инсульта ниже, инфекций – выше	20,17 ± 0,11	19,28 ± 0,09	дефицит	11,96	10^{-7}	(Boldt *et al*., 2006) (Sziller *et al*., 2007) (Cervera *et al*., 2010)
NOS2	№ 1	t-21c	TATAAATAc(t→c)t	рассеянный склероз, но устойчивость к малярии	20,17 ± 0,10	20,38 ± 0,10	избыток	2,90	10^{-2}	(Охотин и др., 2002) (Clark *et al*., 2003)
IL1b	№ 4	C-31t	gc(C→t)ATAAAA	рак легких, рак печени, язва желудка, гастрит, депрессии	19,21 ± 0,08	20,15 ± 0,09	избыток	14,55	10^{-6}	(Wang *et al*., 2003) (Wu *et al*., 2010) (Martinez-Carrillo *et al*., 2010) (Borkowska *et al*., 2011)
TAF5L	№ 2	C-25t	cc(C→t)AGCTGAg	диабет I типа	15,73 ± 0,07	16,83 ± 0,10	избыток	17,82	10^{-8}	(Chistiakov *et al*., 2005)
F3	№ 3	C-21t	cTTTATAg(c→t)gc	инфаркт миокарда	19,59 ± 0,10	19,92 ± 0,10	избыток	4,57	10^{-3}	(Arnaud *et al*., 2000)

были получены аналогичные результаты для генов *SOD1*, *CYP2A6* и *EDH17B2*, дефицит белковых продуктов которых является биохимической причиной развития бокового амиотрофического склероза, рака легких и рака молочной железы соответственно.

В нижней части табл. показаны результаты работы SNP_TATA_Comparator в случае диаметрально противоположного биохимического проявления связанных с болезнями человека SNPs промоторов генов *IL1b*, *TAF5L* и *F3*, избыток продуктов которых вызвал патологии. Однако решения, полученные SNP_TATA_Comparator с помощью формул (1–4), оказались вновь в согласии с данными экспериментов.

Наконец, в центральной части табл. даны примеры нетривиальных «слабоповреждающих/слабоулучшающих» SNP промоторов генов *DARC*, *MBL2* и *NOS2* человека, вследствие многофункциональности продуктов которых оба, и избыток, и дефицит, оказываются «двуликими», препятствуя одним заболеваниям и способствуя другим. Тем не менее решения (формулы 1–4) предлагаемого Web-сервиса SNP_TATA_Comparator снова оказались в согласии с экспериментами.

ОБСУЖДЕНИЕ

Предсказанный теоретически (формула 1) факт связывания TBP с TATA-боксом за три последовательных шага (Пономаренко и др., 2008) был через год независимо установлен в эксперименте Delgadillo с соавт. (2009).

Ранее (Пономаренко и др., 2010) были установлены достоверные корреляции между оценками *in silico* формулы (1) и всеми опубликованными данными 68 экспериментов по измерению влияния локального окружения TATA-боксов на сродство к ним TBP в условиях 16 типов клеток из 19 видов эукариот, включая простейшие, дрожжи, растения, животных, а также их вирусы. Уравнение (1) было верифицировано экспериментально как для представительной выборки связанных с заболеваниями человека SNPs (Savinkova *et al*., 2013), так и отдельно для самых сильных повреждений TATA-боксов промоторов генов человека (Drachkova *et al*., 2011).

Все это вместе взятое обосновывает возможность рассмотрения оценок формулы (1) для предсказания генетически обусловленных изменений экспрессии генов человека. Эти предсказания могут быть приняты во внимание при анализе индивидуальных отклонений генома пациента от референсного генома человека при диагностике, мониторинге, профилактике и лечении для учета генетической предрасположенности пациента к заболеваниям и чувствительности/устойчивости к определенным лекарственным препаратам и терапевтическим воздействиям в рамках развития персонифицированной медицины. Именно эта прикладная задача была решена в настоящей работе путем создания Web-сервиса SNP_TATA_Comparator.

БЛАГОДАРНОСТИ

Работа была поддержана Госконтрактом № 14.512.11.0094 Минобрнауки РФ и Соглашением № 8740 Минобрнауки РФ; Проектом поддержки ведущих научных школ НШ-5278.2012.4; Интеграционным проектом СО РАН № 136 и Программой Президиума РАН «Молекулярная и клеточная биология» (проект 6.6).

ЛИТЕРАТУРА

Колчанов Н.А., Пономаренко М.П., Пономаренко Ю.В. и др. Функциональные сайты геномов про- и эукариот: компьютерное моделирование и предсказание активности // Молекуляр. биология. 1998. Т. 32. Вып. 2. С. 255–267.

Охотин В.Е., Калиниченко С.Г., Дудина Ю.В. NO-ергическая трансмиссия и NO как объемный нейропередатчик. Влияние NO на механизмы синаптической пластичности и эпилептогенез // Усп. физиол. наук. 2002. Т. 33. № 2. С. 41–55.

Пономаренко М.П., Савинкова Л.К., Пономаренко Ю.В. и др. Моделирование последовательностей TATA-боксов генов эукариот // Молекуляр. биология. 1997. Т. 31. Вып. 4. С. 726–732.

Пономаренко П.М., Савинкова Л.К., Драчкова И.А. и др. Пошаговая модель связывания TBP/TATA-бокс позволяет предсказать наследственное заболевание человека по точечному полиморфизму // Докл. АН. 2008. Т. 419. Вып. 6. С. 828–832.

Пономаренко П.М., Суслов В.В., Савинкова Л.К. и др. Точное уравнение равновесия четырех шагов связывания TBP с TATA-боксом для прогноза фенотипического проявления мутаций // Биофизика. 2010. Т. 55. Вып. 3. С. 400–414.

Савинкова Л.К., Пономаренко М.П., Пономаренко П.М. и др. Полиморфизмы TATA-боксов промоторов генов человека и ассоциированные с ними наследственные патологии // Биохимия. 2009. Т. 74. Вып. 2. С. 149–163.

Arnaud E., Barbalat V., Nicaud V. *et al*. Polymorphisms in the 5′ regulatory region of the tissue factor gene and the risk of myocardial infarction and venous thromboembolism: the ECTIM and PATHROS studies. Etude Cas-Temoins de l'Infarctus du Myocarde. Paris Thrombosis case-control Study // Arterioscler. Thromb. Vasc. Biol. 2000. V. 20. No. 3. P. 892–898.

Auble D.T. The dynamic personality of TATA-binding protein // Trends Biochem. Sci. 2009. V. 34. No. 2. P. 49–52.

Berg O.G., von Hippel P.H. Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters // J. Mol. Biol. 1987. V. 193. No. 4. P. 723–750.

Boldt A.B., Culpi L., Tsuneto L.T. *et al*. Diversity of the MBL2 gene in various Brazilian populations and the case of selection at the mannose-binding lectin locus // Hum. Immunol. 2006. V. 67. No. 9. P. 722–734.

Borkowska P., Kucia K., Rzezniczek S. *et al*. Interleukin-1β promoter (-31T/C and -511C/T) polymorphisms in major recurrent depression // J. Mol. Neurosci. 2011. V. 44. No. 1. P. 12–16.

Bucher P. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences // J. Mol. Biol. 1990. V. 212. No. 4. P. 563–578.

Cervera A., Planas A.M., Justicia C. *et al*. Genetically-defined deficiency of mannose-binding lectin is associated with protection after experimental stroke in mice and outcome in human stroke // PLoS ONE. 2010. V. 5. No. 2. P. e8433.

Chistiakov D.A., Chernisheva A., Savost'anov K.V. *et al*. The TAF5L gene on chromosome 1q42 is associated with type 1 diabetes in Russian affected patients // Autoimmunity. 2005. V. 38. No. 4. P. 283–293.

Clark I.A., Rockett K.A., Burgner D. Genes, nitric oxide and malaria in African children // Trends Parasitol. 2003. V. 19. No. 8. P. 335–337.

Coleman R.A., Pugh B.F. Evidence for functional binding and stable slid-ing of the TATA binding protein on nonspecific DNA // J. Biol. Chem. 1995. V. 270. No. 23. P. 13850–13859.

Davison B.L., Egly J.M., Mulvihill E.R., Chambon P. Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences // Nature. 1983. V. 301. No. 5902. P. 680–686.

Delgadillo R.F., Whittington J.E., Parkhurst L.K., Parkhurst L.J. The TATA-binding protein core domain in solution variably bends TATA sequences via a three-step binding mechanism // Biochemistry. 2009. V. 48. No. 8. P. 1801–1809.

Drachkova I.A., Ponomarenko P.M., Arshinova T.V. *et al*. *In vitro* examining the existing prognoses how TBP binds to TATA with SNP associated with human diseases // Health. 2011. V. 3. No. 9. P. 577–583.

Drew H.R., Wing R.M., Takano T. *et al*. Structure of a B-DNA dodecamer: conformation and dynamics // Proc. Natl Acad. Sci. USA. 1981. V. 78. No. 4. P. 2179–2183.

Fire A., Samuels M., Sharp P.A. Interactions between RNA polymerase II, factors, and template leading to accurate transcription // J. Biol. Chem. 1984. V. 259. No. 4. P. 2509–2516.

Flicek P., Amode M.R., Barrell D. *et al*. Ensembl 2011 // Nucl. Acids Res. 2011. V. 39. Database issue. P. D800–D806.

Hahn S., Buratowski S., Sharp P.A., Guarente L. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and non nonconsensus DNA sequences // Proc. Natl Acad. Sci. USA. 1989. V. 86. No. 15. P. 5718–5722.

Harrow J., Frankish A., Gonzalez J.M. *et al*. GENCODE: the reference human genome annotation for The ENCODE Project // Genome Res. 2012. V. 22. No. 9. P. 1760–1774.

Karas H., Knuppel R., Schulz W. *et al*. Combining structural analysis of DNA with search routines for the detection of transcription regulatory elements // Comput. Applic. Biosci. 1996. V. 12. No. 5. P. 441–446.

Kim J.L., Nikolov D.B., Burley S.K. Co-crystal structure of TBP recognizing the minor groove of a TATA element // Nature. 1993. V. 365. No. 6446. P. 520–527.

Kim Y., Gieger J.H., Hahn S., Sigler P.B. Crystal structure of a yeast TBP/TATA-box complex // Nature. 1993. V. 365. No. 6446. P. 512–520.

Lifton R., Goldberg M., Karp R., Hogness D. The organization of the histone genes in Drosophila melanogaster: functional and evolutionary implications // Cold Spring Harb. Symp. Quant. Biol. 1978. V. 42. Pt. 2. P. 1047–1051.

Martinez-Carrillo D.N., Garza-Gonzalez E., Betancourt-Linares R. *et al*. Association of IL1B -511C/-31T haplotype and Helicobacter pylori vacA genotypes with gastric ulcer and chronic gastritis // BMC Gastroenterol. 2010. V. 10. P. 126.

Niemann S., Broom W.J., Brown R.H. Jr. Analysis of a genetic defect in the TATA box of the SOD1 gene in a patient with familial amyotrophic lateral sclerosis // Muscle Nerve. 2007. V. 36. No. 5. P. 704–707.

Parker C.S., Topol J. A Drosophila RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene // Cell. 1984. V. 37. No. 1. P. 273–283.

Peltoketo H., Piao Y., Mannermaa A. *et al*. A point mutation in the putative TATA box, detected in nondiseased individuals and patients with hereditary breast cancer, decreases promoter activity of the 17 beta-hydroxysteroid dehydrogenase type 1 gene 2 (EDH17B2) in vitro // Genomics. 1994. V. 23. No. 1. P. 250–252.

Penner C.G., Davie J.R. Transcription factor GATA-1-multiprotein complexes and chicken erythroid development // FEBS Lett. 1994. V. 342. No. 3. P. 273–277.

Pitarque M., von Richter O., Oke B. *et al*. Identification of a single nucleotide polymorphism in the TATA box of the CYP2A6 gene: impairment of its promoter activity // Biochem. Biophys. Res. Commun. 2001. V. 284. No. 2. P. 455–460.

Ponomarenko M., Mironova V., Gunbin K., Savinkova L. Hogness Box // Brenner's Encyclopedia of Genetics. 2nd edn. / Eds S. Maloy, K. Hughes. San Diego: Academic Press, Elsevier Inc., 2013. V. 3. P. 491–494.

Ponomarenko M.P., Ponomarenko J.V., Frolov A.S. *et al*. Identification of sequence-dependent features correlating to activity of DNA sites interacting with proteins // Bioinformatics. 1999. V. 15. No. 7/8. P. 687–703.

Powell R.M., Parkhurst K.M., Parkhurst L.J. Comparison of TATA-binding protein recognition of a variant and consensus DNA promoters // J. Biol. Chem. 2002. V. 277. No. 10. P. 7776–7784.

Savinkova L.K., Drachkova I.A., Arshinova T.V. *et al*. An experimental verification of the predicted effects of promoter TATA-box polymorphisms associated with human diseases on interactions between the TATA boxes and TATA-binding protein // PLoS ONE. 2013. V. 8. No. 2. P. e54626.

Schmidt M.C., Kao C.C., Pei R., Berk A.J. Yeast TATA-box transcription factor gene // Proc. Natl Acad. Sci. USA. 1989. V. 86. No. 20. P. 7785–7789.

Sziller I., Babula O., Hupuczi P. *et al*. Mannose-binding lectin (MBL) codon 54 gene polymorphism protects against development of pre-eclampsia, HELLP syndrome and pre-eclampsia-associated intrauterine growth restriction // Mol. Hum. Reprod. 2007. V. 13. No. 4. P. 281–285.

Takihara Y., Nakamura T., Yamada H. *et al*. A novel mutation in the TATA box in a Japanese patient with beta +-thalassemia // Blood. 1986. V. 67. No. 2. P. 547–550.

Tournamille C., Colin Y., Cartron J.P., Le Van Kim C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals // Nat. Genet. 1995. V. 10. No. 2. P. 224–228.

Wang Y., Kato N., Hoshida Y. *et al*. Interleukin-1beta gene polymorphisms associated with hepatocellular carcinoma in hepatitis C virus infection // Hepatology. 2003. V. 37. No. 1. P. 65–71.

Wieczorek E., Brand M., Jacq X., Tora L. Function of TAF(II)-containing complex without TBP in transcription by RNA polymerase II // Nature. 1998. V. 393. No. 6681. P. 187–191.

Wu K.S., Zhou X., Zheng F. *et al*. Influence of interleukin-1 beta genetic polymorphism, smoking and alcohol drinking on the risk of non-small cell lung cancer // Clin. Chim. Acta. 2010. V. 411. No. 19/20. P. 1441–1446.

Yang M.Q., Laflamme K., Gotea V. *et al*. Genome-wide detection of a TFIID localization element from an initial human disease mutation // Nucl. Acids Res. 2011. V. 39. No. 6. P. 2175–2187.

SNP_TATA_COMPARATOR: WEB SERVICE FOR COMPARISON OF SNPs WITHIN GENE PROMOTERS ASSOCIATED WITH HUMAN DISEASES USING THE EQUILIBRIUM EQUATION OF THE TBP/TATA COMPLEX

D.A. Rasskazov[1], K.V. Gunbin[1], P.M. Ponomarenko[1], O.V. Vishnevsky[1, 2], M.P. Ponomarenko[1], D.A. Afonnikov[1, 2]

[1] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia, e-mail: pon@bionet.nsc.ru;
[2] Novosibirsk National Research State University, Novosibirsk, Russia

Summary

Web-service SNP_TATA_Comparator has been developed for using experimentally proven *in silico* evaluation of *in vivo* biochemical manifestations of SNPs in the core promoter regions of human genes associated with diseases on the base of the TBP/TATA complex equilibrium equation (TBP is TATA binding protein). Data of individual deviations from the reference human genome (Ensembl, rel. 68, i.e. the variant of human genome commonly accepted as datum in comparison analysis) are accessible for free in the real-time mode. Also, data from the GENCODE database (rel. 17) are automatically invoked. The reported Web service SNP_TATA_Comparator is designed for personalized medicine using individual genomes in diagnosis, monitoring, prevention, and treatment.

Key words: SNP, TATA box, TBP, gene expression, human diseases.

УДК 004.75

BioUniWA – СИСТЕМА ГЕНЕРАЦИИ WEB-СЕРВИСОВ И КОНВЕЙЕРОВ ДЛЯ УНИФИЦИРОВАННОГО ДОСТУПА К РЕСУРСАМ В ОБЛАСТИ БИОИНФОРМАТИКИ

© 2013 г. **Е.Г. Комышев[1, 2], М.А. Генаев[2], К.В. Гунбин[2], Д.А. Афонников[1, 2]**

[1] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия, e-mail: komyshev@bionet.nsc.ru;
[2] Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

Представлена компьютерная система BioUniWA, предназначенная для автоматической генерации Web-сервисов для унифицированного доступа к ресурсам в области биоинформатики. Система BioUniWA изначально разрабатывалась как развитие системы BioInfoWF для поддержки доступа к вычислительным модулям и конвейерам посредством Web-сервисов.

Система BioUniWA способна автоматически генерировать Web-сервисы для вычислительных модулей и конвейеров, формальные описания которых определяются посредством языка XML. Данные Web-сервисы в дальнейшем могут использоваться как в различных биоинформационных системах, таких как Taverna, Galaxy, так и непосредственно в программном коде разрабатываемых приложений. Разработанный нами инструмент существенно упрощает аннотацию вычислительных модулей и конвейеров, а также их публикацию в сети Интернет.

Система BioUniWA распространяется под свободной лицензией GNU GPL. Дистрибутив и пользовательская документация системы BioUniWA доступны на сайте http://bioinfowf.bionet.nsc.ru.

Ключевые слова: BioUniWA, унификация доступа, интеграция данных, конвейерная обработка данных, описание вычислительных модулей, Web-интерфейс, Web-сервис, биоинформатика.

ВВЕДЕНИЕ

Для решения конкретной биологической задачи рутинными процедурами являются обращение к базам данных, представленных в разных форматах, использование большого числа программ, объединенных в цепочки, доступ к которым осуществляется различными способами. Все это делает актуальным в биоинформатике использование конвейерных систем для обработки данных (Deelman *et al.*, 2009), а для получения доступа к ресурсам – применение средств, упрощающих интеграцию программных систем и данных.

Унификация доступа подразумевает его единообразие. При использовании стандартизованных интерфейсов биоинформатику не нужно заботиться о множестве различных нюансов при получении доступа к какому-либо ресурсу из систем, таких как Taverna (Hull *et al.*, 2006) или Galaxy (Goecks *et al.*, 2010). Унификацию доступа можно осуществлять с помощью Web-сервисов. Актуальной задачей является также обеспечение возможности повторного использования разработанных конвейеров, их доступности другим пользователям, в том числе и в виде отдельных модулей.

СИСТЕМА BioInfoWF

BioInfoWF – это система генерации Web-интерфейсов и пакет конвейерной обработки данных для решения задач биоинформатики (Генаев и др., 2012). Система BioInfoWF предоставляет простой и удобный способ формального описания вычислительных модулей

с помощью языка, основанного на XML (Bray *et al*., 2006). С помощью таких XML-описаний BioInfoWF способна объединять вычислительные модули данной системы в конвейеры, для которых автоматически генерируется пользовательский Web-интерфейс. Недостаток данной системы заключается в невозможности интегрировать созданные конвейеры и аннотированные вычислительные модули в другие системы, а также в отсутствии графического интерфейса для аннотирования вычислительных модулей и создания новых конвейеров. Одним из решений вышеописанных недостатков BioInfoWF является применение Web-сервисов для унифицированного доступа к вычислительным модулям и их организации в виде конвейеров.

WEB-СЕРВИСЫ

Web-сервис – это программная система, идентифицируемая строкой URI. Ее общедоступные интерфейсы основаны на базе открытых стандартов и протоколов. Web-сервис является единицей модульности при использовании сервис-ориентированной архитектуры приложений и обеспечивает взаимодействие программных систем независимо от платформы. Во множестве современных языков программирования существуют необходимые библиотеки для работы с Web-сервисами (Cerami, 2002; Richardson, Ruby, 2008).

Web-сервисы различаются по типу используемого при реализации протокола, виду запросов и т. д.

В Web-сервисах для передачи данных используются различные протоколы, такие как HTTP (Fielding *et al*., 1999), SMTP (Postel, 1982), FTP (Postel, Reynolds, 1985), SOAP (Simple Object Access Protocol) и другие (Gudgin *et al*., 2003). Часто в качестве транспортного протокола используется HTTP, реже SMTP.

Один из наиболее применяемых типов Web-сервисов – Web-сервисы, основанные на REST (Representational State Transfer), так называемые RESTful Web-сервисы. REST не является протоколом или стандартом, так как это, скорее, стиль построения архитектуры (Pautasso *et al.*, 2008). RESTful Web-сервисы используют основные методы запросов HTTP протокола для реализации разделения логики их запросов. В основной четверке типов запросов HTTP протокола присутствуют самые необходимые методы для реализации базовых запросов, обращенных к Web-приложению. Это методы GET (получение доступа), POST (создание), PUT (обновление), DELETE (удаление). Остальная логика раскрывается более подробно с помощью других конструкций протокола, таких как заголовки, либо при использовании других методов HTTP протокола. Также REST подразумевает обмен сообщениями без сохранения состояния (англ. stateless). Использование этих и других особенностей подхода REST обеспечивает простоту реализации и использования RESTful Web-сервисов.

Другой, получивший распространение, вид Web-сервисов – это SOAP Web-сервисы.

Разработанный консорциумом W3C протокол SOAP (Gudgin *et al*., 2003) является протоколом более высокого уровня, чем другие, ранее упомянутые. При этом SOAP использует один из этих протоколов в качестве транспортного протокола. Протокол SOAP основывается на расширении языка XML и предназначен для обмена структурированными сообщениями в распределенной вычислительной среде. Основное преимущество SOAP в том, что он является стандартом, а использование XML для передачи структурных сообщений позволяет более просто включать сложные конструкции в запросы. Также консорциум W3C разработал язык WSDL (Web Services Description Language) для описания Web-сервисов и доступа к ним, который хорошо сочетается с протоколом SOAP, что позволяет не только реализовывать Web-сервисы, соответствующие стандарту, но и подробно и формально описывать их (Christensen *et al*., 2001). Все это позволяет формировать запросы к таким Web-сервисам автоматически. Существуют различные системы, в которых реализован доступ к SOAP Web-сервисам при использовании описания WSDL (Kawashima *et al.*, 2003; Aloisio *et al.*, 2005).

По типу обмена запросами Web-сервисы бывают синхронными и асинхронными (Erl, 2004). Синхронные Web-сервисы выполняют весь цикл задач за один запрос. Это удобно с точки зрения использования Web-сервиса, так как взаимодействие с ним сводится к одному запросу

и ответу для каждой задачи. С другой стороны, при выполнении длительных вычислений, тем более в конвейере, такие сервисы могут не работать из-за превышения времени соединения на запрос–ответ Web-сервиса. В таких случаях необходимо использовать асинхронные Web-сервисы. Такие сервисы запрашивают каждый этап операции отдельным запросом: начать выполнение, получить статус выполнения, забрать результат. В этом случае выполнение длительных вычислений происходит между запросами, и обрыва соединения не произойдет.

В настоящей работе мы представляем систему BioUniWA – надстройку над системой BioInfoWF (Генаев и др., 2012) для автоматической генерации Web-сервисов и их организации в виде конвейеров. Система BioInfoWF предоставляет простой и удобный способ формального описания биоинформатических ресурсов с помощью языка, основанного на XML. На основе таких XML описаний BioInfoWF автоматически генерирует Web-интерфейсы, а система BioUniWA может генерировать Web-сервисы, которые в дальнейшем могут использоваться как в различных биоинформационных системах, так и непосредственно в программном коде разрабатываемого приложения.

Для ускорения (упрощения) процесса аннотации (формального описания) вычислительных модулей и конвейеров нами было разработано приложение, предоставляющее интерактивный Web-интерфейс, который обеспечивает генерацию аннотаций вычислительных модулей и конвейеров на основе языка XML.

АРХИТЕКТУРА СИСТЕМЫ BioUniWA

Предлагаемая система включает (рис. 1):

1. Web-сервисы для модулей и конвейеров BioInfoWF;
2. Web-интерфейс BioUniWA;
3. Модуль генерации Web-сервисов;
4. Репозитории, в которых хранятся:
   а) вычислительные модули;
   б) описания вычислительных модулей и схем конвейеров, которые являются наборами XML файлов.

Также существует возможность извлечения информации из баз данных.

СТРУКТУРА WEB-СЕРВИСОВ И ИХ ИСПОЛНЕНИЕ

Web-сервисы системы BioUniWA основаны на Java сервлетах ver. 3.0, которые исполняются в рамках контейнера сервлетов. При реализации Web-сервисов были применены как REST (Pautasso *et al.*, 2008) подход, так и протокол SOAP ver 1.2 (Gudgin *et al*., 2003), что позволяет использовать их в различных системах, которые поддерживают только один из данных подходов. В качестве транспортного протокола был выбран HTTP ver. 1.1. Одни модули BioInfoWF выполняются довольно быстро, другие – намного медленнее, поэтому нами были реализованы как синхронные, так и асинхронные типы Web-сервисов.

Обычный запрос методом GET используется для получения простого текстового описания Web-сервиса. Так можно получить информацию о том, как посылать запросы к REST Web-сервису. Метод GET с параметром «wsdl=get» в URL строке используется для получения WSDL описания. Таким образом, имея URL адрес, сервис можно использовать и как REST, и как SOAP. Для реализации SOAP запросов используется метод POST.

Java Web-сервисы находятся под управлением контейнера сервлетов Tomcat ver. 7 (Apache Software Foundation, 2013), который загружает их в память сервера и инициализирует при первом обращении к ним. После инициализации Web-сервис ожидает запросы к нему. При получении

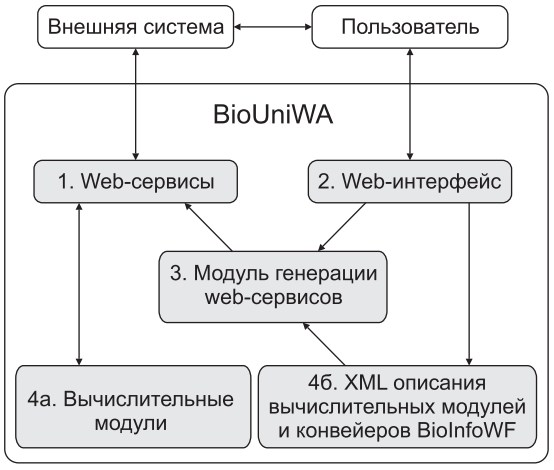


Рис. 1. Основные структурные элементы системы BioUniWA.

запроса Web-сервис определяет вид (REST или SOAP) и тип запроса (создание и выполнение задачи, проверку статуса выполнения, получение результатов выполнения). Таким образом, Web-сервис предоставляет необходимые данные для исполнения, запускает исполнение и возвращает результаты работы модулей или конвейеров BioInfoWF ver. 0.1 (Генаев и др., 2012).

ОПИСАНИЕ МОДУЛЯ ГЕНЕРАЦИИ WEB-СЕРВИСОВ

Модуль генерации запускается из командной строки, принимая на вход команду, опции и пути к необходимым файлам с описаниями модулей и конвейеров. Он анализирует описания вычислительных модулей и конвейеров, производит копирование, модификацию и сборку генерируемого Web-сервиса, упаковывает его в war-файл и помещает в соответствующую директорию, указанную в конфигурации. В случае помещения war-файла в директорию webapps контейнера сервлетов Web-сервис становится доступным для запросов к нему. Формат команды генерации сервисов:
java jar webServicesGenerator.jar <--pipeline> <-a> <moduleName> description.xml…

• Опция pipeline используется при генерации Web-сервиса для конвейера.

• Опция -a используется при генерации асинхронного Web-сервиса.

При генерации Web-сервиса для вычислительного модуля после опций следуют название модуля и файл с описаниями модулей. Если генерация производится для конвейера, то после опций следует список файлов с описаниями. Первый файл в списке должен содержать описание конвейера, а остальные (опционально) должны содержать описания вычислительных модулей, входящих в конвейер. Сначала проверяются описания в указанных файлах, и в случае если они не были найдены в файлах или файлы не указывались, то описания проверяются в хранилище.

WEB-ИНТЕРФЕЙС BioUniWA

Web-интерфейс системы позволяет генерировать описания вычислительных модулей и схем конвейеров, а также производить генерацию Web-сервисов для аннотированных вычислительных модулей и описанных конвейеров.

До генерации описаний вычислительных модулей их установка или обновление на сервере должны быть произведены администратором. При использовании Web-интерфейса BioUniWA необходимо заполнить соответствующие поля, описывающие данный вычислительный модуль. Практически все необходимые данные можно найти в справке по используемой программе, предоставляемой разработчиком. После заполнения полей о входных/выходных файлах и опциях модуля производится генерация описания на специальном языке, основанном на XML. На рис. 2, а продемонстрирован пример описания вычислительного модуля из пакета анализа молекулярной эволюции генов и белков SAMEM (Gunbin *et al.*, 2012) с использованием Web-интерфейса системы BioUniWA. В результате был создан файл с описанием данного модуля (рис. 2, б).

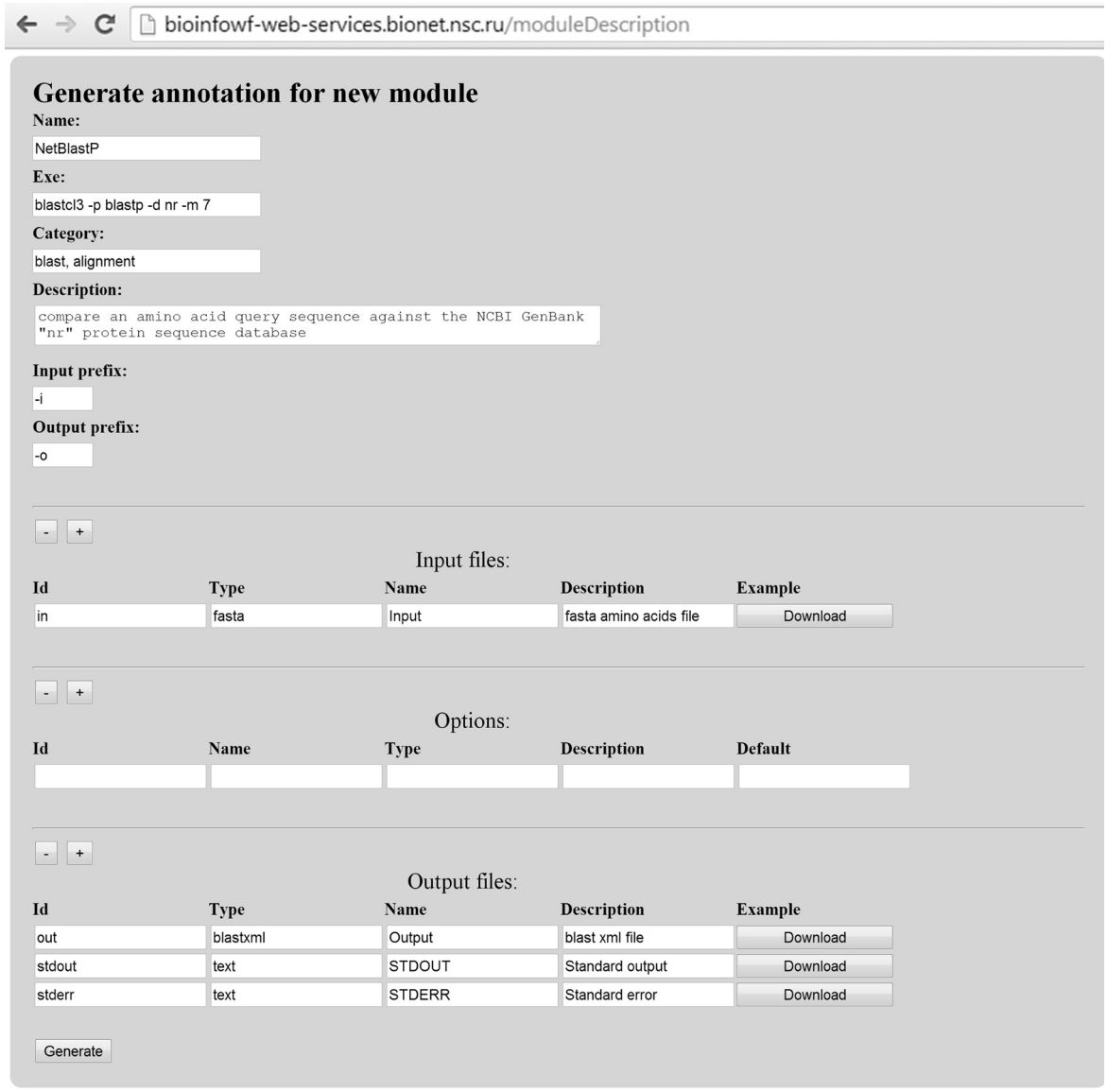
Для генерации описания схемы конвейера необходимо вписать названия аннотированных вычислительных модулей, вовлекаемых в конвейер, после чего будет произведен поиск аннотаций соответствующих вычислительных модулей и создана страница с их описанием (рис. 3). Данный интерфейс позволяет соединять выходы одних вычислительных модулей с входами других и указывать используемые опции модулей при вычислениях. После отправки формы на сервер будет произведена генерация описания схемы конвейера.

Запрос на генерацию Web-сервиса для вычислительного модуля или созданного конвейера можно отправить, открыв страницу для генерации Web-сервисов. При этом будет произведен поиск необходимых описаний и создан соответствующий Web-сервис.

ХРАНИЛИЩЕ АННОТАЦИЙ ВЫЧИСЛИТЕЛЬНЫХ МОДУЛЕЙ И ОПИСАНИЙ КОНВЕЙЕРОВ

Аннотации вычислительных модулей и описания конвейеров хранятся в специальной директории на сервере. При создании новых описаний они сохраняются отдельно до предстоящей проверки. Хранилище накапливает описания конвейеров для их последующего использования в компьютерных экспериментах.

а



б

```
<programs>
  <program name="NetBlastP" exe=".../blastcl3" category="blast, alignment" >
    <description>compare an amino acid query sequence against
          the NCBI GenBank "nr" protein sequence database</description>
    <input>
      <file id="in" type="fasta" name="Input" description="fasta amino acids file"/>
    </input>
    <output>
      <file id="out" type="blastxml" name="Output" description="blast xml file" />
      <file id="stdout" type="text" name="STDOUT" description="Standard output" />
      <file id="stderr" type="text" name="STDERR" description="Standard error" />
    </output>
    <cmdline>
      $cmd = "\"-p\" \"blastp\" \"-d\" \"nr\" \"-i\" \"$files{in}\" \"-m\" \"7\" \"-o\" \"$files{out}\"".
      " 1>\"$files{stdout}\" 2>"$files{stderr}\"";
    </cmdline>
  </program>
</programs>
```

Рис. 2. Страница Web-интерфейса для генерации описания вычислительного модуля (а); пример XML описания вычислительного модуля (б).

В результате генерации аннотации вычислительного модуля Web-интерфейсом системы BioUniWA создается файл с содержанием, описывающим вычислительный модуль. Заполненные поля «in», «out», «stdout», «stderr» и др. в Web-интерфейсе соответствуют XML-элементам и атрибутам в файле.

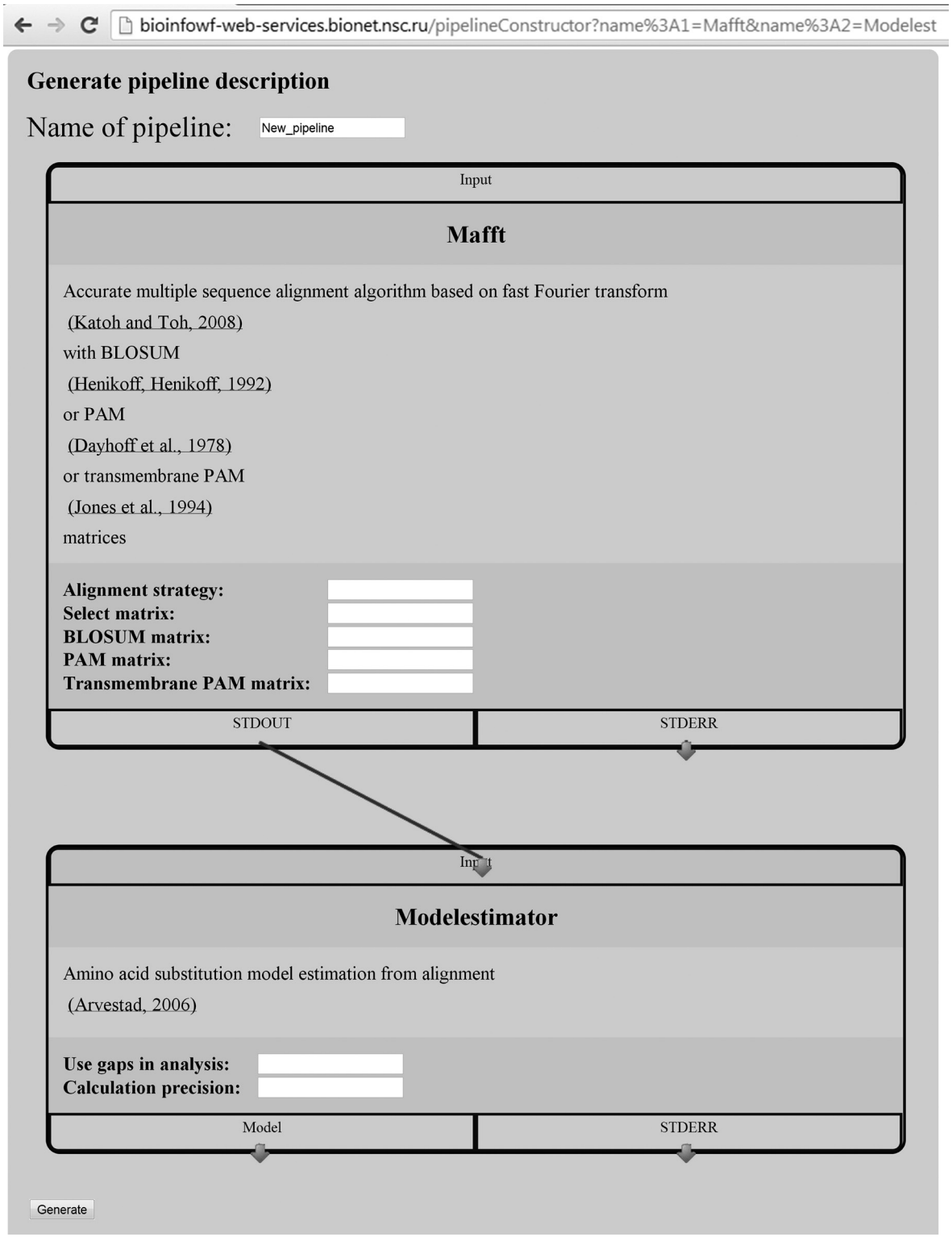


Рис. 3. Страница Web-интерфейса для генерации схемы конвейера.

Страница содержит визуальные представления вычислительных модулей, вовлекаемых в конвейер. Представления содержат краткое описание вычислительного модуля, идентификаторы входных файлов модуля – сверху, и идентификаторы выходных файлов – снизу. Данный Web-интерфейс позволяет указывать связи выходных файлов одних вычислительных модулей с входными файлами других, ассоциируя стрелками эти файлы.

ИЗВЛЕЧЕНИЕ ИНФОРМАЦИИ ИЗ БАЗ ДАННЫХ

Извлечение информации из баз данных осуществляется путем создания клиентов для этих баз данных. При этом вся бизнес-логика определена внутри такого клиента. Система BioUniWA может рассматривать клиента как вычислительный модуль. Таким образом, наша система позволяет реализовывать API для любых существующих баз данных.

ПРИМЕНЕНИЕ СИСТЕМЫ BioUniWA

Основная задача системы BioUniWA – унификация доступа к различным ресурсам в области биоинформатики при помощи автоматической генерации Web-сервисов для данных ресурсов. На данный момент в BioUniWA реализована генерация Web-сервисов для вычислительных модулей и конвейеров системы BioInfoWF с помощью Web-интерфейса. Это упрощает процедуру обновления вычислительных модулей по мере появления новых программ или выхода обновлений к программам, описанным модулями BioInfoWF.

Компьютерная система анализа режимов эволюции белок-кодирующих генов SAMEM (Gunbin *et al.*, 2012) построена на основе вычислительных модулей BioInfoWF, что позволяет непосредственно протестировать возможности BioUniWA. SAMEM состоит из двух основных конвейеров, анализа эволюции генов (состоящего из 12 вычислительных модулей) и анализа эволюции белков (состоящего из 10 вычислительных модулей) и двух дополнительных, собирающих выборки генов и белков и производящих их первичный анализ (состоят в общей сложности из 6 вычислительных модулей). Благодаря разработанной системе BioUniWA в виде Web-сервисов были созданы конвейеры SAMEM, использующие как ранее установленное, так и обновленное программное обеспечение. Управление одним из конвейеров SAMEM частично представлено на рис. 3.

Дополнительная задача BioUniWA – организация доступа к базам данных посредством программ-клиентов. В настоящее время для базы данных PEFF DB (авторское свидетельство № 2012620659), интегрирующей информацию о режимах эволюции белок-кодирующих генов и данные о динамике функционирования генных сетей, в которых функционируют эти гены, разрабатывается возможность авторизованного доступа к информации этой базы данных опосредованного Web-сервисами BioUniWA.

ЗАКЛЮЧЕНИЕ

В настоящей работе предложена система BioUniWA, разрабатываемая для унификации доступа к ресурсам в области биоинформатики, которая позволяет генерировать Web-сервисы для решения биоинформатических задач, расширять функционал существующей системы BioInfoWF, создавая с помощью Web-интерфейса системы необходимые описания вычислительных модулей и конвейеров BioInfoWF (Генаев и др., 2012). Удобство системы заключается в быстром получении доступа к вычислительным модулям и конвейерам BioInfoWF посредством Web-сервисов. Основным преимуществом представленной системы является простота работы с ней. Таким образом, для создания собственных Web-сервисов пользователю не требуется специальных технических знаний, достаточно воспользоваться разработанным Web-интерфейсом системы BioUniWA. Использование системы было продемонстрировано на примере создания описаний новых вычислительных модулей и конвейеров BioInfoWF, а также генерации Web-сервисов уже существующих вычислительных модулей для решения задач по анализу молекулярной эволюции генов и белков – SAMEM (Gunbin *et al.*, 2012).

Система BioUniWA распространяется под свободной лицензией GNU GPL ver. 3 (Free Software Foundation …, 2007). Дистрибутив и пользовательская документация системы доступны на официальном сайте http://bioinfowf.bionet.nsc.ru.

БЛАГОДАРНОСТИ

Работа поддержана интеграционным проектом СО РАН 39, грантом РФФИ 11-04-01771-а, программами РАН «Происхождение и эволюция биосферы» и «Молекулярная и клеточная биология» (проект 6.6), программой поддержки ведущих научных школ НШ-5278.2012.4.

ЛИТЕРАТУРА

Генаев М.А., Комышев Е.Г., Гунбин К.В., Афонников Д.А. BioInfoWF – система автоматической генерации Web-интерфейсов и Web-сервисов для биоинформационных исследований // Вавилов. журн. генет. и селекции. 2012. Т. 16. № 4/1. С. 849–857.

Aloisio G., Cafaro M., Fiore S., Mirto M. A WorkFlow management system for bioinformatics grid // Proc. of the Network Tools and Applications in Biology (NETTAB) Workshops, Naples, Italy, 2005.

Apache Software Foundation, Apache Tomcat. 2013. http://tomcat.apache.org

Bray T., Paoli J., Sperberg-McQueen C.M. *et al.* Extensible Markup Language (XML) (2006). Available at: www.w3.org/TR/xml.

Cerami E. Web services essentials: distributed applications with XML-RPC, SOAP, UDDI and WSDL. O'Reilly Media, Inc., 2002.

Christensen E., Curbera F., Meredith G., Weerawarana S. Web services description language (WSDL) 1.1. 2001. http://www.w3.org/TR/wsdl

Deelman E., Gannon D., Shields M., Taylor I. Workflows and e-Science: An overview of workflow system features and capabilities // Future Generation Computer Systems. 2009. V. 25. No. 5. P. 528–540.

Erl T. Service-oriented architecture. Englewood Cliffs: Prentice Hall, 2004.

Fielding R., Gettys J., Mogul J. *et al*. Hypertext transfer protocol–HTTP/1.1. 1999.

Free Software Foundation, Inc., GNU General Public License, 2007. http://www.gnu.org/licenses/gpl.html

Goecks J., Nekrutenko A., Taylor J. Galaxy a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences // Genome Biol. 2010. V. 11. No. 8. P. R86.

Gudgin M., Hadley M., Moreau J. *et al*. Simple Object Access Protocol (SOAP) 1.2. W3C, 2003.

Gunbin K.V., Suslov V.V., Genaev M.A., Afonnikov D.A. Computer system for analysis of molecular evolution modes (SAMEM): analysis of molecular evolution modes at deep inner branches of the phylogenetic tree // In Silico Biol. 2012. V. 11. No. 3. P. 109–123.

Hull D., Wolstencroft K., Stevens R. *et al.* Taverna: a tool for building and running workflows of services // Nucl. Acids Res. 2006. V. 34. Suppl. 2. W729–W732.

Kawashima S., Katayama T., Sato Y., Kanehisa M. KEGG API: A web service using SOAP/WSDL to access the KEGG system // Genome Inform. Ser. 2003. P. 673–674.

Pautasso C., Zimmermann O., Leymann F. Restful web services vs. big'web services: making the right architectural decision // Proc. 17th Intern. Conf. on World Wide Web. ACM, 2008. P. 805–814.

Postel J., Reynolds J. File transfer protocol. 1985., available at http://tools.ietf.org/html/rfc959

Postel J. Simple mail transfer protocol // Inform. Sci. 1982.

Richardson L., Ruby S. RESTful web services. O'Reilly, 2008.

BioUniWA – WEB SERVICES GENERATION SYSTEM AND PIPELINES FOR UNIFIED ACCESS TO RESOURCES IN THE FIELD OF BIOINFORMATICS

E.G. Komyshev[1, 2], M.A. Genaev[2], K.V. Gunbin[2], D.A. Afonnikov[1, 2]

[1] Novosibirsk National Research State University, Novosibirsk, Russia,
e-mail: komyshev@bionet.nsc.ru;
[2] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

Summary

We present the BioUniWA system for automatic generation of Web services for unified access to resources in the field of bioinformatics. The BioUniWA system was originally designed as the BioInfoWF development system to support access to computing modules and pipelines through Web services. BioUniWA can automatically generate Web-based applications for computational modules and pipelines whose formal descriptions are defined by a language based on XML. In the future, one will be able to use these Web-based applications in a variety of bioinformational systems, such as Taverna or Galaxy, as well as directly in the source code of applications to be developed. We have designed a tool that greatly simplifies the annotation of computing modules and schemes of pipelines, as well as their publication via Internet.

BioUniWA is distributed under a free GNU general public license. The distribution package and user documentation are available at BioUniWA http://bioinfowf.bionet.nsc.ru.

Keywords: BioUniWA, unification of access, data integration, pipelined data processing, descriptions of computational modules, Web interface, Web services, bioinformatics.

УДК 577.214.626+316.452

NETINFERENCE: ПРОГРАММЫ ДЛЯ АНАЛИЗА СТРУКТУРЫ И ДИНАМИКИ СЕТЕЙ

© 2013 г. **И.И. Титов[1, 2], А.А. Блинов[2], К.А. Рудниченко[3], П.В. Крутов[2], А.Л. Казанцев[2], А.И. Куликов[2, 3]**

[1] Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия, e-mail: titov@bionet.nsc.ru;
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия;
[3] Федеральное государственное бюджетное учреждение науки Институт вычислительной математики и математической геофизики Сибирского отделения Российской академии наук, Новосибирск, Россия

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

В работе представлен пакет компьютерных программ для анализа структурно-функциональной организации и эволюции во времени биологических, социальных и других сетей. Пакет позволяет исследовать как глобальную архитектуру сетей, так и их локальные свойства, при этом выявлять ключевые регуляторы и структурно-функциональные модули, а также проследить развитие сетей во времени. Работа пакета иллюстрирована на примере нескольких генных сетей, сети соавторства научных публикаций в области биологии и медицины, а также сети терминов и ключевых слов из этой же области знаний.

Ключевые слова: генная сеть, сеть соавторства, сеть научных терминов, структура сети, динамика сети, синхронная булева модель, компьютерный анализ.

ВВЕДЕНИЕ

Удобным способом представления сложных систем являются сети. Большинство биологических, социальных, технологических и других сетей не являются регулярными или случайными, а обладают сходной сложной архитектурой связей. Для устройства этих сетей характерны сильная кластеризация и малый диаметр, свойство так называемого «малого мира». В результате такие сети демонстрируют интересные динамические свойства (Newman, 2003).

Для понимания организации и функционирования столь сложных сетей необходимо исследование их архитектуры с разных сторон и на разных масштабах рассмотрения: изучение глобальной и локальной структуры, выявление модулей и ключевых элементов, моделирование динамики и эволюции. В статье представлен пакет компьютерных программ, направленных на решение этой задачи, его работа продемонстрирована на примере некоторых генных, социальных и словарных сетей.

МАТЕРИАЛЫ И МЕТОДЫ

Пакет реализован в виде трех программ. Первая – программа для анализа глобальной и локальной архитектуры сетей. Программа реализована на языке C# и производит следующие вычисления: рассчитываются глобальные характеристики сети – распределение вершин по связям и аппроксимация этого распределения степенной зависимостью, диаметр сети и глобальный коэффициент кластеризации C1. В качестве локальных характеристик сети рассчитываются локальный коэффициент кластеризации C2 и коэффициент корреляции по степеням вершин. Неслучайно часто повторяющиеся модули сети (мотивы) находятся с помощью алгоритма

FANMOD (Wernicke, Rasche, 2006). Для ускорения расчетов на сетях больших размеров этот алгоритм модифицирован и использует библиотеку изоморфных графов, которая построена при помощи алгоритма Nauty (McKay, Piperno, 2013). Важно, что группы графов из библиотеки оказываются очень неравномерными по численности, что обосновывает необходимость точной оценки ожидаемой встречаемости при расчете статистической значимости мотива. В целом описанная программа носит общий характер и используется как дополнение для анализа конкретных сетей при помощи второй и третьей программ.

Вторая программа моделирует динамику и выявляет структурные модули и ключевые элементы генных сетей на основе синхронной булевой модели (Kauffman, 1969). Программа осуществляет полный перебор пространства состояний сети и определяет соседние во времени состояния, аттракторы, бассейны притяжения аттракторов и скорости переходов между бассейнами под действием шума заданной величины. Для ускорения расчетов и выявления структурно-функциональных организаций сети используются два подхода. В первом, статическом, сеть разбивается на полунезависимые кластеры одним из выбранных методов: «жадной» оптимизацией модульности, алгоритмом на основе определения смежности и случайных блужданий. После декомпозиции сети моделируется динамика каждого кластера по отдельности, затем на основе динамики отдельных кластеров восстанавливается динамический портрет всей сети. Во втором, динамическом, вершины сети начиная от полностью забуференных, рекурсивно удаляются в зависимости от степени их «канализированности» (Kauffman, 1969), пока не останется только «вычислительное ядро» сети, которое однозначно определяет динамику всей системы. В обратной процедуре динамика сети восстанавливается по динамике ее ядра. Влияние шума экспрессии генов на динамику сети моделируется методом Монте-Карло, в результате чего производится классификация вершин по степени их влияния на переходы между бассейнами притяжения стационарных состояний. Программа реализована на языке C++.

Третья программа реализована на языках SQL и Java и направлена на исследование ландшафта и временной эволюции тех сетей, которые могут быть получены из базы данных научных публикаций по биологии и медицине PubMed – сети соавторства и сети научных терминов. Кластеризация сети осуществляется при помощи модифицированного алгоритма SCAN (Xu *et al*., 2007). Восстановление эволюции сети осуществляется на основе определения соответствия между кластерами на соседних временных срезах. Периоды повышенного интереса к научной области определялись с помощью скрытой марковской модели для временного профиля частоты использования научных терминов.

РЕЗУЛЬТАТЫ

Выявление вычислительной архитектуры генных сетей на основе синхронной булевой модели

Тестирование программы рекурсивной редукции генной сети проводилось на сети ответа на стресс *E. coli*, исходно содержащей 73 вершины (Stepanenko, Titov, 2010), что соответствует 2^{73} возможным состояниям сети при полном переборе. Двукратное применение декомпозиции и редукции графа сократило его размер сначала до 32, а затем и до вполне вычислимого графа из 10 вершин, по состояниям которого восстанавливается полный динамический портрет сети.

Влияние шума экспрессии генов на динамику генной сети моделировалось для хорошо изученной генной сети морфогенеза цветка *Arabidopsis thaliana* (Alvarez-Buylla *et al*., 2008). Были выявлены критические динамические состояния генной сети, т. е. такие пары соседних состояний, которые принадлежат разным бассейнам. Информация о точках бифуркаций динамических траекторий генной сети была обобщена в виде ранжирования генов по степени влияния на неустойчивость траекторий (табл.). Из таблицы видно, что степень влияния шума может сильно варьировать от вершины к вершине. Более половины всех переходов между бассейнами были спровоцированы шумом в вершинах LFY и UFO. Первая из них соответствует гену Leafy, который инициирует развитие недифференцированных клеток, а вторая – F-box протеин, отвечающий за дифференциацию аттракторов Pet1–Pet2 и

Таблица

Относительная роль вершин, шум в которых инициирует переходы между бассейнами притяжения

Вершины генной сети, морфогенеза цветка	AG	AP1	AP2	AP3	EMF1	Ft	FUL	LFY	P1	Sep	TFl1	UFO	WUS
% от общего числа переходов	9,2	7,0	5,3	7,1	4,8	0,7	0	22,1	0,3	0,3	8,6	34,4	0

Stm1–Stm2. Известно, что мутагенное влияние на эти гены влечет различные нарушения в процессе развития цветка и приводит к изменению его внешнего вида.

Моделирование динамики генной сети методом Монте-Карло при заданном уровне шума позволяет построить матрицу Маркова. Эта матрица используется для определения кинетики системы в терминах населенности бассейнов (Alvarez-Buylla *et al*., 2008): порядка и скоростей переходов, устойчивости бассейнов и крупнозернистых кинетических мод, определяемых собственными числами матрицы.

Исследование развития научных коллективов ИЦиГ СО РАН на основе анализа сети соавторов научных публикаций

Построение временных срезов статистических характеристик сети и численности коллективов соавторов ИЦиГ СО РАН показывает, что, несмотря на нестационарный характер их поведения, можно выделить наиболее равномерный период развития, относящийся к 1998–2004 гг. (рис. 1, 2). При этом на протяжении всего рассмотренного времени наблюдаются возникновение, слияние, разделение и исчезновение кластеров соавторов (рис. 1). Наиболее существенное изменение структуры сети датируется 2007–2009 гг., сопряженными с выделением из института части лабораторий. Начиная с 2009 г. сеть возвращается к более плавной эволюции во времени (рис. 1, 2).

Построение динамики и ландшафта научных направлений на основе анализа сети научных терминов

В сравнении с сетями, которые были рассмотрены выше, еще более выраженно дисассортативными и кластеризованными (с высокими значениям коэффициента кластеризации и низким коэффициентом корреляции степеней вершин) оказались сети терминов научных публикаций. В качестве примера эволюции научной области

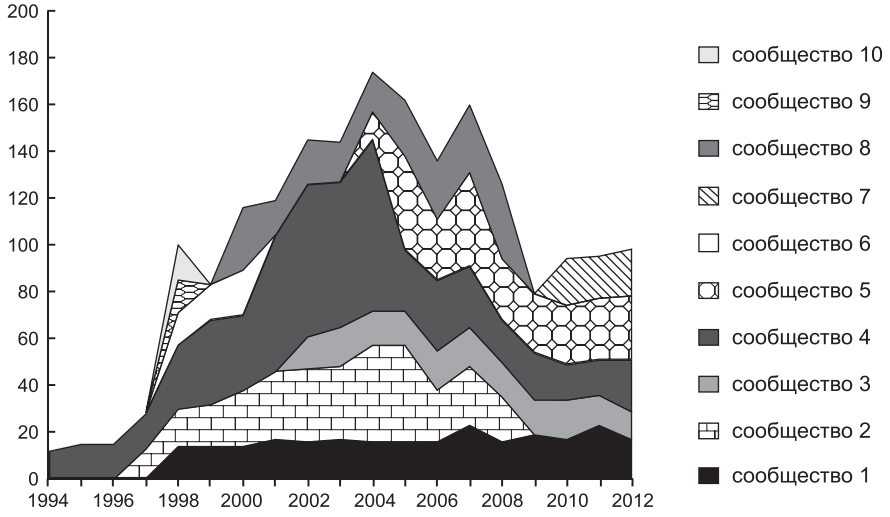


Рис. 1. Динамика численности наиболее крупных кластеров.

Каждый кластер соответствует сообществу внутри ИЦиГ СО РАН, которое образовано соавторством в научных публикациях, аннотированных в базе PubMed.

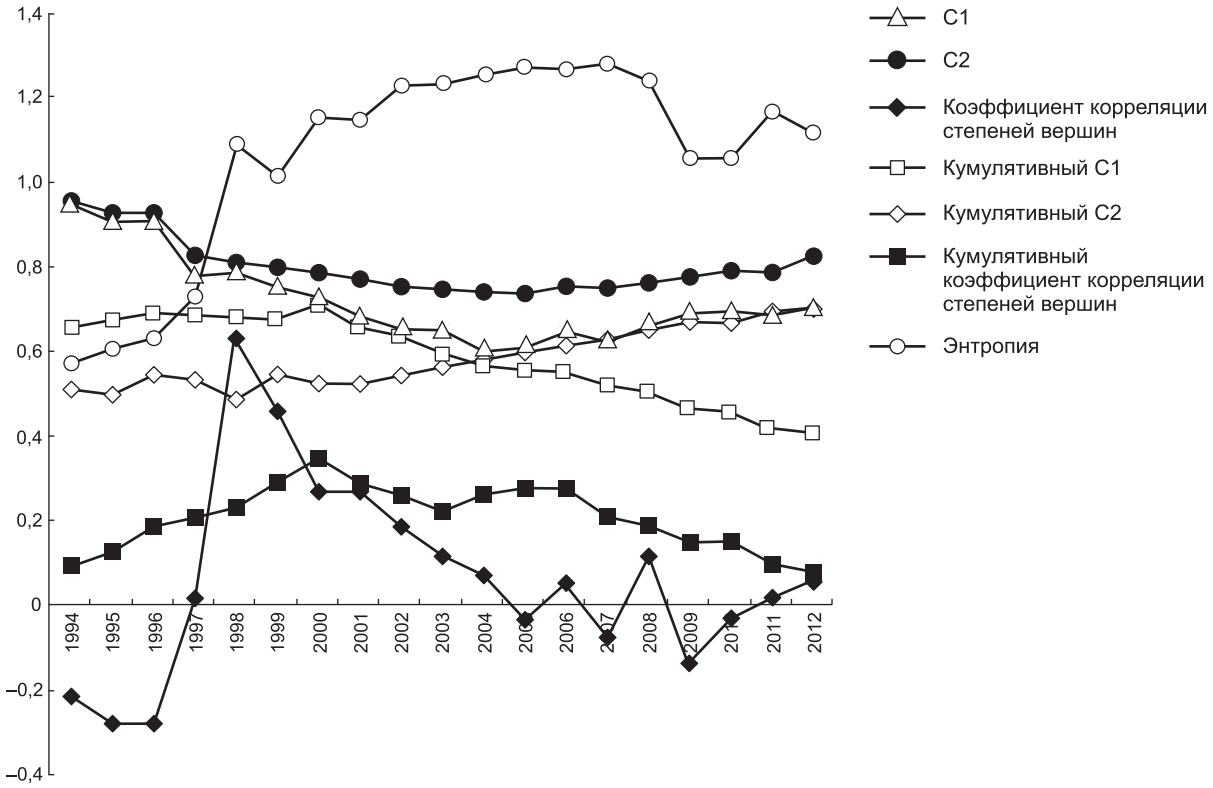


Рис. 2. Кумулятивные и мгновенные статистические характеристики сети соавторов ИЦиГ СО РАН.

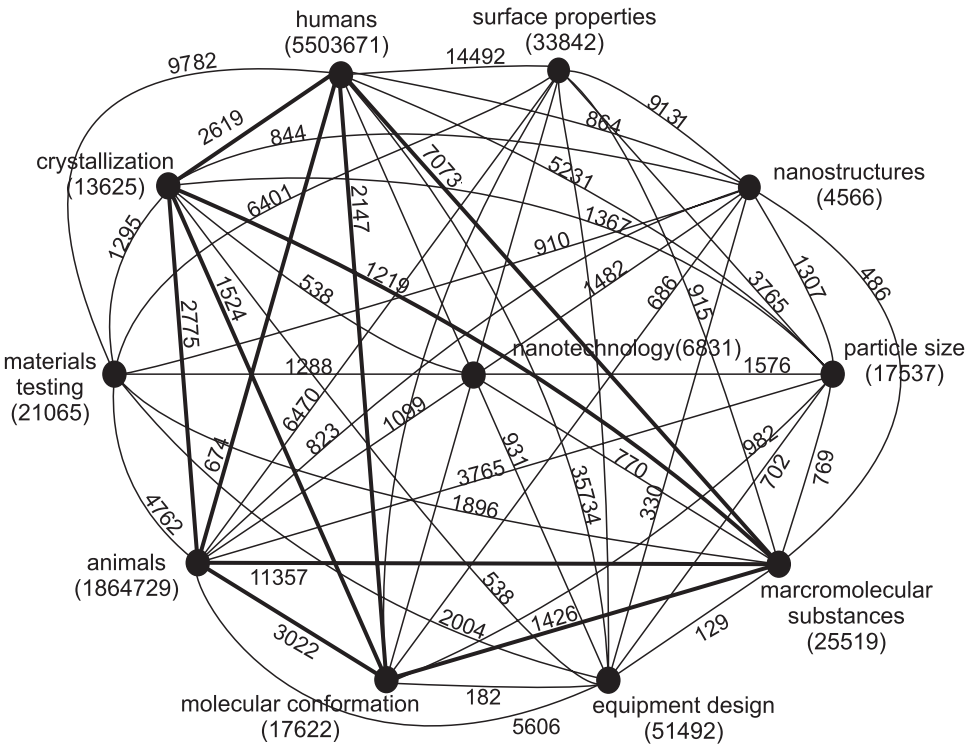


Рис. 3. Фрагмент кумулятивного ландшафта научных направлений в области нанотехнологий, построенного по аннотациям научных статей из базы PubMed.

Веса вершин соответствуют встречаемости терминов в аннотациях, веса ребер – совместной встречаемости терминов. Жирными ребрами показан кластер терминов, соответствующий исследованиям живой природы.

была рассмотрена область нанотехнологий. Хотя область науки «нанотехнология» имеет историю в несколько десятилетий и восходит к лекции Р. Фейнмана, само это слово возникло лишь около 30 лет назад. Еще более недавним является использование термина «nanotechnology» в кратком содержании научных статей в базе PubMed. Впервые термин употребляется в 1995 г., но с того момента частота его использования растет экспоненциально, что отражает взрывной всплеск интереса к этой области науки и ее присутствие в начальной стадии кривой Гартнера развития технологий. Построение сети терминов в области нанотехнологий показывает разделение на области знаний живой и неживой природы (рис. 3).

ЗАКЛЮЧЕНИЕ

Исследование сложных систем часто невозможно представить без изучения свойств сетей, моделирующих эти системы. Такие сети обычно обладают необычной топологией и характеризуются богатой динамикой. В работе представлен набор компьютерных программ для изучения биологических, социальных и других сетей. Разработанные программы предназначены для изучения глобальных и локальных свойств сложных систем, а также их развития во времени.

БЛАГОДАРНОСТИ

Работа поддержана Междисциплинарным интеграционным проектом СО РАН № 21 и Президентской программой по государственной поддержке ведущих научных школ РФ НШ-5278.2012.4.

ЛИТЕРАТУРА

Alvarez-Buylla E.R., Chaos A., Aldana M. *et al*. Padilla-longoria floral morphogenesis: stochastic explorations of a gene network epigenetic landscape // PLoS ONE. 2008. V. 3. No. 11.

Kauffman S.A. Metabolic stability and epigenesis in randomly constructed genetic nets // J. Theor. Biol. 1969. V. 22. P. 437–467.

McKay B.D., Piperno A. Practical graph isomorphism. II. 2013. 22 p. http://arxiv.org/abs/1301.1493.

Newman M.E.J. The structure and functions of complex networks // SIAM Rev. 2003. V. 45. No. 2. P. 167–256.

Stepanenko I.L., Titov I.I. Computer analysis of stress response network *E. coli* // Proc. 7th Int. Conf. on Bioinformatics of Genome Regulation and Structure\Systems Biology BGRS\SB.10. Novosibirsk, Russia, June 20–27 2010. Novosibirsk. P. 278.

Wernicke S., Rasche F. FANMOD: a tool for fast network motif detection // Bioinformatics. 2006. V. 22. No. 9. P. 1152–1153.

Xu X., Yuruk N., Feng Zh., Schweiger T.A.J. SCAN: a structural clustering algorithm for networks // Proc. KDD ‘07 Proc. of the 13th ACM SIGKDD Intern. Conf. on Knowledge discovery and data mining. N.Y., 2007. P. 824–833.

NETINFERENCE: COMPUTER PROGRAMS FOR REVEALING NETWORK STRUCTURE AND DYNAMICS

I.I. Titov[1,2], A.A. Blinov[2], K.A. Rudnichenko[3], P.V. Krutov[2], A.L. Kazantsev[2], A.I. Kulikov[2,3]

[1] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: titov@bionet.nsc.ru;
[2] Novosibirsk National Research State University, Novosibirsk, Russia;
[3] Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia

Summary

We present a computer package for analyzing the structure-functional organization and evolution of biological, social and other networks. The programs allows investigation of not only the global network architecture, but also its local properties, revealing key regulators and structure-functional modules. Also, the network evolution can be traced. The package has been tested with two gene networks: the co-authorship network of biomedical papers and the biomedical term network.

Key words: gene network, co-authorship network, term network, network structure, network dynamics, synchronous Boolean model, computer analysis.

УДК 576.32/36:612.014

МЕЖМОЛЕКУЛЯРНЫЕ ВЗАИМОДЕЙСТВИЯ В ФУНКЦИОНАЛЬНЫХ СИСТЕМАХ НЕЙРОНА

© 2013 г. **А.Л. Проскура[1], И.А. Малахин[1], И.И. Турнаев[2], В.В. Суслов[2], Т.А. Запара[1], А.С. Ратушняк[1]**

[1] Конструкторско-технологический институт ВТ СО РАН, Новосибирск, Россия;
[2] Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия,
e-mail: zapara_t@mail.ru

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

Синаптические нейрональные контакты являются одним из основных элементов, обеспечивающих пластичность нервной системы, а изменение эффективности синаптической передачи ответственно за такие реакции, как восприятие, проведение возбуждения, обучение и память. Дендритные шипики представляют постсинаптическую часть возбуждающих синапсов высших отделов мозга млекопитающих. Белок-белковые сети микродоменов шипиков формируют функциональную систему синапсов нейрона. Проведена реконструкция концептуальной модели межмолекулярных взаимодействий, обеспечивающих изменение эффективности синаптической передачи вслед за активацией синапса, интеграцию возбуждения в локальной дендритной сети нейрона и длительное поддержание нового уровня нейротрансмиссии.

Ключевые слова: синаптическая пластичность, долговременная потенциация, глутаматные рецепторы.

ВВЕДЕНИЕ

Общепринятой клеточной моделью для изучения синаптической пластичности *in vitro* является долговременная потенциация (ДВП) – усиление синаптической передачи между нейронами, возникающее после интенсивного и непродолжительного выброса нейротрансмиттера и сохраняющееся на протяжении длительного времени (Bliss, Collingridge, 1993). Индукция ДВП происходит, например, в результате высокочастотной стимуляции афферентных входов пирамидных нейронов гиппокампа. Гиппокамп активно вовлечен в процессы восприятия информации, ее распознавания, анализа и запоминания (Kjelstrup *et al*., 2008; Hawley *et al*., 2012). Поле CA1 гиппокампа обладает выраженной ламинарной организацией клеточных связей и малым числом рекуррентных взаимодействий (Szirmai *et al*., 2012), что делает его перспективным объектом для исследования ДВП.

Изменение уровня синаптической передачи (синаптическая эффективность) зависит от пре- и постсинаптических механизмов, многие из которых детально изучены (Bliss, Collingridge, 1993; Petersen *et al*., 1998; O'Connor *et al*., 2005). Важная роль в этих процессах, по современным представлениям, отводится активности ионотропных глутаматных рецепторов (Shepherd, Huganir, 2007).

В пирамидальных клетках CA1 поля гиппокампа по чувствительности к агонистам выделяют рецепторы НМДА (агонистом является N-метил-D-аспарагиновая кислота) и АМПА (агонист – α-амино-3-гидрокси-5-метил-4-изоксазолпропионовая кислота) типа.

НМДА рецепторы (НМДАР) состоят из комбинации 5 субъединиц, кодируемых отдельными генами (Grin1 (субъединица zeta), Grin2a-2d (субъединицы epsilon 1-4) (Nagasawa *et al*., 1996)), которые объединяются в рецепторно-ионофорный комплекс и обладают рядом

особенностей: одновременно хемо- и потенциалчувствительностью, медленной динамикой запуска и длительностью эффекта, способностью к временной суммации. Каналы НМДАР пропускают ионы кальция (Nowak *et al*., 1984).

В состав АМПА рецепторов (АМПАР) взрослых животных входят 3 субъединицы (Glu 1-3). Тетрамеры АМПАР формируются из двух гомодимеров одной из субъединиц. Так, Glu1/1 являются гомотетрамерами гомодимеров субъединиц Glu1, а гетеротетрамеры Glu1/2 формируются из гомодимеров субъединиц Glu1 и Glu2. Эти два типа рецепторов часто объединяют под общим термином – Glu1 АМПАР. Гетеротетрамеры Glu2/3 собираются из гомодимеров субъединиц Glu2 и Glu3, их часто обозначают как Glu2 АМПАР. Каждая субъединица кодируется отдельным геном (Gria 1-3) (Palmer *et al*., 2005). Рецепторы, содержащие в своем составе Glu2 (Glu1/2, Glu2/3), проницаемы для ионов натрия, но не ионов кальция. Гомотетрамеры Glu1/1 проницаемы для ионов кальция (Сергеев и др., 1999; Palmer *et al*., 2005).

Различают две основные стадии ДВП: раннюю фазу, сопровождающуюся модификациями уже существующих синаптических белков, и позднюю, которая коррелирует с увеличением белкового синтеза и генетической экспрессией (Steward, Schuman, 2001). Показано, что развитие ДВП сопровождается изменениями морфологии дендритных шипиков – небольших (< 1 μm^3) выростов мембраны осевого дендрита, богатых актином и образующих постсинаптическую часть большинства возбуждающих синапсов мозга млекопитающих (Hotulainen, Hoogenraad, 2010). Эти структуры имеют микродоменную организацию. Микродомены – функциональные комплексы белков и липидов, в которых осуществляется физическое взаимодействие и позиционирование молекул партнеров определенного процесса в надмолекулярные комплексы (Tsunoda *et al*., 1997). Идентифицировано более 1100 белков в синаптической терминали мозга мыши, взаимодействия которых определяют функциональное равновесие между пластичностью и стабильностью эффективности синаптических связей (Collins *et al*., 2006).

НМДАР-зависимый вход ионов кальция после паттерн-зависимого активирования синапса (интенсивного непродолжительного выброса медиатора) считается главным событием, определяющим развитие ДВП в СА1 поле гиппокампа (Raghuram *et al*., 2012). Блокада НМДАР нарушает процессы синаптической пластичности и формирование пространственной памяти (Tsien *et al*., 1996). Число АМПАР на синаптической мембране и их субъединичный состав подвергаются динамичным изменениям в зависимости от состояния синапса, что служит одним из ключевых механизмов изменения синаптической эффективности в поле СА1 гиппокампа (Shepherd, Huganir, 2007).

Исследования пространственно-временной динамики белок-белковых взаимодействий при изменении и поддержании эффективности синаптической передачи могут привести к более глубокому пониманию процессов обработки информации на клеточном уровне. На основании обширных гетерогенных экспериментальных данных впервые проведен комплексный анализ межмолекулярных регуляторных взаимодействий, позволивший реконструировать процессы синхронизации молекулярных ансамблей различных сигнальных систем, реорганизации цитоскелета и транспортной системы клетки.

РЕКОНСТРУКЦИЯ МЕЖБЕЛКОВЫХ ВЗАИМОДЕЙСТВИЙ ДЕНДРИТНОГО ШИПИКА В ТЕЧЕНИЕ ИЗМЕНЕНИЯ ЭФФЕКТИВНОСТИ СИНАПТИЧЕСКОЙ ПЕРЕДАЧИ

Реконструкция межбелковых взаимодействий в дендритных шипиках СА1 поля гиппокампа проводилась с использованием технологии GeneNet (РОСПАТЕНТ № 990006 от 15/02/1999 (Ananko *et al*., 2005)).

В белок-белковой сети «Intermediate-LTP» (http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/AMPA.html) описаны процессы, обеспечивающие усиление нейротрансмиссии, а также поддержание ее нового уровня в течение ДВП. Отдельно реконструированы процессы регулирования формирования везикул, переносящих вновь синтезированные белки из сомы в дендрит (сеть «Vesicle's trafficking», локальная версия). В таблице представлена суммарная по двум сетям информация о белках и процессах, задействованных на различных этапах развития и поддержания ДВП (табл.).

Таблица

Белки и кодирующие их гены, участвующие в процессах поддержания нового уровня нейротрансмиссии в течение ДВП

Группа белков и кодирующих их генов*	Процесс
Биосинтез в соме и доставка к ПМ	
GluR1 (*Gria1*), GluR2 (*Gria2*), GluR3 (*Gria3*)	синтез мРНК формирование димеров AMPA рецепторов из мономеров, формирование тетрамеров
FMR1 (Fmr1), CPEB1 (Cpeb1)	локальный синтез в дендрите
Sec12 (*Preb*), Sar1b (*sar1b*), Sec16a (*sec16a*), Sec23A (*sec23A*), Sec24A (*sec24A*), Sec13 (*sec13*), Sec31a (*sec31a*), p125A (*sec23ip*)	формирование транспортной везикулы на мембранах ЭР (COPII покрытые везикулы)
Arf1 (*Arf1*), GIT1 (*Git1*), BIG1 (*Arfgef1*), AP-1 complex (AP1G1 (*Ap1g1*), AP1B1 (*Ap1b1*), AP1M1 (*Ap1m1*), AP1S1 (*Ap1s1*)), clathrin (*Cltc*, *Clta*, *Cltb*)	формирование покрытых клатрином везикул на мембранах ТГС
CYFIP/NCKAP1 (CYFIP (*Cifip1*), NCKAP1 (*Nckap1*)), WAVE1/ABI2/BRK1 (WAVE1 (*Wasf1*), ABL2 (*Abi2*), BRICK (*Brk1*)), Arp2/3 complex (*Arpc2*, *Arpc3*), Rac1 (*rac1*), MEGAP (*Srgap3*), HIP1R (*Hip1r*)	отпочковывание везикулы от мембраны ТГС
KIF5A (*Kif5a*), Rab8 (*rab8a*), GRIP1, MYO5A (*Myo5a*) Stx4 (*Stx4*), Exo70 (*Exoc7*), NSF (*nsf*)	транспорт везикул из сомы к мембране и встраивание экзоцитозных везикул в ПМ
Сортировка АМПАР в дендрите	
RAB5A (*rab5a*), GRIP1, PICK1 (*Pick1*), RAB4 (*rab4a*), GRASP-1 (*Gripap1*), ARF6 (*Arf6*), ARNO (*Cyth3*), ACAP1 (*Acap1*), GULP1 (*Gulp1*), Rab11FIP2 effector complex (RAB11 (*Rab11A*), RAB11-FIP2 (*Rab11Fip2*)), MYO5b (*Myo5b*)	рециклинг рецепторов
NEEP21 (*nsg1*), ARF1, BIG2 (*ARFGEF2*), AGAP1 (*Agap1*), AP-3 complex, KIFC2 (*Kifc2*)	вывод рецепторов в путь деградации белков
Эндоцитоз АМПАР	
PIP5K1g (*Pip5k1c*), AP-2 complex (AP2A1 (*ap2a1*), AP2B1 (*ap2b1*), AP2M1 (*ap2m1*), AP2S1 (*ap2s1*), clathrin, EPS15, EPN1, HIP1 (*Hip1*), AP180, cortactin (*Cttn*)	формирование покрытой клатрином везикулы
N-WASP (*Wasl*), Endophilin 3 (*Sh3gl3*), DYN3 (*Dnm3*), Homer2 (*Homer2*)	отпочковывание везикулы от ПМ
MYO6 (*Myo6*)	транспорт в дендрит
Регулирование динамики АМПАР в СМ	
SAP97 (*Dlg1*), PSD95 (*Dlg4*), AKAP5 (*Akap5*), Moesin (*Msn*), I-CAM5 (*Icam5*), PKC alpha (*Prkca*), PICK1, PKAalpha (*Prkaca*), CaN (*Calnb*, *Calna*, *Ppp3r1*), CAMKII (*Camk2b*, *Camk2a*), PP1 (*Ppp1cc*, *Ppp1r9b*), spinophilin, nNOS (*Nos1*), cGK2 (*Prkg2*), NPRAP (*Ctnnd2*), N-cadherin (*Cdh2*), GRIP2 (*Grip2*), P4.1 (*Epb41*), SPT (*Sptan1*), F-actin, RGRF1 (*Rasgrf1*), SynGAP (*Syngap1*), RASH (*Hras1*)	вход АМПА рецепторов в перисинаптическую зону, закрепление в синаптической зоне и выход из нее
Усложнение подмембранной актиновой сети	
ARF6 (*Arf6*), EFA6A (*Psd*), BRAG (*Iqsec2*), GIT, beta PIX (*Arhgef7*), CaM (*Calm*), Kalrn7 (*Kalrn*), TIAM1 (*Tiam1*), RAC1 (*Rac1*)	регуляция динамики нитей актина в головке шипика
Масштабирование АМПАР на СМ	
Arc (*Arg3.1*)	регулирование возбуждения в локальной дендритной сети

* Курсивный шрифт – гены, прямой – белки. ПМ – плазматическая мембрана, СМ – синаптическая мембрана ПМ дендритного шипика.

В сети «Intermediate-LTP» впервые продемонстрировано, как структура дендритного шипика обеспечивает пространственно-временную динамику функциональных взаимодействий каскадов белков, отвечающих за процессы ДВП.

Мы реконструировали высокоупорядоченную микродоменную организацию шипиков как на горизонтальном (мембрана шипика – синаптическая, перисинаптическая, экстрасинаптическая), так и вертикальном уровнях (межбелковые сети сигнальных и структурных белков – постсинаптическое уплотнение (ПСУ)). Патология их морфологии связана с рядом психических расстройств (Kasai *et al*., 2010).

Показано, что каждый микродомен содержит свой набор белков, обеспечивающих специфические процессы в течение ДВП. Так, синаптическая мембрана (СМ) – постсинаптическая часть синапса, содержит кластеры ионотропных глутаматных рецепторов (Chen *et al*., 2008). Ее внутриклеточная поверхность связана с ПСУ. В составе ПСУ обнаружено около 620 белков, 9 из них связаны с АМРАР, более 450 – кластеризуются с НМДАР (Collins *et al*., 2006).

Размером ПСУ и ее стабильностью управляют скаффолдные (структурообразующие) белки, которые обеспечивают пространственную организацию функциональных взаимодействий других белков (Okabe, 2007; Sheng, Hoogenraad, 2007).

Микродомен перисинаптической мембраны (ПСМ) примыкает к СМ и выступает как зона динамического депо в процессах траффикинга глутаматных рецепторов (Tardin *et al*., 2003; Ashby *et al*., 2004).

Микродомен экстрасинаптической мембраны (ЭСМ) включает в себя сайты экзоцитоза, эндоцитоза и оставшуюся часть мембраны шипика. Зоны экзоцитоза и эндоцитоза являются независимыми компартментами мембраны, отличающимися специфическим набором белков (Petrini *et al*., 2009; Kennedy *et al*., 2010), участвующих в круговороте мембранных белков как при базовых условиях, когда синапс пластически неактивен (Gerges *et al*., 2004), так и после индукции ДВП синапса (Steiner *et al*., 2005; Opazo, Choquet, 2011).

Проведенный нами анализ регуляторных взаимодействий сети «Intermediate-LTP» позволил выделить ключевые процессы, лежащие в основе индукции и развития ДВП в СА1 поле гиппокампа после ВЧС (рис. 1).

В течение нескольких минут происходит запуск всей сети регуляторных каскадов, обеспечиваемых, как можно предположить,

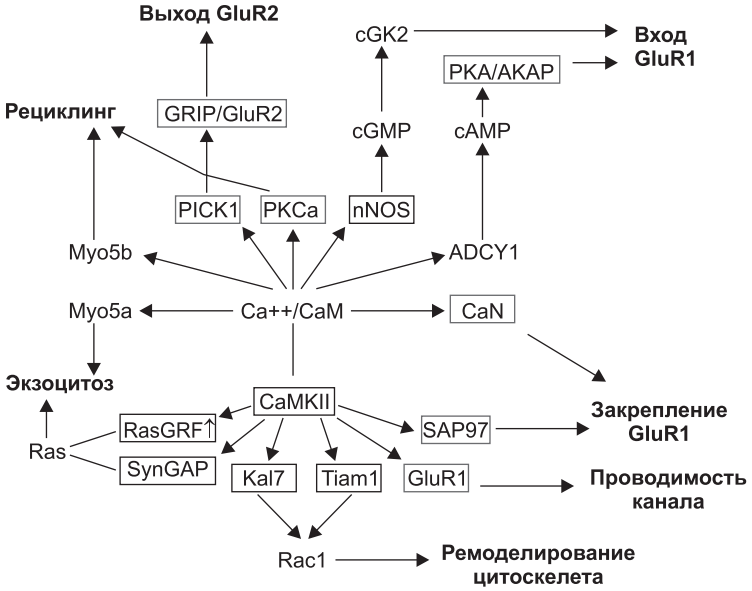


Рис. 1. Анализ сети «Intermediate LTP» базы данных GeneNet.

Кальций – основной регулятор процессов, обеспечивающих развитие ДВП. Шрифтом выделены основные процессы, протекающие в дендритном шипике в течение индукции и развития ДВП в СА1 поле гиппокампа. GluR1 – Glu1-содержащие АПМА рецепторы; GluR2 – Glu2/3 АМПА рецепторы.

предсуществующей до индукции ДВП синапса структурой белковых взаимодействий между микродоменами шипика (O'Connor *et al*., 2005; Patterson *et al*., 2010) (рис. 2).

Входящие через расположенные в центре СМ НМДА рецепторы ионы кальция (Chen *et al*., 2008; Raghuram *et al*., 2012) запускают каскады регуляторных взаимодействий, что обеспечивает в итоге быстрое увеличение эффективности синаптической передачи (рис. 1). Кальций запускает доставку, экзоцитоз, встраивание в ПСМ, латеральное перемещение в СМ (Shi *et al*., 1999; Passafaro *et al*., 2001; Correia *et al*., 2008; Nikandrova *et al*., 2010) и закрепление на ПСУ Glu1 АМПАР (Shen *et al*., 2000; Honkura *et al*., 2008), а также вывод Glu2/3 АМПАР из СМ в ПСМ и их эндоцитоз в составе эндосом в дендрит (Lu, Ziff, 2005; Eyster, 2007; Opazo, Choquet, 2011). Итогом является замена субъединичного состава АМПАР в СМ (рис. 2).

Считается, что этот процесс является ключевым для экспрессии ДВП в первые минуты после индукции (Newpher, Ehlers, 2008). Вероятно, это объясняет наблюдаемое, по крайней мере в ряде экспериментов, увеличение объема шипика (Hanse, Gustafsson, 1992). Высокочастотная стимуляция также сразу вызывает значительное усиление синаптических токов (Hanse, Gustafsson, 1992; O'Connor *et al*., 2005).

Геометрия шипика и быстрое связывание кальция кальмодулином (и иными кальций-связывающими белками) обеспечивают быстрое затухание волны возбуждения в каскадах киназ/фосфатаз (Raghuram *et al*., 2012). В связи с этим напрашивается вывод о важности пространственного позиционирования активируемых входящим кальцием молекул с их субстратами. Так, считается, что пространственное сближение источника синтеза цАМФ (активируемой кальцием циклазы) (Mons *et al*., 2003) и РКА (цАМФ-зависимой белковой киназы) играет ключевую роль для процесса развития ДВП в СА1 поле гиппокампа (Kim *et al*., 2011) (рис. 2).

«Intermediate-LTP» является концептуальной моделью, отражающей принцип формирования функциональной системы межбелковых взаимодействий между всеми микродоменами шипика после индукции ДВП.

Белковый набор микродоменов в течение базовой нейротрансмиссии конститутивно поддерживает баланс киназ и фосфатаз, малых ГТФаз и их регуляторов, липидный состав мембраны, плотность трансмембранных белков,

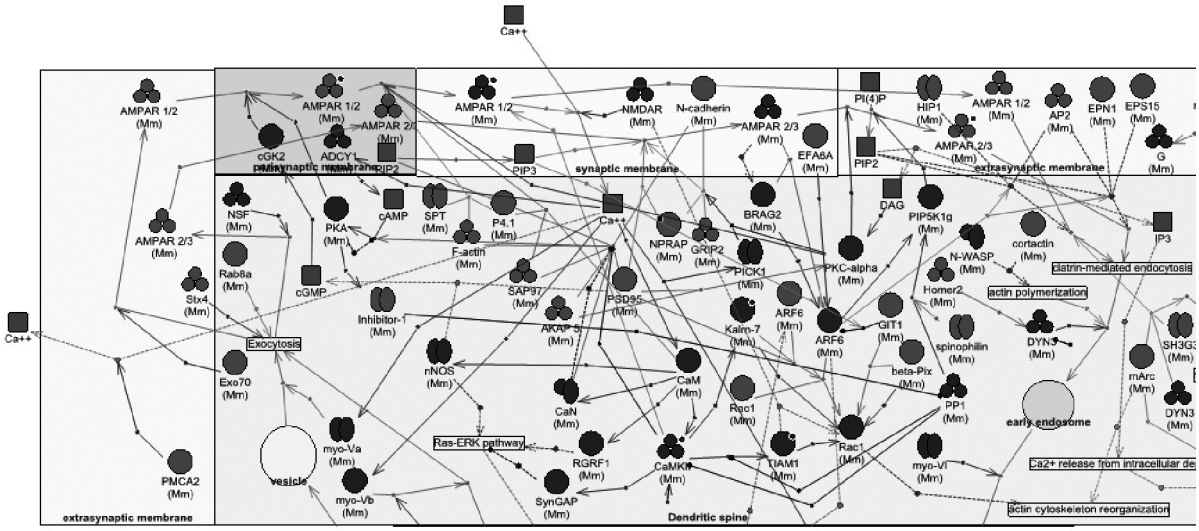


Рис. 2. Фрагмент сети «Intermediate LTP». Основные межмолекулярные регуляторные взаимодействия после ВЧС.

Synaptic membrane – синаптическая мембрана; perisynaptic membrane – перисинаптическая мембрана; extrasynaptic membrane – внесинаптическая мембрана (зона эндоцитоза). ■ – низкомолекулярные вещества; ● – мономер; ●● – димер; ●●● – мультимерный белок; стрелки – взаимодействия белков.

Белки и реакции описаны в формате базы данных GeneNet (http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/AMPA.html).

в том числе и Glu2/3-АМПАР, ответственных за деполяризацию синаптической мембраны, необходимую для активирования НМДАР при индукции ДВП (Opazo, Choquet, 2011). Все это обеспечивает, как можно предположить, предподготовленное состояние дендритных шипиков для восприятия приходящего на синаптическую мембрану информационного сигнала и его первичную обработку.

В течение экспрессии ДВП запускаются регуляторные взаимодействия, приводящие к нарушению существующих до индукции межмолекулярных взаимодействий. Ключевым событием, по нашему мнению, при этом является нарушение существовавшей до индукции цитоархитектуры шипика – происходят разрывы и появление открытых концов актиновых нитей (Ф-актин, рис. 3) и формируется пул глобулярного актина (G-актин, рис. 3) (Sellers, 2000). Таким образом, складываются условия для формирования новой функциональной сети головки шипика адекватно пришедшему на синапс сигналу (рис. 3).

Запускаются контуры обратной негативной регуляции, препятствующие избыточному увеличению синаптической эффективности. Например, через активирование фосфатаз кальцинейрина (Jouvenceau *et al*., 2003; Li *et al*., 2011) и PP1 (белковая фосфатаза 1) (Genoux *et al*., 2002) (рис. 2).

По мере развития ДВП запускаются процессы, обеспечивающие замену встроенных в первые минуты после индукции Glu1 АМПАР на Glu2 АМПАР. Так, известно, что уже через 20 минут после индукции ДВП происходит замена Glu1 АМПАР на Glu2 АМПАР (Plant *et al*., 2006) (рис. 2).

Работа каскадов малых ГТФаз обеспечивает элонгацию и ветвление актиновых нитей (Wegner *et al*., 2008) (рис. 3). Таким образом, восстанавливается целостность цитоархитектуры шипика, что, как мы предполагаем, работает на каркасное закрепление новой, сформированной в результате активности синапса, структуры межбелковых взаимодействий шипика – пространственного позиционирования молекул-партнеров, определяющих эффективность синаптической передачи, которая далее поддерживается в течение продолжительного времени. Поддержание новой структуры шипика и нового уровня синаптической эффективности тесно связано с процессами синтеза и созревания новых белков в соме нейрона (Малахин и др., 2012).

В «Intermediate-LTP» показано, что формирование новой межбелковой сети шипика контролируется на уровне всего дендритного ответвления через регуляцию процессов локального синтеза и синаптического масштабирования.

Локальный синтез в дендрите поддерживает необходимый для шипиков состав белков (Liu

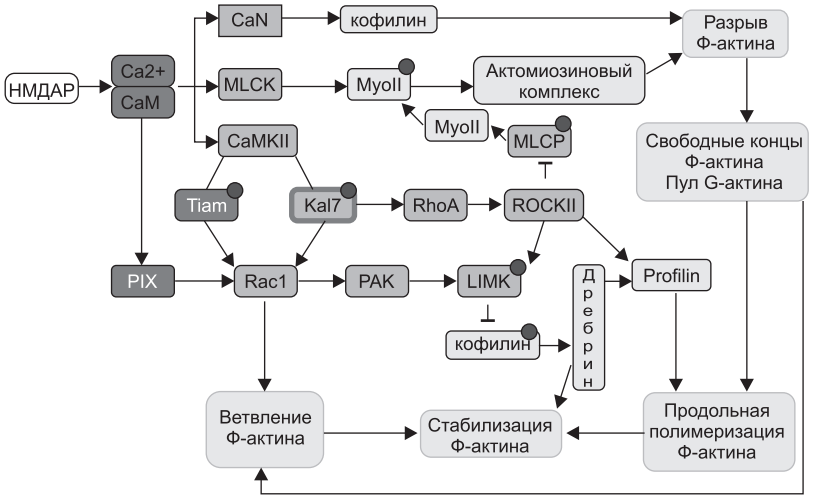


Рис. 3. Схематичное представление молекулярных механизмов, транслирующих нейротрансмиссию в изменение цитоархитектуры дендритного шипика в течение индукции и экспрессии ДВП.

⊣ – выключение; → – активирование белка/запуск процесса; ● – фосфорилирование.

et al., 2006; Verpelli *et al*., 2010). Субъединицы Glu1 и Glu2 локально синтезируются после активирования НМДАР и поступают в шипик, поддерживая кругооборот АМПАР в течение развития ДВП (Martin, Zukin, 2006).

Известно, что в диапазоне 10–20 мин после индукции ДВП происходит процесс масштабирования на СМ АМПАР – ускорение их эндоцитоза (Peebles *et al.*, 2010). Главным регулятором является локально синтезируемый после индукции ДВП белок Arc (activity regulated protein) (Messaoudi *et al*., 2007), который заходит во все активные в этот момент шипики дендрита. Чем выше уровень эффективности синапса в шипике, тем больше молекул Arc заходит в него и тем эффективнее ускоряются вывод АМПАР из СМ и их уход в дендрит. Таким образом, выравнивается итоговый уровень синаптической эффективности во всей группе активных шипиков дендрита (Peebles *et al*., 2010). Дефекты гена *Arc* приводят к бифазным изменениям гиппокампальной ДВП в поле СА1 с повышением ранней и отсутствием поздней фаз ДВП (Plath *et al*., 2006).

Имеются данные, что после НМДАР-зависимой индукции отдельного шипика нейронов поля СА1 гиппокампа наблюдаются выход ряда активных молекул семейства малых ГТФаз и их заход в соседние шипики дендрита. Считается, что данные процессы могут способствовать усилению эффективности их синаптической передачи, а также понижению порога возбуждения неактивных близко расположенных шипиков (Harvey *et al*., 2008; Patterson *et al*., 2010; Murakoshi *et al*., 2011).

ЗАКЛЮЧЕНИЕ

Активирование синапса запускает функциональные межбелковые взаимодействия микродоменов дендритного шипика, что обеспечивает прием сигнала и быстрое увеличение эффективности синаптической передачи в минутном интервале после индукции ДВП. Структурная пластичность шипика тесно связана с функциональной пластичностью синапса, что опосредует формирование новой межбелковой сети взаимодействий адекватно пришедшему сигналу. Поддержание нового уровня синаптической эффективности контролируется процессами возбуждения/торможения в локальной сети дендрита.

Работа выполнена при поддержке базового проекта фундаментальных исследований РАН № 35.1.5; гранта РФФИ 12-01-00639-а, интеграционного проекта СО РАН № 136; Минобрнауки РФ (соглашение 8740).

ЛИТЕРАТУРА

Малахин И.А., Проскура А.Л., Запара Т.А., Ратушняк А.С. Влияние сборки транспортных везикул на процессы сохранения эффективности синаптической передачи // Вестн. НГУ. 2012. Т. 10. № 4. С. 14–20.

Сергеев П.В., Шимановский Н.Л., Петров В.И. Рецепторы физиологически активных веществ: Монография. Волгоград: Семь ветров, 1999. 640 с.

Ananko E.A., Podkolodny N.L., Stepanenko I.L. *et al.* GeneNet in 2005 // Nucl. Acids Res. 2005. V. 33. P. 425–427.

Ashby M.C., De La Rue S.A., Ralph G.S. *et al*. Removal of AMPA receptors (AMPARs) from synapses is preceded by transient endocytosis of extrasynaptic AMPARs AMPARs // J. Neurosci. 2004. V. 24. No. 22. P. 5172–5176.

Bliss T.V., Collingridge G.L. A synaptic model of memory: long-term potentiation in the hippocampus // Nature. 1993. V. 361. No. 6407. P. 31–39.

Chen X., Winters C., Azzam R. *et al.* Organization of the core structure of the postsynaptic density // Proc. Natl Acad. Sci. USA. 2008. V. 105. No. 11. P. 4453–4458.

Collins M.O., Husi H., Yu L., Brandon J.M. *et al.* Molecular characterization and comparison of the components and multiprotein complexes in the postsynaptic proteome // J. Neurochem. 2006. V. 97. P. 16–23.

Correia S.S., Bassani S., Brown T.C. *et al*. Motor protein-dependent transport of AMPA receptors into spines during long-term potentiation // Nat. Neurosci. 2008. V. 11. No. 4. P. 457–466.

Eyster K.M. The membrane and lipids as integral participants in signal transduction: lipid signal transduction for the non-lipid biochemist // Adv. Physiol. Educ. 2007. V. 31. No. 1. P. 5–16.

Genoux D., Haditsch U., Knobloch M. *et al*. Protein phosphatase 1 is a molecular constraint on learning and memory // Nature. 2002. V. 418. No. 6901. P. 970–975.

Gerges N.Z., Tran I.C., Backos D.S. *et al.* Independent functions of hsp90 in neurotransmitter release and in the continuous synaptic cycling of AMPA receptors // J. Neurosci. 2004. V. 24. No. 20. P. 4758–66.

Hanse E., Gustafsson B. Postsynaptic, but not presynaptic, activity controls the early time course of long-term potentiation in the dentate gyrus // J. Neurosci. 1992. V. 12. No. 8. P. 3226–3240.

Harvey C.D., Yasuda R., Zhong H., Svoboda K. The spread of Ras activity triggered by activation of a single dendritic spine // Science. 2008. V. 321. No. 5885. P. 136–140.

Hawley D.F., Morch K., Christie B.R., Leasure J.L. Differential response of hippocampal subregions to stress and

learning // PLoS One. 2012 V. 7. No. 12. P. E53126.
Honkura N., Matsuzaki M., Noguchi J. *et al*. The subspine organization of actin fibers regulates the structure and plasticity of dendritic spines // Neuron. 2008. V. 57. No. 5. P. 719–729.
Hotulainen P., Hoogenraad C.C. Actin in dendritic spines: connecting dynamics to function // J. Cell. Biol. 2010. V. 189. P. 619–629.
Jouvenceau A., Billard J.M., Haditsch U. *et al*. Different phosphatase-dependent mechanisms mediate long-term depression and depotentiation of long-term potentiation in mouse hippocampal CA1 area // Eur. J. Neurosci. 2003. V. 18. No. 5. P. 1279–1285.
Kasai H., Fukuda M., Watanabe S. *et al.* Structural dynamics of dendritic spines in memory and cognition // Trends. Neurosci. 2010. V. 33. No. 3. P. 121–129.
Kennedy M.J., Davison I.G., Robinson C.G., Ehlers M.D. Syntaxin-4 defines a domain for activity-dependent exocytosis in dendritic spines // Cell. 2010. V. 141. 3. P. 524–535.
Kim M., Park A.J., Havekes R. *et al*. Colocalization of protein kinase A with adenylyl cyclase enhances protein kinase A activity during induction of long-lasting long-term-potentiation // PloS. Comput. Biol. 2011. V. 7. No. 6. P. E1002084.
Kjelstrup K.B., Solstad T., Brun V.H. *et al*. Finite scale of spatial representation in the hippocampus // Science. 2008. V. 321. No. 5885. P. 140–143.
Li H., Rao A., Hogan P.G. Interaction of calcineurin with substrates and targeting proteins // Trends. Cell. Biol. 2011. V. 21. No. 2. P. 91–103.
Liu S.H., Cheng H.H., Huang S.Y. *et al.* Studying the protein organization of the postsynaptic density by a novel solid phase- and chemical cross-linking-based technology // Mol. Cell. Proteomics. 2006. V. 5. No. 6. P. 1019–1032.
Lu W., Ziff E.B. PICK1 interacts with ABP/GRIP to regulate AMPA receptor trafficking // Neuron. 2005. V. 47. No. 3. P. 407–421.
Martin K.C., Zukin R.S. RNA trafficking and local protein synthesis in dendrites: an overview // J. Neurosci. 2006. V. 26. No. 27. P. 7131–7134.
Messaoudi E., Kanhema T., Soulé J. *et al.* Sustained Arc/Arg3.1 synthesis controls long-term potentiation consolidation through regulation of local actin polymerization in the dentate gyrus *in vivo* // J. Neurosci. 2007. V. 27. No. 39. P. 10445–10455.
Mons N., Guillou J.L., Decorte L., Jaffard R. Spatial learning induces differential changes in calcium/calmodulin-stimulated (ACI) and calcium-insensitive (ACII) adenylyl cyclases in the mouse hippocampus // Neurobiol. Learn. Mem. 2003. 79. No. 3. P. 226–235.
Murakoshi H., Wang H., Yasuda R. Local, persistent activation of Rho GTPases during plasticity of single dendritic spines // Nature. 2011. V. 472. No. 7341. P. 100–104.
Nagasawa M., Sakimura K., Mori K.J. *et al.* Gene structure and chromosomal localization of the mouse NMDA receptor channel subunits // Brain Res. Mol. Brain. Res. 1996. V. 36. No. 1. P. 1–11.
Newpher T.M., Ehlers M.D. Glutamate receptor dynamics in dendritic microdomains // Neuron. 2008. V. 58. No. 4. P. 472–497.
Nikandrova Y.A., Jiao Y., Baucum A.J. *et al*. Ca2+/calmodulin-dependent protein kinase II binds to and phosphorylates a specific SAP97 splice variant to disrupt association with AKAP79/150 and modulate alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid-type glutamate receptor (AMPAR) activity // J. Biol. Chem. 2010. V. 285. No. 2. P. 923–934.
Nowak L., Bregestovski P., Ascher P. *et al.* Magnesium gates glutamate-activated channels in mouse central neurones // Nature. 1984. V. 307. No. 5950. P. 462–465.
O'Connor D.H., Wittenberg G.M., Wang S.S-H. Graded bidirectional synaptic plasticity is composed of switch-like unitary events // Proc. Natl Acad. Sci. USA. 2005. V. 102. P. 9679–9684.
Okabe S. Molecular anatomy of the postsynaptic density // Mol. Cell. Neurosci. 2007. V. 34. P. 503–518.
Opazo P., Choquet D. A three-step model for the synaptic recruitment of AMPA receptors // Mol. Cell. Neurosci. 2011. V. 46. No. 1. P. 1–8.
Palmer C.L., Cotton L., Henley J.M. The molecular pharmacology and cell biology of alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptors// Pharmacol. Rev. 2005. V. 57. No. 2. P. 253–277.
Passafaro M., Piëch V., Sheng M. Subunit-specific temporal and spatial patterns of AMPA receptor exocytosis in hippocampal neurons // Nat. Neurosci. 2001. V. 4. No. 9. P. 917–926.
Patterson M.A., Szatmari E.M., Yasuda R. AMPA receptors are exocytosed in stimulated spines and adjacent dendrites in a Ras-ERK-dependent manner during long-term potentiation // Proc. Natl Acad. Sci. USA. 2010. V. 107. No. 36. P. 15951–15956.
Peebles C.L., Yoo J., Thwin M.T. *et al*. Arc regulates spine morphology and maintains network stability *in vivo* // Proc. Natl Acad. Sci. USA. 2010. V. 107. No. 42. P. 18173–18178.
Petersen C.C., Malenka R.C., Nicoll R.A., Hopfield J.J. All-or-none potentiation at CA3-CA1 synapses // Proc. Natl Acad. Sci. USA. 1998. V. 95. No. 8. P. 4732–4737.
Petrini E.M., Lu J., Cognet L. *et al.* Endocytic trafficking and recycling maintain a pool of mobile surface AMPA receptors required for synaptic potentiation // Neuron. 2009. V. 63. No. 1. P. 92–105.
Plant K., Pelkey K.A., Bortolotto Z.A. *et al*. Transient incorporation of native GluR2-lacking AMPA receptors during hippocampal long-term potentiation // Nat. Neurosci. 2006. V. 9. No. 5. P. 602–604.
Plath N., Ohana O., Dammermann B. *et al.* Arc/Arg3.1 is essential for the consolidation of synaptic plasticity and memories // Neuron. 2006. V. 52. No. 3. P. 437–444.
Raghuram V., Sharma Y., Kreutz M.R. Ca^{2+} sensor proteins in dendritic spines: a race for Ca^{2+} // Front. Mol. Neurosci. 2012. V. 5. P. 61.
Sellers J.R. Myosins: a diverse superfamily // Biochim. Biophys. Acta. 2000. V. 1496. P. 3–22.
Shen L., Liang F., Walensky L.D., Huganir R.L. Regulation of AMPA receptor GluR1 subunit surface expression by a 4. 1N-linked actin cytoskeletal association // J. Neurosci. 2000. V. 20. No. 21. P. 7932–7940.
Sheng M., Hoogenraad C.C. The postsynaptic architecture of excitatory synapses: a more quantitative view // Annu.

Rev. Biochem. 2007. V. 76. P. 823–847.
Shepherd J.D., Huganir R.L. The cell biology of synaptic plasticity: AMPA receptor trafficking // Annu. Rev. Cell. Dev. Biol. 2007. V. 23. P. 613–643.
Shi S.H., Hayashi Y., Petralia R.S. *et al*. Rapid spine delivery and redistribution of AMPA receptors after synaptic NMDA receptor activation // Science. 1999. V. 284. No. 5421. P. 1811–18116.
Steiner P., Alberi S., Kulangara K. *et al*. Interactions between NEEP21, GRIP1 and GluR2 regulate sorting and recycling of the glutamate receptor subunit GluR2 // EMBO J. 2005. V. 24. No. 16. P. 2873–2884.
Steward O., Schuman E.M. Protein synthesis at synaptic sites on dendrites // Annu. Rev. Neurosci. 2001. V. 24. P. 299–325.
Szirmai I., Buzsáki G., Kamondi A. 120 years of hippocampal Schaffer collaterals // Hippocampus. 2012. V. 22. No. 7. P. 1508–1516.
Tardin C., Cognet L., Bats C. *et al*. Direct imaging of lateral movements of AMPA receptors inside synapses // EMBO J. 2003. V. 22. No. 18. P. 4656–4665.
Tsien J.Z., Huerta P.T., Tonegawa S. The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory // Cell. 1996. V. 87. No. 7. P. 1327–1338.
Tsunoda S., Sierralta J., Sun Y. *et al*. A multivalent PDZ-domain protein assembles signaling complexes in a G-protein-coupled cascade // Nature. 1997. V. 388. P. 243–249.
Verpelli C., Piccoli G., Zanchi A. *et al*. Synaptic activity controls dendritic spine morphology by modulating eEF2-dependent BDNF synthesis // J. Neurosci. 2010. V. 30. No. 17. P. 5830–5842.
Wegner A.M., Nebhan C.A., Hu L. *et al*. N-Wasp and the arp2/3 complex are critical regulators of actin in the development of dendritic spines and synapses // J. Biol. Chem. 2008. V. 283. P. 15912–15920.

INTERMOLECULAR INTERACTIONS IN NEURONAL FUNCTIONAL SYSTEMS

A.L. Proskura[1], I.A. Malachin[1], T.A. Zapara[1], I.I. Turnaev[2], V.V. Suslov[2], A.S. Ratuschnyak[1]

[1] Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia;
[2] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: zapara_t@mail.ru

Summary

Neuronal synaptic contacts are among the basic elements that determine the plasticity of the nervous system. Changes in the efficiency of synaptic transmission mediate sensation, conduction of excitation, learning, and memory. Dendritic spines are the postsynaptic part of excitatory synapses in higher divisions of mammalian brains. Protein–protein networks of spine microdomains form the functional system of neuronal synapses. Reconstruction of the conceptual model of molecular interactions has been performed. The model represents activity-dependent changes of the synaptic transmission efficiency, integration of excitation in the local dendritic network of a neuron, and prolonged maintenance of the new level of neurotransmission.

Key words: synaptic plasticity, long-term potentiation, glutamate receptors.

УДК 577.21:577.29:004.42

КОМПЬЮТЕРНЫЙ АНАЛИЗ ДАННЫХ ЭКСПРЕССИИ ГЕНОВ В КЛЕТКАХ МОЗГА, ПОЛУЧЕННЫХ С ПОМОЩЬЮ МИКРОЧИПОВ И ВЫСОКОПРОИЗВОДИТЕЛЬНОГО СЕКВЕНИРОВАНИЯ

© 2013 г. **И.В. Медведева, О.В. Вишневский, Н.С. Сафронова, О.С. Кожевникова, М.А. Генаев, Д.А. Афонников, А.В. Кочетов, Ю.Л. Орлов**

Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, e-mail: orlov@bionet.nsc.ru

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

В последние годы происходит стремительное расширение фронта нейробиологических исследований, сопровождающееся бурным ростом объема экспериментальных данных по структуре, функции и эволюции нервной системы на различных уровнях ее иерархической организации. Использование технологий высокопроизводительного секвенирования и микрочипов позволяет проводить сравнительный статистический анализ экспрессии тысяч генов одновременно, учитывая при этом пространственное расположение клеток в структурах мозга. Дан краткий обзор основных подходов анализа экспрессии генов в клетках мозга. Проанализированы особенности структуры генов, имеющих дифференциальную экспрессию в клетках мозга. Оценивалось число экзонов, альтернативных транскриптов и его соотношение с уровнем экспрессии. Показано статистическое различие числа альтернативных транскриптов для генов, активных в структурах головного мозга и других органов. Найдены гены, экспрессия которых повышена в структурах мозга и связана с нейродегенеративными заболеваниями.

Ключевые слова: биоинформатика, мозг, экспрессия генов, микрочипы, секвенирование.

ВВЕДЕНИЕ

Компьютерный анализ молекулярных механизмов деятельности высшей нервной системы имеет огромное фундаментальное значение для биологии, генетики и для исследования процессов познания. При этом компьютерные методы используются для изучения структуры генов, их взаимосвязи, координации их работы (экспрессии) в клетках мозга (Орлов и др., 2013). Использование технологий высокопроизводительного секвенирования и анализа данных экспрессии генов на микрочипах позволяет ставить задачи исследования на качественно более высоком уровне (Xie *et al*., 2011; Lipovich *et al*., 2012). Целью данной работы были обзор основных биоинформационных подходов и баз данных по анализу экспрессии генов в клетках мозга и статистический анализ распределения параметров структуры гена по отношению к его экспрессии в структурах мозга при использовании разработанных ранее биоинформационных методов.

Одним из крупнейших достижений в области нейронаук является создание трехмерных атласов экспрессии генов в мозге мыши и мозге человека (рис. 1), среди которых наиболее детальным ресурсом является компьютерная база данных Allen Brain Atlas (Lein *et al*., 2007; Hawrylycz *et al*., 2012).

В настоящее время разработана серия баз данных по экспрессии генов в мозге: GENSAT (Gene Expression Nervous System Atlas) (http://www.gensat.org), MGI (Mouse Genome Informatics) (http://www.informatics.jax.org/), BGEM (Brain Gene Expression Map) (http://www.stjude-bgem.org). Разрабатываются и базы данных, описывающие активность структур мозга, ос-

Рис. 1. Схема обработки данных по построению трехмерных карт экспрессии (левая панель) и пример поиска гена по имени (ген TYMS) с визуализацией экспрессии генов на микрочипах в структурах мозга в Allen Brain Atlas (http://connectivity.brain-map.org/projection) (правая панель).

нованные на методах магнитного резонанса и томографии. Методы электроэнцефалографии, топографического картирования электрической активности мозга и компьютерной томографии связаны с общим измерением активности структур мозга на более высоком уровне и выходят за рамки настоящей работы.

Опыт биоинформационных исследований в области анализа молекулярных механизмов регуляции экспрессии генов, использования технологий секвенирования и микрочипов, в том числе относящихся к экспрессии генов в тканях мозга, изучения роли серотониновой системы мозга в регуляции поведения, накопленный в ИЦиГ СО РАН (Ananko *et al.*, 2005; Витяев и др., 2001; Demenkov *et al.*, 2011; Naumenko *et al.*, 2011), используется при обработке новых полногеномных данных, полученных с помощью высокопроизводительных транскриптомных технологий.

Экспрессия генов в клетках зависит от внешних стимулов и внутренней генетической программы клеток (нейроны, клетки структур мозга, клетки крови, клетки внутренних органов). Регуляция проявления функции генов внутри клетки осуществляется на уровне транскрипции и трансляции.

Контроль экспрессии генов на уровне трансляции мРНК важен как для развития и морфогенеза нейронов (Jung *et al.*, 2011), так и для функционирования специфических генных сетей в зрелых клетках различных разделов мозга. Нейроны характеризуются высокой степенью компартментализации (аксоны, дендриты, синапсы), при этом отдельные части клеток могут быть удалены на очень большое расстояние от ядра. Часть трансляционного аппарата локализована в удаленных районах клетки (вблизи синапсов) и мРНК генов, специфически задействованных в контроле передачи нервного импульса, транспортируется в эти районы (Liu-Yesucevitz *et al.*, 2011). Известны некоторые сигналы, локализованные в мРНК и опосредующие такой транспорт (Willis, Twiss, 2010; Wei, 2011).

Считается, что контроль экспрессии генов на уровне трансляционной активности специфических мРНК может иметь отношение к механизмам высшей нервной деятельности, такой, как физическая основа процесса запоминания (Darnell, 2011; Sidrauski *et al.*, 2013). Связь между трансляцией мРНК и процессами высшей нервной деятельности вызывает очень большой интерес в последние годы. Трансляция мРНК связана с механизмами, близкими к универсальным механизмам стрессового контроля экспрессии генов (фосфорилирование eIF2a, mTOR, eIF4E-BP и т. д.) (Gerashchenko

et al., 2012), адаптированными для решения специфических для клеток мозга задач (Sun *et al*., 2013).

Трансляционный контроль используется как средство регуляции экспрессии генов у многих видов. Посттранскрипционные регуляторные механизмы играют важную роль в метаболических путях стрессового ответа и могут вести к нарушению физиологических функций при нарушении таких механизмов мутациями (Lohse *et al*., 2011). Определенную роль в специфическом посттранскрипционном контроле экспрессии генов в клетках мозга также может играть цитоплазматическое полиаденилирование (Kundel *et al*., 2009). Развитие методов высокопроизводительного секвенирования (Ribo-Seq) и протеомики (Menschaert *et al*., 2013) существенно расширяет имеющиеся возможности для выявления молекулярных механизмов функционирования нейронов и мозга в целом.

Использование компьютерных технологий, таких, как GeneNet (Ananko *et al*., 2005) и AndVisio (Demenkov *et al*., 2011), позволяет реконструировать на основе данных научных публикаций генные сети – ансамбли координированно функционирующих генов, контролирующих биохимические, молекулярно-генетические, физиологические процессы. С помощью GeneNet реконструированы генные сети, контролирующие различные системы и процессы, в том числе в мозговых тканях, включая генную сеть «Early long-term potentiation», отражающую белковые взаимодействия в дендритных шипиках зоны CA1 гиппокампа.

Конкретной целью нашей работы было выявление особенностей генов, активно экспрессирующихся в тканях мозга, с помощью комплексного компьютерного исследования, включая разработанные ранее авторским коллективом методы и программы, такие, как база данных качества проб микрочипов Affymetrix U133 (Orlov *et al*., 2007), программный комплекс ICGenomics для функциональной аннотации генов (Орлов и др., 2012). Были подготовлены выборки генов, экспрессия которых повышена как в целом в структурах мозга, так и в отдельных районах головного мозга человека с использованием базы данных Allen Brain Atlas и BioGPS (Su *et al*., 2009; Wu *et al*., 2009). Из банка данных UCSC (genome.ucsc.edu) были загружены данные нуклеотидных последовательностей, содержащих эти гены и их регуляторные районы. Исследовались встречаемость открытых рамок считывания (ОРС) в промоторных последовательностях, присутствие сигналов трансляции. Выполнен анализ контекстных особенностей нуклеотидных последовательностей таких генов, присутствие коротких некодирующих РНК, в том числе в противоположной ориентации. Для выборок генов, дифференциально экспрессирующихся в различных органах, оценивались число экзонов и его соотношение с уровнем экспрессии. Исследованы данные по экспрессии генов, консервативных для мыши, крысы и человека, на микрочипах; рассмотрены данные секвенирования полных транскриптом.

ИССЛЕДОВАНИЕ ЭКСПРЕССИИ ГЕНОВ В КЛЕТКАХ ГОЛОВНОГО МОЗГА

Экспрессия генов, т. е. проявление их функции в клетках мозга, является базисом работы нейрона. Обычно безотносительно к типу клеток активно транскрибируются в клетке не все гены одновременно, а какая-то небольшая их часть, около 5 %, что позволяет выделить органоспецифичные группы генов. Для поиска генов, экспрессия которых специфична для структур головного мозга, мы использовали базу данных Allen Brain Atlas и базу BioGPS (Wu *et al*., 2009), содержащую данные генной экспрессии в широком круге тканей и органов. Экспрессия генов определялась на микрочипах Affymetrix U133 при использовании фильтрации по качеству (Orlov *et al*., 2007).

Среди проб микрочипа Affymetrix U133, представленных в базе данных BioGPS, были выделены пробы с высокой экспрессией (по ранговым значениям проб всех генов, верхние 1 %) и гены, экспрессия которых представлена в структурах мозга (гипоталамус, префронтальный кортекс и др., всего 12 видов структур), но не в других органах (почки, печень, гладкие мышцы и т. д.). Всего после удаления дублирующих проб микрочипа было отобрано 11830 имен (уникальных идентификаторов) генов. Пример распределения экспрессии генов по органам дан на рис. 2.

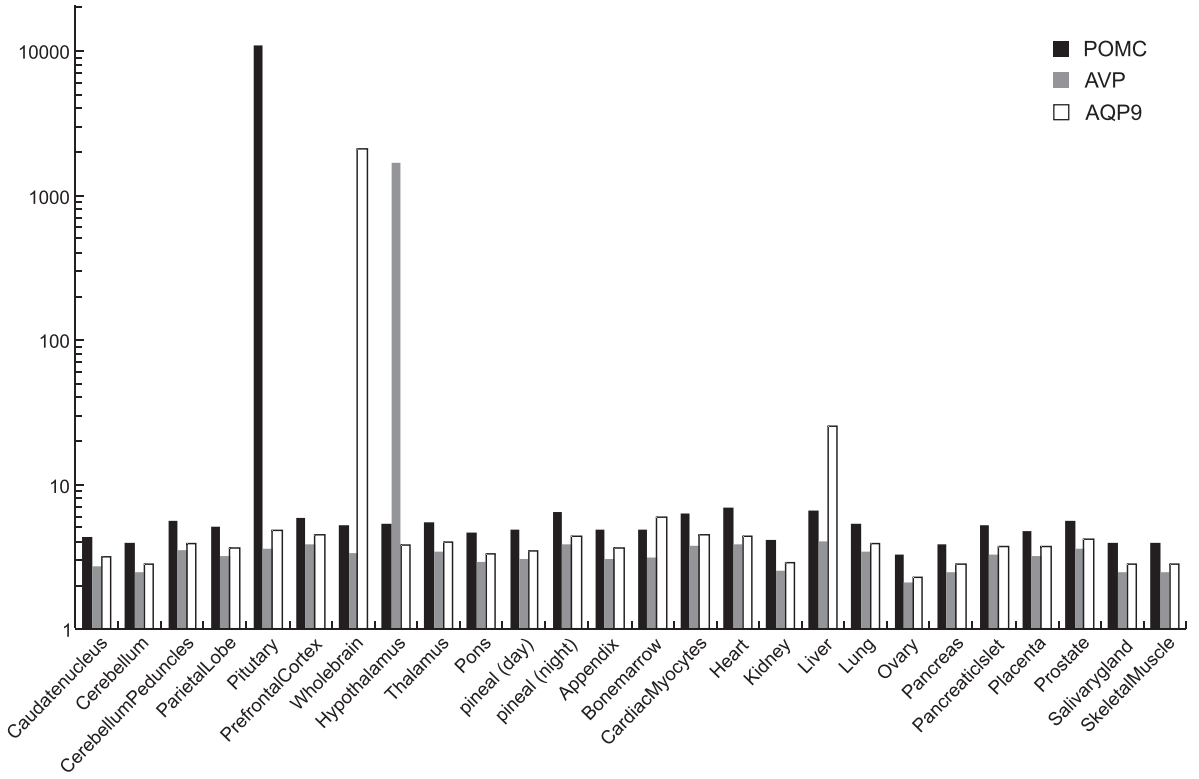


Рис. 2. Распределение экспрессии трех генов, наиболее высокоэкспрессирующихся в структурах головного мозга и других тканях.

По оси Y – уровень экспрессии на микрочипе Affymetrix U133 (база данных BioGPS). По оси X – исследованные ткани. Первые 12 групп слева соответствуют структурам головного мозга.

Из них 1801 ген (15 %) показывали высокую экспрессию хотя бы в одной из структур мозга. Большинство оставшихся генов (9253) не показывали значимую экспрессию (верхние 1 %) ни в одной из структур. Отметим, что использовались только белок-кодирующие гены, представленные на микрочипе.

Далее из выборки генов, входящих в верхние 1 % (1801), были выделены гены, показывающие высокую экспрессию в исследованных структурах мозга в среднем по сравнению со всеми остальными органами. Для полученных списков генов были проанализированы их геномное окружение, контекстная структура регуляторных районов, перекрывание с микроРНК и короткими некодирующими транскриптами в противоположной ориентации, число экзонов, эволюционная консервативность. Функциональная аннотация 1382 генов с высокой экспрессией только в структурах мозга выполнялась с помощью программы анализа генных онтологий DAVID (http://david.abcc.ncifcrf.gov). Результаты представлены в таблице.

Интересно отметить наличие категорий белкового транспорта, фосфопротеинов, нуклеотидного связывания, но не транскрипции. Присутствуют категории передачи нервного импульса, развития нейронов, что ожидаемо для структур головного мозга. Почти половина (45,7 %) генов из списка связана с альтернативным сплайсингом.

АНАЛИЗ ЧИСЛА ЭКЗОНОВ И ПЛАСТИЧНОСТИ ЭКСПРЕССИИ ГЕНОВ

На основе ранее полученных выборок генов мы проанализировали число экзонов, приходящихся на группы генов с высокой экспрессией в структурах головного мозга, с повышенной экспрессией в других органах и группу оставшихся генов на микрочипе Affymetrix U133, экспрессия которых не была статистически значима ни в одном из исследованных органов, по данным BioGPS, представленным в предыдущем разделе. Гистограммы распределения по числу экзонов

Таблица

Функциональная аннотация генов с высокой экспрессией в клетках мозга

Описание белков и их функций	Процент генов	Значимость, (P-value)	Коррекция Бонферрони
Фосфопротеины	56,4	5,5E−50	3,5E−47
Альтернативный сплайсинг	45,7	2,7E−08	1,7E−05
Ацетилирование	24,6	2,2E−30	1,4E−27
Связывание нуклеотидов	14,5	1,4E−13	8,7E−11
Метаболизм фосфатов	9,9	1,1E−08	3,7E−05
Межклеточные сигналы	8,0	1,2E−14	4,1E−11
Передача нервного импульса	7,3	5,4E−28	1,9E−24
Киназы	7,3	6,8E−12	4,3E−09
Структурные молекулы	7,2	7,0E−09	8,2E−06
Рибонуклеиновый комплекс	6,5	8,5E−12	4,9E−09
Синаптическая передача	6,4	6,1E−26	2,2E−22
Развитие нейронов	6,3	1,6E−20	5,9E−17

и транскриптов представлены на рис. 3. Количество экзонов и транскриптов было подсчитано согласно базе данных Ensembl (http://www.ensembl.org). Таким образом, конечная выборка для обработки составила 8219 генов.

Сравнение последовательностей генов, высокоэкспрессирующихся в структурах мозга, по количеству экзонов в последовательности показало, что такие последовательности содержат меньшее количество экзонов (среднее 12, $t = 4{,}5$; $d.f. = 8210$; $p = 10^{-6}$), чем другие (среднее 13,5). Также показано, что для таких генов экспрессируется меньшее количество различных транскриптов ($t = 9{,}1$; $d.f. = 8176$; $p << 0{,}001$). Среднее количество транскриптов, приходящееся на высокоэкспрессирующиеся гены в структурах мозга – 4,6, в то время как для низкоэкспрессирующихся генов – 5,7. Это согласуется с данными о том, что высокоэкспрессирующиеся гены, особенно экспрессирующиеся в различных тканях, обладают большей компактностью (Woody, Shoemaker, 2011; Park *et al*., 2012).

Гены, имеющие высокий уровень экспрессии в широком круге органов, имеют высокий уровень экспрессии и в структурах головного мозга. На основе данных об экспрессии генов в различных органах был проведен следующий сравнительный анализ. Для каждого гена опре-

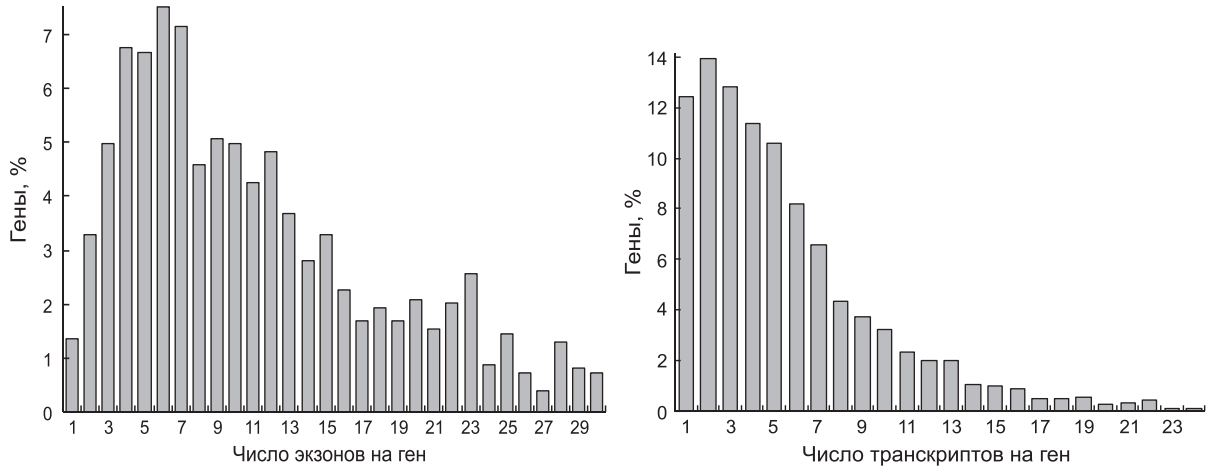


Рис. 3. Распределение последовательностей генов, высокоэкспрессирующихся в структурах мозга, по количеству экзонов в последовательности (слева) и количеству транскриптов (справа).

делялись наиболее высокие значения, лежащие вне доверительного интервала 99 %. Если они обнаруживались хотя бы для одной ткани головного мозга, то такие гены группировались. Всего в эту группу вошло 55 генов, и они имеют повышенную экспрессию во всех изучаемых тканях.

Оказалось, что по сравнению с другими генами количество транскриптов, соответствующих генам с повышенной экспрессией в структурах головного мозга, здесь больше (среднее 5,9, $t = 7{,}7$; $d.f. = 8176$; $p << 0{,}001$), чем в других органах (4,7). Также эта группа генов показала большее значение количества экзонов в последовательности (среднее 13,7 в отличие от среднего 11,9, $t = 5{,}5$; $d.f. = 8210$; $p << 0{,}001$). Ранее было показано, что более низкий уровень экспрессии легче поддается регуляции, в том числе и за счет увеличения количества экзонов и длины последовательности. Таким образом, повышенные значения для количества экзонов и количества транскриптов, вероятно, связаны с высокой специфичностью этих высокоэкспрессирующихся генов (Woody, Shoemaker, 2011).

КОМПЬЮТЕРНЫЙ АНАЛИЗ ВАРИАБЕЛЬНОСТИ УРОВНЯ ЭКСПРЕССИИ ГЕНОВ В МОЗГЕ МЫШИ

В дополнение к анализу распределения экспрессии генов по тканям и структуры генов обрабатывалась информация о распределении экспрессии в отдельных структурах и участках мозга. Информация о генной экспрессии была получена из Allen Brain Atlas (ABA), содержащего данные колометрической *in situ* гибридизации (ISH) об экспрессии генов мыши в ~5×10^4 вокселях (кубических ячейках объемом 200 μm^3) мозга. Мы провели компьютерную оценку зависимости вариабельности уровня экспрессии 12932 генов от степени распространенности их экспрессии в различных районах мозга мыши.

Анализ взаимозависимостей между средним уровнем экспрессии генов E_{Avr} в вокселях мозга мыши с числом вокселей N_{Vox} на трехмерной карте мозга (см. рис. 1), в которых наблюдалась ненулевая экспрессия, и стандартным отклонением уровня экспрессии генов σ на микрочипе показал, что наблюдается достоверная отрицательная корреляция между коэффициентом вариации экспрессии $CV = \sigma/E_{Avr}$ и N_{Vox} ($R = -0{,}69$, $p < 0{,}001$). То есть чем шире ген экспрессируется в различных участках мозга, тем ниже относительная вариабельность его экспрессии (тренд представлен на рис. 4).

ГЕНОМНЫЙ КОНТЕКСТ: ЦИС-АНТИСЕНС ТРАНСКРИПТЫ И НЕКОДИРУЮЩИЕ РНК

Причиной большинства нейродегенеративных заболеваний является прогрессирующая гибель нейронов в определенных отделах головного мозга (Manfredsson *et al*., 2012; Lazarev *et al*., 2013). Традиционные инструменты для определения нейродегенеративных заболеваний, таких, как болезнь Альцгеймера, включают нейропсихологическое тестирование пациентов и специализированные технологии сканирования головного мозга. Поскольку нейродегенеративные изменения начинаются до проявления заметных клинических изменений, большое значение имеет поиск экспрессионных биомаркеров для ранней диагностики заболевания. Хорошими кандидатами для таких диагностик являются микроРНК – малые некодирующие РНК, вовлеченные в посттранскрипционную регуляцию генов (Cheng *et al*., 2013). Они способны циркулировать в крови, тканеспецифические профили их экспрессии могут быть определены в жидкостях тела – крови, слюне,

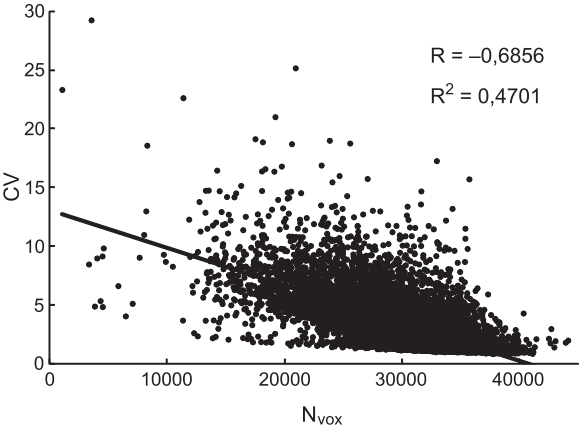


Рис. 4. Зависимость коэффициента вариации экспрессии CV (ось Y) от количества вокселей N_{Vox} (ось X).

моче. Для болезни Альцгеймера, в частности, кандидатами являются miR-9, miR-20a, и miR-132 (Cheng *et al*., 2013).

По данным полного секвенирования транскриптома, до 50 % транскриптов не кодируют белки (Hung, Chang, 2010). Длинные некодирующие РНК (более 300 нуклеотидов) играют большую роль в развитии глиом мозга, их экспрессия также имеет предсказательное значение (Zhang *et al*., 2012). Известен ряд некодирующих РНК генов (в частности, ген HULC, ген внутриядерной РНК MALAT-1), экспрессия которых значительно повышена при раке и может служить строгим маркером для диагностики (Lai *et al*., 2012; Lipovich *et al*., 2012).

По отношению к белок-кодирующим генам их можно классифицировать как межгенные, внутригенные (в интронах) и антисенс-транскрипты (в противоположной ориентации к экзонам кодирующих генов). Цис-антисенс транскрипты – это кодируемые последовательности, которые транскрибируются в противоположных направлениях и частично перекрываются в геномных координатах. Интересно отметить, что для большого числа цис-антисенс транскриптов в геноме человека характерна активность в клетках мозга.

Для анализа мы использовали базу данных антисенс-транскриптов в геноме человека NATsDB (natsdb.cbi.pku.edu.cn/) и USAGDP (Grinchuk *et al*., 2010). Была исследована экспрессия генов на микрочипах Affymetrix U133 на выборке здоровых тканей мозга (21 пациент, данные GEO NCBI) (рис. 5). Как видно из рис. 5,

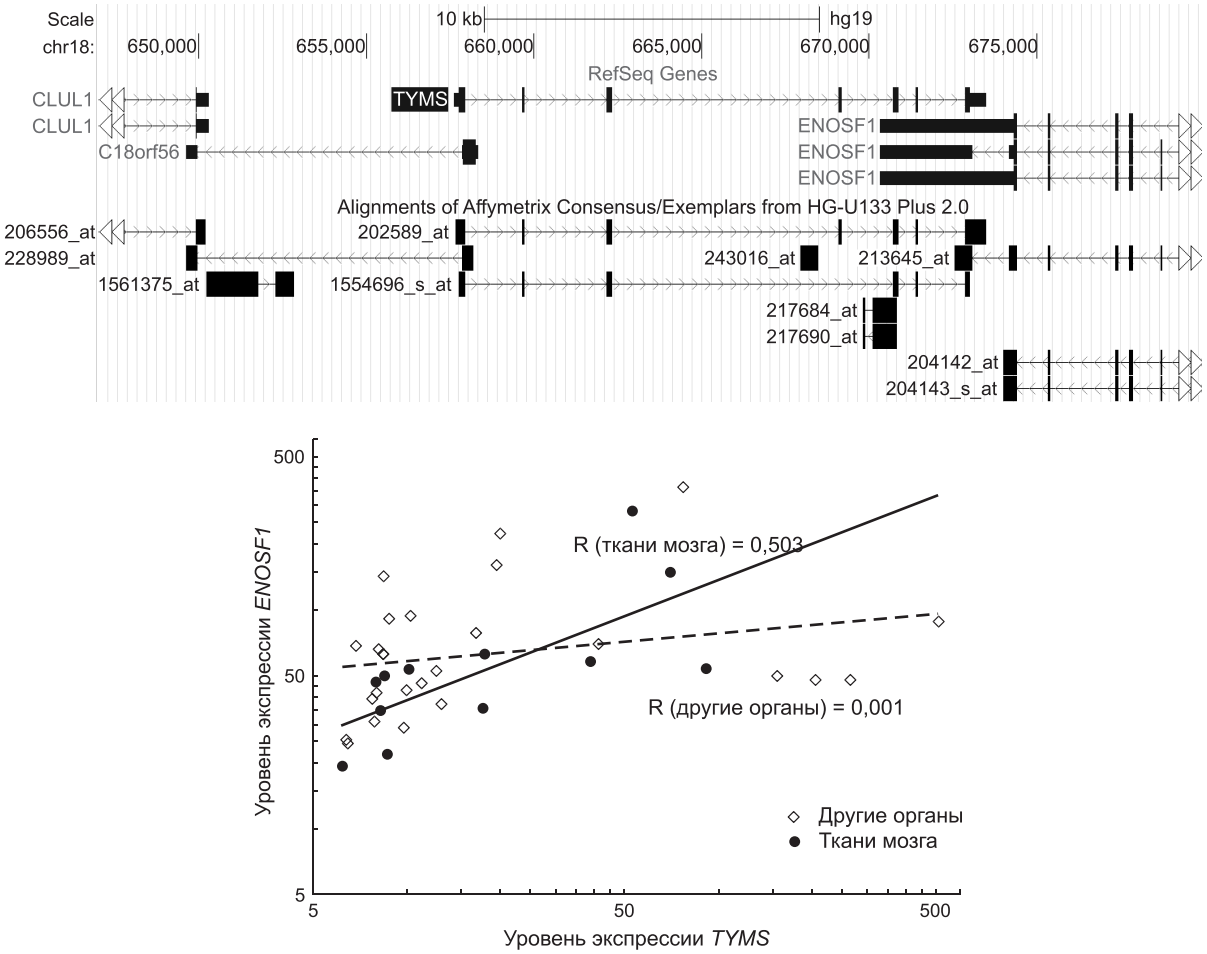


Рис. 5. Верхняя панель – гены *CLUL1*, *C18orf56*, *TYMS*, *ENOSF1* в противоположной ориентации в геноме человека (хромосома 18) и соответствующие пробы микрочипа Affymetrix (UCSC Genome Browser). Нижняя панель – корреляция экспрессии генов *ENOSF1* и *TYMS* в тканях мозга (R = 0,5) и в других органах (R = 0,001), по микрочиповым данным BioGPS.

существует несколько вариантов антисенс-расположения транскриптов (представленных наборами проб). 3 набора проб (Affymetrix) попадают в интроны гена *TYMS*, 2 набора проб находятся в противоположной ориентации (представлены EST).

Интересно отметить положительную корреляцию уровня экспрессии генов, расположенных в противоположной ориентации на клинической выборке микрочиповых данных (GEO NCBI GSE4290). Таким образом, это свидетельствует о повышенной транскрипционной активности всего геномного района в клетках мозга.

Иллюстрация корреляции экспрессии генов *TYMS* и *ENOSF1*, которые пересекаются по расположению транскриптов, на выборке тканей по базе BioGPS дана на рис. 5 (нижняя панель). Каждая точка соответствует ткани (всего 38). Корреляция экспрессии по 12 тканям мозга составляет около 0,5, в то же время по набору других тканей корреляция экспрессии практически не значима.

МОДЕЛИ ЭКСПРЕССИИ НА ЛАБОРАТОРНЫХ ЖИВОТНЫХ

Наибольший интерес представляет исследование экспрессии генов в клетках структур головного мозга человека, однако гораздо больший объем материала идет от лабораторных животных, в том числе в базе данных Allen Brain Atlas. Разработаны специальные модели животных, например, линии крыс, отличающихся по поведению, в частности крысы линии OXYS, выведенные в ИЦиГ СО РАН, (Kolosova *et al*., 2006). Достаточное физиологическое сходство и эволюционная консервативность предполагают возможным исследовать экспрессию генов животных. Рассматривалась задача анализа экспрессии генов в тканях мозга лабораторных животных – крыс, селектированных по генетическим особенностям предрасположенности к стрессу и отличающихся, в том числе, по когнитивным функциям (способности к обучению). Для крыс линии OXYS был выделен набор генов, расположенных на хромосоме 1 и связанных с фенотипом животных (Кожевникова и др., 2012; Kozhevnikova *et al*., 2013). Среди биологических процессов, обогащенных в категориях генных онтологий для генов данного локуса, интересно отметить передачу сигнала, неврологические процессы и визуальное восприятие. Присутствие генов из метаболических путей, связанных с болезнью Альцгеймера, может свидетельствовать об экспрессии генов данного локуса в нейронах и тканях мозга (Kozhevnikova *et al*., 2013).

ВЫВОДЫ И ЗАКЛЮЧЕНИЕ

Развитие новых экспериментальных методов секвенирования привело к стремительному росту объемов данных в геномике в целом и в области экспрессии генов в структурах мозга в частности. Обработка и анализ таких данных требуют разработки специализированных компьютерных средств. Нами использовались собственные компьютерные программы, базы данных и программный комплекс ICGenomics, предназначенный для анализа данных секвенирования и функциональной аннотации генов, компьютерной поддержки исследований в геномике и биомедицине (Орлов и др., 2012).

Проведено компьютерное исследование генов человека, высокоэкспрессирующихся в тканях мозга по сравнению с другими тканями и в отдельных районах головного мозга (по данным BioGPS и Allen Brain Atlas). По выборкам генов, дифференциально экспрессирующихся в различных органах, рассчитаны число экзонов и его соотношение с уровнем экспрессии. Показано статистическое различие числа альтернативных транскриптов для генов, активных в тканях мозга и других органов. Компьютерный анализ вариабельности экспрессии генов показал, что чем шире ген экспрессируется в мозге мыши, тем меньшим коэффициентом вариации экспрессии он характеризуется. Рассмотрены особенности нетранслируемых районов генов, связанные с тканеспецифичной регуляцией экспрессии. Дан обзор особенностей геномной структуры генов, имеющих дифференциальную экспрессию в клетках мозга человека и лабораторных животных, показана связь с дополнительными механизмами регуляции экспрессии генов на уровне транскрипции и трансляции, включающими присутствие микроРНК, коротких некодирующих транскриптов. Системное исследование экспрессии генов

в клетках мозга с помощью взаимодополняющих экспериментальных подходов является необходимой основой междисциплинарных нейробиологических исследований и должно быть продолжено с использованием новых транскриптомных данных.

БЛАГОДАРНОСТИ

Работа поддержана Министерством образования и науки России (соглашение 8740), ИП СО РАН № 136.

ЛИТЕРАТУРА

Витяев Е.Е., Орлов Ю.Л., Вишневский О.В. и др. Компьютерная система «GENE DISCOVERY» для поиска закономерностей организации регуляторных последовательностей эукариот // Молекуляр. биология. 2001. Т. 35. № 6. С. 952–960.

Кожевникова О.С., Мартыщенко М.К., Генаев М.К. и др. RatDNA: база данных микрочиповых исследований на крысах для генов, ассоциированных с заболеваниями старения // Вавилов. журн. генет. и селекции. 2012. Т. 16. № 4/1. P. 756–765.

Орлов Ю.Л., Брагин А.О., Медведева И.В. и др. ICGenomics: программный комплекс анализа символьных последовательностей геномики // Вавилов. журн. генет. и селекции. 2012. Т. 16. № 4/1. P. 732–741.

Орлов Ю.Л., Вишневский О.В., Витяев Е.Е. и др. Биоинформационный анализ экспрессии генов в клетках мозга // Тр. XV Всерос. науч.-техн. конф. «Нейроинформатика-2013». 21–25 января 2013 г. М.: Национальный исследовательский ядерный ун-т «МИФИ», 2013. С. 74–85.

Ananko E.A., Podkolodny N.L., Stepanenko I.L. *et al.* GeneNet in 2005 // Nucl. Acids Res. 2005. 33(Database issue). D425–427.

Cheng L., Quek C., Sun X. *et al.* Deep-sequencing of microRNA associated with Alzheimer's disease in biological fluids: From biomarker discovery to diagnostic practice // Frontiers in Genetics. 2013. V. 4. 00150.

Darnell J.C. Defects in translational regulation contributing to human cognitive and behavioral disease // Curr. Opin. Genet. Dev. 2011. V. 21. No. 4. P. 465–473.

Demenkov P.S., Ivanisenko T.V., Kolchanov N.A., Ivanisenko V.A. ANDVisio: A new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem // In Silico Biol. 2011. V. 11. No. 3. P. 149–161.

Gerashchenko M.V., Lobanov A.V., Gladyshev V.N. Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress // Proc. Natl Acad. Sci. USA. 2012. V. 109. No. 43. P. 17394–17399.

Grinchuk O.V., Jenjaroenpun P., Orlov Y.L. *et al.* Integrative analysis of the human cis-antisense gene pairs, miRNAs and their transcription regulation patterns // Nucl. Acids Res. 2010. V. 38. No. 2. P. 534–547.

Hawrylycz M.J., Lein E.S., Guillozet-Bongaarts A.L. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome // Nature. 2012. V. 489. No. 7416. P. 391–399.

Hung T., Chang H.Y. Long noncoding RNA in genome regulation: prospects and mechanisms // RNA Biol. 2010. V. 7. No. 5. P. 582–585.

Jung H., O'Hare C.M., Holt C.E. Translational regulation in growth cones // Curr. Opin. Genet. Dev. 2011. V. 21. No. 4. P. 458–464.

Kolosova N.G., Trofimova N.A., Fursova A. Opposite effects of antioxidants on anxiety in Wistar and OXYS rats // Bull. Exp. Biol. Med. 2006. V. 141. P. 734–737.

Kozhevnikova O.S., Korbolina E.E., Stefanova N.A. *et al.* Association of AMD-like retinopathy development with an Alzheimer's disease metabolic pathway in OXYS rats // Biogerontology. 2013. DOI 10.1007/s10522-013-9439-2 [Epub ahead of print].

Kundel M., Jones K.J., Shin C.Y., Wells D.G. Cytoplasmic polyadenylation element-binding protein regulates neurotrophin-3-dependent beta-catenin mRNA translation in developing hippocampal neurons // J. Neurosci. 2009. V. 29. No. 43. P. 13630–13639.

Lai M.C., Yang Z., Zhou L. *et al.* Long non-coding RNA MALAT-1 overexpression predicts tumor recurrence of hepatocellular carcinoma after liver transplantation // Med. Oncol. 2012. V. 29. No. 3. P. 1810–1816.

Lazarev V.F., Sverchinskyi D.V., Ippolitova M.V. *et al.* Factors affecting aggregate formation in cell models of Huntington's disease and amyotrophic lateral sclerosis // Acta Naturae. 2013. V. 5. No. 2. P. 81–89.

Lein E.S., Hawrylycz M.J., Ao N. *et al.* Genome-wide atlas of gene expression in the adult mouse brain // Nature. 2007. V. 445. No. 7124. P. 168–176.

Lipovich L., Dachet F., Cai J. *et al.* Activity-dependent human brain coding/noncoding gene regulatory networks // Genetics. 2012. V. 192. No. 3. P. 1133–1148.

Liu-Yesucevitz L., Bassell G.J., Gitler A.D. *et al.* Local RNA translation at the synapse and in disease // J. Neurosci. 2011. V. 31. No. 45. P. 16086–16093.

Lohse I., Reilly P., Zaugg K. The CPT1C 5'UTR contains a repressing upstream open reading frame that is regulated by cellular energy availability and AMPK // PLoS One. 2011. V. 6. No. 9. e21486.

Manfredsson F.P., Bloom D.C., Mandel R.J. Regulated protein expression for in vivo gene therapy for neurological disorders: progress, strategies, and issues // Neurobiol. Dis. 2012. V. 48. No. 2. P. 212–221.

Menschaert G., Van Criekinge W., Notelaers T. *et al.* Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events // Mol. Cell Proteomics. 2013. V. 12. No. 7. P. 1780–1790.

Naumenko V.S., Kondaurova E.M., Popova N.K. On the role of brain 5-HT7 receptor in the mechanism of hypothermia: comparison with hypothermia mediated via 5-HT1A and 5-HT3 receptor // Neuropharmacology. 2011. V. 61. No. 8. P. 1360–1365.

Orlov Y.L., Zhou J., Lipovich L. *et al.* Quality assessment of the Affymetrix U133A&B probesets by target sequence mapping and expression data analysis // In Silico Biol. 2007. V. 7. No. 3. P. 241–260.

Park J., Xu K., Park T., Yi S.V. What are the determinants of gene expression levels and breadths in the human genome? // Hum. Mol. Genet. 2012. V. 21. No. 1. P. 46–56.

Savinkova L., Drachkova I., Arshinova T. *et al.* An experimental verification of the predicted effects of promoter TATA-box polymorphisms associated with human diseases on interactions between the TATA boxes and TATA-binding protein // PLoS One. 2013. V. 8. No. 2. e54626.

Sidrauski C., Acosta-Alvear D., Khoutorsky A. *et al.* Pharmacological brake-release of mRNA translation enhances cognitive memory // eLife. 2013. V. 28. e00498.

Su A.I., Wiltshire T., Batalov S. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes // Proc. Natl Acad. Sci. USA. 2009. V. 101. No. 16. P. 6062–6067.

Sun X., Liu J., Crary J.F. *et al.* ATF4 protects against neuronal death in cellular Parkinson's disease models by maintaining levels of parkin // J. Neurosci. 2013. V. 33. No. 6. P. 2398–2407.

Wei L.N. The RNA superhighway: axonal RNA trafficking of kappa opioid receptor mRNA for neurite growth // Integr. Biol. (Camb). 2011. V. 3. No. 1. P. 10–16.

Willis D.E., Twiss J.L. Regulation of protein levels in subcellular domains through mRNA transport and localized translation // Mol. Cell Proteomics. 2010. V. 9. No. 5. P. 952–962.

Woody J.L., Shoemaker R.C. Gene expression: sizing it all up // Front Genet. 2011. V. 2. P. 70.

Wu C., Orozco C., Boyer J. *et al.* BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources // Genome Biol. 2009. V. 10. No. 11. R130.

Xie J., Zhao T., Lee T. *et al.* Anisotropic path searching for automatic neuron reconstruction // Med. Image Anal. 2011. V. 15. No. 5. P. 680–689.

Zhang X., Sun S., Pu J.K. *et al*. Long non-coding RNA expression profiles predict clinical phenotypes in glioma // Neurobiol. Dis. 2012. V. 48. No. 1. P. 1–8.

COMPUTER ANALYSIS OF DATA ON GENE EXPRESSION IN BRAIN CELLS OBTAINED BY MICROARRAY TESTS AND HIGH-THROUGHPUT SEQUENCING

I.V. Medvedeva, O.V. Vishnevsky, N.S. Safronova, O.S. Kozhevnikova, M.A. Genaev, A.V. Kochetov, D.A. Afonnikov, Y.L. Orlov

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: orlov@bionet.nsc.ru

Summary

The scope of neurobiological studies has been greatly expanded in the last years. It is accompanied by accumulation of a huge body of experimental data on the structure, function and evolution of the nervous system at different hierarchical levels of its organization. High-throughput sequencing technologies and microarray tests permit the expression of thousands of genes to be analyzed with regard to cell location in the brain. Methods of gene expression analysis are briefly reviewed in the context of brain research. We have analyzed specific features of genes differentially expressed in brain cells. Some genes overexpressed in brain tissues are associated with neurological diseases. The numbers of exons and active transcripts in genes differentially expressed in different organs are considered. Statistically significant difference in such parameters is shown for genes intensely expressed in the brain and other organs. Examples of such differentially expressed genes associated with neurological deceases are presented.

Key words: bioinformatics, brain, gene expression, microarray, sequencing.

УДК 575.117.2:577.217.522:579.88

ЭКСПРЕССИЯ ГЕНОВ И ВТОРИЧНЫЕ СТРУКТУРЫ В мРНК В РАЗНЫХ ВИДАХ *MYCOPLASMA*

© 2013 г. **В.С. Соколов[1], В.А. Лихошвай[1, 2], Ю.Г. Матушкин[1, 2]**

[1] Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия,
e-mail: sokovlad1@bionet.nsc.ru;
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

Определение эффективности экспрессии генов организма является актуальной и важной задачей современной биологии. Особый интерес представляют микроорганизмы, паразитирующие на человеке и домашних животных. В работе был проведен биоинформатический анализ геномов 62 штаммов бактерий, принадлежащих к роду *Mycoplasma*. Показано, что эффективность трансляции генов у этих организмов зависит от количества потенциальных вторичных структур в них и не зависит от кодонного состава. Обнаружены виды с пониженным содержанием локальных инвертированных повторов в генах. Анализ филогении показал возможную связь этой особенности со средой обитания данных организмов. Обнаружена не свойственная остальным микоплазмам высокая концентрация локальных инвертированных повторов в районе старт-кодона трансляции в генах *M. haemofelis*.

Ключевые слова: *Mycoplasma*, частоты кодонов, вторичные структуры, эффективность трансляции.

ВВЕДЕНИЕ

Определение эффективности экспрессии генов организма является актуальной и важной задачей современной биологии. Для ее решения были разработаны специальные экспериментальные методы, основными из которых на данный момент являются ДНК микрочипы и ПЦР в реальном времени. Однако данные методы требуют специального оборудования и реактивов. Поэтому разработка методов оценки предполагаемого уровня экспрессии генов на основе биоинформатического анализа их нуклеотидных последовательностей является актуальной и полезной.

Экспрессия гена – это процесс, в ходе которого наследственная информация из последовательности нуклеотидов ДНК преобразуется в функциональный продукт – РНК или белок. Процесс экспрессии генов состоит из нескольких стадий: транскрипция, трансляция и посттрансляционная модификация белков. Подробного изучения требуют все перечисленные стадии. Данная работа посвящена изучению именно стадии трансляции. Большое количество времени и энергии в процессе трансляции затрачивается на стадию элонгации, движение комплекса рибосомных белков вдоль мРНК с одновременным синтезом закодированной в ней молекулы белка. Поэтому изучение особенностей нуклеотидных последовательностей генов, связанных со скоростью прохождения стадии элонгации, может помочь в определении итогового уровня их трансляции.

Во многих организмах была обнаружена неравномерность в использовании синонимичных кодонов при кодировании аминокислот в белках (Grantham *et al*., 1980; Sharp, Li, 1987; Andersson, Kurland, 1990; Wada *et al*., 1990; Stenico *et al*., 1994). Установлено, что частоты кодонов коррелируют с концентрациями соответствующих им молекул тРНК (Bennetzen, Hall, 1982; Gouy, Gautier, 1982; Ikemura, 1985). Чем больше в мРНК наиболее часто используемых кодонов, тем быстрее проходит стадия элонгации трансляции для данного гена, так

как не происходит задержки рибосомы на кодонах, которым соответствуют тРНК с низкой концентрацией (Varenne *et al*., 1984; Sorensen *et al*., 1989).

Было показано, что кроме частот кодонов на скорость движения рибосомы по мРНК могут влиять вторичные структуры (шпильки), образующиеся перед ней (Jacks *et al*., 1988; Dam *et al*., 1990; Thanaraj, Argos, 1996; Lopinski *et al*., 2000; Takyar *et al*., 2005). Поэтому если в нуклеотидной последовательности гена встречается много локальных инвертированных повторов, которые потенциально могут образовать шпильки, то скорость трансляции такого гена может быть ниже, чем у других.

В опубликованных статьях есть данные о различных организмах, которые по-разному оптимизировали первичную структуру своих генов для повышения эффективности трансляции. Так, например, эффективность трансляции у *E. coli* и *S. cerevisiae* коррелирует с неравномерностью использования кодонов в их генах (Bennetzen, Hall, 1982; Gouy, Gautier, 1982; Li, Luo, 1996). В то же время для *H. pylori* такой корреляции не обнаружено, но обнаружена корреляция с количеством вторичных структур в мРНК (Vladimirov *et al*., 2007). У *Mycoplasma gallisepticum* была обнаружена корреляция между количеством в клетке белка и количеством вторичных структур в кодирующем его гене (данных нет в широкой печати, но присутствуют в отчете по гранту РФФИ № 06-04-49556).

Основанием к проведению анализа организмов, принадлежащих именно к роду *Mycoplasma*, послужил размер их геномов. Поскольку большинство данных организмов являются паразитами, их геном значительно редуцирован, что позволяет упростить изучаемую систему и, возможно, обнаружить более явно присущие ей закономерности (табл. 1). А в свете того, что на данный момент процесс трансляции и механизмы его регуляции у микоплазм мало изучены, в том числе экспериментально, данная работа является весьма актуальной.

МАТЕРИАЛЫ И МЕТОДЫ

Анализируемые последовательности белок-кодирующих генов с фланкирующими районами длиной 600 нуклеотидов экстрагировались из файлов в формате gbk, содержащих полные геномы исследуемых организмов. Данные файлы были получены из базы данных GenBank (http://www.ncbi.nlm.nih.gov/genbank/). Из итогового списка удалялись гены, отмеченные как псевдогены, а также гены длиной менее 30 кодонов. Дальнейшие расчеты проводились на оставшихся в списке генах.

Суть работы алгоритма заключается в оценке среднего времени, затрачиваемого рибосомой на стадию элонгации трансляции (Лихошвай, Матушкин, 2000). Для этого для каждого гена рассчитывается специальный индекс эффективности элонгации (EEI – elongation efficiency

Таблица 1

Организмы, на которых паразитируют различные виды *Mycoplasma*

Хозяин	Паразит	Поражаемые клетки
Человек	*M. genitalium*	Реснитчатый эпителий дыхательных и половых путей
	M. pneumoniae	Реснитчатый эпителий трахей
Кошки	*C.M. haemominutum*	Зрелые эритроциты
	M. haemofelis	
	C.M. turicensis	
Собаки	*C.M. haematoparvum*	
	M. haemocanis	
Овцы и козы	*M. ovis*	
Свиньи	*M. suis*	
Коровы	*M. wenyonii*	
Ламы и альпаки	*C.M. haemolamae*	

index), пропорциональный скорости элонгации (Likhoshvai, Matushkin, 2002). Чем выше значение EEI, тем быстрее рибосома движется по мРНК и тем быстрее синтезируется белок. Кодонный состав гена и количество (и «прочность») потенциальных вторичных структур в мРНК – два основных фактора, учитываемых при расчете индекса эффективности элонгации.

Существуют пять типов индекса эффективности элонгации:

1. EEI1 – зависит только от кодонного состава гена;
2. EEI2 – зависит от количества потенциальных вторичных структур в мРНК без учета их энергии;
3. EEI3 – зависит от количества и энергии (стабильности) потенциальных вторичных структур в мРНК;
4. EEI4 – зависит от кодонного состава и от количества потенциальных вторичных структур;
5. EEI5 – зависит от кодонного состава и от количества и энергии потенциальных вторичных структур (Vladimirov *et al*., 2007).

Для определения, какой из 5 типов индекса адекватно определяет эффективность элонгации в конкретном организме, все его гены сортируются по EEI и рассчитываются параметры M в интервале [–100; 100] и R в интервале [0; 100]. M имеет смысл среднего положения генов рибосомных белков в отсортированном списке, а R – стандартного отклонения от среднего значения. Если рибосомные гены оказываются близко к тому краю списка, где располагаются гены с наибольшей скоростью элонгации, тогда параметр M у такого организма и данного типа индекса будет близок к значению 100. Таким образом, для конкретного организма адекватен тот тип индекса, для которого значение параметра M наибольшее. Если у двух типов индекса значения параметра M совпадают, тогда выбирается индекс с меньшим значением параметра R, что говорит о более компактном расположении рибосомных генов.

Для более детального анализа последовательностей генов на наличие вторичных структур для каждого гена организма рассчитывались индексы локальной комплементарности (local complementary index – LCI, один из составляющих EEI) (Likhoshvai, Matushkin, 2002). Данный индекс имеет смысл среднего числа локальных инвертированных повторов на один ген. Таким образом, чем больше таких повторов встречается в последовательности (т. е. чем больше потенциальных вторичных структур может в ней образоваться), тем выше значение индекса LCI. Далее последовательности всех генов одного организма выравнивались по старту (или стоп-кодону) трансляции и по ним рассчитывались средние значения в области [–500; +500] относительно старта (стоп-кодона) трансляции. По полученным данным строились профили LCI для всех организмов.

Для проведения исследований была написана специальная программа, реализующая описанный выше алгоритм. Программа в ближайшее время будет доступна на сайте ИЦиГ СО РАН по адресу: http://www.bionet.nsc.ru/razrabotki/prikladnyie-razrabotki/programmyi-dlya-evm.html.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

При помощи программы были проанализированы геномы 62 штаммов *Mycoplasma*. Часть результатов представлена в табл. 2 (полная версия табл. 2 представлена в Приложении).

Суммарное распределение штаммов по типам индекса представлено на рис. 1.

Как видно из рис. 1, в большинстве штаммов адекватно работает второй тип индекса (EEI2) – эффективность элонгации зависит только от количества вторичных структур в мРНК и не зависит от кодонного состава генов. Почти у всех видов рибосомные гены хорошо определяются как высокоэкспрессируемые, значения параметра смещения M для них высоки. Но есть виды со значительно более низкими значениями параметра M: *C.M. haemominutum*, *M. suis*, *M. pneumoniae* и особенно *M. haemocanis* и *M. haemofelis*. У данных видов организмов практически отсутствует смещение рибосомных генов в сторону высокоэкспрессирующихся.

Объяснить этот факт можно с помощью двух альтернативных предположений.

1. Рибосомные гены у данных видов не являются высокоэкспрессируемыми.
2. Все остальные гены у данных видов характеризуются повышенным уровнем экспрессии, что нивелирует уровень экспрессии рибосомных генов.

Таблица 2

Типы индексов для штаммов *Mycoplasma*

Организм	Тип EEI	M1 (R1)	M2 (R2)	M3 (R3)	M4 (R4)	M5 (R5)
Mycoplasma wenyonii str. Massachusetts	**1**	**70**	–9	–21	69	25
Candidatus Mycoplasma haemolamae str. Purdue	**1**	**30**	5	20	25	29
Mycoplasma capricolum subsp. capricolum ATCC 27343	**2**	–60	**78**	–26	40	–52
Mycoplasma fermentans JER	**2**	–40	**77**	–42	71	–62
Mycoplasma mycoides subsp. capri LC str. 95010	**2**	–53	**76**	–27	25	–51
Mycoplasma leachii PG50	**2**	–56	**75**	–31	41	–56
Mycoplasma cynos C142	**2**	–39	**74**	–29	47	–47
Mycoplasma synoviae 53	**2**	–37	**71**	–8	50	–32
Mycoplasma penetrans HF 2	**2**	–1	**71**	–32	60	–37
Mycoplasma hominis ATCC 23114	**2**	–7	**69**	–28	67	–38
Mycoplasma putrefaciens KS1	**2**	–20	**68**	–21	59	–41
Mycoplasma pulmonis UAB CTIP	**2**	–13	**68**	5	48	–16
Mycoplasma hyopneumoniae 168	**2**	–67	**68**	–28	–33	–60
Mycoplasma genitalium G37	**2**	–66	**59**	–39	–40	–64
Mycoplasma gallisepticum CA06 2006 052-5-2p	**2**	23	**49**	–7	47	–1
Mycoplasma suis KI3806	**2**	18	**27**	0	23	3
Mycoplasma suis str. Illinois	**2**	6	**26**	0	17	2
Mycoplasma pneumoniae M129-B7	**2**	–10	**25**	–27	23	–30
Mycoplasma pneumoniae FH	**2**	–13	**24**	–32	21	–31
Candidatus Mycoplasma haemominutum Birmingham 1	**2**	–8	**24**	2	8	–9
Mycoplasma haemocanis str. Illinois	**2**	–15	**–6** (63)	–6 (70)	–19	–24
Mycoplasma haemofelis Ohio2	**3**	–23	–21	**7**	–36	–3
Mycoplasma haemofelis str. Langford 1	**3**	–25	–16	**3**	–35	–12
Mycoplasma agalactiae	**4**	26	64	–24	**66**	–26
Mycoplasma bovis PG45	**4**	13	63	–40	**66**	–35
Mycoplasma pneumoniae 309	**4**	–9	23	–26	**25**	–29
Mycoplasma pneumoniae M129	**4**	–10	22	–25	**25**	–29

Примечание. Серым цветом в строке выделен тот тип индекса EEI, который работает в соответствующем штамме (наибольшее значение параметра M в строке). Темно-серым выделены штаммы с наименьшим смещением генов рибосомных белков в сторону высокоэкспрессирующихся генов ($M \leq 30$). Для штаммов с одинаковыми значениями параметра M для разных типов индекса в скобках дополнительно приведены значения параметра R.

Для выбора одного из двух объяснений были рассчитаны средние значения количества вторичных структур на один ген для исследуемых 62 представителей рода *Mycoplasma*. Полученные результаты были отсортированы по увеличению среднего количества вторичных структур на один нерибосомный ген и отображены на графике, представленном на рис. 2.

Из графика видно, что рибосомные гены у всех штаммов мало отличаются друг от друга по количеству вторичных структур, что хорошо согласуется с предположением о высокой консервативности нуклеотидных последовательностей данных генов. С другой стороны, нерибосомные гены у разных штаммов могут значительно различаться по среднему количеству вторичных структур на один ген. В частности, интересующие нас «особые» штаммы с низкими значениями параметра M попали в самую крайнюю группу с наименьшими значениями количества

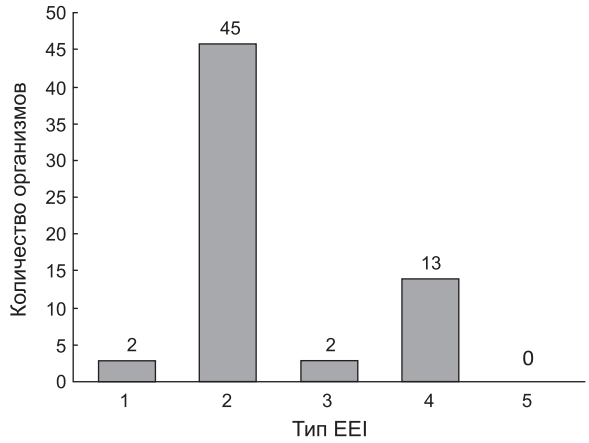


Рис. 1. Распределение 62 штаммов микоплазм по типам индекса EEI.

вторичных структур на один ген (ограничена пунктиром на рис. 2). Благодаря этому разница между рибосомными и нерибосомными генами у данных штаммов значительно меньше, чем у других, что и объясняет их особенность.

Установленный факт неравномерности по среднему числу локальных инвертированных повторов в генах разных штаммов *Mycoplasma* является новым и ранее неизученным, поэтому требует более подробного исследования.

У четырех штаммов (*M. haemofelis* Ohio2, *M. haemofelis* Langford1, *M. haemocanis Illinois*, *C.M. haemolamae* Purdue) с самым низким значением среднего количества локальных инвертированных повторов на один нерибосомный ген (крайние слева на рис. 2) были рассмотрены 100 генов с самым низким количеством локальных инвертированных повторов для определения их функций. Большинство из данных 100 генов отмечены «hypothetical protein» и их функции не известны. Но среди генов с известными функциями встречаются следующие: субъединицы рибосом, субъединицы ДНК полимераз, относящиеся к синтезу АТФ и синтезу тРНК, переносчики АТФ и др. Таким образом, самое низкое содержание локальных инвертированных повторов наблюдается в генах «домашнего хозяйства».

Чтобы понять, почему же именно у данных штаммов (ограничены пунктиром на рис. 2) наблюдается такое низкое значение среднего количества локальных инвертированных повторов на один ген, мы рассмотрели исследуемые штаммы *Mycoplasma* с точки зрения филогении.

Филогенетическое дерево *Mycoplasma*, построенное на основе анализа последовательностей 16S рРНК, было взято из статьи Peters с соавт. (2008) и представлено на рис. 3. На дереве серым цветом отмечены виды со значением параметра M > 30, а черным – с M ≤ 30. Видно, что почти все виды, отмеченные черным (кроме *M. pneumoniae*), попадают в группу гемоплазм. Гемоплазмы – это гемотрофные

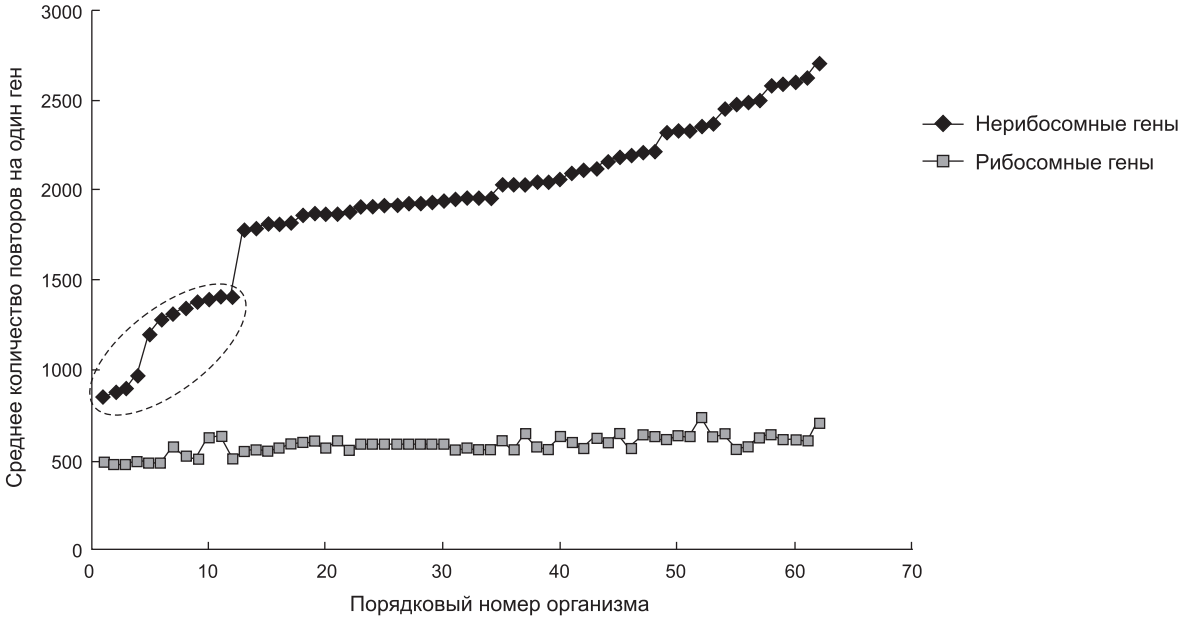


Рис. 2. Среднее число вторичных структур на один ген для каждого штамма.

Пунктиром выделены штаммы с параметром M ≤ 30.

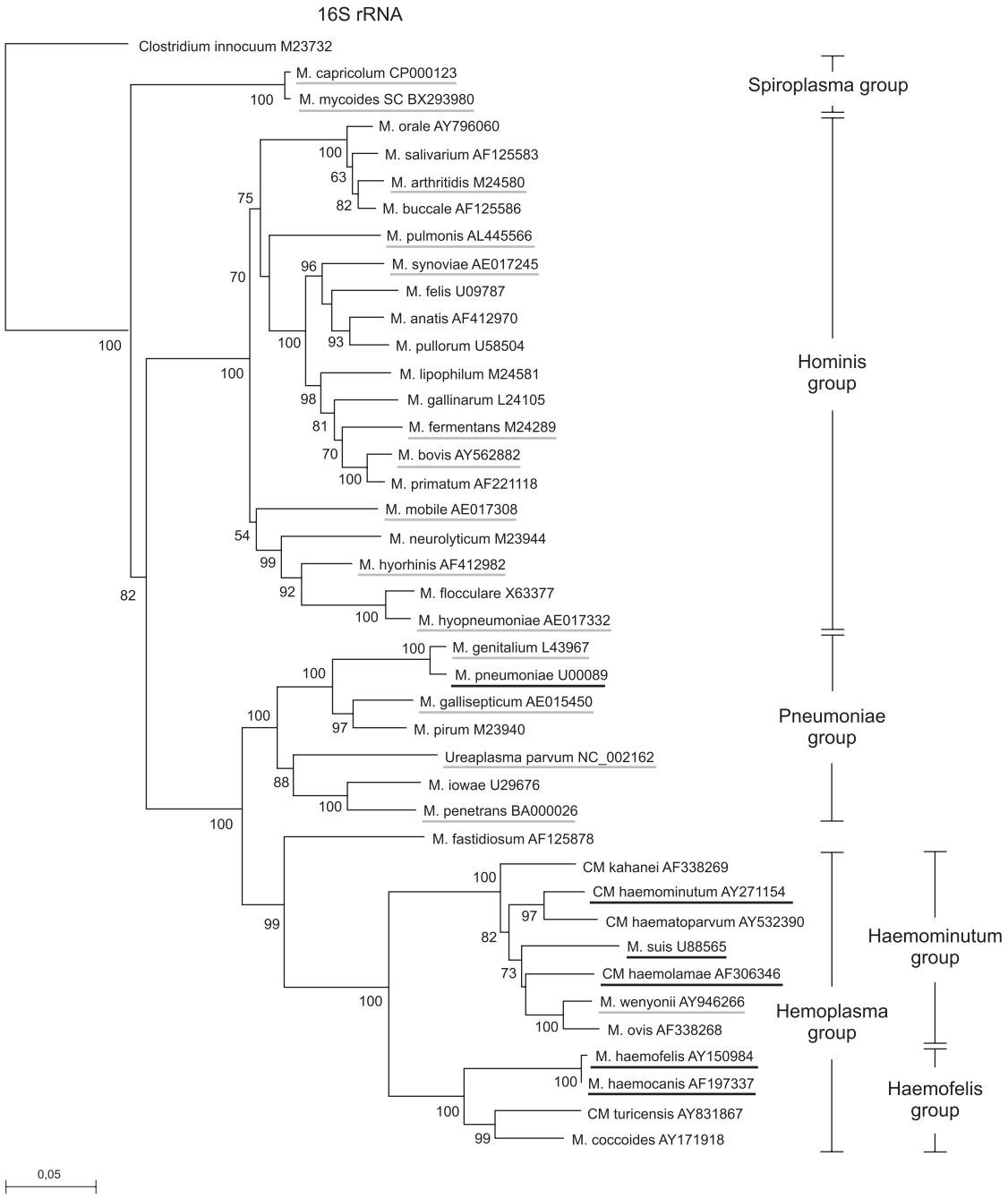


Рис. 3. Филогенетическое дерево *Mycoplasma*, построенное на основе анализа последовательностей 16S рРНК (из: (Peters *et al*., 2008)).

Серым цветом отмечены штаммы с параметром **M > 30, а черным – с M ≤ 30.**

организмы, их жизнь связана с красными кровяными тельцами (эритроцитами). Считается, что эти микоплазмы паразитируют на поверхности эритроцитов и даже могут проникать внутрь них.

На основании всего вышеизложенного мы предполагаем, что в особых условиях обитания на поверхности или внутри эритроцитов данные штаммы эволюционировали в сторону уменьшения количества вторичных структур в своих генах. Возможно, таким образом они уменьшили энергетические затраты на процесс трансляции, чтобы повысить эффективность экспрессии.

Вместе с другими микоплазмами в группу гемоплазм попала *M. wenyonii*, паразит крупного рогатого скота. У данного организма тоже отмечается низкое значение среднего количества вторичных структур на один ген (также попадает в выделенную пунктиром область на рис. 2). Однако в нем работает первый тип индекса EEI и значение параметра M = 70, т. е. рибосомные гены располагаются в области высокоэкспрессирующихся генов. Возможно, данный организм эволюционировал в сторону оптимизации кодонного состава, а пониженное содержание вторичных структур в генах досталось ему от предка.

Как было отмечено выше, *M. pneumoniae* не попала в группу гемоплазм, но у данного организма также наблюдается низкое значение параметра M и сниженное среднее количество вторичных структур на один ген. *M. pneumoniae* является паразитом верхних и нижних дыхательных путей человека. Она может взаимодействовать с поверхностью клеток дыхательного эпителия (реснитчатый эпителий трахей и клетки, выстилающие подслизистые железы (Collier, Clyde, 1971; Powell *et al*., 1976)) и, возможно, проникать внутрь них (Waites *et al*., 2008). Возможно, условия обитания данного организма чем-то схожи с условиями обитания на поверхности эритроцитов, что и способствовало эволюции данного организма в сторону уменьшения количества вторичных структур в его генах. Для получения более точного ответа на данный вопрос необходимы дополнительные исследования.

Для более подробного изучения распределения вторичных структур в генах организмов рода *Mycoplasma* были рассчитаны специальные индексы локальной комплементарности для каждого нуклеотида в гене и на его флангах (LCI(i, j), где i – номер гена, j – номер нуклеотида в гене). После расчетов все гены одного организма выравнивались по старт- (стоп-) кодону трансляции и рассчитывались средние значения индексов LCI(i, j). Полученные результаты показаны на рис. 4–7.

У большинства представителей рода *Mycoplasma* средний профиль LCI(i, j) для 5′- и 3′-районов гена имеет вид, как у *M. fermentans* JER (рис. 4, 5). В районе старт-кодона трансляции наблюдается характерный спад профиля, а в районе стоп-кодона – пик. Графики схожи с LCI профилями для других организмов

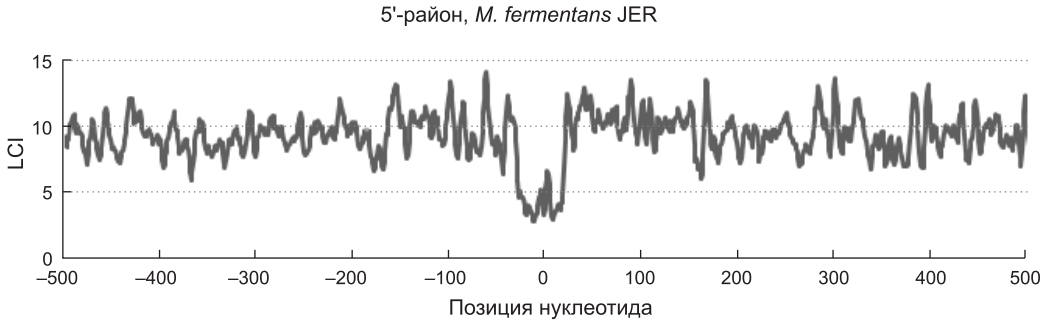


Рис. 4. Средний профиль LCI(i, j) по всем генам *M. fermentans* JER (0 – старт-кодон).

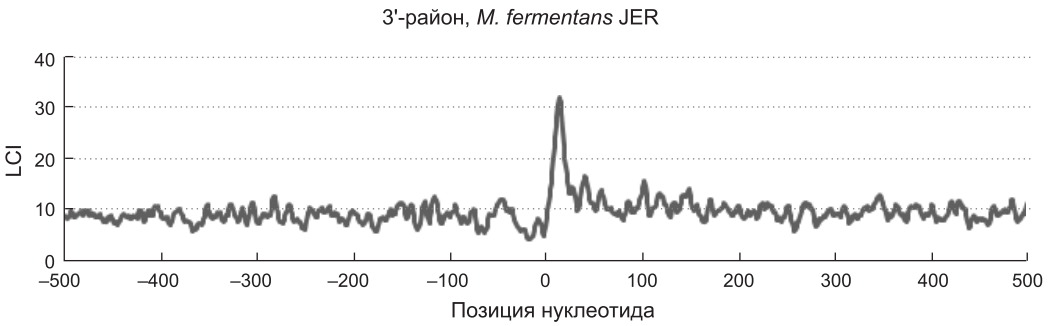


Рис. 5. Средний профиль LCI(i, j) по всем генам *M. fermentans* JER (0 – стоп-кодон).

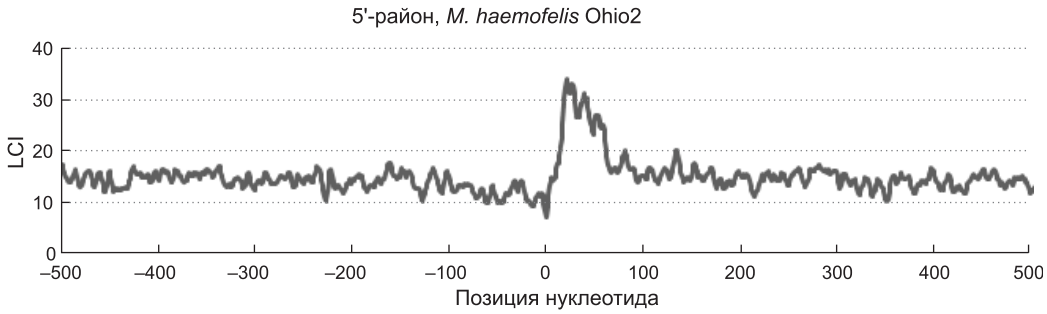


Рис. 6. Средний профиль LCI(*i*, *j*) по всем генам *M. haemofelis* Ohio2 (0 – стоп-кодон).

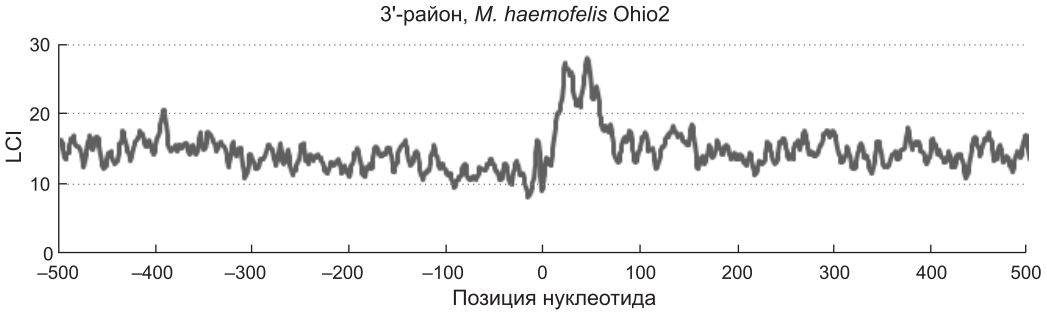


Рис. 7. Средний профиль LCI(*i*, *j*) по всем генам *M. haemofelis* Ohio2 (0 – стоп-кодон).

(*E. coli*, *S. cerevisiae*, *H. sapiens* и др.) (Matushkin *et al.*, 2004).

Спад профиля в 5′-районе, т. е. пониженная вероятность образования шпилек, вероятнее всего, способствует сборке рибосомного комплекса и началу трансляции. Наличие пика в 3′-районе говорит о повышенной вероятности образования шпилек в данной области, которые могут отвечать за терминацию трансляции или, возможно, транскрипции.

Интересной особенностью обладают профили для 5′-районов у *M. haemofelis*. У данных организмов вместо спада профиля, наоборот, наблюдается его повышение (рис. 6), что говорит о повышенной вероятности образования шпилек в данной области. Пока сложно сказать что-либо о причинах данного явления и почему именно *M. haemofelis* обладает данной особенностью. Но, как уже было сказано выше, данный организм обитает в особых условиях (поверхность или внутриклеточное пространство эритроцитов), которые, возможно, способствовали эволюции первичной структуры его генов в сторону уменьшения количества вторичных структур. Если предположить, что скорость прохождения элонгации трансляции у данного штамма возросла, возможно, лимитирующей стадией стала именно стадия начала трансляции. Поэтому мы предполагаем, что шпильки в 5′-районе могут отвечать именно за регуляцию начала трансляции у данного штамма. Возможно, у *M. haemofelis* есть особенный механизм регуляции начала трансляции, отличный от механизмов в других *Mycoplasma*. Для выяснения причин наличия пика в 5′-районе профиля LCI (*i*, *j*) у *M. haemofelis* необходимы дополнительные исследования, в частности более детальное изучение распределения и строения вторичных структур в данном районе генов этого организма.

ЗАКЛЮЧЕНИЕ

Таким образом, проведен биоинформатический анализ геномов 62 штаммов, принадлежащих к роду *Mycoplasma*. Установлено, что эволюционный отбор по эффективности трансляции генов у данных организмов шел на минимизацию количества потенциальных вторичных структур в них, а не на оптимизацию кодонного состава. Получена новая информация о количестве локальных инвертированных повторов в генах разных штаммов *Mycoplasma*. Была обнаружена группа мико-

плазм (*C.M. haemolamae*, *C.M. haemominutum*, *M. haemocanis*, *M. haemofelis*, *M. pneumoniae*, *M. suis*) с пониженным содержанием в генах локальных инвертированных повторов. Было показано, что почти все они (кроме *M. pneumoniae*) относятся к одной филогенетической группе гемоплазм и характеризуются обитанием на поверхности или внутри эритроцитов. Также при построении профилей распределения локальных инвертированных повторов в районах старт- и стоп-кодонов трансляции у *M. haemofelis* обнаружен нехарактерный для остальных микоплазм пик в районе старт-кодона. Пик свидетельствует о повышенной вероятности образования шпилек в данном районе гена. По нашим предположениям, это может быть связано с каким-то альтернативным механизмом регуляции трансляции, отличным от механизмов у других видов микоплазм.

БЛАГОДАРНОСТИ

Работа выполнена при частичной поддержке программ Президиума РАН «Молекулярная и клеточная биология» (проект 6.6) и «Происхождение биосферы и эволюция гео-биологических систем» (№ 15), гранта НШ-5278.2012.4 и гранта РФФИ № 13-04-0062013.

ЛИТЕРАТУРА

Лихошвай В.А., Матушкин Ю.Г. Предсказание эффективности экспрессии генов по их нуклеотидному составу // Молекуляр. биология. 2000. Т. 34. № 3. С. 406–412.

Andersson S.G.E., Kurland C.G. Codon preferences in free-living microorganisms // Microbiol. Rev. 1990. V. 54. P. 198–210.

Bennetzen J.L., Hall B.D. Codon selection in Yeast // J. Biol. Chem. 1982. V. 257. P. 3026–3031.

Dam E.B., Pleij C.W., Bosch L. RNA pseudoknots: translational frameshifting and readthrough on viral RNAs // Virus Genes. 1990. V. 4. P. 121–136.

Collier A.M., Clyde W.A. Jr. Relationships between Mycoplasma pneumoniae and human respiratory epithelium // Infect. Immun. 1971. V. 3. No. 5. P. 694–701.

Gouy M., Gautier C. Codon usage in bacteria: correlation with gene expressivity // Nucl. Acids Res. 1982. V. 10. P. 7055–7070.

Grantham R., Gautier C., Gouy M. *et al.* Codon catalog usage and the genome hypothesis // Nucl. Acids Res. 1980. V. 8. r49–r62.

Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms // Mol. Biol. Evol. 1985. V. 2. P. 13–34.

Jacks T., Madhani H.D., Masiarz F.R., Varmus H.E. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region // Cell. 1988. V. 55. P. 447–458.

Li H., Luo L. The relation between codon usage, base correlation and gene expression level in *Escherichia coli* and Yeast // J. Theor. Biol. 1996. V. 181. Iss. 2. P. 111–124.

Likhoshvai V.A., Matushkin Yu.G. Differentiation of single-cell organisms according to elongation stages crucial for gene expression efficacy // FEBS Lett. 2002. V. 516. P. 87–92.

Lopinski J.D., Dinman J.D., Bruenn J.A. Kinetics of ribosomal pausing during programmed –1 translational frameshifting // Mol. Cell. Biol. 2000. V. 20. P. 1095–1103.

Matushkin Yu.G., Likhoshvai V.A., Kochetov A.V. Local secondary structure may be a critical characteristic influencing translation of unicellular organisms mRNA // Bioinformatics of Genome Regulation and Structure. Boston a.o.: Kluwer Acad. Publ., 2004. P. 103–114.

Peters I.R., Helps C.R., McAuliffe L. *et al.* RNase P RNA gene (*rnp*B) phylogeny of Hemoplasmas and other *Mycoplasma* species // J. Clin. Microbiol. 2008. V. 46. No. 5. P. 1873–1877.

Powell D.A., Hu P.C., Wilson M. *et al.* Attachment of *Mycoplasma pneumoniae* to respiratory epithelium // Infect. Immun. 1976. V. 13 No. 3. P. 959–966.

Sharp P.M., Li W.H. The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications // Nucl. Acids Res. 1987. V. 15. P. 1281–1295.

Sorensen M.A., Kurland C.G., Pedersen S. Codon usage determines translation rate in *Escherichia coli* // J. Mol. Biol. 1989. V. 207. P. 365–377.

Stenico M., Lloyd A.T., Sharp P.M. Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases // Nucl. Acids Res. 1994. V. 22. P. 2437–2446.

Takyar S., Hickerson R.P., Noller H.F. mRNA helicase activity of the ribosome // Cell. 2005. V. 120. P. 49–58.

Thanaraj T.A., Argos P. Ribosome-mediated translational pause and protein domain organization // Protein Sci. 1996. V. 5. P. 1594–1612.

Varenne S., Buc J., Lloubes R., Lazdunski C. Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains // J. Mol. Biol. 1984. V. 180. P. 549–576.

Vladimirov N.V., Likhoshvai V.A., Matushkin Yu.G. Correlation of codon biases and potential secondary structures with mRNA translation efficiency in unicellular organisms // Mol. Biol. 2007. V. 41. No. 5. P. 926–933.

Wada K.S., Aota R., Tsuchiya F. *et al.* Codon usage tabulated from GenBank genetic sequence data // Nucl. Acids Res. 1990. V. 18. (Suppl.). P. 2367–2411.

Waites K.B., Balish M.F., Atkinson T.P. New insights into the pathogenesis and detection of *Mycoplasma pneumoniae* infections // Future Microbiol. 2008. V. 3. No. 6. P. 635–648.

Приложение
Таблица

Типы индексов для 62 штаммов *Mycoplasma*

Организм	Тип EEI	M1 (R1)	M2 (R2)	M3 (R3)	M4 (R4)	M5 (R5)
Mycoplasma wenyonii str. Massachusetts	**1**	**70**	–9	–21	69	25
Candidatus Mycoplasma haemolamae str. Purdue	**1**	**30**	5	20	25	29
Mycoplasma fermentans M64	**2**	–57	**79**	–50	67	–74
Mycoplasma hyorhinis MCLD	**2**	–20	**79**	–23	45	–42
Mycoplasma fermentans PG18	**2**	–53	**78**	–50	75	–66
Mycoplasma hyorhinis HUB 1	**2**	–9	**78**	–16	46	–29
Mycoplasma capricolum subsp. capricolum ATCC 27343	**2**	–60	**78**	–26	40	–52
Mycoplasma fermentans JER	**2**	–40	**77**	–42	71	–62
Mycoplasma hyorhinis SK76	**2**	–15	**77**	–24	48	–40
Mycoplasma hyopneumoniae 232	**2**	–61	**77**	–24	–17	–57
Mycoplasma crocodyli MP145	**2**	–44	**76**	–44	56	–57
Mycoplasma mycoides subsp. mycoides SC str. Gladysdale	**2**	–37	**76**	–24	61	–42
Mycoplasma mycoides subsp. mycoides SC str. PG1	**2**	–45	**76**	–21	65	–43
Mycoplasma leachii 990146	**2**	–59	**76**	–29	70	–49
Mycoplasma mycoides subsp. capri LC str. 95010	**2**	–53	**76**	–27	25	–51
Mycoplasma leachii PG50	**2**	–56	**75**	–31	41	–56
Mycoplasma cynos C142	**2**	–39	**74**	–29	47	–47
Mycoplasma synoviae 53	**2**	–37	**71**	–8	50	–32
Mycoplasma penetrans HF 2	**2**	–1	**71**	–32	60	–37
Mycoplasma hyopneumoniae 7448	**2**	–69	**71**	–29	–21	–63
Mycoplasma putrefaciens Mput9231	**2**	–17	**70** (36)	–24	70 (49)	–37
Mycoplasma hyorhinis GDL 1	**2**	–20	**70**	–23	40	–36
Mycoplasma hyopneumoniae J	**2**	–70	**70**	–30	–23	–62
Mycoplasma mobile 163K	**2**	–48	**70**	–9	53	–30
Mycoplasma hominis ATCC 23114	**2**	–7	**69**	–28	67	–38

Продолжение таблицы

Организм	Тип EEI	M1 (R1)	M2 (R2)	M3 (R3)	M4 (R4)	M5 (R5)
Mycoplasma hyopneumoniae 168 L	**2**	–64	**69**	–26	–27	–58
Mycoplasma putrefaciens KS1	**2**	–20	**68**	–21	59	–41
Mycoplasma pulmonis UAB CTIP	**2**	–13	**68**	5	48	–16
Mycoplasma hyopneumoniae 168	**2**	–67	**68**	–28	–33	–60
Mycoplasma conjunctivae HRC/581	**2**	–8	**67**	–18	50	–29
Mycoplasma arthritidis 158L3-1	**2**	9	**62**	–23	52	–21
Mycoplasma genitalium G37	**2**	–66	**59**	–39	–40	–64
Mycoplasma genitalium M6282	**2**	–52	**53**	–24	–29	–48
Mycoplasma gallisepticum str. R (high)	**2**	18	**53**	2	46	–3
Mycoplasma gallisepticum str. R (low)	**2**	19	**53**	2	46	–2
Mycoplasma genitalium M2288	**2**	–53	**52**	–30	–30	–51
Mycoplasma genitalium M6320	**2**	–50	**50**	–26	–29	–47
Mycoplasma genitalium M2321	**2**	–51	**50**	–26	–30	–46
Mycoplasma gallisepticum NY01 2001.047-5-1P	**2**	24	**49**	–7	49	–2
Mycoplasma gallisepticum ca06 2006 052-5-2p	**2**	23	**49**	–7	47	–1
Mycoplasma gallisepticum NC96 1596-4-2P	**2**	24	**48**	–6	47	–2
Mycoplasma suis KI3806	**2**	18	**27**	0	23	3
Mycoplasma suis str. Illinois	**2**	6	**26**	0	17	2
Mycoplasma pneumoniae M129-B7	**2**	–10	**25**	–27	23	–30
Mycoplasma pneumoniae FH	**2**	–13	**24**	–32	21	–31
Candidatus Mycoplasma haemominutum Birmingham 1	**2**	–8	**24**	2	8	–9
Mycoplasma haemocanis str. Illinois	**2**	–15	**–6** (63)	–6 (70)	–19	–24
Mycoplasma haemofelis Ohio2	**3**	–23	–21	**7**	–36	–3
Mycoplasma haemofelis str. Langford 1	**3**	–25	–16	**3**	–35	–12
Mycoplasma bovis Hubei-1	**4**	25	66	–44	**80**	–27
Mycoplasma agalactiae PG2	**4**	22	64	–27	**73**	–27

Окончание таблицы

Организм	Тип EEI	M1 (R1)	M2 (R2)	M3 (R3)	M4 (R4)	M5 (R5)
Mycoplasma bovis HB0801	**4**	15	67	–38	**69**	–34
Mycoplasma agalactiae	**4**	26	64	–24	**66**	–26
Mycoplasma bovis PG45	**4**	13	63	–40	**66**	–35
Mycoplasma gallisepticum str. F	**4**	31	54	1	**62**	3
Mycoplasma gallisepticum NC08 2008.031-4-3P	**4**	26	49	–7	**53**	–1
Mycoplasma gallisepticum WI01 2001.043-13-2P	**4**	26	50	–7	**52**	–1
Mycoplasma gallisepticum NC06 2006.080-5-2P	**4**	26	49	–7	**51**	–2
Mycoplasma gallisepticum NC95 13295-2-2P	**4**	26	49	–7	**51**	–2
Mycoplasma gallisepticum VA94 7994-1-7P	**4**	26	48	–7	**51**	–2
Mycoplasma pneumoniae 309	**4**	–9	23	–26	**25**	–29
Mycoplasma pneumoniae M129	**4**	–10	22	–25	**25**	–29

Примечание. Серым цветом в строке выделен тип индекса EEI, который работает в соответствующем штамме (наибольшее значение параметра M в строке). Темно-серым выделены штаммы с наименьшим смещением генов рибосомных белков в сторону высокоэкспрессирующихся генов ($M \leq 30$). Для штаммов с одинаковыми значениями параметра M для разных типов индекса в скобках дополнительно приведены значения параметра R.

GENE EXPRESSION AND mRNA SECONDARY STRUCTURES IN DIFFERENT *MYCOPLASMA* SPECIES

V.S. Sokolov[1], V.A. Likhoshvai[1, 2], Yu.G. Matushkun[1, 2]

[1] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: sokovlad1@bionet.nsc.ru;
[2] Novosibirsk National Research State University, Novosibirsk, Russia

Summary

Evaluation of gene expression efficiency in different organisms is a vital task of modern biology. Microorganisms that feed on humans and pets are particularly interesting. In this work, bioinformatical analysis of 62 *Mycoplasma* strains is performed. It has been shown that translation efficiency in these organisms depends on the number of potential secondary structures in genes and does not depend on codon compositions. Several species with low concentrations of local inverted repeats in genes have been found. Phylogenetic analysis shows that this feature may be associated with their environment. High concentrations of local inverted repeats, not typical of other *Mycoplasma* species, have been found in the translation start regions of *M. haemofelis* genes.

Key words: *Mycoplasma*, codon frequencies, secondary structures, translation efficiency.

УДК 579.222.3: 579.66: 579.8.06

ВЫДЕЛЕНИЕ И ИССЛЕДОВАНИЕ СВОЙСТВ БАКТЕРИЙ ТЕРМАЛЬНЫХ ИСТОЧНИКОВ СЕВЕРНОГО ПРИБАЙКАЛЬЯ, ОБЛАДАЮЩИХ ЛИПОЛИТИЧЕСКОЙ АКТИВНОСТЬЮ

© 2013 г. **К.Н. Сорокина, А.С. Розанов, А.В. Брянская, С.Е. Пельтек**

Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия,
e-mail: peltek@bionet.nsc.ru

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

В работе проведено исследование свойств бактериальной микрофлоры ряда термальных источников Северного Прибайкалья (Байкальская рифтовая зона), характеризующихся широким диапазоном природных условий (pH, температур, органического компонента). Проведена таксономическая идентификация бактерий, показано, что липолитической активностью обладают изоляты, относящиеся к видам *Geobacillus stearothermophilus*, *Anoxybacillus flavithermus* и *Thermoactinomyces vulgaris*. Исследованы их морфологические, биохимические и физиологические свойства. Показано, что в целом полученные штаммы *Geobacillus stearothermophilus* растут при температурах до 70 °С и широком диапазоне значений pH (5–10). Штаммы, отнесенные к роду *Anoxybacillus flavithermus*, росли при температурах 60–70 °С и до pH 11, а *Thermoactinomyces vulgaris* Gus-2-1 имел более узкий диапазон роста при 50–60 °С и pH 7–10. В целом из выделенных штаммов, обладающих липолитической активностью, наиболее интересными для изучения свойств продуцируемых липаз являются представители *Geobacillus stearothermophilus*.

Ключевые слова: термофильные микроорганизмы, липолитическая активность, Байкальская рифтовая зона.

ВВЕДЕНИЕ

Исследование термальных источников для поиска микроорганизмов с уникальными свойствами является актуальной задачей современной микробиологии. Микроорганизмы, а также продуцируемые ими ферменты находят широкое применение в современной биотехнологии. Одними из таких ферментов являются липазы, относящиеся к классу гидролаз, катализирующих реакцию расщепления природных триацилглицеридов с образованием глицерина, жирных кислот и воды. Уникальные свойства этих ферментов позволяют применять их в качестве высокоэффективных регио- и энантиоселективных биокатализаторов (Fishman *et al*., 1998) в реакции переэтерификации пищевых жиров, а также аммонолизе и других химических процессах (Joseph *et al*., 2008). В связи с этим липазы нашли широкое применение в промышленности, в том числе пищевой, фармацевтической, текстильной, в производстве бытовой химии и других областях (Houde *et al*., 2004), а изучение природных микробных сообществ для поиска новых вариантов термостабильных ферментов представляет большой практический интерес.

В ряде исследований показано, что термальные источники Прибайкалья и Забайкалья являются богатыми по составу микрофлоры (Микробные сообщества ..., 2006). Вопрос о поиске в данных источниках микроорганизмов, обладающих липолитической активностью, а также изучении их свойств остается открытым.

В данной работе были проведены исследование состава микробиологических сообществ горячих источников Северного Прибайкалья,

поиск, а также выделение и скрининг активности термофильных микроорганизмов, проявляющих липолитическую активность. Выделенные штаммы были идентифицированы, изучены их филогенетическое родство, спектр активности и морфологические характеристики.

МАТЕРИАЛЫ И МЕТОДЫ

Выделение микроорганизмов с липолитической активностью из термальных источников

Для получения активных накопительных и чистых культур термофильных микроорганизмов и последующего выделения чистых культур микроорганизмов-продуцентов липолитических ферментов проводили сбор полевого материала в горячих источниках Байкальской рифтовой зоны. Образцы отбирали с соблюдением условий стерильности. Собранные образцы хранили при +4 °С, после чего проводили посев на элективные среды.

Культивирование природного материала проводили на агаризованных средах Лурия–Бертани (LB), среде А (г/л): (0,5 NaCl, 0,5 пептона, 0,4 мясного экстракта, 0,2 дрожжевого экстракта), мясопептонном агаре (МПА). Культивирование проводили при температурах 45–70 °С в течение 1–3 суток. Определение липолитической активности штаммов проводили на агаризованной среде Пфеннига (г/л): KH_2PO_4 – 0,5; NH_4Cl – 0,5; $MgSO_4 \times 7H_2O$ – 0,5; KCl – 0,5; NaCl – 0,5; $CaCl_2 \times 2H_2O$ – 0,05; $NaHCO_3$ – 1,5, содержащей 1,5 % твин-20, -40 и -80.

Идентификация штаммов по нуклеотидной последовательности гена 16S рРНК

Чистую бактериальную культуру выращивали при температуре 55–65 °С на среде LB, содержащей 2 % агара, в течение 12 ч. Выделение геномной ДНК проводили с использованием набора Wizard SV Genomic DNA Purification System (Promega, США), согласно инструкции производителя. Амплификацию фрагмента 16s рРНК проводили в реакционной смеси общим объемом 50 мкл, содержащей: 20 нг, 0,2 мМ смеси четырех дезокситрифосфатов, 1,5 мМ $MgCl_2$, 0,1 мкМ праймеров 16s-8-f-B 5′-AGRGTTTGATCCTGGCTCA-3′ и 16s-1350-r-B 5′-GACGGGCGGGTGTACAAG-3′, буфер для Taq ДНК-полимеразы и 1 ЕА TaqSE ДНК-полимеразы («Сибэнзим», Россия). ПЦР проводили по следующей программе: 95 °С – 3 мин; 40 циклов (95 °С – 30 с, 54 °С – 20 с, 72 °С – 1 мин 30 с); 72 °С – 10 мин. Определение нуклеотидной последовательности проводили на приборе 3130XL Genetic Analyzer (Applied Biosystems) с использованием праймера 16s-8-f-B. Полученные последовательности 16S рРНК сравнивали с известными последовательностями, содержащимися в базе Genbank, с использованием алгоритма BLAST (http://www.ncbi.nlm.nih.gov/nuccore).

Филогенетический анализ

Выравнивание полученных последовательностей гена 16S рРНК проводили с использованием программы ClustualW. Построение филогенетического древа было выполнено с использованием алгоритма ближайших соседей, реализованного в программе MEGA4. Проверку статистической достоверности проводили при помощи бутстреп теста (Tamura *et al*., 2007).

Микроскопический анализ штаммов

Исследование морфологии выделенных штаммов микроорганизмов методами микроскопии проводили с использованием светового и люминесцентных микроскопов фирмы Karl Zeiss (Axioskop 2 Plus и Axio Skope. A1, Германия). Препараты готовили стандартными методами (Практикум по микробиологии, 2005).

Биохимическая характеристика культур

Биохимическую характеризацию штаммов проводили с использованием тест-системы Enterotest-24 (MicroTest, Lachema) в соответствии с инструкцией производителя. Проводили тест на активность уреазы, каталазы и оксидазы, индола, сероводорода, ацетоина; способность к восстановлению нитратов, к гидролизу казеина, желатина, крахмала, деградации тирозина, дезаминированию фенилаланина.

Скрининг физиологических свойств штаммов, обладающих липолитической активностью

Определение диапазона значений pH и температуры, при которых наблюдался рост культур выделенных микроорганизмов в жидкой среде, проводили с использованием системы MicroFlask (Applikon Biotechnology). Посевной материал выращивали путем внесения материала отдельных колоний исследуемых штаммов в лунки 96-луночного планшета в объем 180 мкл среды LB. Культивирование проводили на термостатированном шейкере KS 4000 IC control с установленными держателями для системы MicroFlask в течение 16 ч при 250 об./мин, при температурах от 40 до 70 °C. По окончании культивирования визуально отмечали лунки, в которых наблюдалось увеличение содержания биомассы.

РЕЗУЛЬТАТЫ

Выделение термофильных бактерий и исследование их свойств

В данной работе для выделения термофильных микроорганизмов были отобраны пробы донных осадков высокотемпературных источников Алла, Гарга, Гусиха, Сея, Уро (Баргузинская долина, Северное Прибайкалье). В работу были взяты образцы из трех источников, отличавшихся от остальных высокими значениями температур на выходе термальных ручьев, обильным развитием микробных сообществ и богатых органическим субстратом. В табл. 1 приведено описание условий, при которых были отобраны природные образцы, и названия источников.

Культуры микроорганизмов, полученные из природных образцов, очищали от сопутствующих организмов путем многократного пересева на агаризованной среде LB. Полученные колонии тестировали на активность по отношению к твин-содержащим субстратам в диапазоне температур 37–65 °C. Всего было отобрано 11 штаммов, обладавших наибольшей липолитической активностью, для которых была выполнена таксономическая идентификация путем исследования гена 16S рРНК, результаты приведены в табл. 2.

Таким образом, выделенные бактерии были отнесены к видам *Anoxybacillus flavithermus*, *Geobacillus stearothermophilus* и *Thermoactinomyces vulgaris*.

В ходе работы был выполнен филогенетический анализ родства выделенных штаммов. На рис. представлена дендрограмма, отражающая филогенетические отношения использованных в работе штаммов родов *Geobacillus* и *Anoxybacillus*.

Штаммы *G. stearothermophilus* B28, B8, B7, B27 в целом оказались близкими со штаммом *G. stearothermophilus* mt-10 и образовывали

Таблица 1

Описание точек забора образцов воды и грунта термальных источников Северного Прибайкалья

№	Название источника	Геохимические параметры	Описание пробы	Количество образцов
1	Алла (правый берег)	52–75 °C pH 8,1	обрастания (маты): зеленые, оранжево-бурые и белые	30
	Алла (левый берег)	43–72 °C	сообщество микроорганизмов, цвет – зеленый	10
2	Гарга	36–74 °C pH 7,3	мощное поле разноцветных матов (розовые, черные, желто-зеленые). Толщина 5–7 мм	15
3	Уринский	25–69 °C	многослойные разноцветные цианобактериальные маты	12
4	Гусиха	43–74 °C pH 8,5	накипные и плавающие цианобактериальные маты	5
5	Сеюйский	49–51 °C pH 9,7	многослойные донные и поверхностные цианобактериальные маты. Толщина до 6 см	7

одну группу за исключением штамма *G. stearothermophilus* Gus 2-3. Бактерии рода *Anoxybacillus*, выделенные в ходе работы, образовывали две разные внутривидовые группы.

Исследование свойств выделенных штаммов бактерий, обладающих липолитической активностью

Морфотипы наиболее активных в культуре штаммов бактерий родов *Geobacillus* и *Anoxybacillus*, а также состав и количество клеток в образцах изучали с помощью микроскопии. Описание колоний пяти выделенных штаммов, обладающих липолитической активностью (B7, B18, B22, B25, B27), и морфологическая характеристика клеток, согласно определителю бактерий Берджи (Boone, Castenholz, 2001), представлены в табл. 3.

Исследуемые штаммы образуют округлые колонии кремового цвета. Края колоний волнистые или слегка волнистые. Профиль колоний слегка выпуклый. Размеры колоний варьируют от 3–5 до 8 мм. Клетки изолятов представлены палочками, размеры которых варьировали от 2–3 × 7–10 мкм до 5 × 8–12 мкм. У исследуемых штаммов выявлено спорообразование. Эндоспоры овальные или сферические, от 1 до 2,5 мкм.

В экспериментах по исследованию биохимических характеристик выделенных штаммов за основу была взята среда Луриа–Бертрани (LB). Определение физиолого-биохимических характеристик проводили на основе методик, описанных в Методах общей бактериологии (Герхард, 1984) и Практикуме по микробиологии (2005).

Таблица 2

Таксономическая идентификация штаммов, выделенных из термальных источников Баргузинской долины

Штамм	Таксон	Источник
B7	*Geobacillus stearothermophilus*	Алла
B18	*Geobacillus stearothermophilus*	Алла
B22	*Geobacillus stearothermophilus*	Алла
B25	*Anoxybacillus* sp.	Алла
B27	*Geobacillus stearothermophilus*	Гаргинский
Uro-2-1	*Anoxybacillus flavithermus*	Уринский
Se-1	*Anoxybacillus flavithermus*	Сеюйский
Ga-1-1	*Anoxybacillus flavithermus*	Гаргинский
Gus-2-1	*Thermoactinomyces vulgaris*	Гусихинский
Gus-2-2	*Anoxybacillus flavithermus*	Гусихинский
Gus-2-3	*Geobacillus stearothermophilus*	Гусихинский

Спорообразование у бактерий выявляли: 1) прогреванием культуры до 80 °С в течение 10 мин и последующим посевом на питательную среду того же состава; 2) микроскопированием старых культур. Для тестирования брали культуры липолитиков, инкубируемых при температуре 60 °С на среде LB в течение 24 ч. Штаммы были проверены на способность к образованию сероводорода, уреазную активность и использование следующих субстратов: лизин,

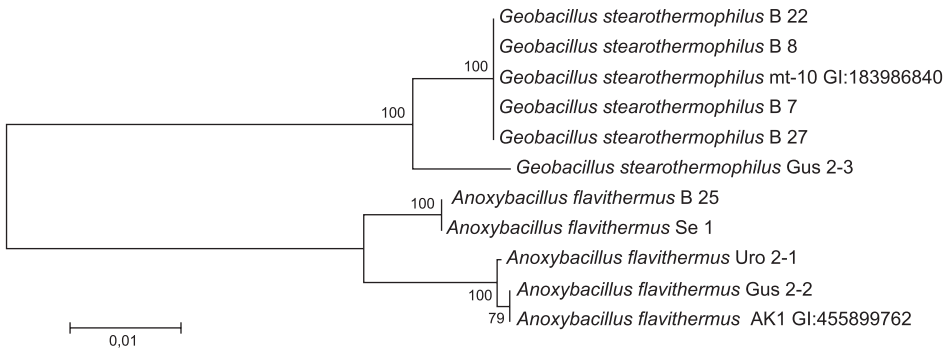


Рис. Филогения штаммов бактерий р. *Geobacillus* и *Anoxybacillus*, выделенных в ходе работы.

Таблица 3

Морфологическая характеристика бактериальных штаммов, обладающих липолитической активностью

Штамм	Морфология колоний					Морфология клеток	
	Форма	Размер, мм	Цвет	Профиль	Край	Морфотип клеток	Размеры клеток, мкм
B7	круглая	3–5	кремовые	слегка выпуклый	слегка волнистый	палочки	2–3 × 7–10 споровые, d = 1,5
B18	круглая	3–5	кремовые	слегка выпуклый	слегка волнистый	палочки	2–3 × 7–10 споровые, d = 1,5
B22	круглая	3–5	кремовые	слегка выпуклый	слегка волнистый	палочки	2–3 × 7–10 споровые, d = 1,5
B25	круглая	5	кремовые	слегка выпуклый	слегка волнистый	палочки	5 × 8–12 споровые
B27	круглая	3–5	кремовые	слегка выпуклый	слегка волнистый	палочки	2–3 × 7–10 споровые, d = 1,5
Uro-2-1	круглая	7	кремовые	слегка выпуклый	слегка волнистый	палочки	2–3 × 7–10 споровые, d = 2,5
Se-1	круглая	5	кремовые	слегка выпуклый	волнистый	палочки	2–3 × 7–10 споровые, d = 2,5
Ga-1-1	круглая	5	кремовые	слегка выпуклый	слегка волнистый	палочки	2–3 × 7–10 споровые, d = 2,5
Gus-2-1	круглая	до 8	кремовые с белым налетом	слегка выпуклый	слегка волнистый	мицелий	Субстратный мицелий 0,4–0,8 мкм в диаметре; споровые, d = 1,0
Gus-2-2	круглая	5	кремовые	слегка выпуклый	слегка волнистый	палочки	2–3 × 7–10 споровые, d = 2,5
Gus-2-3	круглая	3–5	кремовые	слегка выпуклый	слегка волнистый	палочки	2–3 × 7–10 споровые, d = 1,5

орнитин, аргинин, цитрат, малонат, инозитол, адонитол, целлобиоза, сахароза, трегалоза, маннитол, эскулин, сорбитол, рамноза, мелибиоза, раффиноза, дульцит, глюкоза. Липазную активность определяли по образованию зон кристаллизации твина-80 вокруг колоний на плотных агаризованных средах. Все штаммы обладали способностью к росту на среде с комплексными субстратами, внесенными в качестве единственных источников углерода и энергии – дрожжевом экстракте и пептоне. Спектр соединений, утилизируемых штаммами в аэробных и анаэробных условиях, представлен в табл. 4.

Исследуемые штаммы не образуют сероводород, не используют цитрат, малонат, инозитол, адонитол, целлобиозу, сахарозу, трегалозу, маннитол, сорбитол, рамнозу, мелибиозу, раффинозу, дульцит, глюкозу. Используют лизин и орнитин. Дают положительную реакцию на β-галактозидазу и эскулин, в ряде случаев – на уреазу. Данные реакции соответствуют характеристике видов *Bacillus*, приведенных в определителе Берджи. Все тестируемые изоляты были схожи по биохимическим характеристикам и способности использовать различные соединения углерода для конструктивного и энергетического метаболизма.

Также для выделенных штаммов проводили исследование оптимума роста культуры при различных температурах и значениях pH среды, результаты приведены в табл. 5.

Для штаммов вида *Geobacillus* ***stearothermophilus*** в зависимости от источника свойства незначительно отличались, в том числе штаммы B7, B18, B22, выделенные из источника Алла, и

Таблица 4

Биохимическая активность выделенных штаммов

Активность/ субстрат	Штаммы										
	Geobacillus stearothermophilus					*Anoxybacillus* sp.	*Anoxybacillus flavithermus*				*Thermoactinomyces vulgaris*
	B7	B18	B22	B27	Gus-2-3	B25	Uro-2-1	Se-1	Ga-1-1	Gus-2-2	Gus-2-1
Сероводород	–	–	–	–	–	–	–	–	–	–	–
Лизин	+–	+–	+–	+–	+	+–	+	+	+	+	+
Орнитин	+	+–	+–	+–	+	+–	+	+	+	+	+
Уреаза	+	+	+	+	–	+	–	–	–	–	+
Аргинин	–	–	–	–	–	–	+	–	–	–	+
Цитрат Симмонса	–	–	–	–	–	–	–	–	–	–	–
Малонат	–	–	–	–	–	–	–	–	–	–	–
β-галактозидаз	+	+	+	+	+	+	+	+	+	+	+
Инозитол	–	–	–	–	–	–	–	–	–	–	–
Адонитол	–	–	–	–	–	–	–	–	–	–	–
Целлобиоза	–	–	–	–	–	–	–	–	–	–	–
Сахароза	–	–	–	–	–	–	–	–	–	–	–
Трегалоза	–	–	–	–	–	–	–	–	–	–	–
Маннитол	–	–	–	–	–	–	–	–	–	–	–
Эскулин	+	+	+	+	+	+	+	+	+	+	+
Сорбитол	–	–	–	–	–	–	–	–	–	–	–
Рамноза	–	–	–	–	–	–	–	–	–	–	–
Мелибиоза	–	–	–	–	–	–	–	–	–	–	–
Раффиноза	–	–	–	–	–	–	–	–	–	–	–
Дульцит	–	–	–	–	–	–	–	–	–	–	–
Глюкоза	–	–	–	–	–	–	–	–	–	–	–

Примечание. «+» Положительный рост; «–» отрицательный рост; «+–» слабый рост.

штамм *Geobacillus stearothermophilus* B27, выделенный из Гаргинского источника, в зависимости от температуры (40–70 °С) росли в диапазоне pH 6–10. В отличие от них, штамм *Geobacillus stearothermophilus* Gus-2-3, выделенный из источника Гусихинский, обладал способностью к росту при более низких значениях pH 5–9.

Штаммы *Anoxybacillus flavithermus* Uro-2-1 и *Anoxybacillus flavithermus* Ga-1-1, выделенные из источников Уринский и Гарга, соответственно обладают схожими свойствами (рост при pH 6–11 и температуре 40–70 °С). Штамм *Anoxybacillus flavithermus* Se-1, выделенный из источника Сеюйский, и штамм *Anoxybacillus* sp. B25, выделенный из источника Алла, имели склонность к росту в более щелочных условиях (pH 8–11) и при температуре 40–60 °С.

Штамм *Thermoactinomyces vulgaris*, исследованный в работе, имел узкий температурный диапазон роста (50–60 °С) при pH 7–10.

ОБСУЖДЕНИЕ

Ряд ранее проведенных работ по изучению состава микрофлоры термальных источников Северного Прибайкалья указывает на развитие в них микроорганизмов, обладающих различными активностями. Так, в работе М.Ю. Сусловой с соавт. (2008) показано, что из термальных источников в основном выделяются бактерии рода *Bacillus*, обладающие фосфатазной, протеиназной, липазной и другими активностями. Максимальной протеолитической активностью, как и липолитической, обладали бактерии, выделен-

Таблица 5

Исследование роста культур выделенных микроорганизмов при различных значениях pH и температуры

Штамм	Температура, °C				
	40	45	50	60	70
Geobacillus stearothermophilus B7	8*	6–8*	6–8*	6–10	7–10
Geobacillus stearothermophilus B18	8*	6–8*	6–8*	6–10	8–10
Geobacillus stearothermophilus B22	8*	6–8*	6–8*	6–10	8–10
Anoxybacillus sp. B25	8–9	8–10	7–11	8–11	–
Geobacillus stearothermophilus B27	8*	6–8*	6–8*	6–10	8–10
Anoxybacillus flavithermus Uro-2-1	7–11	7–11	7–11	7–11	8–9
*Anoxybacillus flavithermus*Se-1	8–9	8–10	8–11	8–11	–
Anoxybacillus flavithermus Ga-1-1	6–11	7–11	7–11	7–11	7–9
Thermoactinomyces vulgaris Gus-2-1	–	–	8–9	7–10	–
Anoxybacillus flavithermus Gus-2-2	6–11	7–11	7–11	7–11	7–9
Geobacillus stearothermophilus Gus-2-3	6	6–7	5–8	5–9	6–9

Примечание. * Отмечен слабый рост.

ные из источников Котельниковский и Хакусы. В работе Е.В. Лаврентьевой с соавт. (2009) также было проведено исследование состава микрофлоры термальных источников Прибайкалья. Показано, что выделенные представители также относятся к р. *Bacillus* (*B. hemicellulosolyticum*, *B. licheniformis*, *Anoxybacillus flavithermus*, *Anoxybacillus pushchinoensis*) и обладают протеазной активностью. В работе А.А. Раднагуруевой (2009) была исследована протеазная активность микроорганизмов, входящих в состав микробных матов и илов термальных источников Алла, Сея, Умхей, Гусиха и Гарга. Показано, что микроорганизмы из одной и той же станции обладают различной секрецией протеаз, максимальная из них была отмечена у бактерий источника Сея.

В данной работе было выделено и охарактеризовано 11 штаммов термофильных микроорганизмов, обладающих липолитической активностью. Филогенетический анализ родства показал, что выделенные микроорганизмы относятся к видам *Anoxybacillus flavithermus*, *Geobacillus stearothermophilus* и *Thermoactinomyces vulgaris*.

Исследование физиологических характеристик штаммов (влияние pH и температуры на рост культур) показало, что в целом штаммы *Geobacillus stearothermophilus* обладали способностью к росту в широком диапазоне pH 5–10 и температуры (до 70 °C). Штаммы р. *Anoxybacillus* росли до температуры 60–70 °C и pH 11. Выделенный штамм *Thermoactinomyces vulgaris* отличался от остальных по своим характеристикам.

Таким образом, по результатам проведенных исследований были выявлены штаммы, продуцирующие термостабильные липазы, которые являются перспективными для использования в биотехнологии, в том числе для процессов переэтерификации. Разнообразие условий роста штаммов позволяет заключить, что их липазы будут обладать широким набором полезных характеристик, в том числе термостабильностью и устойчивостью в широком диапазоне значений pH (от 5 до 11). Наиболее перспективные из изолированных штаммов-продуцентов липаз, таким образом, относятся к виду *Geobacillus stearothermophilus*.

БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации (ГК № 14.512.11.0065).

ЛИТЕРАТУРА

Герхард Ф.М. Методы общей бактериологии. Т. 2., Т. 3. М.: Мир, 1984. С. 396.

Лаврентьева Е.В., Раднагуруева А.А., Намсараев Б.Б., Дунаевский Я.Е. Биохимические характеристики микроорганизмов щелочных гидротерм Прибайкалья // Вестн. Бурят. гос. ун-та. 2009. Вып. 3. Химия, физика. С. 11–14.

Микробные сообщества щелочных гидротерм / З.Б. Намсараев и др. / Отв. ред. М.Б. Вайнштейн. Рос. акад. наук, Ин-т микробиол. им. С. Н. Виноградского, Сиб. отд-ние, Ин-т общей и эксперим. биол. Новосибирск: Изд-во СО РАН, 2006. 110 с.

Практикум по микробиологии / Под ред. А.И. Нетрусова. М.: Академия, 2005. 608 с.

Раднагуруева А.А., Лаврентьева Е.В. Внеклеточная протеазная активность в природных образцах термальных источников Прибайкалья // Изв. Иркутского гос. ун-та. 2009. Сер. Науки о Земле. Т. 2. № 2. С. 162–166.

Суслова М.Ю., Парфенова В.В., Теркина И.А. и др. Бактерии рода *Bacillus* в экосистемах горячих источников Прибайкалья и Забайкалья // Ecology and Safety. Intern. Sci. Publ. (Bulgaria). 2008. V. 2. No. 2. P. 54–60.

Boone D.R., Castenholz R.W. Bergey's manual of systematic bacteriology. 2nd ed. N.Y. a.o.: Springer-Verlag, 2001. V. 1. 721 p.

Fishman A., Basheer S., Shatzmiller S., Cogan U. Fatty-acid-modified enzymes as effective enantioselective catalysts in microaqueous organic media // Biotechnol. Lett. 1998. V. 20. No. 6. P. 535–538.

Joseph B., Ramteke P.W., Thomas G. Cold active microbial lipases: some hot issues and recent developments // Biotechnol. Adv. 2008. V. 26. No. 5. P. 457–470.

Houde A., Kademi A., Leblanc D. Lipases and their industrial applications: an overview // Appl. Biochem. Biotechnol. 2004. V. 118. No. 1/3. P. 155–170.

Tamura K., Dudley J., Nei M., Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 // Mol. Biol. Evol. 2007. V. 24. P. 1596–1599.

ISOLATION AND INVESTIGATION OF BACTERIA WITH LIPOLYTIC ACTIVITY FROM HOT SPRINGS IN THE NORTHERN BAIKAL REGION

K.N. Sorokina, A.S. Rozanov, A.V. Bryanskaya, S.E. Peltek

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: peltek@bionet.nsc.ru

Summary

We have studied properties of bacterial strains isolated from hot springs in the Northern Baikal region, Baikal Rift Zone, known to have a wide range of growth conditions (pH, temperature, and carbon sources). The phylogenetic analysis and microbiological studies show that thermophilic strains belonging to the *Geobacillus stearothermophilus*, *Anoxybacillus flavithermus*, and *Thermoactinomyces vulgaris* species express lipolytic activity. The isolated *Geobacillus stearothermophilus* strains grow at up to 70 °C in a wide pH range (5–10). The isolates of the *Anoxybacillus* genus can grow at 60–70 °C and pH ≤ 11. *Thermoactinomyces vulgaris* Gus-2-1 has a narrower growth condition range: 50–60 °C and pH 7–10. Of the strains with lipolytic activity isolated in this study, *Geobacillus stearothermophilus* is the most promising for further studies of secreted lipases.

Key words: thermophilic bacteria, lipolytic activity, Baikal Rift Zone.

УДК 579.222.7: 579.252.2

БИОИНФОРМАТИЧЕСКИЙ АНАЛИЗ ГЕНОМА ШТАММА *GEOBACILLUS STEAROTHERMOPHILUS* 22, ВЫДЕЛЕННОГО ИЗ ГОРЯЧЕГО ИСТОЧНИКА ГАРГА (ПРИБАЙКАЛЬЕ)

© 2013 г. **А.С. Розанов[1], Т.В. Иванисенко[1], А.В. Брянская[1], С.В. Шеховцов[1], М.Д. Логачева[2], О.В. Сайк[1], Т.К. Малуп[1], П.С. Деменков[1], Т.Н. Горячковская[1], В.А. Иванисенко[1], С.Е. Пельтек[1]**

[1] Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия, e-mail: peltek@bionet.nsc.ru;
[2] Московский государственный университет им. М.В. Ломоносова, Москва, Россия

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

Новый штамм *Geobacillus stearothermophilus* 22 был выделен из термального источника Гарга, расположенного в Баргузинской долине Прибайкалья. Были проанализированы морфологические и биохимические особенности *G. stearothermophilus* 22, проведено полногеномное секвенирование с последующим биоинформатическим анализом. Показана высокая степень сходства нуклеотидных последовательностей (контигов) анализируемого штамма с геномом бактерии-термофила *G. kaustophilus* Y412MC52. Охарактеризован протеом выделенной бактерии. Обнаружены ферменты, относящиеся к гемицеллюлазам (эндоксиланаза, бета-ксилозидаза, арабинофуранозидаза), и фермент эндоксиланаза.

Ключевые слова: *Geobacillus*, термофилы, полногеномное секвенирование, биоинформатический анализ.

ВВЕДЕНИЕ

Микроорганизмы, обитающие в высокотемпературных условиях, представляют огромный интерес с точки зрения получения термостабильных ферментов, а также как потенциальные клеточные катализаторы. Для разработки клеточных катализаторов перспективными являются бактерии рода *Geobacillus*, который включает широкий спектр термофильных микроорганизмов с различной физиологией. Известно, что представители рода *Geobacillus* являются хемоорганотрофами, аэробами или факультативными анаэробами, термофилами, имеющими температурный диапазон роста 40–75 °С с оптимумом 55–65 °С, диапазон pH 6,0–8,5 с оптимумом 6,2–7,5. Данные микроорганизмы представляют интерес ввиду их высокой скорости роста и способности утилизировать широкий круг субстратов, включая пентасахара. Кроме того, эти микроорганизмы обладают уникальными гемицеллюлолитическими системами, что позволяет рассматривать их в качестве потенциальных источников высокоактивных и термостабильных ферментов для эффективного гидролиза биомассы (Brock *et al.*, 1978; Бонч-Осмоловская и др., 2004).

В настоящее время опубликовано несколько работ, в которых было выполнено изменение геномов для улучшения целевых свойств микроорганизмов (Taylor *et al*., 2008). В том числе было показано, что возможна трансформация ранее нетрансформируемых или очень плохо трансформируемых бактерий после использования модифицированной ДНК, в соответствии с системой рестрикции-модификации штамма-реципиента (Suzuki, Yoshida, 2012). В этой работе было наглядно продемонстрировано, как знания о геноме бактерии могут помочь в разработке методики модификации генома

микроорганизма. Полногеномный анализ необходим в силу того, что генетические последовательности бактерий, выделенных из различных источников, могут значительно различаться по составу генов, несмотря на близость по стандартным филогенетическим маркерам. Кроме того, знание последовательности генома позволяет направленно модифицировать гены для получения штаммов-продуцентов с заданными свойствами (Cripps *et al*., 2009).

Целью данной работы являлось описание штамма *G. stearothermophilus* 22, его полногеномное секвенирование и биоинформатический анализ полученных данных для выявления особенностей катаболизма и определения генетических и белковых последовательностей, кодируемых в геноме. В настоящее время доступны полные геномные последовательности двух видов, относящихся к роду *Geobacillus: G. kaustophilus* и *G. thermodenitrificans*, а также частично расшифрованы геномы некоторых штаммов, относящихся к виду *G. stearothermophilus*.

С помощью биоинформатического анализа вновь секвенированного генома нами была установлена наибольшая степень его сходства среди всех аннотированных геномов с геномом *G. kaustophilus* Y412MC52. В геноме вновь секвенированной бактерии были идентифицированы все гены, ответственные за гликолитический метаболизм, включая гены лактат дегидрогеназы, ацетальдегид дегидрогеназы, алкоголь дегидрогеназы, ацетат киназы и пируват дегидрогеназы.

МАТЕРИАЛЫ И МЕТОДЫ

Штамм *G. stearothermophillus* 22 был выделен из проб донных отложений, отобранных в горячем источнике Гарга, расположенном в Баргузинской долине Прибайкалья. Температура воды источника на изливе достигала 75 °С на момент отбора проб. Начальное выделение штаммов из природного материала проводили на агаризованной среде Лурия-Бертани (LB). Для этого на чашки высевали по 50 мкл суспензии, культивирование проводили при температурах 60–70 °С в течение 1–3 суток. Культуры, полученные из природных образцов, очищали от сопутствующих организмов путем многократного пересева на агаризованной среде LB.

Исследование морфологии штамма проводили с использованием световых и люминесцентных микроскопов фирмы «Karl Zeiss» ЦКП микроскопического анализа биологических объектов СО РАН. Препараты готовили стандартными методами (Нетрусова, 2005).

Биохимическую характеризацию штамма проводили с использованием тест-системы Enterotest-24 (MicroTest, Lachema) в соответствии с инструкцией производителя. Штамм был проверен на способность к образованию сероводорода, уреазную активность и использование следующих субстратов: лизина, орнитина, аргинина, цитрата, малоната, инозитола, адонитола, целлобиозы, сахарозы, трегалозы, маннитола, эскулина, сорбитола, рамнозы, мелибиозы, раффинозы, дульцита, глюкозы.

Препараты ДНК для полногеномного секвенирования были получены с использованием набора genjet DNA purification kit (fermentas) в соответствии с инструкцией производителя.

Секвенирование геномной ДНК проводили при помощи прибора MiSeq фирмы «Illumina» с использованием набора реагентов Miseq reagent kit v.2 в лаборатории эволюционной геномики факультета биоинженерии и биоинформатики Московского государственного университета им. М.В. Ломоносова.

De novo ассемблирование коротких последовательностей в контиги проводилось с использованием пакета программ CLC Genomics workbench v.6.0.4, использующего алгоритм, основанный на графах де Брюйна (de Bruijn graphs). В последующем анализе использовали контиги длиной не менее 1000 нуклеотидов.

Сравнение контигов с базой нуклетидных последовательностей NT проводили с помощью BLASTN. Поиск открытых рамок считывания в контигах и соответствующих им потенциальных белков проводили с помощью BLASTX, который осуществляет трансляцию нуклеотидных последовательностей в аминокислотные и их сравнение с базой белковых последовательностей NR. С помощью BLASTX также проводили сравнение потенциальных белков, кодируемых рамками считывания, с известными белками *Geobacillus*. Для этого была сформирована выборка последовательностей белков *Geobacillus*, на основе которой была создана индексированная база данных в формате BLAST. Сравнение

аминокислотных последовательностей с базой данных последовательностей белков NR проводили с помощью программы BLASTP. В работе были использованы локальные версии программ пакета BLAST (ftp://ftp.ncbi.nlm.nih.gov/blast/).

Точную филогенетическую идентификацию полученного штамма проводили при помощи анализа 16S рРНК, для этого использовали построение филогенетического древа при помощи метода минимальной эволюции, реализованного в пакете программ MEGA 5.

Поиск типовых штаммов видов рода *Geobacillus* и соответствующих номеров последовательностей генов 16S рРНК проводили в базе данных StrainInfo (www.straininfo.net). Последовательности генов 16S рРНК брали из базы данных GenBank (www.ncbi.nlm.nih.gov/nucleotide).

Основные расчеты проводились на вычислительном кластере ЦКП «Биоинформатика» ИЦиГ СО РАН.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В 2007 г. Институтом цитологии и генетики СО РАН была проведена экспедиция на горячий источник Гаргинский, расположенный в долине р. Баргузин в Прибайкалье. В ходе экспедиции были отобраны образцы воды и донных отложений источника. В ходе выделения были получены штаммы, определенные как относящиеся к роду *Geobacillus*.

Исследуемый штамм образует округлые колонии кремового цвета. Края колоний волнистые или слегка волнистые. Профиль колоний слегка выпуклый. Размеры колоний варьируют от 3 до 5 мм. Клетки штамма представлены палочками, размеры которых составили 2–3 ×7–10 мкм (табл. 1). У исследуемого штамма выявлено спорообразование. Эндоспоры овальные или сферические 1,5 мкм (рис. 1).

Для исследуемого штамма проведено изучение биохимических характеристик. Штамм обладал способностью к росту на средах с комплексными субстратами. Спектр соединений, утилизируемых штаммом в аэробных и анаэробных условиях, представлен в табл. 2.

Установлено, что исследуемый штамм не образует сероводород и не использует большинство предложенных субстратов, однако использует глюкозу и эскулин. Штамм дает положительную реакцию на уреазу и β-галактозидазу.

Для точной видовой идентификации выделенного штамма был проведен филогенетический анализ. По предварительным данным анализа полученной в результате секвенирования последовательности гена 16S рРНК штамм был отнесен к роду *Geobacillus* (табл. 3).

Для более точной его идентификации было проведено филогенетическое сравнение с последовательностями 16S рРНК типовых штаммов видов рода *Geobacillus* (рис. 2).

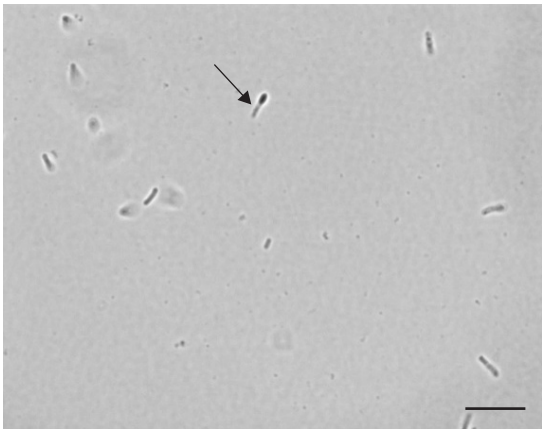


Рис. 1. Микрофотография клеток штамма *G. stearothermophilus* 22.

Клетка со спорой показана стрелкой. Масштабный отрезок – 10 мкм.

Таблица 1

Морфологическая характеристика штамма

Штамм	Морфология колоний					Морфология клеток	
	Форма	Размер, мм	Цвет	Профиль	Край	Морфотип клеток	Размеры клеток, мкм
22	круглая	3–5	кремовый	слегка выпуклый	слегка волнистый	палочки	2–3 × 7–10 споровые, d = 1,5

Таблица 2

Биохимическая характеристика штамма *G. stearothermophilus* 22

Субстрат/ Активность	Штамм 22	Субстрат/ Активность	Штамм 22
Сероводород	–	Сахароза	–
Лизин	–	Трегалоза	–
Орнитин	–	Маннитол	–
Уреаза	+	Эскулин	+
Аргинин	–	Сорбитол	–
Цитрат Симмонса	–	Рамноза	–
Малонат	–	Мелибиоза	–
β-галактозидаза	+	Раффиноза	–
Инозитол	–	Дульцит	–
Адонитол	–	Глюкоза	+
Целлобиоза	–	Индол	–
Сахароза	–	Фенилаланин	–
Целлобиоза	–	Ацетоин	–

Сравнение с последовательностями 16S рРНК типовых штаммов видов рода *Geobacillus* показало, что последовательность штамма *Geobacillus* 22 наиболее близка к последовательности типового штамма *G. stearothermophilus*. Таким образом, мы относим штамм 22 к виду *G. stearothermophilus*.

БИОИНФОРМАТИЧЕСКИЙ АНАЛИЗ

В настоящее время в базах данных отсутствует расшифрованный полный геном *G. stearothermophilus*. Сравнение контигов штамма 22 с базой данных нуклеотидных последовательностей Европейской молекулярно-биологической лаборатории EMBL (http://www.ebi.ac.uk/embl/) показало высокую степень сходства анализируемого штамма с геномом *G. kaustophilus* Y412MC52 (идентичность составила 93,6 %). В связи с этим для идентификации потенциальных белков в штамме 22 использовали

Таблица 3

Уровень сходства между последовательностью гена 16S рРНК штамма *G. stearothermophilus* 22 и последовательностями гена 16S рРНК типовых штаммов

Последовательность типового штамма	Количество идентичных нуклеотидов при попарном выравнивании (%)
G. stearothermophilus (AB021196)	1349/1352 (99)
G. subterraneus (AF276306)	1341/1352 (99)
G. subterraneus aromaticivorans (HE613733)	1336/1352 (99)
G. gargensis (FR749979)	1336/1352 (99)
G. lituanicus (AY044055)	1333/1352 (99)
G. uzenensis (AF276304)	1332/1352 (99)
G. kaustophilus (X60618)	1326/1352 (98)

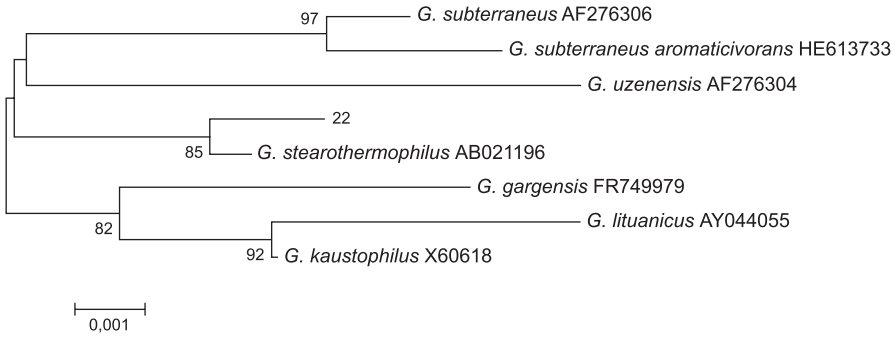


Рис. 2. Филогенетическое дерево последовательностей 16S рРНК штамма 22 и типовых штаммов видов рода *Geobacillus*, построенное методом минимальной эволюции в программе MEGA v.5.0.

Цифры возле ветвей обозначают бутстрепную поддержку.

аминокислотные последовательности охарактеризованных белков *G. kaustophilus* Y412MC52 (http://biocyc.org/GSP550542/organism-summary?object=GSP550542). Удалось охарактеризовать 3186 открытых рамок считывания из штамма 22, которые имели высокое сходство с белками *Geobacillus* sp. Y412MC52 (всего для него известно 3639 белков).

АНАЛИЗ МЕТАБОЛИЧЕСКИХ ПУТЕЙ ШТАММА 22

Большое внимание привлекает к себе проблема получения биотоплива, в частности биоэтанола, из растительного сырья. Перспективным сырьем для выработки биоэтанола является лигноцеллюлозная биомасса, которая может быть

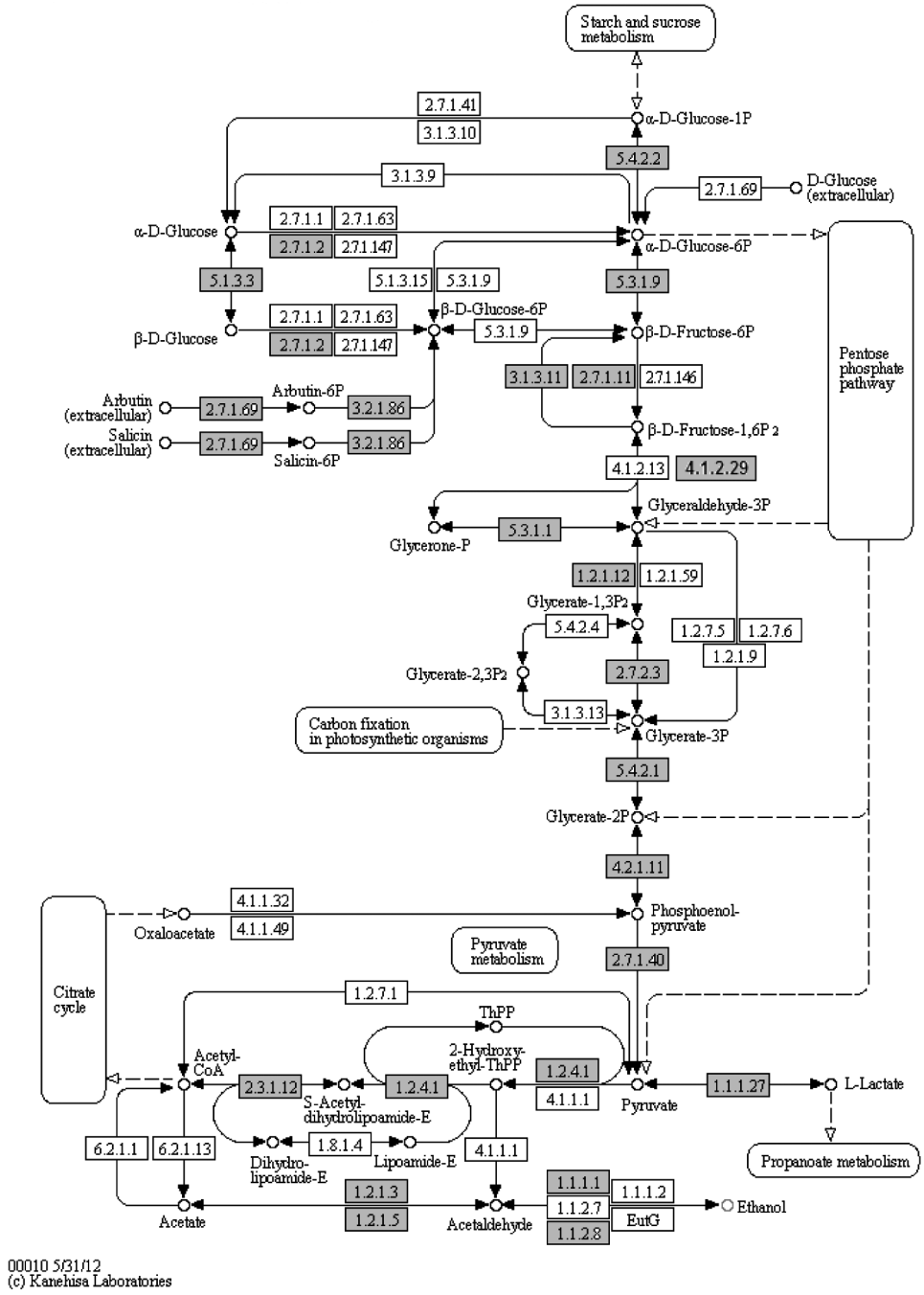


Рис. 3. Гликолитический метаболический путь, описанный в базе данных KEGG (Kanehisa, Goto, 2000; Kanehisa *et al*., 2012).

Серыми прямоугольниками отмечены ферменты, обнаруженные у штамма *G. stearothermophilus* 22.

утилизирована различными микроорганизмами (Tanimura *et al*., 2012; Zhu *et al*., 2013). Однако использование природных штаммов микроорганизмов для наработки биоэтанола в промышленном масштабе затруднено в связи с низким уровнем переработки целлюлозосодержащего сырья, низким выходом этанола и образованием большого количества побочных продуктов. Наиболее перспективным путем считается гидролиз лигноцеллюлозной биомассы до сахаров с последующей ферментацией микроорганизмами. Для получения продуцента, эффективно дающего продукт метаболизма, необходимо, чтобы используемый в работе штамм получал только один, целевой, продукт в результате катаболизма. В большинстве случаев для этого необходимо проведение модификации метаболизма природных продуцентов. Для чего необходима информация об имеющихся в клетке путях катаболизма, о последовательностях ферментов, закодированных в геноме. Таким образом, реконструкция и анализ метаболических путей могут быть основой для проведения генных модификаций, обеспечивающих получение мутантных штаммов микроорганизмов, способных с высокой эффективностью перерабатывать гидролизаты лигноцеллюлозной биомассы до этанола.

Ключевыми для наработки биоэтанола являются реакции гликолитического метаболического пути (рис. 3). В основном это цепочка превращений D-глюкозы-1-фосфат в D-глюкозу-6-фосфат, затем в бета-D-фруктозу-6-фосфат и бета-D-фруктозу-1,6-бисфосфат, далее в глицеральдегид-3-фосфат, глицерат 1,3-дифосфат, 3-фосфоглицерат и 2-фосфоглицерат, фосфоенолпируват, пируват, ацетил-CoA и затем в этанол (рис. 3). Практически все основные ферменты, ответственные за образование этанола из растительного сырья, были выявлены у вновь секвенированного штамма *G. stearothermophilus* 22 (рис. 3) на основе полученных последовательностей.

Кроме того, в геноме бактерии обнаружены фермент гидролиза целлюлозы, эндо-1,4-бета глюканаза и ферменты, участвующие в гидролизе гемицеллюлозы, – эндо-1,4-бета ксиланаза, бета ксилозидаза и альфа глюкоронидаза. Присутствие в геноме этих ферментов говорит о возможности использования данного штамма для гидролиза гемицеллюлозы, которая могла попадать в его среду обитания вместе с опадающими листьями, богатыми ксиланами.

Таким образом, у вновь секвенированного штамма *G. stearothermophilus* 22, обитающего в природном термальном источнике Гарга, расположенном в Баргузинской долине Прибайкалья, были выявлены все ключевые ферменты, ответственные за синтез этанола из растительного сырья, что открывает возможность для использования в промышленности штамма *G. stearothermophilus* 22 для производства биотоплива. Штамм обладает ограниченным набором ферментов карбогидраз, что может положительно сказаться при ферментации лигноцеллюлозных гидролизатов.

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации (ГК 14.512.11.0072 от 19.04.2013 г.) в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 гг.».

ЛИТЕРАТУРА

Бонч-Осмоловская Е.А., Мирошниченко М.Л., Соколова Т.Г., Слободкин А.И. Термофильные микробные сообщества: новые физиологические группы, новые местообитания // Тр. Ин-та микробиологии им. С.Н. Виноградского. М.: Наука, 2004. Вып. 12.

Нетрусова А.И. Практикум по микробиологии. М.: Академия, 2005. 608 с.

Brock T.D. Thermophilic microorganisms and life at high temperatures. N.Y.: Springer-Verlag, 1978. 465 p.

Cripps R.E., Eley K., Leak D.J. *et al.* Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production // Metab. Eng. 2009. V. 11. P. 398–408.

Kanehisa M., Goto S. KEGG: kyoto encyclopedia of genes and genomes // Nucl. Acids Res. 2000. V. 28. P. 27–30.

Kanehisa M., Goto S., Sato Y. *et al.* KEGG for integration and interpretation of large-scale molecular datasets // Nucl. Acids Res. 2012. V. 40. P. D109–D114.

Suzuki H., Yoshida K. Genetic transformation of *Geobacillus kaustophilus* HTA426 by conjugative transfer of host-mimicking plasmids // J. Microbiol. Biotechnol. 2012. V. 22. P. 1279–1287.

Tanimura A., Nakamura T., Watanabe I. *et al.* Isolation of a novel strain of *Candida shehatae* for ethanol production at elevated temperature // Springerplus. 2012. V. 4. P. 1–27.

Taylor M.P., Esteban C.D., Leak D.J. Development of a versatile shuttle vector for gene expression in *Geobacillus* ssp. // Plasmid 60. 2008. P. 45–52.

Zhu X., Cui J., Feng Y. *et al.* Metabolic adaption of ethanol-tolerant *Clostridium thermocellum* // PLoS ONE. 2013. V. 8. P. E70631.

BIOINFORMATIC ANALYSIS OF THE GENOME OF THE *GEOBACILLUS STEAROTHERMOPHILUS* 22 STRAIN ISOLATED FROM THE GARGA HOT SPRING, BAIKAL REGION

A.S. Rozanov[1], T.V. Ivanisenko[1], A.V. Bryanskaya[1], S.V. Shekhovtsov[1], M.D. Logacheva[2], O.V. Saik[1], T.K. Malup[1], P.S. Demenkov[1], T.N. Goryachkovskaya[1], V.A. Ivanisenko[1], S.E. Peltek[1]

[1] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: peltek@bionet.nsc.ru;
[2] Lomonosov Moscow State University, Moscow, Russia

Summary

A new strain, *Geobacillus stearothermophilus* 22 was isolated from the Garga hot spring in the Bargusin Valley, Baikal Region, Russia. The morphology and biochemistry of the strain were analyzed, and the genome-wide sequencing was conducted. The sequence was subjected to bioinformatic analysis. Nucleotide sequences (contigs) of the strain were found to be similar to the genome of the thermophilic strain *G. kaustophilus* Y412MC52. The proteome of the new strain was analyzed. Fragments associated with hemicellulases (endoxylanase, beta xylosidase, and arabinofuranosidase) and the endoxylanase enzyme were detected.

Key words: *Geobacillus*, thermophile, genome-wide sequencing, bioinformatic analysis.

УДК 577.152.311, 577.322.4

КОМПЬЮТЕРНЫЙ АНАЛИЗ СТРУКТУРЫ ЛИПАЗ БАКТЕРИЙ РОДА *GEOBACILLUS* И ВЫЯВЛЕНИЕ МОТИВОВ, ВЛИЯЮЩИХ НА ИХ ТЕРМОСТАБИЛЬНОСТЬ

© 2013 г. **К.Н. Сорокина, М.А. Нуриддинов, А.С. Розанов, В.А. Иванисенко, С.Е. Пельтек**

Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия,
e-mail: sorokina@bionet.nsc.ru

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

В работе проведен анализ свойств термостабильных липаз бактерий рода *Geobacillus*. Проведена классификация ферментов по группам термостабильности. Показано наличие согласованных аминокислотных замен у группы липаз с высокой термостабильностью (со временем полуинактивации свыше 1000 мин): V198A, Q203E, V204I, Q217E и V294I, P306A, T307A, D312S, R313H, E316G, V324I, S334N, A343T. Наибольшее достоверное влияние на термостабильность ферментов оказывал гидрофильный момент α-спиралей, который коррелировал с зарядом и полярностью аминокислотных остатков данных регионов. Показано, что у липаз с наибольшей термостабильностью происходит стабилизация «lid»-домена на структурном уровне в районе 198A-217E.

Ключевые слова: биоинформатика, липазы, термостабильность.

ВВЕДЕНИЕ

В современной биотехнологии одним из наиболее востребованных направлений является разработка новых ферментов с определенными свойствами. Так, например, ферменты липазы (EC 3.1.1.3) нашли широкое применение в различных отраслях промышленности, в том числе пищевой, фармацевтической и химической (Jaeger, Eggert, 2002). Одним из важных свойств, определяющих их применимость в тех или иных процессах, является их термостабильность. Например, термостабильные липазы используются в процессах переэтерификации пищевых жиров для получения продуктов с контролируемым составом триглицеридов.

В ряде работ показано, что липазы бактерий рода *Geobacillus* обладают высокой активностью и термостабильностью, что обусловливает их применимость для получения различных химических веществ с высокой селективностью, в том числе эфиров и амидов (Baldessari, 2012). Однако у липаз бактерий рода *Geobacillus* существует ряд существенных отличий в термостабильности ферментов, что делает их удобным объектом исследования связи структура–термостабильность и выявления функционально значимых мотивов в их структуре с использованием компьютерных подходов. С этой целью на базе аминокислотных последовательностей белков широко применяются методы, основанные на предсказании функционально значимых остатков в соответствии с консенсусным и филогенетическими подходами. Другим часто применяемым методом является статистический анализ взаимосвязи структура–активность в соответствии с физико-химическими свойствами белков (гидрофобность, заряд на поверхности и др.). Однако для наиболее точного выявления связи между определенными структурными мотивами и свойствами белков целесообразно исследовать и определять характер физико-химических взаимодействий в функционально значимых структурных мотивах белков. Этот подход к

исследованию связи структура–активность позволяет выявлять в белках отдельные районы, являющиеся существенными для проявляемых ими свойств и активности.

В данной работе для изучения связи структура–термостабильность у липаз бактерий рода *Geobacillus* был использован консенсусный подход и количественный многофакторный анализ физико-химических свойств структурных мотивов с использованием программы WebProAnalyst (Ivanisenko *et al*., 2005). Полученные данные позволили выделить мотивы и отдельные аминокислотные остатки, а также некоторые структурные особенности ферментов, значимые для их термостабильности.

МАТЕРИАЛЫ И МЕТОДЫ

Последовательности липаз, использованные в работе, и филогенетический анализ

В работе использовано 36 уникальных аминокислотных последовательностей липаз бактерий рода *Geobacilllus*, опубликованных в открытой печати. Выравнивание аминокислотных последовательностей белков проводили с использованием программы ClustualW. Построение филогентического древа было выполнено с использованием алгоритма ближайших соседей в программе MEGA5; проверку статистической достоверности проводили при помощи бутстреп теста.

Моделирование трехмерной структуры липаз

Моделирование структуры липаз для последующего анализа структуры проводили с использованием программы Modeller 9.10. В качестве шаблона для моделирования были взяты липазы бактерий: *G. stearothermophilus* P1 (PDB ID 1JI3, цепи A и B), *G. stearothermophilus* L1 (PDB ID 1KU0, цепи A и B), *G. zalihae* GZL-T1 (PDB ID 2DSN, цепи A и B) и *Geobacillus* sp. SBS-4S (PDB ID 3AUK). Визуальный анализ структур моделей, расчет и определение связей между участками ферментов выполняли в программе Swiss-PDB Viewer (Guex, Peitsch, 1997) и PyMOL (http://pymol.sourceforge.net/).

РЕЗУЛЬТАТЫ

По литературным данным, для компьютерного анализа свойств были отобраны следующие липазы бактерий: *G. lituanicus* (GLL); *G. stearothermophilus* (GSL); *Geobacillus* sp. (GspL); *G. thermocatenulatus* (GspL); *G. thermoleovorans* (GTL); *G. zalihae* (GZL) и липаза, выделенная из метагенома (ML). Первоначально была проведена классификация липаз по термостабильности, данные приведены в табл. 1. Для классификации липаз в качестве основного параметра использовали значения времен полуинактивации ($t_{1/2}$) при температурах 60, 65 и 70 °C. Однако в литературе для ряда ферментов отсутствовали некоторые данные по термостабильности. Для решения этой проблемы для некоторых липаз бактерий рода *Geobacillus* (GTcL BTL2, GTL ID-1, GZL T1 D311E, ML N355K, GSL YN) значения $t_{1/2}$ были предсказаны в соответствии со значениями гомологичных ферментов.

Липазы были разделены на группы стабильности в соответствии со значениями температурного оптимума активности и стабильности. При разделении учитывали значение $t_{1/2}$ при 60 °C, так как ему соответствовало наибольшее количество экспериментальных данных. Таким образом, всего было получено 4 группы: группа 1 ($t_{1/2}$ < 60 мин), группа 2 ($t_{1/2}$ = 60–200 мин), группа 3 ($t_{1/2}$ = 200–1000 мин), группа 4 ($t_{1/2}$ > 1000 мин). Ряд ферментов относили в группу с большей термостабильностью, если их стабильность при 70 °C была выше по сравнению с другими ферментами этой группы. Таким образом, липаза GTL ID-1 была отнесена к группе 2, GSL P1 – к группе 3, а GTL Toshki и GspL-RD2-Y224C – к группе 4.

Первоначально для определения различий между аминокислотными последовательностями термостабильных липаз, использованных в работе, был проведен филогенетический анализ, результаты приведены на рис. 1. В целом все исследуемые липазы могут быть разделены на два кластера, первый включает липазы, близкие к *G. stearothermophilus* и прочим видам; вторая группа – липазы, родственные только *G. zalihae*. Каждый из кластеров, в свою очередь, подразделяется еще на несколько групп. Однако в результате проведенного анализа не было отмечено зависимости между термоста-

Таблица 1

Значения температурных оптимумов активности
и времени полуинактивации липаз бактерий рода *Geobacillus*

Штамм	Оптимум активности, °С	$t_{1/2}$, мин			Группа стабильности
		60 °С	65 °С	70 °С	
GSL ARM 1 (Ebrahimpour *et al*., 2011)	65	0	0	0	1
GSL L1 WT (Kim *et al*., 1998)	60	90	15	10	1
GSL L1 H87A (Choi *et al*., 2005)	50	0	0	0	1
GSL L1 D61A/H87A (Choi *et al*., 2005)	45	0	0	0	1
GSL L1 D61A/H81A/H87A/D238A (Choi *et al*., 2005)	45	0	0	0	1
GspL RD-2 Y224C (Wu *et al*., 2010)	35	0	0	0	1
GspL SBS-4S (Tayyab *et al*., 2011)	60	20	0	0	1
GSL TW1 (Hebin, Xiaobo, 2005)	50	36	18	0	1
ML WT (Sharma *et al*., 2012)	50	5	0	0	1
GSL L2 (Sabri *et al*., 2009)	70	200	100	20	2
GTcL BTL2 WT (Quyen *et al*., 2003)	55	40	30*	20	2*
GTcL BTL2 F181A (Karkhane *et al*., 2009)	65	н/д	н/д	н/д	н/д
GTcL BTL2 F182A (Karkhane *et al*., 2009)	55	н/д	н/д	н/д	н/д
GTcL BTL2 F181A/F182A (Karkhane *et al*., 2009)	55	н/д	н/д	н/д	н/д
GTL CCR11 (Quintana-Castro *et al*., 2009)	60	н/д	н/д	н/д	н/д
GTL ID-1 (Cho *et al*., 2000)	75	60	42*	35	2*
GSL P1 (Tyndall *et al*., 2002)	55	300	90	45	3*
GspL RD-2 WT (Wu *et al*., 2010)	55	720	360	н/д	4
GspL RD-2 Y224P (Wu *et al*., 2010)	65	1000	480	н/д	4
GZL T1 (Ruslan *et al*., 2012)	70	600	300	30	3
GZL T1 D311E (Ruslan *et al*., 2012)	70	720	415*	110	3
ML N355K (Sharma *et al*., 2012)	40	720	400*	75	3
GspL Toshki (Abdel-Fattah, Gaballa, 2008)	65	н/д	н/д	185	4
GSL YN (Soliman *et al*., 2007)	70	1000	600*	200	4
GspL NTU03 WT (Shih, Pan, 2011)	55	>> 7000	7000	н/д	4
GspL NTU03 E189I (Shih, Pan, 2011)	55	>> 1400	1400	н/д	4
GspL NTU03 E189L (Shih, Pan, 2011)	55	>> 1400	1400	н/д	4
GspL NTU03 E189W (Shih, Pan, 2011)	55	>> 1400	1400	н/д	4
GspL NTU03 E189F (Shih, Pan, 2011)	45	н/д	н/д	н/д	н/д
GspL NTU03 E189K (Shih, Pan, 2011)	45	н/д	н/д	н/д	н/д
GspL NTU03 E189R (Shih, Pan, 2011)	50	2800	н/д	н/д	н/д
GspL NTU03 E189V (Shih, Pan, 2011)	50	1700	н/д	н/д	н/д
GspL NTU03 E189Y (Shih, Pan, 2011)	45	н/д	н/д	н/д	н/д

Примечание. * Аппроксимация значения, н/д – нет данных.

бильностью и филогенетическим положением фермента.

Из представленных в табл. 1 значений для последующего компьютерного анализа были отобраны липазы с известными значениями $t_{1/2}$ при 60 °С. Для этих белков было проведено множественное выравнивание 19 аминокислотных последовательностей зрелых форм ферментов

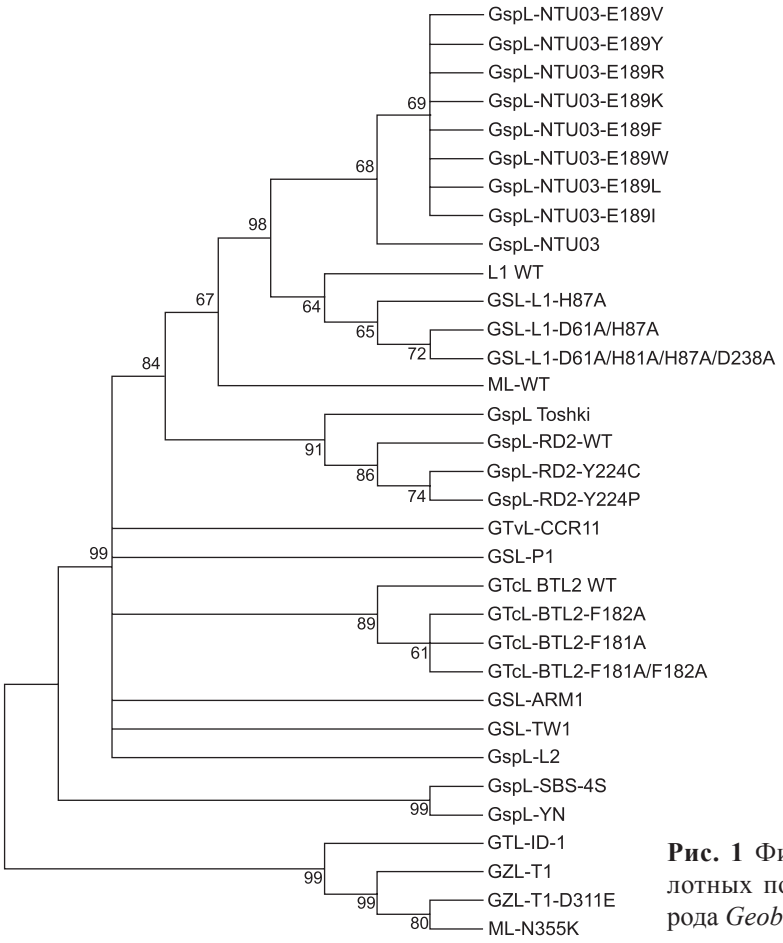


Рис. 1 Филогенетический анализ аминокислотных последовательностей липаз бактерий рода *Geobacillus*, использованных в работе.

(не содержащих сигнальный пептид). Результат приведен на рис. 2. В целом отмечено наличие согласованных аминокислотных замен для липаз бактерий рода *Geobacillus*, в частности у мотивов 198-217 AEIE (V198A, Q203E, V204I, Q217E), 248-269 PTRKS (S248P, K252T, Q255R, Q258K, A269S) и 294-343 IAASHGINT (V294I, P306A, T307A, D312S, R313H, E316G, V324I, S334N, A343T) для липаз 4-й группы термостабильности.

Для исследования связи выявленных отдельных замен и мотивов с термостабильностью ферментов был проведен компьютерный анализ их аминокислотных последовательностей с использованием программного пакета «WebProAnalyst». Для анализа была взята ранее сформированная выборка из 19 аминокислотных последовательностей липаз (см. рис. 2), в соответствие которым были поставлены их характеристические времена $t_{1/2}$ при 60 °C. Первоначально был проведен анализ значений параметра SADC, позволяющий выявить внутри аминокислотной последовательности белка сайты с достоверным уровнем корреляции физико-химических свойств и временем $t_{1/2}$ фермента. Поиск был произведен для окна в 6 аминокислотных остатков, и было выявлено, что максимальная достоверность (p = 0,001–0,003) наблюдается для корреляции значений времени полуинактивации и гидрофильности α-спиралей для следующих позиций: 188–210 (SADC=0,593), 212–223 (SADC=0,812), 300–332 (SADC = 0,695–0,868).

Далее с использованием программы WebProAnalyst был проведен многофакторный дисперсионный анализ внутригрупповой корреляции физико-химических свойств выявленных мотивов с высоким значением SADC, позволивший рассчитать значения и определить уровень достоверности посредством множественной линейной регрессии для отдельных комбинаций параметров. Для анализа использовали значение ширины окна поиска в 20 аминокислотных остатков, а в качестве параметров использовали

№	Липаза	Позиция																																		
		26	**10 8**	**14 0**	**15 4**	**15 8**	**18 1**	**18 2**	**18 7**	**19 0**	**19 8**	**20 3**	**20 4**	**20 7**	**21 9**	**22 0**	**22 5**	**22 7**	**24 8**	**25 2**	**25 5**	**25 8**	**26 9**	**27 4**	**28 2**	**29 4**	**30 6**	**30 7**	**31 2**	**31 3**	**31 6**	**32 4**	**33 4**	**34 3**	**35 5**	**36 2**	**37 2**
2	GTL-ID-1	F	I	A	H	S	F	F	A	E	V	Q	V	Q	G	E	Y	V	S	K	Q	Q	A	Y	Y	V	P	T	D	R	E	V	S	A	Y	I	D
1	GSL-TW1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1	GspL-SBS-4S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
2	GspL-L2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.	H	.	.	.	.	.	.	.	.	T	.	.	.
2	GTcL-BTL2-F181A	L	V	.	.	.	A	.	.	K	.	.	.	.	.	.	.	I	P	.	.	.	S	H	.	.	E	A	.	.	.	.	.	T	C	V	.
2	GTcL-BTL2-F181A/F182A	L	V	.	.	.	A	A	.	K	.	.	.	.	.	.	.	I	P	.	.	.	S	H	.	.	E	A	.	.	.	.	.	T	C	V	.
3	GZL-T1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.	H	.	.	.	.	.	.	.	.	T	.	.	.
3	GZL-T1-D311E	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.	H	.	.	.	E	.	.	.	.	T	.	.	.
3	GSL-P1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.	H	.	.	.	S	H	.	I	N	T	.	.	.
4	GspL-NTU03	L	.	E	R	.	.	.	.	.	A	E	I	E	.	.	.	.	P	T	R	K	S	.	.	I	A	A	S	H	G	I	N	T	.	V	N
4	GspL-NTU03-E189I	L	.	E	R	.	.	.	.	.	A	E	I	E	.	I	.	.	P	T	R	K	S	.	.	I	A	A	S	H	G	I	N	T	.	V	N
4	GspL-NTU03-E189L	L	.	E	R	.	.	.	.	.	A	E	I	E	.	L	.	.	P	T	R	K	S	.	.	I	A	A	S	H	G	I	N	T	.	V	N
4	GspL-NTU03-E189R	L	.	E	R	.	.	.	.	.	A	E	I	E	.	R	.	.	P	T	R	K	S	.	.	I	A	A	S	H	G	I	N	T	.	V	N
4	GspL-NTU03-E189V	L	.	E	R	.	.	.	.	.	A	E	I	E	.	V	.	.	P	T	R	K	S	.	.	I	A	A	S	H	G	I	N	T	.	V	N
4	GspL-NTU03-E189W	L	.	E	R	.	.	.	.	.	A	E	I	E	.	W	.	.	P	T	R	K	S	.	.	I	A	A	S	H	G	I	N	T	.	V	N
4	GspL-Toshki	.	.	.	.	.	.	.	.	.	A	E	I	E	.	.	.	.	P	T	R	K	S	.	.	I	A	A	S	H	.	I	.	.	.	.	.
4	GspL-YN	.	.	.	.	.	.	.	.	.	A	E	I	E	.	.	.	.	P	T	R	K	S	.	.	I	A	A	S	H	.	I	.	.	.	.	.
4	GspL-RD2-Y224C	.	.	.	.	.	.	.	T	.	A	E	I	E	.	.	C	.	P	T	R	K	S	.	.	I	A	A	S	H	.	I	.	.	.	.	.
4	GspL-RD2-WT	.	.	.	.	.	.	.	T	.	A	E	I	E	.	.	.	.	P	T	R	K	S	.	.	I	A	A	S	H	.	I	.	.	.	.	.

Рис. 2. Множественное выравнивание аминокислотных последовательностей липаз бактерий рода *Geobacillus*, отнесенных к различным группам термостабильности. Использованы последовательности с известными значениями времен полуинактивации при 65 °C.

* № – номер группы термостабильности.

комбинацию двух факторов: свойства (заряд и полярность) и гидрофильного момента α-спиралей. Результаты приведены в табл. 2.

Компьютерный анализ показал, что в целом наблюдается высокая степень корреляции двух параметров с гидрофильным моментом α-спирали: полярности аминокислотных остатков и их заряда. Все полученные значения корреляции свойств имели высокий уровень достоверности (p = 0,003), однако наиболее достоверные значения обоих параметров отмечены для района 321–340. Стоит отметить, что район 188–207, являющийся частью «lid»-домена липаз, имеет уровень достоверности корреляции гидрофильного момента α-спирали и заряда/полярности несколько меньший, чем для других районов.

С целью оценки связи физико-химических свойств выявленных согласованных аминокислотных замен и гидрофильного момента α-спирали был проведен аналогичный анализ для районов 198–218, 248–268, 294–335, которые по своему положению незначительно отличались от районов, предсказанных программой. Результаты расчетов представлены в табл. 3.

Как видно из приведенных данных, в целом значения уровня достоверности для данных районов также были высокими, однако чуть меньшими, чем для районов, предсказанных по значениям SACD. Для мотива 315–335 также была предсказана наибольшая достоверность по корреляции гидрофильного момента α-спирали с зарядом (p = 0,006) и полярностью (p = 0,010) аминокислотных остатков.

Таблица 2

Многофакторный дисперсионный анализ внутригрупповой корреляции физико-химических свойств мотивов с высоким значением SADC посредством программы WebProAnalyst

Фактор 1	Фактор 2	Позиция	R	F	P	SACC
Гидрофильный момент α-спирали	Полярность	188–207	0,627	4,545	0,030	0,757
		212–222	0,650	5,121	0,021	0,661
		300–319	0,625	4,488	0,031	0,667
		321–340	0,750	9,026	0,003	0,755
	Заряд	188–207	0,589	3,718	0,051	0,757
		212–222	0,661	5,432	0,018	0,661
		300–319	0,620	4,362	0,034	0,667
		321–340	0,755	9,253	0,003	0,755

Таблица 3

Многофакторный дисперсионный анализ внутригрупповой корреляции физико-химических свойств районов консервативных замен термостабильных липаз бактерий рода *Geobacillus* посредством программы WebProAnalyst

Фактор 1	Фактор 2	Позиция	R	F	P	SACC
Гидрофильный момент α-спирали	Полярность	198–218	0,661	5,434	0,018	0,661
		248–268	0,655	5,270	0,020	0,661
		294–314	0,576	3,483	0,059	0,666
		315–335	0,693	6,476	0,010	0,762
	Заряд	198–218	0,661	5,434	0,018	0,661
		248–268	0,661	5,433	0,018	0,661
		294–314	0,577	3,501	0,059	0,666
		315–335	0,720	7,536	0,006	0,762

Для исследования структурных особенностей ферментов было проведено построение трехмерной структуры термостабильных липаз *Geobacillus*. Моделирование трехмерной структуры ферментов проводилось по гомологии с известными трехмерными PDB структурами липаз бактерий рода *Geobacillus* с учетом гетероатомов – Zn^{2+} и Ca^{2+}. Кроме того, следует отметить, что, по данным рентгеновской кристаллографии, липаза GSL L1 (PDB ID 1KU0) представлена двумя цепями, поэтому полученные модели ферментов также присутствовали в двух конформациях (обозначенных A и B1). Для остальных липаз таких различий не было обнаружено, и их конформация была обозначена как B2.

По результатам структурного выравнивания ферментов, показано отличие по положению атомов основной цепи ферментов не более чем на 0,5Å, что говорит об их высокой степени гомологии. Единственное значимое отличие в структуре липаз группы 1 от липаз группы 4 было обнаружено в регионе 191–195 аминокислотного остатка и соответствует региону «lid»-домена, включающему α-спираль и прилежащему к ней району петли, структура которого приведена на рис. 3.

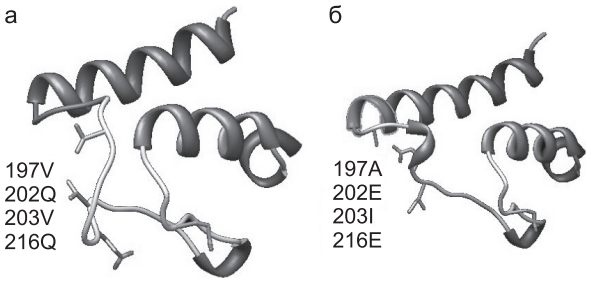


Рис. 3. Различия в третичной структуре «lid»-домена у липаз группы 1 и группы 4, по данным моделирования третичной структуры.

а – липазы группы 1; б – липазы группы 4.

ОБСУЖДЕНИЕ

Липазы бактерий рода *Geobacillus* представляют собой обширную группу ферментов, имеющих, тем не менее, различия в термостабильности. В литературе описано достаточно данных о свойствах этих ферментов, однако полученные данные ранее не подвергались системному компьютерному анализу. В целом ранее было показано, что для липаз большое значение для термостабильности имеет «lid»-домен (Chakravorty, 2011).

Проведенный филогенетический анализ аминокислотных последовательностей не позволил определить четкую связь между термостабильностью и филогенетическим положением белка. Например, ферменты с высокой термостабильностью (Gsp-NTU и его варианты, относящиеся к 4-й группе термостабильности) имели высокую степень гомологии с GSL-L1, относящимся к группе 1, и т. д. Это может быть объяснено тем, что у липаз как у ферментов с высокой термостабильностью незначительные изменения в аминокислотной последовательности приводят к значительным изменениям в структуре белка, что выражается в изменении его свойств. В том числе в этом случае происходит изменение ядер стабильности в структуре и потеря «жесткости», что приводит к значительному снижению термостабильности фермента.

Проведенная в данной работе классификация исследуемых ферментов по группам термостабильности позволила выявить ряд замен, характерных для отдельных белков с характеристическими временами $t_{1/2}$. В целом у липаз 4-й группы, имеющих наибольшую термостабильность, было отмечено наличие двух мотивов с согласованными заменами (мотив AEIE в позициях 198–217 и мотив IAASHGINT в позициях 294–343). Стоит отметить, что у липаз группы 1 данные замены не представлены, у липаз 2-й и 3-й групп представлены лишь частично. Мотив IAASHGINT у липаз 4-й группы также представлен лишь отчасти. У всех последовательностей также были выявлены отдельные аминокислотные замены с разной степенью гомологии внутри групп.

Однако выявленные по данным множественного выравнивания согласованные аминокислотные замены у группы 4 не дают представления об их вкладе во внутрибелковые физико-химические взаимодействия, поэтому далее был проведен статистический анализ взаимосвязи структуры и активности в соответствии с физико-химическими свойствами белков с использованием программы «WebProAnalyst». Наиболее достоверное влияние на взаимосвязь структуры и термостабильности (SADC > 0,5) оказывал гидрофильный

момент α-спиралей. Таким образом, показано, что характер элементов вторичной структуры, а именно наличие α-спиралей с определенными физико-химическими характеристиками, вносит существенный вклад в термостабильность белка. Также стоит отметить, что мотив с позицией 188–207 является частью «lid»-домена липазы, а остальные выявленные участки содержат выявленные ранее согласованные консервативные замены в белках.

Для определения влияния отдельных мотивов, содержащих отдельные аминокислотные замены, на термостабильность липаз был проведен многофакторный дисперсионный анализ внутригрупповой корреляции физико-химических свойств выявленных мотивов с высоким значением SADC, а также консервативных мотивов, выявленных по данным множественного выравнивания аминокислотных последовательностей. Показано, что все три выявленных консервативных района имеют высокий уровень корреляции между гидрофильным моментом α-спирали и полярностью и зарядом остатков. Наибольшее значение достоверности отмечено для района 315–335, включающего участок α16-спирали. Однако также был выявлен и высокий вклад в термостабильность ферментов участка «lid»-домена (198–218), включающего α9-спираль. Последняя особенность соответствует ранее описанным в литературе данным (Chakravorty *et al*., 2011). Для данного региона, по результатам моделирования третичной структуры, также было обнаружено, что у ферментов с заменами, характерными для липаз 4-й группы термостабильности (197A-202E-203I-216E), происходит стабилизация «lid»-домена на уровне макроструктуры. При этом в конформации A наблюдается появление дополнительного витка в α9-спирали «lid»-домена и более компактная укладка петли в районе 196–217. Несмотря на то что конформация B1 практически не отличалась от конформации B2 по своей структуре, для нее, тем не менее, было обнаружено наличие дополнительного солевого мостика 202E-109R для липаз группы 4, не формирующегося у прочих липаз, имеющих мотив 197V-202Q-203V-216Q.

Проведенное исследование позволило выявить два района согласованных замен у липаз бактерий рода *Geobacillus*, характерных для ферментов с большими временами полуинактивации. Данные замены также входят в структуру α-спиралей, чей гидрофильный момент с высоким уровнем достоверности коррелирует с зарядом и полярностью входящих в нее аминокислотных остатков. Помимо этого, свой вклад в термостабильность вносят отдельные структурные особенности, в том числе происходит стабилизация «lid»-домена за счет компактной укладки данного региона или появления дополнительного солевого мостика у наиболее термостабильных липаз. Таким образом, на термостабильность липаз бактерий рода *Geobacillus* оказывает влияние комбинация ряда факторов, в том числе физико-химические характеристики отдельных мотивов и особенности структурной организации ферментов.

БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации (ГК № 14.512.11.0065).

ЛИТЕРАТУРА

Abdel-Fattah Y.R., Gaballa A.A. Identification and overexpression of a thermostable lipase from *Geobacillus thermoleovorans* Toshki in *Escherichia coli* // Microbiol. Res. 2008. V. 163. No. 1. P. 13–20.

Chakravorty D., Parameswaran S., Dubey V.K., Patra S. *In silico* characterization of thermostable lipases // Extremophiles. 2011. V. 15. P. 89–103.

Cho A.R., Yoo S.K., Kim E.J. Cloning, sequencing and expression in *Escherichia coli* of a thermophilic lipase from *Bacillus thermoleovorans* ID-1 // FEMS Microbiol. Lett. 2000. V. 186. No. 2. P. 235–238.

Choi W.C., Kim M.H., Ro H.S. *et al.* Zinc in lipase L1 from *Geobacillus stearothermophilus* L1 and structural implications on thermal stability // FEBS Lett. 2005. V. 579. No. 16. P. 3461–3466.

Ebrahimpour A., Rahman R.N.Z.R.A., Basri M., Salleh A.B. High level expression and characterization of a novel thermostable, organic solvent tolerant, 1,3-regioselective lipase from *Geobacillus* sp. strain ARM // Bioresource Technol. 2011. V. 102. No. 13. P. 6972–6981.

Guex N., Peitsch M.C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling // Electrophoresis. 1997. V. 18. P. 2714–2723.

Hebin L., Xiaobo Z. Characterization of thermostable lipase from thermophilic *Geobacillus* sp. TW1 // Protein Expres. Purif. 2005. V. 42. No. 1. P. 153–159.

Ivanisenko V.A., Eroshkin A.M., Kolchanov N.A. WebPro-Analyst: an interactive tool for analysis of quantitative structure-activity relationships in protein families // Nucl.

Acids Res. 2005. V. 33. P. 99–104.
Jaeger K.E., Eggert T. Lipases for biotechnology // Curr. Opin. Biotechnol. 2002. V. 13. P. 390–397.
Karkhane A.A., Yakhchali B., Jazii F.R., Bambai B. The effect of substitution of Phe181 and Phe182 with Ala on activity, substrate specificity and stabilization of substrate at the active site of *Bacillus thermocatenulatus* lipase // J. Mol. Catalysis B: Enzymatic. 2009. V. 61. No. 3/4. P. 162–167.
Kim H.K., Park S.Y., Lee J.K., Oh.T.K. Gene cloning and characterization of thermostable lipase from *Bacillus stearothermophilus* L1 // Biosci., Biotechnol. Biochem. 1998. V. 62. No. 1. P. 66–71.
Quintana-Castro R., Dнaz P., Valerio-Alfaro G. *et al*. Gene cloning, expression, and characterization of the *Geobacillus thermoleovorans* CCR11 thermoalkaliphilic lipase // Mol. Biotechnol. 2009. V. 42. No. 1. P. 75–83.
Quyen D.T., Schmidt-Dannert C., Schmid R.D. High-level expression of a lipase from *Bacillus thermocatenulatus* BTL2 in *Pichia pastoris* and some properties of the recombinant lipase // Protein Expres. Purif. 2003. V. 28. No. 1. P. 102–110.
Ruslan R., Rahman R.N.Z.R.A., Leow T.C. *et al.* Improvement of thermal stability via outer-loop ion pair interaction of mutated T1 lipase from *Geobacillus zalihae* strain T1 // Intern. J. Mol. Sci. 2012. V. 13. No. 1. P. 943–960.
Sabri S., Rahman R.N.Z.R.A., Leow T.C. *et al*. Secretory expression and characterization of a highly Ca^{2+}-activated thermostable L2 lipase // Protein Expres. Purif. 2009. V. 68. No. 2. P. 161–166.
Sharma P.K., Kumar R., Kumar R. *et al*. Engineering of a metagenome derived lipase toward thermal tolerance: Effect of asparagine to lysine mutation on the protein surface // Gene. 2012. V. 491. No. 2. P. 264–271.
Shih T.W., Pan T.M. Substitution of Asp189 residue alters the activity and thermostability of *Geobacillus* sp. NTU 03 lipase // Biotechnol. Lett. 2011. V. 33. No. 9. P. 1841–1846.
Baldessari A. Lipases as catalysts in synthesis of fine chemicals // Methods Mol. Biol. 2012. V. 861. P. 445–456.
Soliman N.A., Knoll M., Abdel-Fattah Y.R. *et al.* Molecular cloning and characterization of thermostable esterase and lipase from *Geobacillus thermoleovorans* YN isolated from desert soil in Egypt // Process Biochem. 2007. V. 42. No. 7. P. 1090–1100.
Tayyab M., Rashid N., Akhtar M. Isolation and identification of lipase producing thermophilic *Geobacillus* sp. SBS-4S: Cloning and characterization of the lipase // J. Biosci. Bioengineer. 2011. V. 111. No. 3. P. 272–278.
Tyndall J.D.A., Sinchaikul S., Fothergill-Gilmore L.A., Taylor P. Crystal structure of a thermostable lipase from *Bacillus stearothermophilus* P1 // J. Mol. Biol. 2002. V. 323. No. 5. P. 859–869.
Wu L., Liu B., Hong Y. *et al.* Residue Tyr224 is critical for the thermostability of *Geobacillus* sp. RD-2 lipase // Biotechnol. Lett. 2010. V. 32. No. 1. P. 107–112.

COMPUTER ANALYSIS OF THE STRUCTURES OF LIPASES FROM *GEOBACILLUS* BACTERIA AND IDENTIFICATION OF MOTIFS DETERMINING THEIR THERMOSTABILITY

K.N. Sorokina, M.A. Nuriddinov, A.S. Rozanov, V.A. Ivanisenko, S.E. Peltek

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: k.sorokina@gmail.com

Summary

Properties of thermostable lipases from *Geobacillus* bacteria are considered. The enzymes are divided into groups with regard to their thermostability. Coordinated amino acid substitutions are demonstrated in the highly thermostable group (half-inactivation time > 1000 min); V198A, Q203E, V204I, Q217E; and V294I, P306A, T307A, D312S, R313H, E316G, V324I, S334N, A343T. The hydrophilic moment of α helices, correlating with the charge and polarity of amino acid in the region, exerts the most significant influence on enzyme thermostability. Most thermostable lipases are characterized by structural stabilization of the lid domain in the 198A-217E region.

Key words: bioinformatics, lipases, thermostability.

УДК 579.66

СОВРЕМЕННОЕ СОСТОЯНИЕ ИССЛЕДОВАНИЙ В ОБЛАСТИ ГЕНЕТИЧЕСКОЙ И МЕТАБОЛИЧЕСКОЙ ИНЖЕНЕРИИ БАКТЕРИЙ РОДА *GEOBACILLUS*, НАПРАВЛЕННЫХ НА ПОЛУЧЕНИЕ ЭТАНОЛА И ОРГАНИЧЕСКИХ КИСЛОТ

© 2013 г. **А.С. Розанов, И.А. Мещерякова, С.В. Шеховцов, С.Е. Пельтек**

Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия,
e-mail: miren@ngs.ru, asroza@gmail.com, shekhovtsov@bionet.nsc.ru, peltek@bionet.nsc.ru

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

Термофильные бактерии находят все более широкое применение в биотехнологии. Одними из наиболее перспективных термофилов являются представители рода *Geobacillus*. В статье рассмотрены известные на данный момент методики генетической и метаболической инженерии этих микроорганизмов, а также примеры их использования в различных отраслях биотехнологии.

Ключевые слова: *Geobacillus*, термофилы, биотехнология, генетическая инженерия.

ВВЕДЕНИЕ

Современные химическая, нефтехимическая и топливная отрасли промышленности базируются на использовании огромного количества органических веществ. На сегодняшний день значительная часть органических соединений добывается из ископаемых источников, что отрицательно влияет на экологию планеты. Кроме того, полезные ископаемые оказываются во все более трудноизвлекаемой форме, а объем потребления непрерывно растет, что приводит к увеличению затрат на добычу сырья и росту цен на продукты, полученные из него.

Ввиду постоянного роста цен на ископаемые источники, а также по причине развития биологических дисциплин регулярно появляются новые и развиваются уже известные биотехнологические процессы получения органических веществ. С момента возникновения промышленных производств микроорганизмы активно использовались в различных процессах, изначально это были процессы получения кисломолочных продуктов, продуктов спиртового, уксусного брожения, силосования, а также многие другие. В настоящее время актуальным направлением развития биотехнологий является разработка процессов, которые в перспективе смогут сократить использование полезных ископаемых, а в отдаленном будущем полностью исключить их использование.

РАЗВИТИЕ БИОТЕХНОЛОГИИ КЛЕТОЧНЫХ КАТАЛИЗАТОРОВ

В 20-м веке происходило бурное развитие биологических наук: общей биологии, микробиологии, генетики и молекулярной биологии. Знаковым событием для биологии как академической науки и биотехнологии как отрасли экономики стало развитие методов секвенирования. В 1970-е годы секвенирование фрагментов генов 16S рРНК показало, что настоящее разнообразие микроорганизмов многократно превышает возможности его изучения традиционными морфологическими методами (Woese *et al*., 1975). Секвенирование геномов микроорганизмов, сначала прокариотического, затем эукариотического происхождения пролило свет на процессы, происходящие в клетке.

Дальнейшее развитие технологий привело к появлению метода параллельного секвенирования и сделало возможным использование методов полногеномного секвенирования в повседневной работе отдельной лаборатории. Полученный таким образом материал о генетических последовательностях, нарастающий лавинообразно с каждым годом, дал толчок к изучению свойств белков, межгенных взаимодействий и способов регуляции в геноме. Значительный вклад в развитие биологического знания привнесло усовершенствование вычислительных алгоритмов, связанное с ростом вычислительных мощностей, в результате чего возникла отдельная наука – биоинформатика.

Накопленные знания о клеточных процессах вместе с применением методов генетической инженерии в настоящее время позволяют манипулировать метаболическими процессами клетки путем внесения изменений в ее геном. При этом, благодаря применению биоинформатических алгоритмов, возможно предварительно смоделировать наиболее перспективный вариант преобразований генома для получения целевого продукта метаболизма.

ЗАДАЧИ, СТОЯЩИЕ ПЕРЕД СОВРЕМЕННОЙ БИОТЕХНОЛОГИЕЙ

Для любых технологических процессов, в том числе биотехнологических, основными критериями успешного применения являются эффективность и экономическая обоснованность. В настоящее время одним из ключевых направлений биотехнологических исследований в области получения спирта и органических кислот является проблема использования в качестве первоначального источника вещества растительной биомассы или выделенных из нее сахаров. В отличие от традиционно используемых для этой цели зерен или плодов, состоящих преимущественно из крахмала или сахарозы, растительная биомасса состоит из целлюлозы и гемицеллюлозы, при этом до 40 % всех сахаров в ней составляют пентасахариды, в основном ксилоза. Эти сахара отсутствуют в крахмалосодержащем сырье. В связи с этим возникла задача поиска и разработки продуцентов, способных усваивать пентозы, поскольку метаболизм микроорганизмов, традиционно используемых для получения биоспирта и органических кислот, к этому не приспособлен. Другим приоритетным направлением развития биотехнологии является разработка процессов и схем, не требующих больших затрат энергии на их выполнение. Большинство биотехнологических процессов полного цикла включают высокотемпературную обработку либо на стадии подготовки сырья, либо на стадии выделения продукта. Особенно актуальна высокотемпературная обработка при использовании целлюлозы в качестве источника биоэтанола, так как ее гидролиз значительно облегчается при повышенных температурах. Выделение биоэтанола также проводится при повышенных температурах. При этом ферментирование сахаров, полученных при гидролизе лигноцеллюлозы мезофильными микроорганизмами, будет требовать сначала охлаждения, потом нагрева. Избежать необходимости постоянного охлаждения больших объемов растворов можно при проведении ферментирования с использованием термофильных микроорганизмов в качестве клеточных катализаторов. Использование термофилов позволяет проводить многие стадии биотехнологических процессов при температурах свыше 50 °С, в результате чего облегчается экстракция некоторых летучих продуктов (например спиртов) и поддержание анаэробных условий из-за меньшей растворимости кислорода. Кроме того, высокие температуры приводят к снижению вероятности контаминации. В настоящее время в этом направлении активно изучаются следующие микроорганизмы: *Clostridia*, *Thermoanaerobacter*, *Caldicellulosiruptor*, *Thermotoga*, *Pyrococcus*, *Anoxybacillus u Geobacillus* (Taylor *et al*., 2009; Verhaart *et al*., 2010; Goh *et al*., 2013). Из вышеприведенного списка особый интерес представляют бактерии рода *Geobacillus* и *Anoxybacillus*, которые, в отличие от бактерий рода *Clostridia*, способны жить в присутствии кислорода. В то же время бактерии рода *Geobacillus* в отличие от *Thermotoga* и *Pyrococcus* живут при температурах, близких к умеренным, что позволяет применять для селекции некоторые маркеры устойчивости к антибиотикам, используемые для манипуляции с мезофильными микроорганизмами. Род *Anoxybacillus* практически не изучен в плане применения методов манипуляций с геномом и применения в биотехнологии.

СИСТЕМАТИКА РОДА *GEOBACILLUS*

Представители рода *Geobacillus* являются грамположительными термофильными палочковидными бактериями, аэробными или факультативно анаэробными. Довольно долго единственным облигатным термофилом, относящимся к роду *Bacillus*, был *Bacillus stearothermophilus* (Donk, 1920). В 1949 г. Gordon и Smith (1949) провели ревизию 206 термофильных штаммов рода *Bacillus*, в результате которой 46 из них они отнесли к мезофильным видам, а остальные разделили между двумя видами, *B. stearothermophilus* и *B. coagulans*. В последующие десятилетия число видов постепенно росло. В 1993 г. White с соавт. исследовали 234 термофильных штамма рода *Bacillus* при помощи нескольких методов (фенотипические характеристики, различные физиологические тесты, гибридизация ДНК) и заключили, что их выборку можно разделить на 18 видов (White *et al*., 1993).

Методы секвенирования ДНК изменили систематику рода *Bacillus*. Ash с соавт. построили филогенетические деревья по генам 16S рРНК 51 штамма рода *Bacillus*. Род *Bacillus* оказался полифилетичным и разделился на 5 ветвей, причем термофильные его представители: *B. stearothermophilus*, *B. kaustophilus* и *B. thermoglucosidasius* были отнесены к группе 5. В результате этой и последующих работ из рода *Bacillus* было выделено более десятка новых родов (Ash *et al*., 1991).

В 2001 г. Т. Назина с соавт. выделили новый род *Geobacillus*, к которому отнесли 6 видов, ранее относимых к роду *Bacillus* (*Geobacillus stearothermophilus*, *Geobacillus thermocatenulatus*, *Geobacillus thermoleovorans*, *Geobacillus kaustophilus*, *Geobacillus thermoglucosidasius u Geobacillus thermodenitrificans*), и два новых вида, выделенных в данной работе: *Geobacillus subterraneus* и *Geobacillus uzunensis* (Nazina *et al*., 2001). Позднее к роду *Geobacillus* были отнесены несколько новых видов: *Geobacillus toebii* (Sung *et al*., 2002), *Geobacillus debilis* (Banat *et al*., 2004), *Geobacillus lithuanicus* (Kuisiene *et al*., 2004), *Geobacillus gargensis* (Nazina *et al*., 2004), еще несколько предполагаемых видов имеют неопределенный статус. Кроме того, к роду *Geobacillus* причислены из других родов: *Saccharococcus caldoxylosilytic* (Fortina *et al*., 2001), *Bacillus pallidus* (Banat *et al*., 2004), *Bacillus Vulcani* (Nazina *et al*., 2004), *Bacillus thermantarcticus* (Coorevits *et al*., 2011). В дальнейшем был выделен новый род *Aeribacillus*, в который был перенесен вид *B. pallidus* (Micana-Galbis *et al*., 2010). Coorevits с соавт. перенесли виды *Geobacillus caldoproteolyticus* и *Geobacillus tepidamans* в род *Anoxybacillus*, а вид *G. debilis* выделили в новый род *Caldibacillus* (Coorevits *et al*., 2011). Надо отметить, что валидность некоторых новых видов подвергается сомнению. Так, Dinsdale с соавт. считают, что *G. kaustophilus*, *G. lituanicus* и *G. vulcani* следует объявить синонимами *G. thermoleovorans*, а *G. gargensis* является синонимом *G. thermocatenulatus* (Dinsdale *et al*., 2011). В целом можно заключить, что систематика рода *Geobacillus* находится лишь в начале своего становления.

БИОТЕХНОЛОГИЧЕСКОЕ ПРИМЕНЕНИЕ БАКТЕРИЙ РОДА *GEOBACILLUS*

Термофильные бактерии рассматриваются в первую очередь как источник термостабильных ферментов. В последние годы значительное количество генов термофильных ферментов было выделено из геномов бактерий рода *Geobacillus* для гетерологичной экспрессии в мезофильных бактериях. Среди них, например, липазы (Abdel-Fattah, Gaballa, 2008; Quintana-Castro *et al*., 2009; Cheong *et al*., 2011; Ebrahimpour *et al*., 2011; Balan *et al*., 2012), ферменты, участвующие в расщеплении лигноцеллюлозной биомассы: β-глюкозидазы (Shallom *et al*., 2005; Ben-David *et al*., 2007; Wagschal *et al*., 2009; Ratnadewi *et al*., 2013), эндоглюканазы (Ng *et al*., 2009), ксиланазы (Wu *et al*., 2006; Canakci *et al*., 2007, 2012; Gerasimova, Kuisiene, 2012; Liu *et al*., 2012; Verma *et al*., 2013) и многие другие белки.

В последние несколько лет бактерии рода *Geobacillus* стали активно использовать в качестве клеточных катализаторов и продуцентов белка. Так, штамм *Geobacillus* sp. XT15 способен производить ацетоин и 2,3-бутандиол из различных сахаров (Xiao *et al*., 2012). *Geobacillus* sp. T1 используется как источник смеси целлюлаз, применяемой для расщепления лигноцеллюлоз-

ной биомассы (Assareh *et al.*, 2012). De Bendetti с соавт. иммобилизовали *G. stearothermophilus* CECT 43 на агарозе для получения 2,6-диаминопурин-2′-дезоксирибозида и 2,6-диаминопуринрибозида (De Bendetti *et al.*, 2012). Azfal с соавт. использовали *G. stearothermophilus* для расщепления хенодезоксихолевой кислоты (Azfal *et al.*, 2011). Модификация генома этих бактерий может открыть новые возможности для развития биотехнологических процессов.

Бактерии рода *Geobacillus* ввиду своей способности использовать различные сахара и высокой скорости роста при повышенных температурах как в аэробных, так и в анаэробных условиях представляют особый интерес в качестве продуцентов спиртов из лигноцеллюлозной биомассы. Исследования в этом направлении проводятся достаточно давно. В частности, штамм *B. stearothermophilus* LLD-R, способный расти с высокой скоростью при 70 °C, при анаэробных условиях производит в основном L-лактат и небольшие количества муравьиной и уксусной кислот, а также спирта. Путем селекции резистентного штамма на среде с флуоропируватом был получен мутантный штамм LLD-15, у которого ген L-лактатдегидрогеназы был инактивирован (Payton, Hartley, 1985). Мутантный штамм *B. stearothermophilus* LLD-R был способен производить спирт из сахарозы при 70 °C с эффективностью и скоростью, сопоставимыми с таковыми для дрожжей, традиционно используемых для получения этанола из крахмала и сахарозы (Hartley, Shama, 1987).

Geobacillus thermoglucosidasius. В другой работе была осуществлена попытка получения продуцента этанола из бактерии *G. thermoglucosidasius* путем введения в клетку гена пируватдекарбоксилазы из *Zymomonas mobilis*. Эта стратегия показала свою эффективность в создании этанологенных *Escherichia coli* (Ingram *et al.*, 1987; Alterthum, Ingram 1989) и *Bacillus megaterium* (Talarico *et al.*, 2005). Thompson с соавт. продемонстрировали, что пируватдекарбоксилаза из *Z. mobilis* может экспрессироваться в *G. thermoglucosidasius* в активной форме при температурах до 52 °C, что соответствует минимальной температуре роста этой бактерии. Хотя нативный фермент стабилен при температуре до 60 °C, сборка полноценной термостабильной пируватдекарбоксилазы в *G. thermoglucosidasius* не происходила, что не позволило использовать полученный штамм в качестве продуцента этанола (Thompson *et al.*, 2008).

В 2009 г. Cripps с соавт. опубликовали статью, в которой исследовали возможность получения продуцентов спирта на основе бактерий *G. thermoglucosidasius* NCIMB 11955 и DL33 (Cripps *et al.*, 2009). Авторы разработали систему генетической инженерии для этих штаммов и смогли внести ряд изменений в их геном. Первоначально в геноме обоих штаммов был нокаутирован ген лактатдегидрогеназы. В случае *G. thermoglucosidasius* NCIMB 11955 доминирующим продуктом стал этанол с заметным количеством примеси муравьиной и уксусной кислот, выход этанола возрос с 0,10 г/г до 0,24 г/г глюкозы. Недостатком мутантной линии стало уменьшение скорости роста и, соответственно, переработки глюкозы; время достижения максимальной концентрации этанола возросло с 6,5 до 12,5 ч. В случае *G. thermoglucosidasius* DL33 также наблюдалось заметное увеличение выхода этанола (до 0,31–0,35 г/г глюкозы) с примесью муравьиной и уксусной кислот.

Для дальнейшей работы авторы выбрали штамм *G. thermoglucosidasius* NCIMB 11955. Было показано, что лимитирующим звеном является переработка пирувата до ацетил-КоА. В связи с чем для ускорения этого процесса была проведена замена нативного промотора гена *pdhA* (пируват дегидрогеназы α) на промоторы генов *ldh* (лактат дегидрогеназы), взятых из геномов других штаммов бактерий этого же рода, а также на промотор гена *pfl* (пируватформиатлиазы) *B. cereus* ATCC14579. Во всех случаях было достигнуто заметное увеличение выхода этанола. Для снижения количества примесей в виде формиата и исключения его влияния на клетки авторы провели нокаут гена пируватформиатлиазы *pfl*. В результате в конечном продукте содержание муравьиной кислоты снизилось практически до нуля. Эта работа была выполнена на основании ранее проведенных исследований *G. thermoglucosidasius* M10EXG.

G. thermoglucosidasius M10EXG был выделен путем скрининга в компостных отложениях термофильных микроорганизмов, способных жить при высоких концентрациях спирта (Fong *et al.*, 2006). Выявленным в результате скрининга двум новым штаммам были даны названия

G. thermoglucosidasius M5EXG и M10EXG, что отражает их толерантность к спирту в концентрациях 5 и 10 % (об./об.) соответственно. Эти штаммы способны расти в диапазоне температур 50–80 °C и pH 6,0–8,0, оба штамма могут использовать различные источники углерода, включая арабинозу, галактозу, маннозу, глюкозу и ксилозу, и продуцируют небольшие количества спирта, ацетата и лактата.

Исследования профиля роста *G. thermoglucosidasius* M10EXG на различных питательных средах показали, что этот штамм способен расти на минимальной среде, содержащей глюкозу или ксилозу в качестве единственного источника углерода. *G. thermoglucosidasius* M10EXG может использовать глюкозу и ксилозу одновременно (совместное брожение), хотя при относительно низкой концентрации глюкозы потребление ксилозы снижается, особенно при добавлении дрожжевого экстракта в среду (Riyanti *et al*., 2009). Самый высокий выход биомассы (0,5 г/л) был получен на среде с глюкозой, выход возрастал при добавлении дрожжевого экстракта. Самая высокая удельная скорость роста была получена при выращивании штамма на смеси глюкозы и ксилозы (0,5 %: 0,5 % вес/объем). Диауксический рост был показан на смеси глюкозы, ксилозы и дрожжевого экстракта. Штамм производит этанол (0,1 г/л), а также (0,2 г/л) побочные продукты, L-лактат и ацетат, после 15 ч роста.

Исследования центрального метаболизма *G. thermoglucosidasius* M10EXG при различных условиях роста показали, что при аэробных условиях метаболизм глюкозы протекает через гликолиз, пентозофосфатный путь и цикл трикарбоновых кислот (ЦТК). Когда условия роста были переведены с аэробных на микроаэробные, потоки углерода в ЦТК и пентозофосфатном пути сократились примерно в два раза и были направлены на производство этанола, L-лактата (> 99 % оптической чистоты), ацетата и формиата. При полностью анаэробных условиях *G. thermoglucosidasius* M10EXG использовал смешанный процесс брожения и давал максимальный выход этанола: 0,38 ± 0,07 моль на 1 моль глюкозы. Применение моделирования потоков углерода *G. thermoglucosidasius* M10EXG *in silico* показало, что для повышения производства этанола необходимо модифицировать метаболические пути, поскольку продукция лактата и ацетата уменьшает максимальный выход этанола примерно в 3 раза (Tang *et al*., 2009).

Работу со штаммом *G. thermoglucosidasius* NCIMB 11955 продолжил Bartosiak-Jentys с соавт. Эти исследования были направлены на возможность использования этого штамма в качестве продуцента белков. Авторы разработали вектор для экспрессии и секреции гетерологичных белков в *G. thermoglucosidasius* NCIMB 11955 (Bartosiak-Jentys *et al*., 2013). Этот вектор (pUCG3.8) содержал ген устойчивости к канамицину (*knt*) и полилинкер, включающий различные сайты рестрикции. В полученную конструкцию был встроен ген эндоглюканазы Cel5A *Thermotoga maritima* под промотором гена β-глюкозидазы *G. thermoglucosidasius* NCIMB 11955. С 3′-конца гена эндоглюканазы была помещена последовательность, кодирующая сигнальный пептид β-1,4-ксиланазы *G. thermoglucosidasius* C56-YS93. Было показано, что промотор β-глюкозидазы успешно индуцируется целлобиозой и обеспечивает высокий уровень экспрессии белка.

Авторы также сделали попытку экспрессировать в pUCG3.8 укороченный ген *celA Caldicellulosiruptor saccharolyticus*, содержащий экзоглюканазный домен семейства 48 гликозидгидролаз и С-терминальный углевод-связывающий домен. В результате был достигнут лишь незначительный уровень специфической активности, что говорит или о слабой работе промотора, или о слабой секреции белка.

Так как уровень экспрессии генов под промотором β-глюкозидазы оказался неустойчивым из-за постепенного уменьшения количества целлобиозы, авторы заменили его на модифицированный конститутивный промотор гена урацилфосфорибозилтрансферазы *G. thermoglucosidasius* NCIMB 11955. В результате уровень экспрессии белка Cel5A вырос в 5 раз по сравнению с экспрессией под промотором β-галактозидазы.

МЕТОДЫ ТРАНСФОРМАЦИИ *GEOBACILLUS*

Возможность проведения манипуляций с геномом клетки открывает широкие перспективы для получения эффективных продуцентов

на основе выделенных из природы бактерий. Ключевым этапом в манипуляции с геномом микроорганизма является его трансфекция экзогенной ДНК. Для бактерий рода *Geobacillus* удалось трансформировать очень ограниченный круг штаммов, в большинстве случаев с применением достаточно сложных методов, в связи с чем данный вопрос требует отдельного, внимательного рассмотрения. За время изучения этого вопроса было опробовано достаточно большое число методов трансфекции *Geobacillus* spp.

Трансфекция протопластов

Первые работы по трансформации бактерий, отнесенных в настоящее время к роду *Geobacillus*, появились в 1980-х годах. Трансформация *B. stearothermophilus* CU21 была проведена коллективом японских авторов с применением методики трансформации протопласта в присутствии полиэтиленгликоля, описанной для *Bacillus subtilis* (Chang, Cohen, 1979), с некоторыми модификациями, учитывающими специфику роста термофильной бактерии (Imanaka *et al*., 1982). Максимальная эффективность этой процедуры составила 2×10^7 трансформантов на 1 мкг ДНК, при этом плазмиды pTB19 и pTB90 из термофильного *Bacillus spp.* могли поддерживаться и экспрессироваться в *B. stearothermophilus* до 65 °C, тогда как экспрессия pUB110 из мезофильного *Staphylococcus aureus* происходила при температурах до 55 °C.

Успешная трансформация протопластов другого штамма, *B. stearothermophilus* 1174, была проведена конструкцией, состоящей из термостабильного ориджина репликации плазмиды *B. stearothermophilus* и селективного маркера антибиотикоустойчивости из плазмиды pUB110 (Liao *et al*., 1986). Химерная плазмида была значительно более стабильна, чем pUB110, и поддерживалась при температуре до 70 °C, однако кодируемая плазмидой канамицин-нуклеотидилтрансфераза была неустойчива при температурах выше 55 °C. Исследователями был отобран мутантный вариант плазмиды, устойчивый к канамицину при температурах до 63 °C, и определены нуклеотидные замены, обеспечивающие термостабильность канамицин-нуклеотидилтрансферазы. Эффективность трансформации, достигнутая в данной работе, составила 4×10^4 трансформантов на мкг ДНК.

Эффективная система трансформации протопластов *B. stearothermophilus* NUB3621 плазмидой pTHT15 Tcr, выделенной из термофильного *Bacillus* spp., и pLW05 Cmr, сконструированной на основе плазмиды pUB110 из мезофильного *S. aureus*, была разработана Wu и Welker (1989). Эффективность трансформации этими плазмидами составила 4×10^8 и 2×10^7 трансформантов на мкг ДНК соответственно, что выше, чем для *B. stearothermophilus* CU21 (Imanaka *et al*., 1982) или *B. stearothermophilus* NRRL 1174 (Liao *et al*., 1986). Трансформация проводилась при 50 °C, что близко к минимальной температуре роста этого штамма. Однако pLW05 Cmr не могла стабильно поддерживаться при температурах свыше 50 °C, хотя белок, кодируемый геном антибиотикоустойчивости, был активен до 70 °C. Напротив, pTHT15 Tcr была стабильна в культурах, растущих при температурах более 60 °C, но белок, обеспечивающий устойчивость к тетрациклину, был относительно термолабилен при повышенных температурах.

Электропорация

Одной из первых работ, в которой был описан метод электропорации *B. stearothermophilus*, стала работа Narumi с соавт., которые выделили из образцов почвы штаммы *B. stearothermophilus* и провели скрининг на эффективность электропорации (Narumi *et al*., 1992). В результате был выделен штаммм *B. stearothermophilus* K1041, для которого провели оптимизацию условий электропорации, в результате чего эффективность этой процедуры сравнялась с эффективностью трансформации протопластов *B. stearothermophilus* CU21 плазмидой pUB110 (Imanaka *et al*., 1982). Эффективность трансформации была максимальна для клеток, находящихся в поздней экспоненциальной фазе роста при $ОД_{600} = 0{,}95$, и составила $5{,}8 \times 10^5$ на мкг pUB110. При этих условиях и трансформации рекомбинантной плазмидой pIH41, кодирующей устойчивость к хлорамфениколу, тетрациклину и канамицину, эффективность процедуры немного снижалась и составила 10^4–10^5 на мкг pIH41.

Конъюгация

Метод трансформации при помощи конъюгации широко используется для трансформации актиномицетов. Этот подход был использован Suzuki с соавт. для трансформации бактерий рода *Geobacillus* (Suzuki, Yoshida, 2012). Культуру термостабильных бактерий, в данном случае *G. kaustophilus*, легко отделить от клеток *E. coli* инкубированием при высокой температуре, что очень удобно при использовании метода.

Suzuki с соавт. описали способы трансформации бактерий *G. kaustophilus* (Suzuki, Yoshida, 2012; Suzuki *et al*., 2012, 2013). Штамм *G. kaustophilus* HTA426 был выделен из глубоководных отложений Марианской впадины, он способен к росту в аэробных условиях при высоких температурах (40–74 °C) и высокой концентрации NaCl – до 3 % w/v (Takami *et al*., 2004a). Этот штамм имеет высокую скорость роста при аэробных и анаэробных условиях, сходную с таковой для *E. coli* и *B. subtilis*, и способен использовать в качестве источника углерода широкий спектр веществ: глицерин, казаминовые кислоты, гексозы (D-глюкоза, D-галактоза, D-манноза, миоинозитол), пентозы (L-арабиноза и D-ксилоза), олигосахариды (целлобиоза, мальтоза, сахароза, растворимый крахмал, ксилоолигосахариды) и спирты (этанол, 2-пропанол, н-бутанол).

Значительное влияние на эффективность трансформации могут оказывать системы рестрикции–модификации (R-M). Для преодоления рестрикционного барьера могут быть использованы методы метилирования плазмидной ДНК либо *in vitro*, либо *in vivo*. Получение системы метилирования требует знания последовательностей ключевых ферментов. Для штамма *G. kaustophilus* HTA426 известна полная последовательность генома (Takami *et al*., 2004b). Анализ генома *G. kaustophilus* HTA426 показал, что у этого вида имеются два набора генов системы R-M типа I: GK0343 (M subunit) – GK0344 (S subunit) – GK0346 (R subunit) и GK1380 (M subunit) – GK1381 (S subunit) – GK1382 (R subunit), а также три гена системы R-M типа IV (GK1378, GK1379 и GK1390). Кроме того, плазмида pHTA426, присутствующая в клетках этого штамма, имеет один набор генов системы R-M типа II: GKP09 (эндонуклеаза) – GKP08 (метилаза). На основании полученных данных были разработаны плазмиды, содержащие гены *G. kaustophilus* HTA426 GK0343 (M-субъединица) – GK0344 (S-субъединица) и GK1380 (M-субъединица) – GK1381 (S-субъединица) под контролем промотора гена *hsp90*. Гены *G. kaustophilus*, встроенные в эти плазмиды, успешно метилировали ДНК *E. coli*. В результате экспрессии в клетках *E. coli* белков системы метилирования *G. kaustophilus* HTA426 ДНК, содержащаяся в штамме, имела тот же профиль метилирования, что и геном *G. kaustophilus*. В результате конъюгации с применением штамма *E. coli*, имеющего профиль метилирования, аналогичный профилю метилирования штамма реципиента, удалось получить очень высокий уровень трансформации с использованием плазмиды pUCG18t.

После того как была получена эффективная методика трансфекции *G. kaustophilus* HTA426, была создана система для модификации генома этой бактерии с последующим удалением маркерных генов (Suzuki *et al*., 2012b). Была использована широко применяемая в молекулярной биологии система положительного и отрицательного отбора, основанная на использовании ауксотрофии по урацилу. Ген *pyrF* кодирует оротидин-5′-фосфатдекарбоксилазу. Этот фермент также переводит 5-флуорооротат в токсичный 5-флуороуридин-5′-монофосфат. Мутанты с делецией по гену *pyrF* не способны расти на среде, не содержащей урацил, однако устойчивы к присутствию 5-флуорооротата.

При использовании этой системы селекции авторам удалось получить генетически модифицированные штаммы *G. kaustophilus*, содержащие встройки генов *bgaB* (кодирует β-галактозидазу) и *amyE* (кодирует α-амилазу). Продукты этих генов позволяют быстро детектировать наличие гетерологичной встройки. Та же группа авторов (Suzuki *et al*., 2013) идентифицировала в геноме *G. kaustophilus* HTA426 ряд промоторов, индуцируемых мальтозой, D-галактозой, целлобиозой, L-арабинозой и миоинозитолом, и показала их способность успешно активировать гетерологичный ген β-галактозидазы. При помощи вышеописанной системы для модификации генома авторы смогли встроить в геном *G. kaustophilus* HTA426 несколько гетерологичных генов: α-амилазу *G. stearothermophilus*,

предполагаемую NTPтрансферазу и фрагменты гена целлюлазы *Pyrobaculum horikoshii*, эстеразу *Pyrobaculum calidifontis*, D-лактатдегидрогеназу *Sulfolobus tokodaii* и предполагаемую азоредуктазу *G. kaustophilus*.

ПЛАЗМИДЫ И ЧЕЛНОЧНЫЕ ВЕКТОРА ДЛЯ КЛОНИРОВАНИЯ В *GEOBACILLUS*

Плазмида pUB110 наиболее широко используется для трансфекции бактерий рода *Geobacillus* и других термофильных бактерий типа Bacilli. На его основе было создано несколько векторов, несущих ориджин репликации pUB110 и ген устойчивости к канамицину.

В частности, благодаря использованию таких векторов коллективу во главе с R. Cripps удалось получить стабильные генетически модифицированные штаммы *G. thermoglucosidasius* (Cripps *et al*., 2009). Было показано, что эти вектора не способны реплицироваться в *G. thermoglucosidasius* при температуре выше 65 °C. При этом наблюдалось большое количество случаев гомологичной рекомбинации между плазмидой и бактериальными хромосомами. Таким образом, встроенная в плазмиду нужная последовательность, фланкированная длинными (около 300 п.н.) фрагментами, идентичными последовательностям бактериального генома, интегрировалась в нуклеоид *G. thermoglucosidasius*. Колонии со встройками можно отбирать по фенотипу или методом полимеразной цепной реакции.

Другая векторная система была получена Narumi с соавт. путем вставки фрагмента термофильной плазмиды pIH41, выделенной из *Bacillus* sp., в плазмиду pUC18. Челночный вектор pSTE12 благодаря термофильной вставке был способен реплицироваться в *B. stearothermophilus* K1041 и экспрессировал ген резистентности к тетрациклину. Вектор содержал 10 уникальных сайтов рестрикции из района lacI'OPZ' плазмиды pUC18. pSTE12 со встройкой фрагмента, кодирующего аспартат транскарбамилазу *E. coli*, стабильно поддерживался в *B. stearothermophilus* K1041 при селективных условиях. Этот же коллектив авторов (Narumi *et al*., 1992) разработал термостабильный вектор pSTE33, который состоит из фрагмента плазмиды *E. coli* pUC19, гена, кодирующего термостабильную канамицин-нуклеотидилтрансферазу из плазмиды pKM14, и фрагмента плазмиды pSTK1 *B. stearothermophilus*. Вектор стабильно поддерживался в *B. stearothermophilus* при 67 °C без снижения числа копий.

De Rossi с соавт. на основе маленькой криптической плазмиды из термофильного штамма *B. coagulants* Zu 196 I создали рекомбинантную плазмиду pRP9, содержащую ген устойчивости к хлорамфениколу. Полученная плазмида обладала высокой сегрегационной и структурной стабильностью в *B. stearothermophilus*. Эффективность трансформации *B. stearothermophilus* NUB3621 плазмидой pRP9 составила 4–6 × 10^5 на мкг ДНК. Было показано, что на неселективной среде плазмида стабильно поддерживается по крайней мере на протяжении 100 поколений (De Rossi *et al*., 1991).

О создании челночного вектора для метаболической инженерии *Geobacillus* spp. сообщали Taylor с соавт. На основе плазмиды pUC19, гена устойчивости к канамицину из плазмиды pUB110 и маленькой криптической плазмиды pBST1 был создан вектор, названный pUCG18, который реплицировался в клетках *E. coli* и *G. thermoglucosidasius* DL44 и был стабильным до 68 °C в присутствии селектирующего агента. Эффективность трансформации *G. thermoglucosidasius* составила 1 × 10^4 на мкг ДНК (Taylor *et al*., 2008).

На основе плазмиды pUCG18 Bartosiak-Jentys с соавт. создали систему для скрининга трансформантов *G. thermoglucosidasius*. С этой целью была сконструирована плазмида pGR002, содержащая ген катехол-2,3-диоксигеназы (*pheB*) под промотором гена лактатдегидрогеназы (*ldh*). Катехол-2,3-диоксигеназа катализирует расщепление катехольного ароматического кольца до полуальдегида 2-гидроксимуконовой кислоты. Этот продукт имеет яркий желтый цвет, что позволяет использовать его как маркер. Было показано, что наиболее высокий уровень экспрессии катехол-2,3-диоксигеназы достигается в аэробных условиях, тогда как в анаэробных условиях специфической активности в клетках не наблюдается (Bartosiak-Jentys *et al*., 2012). Позднее та же группа создала вектор для экспрессии и секреции гетерологичных белков в *G. thermoglucosidasius* NCIMB 11955. Этот вектор (pUCG3.8) содержал ген устойчивости к канамицину (*knt*) и полилинкер,

включающий различные сайты рестрикции. В полученную конструкцию был встроен ген эндоглюканазы Cel5A *T. maritima* под промотором гена β-глюкозидазы *G. thermoglucosidasius* NCIMB 11955. С 3′-конца гена эндоглюканазы был помещен сигнальный пептид β-1,4-ксиланазы *G. thermoglucosidasius* C56-YS93. Данная конструкция была клонирована в *G. thermoglucosidasius* NCIMB 11955. Было показано, что промотор β-глюкозидазы успешно индуцируется целлобиозой и обеспечивает высокий уровень экспрессии белка (Bartosiak-Jentys *et al*., 2013).

ЗАКЛЮЧЕНИЕ

Благодаря своим свойствам бактерии рода *Geobacillus* находят применение в качестве клеточных катализаторов. Также в биотехнологии используют выделенные из них ферменты. Отдельный интерес представляет возможность получения спирта напрямую из лигноцеллюлозной биомассы с использованием бактерий рода *Geobacillus*. Однако для получения эффективных продуцентов необходимо ввести в геном бактерий значительное количество модификаций с целью как оптимизации метаболизма, так и создания в клетках способности к гидролизу лигноцеллюлозной биомассы. Эти задачи требуют наличия эффективной системы трансфекции и механизмов модификации генома бактерий. Еще в 1980-х годах были найдены вектора и гены устойчивости к антибиотикам, способные поддерживаться в клетках термофильных бактерий. Были найдены штаммы бактерий рода *Geobacillus*, которые удалось трансформировать плазмидными векторами. Однако большинство штаммов оказались неспособными к трансфекции, возможно, по причине наличия рестрикционного барьера, возможно, по другим, невыясненным пока причинам.

Таким образом, можно заключить, что в области изучения свойств бактерий рода *Geobacillus* имеется значительный потенциал для получения на их основе клеточных катализаторов при помощи методов направленной инженерии.

Работа выполнена при финансовой поддержке гранта № 14.512.11.0057 Министерства образования и науки Российской Федерации.

ЛИТЕРАТУРА

Abdel-Fattah Y.R., Gaballa A.A. Identification and overexpression of a thermostable lipase from *Geobacillus thermoleovorans* Toshki in *Escherichia coli* // Microbiol. Res. 2008. V. 163. P. 13–20.

Afzal M., Oommen S., Al-Awadi S. Transformation of chenodeoxycholic acid by thermophilic *Geobacillus stearothermophilus* // Biotechnol. Appl. Biochem. 2011. V. 58. P. 250–255.

Alterthum F., Ingram L.O. Efficient ethanol production from glucose, lactose, and xylose by recombinant *Escherichia coli* // Appl. Envir. Microbiol. 1989. V. 55. P. 1943–1948.

Ash C., Farrow J.A.E., Wallbanks S., Collins M.D. Phylogenetic heterogeneity of the genus Bacillus revealed by comparative analysis of small-subunit-ribosomal RNA sequences // Lett. Appl. Microbiol. 1991. V. 13. P. 202–206.

Assareh R., Shahbani Zahiri H., Akbari Noghabi K. *et al*. Characterization of the newly isolated *Geobacillus* sp. T1, the efficient cellulase-producer on untreated barley and wheat straws // Biores. Technol. 2012. V. 120. P. 99–105.

Balan A., Ibrahim D., Abdul Rahim R., Ahmad Rashid F.A. Purification and characterization of a thermostable lipase from *Geobacillus thermodenitrificans* IBRL-nra // Enzyme Res. 2012. V. 2012. P. 987523.

Banat I.M., Marchant R., Rahman T.J. *Geobacillus debilis* sp. nov., a novel obligately thermophilic bacterium isolated from a cool soil environment, and reassignment of *Bacillus pallidus* to *Geobacillus pallidus* comb. nov. // Int. J. Syst. Evol. Microbiol. 2004. V. 54. P. 2197–2201.

Bartosiak-Jentys J., Eley K., Leak D.J. Application of pheB as a reporter gene for *Geobacillus* spp., enabling qualitative colony screening and quantitative analysis of promoter strength // Appl. Envir. Microbiol. 2012. V. 78. P. 5945–5947.

Bartosiak-Jentys J., Hussein A.H., Lewis C.J., Leak D.J. Modular system for assessment of glycosyl hydrolase secretion in *Geobacillus thermoglucosidasius* // Microbiology. 2013. V. 159. P. 1267–1275.

Ben-David A., Bravman T., Balazs Y.S. *et al*. Glycosynthase activity of *Geobacillus stearothermophilus* GH52 beta-xylosidase: efficient synthesis of xylooligosaccharides from alpha-D-xylopyranosyl fluoride through a conjugated reaction // Eur. J. Chem. Biol. 2007. V. 8. P. 2145–2151.

Canakci S., Inan K., Kacagan M., Belduz A.O. Evaluation of arabinofuranosidase and xylanase activities of *Geobacillus* spp. isolated from some hot springs in Turkey // J. Microbiol. Biotechnol. 2007. V. 17. P. 1262–1270.

Canakci S., Cevher Z., Inan K. *et al*. Cloning, purification and characterization of an alkali-stable endoxylanase from thermophilic *Geobacillus* sp. 71 // World J. Microbiol. Biotechnol. 2012. V. 28. P. 1981–1988.

Chang S., Cohen S.N. High frequency transformation of *Bacillus subtilis* protoplasts by plasmid DNA // Mol. Gen. Genet. 1979. V. 168. P. 111–115.

Cheong K.W., Leow T.C., Rahman R.N.Z.R.A. *et al*. Reductive alkylation causes the formation of a molten globule-like intermediate structure in *Geobacillus zalihae* strain T1 thermostable lipase // Appl. Biochem. Biotechnol. 2011. V. 164. P. 362–375.

Coorevits A., Logan N.A., Dinsdale A.E. *et al*. *Bacillus thermolactis* sp. nov., isolated from dairy farms, and emended

description of *Bacillus thermoamylovorans* // Int. J. Syst. Evol. Microbiol. 2011. V. 61. P. 1954–1961.

Cripps R.E., Eley K., Leak D.J. *et al*. Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production // Metab. Eng. 2009. V. 11. P. 398–408.

De Benedetti E.C., Rivero C.W., Britos C.N. *et al*. Biotransformation of 2,6-diaminopurine nucleosides by immobilized *Geobacillus stearothermophilus* // Biotechnol. Progr. 2012. V. 28. P. 1251–1256.

De Rossi E., Brigidi P., Rossi M. *et al*. Characterization of gram-positive broad host-range plasmids carrying a thermophilic replicon // Res. Microbiol. 1991. V. 142. P. 389–396.

Dinsdale A.E., Halket G., Coorevits A. *et al*. Emended descriptions of *Geobacillus thermoleovorans* and *Geobacillus thermocatenulatus* // Int. J. Syst. Evol. Microbiol. 2011. V. 61. P. 1802–1810.

Donk P.J. A highly resistant thermophilic organism // J. Bacteriol. 1920. V. 5. P. 373–374.

Ebrahimpour A., Rahman R.N., Basri M., Salleh A.B. High level expression and characterization of a novel thermostable, organic solvent tolerant, 1,3-regioselective lipase from *Geobacillus* sp. strain ARM // Biores. Technol. 2011. V. 102. P. 6972–6981.

Fong J.C.N., Svenson C.J., Nakasugi K. *et al*. Isolation and characterization of two novel ethanol-tolerant facultative-anaerobic thermophilic bacteria strains from waste compost // Extremophiles. 2006. V. 10. P. 363–372.

Fortina M.G., Mora D., Schumann P. *et al*. Reclassification of *Saccharococcus caldoxylosilyticus* as *Geobacillus caldoxylosilyticus* (Ahmad *et al*. 2000) comb. nov. // Int. J. Syst. Evol. Microbiol. 2001. V. 51. P. 2063–2071.

Gerasimova J., Kuisiene N. Characterization of the novel xylanase from the thermophilic *Geobacillus thermodenitrificans* JK1 // Mikrobiologiia. 2012. V. 81. P. 457–463.

Goh K.M., Kahar U.M., Chai Y.Y. *et al*. Recent discoveries and applications of *Anoxybacillus* // Appl. Microbiol. Biotechnol. 2013. V. 97. P. 1475–1488.

Gordon R.E., Smith N.R. Aerobic sporeforming bacteria capable of growth at high temperatures // J. Bacteriol. 1949. V. 58. P. 327–341.

Hartley B.S., Shama G. Novel ethanol fermentations from sugar cane and straw // Phil. Trans. Royal Soc. A. 1987. V. 321. P. 555–568.

Imanaka T., Fujii M., Aramori I., Aiba S. Transformation of *Bacillus stearothermophilus* with plasmid DNA and characterization of shuttle vector plasmids between *Bacillus stearothermophilus* and *Bacillus subtilis* // J. Bacteriol. 1982. V. 149. P. 824–830.

Ingram L.O., Conway T., Clark D.P. *et al*. Genetic engineering of ethanol production in *Escherichia coli* // Appl. Envir. Microbiol. 1987. V. 53. P. 2420–2425.

Kuisiene N., Raugalas J., Chitavichius D. *Geobacillus lituanicus* sp. nov. // Int. J. Syst. Evol. Microbiol. 2004. V. 54. P. 1991–1995.

Liao H., McKenzie T., Hageman R. Isolation of a thermostable enzyme variant by cloning and selection in a thermophile // Proc. Natl Acad. Sci. USA. 1986. V. 83. P. 576–580.

Liu B., Zhang N., Zhao C. *et al*. Characterization of a recombinant thermostable xylanase from hot spring thermophilic *Geobacillus* sp. TC-W7 // J. Microbiol. Biotechnol. 2012. V. 22. P. 1388–1394.

Miñana-Galbis D., Pinzón D.L., Lorén J.G. *et al*. Reclassification of *Geobacillus pallidus* (Scholz *et al*., 1988) Banat et al. 2004 as *Aeribacillus pallidus* gen. nov., comb. nov. // Int. J. Syst. Evol. Microbiol. 2010. V. 60. P. 1600–1604.

Narumi I., Sawakami K., Nakamoto S. *et al*. A newly isolated *Bacillus stearotheromophilus* K1041 and its transformation by electroporation // Biotechnol. Techn. 1992. V. 6. P. 83–86.

Nazina T.N., Tourova T.P., Poltaraus A.B. *et al*. Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans* // Int. J. Syst. Evol. Microbiol. 2001. V. 51. P. 433–446.

Nazina T.N., Lebedeva E.V., Poltaraus A.B. *et al*. *Geobacillus gargensis* sp. nov., a novel thermophile from a hot spring, and the reclassification of *Bacillus vulcani* as *Geobacillus vulcani* comb. nov. // Int. J. Syst. Evol. Microbiol. 2004. V. 54. P. 2019–2024.

Ng I.-S., Li C.-W., Yeh Y.-F. *et al*. A novel endo-glucanase from the thermophilic bacterium *Geobacillus* sp. 70PC53 with high activity and stability over a broad range of temperatures // Extremophiles. 2009. V. 13. P. 425–435.

Payton M.A., Hartley B.S. Mutants of *Bacillus stearothermophilus* lacking NAD-linked l-lactate dehydrogenase // FEMS Microbiol. Lett. 1985. V. 26. P. 333–336.

Quintana-Castro R., Díaz P., Valerio-Alfaro G. *et al*. Gene cloning, expression, and characterization of the *Geobacillus thermoleovorans* CCR11 thermoalkaliphilic lipase // Mol. Biotechnol. 2009. V. 42. P. 75–83.

Ratnadewi A.A.I., Fanani M., Kurniasih S.D. *et al*. β-D-Xylosidase from *Geobacillus thermoleovorans* IT-08: biochemical characterization and bioinformatics of the enzyme // Appl. Biochem. Biotechnol. 2013. V. 170. No. 8. P. 1950–1964.

Riyanti E.I., Rogers P.L. Kinetic evaluation of ethanol-tolerant thermophile *Geobacillus thermoglucosidasius* M10exg for ethanol production // Indones. J. Agric. Sci. 2009. V. 10. No. 1. P. 34–41.

Shallom D., Leon M., Bravman T. *et al*. Biochemical characterization and identification of the catalytic residues of a family 43 beta-D-xylosidase from *Geobacillus stearothermophilus* T-6 // Biochemistry. 2005. V. 44. P. 387–397.

Sung M.H., Kim H., Bae J.W. *et al*. *Geobacillus toebii* sp. nov., a novel thermophilic bacterium isolated from hay compost // Int. J. Syst. Evol. Microbiol. 2002. V. 52. P. 2251–2255.

Suzuki H., Yoshida K.-I. Genetic transformation of *Geobacillus kaustophilus* HTA426 by conjugative transfer of host-mimicking plasmids // J. Microbiol. Biotechnol. 2012. V. 22. P. 1279–1287.

Suzuki H., Murakami A., Yoshida K.-I. Counterselection system for *Geobacillus kaustophilus* HTA426 through disruption of pyrF and pyrR // Appl. Envir. Microbiol. 2012. V. 78. P. 7376–7383.

Suzuki H., Yoshida K.-I., Ohshima T. Polysaccharide-degrading thermophiles generated by heterologous gene expres-

sion in *Geobacillus kaustophilus* HTA426 // Appl. Envir. Microbiol. 2013. V. 79. P. 5151–5158.

Takami H., Nishi S., Lu J. *et al*. Genomic characterization of thermophilic Geobacillus species isolated from the deepest sea mud of the Mariana Trench // Extremophiles. 2004a. V. 8. P. 351–356.

Takami H., Takaki Y., Chee G.-J. *et al*. Thermoadaptation trait revealed by the genome sequence of thermophilic *Geobacillus kaustophilus* // Nucl. Acids Res. 2004b. V. 32. P. 6292–6303.

Talarico L.A., Gil M.A., Yomano L.P. *et al*. Construction and expression of an ethanol production operon in Gram-positive bacteria // Microbiol. 2005. V. 151. P. 4023–4031.

Tang Y.J., Sapra R., Joyner D. *et al*. Analysis of metabolic pathways and fluxes in a newly discovered thermophilic and ethanol-tolerant *Geobacillus* strain // Biotechnol. Bioeng. 2009. V. 102. P. 1377–1386.

Taylor M.P., Esteban C.D., Leak D.J. Development of a versatile shuttle vector for gene expression in *Geobacillus* spp. // Plasmid. 2008. V. 60. P. 45–52.

Taylor M.P., Eley K.L., Martin S. *et al*. Thermophilic ethanologenesis: future prospects for second-generation bioethanol production // Trends Biotechnol. 2009. V. 27. P. 398–405.

Thompson A.H., Studholme D.J., Green E.M., Leak D.J. Heterologous expression of pyruvate decarboxylase in *Geobacillus thermoglucosidasius* // Biotechnol. Lett. 2008. V. 30. P. 1359–1365.

Verhaart M.R.A., Bielen A.A.M., van der Oost J. *et al*. Hydrogen production by hyperthermophilic and extremely thermophilic bacteria and archaea: mechanisms for reductant disposal // Env. Technol. 2010. V. 31. P. 993–1003.

Verma D., Anand A., Satyanarayana T. Thermostable and alkalistable endoxylanase of the extremely thermophilic bacterium *Geobacillus thermodenitrificans* TSAA1: cloning, expression, characteristics and its applicability in generating xylooligosaccharides and fermentable sugars // Appl. Biochem. Biotechnol. 2013. V. 170. P. 119–130.

Wagschal K., Heng C., Lee C.C. *et al*. Purification and characterization of a glycoside hydrolase family 43 beta-xylosidase from *Geobacillus thermoleovorans* IT-08 // Appl. Biochem. Biotechnol. 2009. V. 155. P. 304–313.

White D., Sharp R.J., Priest F.G. A polyphasic taxonomic study of thermophilic bacilli from a wide geographical area // Antonie van Leeuwenhoek. 1993. V. 64. P. 357–386.

Woese C.R., Fox G.E., Zablen L. *et al*. Conservation of primary structure in 16S ribosomal RNA // Nature. 1975. V. 254. P. 83–86.

Wu L.J., Welker N.E. Protoplast transformation of *Bacillus stearothermophilus* NUB36 by plasmid DNA // J. General Microbiol. 1989. V. 135. P. 1315–1324.

Wu S., Liu B., Zhang X. Characterization of a recombinant thermostable xylanase from deep-sea thermophilic *Geobacillus* sp. MT-1 in East Pacific // Appl. Microbiol. Biotechnol. 2006. V. 72. P. 1210–1216.

Xiao Z., Wang X., Huang Y. *et al*. Thermophilic fermentation of acetoin and 2,3-butanediol by a novel *Geobacillus* strain // Biotechnol. Biofuels. 2012. V. 5. P. 88.

THE CURRENT STATE OF GENETIC AND METABOLIC ENGINEERING OF *GEOBACILLUS* BACTERIA AIMED AT THE PRODUCTION OF ETHANOL AND ORGANIC ACIDS

A.S. Rozanov, I.A. Meshcheryakova, S.V. Shekhovtsov, S.E. Peltek

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: miren@bionet.nsc.ru

Summary

Thermophilic bacteria are extensively used in biotechnology. Species of the genus *Geobacillus rank* among the most promising ones. Current methods of the genetic and metabolic engineering of these microorganisms are considered. Examples of their use in various branches of biotechnology are presented.

Key words: *Geobacillus*, thermophiles, biotechnology, genetic engineering.

УДК 57:51-76

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ СИНТЕЗА БИОЭТАНОЛА И МОЛОЧНОЙ КИСЛОТЫ ТЕРМОФИЛЬНЫМИ БАКТЕРИЯМИ РОДА *GEOBACILLUS*

© 2013 г. **М.А. Нуриддинов[1], Ф.В. Казанцев[1], А.С. Розанов[1], К.Н. Козлов[2], С.Е. Пельтек[1], Н.А. Колчанов[1, 3], И.Р. Акбердин[1]**

[1] Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия, e-mail: akberdin@bionet.nsc.ru

[2] Санкт-Петербургский государственный политехнический университет, Санкт-Петербург, Россия

[3] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

В работе представлена разработанная и адаптированная к имеющимся экспериментальным данным математическая модель биосинтеза биоэтанола и молочной кислоты в клетках *Geobacillus* spp. Показано, что математическая модель позволяет осуществлять *in silico* планирование экспериментов с бактерией *Geobacillus* spp. методами молекулярно-генетической инженерии, предсказывать динамику изменения концентрации синтезируемых биоэтанола и молочной кислоты в зависимости от молекулярно-генетических манипуляций с активностью ферментов метаболической системы.

Ключевые слова: математическое моделирование, кинетические данные, молочная кислота, биоэтанол, *Geobacillus*.

ВВЕДЕНИЕ

В настоящее время все большее внимание исследователей и биотехнологических компаний привлекают возобновляемые источники энергии. Предполагается, что наиболее перспективным источником для получения энергии и материалов в ближайшее время станут сахара из лигноцеллюлозной биомассы растений (Kasi, Ragauskas, 2010). В результате гидролиза лигноцеллюлозной биомассы могут быть получены сахара, до 50 % массы которых составляют пентасахара. Это значительная часть сырья, которой нельзя пренебрегать при разработке новой технологии переработки. Использование смеси сахаров предполагает возможность их утилизации микроорганизмами-продуцентами. Большинство продуцентов, используемых в настоящее время, разрабатывались для переработки крахмала и зачастую не способны перерабатывать пентасахара. В результате возникла потребность в поиске новых микроорганизмов, способных их заменить.

Еще одной тенденцией современной биотехнологии является использование термофильных бактерий в качестве продуцентов спиртов и органических кислот. Их использование позволяет получить ряд технологических преимуществ: облегчение поддержания температурных характеристик, снижение вязкости используемых в производстве жидкостей, снижение вероятности контаминации в результате взаимодействия с внешней средой, высокая скорость роста и процессов, которые катализируют микроорганизмы (Sonnleitner *et al.*, 1982).

Существует достаточно широкий круг микроорганизмов, обитающих в геотермальных источниках, способных к гетеротрофному питанию. В основном это прокариотические живые организмы. Среди микроорганизмов, способных жить при температуре выше 60 °С, широко распространены бактерии и археи (Cavicchioli *et al.*, 2011). Представители отдела Fermicutes широко распространены в термальных источниках. Среди них встречаются как полностью, так

и факультативно анаэробные микроорганизмы. Термофильные представители этого отдела хорошо культивируются в лабораторных условиях, в том числе и *Geobacillus* spp. – грамположительные, спорообразующие аэробные или факультативно анаэробные, широко представленные в термальных местах обитания микроорганизмы. Представители рода *Geobacillus* способны жить в температурном диапазоне от 40 до 75 °C. Большинство видов этого рода способны перерабатывать такие сахара, как D-глюкоза, D-ксилоза и L-арабиноза, в диапазоне температур 55–70 °C до смеси продуктов, содержащей лактат, формиат, ацетат и этанол. Первоначально представители этого рода относились к роду *Bacillus*, но в 2001 г. они были переклассифицированы в отдельный род ввиду наличия большого числа метаболических признаков, отличающих их от рода *Bacillus* (Nazina *et al.*, 2001). Часть представителей этого рода обладает хорошо развитым комплексом ферментов гидролиза гемицеллюлозы, что можно видеть в результате исследования аннотированных геномов этих микроорганизмов (Wu *et al.*, 2006; Feng *et al.*, 2007; Zhao *et al.*, 2012). Благодаря возможностям утилизации широкого круга сахаров, эти микроорганизмы являются перспективными продуцентами, способными использовать в качестве субстрата сахара, источником накопления которых является гидролиз лигноцеллюлозной биомассы. В литературе описаны случаи модификации представителей рода *Geobacillus* экзогенной ДНК, что делает его перспективным для применения методов направленного мутагенеза с целью получения термофильных продуцентов (Cripps *et al.*, 2009).

Бактерии широко используются в современной биотехнологии для наработки белков и метаболитов. Первоначально для этого применялись природные продуценты. Такой подход требовал широкомасштабных исследований методами скрининга природных популяций бактерий. После открытия явления мутагенеза в результате химических или физических воздействий на клетку его стали применять для изменения свойств микроорганизмов. Использование методов статистического мутагенеза позволило достичь значительных результатов по наработке целевых продуктов, как белков, так и метаболитов (Parekh *et al.*, 2000). Во второй половине XX столетия происходило поступательное развитие методов исследования и модификации генетического материала как прокариотических, так и эукариотических организмов (Kuipers *et al.*, 1999), в результате чего появились предпосылки, позволяющие направленно изменять генетический материал микроорганизмов, а вместе с ним и их метаболические свойства. Более того, значительный скачок в накоплении знаний о процессах, происходящих внутри клетки, произошел за счет создания и применения методов секвенирования геномов микроорганизмов (Keasling, 2012). Стоит отметить, что молекулярно-генетические процессы осуществляются при участии огромного количества взаимосвязанных компонентов этих систем: ферментов, регуляторных элементов, низкомолекулярных соединений и кофакторов. Выявление закономерностей, поиск новых биологических знаний на основе анализа таких объемных разнородных и разномасштабных экспериментальных данных требуют использования современных теоретических подходов, в частности метода математического моделирования. Математическая модель позволяет описывать динамику изменений внутри- и внеклеточных метаболитов в ответ на генетическую модификацию и/или изменение условий внешней среды. Такая модель позволяет предсказать *in silico* фенотипическое проявление конкретной генетической модификации (нокаут гена, модификация промотора) с целью увеличения выхода конечного продукта. На данный момент уже созданы математические модели, описывающие центральный метаболизм глюкозы для *Saccharomyces cerevisiae* (Rizzi *et al.*, 1997; Smallbone *et al.*, 2010) и *Escherichia coli* (Chassagnole *et al.*, 2002; Kadir *et al*., 2010; Peskov *et al*., 2012), а также метаболизм ксилулозы для *Lactococcus lactis* (Oshiro *et al.*, 2009), которые были успешно экспериментально верифицированы.

Объектом данного исследования является центральный метаболизм бактерии *Geobacillus* spp. Особенностью бактерий данного вида является их высокий оптимум роста, а также способность потреблять ксилозу, что делает их использование более предпочтительным в ряде технологических процессов (Tang *et al.*, 2009; Weber *et al.*, 2010). Для бактерий данного рода экспериментально показан синтез

таких внеклеточных метаболитов, как лактат (из пирувата), ацетат и этанол (из ацетил-КоА) (Cripps *et al.*, 2009).

Анализ путей центрального метаболизма глюкозы для 10 штаммов разных видов бактерий рода *Geobacillus*, представленных в базе данных KEGG, показал их сходство с метаболическими путями *E. coli*, что делает возможным использование уже построенных моделей для относительно хорошо исследованного модельного объекта в качестве стартового приближения кинетики молекулярно-генетических процессов в клетке представителя рода *Geobacillus*. Однако в ходе детального анализа данных, представленных в базе KEGG, также была выявлена и вариабельность в метаболизме пирувата у представителей этого рода. Так, для штаммов *Geobacillus thermoglucosidasius*, *Geobacillus* sp. WCH70 и *Geobacillus* sp. Y4.1MC1 показано участие фермента формиат ацетилтрансферазы в пути превращения пирувата в ацетил-КоА (http://www.genome.jp/kegg-bin/show_pathway?org_name=gwc&mapno=00620&mapscale=1.0&show_description=hide, [EC:2.3.1.54]), а у *Geobacillus thermodenitrificans* присутствует пируват ферродоксин оксидоредуктаза (http://www.genome.jp/kegg-bin/show_pathway?gtn00620, [EC:1.2.7.1]). Для штаммов *Geobacillus kaustophilus*, *Geobacillus thermoleovorans*, *Geobacillus* sp. WCH70 показано отсутствие ацетальдегиддегидрогеназы на пути превращения ацетил-КоА в ацетальдегид (http://www.genome.jp/kegg-bin/show_pathway?org_name=gka&mapno=00620&mapscale=1.0&show_description=hide, [EC:1.2.1.10]). Таким образом, в нескольких узловых точках центрального метаболизма глюкозы – метаболизме пирувата и ацетил-КоА – наблюдаются существенные различия в строении метаболических путей по сравнению со структурно-функциональной организацией одноименного пути у *E. coli*, что не может не сказаться на наработке экстраклеточных метаболитов – лактата, этанола и ацетата. Безусловно, ряд ферментативных реакций в центральном метаболизме в клетке *Geobacillus* sp. остаются пока экспериментально неидентифицированными. Также относительно недавно были опубликованы результаты экспериментального исследования структурно-функциональной организации метаболических путей и динамики потоков веществ в них при культивировании клеток *Geobacillus thermoglucosidasius* M10EX в аэробных, микроаэробных и анаэробных условиях (Tang *et al.*, 2009). Полученные результаты позволяют предположить, что существующие изменения в структурно-функциональной организации центрального метаболизма в клетке *Geobacillus* sp. по сравнению как с аналогичными процессами в клетке *E. coli*, так и между различными видами рода *Geobacillus* могут быть связаны и с особенностями условий культивирования клеток.

В соответствии с этим математическая модель центрального метаболизма бактерии *Geobacillus* spp. может позволить учесть указанные выше изменения в структурно-функциональной организации метаболического пути, выявить ключевые гены-мишени, внесение генно-инженерных модификаций в последовательности которых приводит к изменению динамики синтеза конечных продуктов, и на основе сравнительного анализа с экспериментальными кинетическими данными предсказать вариант структурно-функциональной организации метаболического пути, реализуемого у представителей этого рода.

МАТЕРИАЛЫ И МЕТОДЫ

В результате интеграции данных из KEGG (http://www.genome.jp/kegg/kegg2.html) и экспериментальных фактов об особенностях функционирования генной сети метаболизма пирувата в клетке *Geobacillus* spp. (Cripps *et al.*, 2009) была создана структурная модель биосинтеза этанола, лактата и ацетата в клетке *Geobacillus* spp. (рис. 1). В работе Криппса с соавторами был использован метод метаболической инженерии для оптимизации сети биосинтеза этанола штаммами *Geobacillus thermoglucosidasius*, мутантными по *ldh* (лактатдегидрогеназа) и *pflB* (пируват формиат лиаза) генам, и с повышенной экспрессией гена, кодирующего пируватдегидрогеназу.

В качестве субстратов «на входе» в систему используются такие метаболиты, как глюкоза и ксилоза. Структурная модель включает в себя следующие метаболические процессы: гликолиз, метаболизм пирувата, пентозофосфатный шунт и в обобщенном виде – цикл трикарбоновых кислот (рис. 2).

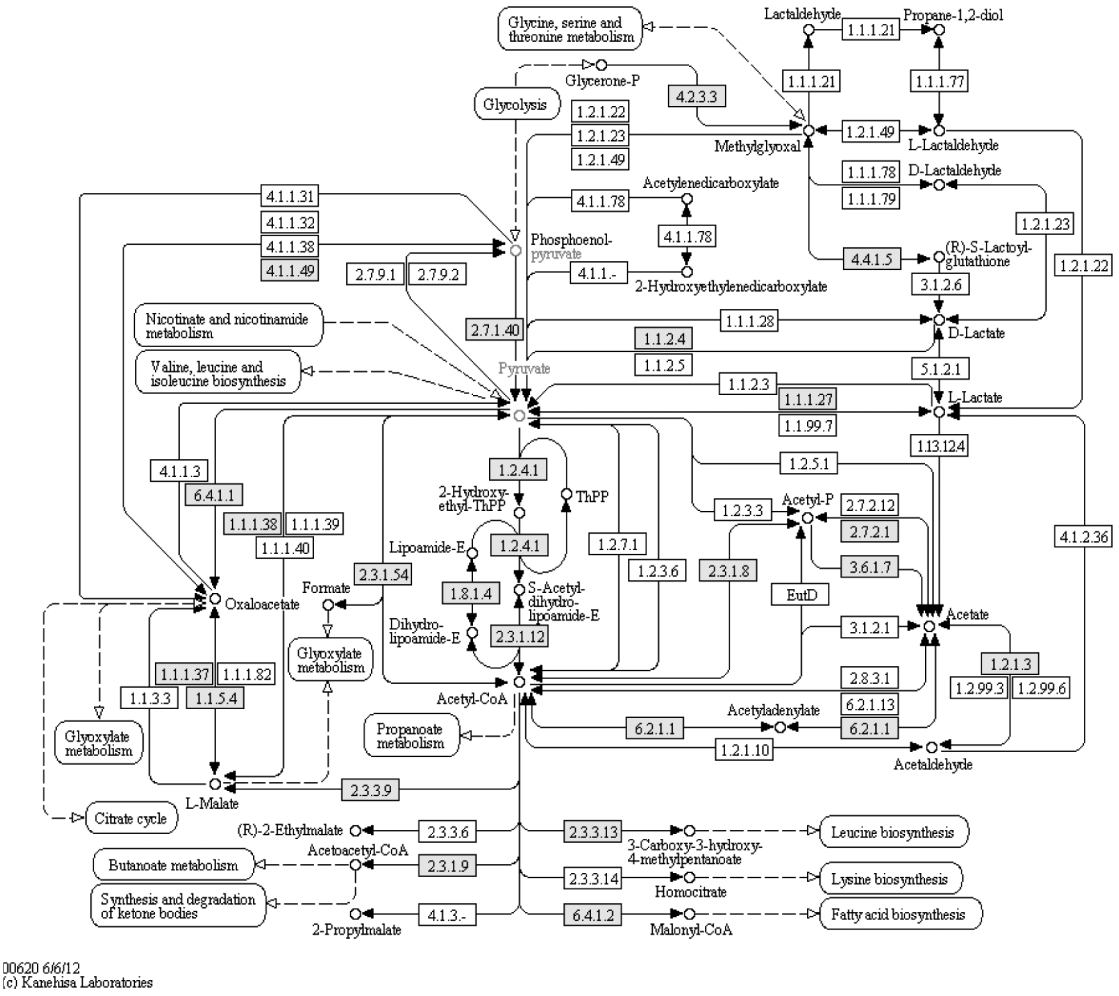


Рис. 1. Генная сеть метаболизма пирувата в клетке *Geobacillus* spp., реконструированная в системе KEGG.

Прямоугольник обозначает фермент, катализирующий конкретную реакцию в метаболизме пирувата для всех типов бактерий, а выделенные прямоугольники – именно в клетке *Geobacillus* spp. Цифры в прямоугольнике обозначают EC номер данного фермента, например [EC:1.2.4.1] – пируватдегидрогеназа.

На основе полученной структурной модели и опубликованных ранее математических моделей (Chassagnole *et al.*, 2002; Oshiro *et al.*, 2009; Kadir *et al.*, 2010; Peskov *et al.*, 2012) была реконструирована интегрированная кинетическая модель биосинтеза этанола, лактата и ацетата в клетке *Geobacillus* spp. Интегрированная кинетическая модель разработана на основе обобщенного химико-кинетического подхода с использованием обобщенных функций Хилла (Likhoshvai, Ratushny, 2007), который позволяет описывать молекулярно-генетические процессы, связывающие компоненты генной сети, системой дифференциальных уравнений, в том числе в формате компьютерной среды SciLab (https://www.scilab.org/). SciLab – свободное программное обеспечение с открытым кодом, предназначенное для реконструкции и исследования математических моделей. Данный инструмент обладает богатой функциональностью, и для него существуют версии для разных операционных систем (ОС). Таким образом, реконструированная в данном инструменте модель однозначно воспроизводит результаты на любом компьютере без применения сторонних программ конвертеров для разных ОС. Модель включает 22 переменные, одна из которых характеризует рост биомассы (формула 2.1), остальные уравнения описывают динамику изменения концентрации метаболитов (формулы 2.2–2.22), представленных на рис. 2 (например концентрацию глюкозы, пирувата, этанола), а также 127 параметров (табл. 1).

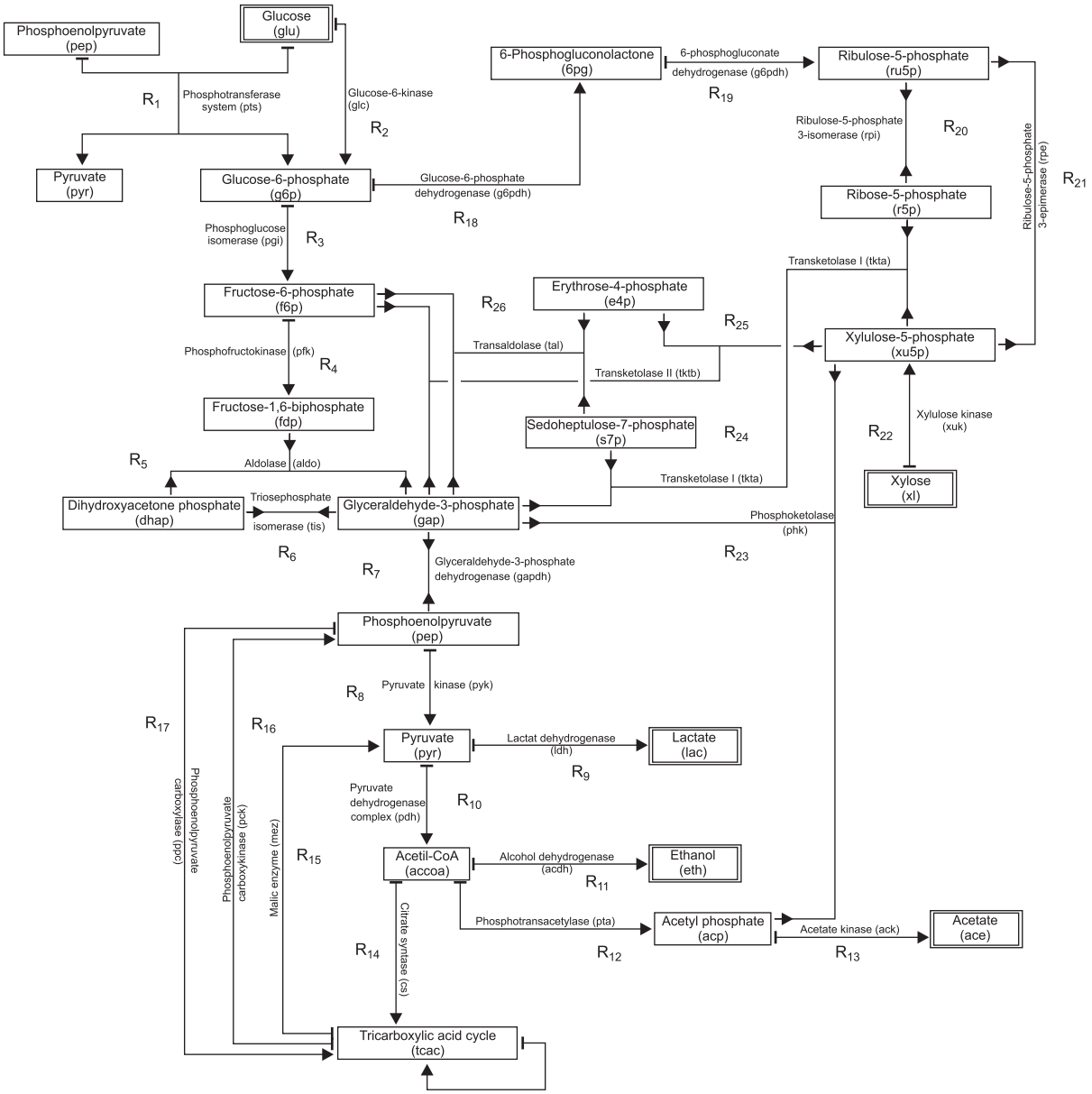


Рис. 2. Структурная модель биосинтеза этанола, лактата и ацетата в клетке *Geobacillus* spp.

Прямоугольник – конкретный метаболит; над каждой стрелкой указано полное и сокращенное название фермента, катализирующего конкретную реакцию. Двойными прямоугольниками обозначены начальные субстраты в метаболической системе (глюкоза и ксилоза) и конечные продукты (лактат, этанол и ацетат). Тупой конец стрелки (⊢) обозначает, где начинается ферментативная реакция, острый конец стрелки (➔) обозначает, во что метаболизируются субстраты реакции; острые стрелки на обоих концах – реакция обратима. R_i – порядковый номер ферментативной реакции R.

Задача поиска параметров модели, которые позволяют адекватно экспериментальным данным воспроизвести кинетику метаболических процессов, была решена с помощью интеграции SciLab и программного комплекса DEEP (http://urchin.spbcas.ru/trac/DEEP) на высокопроизводительном кластере ЦКП «Биоинформатика».

Программный комплекс DEEP был разработан и успешно применен в Санкт-Петербургском политехническом университете для решения задачи выявления генетических взаимодействий в клетках эмбриона на его ранних стадиях развития у дрозофилы (Kozlov, Samsonov, 2011). Основой комплекса является модифицированный метод дифференциальной эволюции – стохастический итеративный алгоритм многомерной математической оптимизации с применением идей генетических

Таблица 1

Параметры интегрированной кинетической модели биосинтеза этанола, лактата и ацетата в клетке *Geobacillus* spp.

Процесс	Параметр, размерность	Значение перед адаптацией модели	Значение после адаптации модели в программном комплексе DEEP
Рост культуры	m	0,6	0,6
	X_m	40	42
	V_x	1,7	1,7
	K_x	8,3	7,7
	K_{xs}	5	4,5
PTS-транспорт глюкозы	V_{pts}*, mmol/gDCW h	26	103
	K_{pts1}, mM	1	1
	K_{pts2}, mM	0,01	0,01
	K_{pts3}, mM	1	1
	K_{pts4}, mM	0,5	0,5
Не PTS-транспорт глюкозы (фосфорилирование глюкозы)	V_{glk}*, mmol/gDCW h	4,5	1,5
	K_{glkm}, mM	0,12	0,12
	K_{glkl}, mM	0,5	0,5
Глюкозофосфатизомераза	V_{pgi}*, mmol/gDCW h	26,4	167,4
	K_{pgieq}, mM	0,43	0,43
	K_{pgi1}, mM	2,5	2,5
	K_{pgi2}, mM	0,2	0,2
	K_{pgi3}, mM	0,2	0,2
	K_{pgi4}, mM	0,2	0,2
Фосфофруктокиназа	V_{pfk}*, mmol/gDCW h	25	154
	L_{pfk}	1000	1000
	K_{atp1}, mM	4,3	4,3
	K_{atp2}, mM	4,7	4,7
	K_{adp1}, mM	1	1
	$K_{adp}2$, mM	99	99
	K_{pfk1}, mM	0,15	0,15
	K_{pfk2}, mM	3,3	3,3
Альдолаза	V_{aldo}*, mmol/gDCW h	2,8	0,1
	K_{aldoeq}, mM	0,14	0,14
	V_{blf}	2	2
	K_{aldo1}, mM	0,133	0,133
	K_{aldo2}, mM	0,01	0,01
	K_{aldo3}, mM	0,01	0,01
	K_{aldo4}, mM	0,6	0,6
Триозофосфатизомераза	V_{tis}*, mmol/gDCW h	200	685
	K_{tiseq}	1,39	1,39
	K_{tis1}, mM	2,8	2,8
	K_{tis2}, mM	0,3	0,3
Глицеральдегид-3-фосфат дегидрогеназа	V_{gapdh}*, mmol/gDCW h	121	55
	$K_{gapdheq}$	0,6	0,6
	K_{gapdh1}, mM	0,15	0,15

Продолжение таблицы 1

Процесс	Параметр, размерность	Значение перед адаптацией модели	Значение после адаптации модели в программном комплексе DEEP
	K_{gapdh2}, mM	0,1	0,1
	K_{gapdh3}, mM	0,45	0,45
	K_{gapdh4}, mM	0,02	0,02
Фосфоенолпируваткарбоксилаза	V_{ppc}*, mmol/gDCW h	0,19	63,5
	K_{ppcm}, mM	0,3	0,3
	K_{ppc1}, mM	0,03	0,03
	K_{ppc2}, mM	1,3	1,3
	K_{ppc3}, mM	0,05	0,05
	K_{ppc4}, mM	0,8	0,8
	K_{ppc5}, mM	0,09	0,09
	K_{ppc6}, mM	0,27	0,27
Фосфоенолпируваткарбоксикиназа	V_{pck}*, mmol/gDCW h	4,5	80
	K_{pckm1}, mM	0,67	0,67
	K_{pckm2}, mM	0,07	0,07
	K_{pcki1}, mM	0,04	0,04
	K_{pcki2}, mM	0,04	0,04
	K_{pcki3}, mM	0,67	0,67
	K_{pcki4}, mM	0,06	0,06
Малатдегидрогеназа	V_{mez}*, mmol/gDCW h	0,07	4,2
	K_{mezeq}	0,1	0,1
	K_{mez1}, mM	0,37	0,37
Пируваткиназа	V_{pyk}*, mmol/gDCW h	1	5
	L_{pyk}	1000	1000
	K_{pyk1}, mM	0,3	0,3
	K_{pyk2}, mM	0,2	0,2
	K_{pyk3}, mM	22,5	22,5
	K_{pyk4}, mM	0,26	0,26
	K_{pyk5}, mM	0,2	0,2
Пируватдегидрогеназный комплекс	V_{pdh}*, mmol/gDCW h	27171	957
	K_{pdh1}, mM	46	46
	K_{pdh2}, mM	1	1
	K_{pdh3}, mM	0,4	0,4
	K_{pdh4}, mM	0,015	0,015
	K_{pdh5}, mM	0,1	0,1
	K_{pdh6}, mM	0,01	0,01
Лактатдегидрогеназа	V_{ldh2}*, mmol/gDCW h	2,5	2
	K_{ldh2m}, mM	0,1	0,1
	K_{ldh2a}, mM	6,4	6,4
	K_{ldh2i}, mM	3,6	3,6
Фосфотрансацетилаза	V_{pta}*, mmol/gDCW h	12,6	10,6
	K_{ptaeq}	0,03	0,03
	K_{ptai1}, mM	0,2	0,2
	K_{ptai2} , mM	0,2	0,2
	K_{ptai3}, mM	0,03	0,03

Окончание таблицы 1

Процесс	Параметр, размерность	Значение перед адаптацией модели	Значение после адаптации модели в программном комплексе DEEP
	K_{ptam}, mM	0,7	0,7
	K_{ptapm}, mM	2,6	2,6
	K_{ptapi}, mM	2,6	2,6
Алькогольдегидрогеназа	V_{acdh}*, mmol/gDCW h	5	190
	K_{acdhm}, mM	0,1	0,1
Ацетаткиназа	V_{ack}*, mmol/gDCW h	2865	39
	K_{ackeq}	174	174
	K_{ack1}, mM	0,16	0,16
	K_{ack2}, mM	7	7
	K_{ack3}, mM	0,07	0,07
	K_{ack4}, mM	0,5	0,5
Цитратсинтаза	V_{cs}*, mmol/gDCW h	17,36	1000
	K_{csm1}, mM	0,04	0,04
	K_{csm2}, mM	0,18	0,18
	K_{csd}, mM	0,1	0,1
	K_{csi1}, mM	0,00033	0,00033
	K_{csi2}, mM	0,0084	0,0084
Глюкозо-6-фосфат дегидрогеназа	V_{g6pdh}*, mmol/gDCW h	1	980
	K_{g6pdh1}, mM	1	1
	K_{g6pdh2}, mM	0,18	0,18
	K_{g6pdh3}, mM	0,015	0,015
	K_{g6pdh4}, mM	0,01	0,01
6-фосфоглюканат дегидрогеназа	V_{6pdh}*, mmol/gDCW h	1,8	423
	K_{6pdh1}, mM	0,1	0,1
	K_{6pdh2}, mM	0,03	0,03
	K_{6pdhi1}, mM	0,01	0,01
	K_{6pdhi2}, mM	3	3
Пентозофосфатэпимераза	V_{rpe}*, mmol/gDCW h	18,5	166
	K_{rpeeq}, mM	1,4	1,4
Пентозофосфатизомераза	V_{rpi}*, mmol/gDCW h	13,3	537
	K_{rpieq}, mM	4	4
Транскетолаза A	V_{tkta}*, mmol/gDCW h	29	0,1
	K_{tktaeq}, mM	1,1	1,1
Транскетолаза B	V_{tktb}*, mmol/gDCW h	316	228
	K_{tktbeq}, mM	10	10
Трансальдолаза	V_{tal}*, mmol/gDCW h	24,5	11,3
	K_{taleq}, mM	1	1
Ксилозакиназа	V_{xuk}, mmol/gDCW h	88	0,1
	K_{xukm}, mM	19	19
	K_{xuki}, mM	0,25	0,25
Фосфокетолаза	V_{phk}*, mmol/gDCW h	20	66
	K_{phkm}, mM	0,19	0,19

Примечание. Использованы значения параметров, полученные в работах ряда авторов (Chassagnole *et al*., 2002; Oshiro *et al*., 2009; Kadir *et al*., 2010) для соответствующих реакций R_1–R_{26} (рис. 2 и см. формулы 1.1–1.26).

алгоритмов (Storn, Price, 1997). В программном комплексе DEEP реализована многопоточная версия метода, обеспечивающая высокую скорость решения на многоядерных вычислительных узлах.

На первом шаге метод генерирует фиксированное количество случайных векторов параметров как начальное условие в первом поколении «популяции». После этого вычисляется вес функционала для каждой «популяции». На каждом последующем шаге происходит «скрещивание» параметров векторов как аналог обмена генами между популяциями с получением новых векторов-«популяций». Если новый вектор параметров имеет меньший вес функционала, чем родительский, то он заменяет родительский вектор в следующем поколении. Вычисления останавливаются при преодолении заданной границы функционала за фиксированное количество итераций. В качестве функции минимизации использовано эвклидово расстояние теоретических расчетов от экспериментальных, опубликованных Cripps с соавт. (2009).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Начальные значения переменных и значения параметров перед адаптацией модели к экспериментальным данным были взяты из данных ряда авторов (Chassagnole *et al.*, 2002; Oshiro *et al.*, 2009; Kadir *et al.*, 2010; Peskov *et al.*, 2012), в которых эти значения оценены или выявлены в экспериментах на клетках *Escherichia coli* и *Lactococcus lactis*. Математическая модель была адаптирована к экспериментальным данным (Cripps *et al.*, 2009, рис. 1, а): временная кинетика потребления глюкозы клеткой *Geobacillus thermoglucosidasius* и синтеза конечных продуктов (рис. 3) с помощью варьирования нескольких параметров (табл. 1, указаны*) в программном комплексе DEEP.

Начальные значения переменных (концентрации метаболитов) представлены в табл. 2.

Таблица 2

Начальные значения переменных модели биосинтеза этанола, лактата и ацетата в клетке *Geobacillus* spp. (Chassagnole *et al*., 2002; Oshiro *et al*., 2009; Kadir *et al*., 2010)

Название метаболита	Сокращенное название метаболита в модели	Начальная концентрация, mM	Название метаболита	Сокращенное название метаболита в модели	Начальная концентрация, mM
АТФ	C_{atp}	4	Фосфоенол пируват	C_{pep}	3
АДФ	C_{adp}	0,6	Пируват	C_{pyr}	3
АМФ	C_{amp}	1	Лактоза	C_{laci}	0
НАД+	C_{nad}	1,5	Ацетил-СоА	C_{accoa}	0,1
НАДН	C_{nadh}	0,1	Оксалоацетат	C_{oaa}	0,1
НАДР+	C_{nadp}	0,2	Ацетилфосфат	C_{acp}	0,1
НАДРН	C_{nadph}	0,06	Ацетат	C_{ace}	0
КоА	C_{coa}	0,001	6-фосфоглюконат	C_{6pg}	0,1
Фосфат	C_{p}	10	Рубило-5-зофосфат	C_{ru5p}	0,1
Концентрация клеток в культуре	X	0,1	Рибозо-5-фосфат	C_{r5p}	0,1
Глюкоза экстраклеточная	C_{glc}	190	Ксилоза экстраклеточная	C_{xl}	0
Глюкоза-6-фосфат	C_{g6p}	0,1	Ксилоза внутриклеточная	C_{xli}	0,1
Фруктоза-6-фосфат	C_{f6p}	0,1	Ксилулоза-5-фосфат	C_{xu5p}	0,1
Фруктоза-1,6-дифосфат	C_{fdp}	0,1	Седогептулоза-7-фосфат	C_{s7p}	0,1
Глицеральдегид фосфат	C_{gap}	0,1	Эритроза-4-фосфат	C_{e4p}	0,1
Дегидроксиацетон фосфат	C_{dhap}	0,1	Этанол	C_{eth}	0

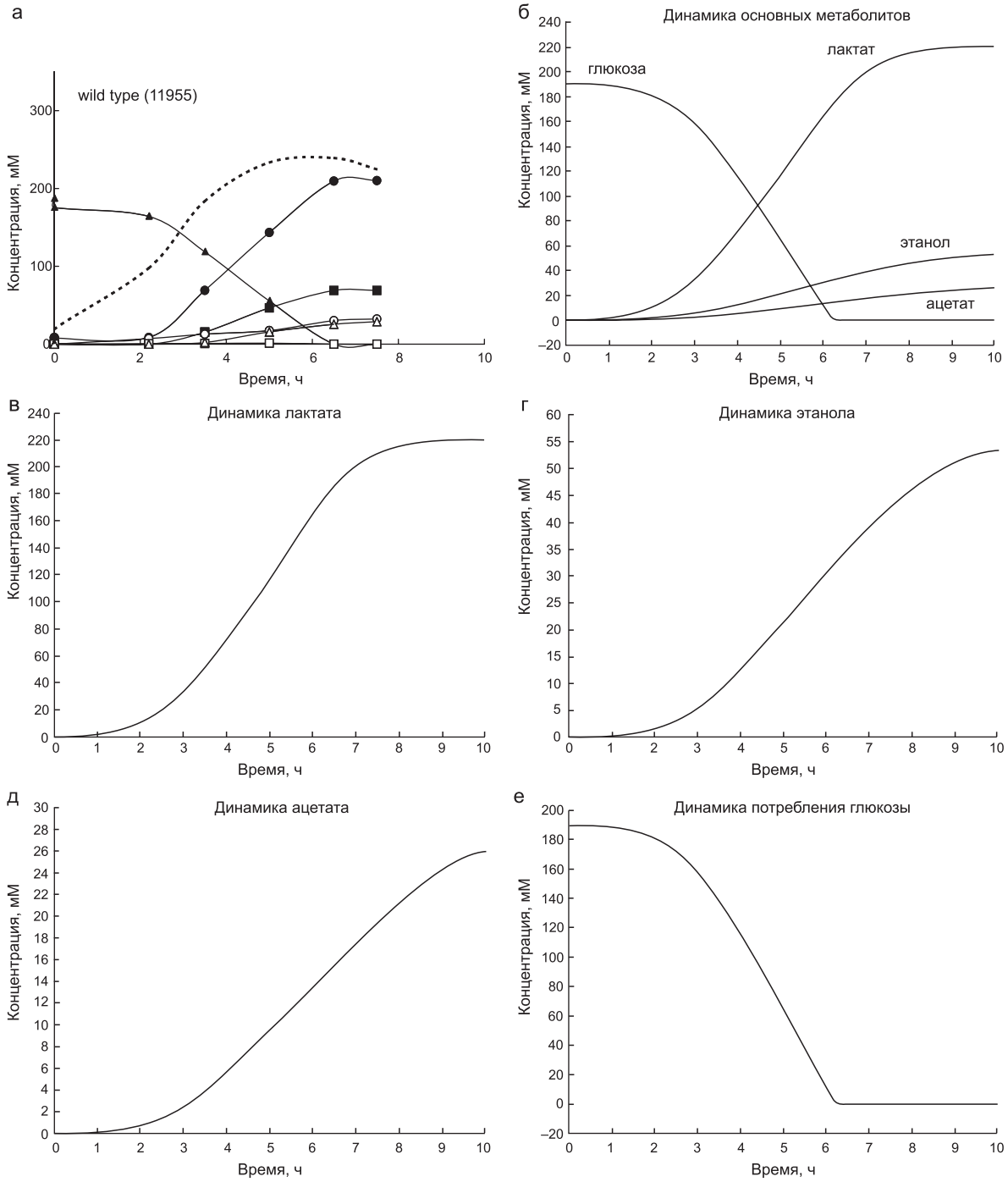


Рис. 3. Кинетика потребления глюкозы клеткой дикого типа *G. thermoglucosidasius* и синтеза конечных продуктов.

а – ось X – время (ч), ось Y – концентрация метаболита (mM); глюкоза (темные треугольники), этанол (темные квадраты), молочная кислота (темные кружки), уксусная кислота (светлые кружки), муравьиная кислота (светлые треугольники) и пировиноградная кислота (светлые квадраты). Пунктиром обозначена динамика изменения плотности культуры (OD 600) (Cripps *et al*., 2009); б – рассчитанные в модели кинетики потребления глюкозы и синтеза конечных продуктов в клетке *Geobacillus* sp., ось X – время (ч), ось Y – концентрация метаболита (mM); в – рассчитанная в модели кинетика биосинтеза молочной кислоты клеткой *Geobacillus* sp., ось X – время (ч), ось Y – концентрация метаболита (mM); г – рассчитанная в модели кинетика биосинтеза этанола клеткой *Geobacillus* sp., ось X – время (ч), ось Y – концентрация метаболита (mM); д – рассчитанная в модели кинетика биосинтеза уксусной кислоты клеткой *Geobacillus* sp., ось X – время (ч), ось Y – концентрация метаболита (mM); е – рассчитанная в модели кинетика потребления глюкозы клеткой *Geobacillus* sp., ось X – время (ч), ось Y – концентрация метаболита (mM).

Уравнения скоростей ферментативных реакций (обозначения-сокращения в нижнем регистре констант соответствуют обозначениям-сокращениям на рис. 2; графическое соответствие указанных скоростей реакций R представлено на рис. 2):

$$R_1 = \frac{V_{pts} \cdot C_{glc} \cdot C_{pep}}{C_{pyr} \cdot \left(1 + \frac{C_{g6p}^{4}}{K_{pts4}}\right) \cdot \left(\frac{C_{glc} \cdot C_{pep}}{C_{pyr}} + C_{glc} \cdot K_{pts3} + \frac{C_{pep} \cdot K_{pts2}}{C_{pyr}} + K_{pts1}\right)} \quad (1.1)$$

(Kadir *et al.*, 2010);

$$R_2 = \frac{V_{glk} \cdot C_{atp} \cdot C_{glc}}{\left(C_{atp} + K_{glk1}\right) \cdot \left(C_{glc} + K_{glkm}\right)} \quad (1.2)$$

(Kadir *et al.*, 2010);

$$R_3 = \frac{V_{pgi} \cdot \left(C_{g6p} - \frac{C_{f6p}}{K_{pgieq}}\right)}{K_{pgi1} \cdot \left(1 + \frac{C_{6pg}}{K_{pgi3}} + \frac{C_{f6p}}{K_{pgi2} \cdot \left(1 + \frac{C_{f6p}}{K_{pgi4}}\right)}\right) + C_{g6p}} \quad (1.3)$$

(Chassagnole *et al.*, 2002);

$$R_4 = \frac{V_{pfk} \cdot C_{f6p} \cdot K_{atp1}}{K_{atp2} \cdot \left(C_{f6p} + \frac{K_{pfk1} \cdot \left(K_{adp2} + \frac{C_{pep}}{K_{pfk2}}\right)}{K_{adp1}}\right) \cdot \left(1 + \frac{L_{pfk}}{\left(1 + \frac{C_{f6p} \cdot K_{adp1}}{K_{pfk1} \cdot \left(K_{adp2} + \frac{C_{pep}}{K_{pfk2}}\right)}\right)^{4}}\right)} \quad (1.4)$$

(Chassagnole *et al.*, 2002; Kadir *et al.*, 2010);

$$R_5 = \frac{V_{aldo} \cdot \left(C_{fdp} - \frac{C_{dhap} \cdot C_{gap}}{K_{aldoeq}}\right)}{K_{aldo1} + C_{fdp} + \frac{C_{dhap} \cdot C_{gap}}{K_{aldoeq} \cdot V_{blf}} + \frac{C_{dhap} \cdot K_{aldo2}}{K_{aldoeq} \cdot V_{blf}} + \frac{C_{fdp} \cdot C_{gap}}{K_{aldo4}} + \frac{C_{gap} \cdot K_{aldo3}}{K_{aldoeq} \cdot V_{blf}}} \quad (1.5)$$

(Chassagnole *et al.*, 2002);

$$R_6 = \frac{V_{tis} \cdot \left(C_{dhap} - \frac{C_{gap}}{K_{tiseq}}\right)}{K_{tis1} \cdot \left(1 + \frac{C_{gap}}{K_{tis2}}\right) + C_{dhap}} \quad (1.6)$$

(Chassagnole *et al.*, 2002);

$$R_7 = \frac{V_{gapdh} \cdot \left(C_{gap} - \frac{C_{nadh} \cdot C_{pep}}{C_{nad} \cdot K_{gapdheq}}\right)}{\left(C_{gap} + K_{gapdh1} \cdot \left(1 + \frac{C_{pep}}{K_{gapdh2}}\right)\right) \cdot \left(1 + \frac{K_{gapdh3} \cdot \left(1 + \frac{C_{nadh}}{K_{gapdh4}}\right)}{C_{nad}}\right)} \quad (1.7)$$

(Chassagnole *et al.*, 2002);

$$R_8 = \frac{V_{pyk} \cdot C_{adp} \cdot C_{pep} \cdot \left(1 + \frac{C_{pep}}{K_{pyk1}}\right)^3}{K_{pyk1} \cdot \left(C_{adp} + K_{pyk3}\right) \cdot \left(L_{pyk} \cdot \left(\frac{1 + \frac{C_{atp}}{K_{pyk2}}}{1 + \frac{C_{amp}}{K_{pyk4}} + \frac{C_{fdp}}{K_{pyk5}}} \right)^4 + \left(1 + \frac{C_{pep}}{K_{pyk1}}\right)^4 \right)} \quad (1.8)$$

(Chassagnole *et al.*, 2002);

$$R_9 = \frac{V_{ldh2} \cdot C_{pyr}}{C_{pyr} \cdot \left(1 + \frac{C_{laci}}{K_{ldh2i}}\right) + K_{ldh2m} \cdot \left(1 + \frac{K_{ldh2a}}{C_{xli}}\right)} \quad (1.9)$$

(Oshiro *et al.*, 2009);

$$R_{10} = \frac{V_{pdh} \cdot C_{coa} \cdot C_{pyr}}{\left(C_{nad} \cdot K_{pdh2} \cdot K_{pdh3} \cdot K_{pdh4} \cdot \left(1 + \frac{C_{nadh} \cdot K_{pdh1}}{C_{nad}}\right)\right) \cdot \left(\left(1 + \frac{C_{pyr}}{K_{pdh2}}\right) \cdot \left(1 + \frac{C_{coa}}{K_{pdh4}} + \frac{C_{accoa}}{K_{pdh6}}\right) \cdot \left(\frac{1}{C_{nad}} + \frac{1}{K_{pdh3}} + \frac{C_{nadh}}{C_{nad} \cdot K_{pdh5}}\right)\right)} \quad (1.10)$$

(Kadir *et al.*, 2010);

$$R_{11} = \frac{V_{acdh} \cdot C_{accoa}}{C_{accoa} + K_{acdhm}} \quad (1.11)$$

(Oshiro *et al.*, 2009);

$$R_{12} = \frac{V_{pta} \cdot \left(C_{accoa} \cdot C_p - \frac{C_{acp} \cdot C_{coa}}{K_{ptaeq}}\right)}{K_{ptai1} \cdot K_{ptapm} \cdot \left(1 + \frac{C_{accoa} \cdot C_p}{K_{ptai1} \cdot K_{ptapm}} + \frac{C_{acp} \cdot C_{coa}}{K_{ptai3} \cdot K_{ptam}} + \frac{C_{accoa}}{K_{ptai1}} + \frac{C_{acp}}{K_{ptai2}} + \frac{C_{coa}}{K_{ptai3}} + \frac{C_p}{K_{ptapi}}\right)} \quad (1.12)$$

(Kadir *et al.*, 2010);

$$R_{13} = \frac{V_{ack} \cdot \left(C_{acp} \cdot C_{adp} - \frac{C_{ace} \cdot C_{atp}}{K_{ackeq}}\right)}{K_{ack1} \cdot K_{ack4} \cdot \left(1 + \frac{C_{ace}}{K_{ack2}} + \frac{C_{acp}}{K_{ack1}}\right) \cdot \left(1 + \frac{C_{adp}}{K_{ack4}} + \frac{C_{atp}}{K_{ack3}}\right)} \quad (1.13)$$

(Kadir *et al.*, 2010);

$$R_{14} = \frac{V_{cs} \cdot C_{accoa} \cdot C_{oaa}}{K_{csd} \cdot K_{csm1} + C_{oaa} \cdot K_{csm2} + C_{accoa} \cdot K_{csm1} \cdot \left(1 + \frac{C_{nadh}}{K_{csi1}}\right) + C_{accoa} \cdot C_{oaa} \cdot \left(1 + \frac{C_{nadh}}{K_{csi2}}\right)} \quad (1.14)$$

(Kadir *et al.*, 2010);

$$R_{15} = \frac{V_{mez} \cdot C_{nadp} \cdot C_{oaa}}{\left(C_{nadp} + K_{mezeq}\right) \cdot \left(C_{oaa} + K_{mezl}\right)} \quad (1.15)$$

(Kadir *et al.*, 2010);

$$R_{16} = \frac{V_{pck} \cdot C_{atp} \cdot C_{oaa}}{C_{adp} \cdot \left(\frac{C_{atp} \cdot C_{oaa}}{C_{adp}} + \frac{C_{atp} \cdot C_{pep} \cdot K_{pcki1} \cdot K_{pckm1}}{C_{adp} \cdot K_{pcki1} \cdot K_{pcki4}} + \frac{C_{atp} \cdot K_{pckm1}}{C_{adp}} + \frac{C_{oaa} \cdot K_{pcki1} \cdot K_{pckm1}}{K_{pcki2} \cdot K_{pcki3}} + \frac{C_{pep} \cdot K_{pcki1} \cdot K_{pckm1}}{K_{pcki2} \cdot K_{pckm2}} + \frac{K_{pcki1} \cdot K_{pckm1}}{K_{pcki2}} \right)} \quad (1.16)$$

(Kadir *et al.*, 2010);

$$R_{17} = \frac{V_{ppc} \cdot C_{pep} \cdot \left(K_{ppc1} + C_{accoa} \cdot K_{ppc2} + C_{fdp} \cdot K_{ppc3} + C_{accoa} \cdot C_{fdp} \cdot K_{ppc4}\right)}{\left(C_{pep} + K_{ppcm}\right) \cdot \left(1 + C_{accoa} \cdot K_{ppc5} + C_{fdp} \cdot K_{ppc6}\right)} \quad (1.17)$$

(Kadir *et al.*, 2010);

$$R_{18} = \frac{V_{g6pdh} \cdot C_{g6p}}{\left(C_{g6p} + K_{g6pdh1}\right) \cdot \left(1 + \frac{C_{nadph}}{K_{g6pdh2}}\right) \cdot \left(1 + \frac{K_{g6pdh3} \cdot \left(1 + \frac{C_{nadph}}{K_{g6pdh4}}\right)}{C_{nadp}}\right)} \quad (1.18)$$

(Chassagnole *et al.*, 2002);

$$R_{19} = \frac{V_{6pgdh} \cdot C_{6pg}}{\left(C_{6pg} + K_{6pgdh1}\right) \cdot \left(1 + \frac{K_{6pgdh2} \cdot \left(1 + \frac{C_{nadph}}{K_{6pgdhi1}}\right) \cdot \left(1 + \frac{C_{atp}}{K_{6pgdhi2}}\right)}{C_{nadp}}\right)} \quad (1.19)$$

(Chassagnole *et al.*, 2002);

$$R_{20} = V_{rpi} \cdot \left(C_{ru5p} - \frac{C_{r5p}}{K_{rpieq}}\right) \quad (1.20)$$

(Chassagnole *et al.*, 2002);

$$R_{21} = V_{rpe} \cdot \left(C_{ru5p} - \frac{C_{xu5p}}{K_{rpeeq}}\right) \quad (1.21)$$

(Chassagnole *et al.*, 2002);

$$R_{22} = \frac{V_{xuk} \cdot C_{xl}}{C_{xl} \cdot \left(1 + \frac{C_{laci}}{K_{xuki}}\right) + K_{xukm} \cdot \left(1 + \frac{C_{laci}}{K_{xuki}}\right)} \quad (1.22)$$

(Oshiro *et al.*, 2009);

$$R_{23} = \frac{V_{phk} \cdot C_{xu5p}}{C_{xu5p} + K_{phkm}} \quad (1.23)$$

(Oshiro *et al.*, 2009);

$$R_{24} = V_{tkta} \cdot \left(C_{r5p} \cdot C_{xu5p} - \frac{C_{gap} \cdot C_{s7p}}{K_{tktaeq}}\right) \quad (1.24)$$

(Chassagnole *et al.*, 2002);

$$R_{25} = V_{tktb} \cdot \left(C_{e4p} \cdot C_{xu5p} - \frac{C_{gap} \cdot C_{f6p}}{K_{tktbeq}}\right) \quad (1.25)$$

(Chassagnole *et al.*, 2002);

$$R_{26} = V_{tal} \cdot \left(C_{gap} \cdot C_{s7p} - \frac{C_{e4p} \cdot C_{f6p}}{K_{taleq}}\right) \quad (1.26)$$

(Chassagnole *et al.*, 2002).

Система обыкновенных дифференциальных уравнений (расшифровка обозначений концентраций метаболитов генной сети (рис. 2) дана в табл. 2):

$$\frac{dX}{dt} = \frac{V_x \cdot C_{glc} \cdot X \cdot \left(1 - \frac{X}{K_x}\right)}{C_{glc} + K_{xs}}. \quad (2.1)$$

$$\frac{dC_{glc}}{dt} = -X \cdot (R_1 + R_2). \quad (2.2)$$

$$\frac{dC_{pep}}{dt} = R_7 + R_{16} - R_1 - R_8 - R_{17} - m \cdot C_{pep}. \quad (2.3)$$

$$\frac{dC_{g6p}}{dt} = R_1 + R_2 - R_3 - R_{18} - m \cdot C_{g6p}. \quad (2.4)$$

$$\frac{dC_{pyr}}{dt} = R_1 + R_8 + R_{15} - R_9 - R_{10} - m \cdot C_{pyr}. \quad (2.5)$$

$$\frac{dC_{f6p}}{dt} = R_3 + R_{25} + R_{26} - R_4 - m \cdot C_{f6p}. \quad (2.6)$$

$$\frac{dC_{fdp}}{dt} = R_4 - R_5 - m \cdot C_{fdp}. \quad (2.7)$$

$$\frac{dC_{gap}}{dt} = R_5 + R_6 - R_7 - R_{26} + R_{23} + R_{24} + R_{25} - m \cdot C_{gap}. \quad (2.8)$$

$$\frac{dC_{dhap}}{dt} = R_5 - R_6 - m \cdot C_{dhap}. \quad (2.9)$$

$$\frac{dC_{lac}}{dt} = R_9 \cdot X. \quad (2.10)$$

$$\frac{dC_{accoa}}{dt} = R_{10} - R_{11} - R_{12} - R_{14} - m \cdot C_{accoa}. \quad (2.11)$$

$$\frac{dC_{eth}}{dt} = R_{11} \cdot X. \quad (2.12)$$

$$\frac{dC_{acp}}{dt} = R_{12} + R_{23} - R_{13} - m \cdot C_{acp}. \quad (2.13)$$

$$\frac{dC_{ace}}{dt} = R_{13} \cdot X. \quad (2.14)$$

$$\frac{dC_{oaa}}{dt} = R_{17} - R_{15} - R_{16} - m \cdot C_{oaa}. \quad (2.15)$$

R_{14} *не учитывается в уравнении концентрации* dC_{oaa}, *поскольку: сколько молекул ACCOA поступает в качестве субстрата в цикле трикарбоновых кислот, столько же за один цикл и образуется продукта ACCOA по стехиометрии*).

$$\frac{dC_{6pg}}{dt} = R_{18} - R_{19} - m \cdot C_{6pg}. \quad (2.16)$$

$$\frac{dC_{ru5p}}{dt} = R_{19} - R_{20} - R_{21} - m \cdot C_{ru5p}. \quad (2.17)$$

$$\frac{dC_{r5p}}{dt} = R_{20} - R_{24} - m \cdot C_{r5p}. \quad (2.18)$$

$$\frac{dC_{xu5p}}{dt} = R_{21} + R_{22} - R_{23} - R_{24} - R_{25} - m \cdot C_{xu5p}. \quad (2.19)$$

$$\frac{dC_{xl}}{dt} = - R_{22} \cdot X. \quad (2.20)$$

$$\frac{dC_{s7p}}{dt} = R_{24} - R_{26} - m \cdot C_{s7p}. \quad (2.21)$$

$$\frac{dC_{e4p}}{dt} = R_{26} - R_{25} - m \cdot C_{e4p}, \quad (2.22)$$

где X – концентрация клеток в культуре, C_i – концентрация метаболита в реакции, катализируемой i-м ферментом ($i = glc, pts \ldots acp$ – совпадает с обозначениями на рис. 2); V_i – максимальная скорость ферментативной реакции, катализируемой i-м ферментом ($i = glc, pts \ldots acp$ – совпадает с обозначениями на рис. 2); R_j – скорость j-й ферментативной реакции ($j = 1 \ldots 26$).

Как видно из табл. 1, для воспроизведения имеющихся кинетических данных для клеток дикого типа *Geobacillus* spp. необходимо изменять активности практически всех ферментов центрального метаболизма: в большинстве ферментативных реакций гликолиза (кроме реакций, катализируемых глюкозокиназой, альдолазой, глицеральдегид-3-фосфат дегидрогеназой, пируватдегидрогеназного комплекса (в его ферментативную активность также включена активность пируватформиатлиазы, катализирующей аналогичную реакцию в клетке *Geobacillus* spp.), фосфотрансацетилазой, ацетаткиназой) необходимо увеличивать их значения по сравнению с аналогичными показателями в клетках *Escherichia coli* и *Lactococcus lactis*. Необходимо отметить, что в большинстве случаев увеличение/уменьшение констант скорости реакций происходит на один и тот же порядок. По-видимому, выявленное «синхронное» изменение ферментативных активностей связано с совершенно другими оптимальными условиями роста культуры *Geobacillus thermoglucosidasius* по сравнению с *Escherichia coli*, что подтверждается и экспериментально (Tang *et al.*, 2009).

Более того, как было указано во введении, метаболизм *E. coli* отличается от метаболизма представителей отдела Fermicutes. В частности, существенным отличием является отсутствие лактатдегидрогеназы в геноме *E. coli*, в то время как молочная кислота является основным продуктом метаболизма большинства представителей *Bacilli*. Возможно, это в совокупности со значительной филогенетической удаленностью представителей двух бактериальных родов связано с тем, что в ходе эволюции прокариот возникли несколько различающиеся последовательности реакций и ветви центрального метаболизма у разных видов бактерий, в то же время обеспечивающих жизнедеятельность клетки. Кроме этого, можно добавить, что представители *Bacilli* как свободноживущие организмы в качестве источников углерода и энергии часто используют пентасахара, и изменение скоростей некоторых реакций может быть связано с обеспечением их успешного усвоения.

Результаты параметрической адаптации математической модели методом DEEP (Kozlov, Samsonov, 2011) к экспериментальным данным (Cripps *et al.*, 2009) приведены на рис. 3. Графики наглядно демонстрируют, что полученные расчеты не только качественно отражают экспериментальную динамику изменения концентрации конечных продуктов (этанол, молочная кислота и ацетат), потребления глюкозы клеткой дикого типа *G. thermoglucosidasius*, но и коли-

чественно воспроизводят значения стационарных концентраций измеренных метаболитов.

Для исследования прогностических возможностей математической модели, адаптированной к экспериментальным данным, были проведены эксперименты *in silico*: 1) «выключение» активности одного из ферментов (например, *ldh*$^{-}$) метаболического пути за счет обнуления константы максимальной скорости; 2) обнуление константы максимальной скорости одной из ферментативных реакций метаболического пути с одновременным повышением («up») активности фермента *pdh* (т. е. *pdh*^{up}); 3) одновременное повышение активностей ферментов *pdh* и *phk*; *pdh* и *pta*. В результате были рассмотрены несколько мутантных вариантов клетки *Geobacillus* spp.: *ldh*$^{-}$; *ldh*$^{-}$ *pdh*^{up}; *pdh*$^{-}$; *pdh*^{up}; *pta*$^{-}$; *pta*$^{-}$*pdh*^{up}; *pta*^{up}*pdh*^{up}; *phk*^{up}*pdh*^{up} (рис. 4).

Как видно из графиков, наибольшие изменения (как увеличение, так и уменьшение) ста-

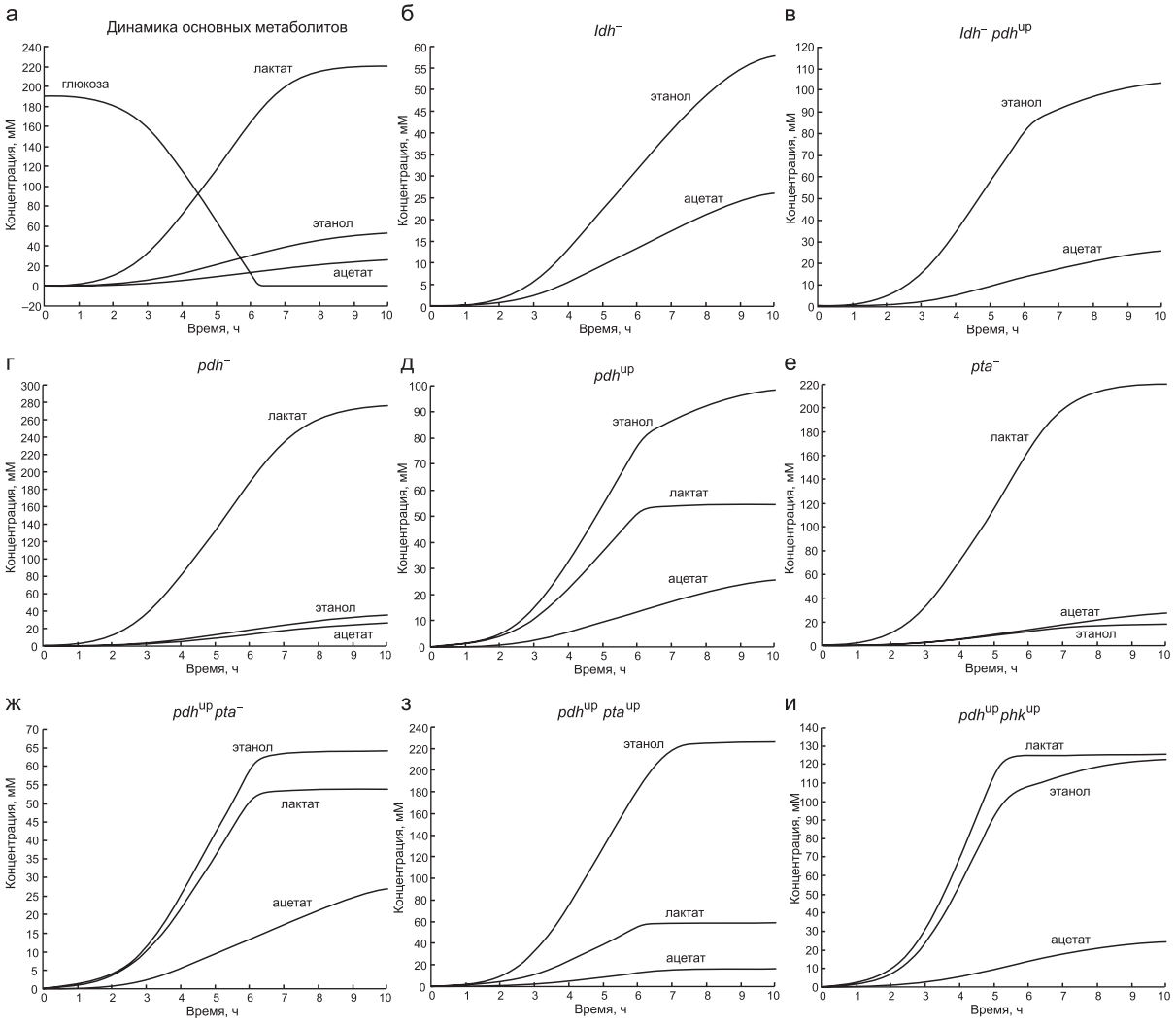


Рис. 4. Кинетика синтеза конечных продуктов клеткой дикого типа и мутантных вариантов *G. thermoglucosidasius*.

В клетке дикого типа: ось X – время (ч); ось Y – концентрация метаболита (**mM) (а); в клетке** *ldh*$^{-}$: ось X – время (ч); ось Y – концентрация метаболита (**mM) (б); в клетке** *ldh*$^{-}$ *pdh*^{up}: ось X – время (ч); ось Y – концентрация метаболита (**mM) (в); в клетке** *pdh*$^{-}$: ось X – время (ч); ось Y – концентрация метаболита (**mM) (г); в клетке** *pdh*^{up}: ось X – время (ч); ось Y – концентрация метаболита (**mM) (д); в клетке pta**$^{-}$: ось X – время (ч); ось Y – концентрация метаболита (**mM) (е);** в клетке *pta*$^{-}$*pdh*^{up}: ось X – время (ч); ось Y – концентрация метаболита (**mM) (ж); в клетке** *pta*^{up}*pdh*^{up}: ось X – время (ч); ось Y – концентрация метаболита (**mM) (з); в клетке** *phk*^{up}*pdh*^{up}: ось X – время (ч); ось Y – концентрация метаболита (**mM) (и)**.

ционарной концентрации этанола происходят при манипуляциях (при увеличении, нокауте) с активностью трех ферментов метаболического пути: лактатдегидрогеназы (*ldh*), пируватдегидрогеназного комплекса (*pdh*) и фосфотрансацетилазы (*pta*). Более того, динамика изменения концентрации этанола в мутантных *in silico* вариантах *Geobacillus* spp. в точности отражает результаты аналогичных экспериментов, полученные при культивировании в хемостате культуры клеток *G. thermoglucosidasius* двух мутантных штаммов, TM89 (ldh^-) и TM180 (ldh^- pdh^{up}) (Cripps *et al.*, 2009). Модель также предсказывает значительное повышение стационарной концентрации этанола при увеличении ферментативной активности фосфотрансацетилазы (рис. 4, з). Несмотря на то что этот фермент катализирует реакцию, в результате которой происходит отток ацетил-КоА – субстрата для синтеза этанола – в сторону наработки другого конечного продукта – ацетата, стационарная концентрация этанола в такой генномодифицированной культуре клеток будет повышена по сравнению с диким фенотипом. Полученный контринтуитивный результат можно объяснить усилением потока субстратов в метаболическом цикле через gapdh-pyk-pdh-pta-phk ферментативные реакции (рис. 1), в результате которого в конечном счете увеличивается и скорость синтеза этанола.

Для планирования направленных экспериментов молекулярно-генетической инженерии с бактерией *Geobacillus* spp. по оптимизации синтеза молочной кислоты с помощью разработанной и адаптированной математической модели был также проведен численный анализ экспериментов *in silico* (рис. 5): 1) «выключение» активности одного из ферментов (например $acdh^-$; acdh – алкогольдегидрогеназа) метаболического пути за счет обнуления константы максимальной скорости; 2) повышение («up») активности фермента системы, например *ldh* (т. е. ldh^{up}; ldh – лактатдегидрогеназа); 3) одновременное повышение активности ферментов, например *pdh* (пируватдегидрогеназный комплекс) и *pyk* (пируваткиназа).

На графике представлены только те примеры расчетов модели, в которых получено значительное изменение динамики синтеза молочной кислоты (лактата, рис. 5). Так, например, нокаут-мутации или увеличение активности ферментов *pta* (фосфат-ацетилтрансфераза), *ack* (ацетаткиназа) и *cs* (цитратсинтаза) практически не влияют на биосинтез молочной кислоты (изменение стационарной концентрации молочной кислоты для клеток мутантных фенотипов по сравнению со стационарной концентрацией лактата в клетке дикого типа составляет 1–2 %) в отличие от изменения динамики синтеза этанола, например, при увеличении активности фосфатацетилтрансферазы *pta* (рис. 4).

График наглядно демонстрирует, что как при уменьшении оттока пирувата в сторону образования этанола через ферментативную реакцию, катализируемую алкогольдегидрогеназой, так и при увеличении активности фермента лактатдегидрогеназы происходит значительное увеличение (~ в 1,5 раза) стационарной концентрации молочной кислоты. Однако увеличение активности пируватдегидрогеназного комплекса приводит к уменьшению концентрации лактата практически в два раза. В результате молекулярно-генетических манипуляций, приводящих к увеличению активности пируваткиназы, согласно экспериментам *in silico*, будет наблюдаться увеличение концентрации молочной кислоты в клетках *Geobacillus* spp., что может нивелироваться при одновременном увеличении активности пируватдегидрогеназного комплекса.

Полученные данные демонстрируют, что метаболиты, которые производят микроорганизмы в результате жизнедеятельности, зависят не только от каталитических свойств ферментов, участвующих в их получении из веществ основного гликолитического пути, но и от концентраций конкретных метаболитов. В частности, фермент лактатдегидрогеназа использует в качестве субстрата пируват. Как можно видеть из представленных данных, динамическое изменение его концентрации в результате изменения активности комплекса *pdh* сильно сказывается на синтезе конечного продукта – молочной кислоты.

ЗАКЛЮЧЕНИЕ

Впервые разработана и адаптирована к имеющимся экспериментальным данным интегрированная кинетическая модель биосинтеза этанола, молочной кислоты и ацетата в клетках *Geobacillus* spp. Рассчитанная в мо-

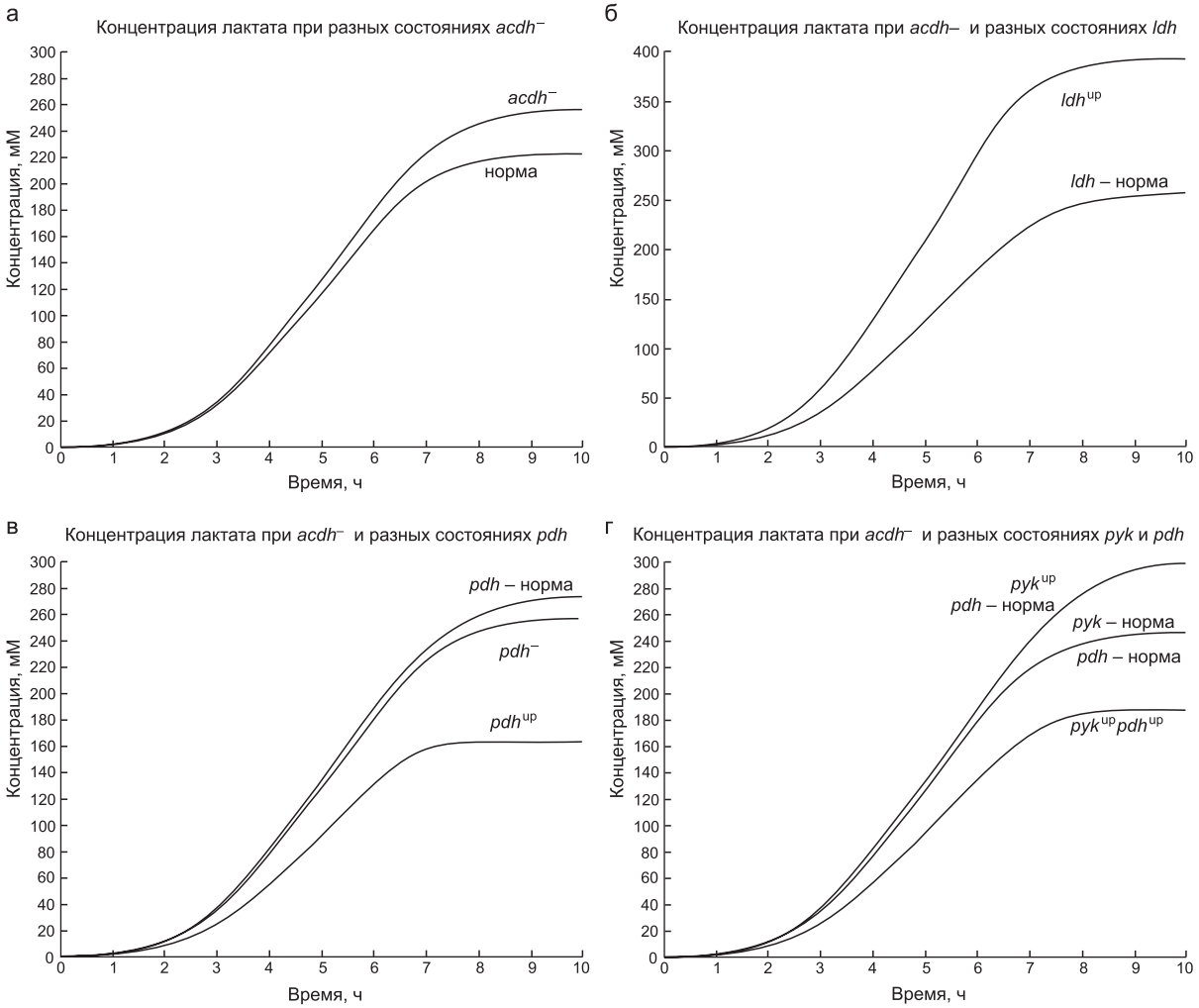


Рис. 5. Кинетика синтеза молочной кислоты (лактата) клеткой дикого типа и мутантных вариантов *G. thermoglucosidasius*.

В клетке дикого типа и в клетке *acdh*⁻: ось X – время (ч); ось Y – концентрация лактата (mM) (а); в клетке дикого типа и в клетке *ldh*^{up}: ось X – время (ч); ось Y – концентрация метаболита (mM) (б); в клетке *pdh*⁻, *pdh*, *pdh*^{up}: ось X – время (ч); ось Y – концентрация метаболита (mM) (в); в клетке дикого типа, в клетке *pdh pyk*^{up} и в клетке *pdh*^{up} *pyk*^{up}: ось X – время (ч) (г).

дели кинетика биосинтеза конечных продуктов воспроизводит экспериментальные данные на качественном и количественном уровнях. Более того, эксперименты *in silico* по созданию мутантных вариантов *Geobacillus* spp. по ферментам метаболического пути синтеза этанола показали, что для повышения выхода биоэтанола в результате культивирования *Geobacillus* spp., в первую очередь, необходимо нокаутировать ген фермента лактатдегидрогеназы, который обеспечивает основной поток катаболизма в направлении молочной кислоты. Вторым геном, который необходимо нокаутировать в геноме *Geobacillus* spp. для сверхпродукции этанола, является ген фермента пируват формиат лиазы. Нокаутирование этого гена важно, так как в результате работы этого фермента образуется формиат, ингибирующий рост культуры клеток. Значительное увеличение концентрации наработанного этанола клетками *Geobacillus* spp. также предсказано при увеличении активности или скорости синтеза ферментативных комплексов: пируватдегидрогеназного комплекса и фосфотрансацетилазы.

Численный анализ модели также показал, что перспективными молекулярно-генетическими экспериментами для увеличения стационарной концентрации лактата в бактериальной клетке

являются те, которые приводят к увеличению ферментативной активности лактатдегидрогеназы или пируваткиназы, либо же нокаут-мутация (или миссенс-мутации, приводящие к резкому уменьшению активности фермента) по гену, кодирующему алкогольдегидрогеназу.

Таким образом, разработанная математическая модель не только позволяет воспроизводить имеющиеся экспериментальные данные по динамике функционирования центрального метаболизма бактериальной клетки *Geobacillus* spp., но также является мощным *in silico* инструментом (Акбердин и др., 2013) для исследования режимов функционирования и метаболических потоков при «компьютерном создании» мутантных генотипов соответствующей бактерии. Безусловно, предложенная интегральная кинетическая модель центрального метаболизма *Geobacillus* spp. является «отправной точкой» в исследовании структурно-функциональных и динамических свойств метаболизма этой бактерии, поскольку при комплексном исследовании необходимо учитывать функционирование метаболических подсистем клетки в условиях повышенных температур, свойственных жизнедеятельности термофильных микроорганизмов; описывать участие промежуточных метаболитов в других ферментативных реакциях, подсистемах метаболизма клетки, что сказывается и на синтезе кофакторов – ключевых элементов для функционирования некоторых ферментов, и на росте бактериальной культуры клеток в целом.

БЛАГОДАРНОСТИ

Работа была выполнена в рамках Государственного контракта № 14.512.11.0050 «Создание методов метаболической инженерии термофильных микроорганизмов для получения штаммов-продуцентов молочной кислоты».

ЛИТЕРАТУРА

Акбердин И.Р., Казанцев Ф.В., Ермак Т.В. и др. «Электронная клетка»: проблемы и перспективы // Мат. биол. биоинф. 2013. Т. 8. № 1. С. 295–315.

Cavicchioli R., Amils R., Wagner D., McGenity T. Life and applications of extremophiles // Environ. Microbiol. 2011. V. 13. No. 8. P. 1903–1907.

Chassagnole C., Noisommit-Rizzi N., Schmid J.W. *et al.* Dynamic modeling of the central carbon metabolism of *Escherichia coli* // Biotechnol. Bioeng. 2002. V. 79. No. 1. P. 53–73.

Cripps R.E., Eley K., Leak D.J. *et al.* Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production // Metab. Eng. 2009. V. 11. No. 6. P. 398–408.

Feng L., Wang W., Cheng J. *et al.* Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir // Proc. Natl Acad. Sci. USA. 2007. V. 104. No. 13. P. 5602–5607.

Kadir T.A., Mannan A.A., Kierzek A.M. *et al.* Modeling and simulation of the main metabolism in *Escherichia coli* and its several single-gene knockout mutants with experimental verification // Microb. Cell Fact. 2010. V. 9. P. 88.

Kasi D., Ragauskas A.J. Switchgrass as an energy crop for biofuel production: A review of its ligno-cellulosic chemical properties // Energy Environ. Sci. 2010. V. 3. No. 9. P. 1182–1190.

Keasling J.D. Synthetic biology and the development of tools for metabolic engineering // Metab. Eng. 2012. V. 14. No. 3. P. 189–195.

Kozlov K., Samsonov A. DEEP – differential evolution entirely parallel method for gene regulatory networks // J. Supercomputing. 2011. V. 57. P. 172–178.

Kuipers O.P. Genomics for food biotechnology: prospects of the use of high-throughput technologies for the improvement of food microorganisms // Curr. Opin. Biotechnol. 1999. V. 10. No. 5. P. 511–516.

Likhoshvai V., Ratushny A. Generalized Hill function method for modeling molecular processes // J. Bioinform. Computat. Biol. 2007. V. 5. No. 2b. P. 521–531.

Nazina T.N., Tourova T.P., Poltaraus A.B. *et al.* Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermodenitrificans* to Geobacillus as the new combinations *G. stearothermophilus*, *G. thermoleovorans*, *G. kaustophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans* // Intern. J. Syst. Evol. Microbiol. 2001. V. 51. P. 433–446.

Oshiro M., Shinto H., Tashiro Y. *et al.* Kinetic modeling and sensitivity analysis of xylose metabolism in *Lactococcus lactis* IO-1 // J. Biosci. Bioeng. 2009. V. 108. No. 5. P. 376–384.

Parekh S., Vinci V.A., Strobel R.J. Improvement of microbial strains and fermentation processes // Appl. Microbiol. Biotechnol. 2000. V. 54. No. 3. P. 287–301.

Peskov K., Mogilevskaya E., Demin O. Kinetic modelling of central carbon metabolism in *Escherichia coli* // FEBS J. 2012. V. 279. No. 18. P. 3374–3385.

Rizzi M., Baltes M., Theobald U., Reuss M. *In viv*o analysis of metabolic dynamics in *Saccharomyces cerevisiae*: II. Mathematical model // Biotechnol. Bioeng. 1997. V. 55. No. 4. P. 592–608.

Smallbone K., Simeonidis E., Swainston N., Mendes P. Towards a genome-scale kinetic model of cellular metabolism // BMC Systems Biol. 2010. V. 4. No. 1. P. 6.

Sonnleitner B., Cometta S., Fiechter A. Equipment and growth

inhibition of thermophilic bacteria // Biotechnol. Bioeng. 1982. V. 24. No. 11. P. 2597–2599.

Storn R., Price K. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces // J. Global Optimization. 1997. V. 11. No. 4. P. 341–359.

Tang Y.J., Sapra R., Joyner D. *et al.* Analysis of metabolic pathways and fluxes in a newly discovered thermophilic and ethanol-tolerant *Geobacillus* strain // Biotechnol. Bioeng. 2009. V. 102. No. 5. P. 1377–1386.

Weber C., Farwick A., Benisch F. *et al.* Trends and challenges in the microbial production of lignocellulosic bioalcohol fuels // Appl. Microbiol. Biotechnol. 2010. V. 87. No. 4. P. 1303–1315.

Wu S., Liu B., Zhang X. Characterization of a recombinant thermostable xylanase from deep-sea thermophilic *Geobacillus* sp. MT-1 in East Pacific // Appl. Microbiol. Biotechnol. 2006. V. 72. No. 6. P. 1210–1216.

Zhao Y., Caspers M.P., Abee T. *et al.* Complete genome sequence of *Geobacillus thermoglucosidans* TNO-09.020, a thermophilic sporeformer associated with a dairy-processing environment // J. Bacteriol. V. 194. 2012. No. 15. P. 4118–4118.

MATHEMATICAL MODELING OF ETHANOL AND LACTIC ACID BIOSYNTHESIS BY THERMOPHILIC *GEOBACILLUS* BACTERIA

M.A. Nuriddinov[1], F.V. Kazantsev[2], A.S. Rozanov[1], K.N. Kozlov[2], S.E. Peltek[1], I.R. Akberdin[1], N.A. Kolchanov[1, 3]

[1] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: akberdin@bionet.nsc.ru;
[2] St-Petersburg State Polytechnicol university, St-Petersburg, Russia;
[3] Novosibirsk National Research State University, Novosibirsk, Russia

Summary

A mathematical model of the ethanol and lactic acid biosynthesis in the cells of *Geobacillus* spp. developed and adapted to the available experimental data is presented. It is shown that the mathematical model allows *in silico* design of genetic engineering experiments with the *Geobacillus* spp. bacterium and prediction of the dynamics of changes in synthesized ethanol and lactic acid concentrations depending on the molecular manipulations with the activity of enzymes of the metabolic system.

Key words: mathematical model, kinetic data, lactic acid, bioethanol, *Geobacillus*.

УДК 004.9; 575.112

СЕГРЕГАЦИОННЫЕ МОДЕЛИ СЛОЖНЫХ КОЛИЧЕСТВЕННЫХ ПРИЗНАКОВ И АНАЛИЗ СЦЕПЛЕНИЯ В РАСШИРЕННЫХ ДИАЛЛЕЛЬНЫХ СКРЕЩИВАНИЯХ ПАНЕЛИ РЕКОМБИНАНТНЫХ ИНБРЕДНЫХ ЛИНИЙ

© 2013 г. **М.С. Дьяков, А.В. Осадчук**

Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия,
e-mail: dkmike@gmail.com

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

Представлены классы сегрегационных моделей, описывающих характер наследования количественного признака, для расширенных нереципрокных диаллельных скрещиваний рекомбинантных инбредных линий. Отличительной особенностью данной работы являются использование многолокусного подхода и учет эпистатических взаимодействий групп локусов между собой. В построенных классах моделей производится поиск решений, которые с точностью до средовых шумов описывают экспериментальные данные. Далее выполняется анализ сцепления, т.е. определение положения модельных локусов найденных решений на генетической карте хромосом. Апробация процедуры поиска в пространстве моделей и подхода к анализу сцепления проводилась на реальных данных, где в качестве количественного признака выступала масса мозжечка лабораторных мышей. Дано краткое описание реализованного программного обеспечения.

Ключевые слова: генетический анализ, количественные признаки, многолокусный подход, эпистатические взаимодействия, анализ сцепления, регрессионный анализ, рекомбинантные инбредные линии, расширенные диаллельные скрещивания, информационные технологии анализа данных.

ВВЕДЕНИЕ

Диаллельные скрещивания представляют собой потомков первого поколения, полученных от скрещивания инбредных линий во всевозможных комбинациях. Диаллельный анализ позволяет получать достаточно точную оценку генотипических значений для каждой комбинации скрещиваний, поскольку в ней используются группы животных, одинаковые по своему генотипу. Это свойство делает его привлекательным подходом в физиологической генетике, так как физиологические признаки часто характеризуются большой средовой изменчивостью. Кроме того, в отличие от других систем генетического анализа, комплекс диаллельных скрещиваний обеспечивает однозначное установление всех генотипов диаллельной матрицы скрещиваний при условии, что известны генотипы инбредных линий. Эти два свойства диаллельных скрещиваний позволяют эффективно конструировать довольно сложные модели генетической детерминации исследуемых признаков с минимально необходимым для этих целей числом генетических локусов.

ПОСТРОЕНИЕ МОДЕЛИ

В данной работе в качестве экспериментального материала используются расширенные диаллельные скрещивания. Матрица расширенных диаллельных скрещиваний содержит: 1) рекомбинантные инбредные (РИ) линии; 2) потомков первого поколения скрещиваний панели РИ линий во всевозможных комбинациях; 3) линии-основатели панели РИ линий; 4) кроссы между

РИ линиями и их линиями-основателями; 5) гибриды первого поколения линий-основателей.

Для анализа характера наследования рассматриваемого количественного признака в расширенных диаллельных скрещиваниях используется метод построения статистических генетических моделей на основе множественного регрессионного анализа диаллельных матриц. Аутосомный локус рассматривается как фактор, имеющий 3 градации (уровня): 2 градации для гомозигот и 1 градацию для гетерозиготы. Генотипическое значение анализируемого признака в модели представлено как результат суммарного влияния некоторого числа вышеуказанных факторов, который описывается линейным регрессионным уравнением. Это уравнение является линейной комбинацией генетических эффектов данных локусов или факторов. Каждый аутосомный локус имеет два главных эффекта: аддитивный и доминантный. При взаимодействии локусов между собой в регрессионное уравнение вводятся соответствующие эффекты. При взаимодействии двух аутосомных локусов имеется 4 вида эффектов: гомо-гомозиготные, гомо-гетерозиготные, гетеро-гомозиготные, гетеро-гетерозиготные. Такого рода уравнения, выражающие генотипическое значение анализируемого признака как линейную регрессионную функцию от генотипа, впервые были введены ван дер Вином (Van der Veen, 1959) и описаны в классической монографии по биометрической генетике К. Мазера и Дж. Джинкса (1985). Эти уравнения использовались ими главным образом для описания компонент фенотипической изменчивости. В нашей работе эти уравнения используются для адекватного описания генотипической изменчивости в диаллельных скрещиваниях на основе минимального числа генетических локусов.

В настоящей работе используется множественная регрессионная модель вида:

$$X_{mf} = \mu + E_1 + E_2 + ... + E_L + \varepsilon, \qquad (1)$$

где X_{mf} – генотипическое значение признака для *mf*-го кросса диаллельной матрицы (*m* обозначает номер отцовской рекомбинантной линии, *f* – материнской); μ – свободный член уравнения, который оказывается равным среднему значению признака по всевозможным кроссам; E_1 – вклад главных (аддитивных и доминантных) генетических эффектов; E_i – вклад эпистатических эффектов взаимодействия всевозможных групп локусов размера *i*.

Генотип каждой рекомбинантной линии представляется в виде вектора $\theta_i = (\theta_i^1, ..., \theta_i^L)$, $i = 1...S$, S – число рекомбинантных линий, L – количество локусов.

Каждая из компонент вектора может принимать значения 0 или 1, т. е. каждый локус представлен двумя аллелями. Для определенности полагаем, что значение 0 будет указывать на аллель, соответствующий гену, унаследованному от материнской гомозиготы, и наоборот значение 1 будет указывать на аллель, соответствующий гену отцовской линии. Таким образом, значения, соответствующие генотипам материнской и отцовской линий, будут выражаться как $\theta_{S+1} = (0, 0, ..., 0)$ и $\theta_{S+1} = (1, 1, ..., 1)$ соответственно.

Из имеющихся значений рассматриваемого множества из $(S + 2)^2$ кроссов, полученных от нереципрокного скрещивания рекомбинантных линий друг с другом и с линиями-основателями во всевозможных сочетаниях с добавлением генотипов линий основателей и их нереципрокного гибрида F1, составляется система линейных уравнений (всего не более $(S + 2)^2$ уравнений). Фиксированием значений генотипов для рекомбинантных линий однозначно определяются значения индикаторных переменных в уравнении (1).

Каждая индикаторная переменная в линейном уравнении (1) умножена на коэффициент, равный соответствующим аддитивному, доминантному или различного рода эпистатическим генетическим эффектам. Кроме того, каждая индикаторная переменная является целочисленной функцией от значений генотипа рекомбинантных линий и может принимать значения: 1, 0 или –1. $A(i)^1_{mf} = D(i)_{mf} = \theta^i_m + \theta^i_f - 1$, $A(i)^2_{mf} = H(i)_{mf} = (\theta^i_m - \theta^i_f)^2$ – индикаторные переменные перед коэффициентами, равными аддитивным и доминантным эффектам соответственно. $B(i,j)^{rs}_{mf} = A(i)^r_{mf} \cdot A(j)^s_{mf}$ – индикаторные переменные перед коэффициентами, равными вышеуказанным эффектам взаимодействия между парами локусов, входящими в $E2$: $B(i,j)^{11}_{mf}$ – перед гомо-гомозиготными эффектами; $B(i,j)^{12}_{mf}$ – перед гомо-гетерозиготными эффектами; $B(i,j)^{21}_{mf}$ – перед гетеро-гомозиготными эффектами; $B(i,j)^{22}_{mf}$ – перед гетеро-гетерозиготными эффектами.

Таким образом, вклад главных генетических эффектов и эпистатических эффектов взаимодействия локусов может быть выражен следующим образом:

$$E_1 = \sum_{r=1}^{2} \sum_{i=1}^{L} [A(i)_{mf}^{r} \cdot a(i)^{r}],$$

$$E_2 = \sum_{r=1}^{2} \sum_{s=1}^{2} \sum_{i=1}^{L} \sum_{j=i+1}^{L} [B(i,j)_{mf}^{rs} \cdot b(i,j)^{rs}],$$

где L – число локусов. Значения вклада генетических эффектов E_i, $2 < i \leq L$ вычисляются аналогично.

Методом множественной линейной взвешенной регрессии (Кобзарь, 2006) определяются такие значения генетических эффектов, которые минимизировали бы отклонения от полученных в эксперименте значений фенотипов, т. е. минимизировали бы ошибку. Для этого решается система линейных уравнений и рассчитывается коэффициент множественной детерминации R^2. В качестве весов w_{mf} берется количество особей, использованное для определения генотипических средних значений признака X_{mf}. Качество полученного решения проверяется с использованием критерия Фишера сравнением остаточной дисперсии, не учтенной множественной регрессией, со средовой вариансой. Если оценка адекватности регрессии больше некоторого уровня, считаем, что полученное решение описывает экспериментальный материал с точностью до средовой вариансы – отклонений, обусловленных случайными средовыми факторами. При этом коэффициент множественный детерминации R^2 указывает на долю межкроссной наследственно обусловленной изменчивости, объясняемую с помощью множественной регрессионной модели. Если выбранное решение не подходит, выбираются другие значения генотипов у рекомбинантных линий и вычислительная процедура повторяется. Если окажется, что ни один из вариантов не описывает адекватно экспериментальные данные, то необходимо выбрать более сложную модель (большее число взаимодействующих локусов L).

Таким образом, при выборе оптимального решения мы приходим к перебору всех возможных вариантов генотипов рекомбинантных линий. Исходя из того, что L векторов $\theta^i = (\theta^i_1, \theta^i_2, ..., \theta^i_S)$, $i = 1 ... L$ должны быть различны и из того что от перестановки номеров локусов решение не изменяется, общий размер пространства решений вычисляется по следующей формуле:

$$N = \frac{2^S \cdot (2^S - 1) \cdot \ldots \cdot (2^S - L + 1)}{L!},$$

где S – число рекомбинантных линий, L – число локусов. Часто на практике в генетических экспериментах это число является очень большим, например, нами была исследована задача с $N \approx 1{,}88 \cdot 10^{14}$.

ПЕРЕБОР ПРОСТРАНСТВА РЕШЕНИЙ

Поскольку при данной постановке задачи не существует методов, которые бы гарантированно находили все адекватные решения из пространства моделей вида (1) быстрее, чем полный перебор вариантов, было применено улучшение алгоритма перебора, которое позволяет гораздо раньше находить адекватные решения. Данный алгоритм лучше всего можно описать как «рандомизированный поиск в ширину с приоритетом». Он основан на одновременном использовании двух методов: метода случайного поиска решений и метода поиска в ширину с приоритетом. Приоритеты определяются посредством ранжирования решений-кандидатов по качеству.

Метод поиска в ширину с приоритетом. Данный метод основывается на том принципе, что соседи более хорошего решения должны перебираться раньше соседей более плохого решения. Метод можно описать следующим образом.

Имеется контейнер («контейнер непросчитанных решений»), представляющий собой очередь с приоритетом, в нем содержатся непросчитанные решения. В качестве приоритета выступает коэффициент множественной детерминации. Каждый раз из контейнера извлекается наилучшее решение и просчитываются значения критерия для его соседей. Соседи, согласно их приоритетам, также кладутся в контейнер. Контейнер ограничен, и поэтому решения, занимающие положение ниже определенного («объем контейнера»), вытесняются. Рассмотренные решения убираются из контейнера непросчитанных решений поиска в ширину и помечаются как просчитанные.

Так как одно решение является соседом нескольких других, то для того чтобы избежать

повторного попадания решения в «контейнер непросчитанных решений», решение помечается как просчитанное после просмотра его соседей и добавляется в «контейнер просчитанных решений». Перед добавлением нового решения в «контейнер непросчитанных решений» проверяется, не просчитывалось ли данное решение ранее. В случае если решение рассматривалось, то оно игнорируется, иначе попадает в «контейнер непросчитанных решений».

Если в результате расчета критерия решение описывает экспериментальный материал с точностью до средовых шумов, то оно добавляется в результирующий набор. Первоначально контейнер заполняется либо указанными исследователем решениями, либо случайным образом.

Метод случайного поиска решений. Для вывода алгоритма из локальных областей необходимо введение случайных решений. В случае наличия большого числа хороших решений у системы поиска в ширину с приоритетом есть недостаток – при добавлении случайного решения в «контейнер непросчитанных решений» поиска очень вероятно, что это решение окажется в самом низу и будет рассмотрено в последнюю очередь или даже не будет рассмотрено вообще. Для решения этой проблемы введен «контейнер индивидуального поиска», в котором воспроизводится процесс поиска в ширину с приоритетом для решений, полученных из одного случайного решения. Все содержимое контейнера через несколько итераций добавляется в «контейнер непросчитанных решений», где некоторые удачные решения, возможно, будут в дальнейшем рассмотрены. Решения из «контейнера индивидуального поиска», которые были просчитаны, перемещаются сразу в «контейнер просчитанных решений».

Особенности программной реализации. На иллюстрации (рис. 1) изображена одна итерация алгоритма поиска решений в пространстве моделей на основе данных о расширенных диаллельных скрещиваниях.

1) В управляющем потоке из «контейнера непросчитанных решений» извлекается группа решений с наибольшим приоритетом. Эти решения помечаются как просчитанные и добавляются в «контейнер просчитанных решений».

2) Из окрестности извлеченных решений формируются решения-соседи.

3) При помощи набора переиспользуемых потоков (thread pool) в модуле расчета регрессии параллельно рассчитываются критерии адекватности для сформированных решений-соседей.

4) После расчета критериев в управляющем потоке новые решения добавляются в «контейнер непросчитанных решений» и происходит выбор решений, описывающих экспериментальные данные с точностью до средовых шумов.

АНАЛИЗ СЦЕПЛЕНИЯ МОДЕЛЬНЫХ ЛОКУСОВ С МАРКЕРАМИ ХРОМОСОМ

После формулирования гипотез о характере наследования локуса по рекомбинантным линиям и поиска адекватных решений на основе этих гипотез актуальной становится задача точного определения положения локуса в исследуемом геноме. Произведенные в последние годы исследования предоставили такую возможность, обеспечив экспериментатора идеальными хромосомными микросателлитными маркерами, плотно картирующими весь геном. Микросателлитные маркеры имеют значительно более высокий уровень полиморфизма, чем ранее используемые для этой цели мутантные аллели генов и полиморфизм по генам ферментов, и их применение позволяет с большой степенью достоверности найти хромосомную локализацию модельного локуса. Характеристика степени близости модельного локуса к определенному месту хромосомы носит название сцепленности и определяется через сходство характеров распределения аллелей модельного и фланкирующих его микросателлитных локусов, а также расстояния между маркерными микросателлитными локусами на генетической карте рекомбинантных линий.

Таким образом, построенная сегрегационная модель будет не только адекватно описывать межкроссную генотипическую изменчивость, но ее модельные локусы будут сцеплены с некоторыми картированными микросателлитными маркерами. Это позволит произвести отсев несцепленных адекватных решений-кандидатов.

Трехлокусное сцепление. Одномерный случай трехлокусного сцепления был описан П. Ньюманом (Neumann, 1991). Рассмотрим тестовый локус *C* и пару сцепленных маркерных

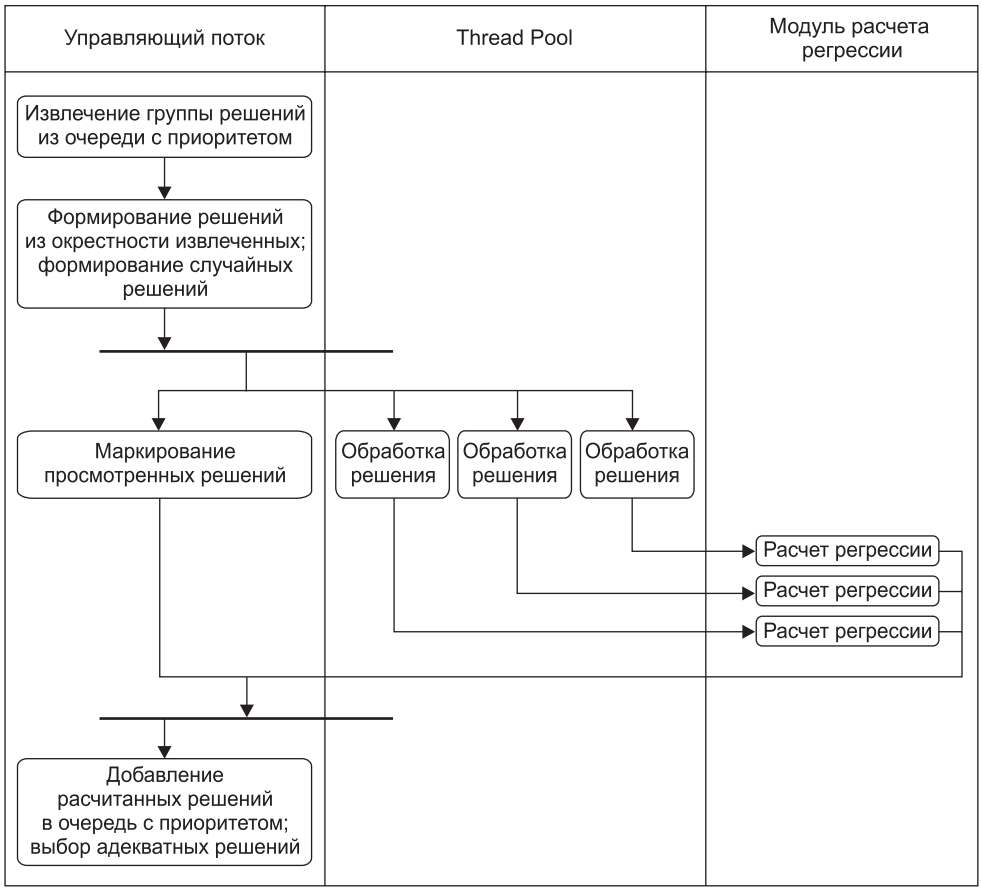


Рис. 1. Диаграмма деятельности (activity diagram). Представлена одна итерация алгоритма поиска решений в пространстве моделей.

локусов A и B (рис. 2). Количество различий (рекомбинаций) в генотипах локусов A и C в S рекомбинантных линиях обозначим как a. Аналогично определим b как количество рекомбинаций между B и C и c как количество различий в генотипах между локусами A и B. Количество двойных рекомбинаций d (число линий, которые имеют различия между локусами A и C и B и C одновременно) можно определить как $d = (a + b - c)/2$. n – разность $(S - c)$.

Вероятность сцепления $P(L)$ представляет собой сумму вероятностей возникновения каждого из трех альтернативных порядков расположения генов: $P(L) = P(CAB) + P(ACB) + P(ABC)$, где C – тестовый локус, A и B – сцепленные маркерные локусы. Таким образом, вероятность (на основе данных о распределении аллелей) того, что локус C сцеплен с A и B, может быть получена из байесовских выражений (2):

$$P(L|a, b) = \frac{[P(CAB)P(a, b|CAB) + P(ACB)P(a, b|ACB) + P(ABC)P(a, b|ABC)]}{P(a, b)} \quad (2)$$

$$P(a, b) = P(CAB)P(a, b|CAB) + P(ACB)P(a, b|ACB) + P(ABC)P(a, b|ABC) + P(\overline{L})P(a, b|\overline{L}).$$

Хромосома, содержащая сцепленные маркерные локусы A и B, имеет длину l (выраженную в морганидах) и состоит из трех частей. Длину сегмента между локусами A и B обозначим как m_{AB}, расстояние от начала хромосомы до локуса A примем за m_A, а расстояние от локуса B до конца хромосомы – за m_B (рис. 2). Таким образом, $l = m_A + m_{AB} + m_B$.

Априорная вероятность того, что тестовый локус C сцеплен с сегментом AB и расположен слева от маркерного локуса A, пропорциональна m_A и равна $P(CAB) = m_A/T$, где

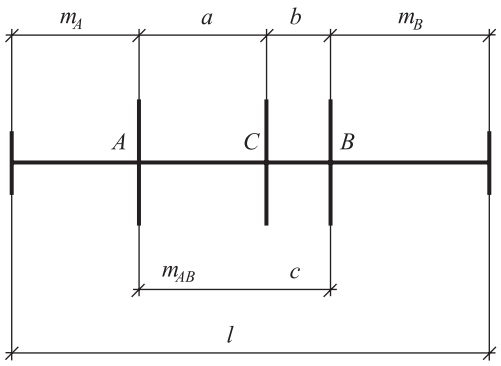


Рис. 2. Схематическое изображение хромосомы.

A и *B* – маркерные локусы, *C* – тестовый локус; *a*, *b* и *c* – количество соответствующих рекомбинаций; m_A, m_B, m_{AB} и *l* – расстояния, выраженные в морганидах.

T – длина всего генома (например, для мыши *T* = 16 морганид). Аналогично определяются априорные вероятности того, что тестовый локус расположен между маркерными локусами или справа от локуса *B*: $P(ACB) = m_{AB}/T$, $P(ABC) = m_B/T$.

Априорная вероятность сцепления тестового локуса с маркерными равна $P(L) = l/T$. Соответственно, вероятность того, что тестовый локус не сцеплен с маркерными, составляет $P(\overline{L}) = 1 - P(L) = (T - l)/T$.

Вероятность того, что тестовый локус *C* имеет *a* и *b* рекомбинаций со сцепленными маркерами *A* и *B* соответственно в наборе из *S* рекомбинантных инбредных линий и не сцеплен с ними, равна (полиномиальное распределение):

$$P(a, b|\overline{L}) = \frac{C \cdot R^c(m_{AB}) \cdot (1 - R(m_{AB}))^n}{2^S},$$

где $C = \dfrac{S!}{[d!(n-d)!(a-d)!(b-d)!]}$,

$R(x) = \dfrac{4 \cdot r(x)}{[1 + 6 \cdot r(x)]}$ – ожидаемая доля различий в распределении аллелей между двумя локусами, которые находятся на расстоянии *x* морганид, в панели РИ линий (Haldane, Waddington, 1931). $r(x) = \dfrac{0{,}5 \cdot (e^{2x} - e^{-2x})}{(e^{2x} + e^{-2x})}$ – картирующая функция Косамби (Kosambi, 1943). $R(m_{AB})$ можно считать равным рекомбинантному соотношению *c*/*S*.

Остальные три условные вероятности определяются как

$$P(a, b|CAB) = \left(\frac{C}{m_A}\right) K(CAB),$$

$$P(a, b|ACB) = \left(\frac{C}{m_{AB}}\right) K(ACB),$$

$$P(a, b|ABC) = \left(\frac{C}{m_B}\right) K(ABC), \text{ где}$$

$$K(CAB) = \int_0^{m_A} R^c(m_{AB}) \cdot \left(1 - R(m_{AB})\right)^n \cdot R^a(x) \cdot \left(1 - R(x)\right)^{S-a} dx,$$

$$K(ACB) = \int_0^{m_{AB}} R^b(m_{AB} - x) \cdot \left(1 - R(m_{AB} - x)\right)^{S-b} \cdot R^a(x) \cdot \left(1 - R(x)\right)^{S-a} dx,$$

$$K(ABC) = \int_0^{m_B} R^c(m_{AB}) \cdot \left(1 - R(m_{AB})\right)^n \cdot R^b(x) \cdot \left(1 - R(x)\right)^{S-b} dx.$$

В многомерном случае при определении вероятности сцепления группы локусов можно считать, что сцепление по каждому локусу в отдельности происходит независимо. Поэтому вероятность в многомерном случае выражается следующим образом: $P(L^k|a^k, b^k) = \prod_{i=1}^{k} P(L_i|a_i, b_i)$, где *k* – размерность задачи (количество локусов), а $P(L_i \mid a_i, b_i)$ вычисляются для одномерного случая описанным выше способом.

Поиск всех точек сцепления. Для каждого адекватного решения, найденного на первом этапе, определяются все возможные точки сцепления на генетической карте хромосом. Используемый метод основывается на поиске с возвратом (backtracking).

Допустим, что для первых *i* из *k* локусов уже установлен возможный вариант сцепления. Для *i* + 1 локуса подбирается возможное положение, и алгоритм рекурсивно повторяется для *i* + 2 локуса и т. д. Данная процедура продолжается до тех пор, пока не найдено возможное положение для всех *k* локусов. Если итоговая *k*-мерная точка удовлетворяет многомерному критерию сцепленности, то она добавляется в результирующий список.

На каждом этапе алгоритма для ускорения поиска применяются два отсечения с использованием оценки снизу на суммарное количество

двойных рекомбинаций и оценки сверху на многомерную вероятность сцепления. Первое отсечение основывается на том, что необходимый уровень многомерного сцепления не может быть достигнут при суммарном количестве двойных рекомбинаций больше определенного. Оценка сверху на вероятность многомерного сцепления также позволяет не рассматривать поддеревья поиска, в листьях которых пороговое значение критерия сцепленности заведомо не может достигаться.

Так как вероятность одномерного сцепления при фиксированном количестве рекомбинантных инбредных линий зависит только от трех параметров (количества соответствующих рекомбинаций a, b и c, описанных выше) и не зависит от внутренней структуры локусов или микросателлитных маркеров, полученное соотношение (2) можно вычислять только один раз для каждой комбинации параметров. Далее эти значения можно хранить в виде таблицы и при необходимости получать их мгновенно.

Применение хеширования. Метод поиска всех точек сцепления (без учета отсечений), описанный выше, обладает временно́й сложностью $O(M^L)$, где M – количество маркеров на генетической карте хромосом, а L – количество локусов. Таким образом, сложность алгоритма растет экспоненциально в зависимости от числа локусов. Однако, если исследователем выбран уровень многомерного сцепления, при котором суммарное количество двойных рекомбинаций не может превышать 0 (т. е. равняется 0), то перед исполнением алгоритма поиска всех точек сцепления можно ввести дополнительную проверку, которая позволяет избавиться от лишних расчетов при отсутствии сцепления модельных локусов с маркерами на генетической карте хромосом.

Рассмотрим пару фланкирующих маркеров, между которыми может потенциально располагаться модельный локус. При условии, что количество двойных рекомбинаций в этой тройке локусов равняется нулю, можно по паре фланкирующих маркеров сформировать все возможные модельные локусы, которые удовлетворяют данному ограничению. Количество таких модельных локусов равняется 2^c, где c – число рекомбинаций между фланкирующими маркерами. Время формирования этих локусов пропорционально их количеству.

Далее для определения возможности сцепления конкретного модельного локуса с фланкирующими маркерами необходимо проверить, находится ли этот модельный локус среди сформированной группы локусов для данных фланкирующих маркеров. Для того чтобы делать это эффективно, для каждого локуса из группы рассчитывается полиномиальный хеш (3) и полученные значения добавляются в хеш-таблицу. Для модельного локуса также вычисляется значение хеш-функции (3) и проверяется, находится или нет полученное значение в хеш-таблице. Если значение присутствует, то существует возможность того, что модельный локус может быть сцеплен с парой фланкирующих маркеров. Если полученное значение хеш-функции отсутствует в таблице, значит количество двойных рекомбинаций в данной тройке больше нуля, следовательно, необходимый уровень многомерного сцепления не может быть достигнут. Таким образом, данный модельный локус может быть сразу исключен из рассмотрения.

Использованная нами полиномиальная хеш-функция выглядит следующим образом:

$$h(\theta) = \sum_{i=1}^{S} (\theta_i \cdot 2^{i-1}) \bmod p, \qquad (3)$$

где θ – распределение аллелей, S – число рекомбинантных линий, p – размер хеш-таблицы.

Для реализации данного отсечения необходимо один раз выполнить предрасчет для всех пар фланкирующих маркеров. Асимптотическая оценка сложности предрасчета равняется $O(M \cdot 2^{c_{max}})$, где M – количество маркеров на генетической карте хромосом, а c_{max} – максимальное число рекомбинаций для пар фланкирующих маркеров. Дополнительная проверка для модельных локусов перед поиском всех точек сцепления при этом будет занимать $O(L)$ времени, где L – число локусов.

На практике эта оптимизация в применении к данным, описанным в разделе «Апробация программного пакета», позволила увеличить производительность в 30 раз. Следует также отметить, что данный подход можно обобщить на случай, если выбранный уровень многомерного сцепления может быть достигнут и при суммарном количестве двойных рекомбинаций больше нуля. Однако это повлечет увеличение времени предрасчета и уменьшение эффек-

тивности отсечения, так как при хешировании возрастет количество коллизий.

АПРОБАЦИЯ ПРОГРАММНОГО ПАКЕТА

Тестирование системы проводилось на реальных данных, где в качестве сложного количественного признака выступала масса мозжечка у мышей. Была использована матрица расширенных диаллельных скрещиваний РИ линий панели СХВ. Количество рекомбинантных инбредных линий S равнялось 13. При использовании одно-, двух- и трехлокусных моделей не было выявлено ни одного решения, адекватно описывающего экспериментальные данные с точностью до средовых шумов (рис. 3). Причем для одно- и двухлокусной модели было просмотрено все пространство допустимых решений. При количестве взаимодействующих локусов L, равном 4, было выявлено приблизительно 1,2 млн адекватных решений, из них сцепленными оказались только 22, учитывая, что все пространство решений содержит $N \approx 1{,}88 \cdot 10^{14}$ кандидатов.

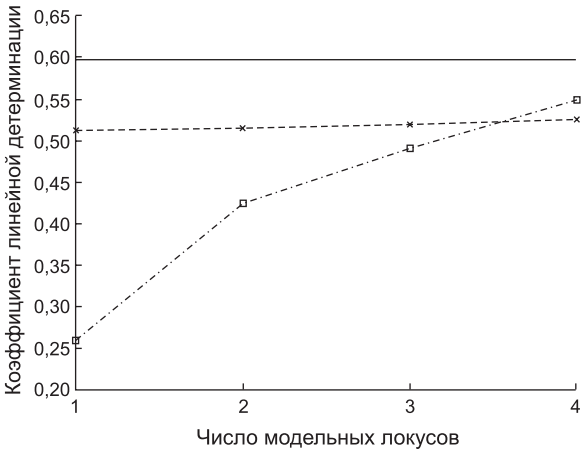


Рис. 3. Описательная способность различных многолокусных моделей.

Сплошной линией отмечен максимально возможный теоретический уровень описания (при помощи регрессионного анализа) экспериментальных данных в расширенных диаллельных скрещиваниях РИ линий СХВ. Он обусловлен различиями в фенотипе особей одного изогенного кросса, т. е. влиянием среды. Штриховой линией отмечены минимально необходимые уровни коэффициента линейной детерминации R^2 для каждой модели при описании данных на уровне $\alpha = 0{,}05$. Штрихпунктирной линией отмечены максимально достижимые коэффициенты R^2 для соответствующих моделей.

Следует отметить, что бóльшая часть найденных решений, описывающих экспериментальные данные с точностью до средовых шумов, выявляется разработанным алгоритмом на начальных итерациях (рис. 4). За динамикой поиска удобно проследить, используя логарифмическую шкалу (рис. 5).

Размещение программного пакета предполагается на современных персональных компьютерах и ноутбуках, поэтому система разрабатывалась с учетом многоядерности процессорных

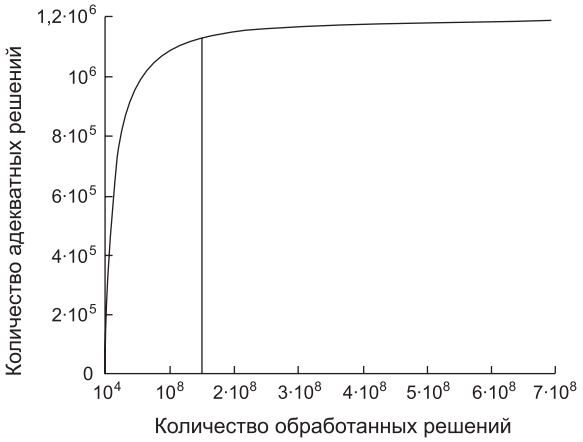


Рис. 4. Количество адекватных решений в зависимости от количества обработанных для поиска решений в пространстве моделей.

Вертикальным отрезком обозначены 95 % всех найденных адекватных решений.

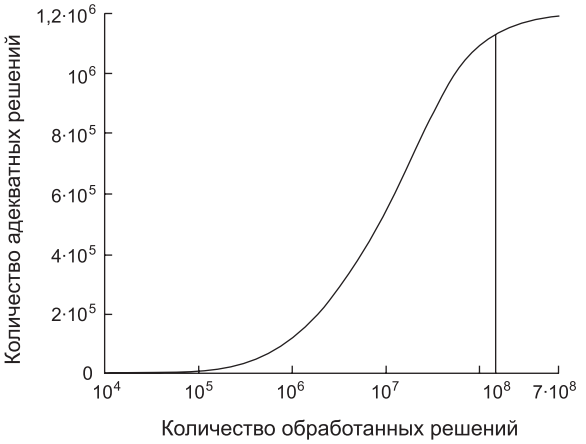


Рис. 5. Количество адекватных решений в зависимости от количества обработанных для поиска решений в пространстве моделей.

Ось абсцисс изображена в логарифмической шкале по основанию 10. Вертикальным отрезком обозначены 95 % всех найденных адекватных решений.

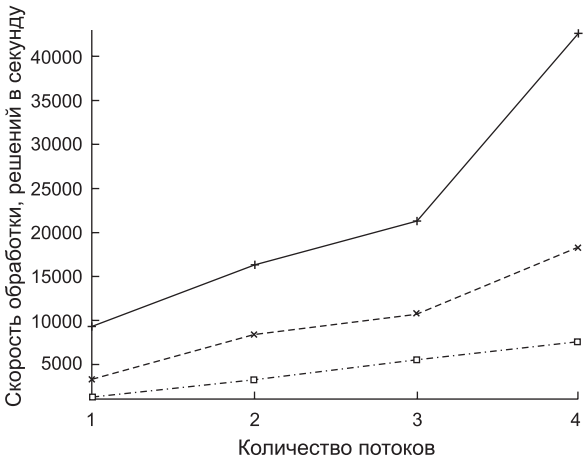


Рис. 6. Зависимость скорости счета от количества используемых потоков для разных моделей.

Сплошной линией обозначена двухлокусная модель, штриховой – трехлокусная, штрихпунктирной – четырехлокусная.

архитектур. За счет распараллеливания вычислений удалось добиться линейного повышения производительности в зависимости от количества физических ядер процессора. На графике (рис. 6) представлены данные о скорости счета с использованием компьютера на базе четырехъядерного процессора Intel Core i7 в зависимости от количества используемых потоков.

ЛИТЕРАТУРА

Кобзарь А.И. Прикладная математическая статистика. М.: ФИЗМАТЛИТ, 2006. 816 с.

Мазер К., Джинкс Дж. Биометрическая генетика. М.: Мир, 1985. 464 с.

Haldane J.B.S., Waddington C.H. Inbreeding and linkage // Genetics. 1931. V. 16. No. 4. P. 357–374.

Kosambi D.D. The estimation of map distance from recombination values // Annual. Eugenics. 1943. V. 12. No. 1. P. 172–175.

Neumann P.E. Three-locus linkage analysis using recombinant inbred strains and bayes' theorem // Genetics. 1991. V. 128. No. 3. P. 631–638.

Van der Veen J.H. Test of non-allelic interaction and linkage for quantitative characters in generations derived from two diploid pure lines // Genetics. 1959. V. 30. No. 3. P. 201–232.

SEGREGATION MODELS OF COMPLEX QUANTITATIVE TRAITS AND LINKAGE ANALYSIS IN EXTENDED RECOMBINANT INBRED CROSSES

M.S. Diakov, A.V. Osadchuk

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia, e-mail: dkmike@gmail.com

Summary

This paper introduces classes of segregation models, which describe the inheritance of a quantitative trait, in extended diallelic recombinant inbred crosses. The distinctive feature of the method is usage of the multiple-QTL approach and incorporation of epistatic interactions of loci groups. Solutions that would describe experimental data set to an accuracy of environmental variance are sought in the constructed classes of models. Then linkage analysis is performed: model loci positions of the found solutions are mapped on chromosomes. The search procedure and linkage analysis have been tested with real data on cerebellum weight in laboratory mice as a quantitative trait. The developed software is briefly described.

Key words: segregation analysis, complex quantitative traits, multiple-QTL approach, linkage analysis, regression analysis, epistatic interactions, recombinant inbred strains, extended diallelic crosses, information technologies of data analysis.

УДК 57.065

ГЕОМЕТРИЧЕСКИЕ СВОЙСТВА ЭВОЛЮЦИОННЫХ ДИСТАНЦИЙ

© 2013 г. **В.М. Ефимов[1, 2, 3], М.А. Мельчакова[4], В.Ю. Ковалева[2]**

[1] Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия, e-mail: efimov@bionet.nsc.ru;
[2] Федеральное государственное бюджетное учреждение науки Институт систематики и экологии животных Сибирского отделения Российской академии наук, Новосибирск, Россия;
[3] Томский национальный исследовательский государственный университет, Томск, Россия;
[4] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

Одним из способов изучения изменчивости биологических объектов является геометризация задачи: представление объектов точками в многомерном пространстве таким образом, чтобы расстояния между точками как можно лучше соответствовали различиям между объектами. Если различия между объектами являются евклидовыми расстояниями, то эта задача (с точностью до переноса, поворота и отражения) решается методами метрического шкалирования. Рассмотрены метрические свойства некоторых широко известных эволюционных дистанций для нуклеотидных последовательностей. Показано, что расстояния Джукса–Кантора и Кимуры не являются метриками. Введено новое расстояние, аналог расстояния Кимуры, – *PQ*-дистанция. Показано, что *p*-дистанция и *PQ*-дистанция являются квадратами евклидовых метрик, названных в статье E_p-дистанцией и E_{PQ}-дистанцией соответственно. Применимость E_{PQ}-дистанции проиллюстрирована на взятом из GenBank множестве нуклеотидных последовательностей цитохрома *b* 12 видов мышевидных грызунов Западной Сибири и Алтая и сравнена с результатами использования *LogDet*-расстояния.

Ключевые слова: нуклеотидные последовательности, модели эволюции, филогенетические реконструкции, генетические расстояния, геометризация, зоологическая систематика.

ВВЕДЕНИЕ

Эволюционные дистанции (генетические расстояния) – это различия генетической информации двух организмов (например, частот аллелей, нуклеотидных или аминокислотных последовательностей и т. д.), возникшие после их дивергенции от общего предка. Эволюционные дистанции между последовательностями могут быть прямо интерпретированы как филогенетические отношения между формами жизни, от которых эти последовательности получены. Иначе говоря, чем меньше эволюционные дистанции между двумя последовательностями, тем вероятнее, что они имели недавнего общего предка и, соответственно, тем более они родственны друг другу (Лукашов, 2009).

В настоящее время под эволюционной дистанцией понимается исключительно число замен нуклеотидов/аминокислот в пересчете на одну позицию, произошедших за время независимой эволюции двух ДНК-последовательностей после их дивергенции от общего предка, или его оценка различными методами (Ней, Кумар, 2004). Мы ограничимся рассмотрением только нуклеотидных последовательностей. Эволюционные дистанции делятся на истинные, наблюдаемые и расчетные (Лукашов, 2009). Истинные дистанции, как правило, неизвестны, так как для анализа доступны только

ныне живущие или недавно жившие формы, а информация об общих предках отсутствует. Наблюдаемые дистанции, например *p*-дистанция, основаны только на различиях между имеющимися последовательностями. Считается, что они занижают истинные, так как не учитывают ни длину путей от общего предка (ни даже его наличие), ни возможные петли на этих путях (множественные замены, приводящие к той же самой последовательности). Поэтому чаще всего используются расчетные эволюционные дистанции, в которых делается поправка на возможную множественность замен и которые опираются на некоторые модели эволюции нуклеотидных последовательностей. Общим для многих моделей является предположение о существовании матрицы переходных вероятностей между нуклеотидами и постоянстве ее во времени. Примерами являются генетические расстояния Джукса–Кантора, Кимуры и т. п. (Felsenstein, 2003; Ней, Кумар, 2004). В более продвинутых моделях эволюции авторы учитывают непостоянство нуклеотидных замен во времени, например в LogDet-дистанции (Lockhart *et al*., 1994; Zharkikh, 1994).

Эволюционные дистанции исходно разрабатывались для реконструкции эволюционной истории видов и изучения природы и сути селективных сил, формирующих эволюцию генов и видов. На практике это свелось к построению филогенетических деревьев (филодендрограмм) и их применению в биологической систематике. В настоящее время более прогрессивным считается построение филодендрограмм через компьютерное моделирование эволюции самих молекулярных последовательностей (Felsenstein, 2003; Ней, Кумар, 2004).

Полученные на этом пути результаты уже привели к значительной перестройке всего здания биологической систематики (см., например, MSW – Mammal species …, 2005; Млекопитающие …, 2012). Однако «за бортом» остались вся предыдущая работа и весь опыт классических систематиков со времен К. Линнея, несмотря на то что первичное определение видов до сих пор производится на основе морфологии, экологии, географии и прочей доступной информации, и без этого первичного определения молекулярная филогенетика все еще обойтись не может.

По нашему мнению, дальнейшие возможности развития современной компьютерно-молекулярной филогенетики существенно ограничены двумя серьезными обстоятельствами – невозможностью присоединения информации других типов, в частности морфологических данных, и особенно представлением получаемых результатов исключительно в виде филодендрограмм. Вопреки широко распространенному мнению, филодендрограммы не являются единственным способом представления филогенетических взаимоотношений и могут быть дополнены, например, отображением взаимного расположения таксонов в многомерном пространстве (Klingenberg, Ekau, 1996; Revell, 2009; Klingenberg, Gidaszewski, 2010; Ковалева и др., 2012, 2013; Polly *et al*., 2013). Это можно сделать всегда, если отнестись к генетическим расстояниям как к обычным мерам сходства/различия и применить методы неметрического шкалирования (Shepard, 1962; Ковалева и др., 2013). Однако это неизбежно приведет к искажению взаимных расстояний, масштаб которого не всегда можно оценить. Было бы удобнее, если бы генетические расстояния сразу были метрическими расстояниями, т. е. удовлетворяли аксиомам метрики. Еще лучше, если бы они были евклидовыми расстояниями. Однако, несмотря на почти полувековую историю генетических расстояний, их геометрическим свойствам уделялось очень мало внимания.

ЯВЛЯЮТСЯ ЛИ ГЕНЕТИЧЕСКИЕ РАССТОЯНИЯ МЕТРИЧЕСКИМИ РАССТОЯНИЯМИ?

В математике неотрицательная действительная функция $d(x,y)$, определенная на множестве X, называется *метрикой*, если она удовлетворяет следующим условиям:

$d(x,y) = d(y,x)$ (аксиома симметрии)

$d(x,x) = 0 <=> x = y$ (аксиома тождества)

$d(x,y) <= d(x,z) + d(z,y)$ (аксиома треугольника).

Свойство неотрицательности $d(x,y) \geq 0$ вытекает из этих аксиом. Числовое значение функции $d(x,y)$ называется *расстоянием* между элементами x и y (Петровский, 2003). Выполнение аксиом метрики обеспечивает возможность помещения элементов множества X без искаже-

ния взаимных расстояний в некоторое геометрическое пространство и наделения точек этого множества координатами в этом пространстве. Это, в свою очередь, позволяет применять весь арсенал методов многомерного анализа для исследования соотношения внутри- и межвидовой изменчивости, визуализации возможных направлений эволюции, объединения данных различных типов, например молекулярных и морфологических, и оценки их конгруэнтности (Ковалева и др., 2012, 2013).

Простейшими примерами генетических расстояний являются (Felsenstein, 2003; Ней, Кумар, 2004):

***p*-дистанция** – наблюдаемая доля различающихся нуклеотидов для двух последовательностей одинаковой длины. *p*-дистанция является метрикой Хэмминга с точностью до домножения на длину последовательности (Hamming, 1950);

расстояние Джукса–Кантора – предполагаемое число замен нуклеотидов в двух последовательностях, происшедших от одного неизвестного предка за эволюционное время, в пересчете на одну позицию, вычисляемое как

$d_{JC} = -\frac{3}{4}\ln(1-\frac{4}{3}p)$ (Jukes, Cantor, 1969);

двупараметрическое расстояние Кимуры – предполагаемое число замен в пересчете на одну позицию, вычисляемое как $d_{K2p} = -\frac{1}{2}\ln(1-2P-Q)-\frac{1}{4}\ln(1-2Q)$, где P – наблюдаемая доля транзиций, Q – наблюдаемая доля трансверсий (Kimura, 1980);

LogDet-расстояние – обобщение расстояния Джукса–Кантора на случай непостоянства вероятностей нуклеотидных замен во времени, вычисляемое как $d_{xy} = -\ln[\det F_{xy}]$, где x, y – последовательности, F_{xy} – матрица 4 × 4 частот совместной встречаемости пар нуклеотидов в каждой позиции для последовательностей x и y (Lockhart *et al.*, 1994).

Рассмотрим последовательности фиксированной длины m. Заметим, что расстояния Джукса–Кантора и Кимуры определены не для всех значений p, P, Q. Для таких значений можем положить значение расстояния, равное ∞. Покажем, что d_{JC} и d_{K2p} не являются метриками, так как для них не выполняется неравенство треугольника.

Запишем неравенство треугольника для расстояния Джукса–Кантора на последовательностях x, y, z:

$$-\frac{3}{4}\ln(1-\frac{4}{3}p(x,y)) \le$$
$$\le -\frac{3}{4}\ln(1-\frac{4}{3}p(x,z)) - \frac{3}{4}\ln(1-\frac{4}{3}p(z,y)).$$

После очевидных преобразований получим:

$$p(x,z) + p(z,y) - p(x,y) \ge \frac{4}{3}\,p(x,z)\,p(z,y).$$

Возьмем такие последовательности длины $m \ge 2$, в которых имеются различия только в позициях 1 и 2: x – AA, y – GG, z – AG. В последнем неравенстве слева будет нуль, а справа – положительное значение. Следовательно, расстояние Джукса–Кантора не является метрикой (Felsenstein, 2003).

Для расстояния Кимуры проводим аналогичные рассуждения для случая, когда доля трансверсий Q равна нулю, т. е. когда

$$d_{K2p} = -\frac{1}{2}\ln(1-2P).$$

Получаем неравенство

$$P(x,z) + P(z,y) - P(x,y) \ge 2\,P(x,z)\,P(z,y),$$

которое нарушается для указанных последовательностей. Следовательно, двупараметрическое расстояние Кимуры тоже не является метрикой (Мельчакова, Ефимов, 2011).

Заметим, что истинная эволюционная дистанция, т. е. число замен нуклеотидов в пересчете на одну позицию, является метрикой (Мельчакова, 2013). Поэтому то обстоятельство, что общеизвестные и широко используемые расчетные оценки этой дистанции, такие как расстояния Джукса–Кантора и Кимуры, теряют ее метрические свойства, вызывает некоторое недоумение и требует дальнейших исследований.

ЧТО ДОЛЖНЫ ОТРАЖАТЬ ЭВОЛЮЦИОННЫЕ ДИСТАНЦИИ?

Все расчетные генетические расстояния оценивают исключительно суммарное число замен нуклеотидов в пересчете на одну позицию, несмотря на то что для многопараметрических расстояний, начиная с расстояния Кимуры, предполагается, что вероятности замен различных типов (например, транзиций и трансверсий) тоже могут быть различными. «Поскольку транзиции в целом более вероятны, более часты, чем

трансверсии, т. е. занимают меньше времени, логично считать, что эволюционная дистанция между последовательностями с одной транзицией меньше, чем между последовательностями с одной трансверсией, а сами последовательности с одной транзицией более родственны друг другу (имеют более недавнего общего предка), чем последовательности с одной трансверсией» (Лукашов, 2009). Но в формулах многопараметрических расстояний это никак не отражено! В расстоянии Кимуры, например, расчетные частоты транзиций и трансверсий оцениваются раздельно по наблюдаемым частотам, а потом просто суммируются. Вообще говоря, логично было бы суммировать их с разными весами, причем вес трансверсий должен быть больше, чем вес транзиций. К сожалению, даже «взвешенное» расстояние Кимуры ни при каких весах не будет являться метрикой.

Заметим, что расстояние Джукса–Кантора получается монотонным преобразованием *p*-дистанции, которая относится к наблюдаемым дистанциям. По аналогии с *p*-дистанцией для «взвешенного» расстояния Кимуры можно предложить его наблюдаемый аналог – *PQ*-дистанцию:

$$d_{PQ} = P + (1 + \alpha)Q,\ \alpha \geq 0,$$

где P – наблюдаемая доля транзиций, Q – наблюдаемая доля трансверсий. Вопрос об их монотонной зависимости рассмотрен в следующем разделе.

Покажем, что и *p*-дистанция, и *PQ*-дистанция являются квадратами евклидовых расстояний. Если закодировать каждый нуклеотид в последовательностях набором чисел в соответствии с табл. 1 и для каждой пары последовательностей вычислить сумму квадратов разностей, т. е. квадрат евклидова расстояния, то, очевидно получим *p*-дистанцию с точностью до постоянного множителя. Если сделать то же самое в соответствии с табл. 2, то очевидно получим *PQ*-дистанцию, также с точностью до постоянного множителя. Отсюда следует, что разумнее всего извлекать из обеих дистанций квадратные корни, что приведет к евклидовым метрикам. Поэтому, по нашему мнению, именно их и следует использовать при применении геометрических методов для филогенетических реконструкций. Назовем их соответственно E_p-дистанция и E_{PQ}-дистанция.

Заметим, что в предположении постоянства нуклеотидных замен во времени *LogDet*-расстояние сводится к *p*-дистанции (Lockhart *et al.*, 1994; Zharkikh, 1994), следовательно, является ее обобщением. Соответственно, само *LogDet*-расстояние не евклидово, а его евклидовым аналогом является E_p-дистанция.

ПРИМЕНЕНИЕ К БИОЛОГИЧЕСКИМ ДАННЫМ

Для иллюстрации применимости предлагаемых нами евклидовых генетических расстояний – E_p-дистанции и E_{PQ}-дистанции – в качестве эмпирических данных были взяты 87 нуклеотидных последовательностей цитохрома *b* митохондриальной ДНК серых полевок (род *Microtus: **M. agrestis**, M. levis = M. rossiaemeridionalis, **M. oeconomus**, **M. gregalis***), лесных полевок (род *Myodes = Clethrionomys: My. Glareolus, My. **rufocanus**, My. **rutilus***), водяной полевки (род *Arvicola: **Arv. amphibious** = **Arv.***

Таблица 1

Кодировка нуклеотидов для *p*-дистанции

	A	G	T	C
A	$\frac{1}{\sqrt{2}}$	0	0	0
G	0	$\frac{1}{\sqrt{2}}$	0	0
T	0	0	$\frac{1}{\sqrt{2}}$	0
C	0	0	0	$\frac{1}{\sqrt{2}}$

Таблица 2

Кодировка нуклеотидов для *PQ*-дистанции

	A	G	T	C	A/G	T/C
A	$\frac{1}{\sqrt{2}}$	0	0	0	$\sqrt{\frac{\alpha}{2}}$	0
G	0	$\frac{1}{\sqrt{2}}$	0	0	$\sqrt{\frac{\alpha}{2}}$	0
T	0	0	$\frac{1}{\sqrt{2}}$	0	0	$\sqrt{\frac{\alpha}{2}}$
C	0	0	0	$\frac{1}{\sqrt{2}}$	0	$\sqrt{\frac{\alpha}{2}}$

terrestris), полевых мышей (род *Apodemus: A. agrarius*, *A. peninsulae*), домовой мыши (род *Mus: Mus musculus*), серой крысы (род *Rattus: R. norvegicus*) длиной 1138 п.н. из базы данных GenBank, ранее использованные в работе В.Ю. Ковалева с соавт. (2012). Для всех последовательностей с помощью пакетов MEGA5 (Tamura *et al*., 2011) и Excel вычислены матрицы расстояний Джукса–Кантора и Кимуры, а также *LogDet*-дистанций, E_p-дистанций и E_{PQ}-дистанций (при $\alpha = 1$). В табл. 3 приведены попарные коэффициенты корреляции между этими матрицами, полученные с помощью теста Мантеля (Mantel, 1967; Mantel, Valand, 1970). Видно, что на данном эмпирическом материале все дистанции отражают фактически одно и то же. На рис. 1 показана зависимость расчетных генетических расстояний от их наблюдаемых аналогов. Монотонная зависимость расстояния Джукса–Кантора от E_p-дистанции следует из определяющих их формул и ожидаема для *LogDet*-расстояния. Монотонная зависимость расстояния Кимуры от E_{PQ}-дистанции из формул не следует, но из графика видно, что их эмпирическая зависимость, невзирая на небольшие отклонения, выглядит точно так же. С точки зрения геометрического подхода E_p и E_{PQ}-дистанции наиболее удобны для применения, так как являются евклидовыми метриками.

Далее для матрицы E_{PQ}-дистанций был применен один из методов многомерного шкалирования – метод главных координат (Torgerson, 1952), включенный в пакет Jacobi4 (Ефимов и др., 2011). Из рис. 2 видно, что внутривидовая изменчивость незначительна на фоне межвидовой и ею можно пренебречь. Однако на взаимное расположение видов может влиять разный объем выборок. Поэтому для всех видов по главным координатам были вычислены их выборочные центроиды и между ними – матрица евклидовых расстояний. Соответственно, все виды получили равные веса. (Заметим, что без геометрического подхода это сделать не так просто.) К полученной матрице расстояний снова был применен метод главных координат (рис. 3, 4). Первая главная координата четко отвечает за различия между семействами Cricetidae и Muridae. Хорошо видна родовая структура, совпадающая с принятой на сегодня зоологической классификацией (MSW – Mam-

Таблица 3

Тест Мантеля
для матриц эволюционных дистанций

r	*JC*	*K2p*	E_p	E_{pq}	LogDet
JC	1,000	0,999	0,961	0,978	0,999
K2p	0,999	1,000	0,962	0,978	0,999
E_p	0,961	0,962	1,000	0,995	0,963
E_{pq}	0,978	0,978	0,995	1,000	0,978
LogDet	0,999	0,999	0,963	0,978	1,000

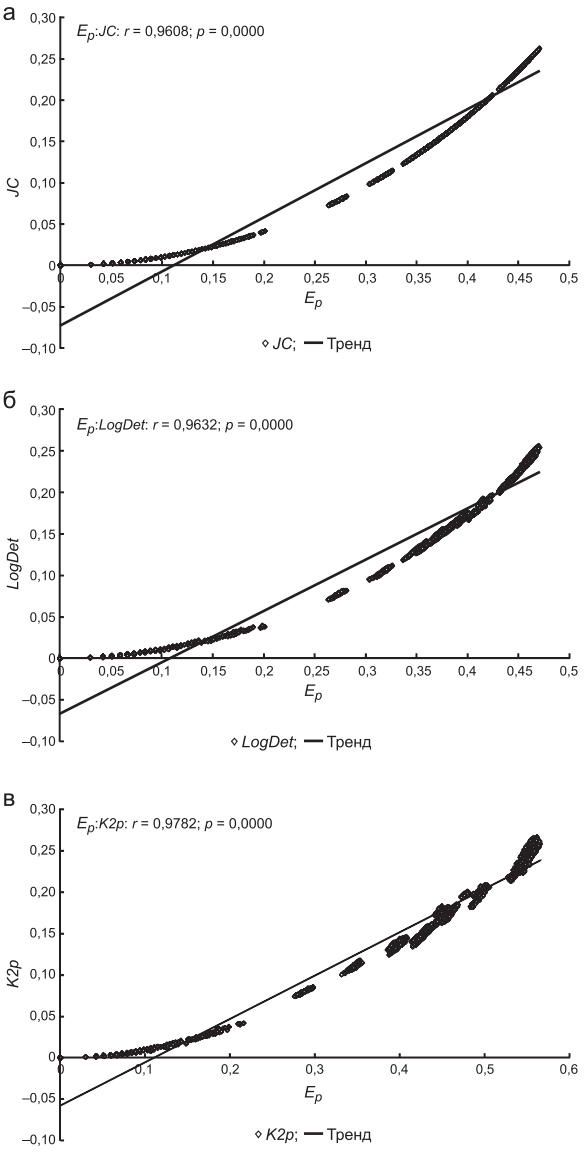


Рис. 1. Зависимость расчетных генетических расстояний от их наблюдаемых аналогов.

а – расстояние Джукса–Кантора от E_p-дистанции; б – *LogDet*-расстояние от E_p-дистанции; в – расстояние Кимуры от E_{PQ}-дистанции.

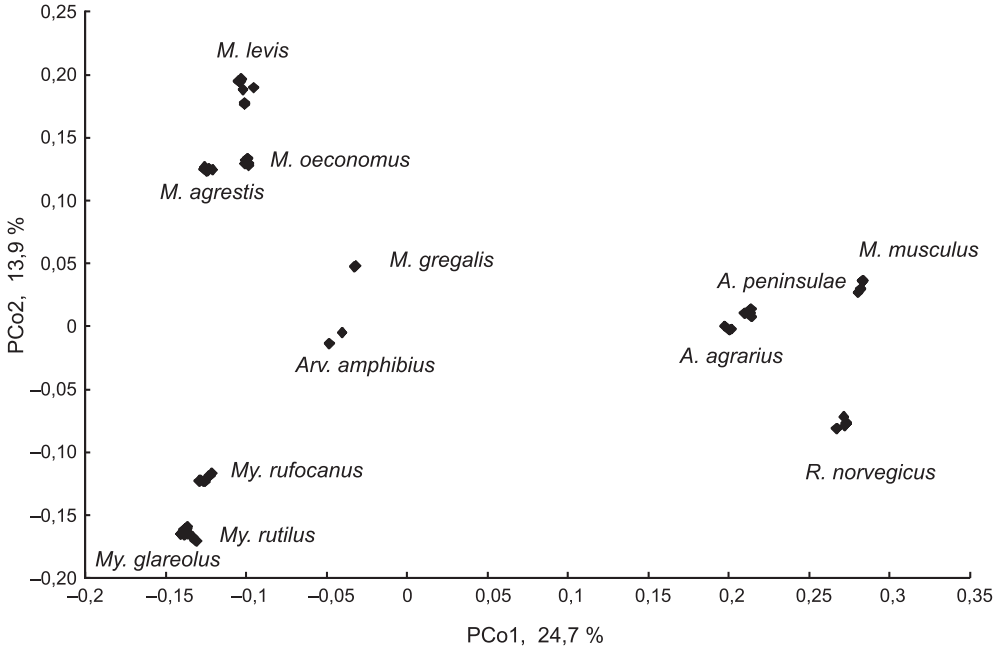


Рис. 2. Расположение нуклеотидных последовательностей на плоскости I–II главных координат.

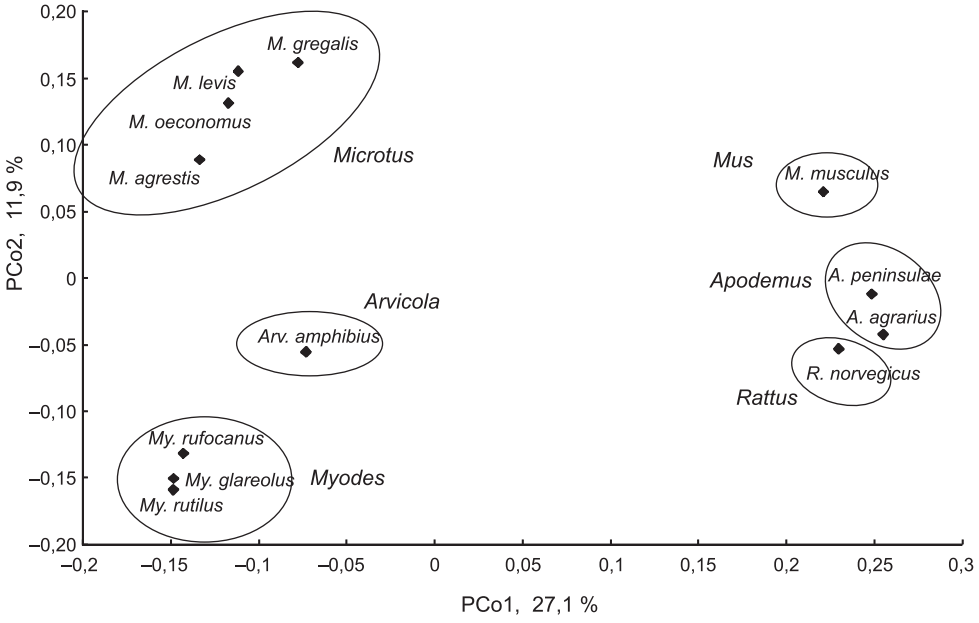


Рис. 3. Расположение видов на плоскости I–II главных координат.

mal species …, 2005), за исключением одного вида – *M. gregalis*, который заметно отличается по третьей главной координате (рис. 4) от других видов рода *Microtus*. Обособленность *M. gregalis* видна и на филодендрограммах, полученных методом UPGMA (невзвешенного попарного среднего) по матрице евклидовых расстояний между видами (рис. 5) и по матрице *LogDet*-расстояний между исходными последовательностями (бутстреп = 1000) (рис. 6). Это совпадает с полученными ранее результатами на основании неметрического шкалирования матрицы расстояний Кимуры (Ковалева и др., 2012) и согласуется с современными тенденциями в зоологической систематике полевок (Абрамсон, Лисовский, 2012).

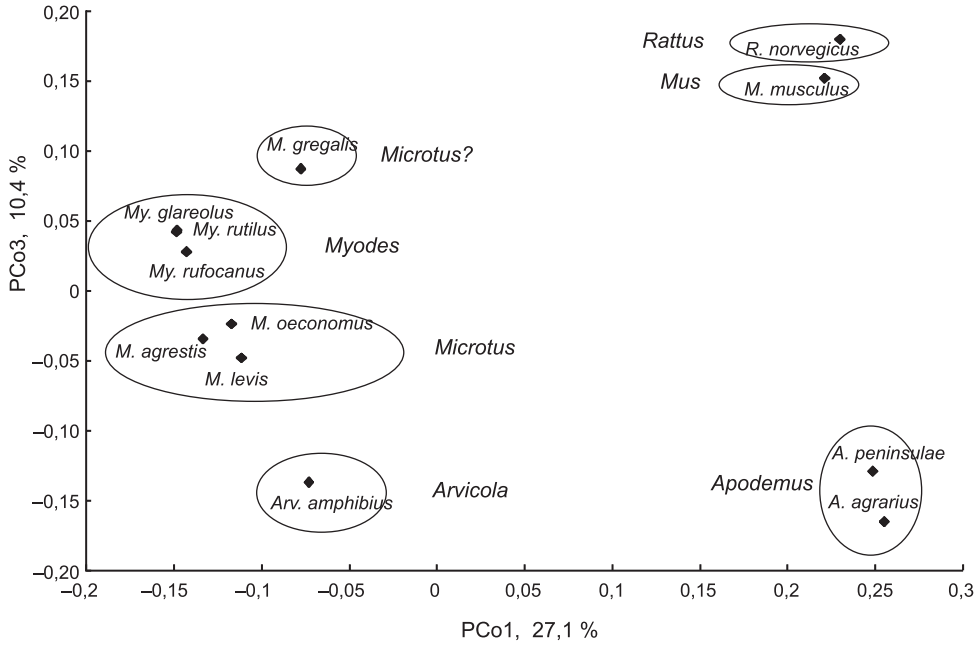


Рис. 4. Расположение видов на плоскости I–III главных координат.

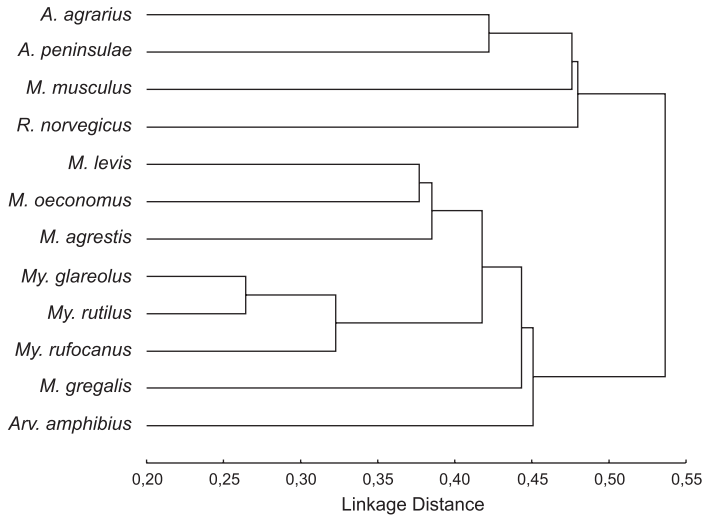


Рис. 5. Филодендрограмма, полученная методом UPGMA (невзвешенного попарного среднего) по матрице евклидовых расстояний между видами.

ЗАКЛЮЧЕНИЕ

Почти все известные генетические расстояния обладают двумя существенными недостатками: даже при разных вероятностях замен нуклеотидов оценивают только суммарное число замен; не являются геометрическими расстояниями. В работе предложены евклидовы аналоги расстояний Джукса–Кантора, *LogDet* и Кимуры – E_p-дистанция и E_{PQ}-дистанция. В E_{PQ}-дистанции числа транзиций и трансверсий взяты с разными весами. Евклидовость предложенных расстояний позволяет дополнительно применять весь арсенал методов многомерного анализа. На реальных данных проиллюстрировано практическое использование E_{PQ}-дистанции. Показана высокая корреляция всех пяти расстояний.

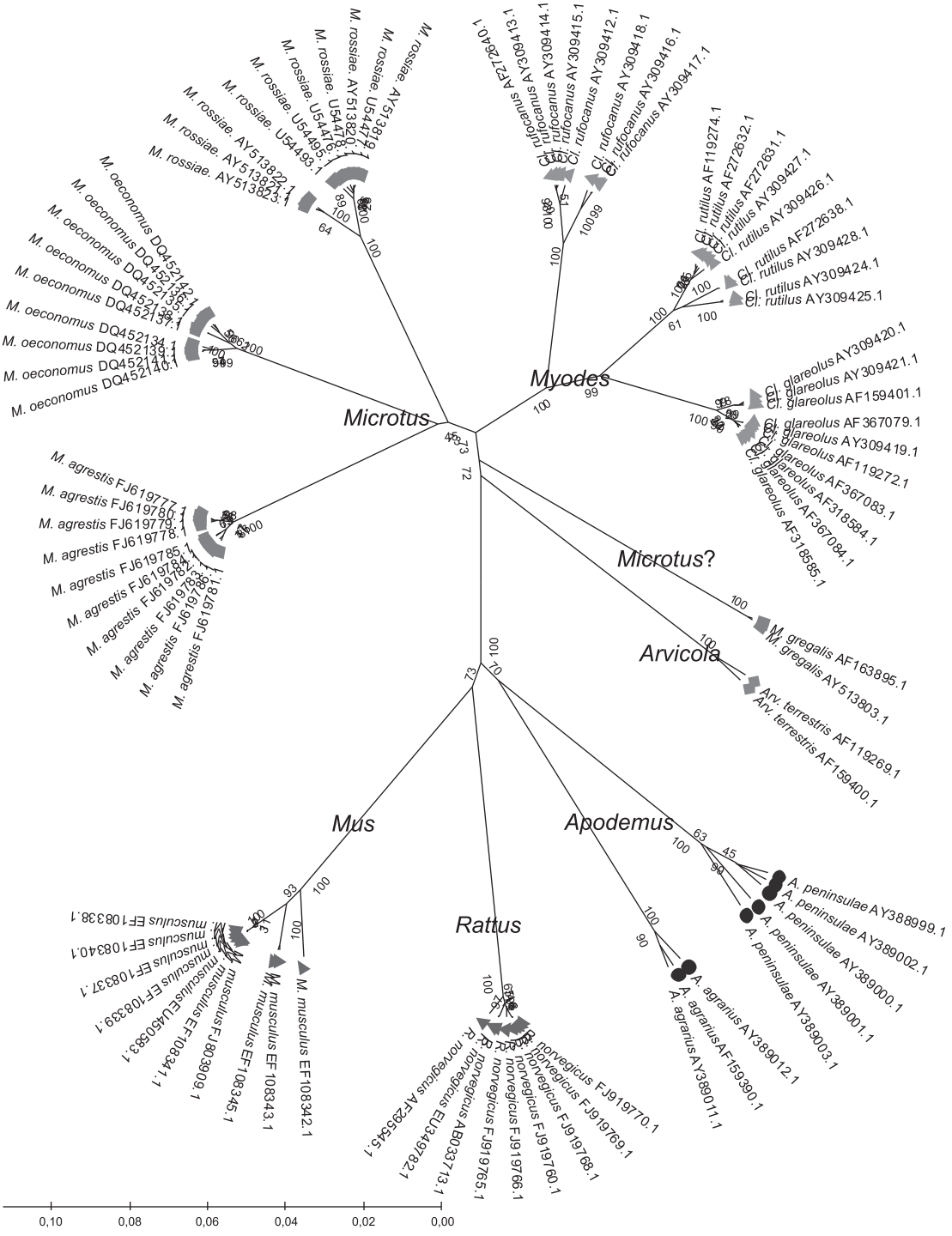


Рис. 6. Филодендрограмма, полученная методом UPGMA (невзвешенного попарного среднего) по матрице *LogDet*-расстояний между исходными последовательностями.

БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке Программы Президиума РАН – Интеграция РАН (6.8 и 28), Президента РФ (НШ-5278 2012.4), Интеграционного проекта СО РАН (18.13), Проекта фундаментальных исследований СО РАН и УрО РАН (70) и РФФИ (11-04-00141а, 13-07-00315а).

ВИДЫ И НОМЕРА ПОСЛЕДОВАТЕЛЬНОСТЕЙ ЦИТОХРОМА *b* мтДНК В БАЗЕ ДАННЫХ GENBANK

M. agrestis – FJ619777–FJ619786; *M. levis* – AY513819–AY513823, U54476–U54478, U54493, U54495; *M. oeconomus* – DQ452134–DQ452142; *M. gregalis* – AF163895, AY513803; *Arv. terrestris* – AF119269, AF159400; *My. glareolus* – AF119272, AF159401, AF318584, AF318585, AF367079, AF367083, AF367084, AY309419–AY309421; *My. rufocanus* – AF272640, AY309412–AY309418; *My. rutilus* – AF119274, AF272631, AF272632, AF272638, AY309424–AY309428; *A. agrarius* – AF159390, AY389011, AY389012; *A. peninsulae* – AY388999, AY389000–AY389003; *Mus musculus* – EF108337 – EF108343, EF108345, EU450583, FJ803909; *R. norvegicus* – AB033713, AF295545, EU349782, FJ919760, FJ919765, FJ919766, FJ919768– FJ919770.

ЛИТЕРАТУРА

Абрамсон Н.И., Лисовский А.А. Полевочьи // Млекопитающие России: систематико-географический справочник / Ред. И.Я. Павлинов, А.А. Лисовский. М.: КМК, 2012. С. 220–276.

Ефимов В.М., Штайгер И.А., Полунин Д.А. и др. Программно-алгоритмический комплекс для многомерного анализа микрочиповых данных // II Междунар. науч.-практ. конф. «Постгеномные методы анализа в биологии, лабораторной и клинической медицине: геномика, протеомика, биоинформатика». Новосибирск, Россия, 14–17 ноября, 2011. С. 120.

Ковалева В.Ю., Абрамов С.А., Дупал Т.А. и др. Анализ соответствия и комбинирование молекулярно-генетических и морфологических данных в зоологической систематике // Изв. РАН. Сер. биол. 2012. Вып. 4. С. 404–414.

Ковалева В.Ю., Литвинов Ю.Н., Ефимов В.М. Землеройки (Soricidae, Eulipotyphla) Сибири и Дальнего Востока: комбинирование и поиск конгруэнтности молекулярно-генетических и морфологических данных // Зоол. журнал. 2013. Т. 92. Вып. 11. С. 1–15.

Лукашов В.В. Молекулярная эволюция и филогенетический анализ. М.: БИНОМ, Лаборатория знаний, 2009. 256 с.

Мельчакова М.А. Геометрические аналоги генетических расстояний: Магистерская диссертация. Новосибирск: НГУ, 2013. 33 с.

Мельчакова М.А., Ефимов В.М. О метрических свойствах эволюционных расстояний // Тез. докл. конф. «Соврем. пробл. математики, информатики и биоинформатики», посвящ. 100-летию А.А. Ляпунова, 11–14 окт. 2011 г. Новосибирск, 2011. С. 88.

Млекопитающие России: систематико-географический справочник / Ред. И.Я. Павлинов, А.А. Лисовский. М.: КМК, 2012. 604 с.

Ней М., Кумар С. Молекулярная эволюция и филогенетика. Киев: КВЩ, 2004. 418 с.

Петровский А.Б. Пространства множеств и мультимножеств. М.: Едиториал УРСС, 2003. 248 с.

Felsenstein J. Inferring phylogenies. Sunderland: Sinauer Associates, 2003. 664 p.

Hamming R.W. Error detecting and error correcting codes // Bell Syst. Tech. J. 1950. V. 29. No. 2. P. 147–160.

Jukes T.H., Cantor C.R. Evolution of protein molecules // Mammalian Protein Metabolism / Ed. H.N. Munro. N.Y.: Acad. Press, 1969. P. 21–132.

Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences // J. Mol. Evol. 1980. V. 16. No. 2. P. 111–120.

Klingenberg C.P., Ekau W. A combined morphometric and phylogenetic analysis of an ecomorphological trend: pelagization in Antarctic fishes (Perciformes: Nototheniidae) // Biol. J. Linn. Soc. 1996. V. 59. No. 2. P. 143–177.

Klingenberg C.P., Gidaszewski N.A. Testing and quantifying phylogenetic signals and homoplasy in morphometric data // Syst. Biol. 2010. V. 59. No. 3. P. 245–261.

Lockhart P.J., Steel M.A., Hendy M.D., Penny D. Recovering evolutionary trees under a more realistic model of sequence evolution // Mol. Biol. Evol. 1994. V. 11. No. 4. P. 605–612.

Mammal Species of the World: a Taxonomic and Geographic Reference / Eds D.E. Wilson, D.M. Reeder. 3rd ed. Baltimore: J. Hopkins Univ. Press, 2005. 2142 p. Available at http://www.departments.bucknell.edu/biology/resources/msw3/ browse.asp

Mantel N. The detection of disease clustering and a generalized regression approach // Cancer Res. 1967. V. 27. P. 209–220.

Mantel N., Valand R.S. A technique of nonparametric multivariate analysis // Biometrics. 1970. V. 26. P. 547–558.

Polly P.D., Lawing A.M., Fabre A.C., Goswami A. Phylogenetic principal components analysis and geometric morphometrics // Hystrix, the Italian J. Mammalogy. 2013. V. 24. No. 1. P. 1–9.

Revell L.J. Size-correction and principal components for interspecific comparative studies // Evolution. 2009. V. 63. P. 3258–3268.

Shepard R.N. The analysis of proximities: multidimensional scaling with an unknown distance function. 1 // Psychometrika. 1962. V. 27. No. 2. P. 125–140.

Tamura K., Peterson D., Peterson N. *et al*. MEGA5: molecular evolutionary genetics analysis using maximum likelihood; evolutionary distance; and maximum parsimony methods // Mol. Biol. Evol. 2011. V. 28. P. 2731–2739.

Torgerson W.S. Multidimensional scaling: I. Theory and method // Psychometrika. 1952. V. 17. No. 4. P. 401–419.

Zharkikh A. Estimation of evolutionary distances between nucleotide sequences // J. Mol. Evol. 1994. V. 39. P. 315–329.

GEOMETRIC PROPERTIES OF EVOLUTIONARY DISTANCES

V.M. Efimov[1, 2, 3], M.A. Melchakova[4], V.Yu. Kovaleva[2]

[1] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia, e-mail: efimov@bionet.nsc.ru;
[2] Institute of Systematics and Ecology of Animals SB RAS, Novosibirsk, Russia;
[3] Tomsk National Research State University, Tomsk, Russia;
[4] Novosibirsk National Research State University, Novosibirsk, Russia

Summary

One way to study the variability of biologic objects is their geometrization: the objects are presented by points in a multidimensional space in such a way that the distances between the points would be best consistent with the dissimilarities between objects. If the dissimilarities between the objects are Euclidean distances, this task (up to translation, rotation and reflection) is solved by metric scaling. We consider the metric properties of some well-known evolutionary distances of nucleotide sequences. It is shown that the Jukes-Cantor and Kimura distances are not metrics. We introduce a new Kimura distance analog, the *PQ*-distance. It is shown that the p and *PQ* distances are the squares of Euclidean metrics named E_p-distance and E_{PQ}-distance, respectively. The applicability of the E_{PQ} distance is illustrated by the example of a cytochrome b sequence set of 12 rodent species from West Siberia and Altai, taken from the GenBank, and compared with the results of the use of the *LogDet*-distance.

Key words: nucleotide sequences, evolution models, phylogenetic reconstructions, genetic distances, geometrization, zoological systematics..

УДК 57.017.64,576.32,573.2

МОДЕЛИРОВАНИЕ БИОМЕХАНИКИ И МОРФОДИНАМИКИ РАСТЕНИЙ В ПАКЕТЕ COMSOL

© 2013 г. **С.В. Николаев**

Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия,
e-mail: nikolaev@bionet.nsc.ru

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

В статье дан краткий обзор пакета COMSOL Multiphysics; показаны способы построения модели и спецификации задачи в пакете COMSOL; продемонстрированы возможности пакета при изучении нескольких конкретных проблем биомеханики и морфодинамики растений.

Ключевые слова: COMSOL Multiphysics, метод конечных элементов, математическое моделирование, биомеханика, морфодинамика, атомная силовая микроскопия, АСМ, клетки растений, стенка растительной клетки.

ВВЕДЕНИЕ

В настоящее время в моделировании биологических процессов преобладают модели с сосредоточенными параметрами, описывающие динамику переменных состояния системы (например, концентрацию биомолекул или генов и их регуляторов) в областях, искусственно стянутых в точку. Хотя очевидно, что для понимания биологических явлений на системном уровне необходимо учитывать тот факт, что биологические процессы в клетках, органах, организмах основаны не только на молекулярно-генетической регуляции. Необходимыми атрибутами биологических процессов являются также транспорт молекул, перемещения материальных масс как сплошных сред и их деформации, которые протекают в пространственных областях, имеющих разнообразную геометрию. Одним из биологически значимых результатов этих процессов является изменение геометрии этих областей – морфодинамика, которая наблюдается при движениях биологических систем и их росте, и должно рассматриваться как механическое явление. Форма целого организма и его анатомических структур является одновременно результатом и диагностическим признаком функционирования организма – от экспрессии генов до транспортных процессов. Учет важности формы и того, что процессы протекают в пространстве, приводит к необходимости развития моделей с распределенными параметрами (в противовес моделям с сосредоточенными параметрами), рассматриваемых на областях с реальной геометрией. Метод конечных элементов является одним из немногих вычислительных методов, которые можно использовать для решения подобных задач и, пожалуй, единственным, который реализован в пакетах, позволяющих использовать его специалистами, не являющимися экспертами в данном вычислительном методе. Среди таких пакетов есть как коммерческие (COMSOL, ANSYS, ABACUS), так и свободные (Elmer). Следует отметить, что некоторые из этих пакетов имеют в названии дополнение «Multiphysics». Это отражает важное обстоятельство: пользовательский интерфейс в этих пакетах позволяет специфицировать не вычислительный алгоритм для некоторой задачи, а саму задачу, принадлежащую к какой-либо области физики, компоновать эти задачи в рамках одной модели, дополняя при необходимости своими задачами, специфицированными в терминах основных типов уравнений математической физики. Несмотря на то что за последние несколько лет продемонстрированы

уникальные возможности метода конечных элементов для моделирования биологических систем разных уровней – от субклеточных структур до целого организма, – метод не стал привычным инструментом моделирования при изучении медико-биологических проблем. Интересные применения метода можно найти в статьях (Atchley, Hall, 1991; You, Harvey, 1993; Missel, 2000; Rayfield, 2005, 2007; Richmond *et al.*, 2005; Tang *et al.*, 2006; Chatziprodromou *et al.*, 2007; Barreira *et al.*, 2011; Kraft *et al.*, 2012).

ПАКЕТ COMSOL MULTIPHYSICS

Пакет COMSOL развивался как средство для решения инженерно-физических задач для систем со сложной геометрией и гетерогенной структурой. Пользовательский интерфейс пакета позволяет специфицировать задачи, приводящие практически ко всем типам уравнений математической физики, таким как электромагнетизм, акустика, транспортные процессы в химии, электрохимия, механика сплошных сред, теплопередача, физика плазмы, структурная механика, – набор этих модулей зависит от лицензии на конкретную комплектацию пакета. Для примера предположим, что надо рассчитать распределение температуры T в некотором организме. Известно, что при изучении экофизиологических проблем правильно использовать этот параметр вместо температуры окружающей среды (Campbell, Norman, 2000). Это распределение можно найти, решая уравнение теплопроводности:

$$\rho C_p \frac{\partial T}{\partial t} + \nabla \cdot (-k\nabla T) = Q.$$

Если надо найти стационарное решение, то $\frac{\partial T}{\partial t} = 0$. Для решения этой задачи надо задать теплопроводность k, теплоемкость C_p, плотность ρ материала тела и источники тепла Q внутри организма. Кроме того, необходимо определить начальные условия и задать типы и интенсивности его теплообмена с окружающей средой, например, $-\overline{n} \cdot \overline{q} = q_0$ на границе тела $\partial\Omega$, где $\overline{n}$ – вектор нормали к границе, $\overline{q} = -k\nabla T$ – вектор теплового потока внутри тела на его границе и q_0 – интенсивность потока тепла внутрь тела через границу. Далее можно запускать

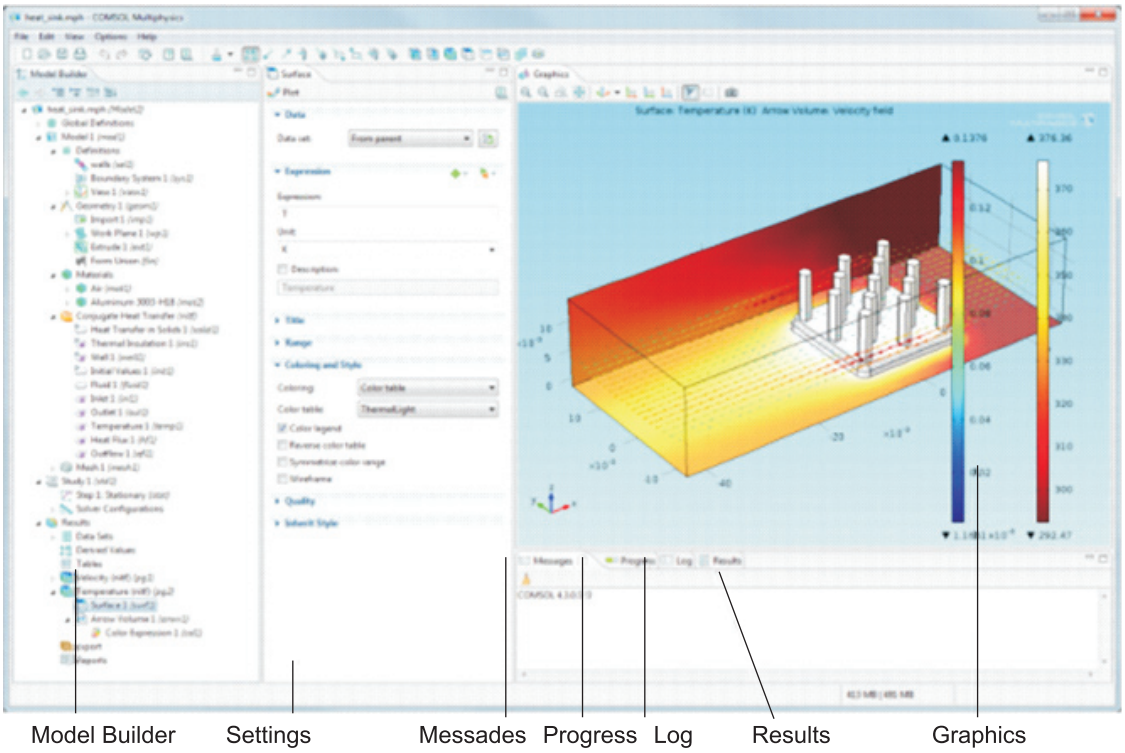


Рис. 1. Графический интерфейс пакета COMSOL 4.3b состоит из 4 окон: «Model Builder» (конструктор модели), «Settings» (спецификации), «Graphics» (графики) и «Messages/Progress/Log/Results» (информация о ходе процесса вычисления). Рисунок взят из документации к пакету COMSOL.

задачу на счет, при этом задача автоматически транслируется в вычислительный алгоритм, основанный на методе конечных элементов.

Построение модели в интерфейсе пакета начинается с задания в окне «Settings» (рис. 1) геометрической размерности задачи: 0-, 1-, 2- или 3-мерная. Затем выбирается «физика» задачи. В вышеприведенном примере надо выбрать «Heat Transfer > Heat Transfer in Solids». Завершается спецификация структуры задачи выбором типа исследования, в нашем примере это «Stationary». После чего надо отметить окончание спецификации задачи (значок судейского флажка). Реакцией системы будет построение начального дерева структуры задачи в окне «Model Builder» (рис. 2).

Для дальнейшего построения дерева задачи надо щелкнуть правой кнопкой мышки на некоторый узел, чтобы раскрыть контекстно зависимое меню, опции из которого могут быть частями поддерева для данного узла. Например, для задания геометрической модели объекта надо правым щелчком мышки открыть меню на узле «Geometry». Используя базовые геометрические объекты и их композиции, можно построить достаточно сложные геометрические тела. Либо можно импортировать геометрическую модель, построенную с использованием какого-либо CAD-пакета.

После задания геометрии тела надо задать свойства материалов, из которых состоят разные части тела. Для этого надо раскрыть меню на узле «Materials» и либо задать свойства материала, либо выбрать их из базы данных. Причем для задания свойств COMSOL предложит таблицу, где будут присутствовать слоты только для тех параметров, которые необходимы для решения выбранной задачи.

Спецификация задачи производится через всплывающее меню интерфейса для соответствующей задачи; в приведенном примере это «Heat Transfer in Solids». Интерфейсы для каждой задачи, разумеется, разные и отражают теоретические основы соответствующей физической задачи. Поскольку в данном примере для формулировки задачи надо задать начальные условия, граничные условия, распределение источников и стоков тепла в теле, то интерфейс в этом случае предлагает опции для внесения данной информации в условия задачи.

Все выборы опций меню и задание спецификаций приводят к достраиванию дерева модели, и его вид в окне «Model Builder» изменяется.

После спецификации задачи можно в меню на узле «Study 1» выбрать опцию «Compute» и тем самым запустить задачу на счет. Многие параметры модели, специфицирующие дета-

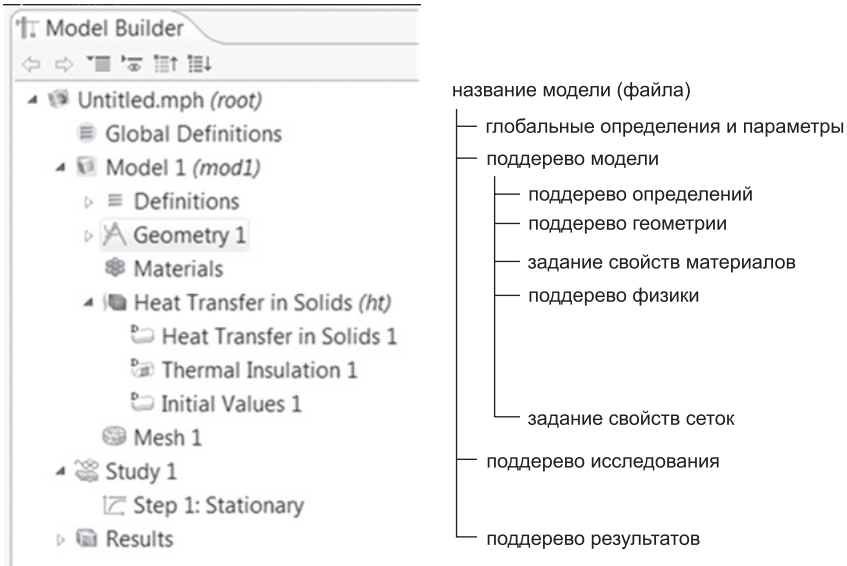


Рис. 2. Начальное состояние дерева модели в окне Конструктора модели.

Слева – фрагмент изображения окна, справа – схема дерева.

ли вычислительного алгоритма, будут заданы автоматически «по умолчанию», во многих случаях (если задача без особенностей) это приведет к успешному решению задачи и получению результата, который, разумеется, надо осмыслить. Иногда надо подкорректировать эти «параметры по умолчанию» (например задать другой размер сетки). Для этого надо выбрать соответствующий узел (например, «Mesh 1») и в его меню выбрать соответствующую опцию для внесения изменений.

Пакет COMSOL предлагает большой выбор способов обработки результатов счета: от построения отображения результатов решения на геометрическом теле и его виртуальных срезах до графиков, представляющих результаты серии вычислений с моделью при изменении параметров (режим «Parametric Sweep»). Эти возможности можно найти в иерархии раскрывающихся меню в узле «Results».

В итоге можно сказать, что с точки зрения пользователя пакет COMSOL Multiphysics можно рассматривать как средство для решения задач математической физики (которые естественным образом постоянно возникают и при моделировании биологических систем) с интуитивно понятным интерфейсом, зависящим от контекста.

МОДЕЛИРОВАНИЕ МЕХАНИКИ РАСТИТЕЛЬНОЙ КЛЕТКИ

Для понимания функционирования клетки при меняющихся внешних условиях, при ее росте и других процессах, в которых изменяется геометрия клетки, необходимо знать механические свойства структурных компонентов клетки, в частности клеточной стенки и собственно клетки, заключенной в клеточную стенку, – протопласта (Spatz *et al.*, 1999; Thompson, 2005; Suslov *et al.*, 2009; Fernandes *et al.*, 2012). Проведение измерений механических свойств (например жесткости и режимов деформирования) отдельно для стенки и протопласта является трудной экспериментальной задачей, особенно, если принять во внимание, что эти свойства активно изменяются при функционировании клетки (Proseus *et al.*, 1999; Thompson, 2001; Hansen *et al.*, 2011; Braybrook *et al.*, 2012; Routier-Kierzkowska *et al.*, 2012). В таком случае для проверки гипотез о предполагаемых механизмах можно использовать математическое моделирование, которое в данном случае должно включать моделирование механического поведения растительной клетки и/или ее компонентов (Bruce, 2003; Boudaoud, 2010; Geitmann, 2010; Dyson *et al.*, 2012).

Моделирование механики клеточной стенки в пакете COMSOL

Известно, что увеличение размеров клетки происходит за счет тургорного давления и пластической деформации клеточной стенки, при этом тургорное давление возникает за счет притока воды (как компонента биомассы) в клетку, и уравновешивается моментальной упругой деформацией клеточной стенки (Proseus *et al.*, 1999; Schopfer, 2006). Эти представления стимулировали моделирование упругих деформаций клеточной стенки при изучении биомеханики растительной клетки (Ortega, 1985; Boudaoud, 2010; Geitmann, 2010).

Такое моделирование в пакете COMSOL легко осуществить, выбрав для спецификации задачи раздел «Structural Mechanics». В данном узле имеются интерфейсы «Solid Mechanics» и «Shell». В первом случае клеточная стенка моделируется как объемное твердое тело, во втором – как двумерная поверхность, толщина которой служит параметром в уравнениях механики оболочек. Моделирование в механике оболочек приводит к построению двумерных сеток, что уменьшает время счета.

Поскольку нас будут интересовать детали деформации клеточной стенки по ее толщине, кратко разберем моделирование в интерфейсе «Solid Mechanics», причем нас будет интересовать стационарное решение задачи.

Основой для моделирования упругой деформации является закон Гука: $F = EA\frac{\Delta l}{l}$. Его можно прочитать следующим образом: чтобы растянуть проволоку с поперечным сечением A и длиной l на величину Δl, надо приложить силу F. Коэффициент пропорциональности E называется модулем упругости материала. Если обе части уравнения разделить на A, получим $\sigma = E\frac{\Delta l}{l}$, где величина σ называется напряжением.

В случае сплошной среды этот закон говорит, что каждая компонента тензора напряжений s_{ij} линейно связана с каждой компонентой тензора деформаций ε_{kl}:

$$s_{ij} = \sum_{k,l} C_{ijkl}\, \varepsilon_{kl}. \qquad (1)$$

В этой формуле компонента тензора напряжений является элементом матрицы размерности 3 × 3 и является *i*-той компонентой силы, действующей на единичной площадке, перпендикулярной оси *j*. Компонента тензора деформаций также является элементом матрицы размерности 3 × 3 и вычисляется по формуле $\varepsilon_{kl} = \frac{1}{2}\left(\frac{\partial u_l}{\partial x_k} + \frac{\partial u_k}{\partial x_l}\right)$, где u_l есть компонента вектора смещения u вдоль оси x_l некоторой точки внутри тела при его деформации. И, наконец, C_{ijkl} являются компонентами тензора четвертого ранга, который называется тензором упругости (Тимошенко, 1965; Фейнман и др., 1967). Другой способ записи равенства (1): $s = C : \varepsilon$, где знак двоеточия обозначает двойную свертку тензоров, представленную правой частью в (1).

Стационарная задача формулируется следующим образом: задано тело (возможно, состоящее из нескольких частей) определенной формы, заданы модули упругости, коэффициенты Пуассона и плотности материала для каждой из частей тела, заданы силы, приложенные на границе тела (поверхностные силы) и к каждому микроскопическому объему тела (объемные силы). Найти деформации и возникающие напряжения, при которых тело (и все его части) находится в механическом равновесии. Разумеется, напряжения и деформации в каком-то одном мысленно выделенном объеме тела зависят от таковых во всех остальных объемах. Поэтому задача нахождения равновесных распределений деформаций и напряжений приводит к вариационной задаче, в которой ищется минимум потенциальной энергии деформированного состояния (Бате, Вилсон, 1982). Для решения этой задачи COMSOL строит пространственную сетку, узлы которой являются вершинами пирамидок, и для каждой пирамидки генерируется уравнение механического равновесия $-\nabla \cdot \sigma = F_v$ с учетом равенства (1), где F_v – объемная сила. Далее COMSOL ищет решение, согласованное с приложенными поверхностными силами.

Для спецификации такой задачи в пакете COMSOL, используя меню на узле «Solid Mechanics», надо задать нагрузки, приложенные на участках поверхности тела, и ограничения на перемещения определенных элементов тела (если таковые имеются). После чего можно запускать задачу на счет.

Далее посмотрим, как это применить для нахождения деформации клеточной стенки на поверхности меристемы.

Биомеханика клетки туники апикальной меристемы побега

Апикальная меристема побега растения состоит из нескольких зон, в том числе центральной зоны (ЦЗ), периферической зоны (ПЗ) и организационного центра (ОЦ). Клетки в ЦЗ являются стволовыми клетками, растут и делятся медленно по сравнению с клетками в ПЗ. В экспериментальных работах по измерению жесткости клеточной стенки методами атомной силовой микроскопии (АСМ) было обнаружено, что клетки ЦЗ имеют более жесткие стенки, чем клетки ПЗ (Milani *et al.*, 2011; Peaucelle *et al.*, 2011; Braybrook *et al.*, 2012). Такой факт может укладываться в следующее представление: быстрый рост – менее жесткая стенка. В то же время измерения, проведенные на клетках корня из зон с разной скоростью роста, не выявили значимых различий в жесткости клеточной стенки (Fernandes *et al.*, 2012). Кроме того, измерения модулей упругости клетки и тургорного давления на образцах одного и того же вида растения даже при, казалось бы, одинаковых условиях дают различающиеся результаты. Это мотивировало разработку моделей механики клеточной стенки и процедуры измерения жесткости. В частности, для выяснения вопроса о корректности интерпретации измерений методами АСМ применяют моделирование самой процедуры измерения. Для демонстрации этого подхода воспроизведем модель (Milani *et al.*, 2011) с некоторыми упрощениями – будем рассматривать изотропный материал вместо ортотропного (у ортотропного материала модуль упругости в направлениях, параллельных к плоскости клеточной стенки, отличается от такового в перпендикулярном направлении). Суть метода измерения с помощью АСМ состоит в том, что производят надавливание на клетку пробником в виде иглы, имеющей

геометрию усеченной пирамиды (или конуса) с закругленной вершиной. Зная силу надавливания и перемещение пробника, можно в рамках определенной модели поведения материала рассчитать модуль упругости материала. В пакете COMSOL 4.3b имеется возможность для моделирования механического взаимодействия между телами – для этого надо задать контактные пары. Это пары поверхностей, которые принадлежат к двум разным телам, не могут пересекаться в пространстве, и при сближении между ними возникают распределенные силы, с которыми каждое из тел действует на другое. Использование контактных пар приводит к существенному увеличению времени счета. Для его снижения можно уменьшить размер моделируемого объема, используя имеющиеся в задаче геометрические симметрии. В данном случае имеется симметрия четвертого порядка относительно вращения вокруг продольной оси Oz. Это позволяет проводить расчет для одной четверти геометрического тела. Симметрия задачи позволяет также сформулировать граничные условия для поверхностей срезов – они могут перемещаться только в плоскостях соответствующих срезов. Для спецификации такого условия надо раскрыть меню в узле «Solid Mechanics» и выбрать опцию «Prescribed Displacement», после чего выбрать срез блока, например, параллельный координатной плоскости Oyz (см. рис. 3), и отметить ограничение на движение этого среза в виде $x = 0$, что и означает, что срез может двигаться только в плоскости Oyz. Внешние грани моделируемого блока клеточной стенки считаются неподвижными. Для спецификации этого условия надо выбрать опцию «Fixed Constraint» в узле «Solid Mechanics». После спецификации остальных параметров задачу можно запускать на счет. Результат счета – смещение пробника АСМ и деформация блока клеточной стенки (рис. 3).

Для того чтобы посмотреть, какой вклад вносит деформация всей клеточной стенки в смещение пробника (от этого зависит величина рассчитанного модуля упругости материала при измерении на АСМ), мы смоделировали процедуру измерения не на блоке, а на всей клеточной стенке. Для уменьшения времени счета контакт между пробником и клеточной стенкой моделировался как локальная нагрузка на стенку в месте контакта. Разумеется, механическая деформация при такой нагрузке отличается от деформации при контакте, однако эти различия носят локальный характер и можно ожидать, что они мало скажутся на деформации в масштабе клетки. Клеточную стенку смоделировали прямоугольной, будто она накрывает прямоугольную в сечении клетку. Клетка зажата со всех сторон соседними клетками, так что верхнюю стенку можно моделировать, считая, что она по периметру жестко зафиксирована. Учитывая симметрию (как в задаче с блоком стенки), считаем задачу для одной четверти стенки. На рис. 4 приведена клеточная стенка, деформированная тургорным давлением.

Получив решение, мы можем для наглядности построить срез модели, проходящий, например, параллельно длинной стороне стенки. Для этого в узле «Stress (solid)» поддерева «Results» раскрываем меню и выбираем опцию «Slice». Здесь можно задать число и положение срезов, а

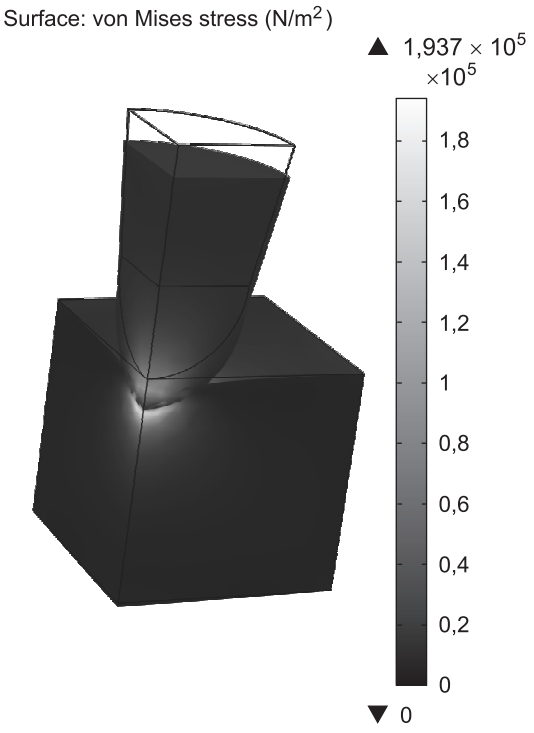


Рис. 3. Результат вычисления смещения пробника атомного силового микроскопа при надавливании на блок материала клеточной стенки.

Черная линия обозначает контуры исходной конфигурации, цветная кодировка – механические напряжения в пробнике и в блоке клеточной стенки в их финальной конфигурации.

также выбрать параметры задачи, которые будут отображаться на этих срезах (рис. 5).

Чтобы получить более полную картину поведения клеточной стенки при серии нажатий с разной силой, можно применить режим «Parametric Sweep» и результаты представить в виде графика зависимости перемещения пробника от приложенной силы (рис. 6).

При использовании данного подхода можно смоделировать процедуру измерения модуля

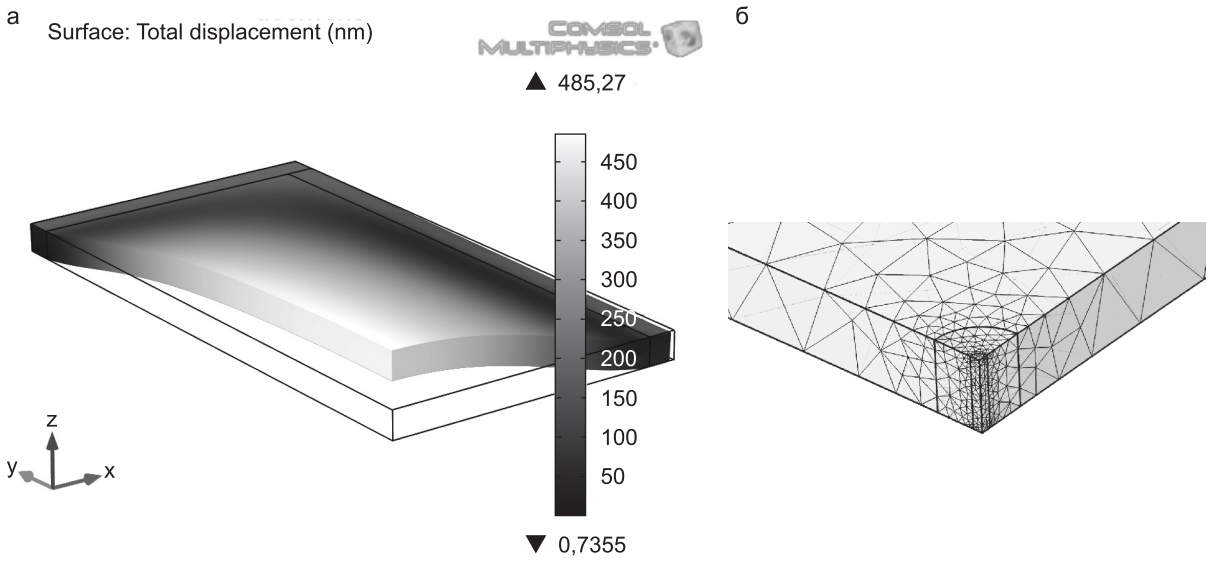


Рис. 4. Изображение расчетного фрагмента клеточной стенки под тургорным давлением (а). Ближний угол на изображении является центром клеточной стенки.

В центре прямоугольной стенки прикладывается сила, имитирующая нажатие пробником АСМ. Область центра смоделирована как объединение вложенных цилиндров. Это сделано для того, чтобы элементы сетки здесь были гораздо мельче, чем в остальной части клеточной стенки. (б) – фрагмент сетки клеточных элементов в районе ближнего угла изображения (а).

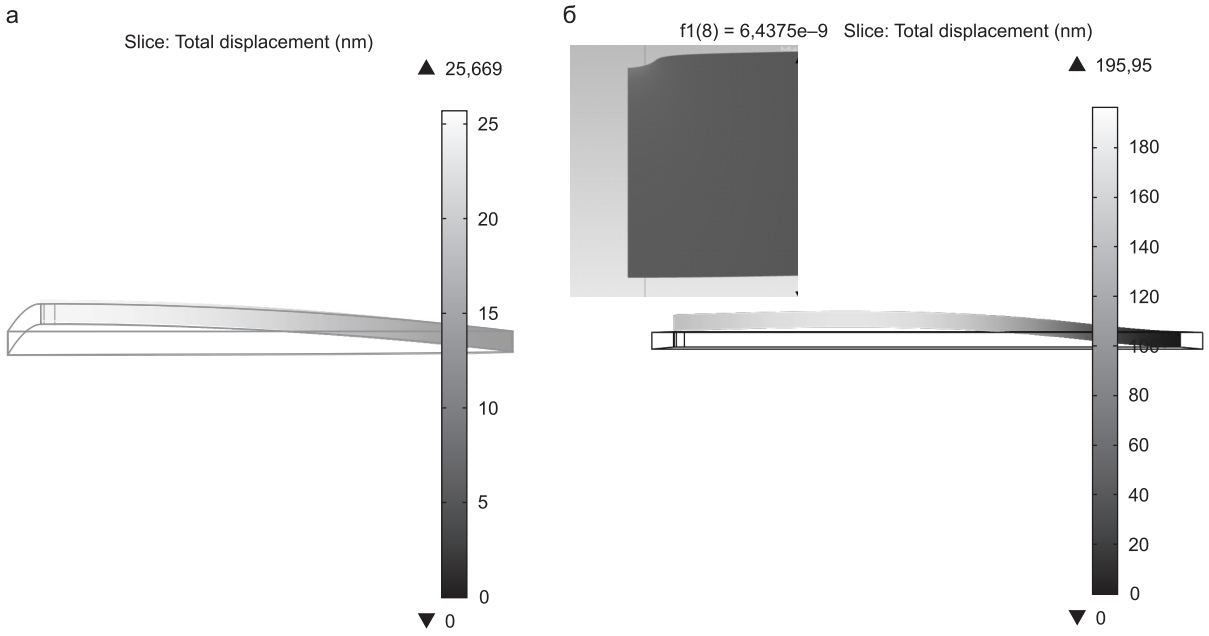


Рис. 5. Продольный срез посередине стенки. Исходная геометрия клеточной стенки, находящейся под тургорным давлением (а) и деформированная нажатием пробника АСМ с силой 6,4 нН (б).

На врезке увеличенная область стенки, деформированная нажатием пробника.

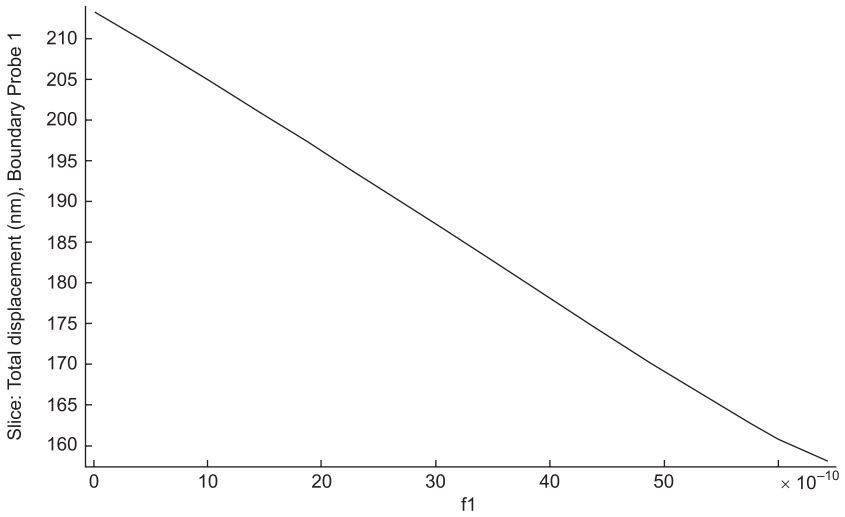


Рис. 6. График зависимости высоты пробника АСМ от приложенной силы.

упругости материала клеточной стенки методом АСМ, выяснить, какой вклад в перемещение пробника вносит деформация клеточной стенки, и таким образом оценить точность измерения данного параметра.

МОДЕЛИРОВАНИЕ МОРФОДИНАМИКИ РОСТА РАСТЕНИЯ

Имитация роста в интерфейсе «Structural Mechanics > Thermal Stress»

Рост организма и сопутствующее изменение его формы – морфодинамика – являются сложным взаимодействием самых разнообразных процессов (Boudaoud, 2010; Heisler *et al.*, 2010). Чтобы понять принципы управления этими процессами, необходимо знать последовательность механических событий: как распределены деформации при росте, какие напряжения они вызывают и как эти напряжения эволюционируют, т. е. либо релаксируют, возможно, вызывая механические деформации, либо остаются такими и даже возрастают и т. д.

В механике роста следует различать два типа деформаций. Во-первых, за счет притока компонентов биомассы и формирования структурной биомассы происходит увеличение объема (деформация). Во-вторых, разные части организма могут расти неравномерно, что приводит к неравномерной деформации. При этом возникают механические напряжения, которые вызывают второй тип деформаций – механические деформации с релаксацией этих напряжений, частичной или полной. Деформация за счет ростового увеличения объема происходит за счет активных биологических процессов с затратой энергии, в то время как сопутствующие релаксационные деформации ведут к уменьшению механической энергии и могут происходить как с затратой энергии, так и без нее. При этом энергия тратится таким образом, что ведет к уменьшению механических напряжений. Для имитации ростового увеличения объема мы использовали аналогию между увеличением объема за счет внедрения новых материальных частиц и увеличением объема за счет нагрева – в обоих случаях среднее расстояние между «старыми» частицами тела увеличивается (Volokh, 2004). Таким образом, чтобы специфицировать модель ростового увеличения объема в готовом интерфейсе «Structural Mechanics > Thermal Stress», можно интерпретировать $\alpha \cdot \theta$ в формуле $s = s_0 + C: (\varepsilon - \varepsilon_0 - \alpha \cdot \theta)$ как тензор прироста ε_g. Например, когда $\varepsilon_g = \varepsilon$, формула описывает деформацию без напряжения, т. е. «чистый рост» области. Для задания величины тензора прироста можно использовать контролируемый источник тепла с заданной целевой температурой и рассматривать материал как среду без теплопроводности. Это позволяет задавать произвольное распределение прироста по области и изучать результирующие деформацию и напряжение.

Мы использовали данный подход к моделированию роста для изучения механических на-

пряжений, возникающих в апикальной меристеме побега во время роста, а также для изучения перехода зародыша растения от глобулярной формы к ранней сердечковидной форме.

Моделирование деформаций при росте апикальной меристемы побега

Апикальной меристемой побега оканчивается каждая растущая ветка растения. В апикальной меристеме находятся стволовые клетки, которые растут и делятся и тем самым обеспечивают увеличение числа клеток растущей ветки. В апикальной меристеме выделяют наружные слои (в арабидопсисе – два слоя) клеток, которые формируют тунику (покров), и внутреннюю часть под туникой, так называемый корпус. Известно, что клетки туники делятся антиклинно, т. е. перпендикулярно поверхности меристемы, что приводит к образованию плоской ткани. В то же время рост и деление клеток корпуса меристемы не имеют преимущественной пространственной ориентации, что приводит к образованию объемной ткани. На основании экспериментальных данных ранее было выдвинуто предположение, что механические напряжения в клетках могут определять ориентацию клеточной стенки при ее делении (Nakielski, 2008; Hamant *et al.*, 2010; Mirabet *et al.*, 2011).

При отсутствии экспериментальных данных о распределении механических свойств в тканях меристемы для изучения распределения деформаций и механических напряжений в растущей меристеме мы построили ряд объемных моделей апикальной меристемы. В этих моделях апикальная меристема представлялась твердым телом, состоящим из двух частей: корпуса и оболочки, имеющих разные механические свойства. В первой серии вычислительных экспериментов были рассмотрены следующие варианты распределения скоростей роста и жесткости тканей меристемы: а) одинаковые модули упругости корпуса и оболочки, одинаковые скорости роста тканей корпуса и оболочки; б) модуль упругости ткани оболочки больше, чем ткани корпуса, одинаковые скорости роста тканей корпуса и оболочки; в) одинаковые модули упругости корпуса и оболочки, скорость роста ткани корпуса больше, чем ткани оболочки.

Выяснено, что в случаях (а) и (б) в тканях меристемы не возникает напряжений в результате роста. Небольшое напряжение в ткани в самой верхней части модели (рис. 7) можно трактовать как артефакт, возникающий при данной геометрии и ограничении движения нижней границы области условием Z = –1.

В случае (в) в ткани корпуса, растущего быстрее оболочки, возникает сжатие, а оболочка оказывается растянутой, о чем свидетельствует распределение давлений. Распределение тензоров напряжений показывает, что первые компоненты тензора в оболочке ориентированы параллельно ее поверхности и по величине гораздо больше вторых компонент. В то же время в корпусе первая и вторая компоненты тензора напряжений примерно одинаковы (рис. 8).

Если компоненты тензора являются сигналами, ориентирующими клеточные стенки при делениях (Nakielski, 2008; Mirabet *et al.*, 2011), то можно предполагать, что в оболочке (слоях L1, L2) эти стенки будут иметь преобладающую ориентацию, а в ткани корпуса такая ориентация будет выражена гораздо слабее, что

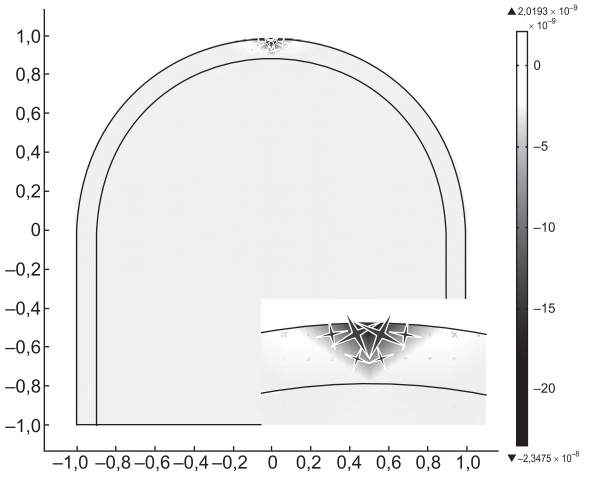


Рис. 7. Давление в тканях меристемы в случае одинакового роста корпуса и оболочки везде равно нулю.

Крайне малое отрицательное давление на верхушке меристемы (увеличено на врезке) – артефакт как следствие выбранной геометрии и ограничений на перемещение нижней границы области. Изменение цвета от черного к белому соответствует увеличению давления приблизительно от -2×10^{-8} до 2×10^{-9} Па, что является чрезвычайно малой величиной, практически неотличимой от нуля. Длина стрелок пропорциональна величине компонентов тензора напряжений.

согласуется с клеточным строением этих тканей (Kwiatkowska, 2004).

Во второй серии вычислительных экспериментов мы изучали влияние неоднородного распределения скоростей роста в корпусе и оболочке на деформацию формы меристемы (рис. 9).

Такая постановка вопроса является актуальной для выяснения роли распределения скоростей роста при формировании примордиев листьев на флангах меристемы. Предварительные результаты показывают, что распределения скоростей роста оказывают большое влияние на распределение тензоров напряжений в оболочке и корпусе меристемы, что в случае сигнальной роли этих напряжений может оказывать специфическое регуляторное воздействие на распределение дифференциальной экспрессии генов и ростовую морфодинамику.

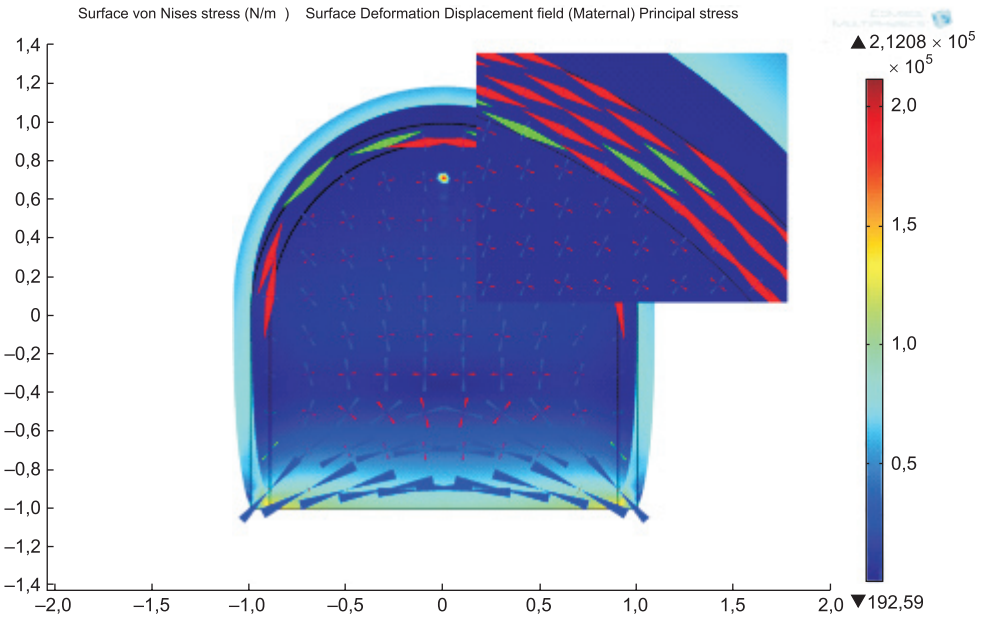


Рис. 8. Распределение компонентов тензора механических напряжений в ткани апикальной меристемы.

Чистый рост (рост области, который наблюдался бы в условиях, когда не возникают упругие напряжения) ткани корпуса много больше чистого роста ткани оболочки. На врезке видно, что первая компонента тензора напряжений (растяжение) в оболочке ориентирована параллельно поверхности меристемы и существенно больше второй компоненты. В то же время, как видно из рисунка, компоненты тензора напряжения (сжатия) ткани корпуса незначительно различаются. Черной линией обозначены контуры исходной формы меристемы. Длина стрелок пропорциональна величине компонентов тензора напряжений. Для наглядности деформация на изображении увеличена в 15 раз.

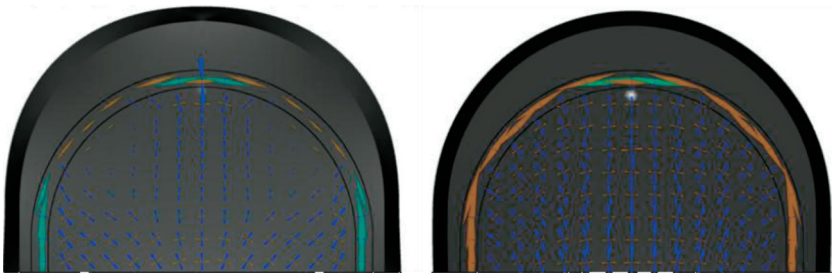


Рис. 9. Влияние неоднородного удельного прироста на ростовую морфодинамику.

Справа – однородный рост корпуса и однородный рост оболочки меристемы при различных величинах скоростей роста корпуса и оболочки. Слева – однородный рост ткани корпуса и неоднородный рост ткани оболочки. Скорость роста оболочки максимальна в областях, расположенных по направлениям +/– 45° от продольной оси (в начальной геометрии области). Видно, что в этих областях уменьшается механическое напряжение в оболочке и по этим направлениям происходит более интенсивный актуальный рост. Для наглядности деформация на изображении увеличена в 15 раз.

Таким образом, тензорные поля механических напряжений в растущей апикальной меристеме с разнообразным распределением главных осей тензоров могут возникать в результате неоднородного изотропного роста ткани.

Моделирование морфодинамики перехода от глобулярной к ранней сердечковидной стадии зародыша растения арабидопсиса

Начиная с одноклеточной стадии развития зародыш растения приобретает округлую форму, которая сохраняется на протяжении нескольких клеточных делений. Затем в результате неравномерного роста частей он начинает приобретать так называемую сердечковидную форму – происходит формирование билатеральной симметрии зародыша (Laux, Jürgens, 1997). На первом этапе изучения такой морфодинамики зародыша было необходимо выяснить, как влияет распределение прироста биомассы в зародыше растения на его форму, чтобы понять механику перехода из глобулярной формы в раннюю сердечковидную форму. Для этого была построена геометрическая модель зародыша в форме глобулы на суспензоре. Глобула состояла из трех слоев – туники, промежуточного слоя и внутренней части. Вычислительные эксперименты проводили по следующему сценарию: задавали модули упругости материала для каждого слоя, распределение прироста по объему зародыша и вычисляли результирующую форму. В итоге удалось подобрать такое распределение изотропного прироста областей зародыша (рис. 10), которое при заданных механических параметрах деформируется в форму, похожую на раннюю сердечковидную (рис. 11).

Конечно, на данном этапе результаты вычислительных экспериментов следует интерпретировать только качественно, поскольку требуются дополнительная оценка параметров модели и дальнейшие вычисления.

ЗАКЛЮЧЕНИЕ И ПЕРСПЕКТИВЫ

Моделирование биомеханики от клеток до тканей, органов и организма является актуальной задачей, возникающей как при решении фундаментальных проблем биологии, так и в рамках прикладных задач в биомедицинской инженерии и биотехнологии. Поскольку предметом

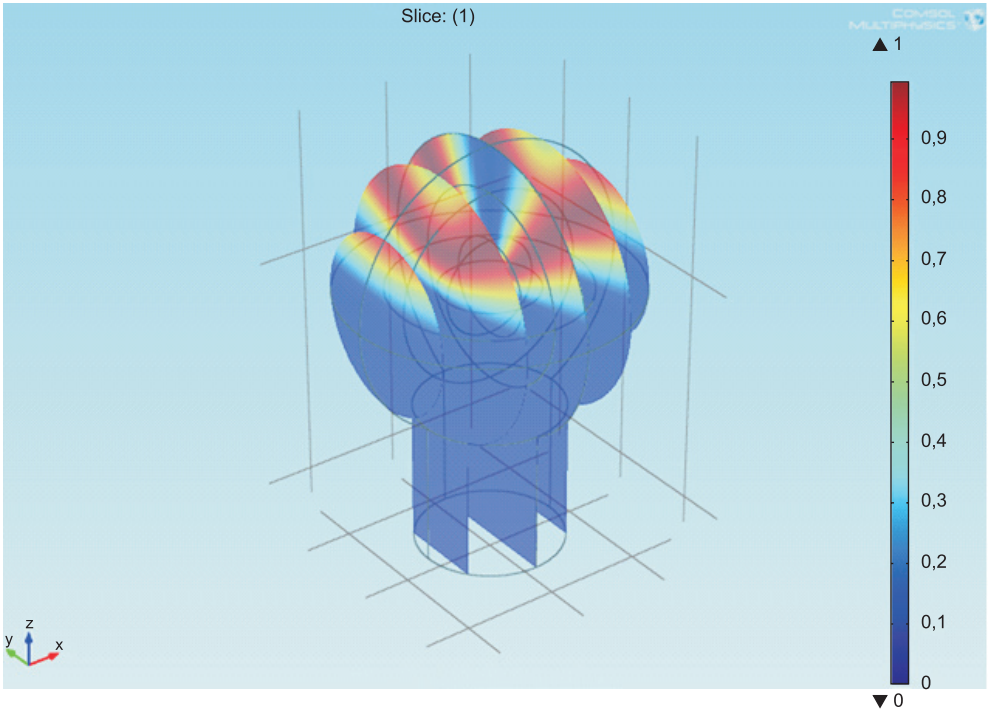


Рис. 10. Распределение удельного прироста областей глобулярного зародыша, приводящее к сердечковидной форме.

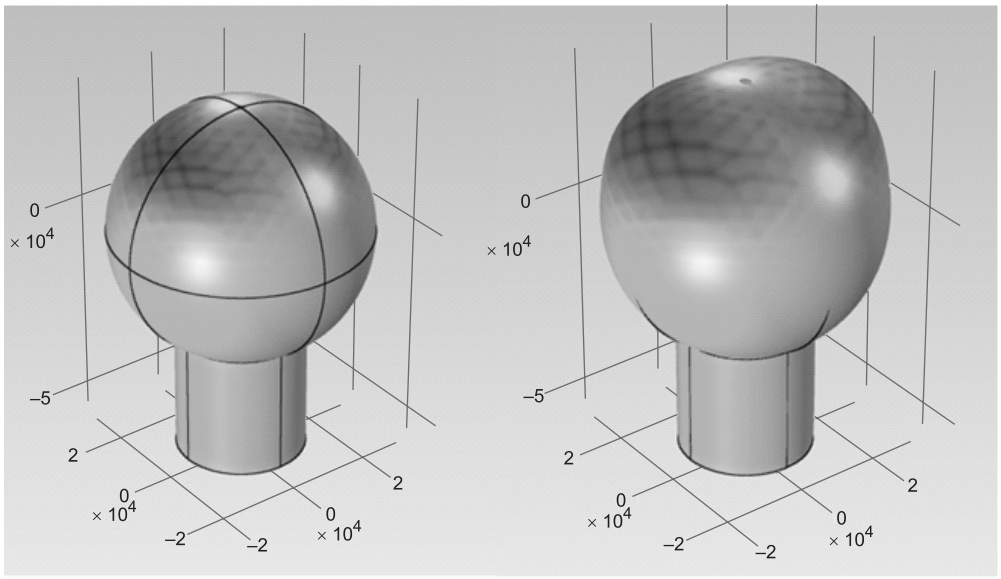


Рис. 11. Исходная глобулярная форма зародыша (слева) в результате неравномерного роста частей зародыша приобрела сердечковидную форму.

изучения биомеханики часто являются активные материалы (например, клетки и ткани), то наряду с механикой формулировки задач включают моделирование процессов управления, реализованных молекулярно-генетическими системами с транспортом веществ и, возможно, с нейрорегуляцией. В пакете COMSOL Multiphysics пользователь может специфицировать такие разнообразные процессы в одной модели и решать их методом конечных элементов для тел со сложной геометрией. Геометрическую модель изучаемого объекта пользователь может построить в пакете COMSOL либо импортировать сеточную модель, построенную по реальным снимкам с использованием специальных пакетов (например Avizo, Simpleware). Такая возможность открывает перспективы создания биоинформационной платформы для организации технологического конвейера с целью моделирования биологических объектов с реальной геометрией – от стека снимков (сканирующая микроскопия, томография и т. п.) до построения геометрической модели, построения на этой геометрии математической модели процессов и проведения вычислительных экспериментов с ней (Николаев и др., 2012). Данный подход позволяет получить информацию, которую невозможно получить с использованием современной экспериментальной техники.

БЛАГОДАРНОСТИ

Автор выражает благодарность Суперкомпьютерному центру Новосибирского государственного университета за предоставленную возможность использовать пакет COMSOL 4.3b.

Работа выполнена при частичной финансовой поддержке грантами РФФИ-НИСИ № 11-04-91397-а, РФФИ № 11–04–01748-а.

ЛИТЕРАТУРА

Бате К., Вилсон Е. Численные методы анализа и метод конечных элементов. М., Стройиздат, 1982. 448 с.

Николаев С.В., Колчанов Н.А., Голушко С.К. и др. Моделирование морфодинамики на ранних стадиях эмбриогенеза растения // Вавилов. журн. генет. и селекции. 2012. Т. 16. Вып. 4/1. С. 805–815.

Тимошенко С.П. Сопротивление материалов. Т. 1. Элементарная теория и задачи. М.: Наука, 1965.

Фейнман Р., Лейтон Р., Сэндс М. Фейнмановские лекции по физике. Т. 7. Физика сплошных сред. М.: Мир, 1967.

Atchley W.R., Hall B.K. A model for development and evolution of complex morphological structures // Biol. Rev. 1991. V. 66. P. 101–157.

Barreira R., Elliott C., Madzvamuse A. The surface finite element method for pattern formation on evolving biological surfaces // J. Mathemat. Biol. 2011. V. 63. P. 1095–1119.

Boudaoud A. An introduction to the mechanics of morphogenesis for plant biologists // Trends Plant Sci. 2010. V. 15. P. 353–360.

Braybrook S.A., Hofte H., Peaucelle A. Probing the mechanical contributions of the pectin matrix: insights for cell

growth // Plant Signal Behav. 2012. V. 7. P. 1037–1041.
Bruce D.M. Mathematical modelling of the cellular mechanics of plants // Philos. Trans. Roy. Soc. Lond. Series B: Biological Sciences. 2003. V. 358. P. 1437–1444.
Campbell G.S., Norman J.M. An Introduction to Environmental Biophysics. 2nd ed. Springer, 2000. 286 p.
Chatziprodromou I., Tricoli A., Poulikakos D., Ventikos Y. Haemodynamics and wall remodelling of a growing cerebral aneurysm: A computational model // J. Biomech. 2007. V. 40. P. 412–426.
Dyson R., Band L., Jensen O. A model of crosslink kinetics in the expanding plant cell wall: Yield stress and enzyme action // J. Theor. Biol. 2012. V. 307. P. 125–136.
Fernandes A.N., Chen X., Scotchford C.A. *et al.* Mechanical properties of epidermal cells of whole living roots of *Arabidopsis thaliana*: an atomic force microscopy study // Phys. Rev. E. 2012. V. 85. P. 021916.
Geitmann A. Mechanical modeling and structural analysis of the primary plant cell wall // Curr. Opin. Plant Biol. 2010. V. 13. P. 693–699.
Hamant O., Traas J. The mechanics behind plant development // New Phytol. 2010. V. 185. P. 369–385.
Hansen S.L., Ray P.M., Karlsson A.O. *et al*. Mechanical properties of plant cell walls probed by relaxation spectra // Plant Physiol. 2011. V. 155. P. 246–258.
Heisler M.G., Hamant O., Krupinski P. *et al.* Alignment between PIN1 polarity and microtubule orientation in the shoot apical meristem reveals a tight coupling between morphogenesis and auxin transport // PLoS Biol. 2010. V. 8. e1000516.
Kraft R.H., Mckee P.J., Dagro A.M., Grafton S.T. Combining the finite element method with structural connectome-based analysis for modeling neurotrauma: connectome neurotrauma mechanics // PLoS Comput Biol. 2012. V. 8. P. e1002619.
Kwiatkowska D. Structural integration at the shoot apical meristem: models, measurements, and experiments // Amer. J. Bot. 2004. V. 91. P. 1277–1293.
Laux T., Jürgens G. Embryogenesis: a new start in life // Plant Cell. 1997. V. 9. P. 989–1000.
Milani P., Gholamirad M., Traas J. *et al. In vivo* analysis of local wall stiffness at the shoot apical meristem in Arabidopsis using atomic force microscopy // Plant J. 2011. V. 67. P. 1116–1123.
Mirabet V., Das P., Boudaoud A., Hamant O. The role of mechanical forces in plant morphogenesis // Annu. Rev. Plant Biol. 2011. V. 62. P. 365–385.
Missel P. Finite element modeling of diffusion and partitioning in biological systems: the infinite composite medium problem // Ann. Biomed. Eng. 2000. V. 28. P. 1307–1317.
Nakielski J. The tensor-based model for growth and cell divisions of the root apex. I. The significance of principal directions // Planta. 2008. V. 228. P. 179–189.
Ortega J.K. Augmented growth equation for cell wall expansion // Plant Physiol. 1985. V. 79. P. 318–320.
Peaucelle A., Braybrook S.A., Le Guillou L. *et al.* Pectin-induced changes in cell wall mechanics underlie organ initiation in Arabidopsis // Curr. Biol. 2011. V. 21. P. 1720–1726.
Proseus T.E., Ortega J.K., Boyer J.S. Separating growth from elastic deformation during cell enlargement // Plant Physiol. 1999. V. 119. P. 775–784.
Rayfield E.J. Using finite-element analysis to investigate suture morphology: A case study using large carnivorous dinosaurs // Anat. Rec. Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology. 2005. V. 283A. P. 349–365.
Rayfield E.J. Finite element analysis and understanding the biomechanics and evolution of living and fossil organisms // Annu. Rev. Earth Planet. Sci. 2007. Book Series: Annu. Rev. Earth Planet. Sci. V. 35. P. 541–576.
Richmond B.G., Wright B.W., Grosse I. *et al.* Finite element analysis in functional morphology // Anat. Rec. Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology. 2005. V. 283A. P. 259–274.
Routier-Kierzkowska A.-L., Weber A., Kochova P. *et al.* Cellular force microscopy for *in vivo* measurements of plant tissue mechanics // Plant Physiol. 2012. V. 158. P. 1514–1522.
Schopfer P. Biomechanics of plant growth // Am. J. Bot. 2006. V. 93. P. 1415–1425.
Spatz H., Kohler L., Niklas K. Mechanical behaviour of plant tissues: composite materials or structures? // J. Experim. Biol. 1999. V. 202. P. 3269–3272.
Suslov D., Verbelen J.-P., Vissenberg K. Onion epidermis as a new model to study the control of growth anisotropy in higher plants // J. Experim. Bot. 2009. V. 60. P. 4175–4187.
Tang Y., Cao G., Chen X. *et al.* A finite element framework for studying the mechanical response of macromolecules: application to the gating of the mechanosensitive channel MscL // Biophys. J. 2006. V. 91. P. 1248–1263.
Thompson D.S. How do cell walls regulate plant growth? // J. Experim. Bot. 2005. V. 56. P. 2275–2285.
Thompson D.S. Extensiometric determination of the rheological properties of the epidermis of growing tomato fruit // J. Experim. Bot. 2001. V. 52. P. 1291–1301.
Volokh K.Y. A simple phenomenological theory of tissue growth // Mech. Chem. Biosyst. 2004. V. 1. P. 147–160.
You T.J., Harvey S.C. Finite element approach to the electrostatics of macromolecules with arbitrary geometries // J. Computat. Chem. 1993. V. 14. P. 484–501.

MODELING OF PLANT BIOMECHANICS AND MORPHODYNAMICS IN THE COMSOL PACKAGE

S.V. Nikolaev

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: nikolaev@bionet.nsc.ru

Summary

This is a short review of the COMSOL Multiphysics package. Model building in COMSOL and methods of biophysical problem specification are demonstrated. Examples of the investigation of several problems in plant biomechanics and morphodynamics are considered.

Key words: COMSOL Multiphysics, finite element method, mathematical modeling, biomechanics, morphodynamics, atomic force microscopy, plant cell, plant cell wall.

УДК 57.032:577.38:5-76

МОДЕЛИ РЕГУЛЯЦИИ СТРУКТУРЫ НИШИ СТВОЛОВЫХ КЛЕТОК В АПИКАЛЬНОЙ МЕРИСТЕМЕ ПОБЕГА

© 2013 г. **У.С. Зубаирова, С.В. Николаев**

Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия,
e-mail: ulyanochka@bionet.nsc.ru

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

Экспериментальные данные, полученные к настоящему времени, привели к определенным представлениям о регуляции ниши стволовых клеток в апикальной меристеме побега. Для проверки непротиворечивости этих представлений и их согласованности с экспериментальными данными применяется математическое моделирование. В статье рассматриваются математические модели регуляции ниши стволовых клеток в апикальной меристеме побега, предложенные разными авторами; анализируются экспериментальные основания и рабочие гипотезы, формализованные в этих моделях, и выявляются методические различия в подходах к построению этих моделей.

Ключевые слова: апикальная меристема побега, ниша стволовых клеток, CLAVATA3, WUSCHEL, математическая модель.

ВВЕДЕНИЕ

Все органы взрослого растения формируются из апикальной меристемы побега (АМП), находящейся на кончике каждого вегетативного побега. АМП является одной из важнейших структур для роста и развития, так как именно здесь находится ниша стволовых клеток, дающих начало всем клеткам наземной части растения (Sharma *et al*., 2003).

Несмотря на то что АМП имеет небольшие размеры, ее структура достаточно сложна. Выделяют зоны (Newman, 1965) и области клеток, отличающиеся по своим свойствам и функциям, маркированные экспрессией определенных генов и находящиеся в определенном пространственном расположении друг относительно друга на протяжении всей жизни растения (Bowman, Eshed, 2000; Brand *et al.*, 2001; Traas, Doonan, 2001). В АМП выделяют следующие зоны (Gross-Hardt, Laux, 2003; Lenhard, Laux, 2003; Kwiatkowska, 2004): центральная (ЦЗ), периферическая (ПЗ), риб-зона (РЗ), зона листовых примордиев (ЛП), а также организационный центр (ОЦ). В ЦЗ содержится запас стволовых клеток, которые в процессе деления вытесняются в ПЗ и РЗ, клетки в этой зоне характеризуются низкими темпами деления. ПЗ является переходной зоной, клетки здесь делятся быстрее, в дальнейшем клетки из ПЗ переходят в ЛП, РЗ или зону между зачатками листьев. Как и ПЗ, ОЦ также является переходной зоной, она лежит непосредственно под ЦЗ и имеет важную функцию в поддержании ее размеров. Клетки, вытесненные из ОЦ, переходят в РЗ, где образуют сосудистую систему и паренхиму стебля.

Помимо этих зон в структуре АМП также выделяют 3 слоя: L1 (эпидермис), L2 (субэпидермис) и L3 (кортекс). Клетки слоев L1 и L2, которые также называют туникой, делятся в основном антиклинно, обеспечивая поверхностный рост организма. Клетки L3 не имеют определенного направления деления и тем самым обеспечивают объемный рост организма. Рост и деление клеток приводят к потоку клеток из ЦЗ в ПЗ и ОЦ и далее по корпусу АМП. Однако, несмотря на такой поток клеток, положения зон относительно верхушки АМП остаются постоянными.

В результате гистоморфологического и клонального анализов было выяснено, что

стволовые клетки расположены в трех верхних слоях АМП (L1, L2, L3), концентрируясь вокруг ее центральной оси (Gross-Hardt, Laux, 2003; Lenhard, Laux, 2003; Kwiatkowska, 2004).

К настоящему времени у *A. thaliana* выявлен ряд генов, мутации которых приводят к изменению структуры АМП и ниши стволовых клеток в ней. Так, характерной особенностью клеток ЦЗ является экспрессия гена *CLV3*, а в клетках ОЦ наблюдается экспрессия гена *WUS* (Yadav *et al.*, 2009). В клетках ОЦ и его ближайшего окружения наблюдается экспрессия генов *CLV1* и *CLV2*, продуктами которых являются субъединицы гетеродимерного рецепторного комплекса CLV1/CLV2, локализованного на клеточной мембране (Schoof *et al.*, 2000; Williams, Fletcher, 2005). Показано, что образующийся в результате процессинга белка CLV3 короткий пептид выходит из клетки, связывается с CLV1/CLV2 в клетках вокруг ОЦ и подавляет в этих клетках экспрессию гена *WUS* (Schoof *et al.*, 2000; Rojo *et al.*, 2002; Lenhard, Laux, 2003; Ogawa *et al.*, 2008). Связывание пептида CLV3 с CLV1/2 в клетках, окружающих ОЦ, приводит к снижению его поступления в клетки ОЦ, в результате чего в норме в клетках ОЦ наблюдается экспрессия гена *WUS* (Lenhard, Laux, 2003). Белок WUS перемещается в клетки ЦЗ и активирует в них экспрессию гена *CLV3* посредством прямого контроля транскрипции (Yadav *et al.*, 2011).

Несмотря на приведенные выше данные, вопрос о механизме регуляции пространственной структуры ниши стволовых клеток в апикальной меристеме побега растения остается открытым. Механизмы, обеспечивающие постоянство структуры АМП, являются предметом интенсивных исследований, как экспериментальных, так и теоретических. Результатом большой экспериментальной работы на протяжении последних 30 лет являются данные о роли отдельных генов в развитии растения. В частности, показано, что отрицательная обратная связь между генами *CLV3* и *WUS* играет центральную роль в поддержании структуры ниши стволовых клеток АМП (Barton, 2010; Sablowski, 2011). К настоящему времени разными авторами (Jönsson *et al.*, 2003, 2005; Николаев и др., 2006, 2007, 2010, 2013; Geier *et al.*, 2008; Gordon *et al.*, 2009; Hohm *et al.*, 2010; Fujita *et al.*, 2011; Sahlin *et al.*, 2011; Yadav *et al.*, 2011, 2013; Chickarmane *et al.*, 2012; Чуб, Синюшин, 2012) предложены математические модели регуляции ниши стволовых клеток в АМП, ядром которых является генная сеть, описывающая взаимодействие между генами *CLV3* и *WUS*.

Схемы рассматриваемых в данном обзоре моделей (Hohm *et al.*, 2010; Yadav *et al.*, 2011; Chickarmane *et al.*, 2012; Николаев и др., 2013) регуляции пространственного паттерна экспрессии генов *CLV3* и *WUS* представлены на рис. 1 в единообразной форме.

Модели базируются на двух типах экспериментальных данных. Во-первых, это данные о взаимодействии генов и их продуктов, получаемые в молекулярно-генетических экспериментах. Другой тип экспериментальных данных – это пространственное расположение зон экспрессии генов в АМП. Соответственно, исходные предположения, формализуемые в моделях, основаны на представлениях о механизмах взаимодействия генов, природе регуляторных сигналов с учетом данных молекулярно-генетических экспериментов и механизме распространения этих сигналов между клетками. Выходные данные этих моделей – пространственное расположение зон экспрессии генов в АМП, которое можно сравнить с экспериментально наблюдаемым. Важным аспектом моделируемой системы является также постоянство зон, несмотря на деление клеток в АМП.

ИЗУЧЕНИЕ РОЛИ ГЕНА *WUS* В РЕГУЛЯЦИИ НИШИ СТВОЛОВЫХ КЛЕТОК В АМП (Jönsson *et al.*, 2005; Yadav *et al.*, 2011, 2013)

В статье H. Jönsson с соавт. (2005) была рассмотрена модель позиционирования экспрессии гена *WUS* на двумерном клеточном ансамбле с центральной симметрией. Благодаря симметрии оказалось естественным поместить в «морфологически выделенный» внешний слой клеток некоторый репрессор для гена *WUS*. Модель воспроизводит локализацию экспрессии *WUS* в центре клеточного ансамбля, а также результаты экспериментов, в которых наблюдалось возобновление экспрессии *WUS* после лазерного разрушения ОЦ.

Компьютерная модель (Yadav *et al.*, 2011) объясняет важность поддержания градиента

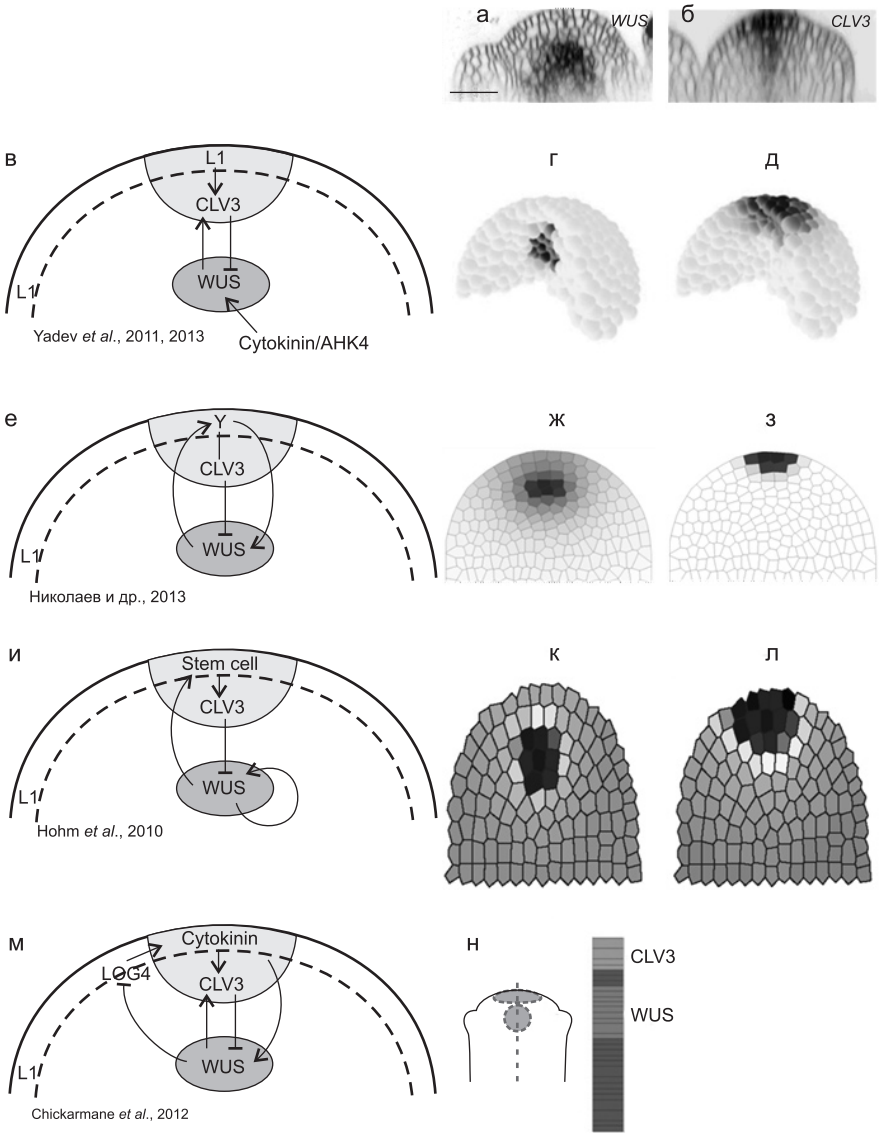


Рис. Модели регуляции структуры ниши стволовых клеток АМП.

Схемы взаимодействия (в, е, и, м – первая колонка) и решения распределения концентраций (г, д, ж, з, к, л, н – вторая (для гена *WUS*) и третья (для гена *CLV3*) колонки) для основных компонент моделей; а, б – распределения экспрессии генов *WUS* (фотография взята из работы (Yadav *et al.*, 2009)) и *CLV3* (фотография взята из работы (Reddy, Meyerowitz, 2005)); в, г, д – модель (Yadav *et al*., 2011); е, ж, з – модель (Николаев и др., 2013); и, к, л – модель (Hohm *et al*., 2010) и (м, н) модель (Chickarmane *et al*., 2012).

белка WUS в регуляции ниши стволовых клеток. Модель (схема модели – рис., в): учитывает прямую активацию транскрипции *CLV3* белком WUS (Yadav *et al.*, 2011) и отрицательную регуляцию *WUS* сигналом CLV3, который происходит из ЦЗ (Brand *et al*., 2000, 2002; Schoof *et al*., 2000). Дополнительно *WUS* активируется локальным сигналом цитокинина (Jqnsson *et al*., 2005; Gordon *et al*., 2009; Hohm *et al*., 2010), а также для активации *CLV3* используется дополнительный гипотетический сигнал, происходящий из клеток слоя L1 (Jönsson *et al*., 2003).

Геометрия АМП описывается в модели как верхняя треть трехмерной сферы, заполненная 1366 пересекающимися сферическими клетками. Клетки считаются соседними (т. е. между ними осуществляется транспорт), если соответствующие сферы пересекаются. Начальное распределение экспрессии генов *CLV3* и *WUS* строится на основе микрофотографии. Экспрес-

сия *CLV3* в клетках ЦЗ на верхушке меристемы (55 клеток) и экспрессия *WUS* в клетках ОЦ под ЦЗ (42 клетки) задаются равными единице.

Для описания регуляции транскрипции используются функции Хилла; синтез белков линейно зависит от уровня РНК. РНК и белки распадаются, распространение *CLV3* и *WUS* между клетками описывается как пассивный транспорт. Модель представлена следующей системой уравнений:

$$\frac{d[C]}{dt} = V_C \frac{[a1]^n}{k_{a1/C}^n + [a1]^n} \frac{[w]^n}{k_{w/C}^n + [w]^n} - g_C[C],$$

$$\frac{d[c]}{dt} = P_c[C] - g_c[c] + D_c\Delta[c],$$

$$\frac{d[W]}{dt} = V_W \frac{[a2]^n}{k_{a2/W}^n + [a2]^n} \frac{k_{c/W}^n}{k_{c/W}^n + [c]^n} - g_W[W],$$

$$\frac{d[w]}{dt} = P_w[W] - g_w[w] + D_w\Delta[w],$$

$$\frac{d[a1]}{dt} = P_{a1}[\mathrm{A1}] - g_{a1}[a1] + D_{a1}\Delta[a1],$$

$$\frac{d[a2]}{dt} = P_{a2}[\mathrm{A2}] - g_{a2}[a2] + D_{a2}\Delta[a2],$$

где $[X]$ и $[x]$ обозначают, соответственно, концентрации мРНК и сигнальных молекул (белка для WUS и пептида для CLV3); Δ – дискретный оператор Лапласа, описывающий пассивный транспорт (основанный на градиенте концентраций). *C* обозначает *CLV3*, *W* – *WUS*. Значения A1 и A2 являются константами, равными единице, в клетках эпидермиса (внешний слой L1) и ОЦ соответственно.

В результате оптимизации были найдены такие значения параметров, при которых решение модели (рис., г, д) соответствует экспериментально наблюдаемым зонам экспрессии генов *CLV3* и *WUS* на трехмерной ткани.

Расчеты были произведены с использованием пакета Organism (http://dev.thep.lu.se/organism), использующего метод Рунге–Кутты 5-го порядка с адаптивной сеткой.

Следующая реализация модели Yadav с соавт. (2013) учитывает непосредственную регуляцию транскрипции гена *KAN1* белком WUS. *KAN1* считается одним из генов периферической зоны, непосредственно репрессируемым транскрипционным фактором WUS (Kerstetter *et al.*, 2001; Yadav *et al.*, 2013). Дополнительно ген *KAN1* активируется неким гипотетическим сигналом из слоя L1 (так же, как *CLV3* в предыдущем варианте модели). Правомерность такой активации объясняется тем, что когда WUS не диффундирует, зона экспрессии *KAN1* расположена близко к внешнему слою (L1) меристемы.

Отметим, что в модели используются два дополнительных геометрических ограничения на зоны возможной экспрессии генов: 1) дополнительный фактор активации для *WUS*, представляющий рецепторы AHK4, расположенный в ОЦ; 2) дополнительный фактор автивации для *CLV3*, представляющий гипотетические эпидермальные сигналы, расположенный во внешнем слое клеток (L1). Оба ограничения моделируются как гены, т. е. экспрессия гипотетического гена приводит к синтезу сигнальных молекул, которые диффундируют и создают градиент концентрации.

МОДЕЛЬ РЕГУЛЯЦИИ СТРУКТУРЫ НИШИ СТВОЛОВЫХ КЛЕТОК В АМП (Николаев и др., 2007, 2010, 2013)

Приведенные выше опубликованные экспериментальные данные о роли системы *CLA3/WUS* в регуляции структуры ниши стволовых клеток АМП представлены в модели (Николаев и др., 2013) в виде следующих постулатов (схема модели – рис., е):

1. Имеется ген *Y*, экспрессия которого разрешена только в слое L1. Продукт экспрессии этого гена (белок Y) диффундирует по АМП с одновременным распадом, в результате чего устанавливается некоторое стационарное неоднородное распределение его концентрации.

2. Экспрессия генов *CLV3* и *WUS* активируется белком Y. При этом порог активации для *CLV3* выше, чем для *WUS*. В результате этого нижняя граница зоны экспрессии гена *CLV3* располагается ближе к верхушке АМП (к слою L1), чем нижняя граница зоны экспрессии *WUS*.

3. Белковый продукт гена *WUS* диффундирует от ОЦ, в том числе к верхушке АМП, где он активирует экспрессию гена *Y*. Кроме того, белок WUS активирует экспрессию генов *CLV1* и *CLV2*. Белки CLV1 и CLV2 образуют комплекс CLV1/CLV2 на поверхности клеток, в которых они синтезируются (Schoof *et al.*, 2000; Rojo *et al.*, 2002; Николаев и др., 2007; Ogawa *et al.*, 2008).

4. Пептид pCLV3 распространяется по внешним слоям АМП быстрее, чем по корпусу. Этот пептид необратимо связывается с рецептором

комплексом CLV1/CLV2 на поверхности клеток (Rojo *et al.*, 2002; Николаев и др., 2007; Ogawa *et al.*, 2008).

5. Связывание пептида pCLV3 с рецептором CLV1/CLV2 запускает путь передачи сигнала, подавляющего экспрессию *WUS* (Lenhard, Laux, 2003).

6. Комплекс CLV1/2+CLV3 поглощается клетками и деградирует. В результате этого концентрация CLV3 уменьшается и внутри АМП возникает зона ОЦ, свободная от CLV1/2+CLV3, где и наблюдается экспрессия гена *WUS* (Lenhard, Laux, 2003; Williams, Fletcher, 2005).

Модель рассматривается на продольном срезе АМП с клеточной структурой, геометрически подобной реальным изображениям продольного среза АМП (построена методом разбиения Вороного).

На такой геометрической модели продольного среза АМП была построена динамическая модель пространственно распределенного механизма регуляции со сосредоточенными параметрами в точках i, соответствующих клеткам области. Модель представлена в виде системы обыкновенных дифференциальных уравнений:

$$\frac{dy_i}{dt} = \frac{\beta_y}{V_i}\sum_{j\in\varepsilon(i)} S_{ij}(y_i - y_j) + v_y \mathrm{I}_{\mathrm{Y}}^{i}\,\mathbf{g}(h_y + \mathrm{T}_{yw}w_i) - \mathrm{d}_y y_i,$$

$$\frac{dc_i}{dt} = \frac{\beta_c}{V_i}\sum_{j\in\varepsilon(i)} S_{ij}(c_i - c_j) + v_c \mathbf{g}(h_c + \mathrm{T}_{cy}y_i) - \mathrm{d}_c c_i - \alpha c_i z_i + \beta u_i,$$

$$\frac{dw_i}{dt} = \frac{\beta_w}{V_i}\sum_{j\in\varepsilon(i)} S_{ij}(w_i - w_j) + v_w \mathbf{g}(h_w + \mathrm{T}_{wy}y_i + \mathrm{T}_{wu}u_i) - \mathrm{d}_w w_i,$$

$$\frac{dz_i}{dt} = v_z \mathbf{g}(h_z + \mathrm{T}_{zw}w_i) - \mathrm{d}_z z_i - \alpha c_i z_i + \beta u_i,$$

$$\frac{du_i}{dt} = \alpha c_i z_i - \beta u_i - \gamma u_i,$$

где y, c, w – концентрации белков Y, CLV3, WUS; z и u – концентрации гетеродимерного рецептора CVL1/2 и комплекса CLV1/2+CLV3 соответственно. β_y, β_c, β_w – коэффициенты проницаемости межклеточных границ для веществ y, c, w. V_i – объем i-й клетки (двумерной), S_{ij} – площадь границы (одномерной) между i-й и j-й клетками. v_y, v_c, v_w – максимальные скорости синтеза веществ y, c, w; d_y, d_c, d_w – коэффициенты распада веществ y, c, w; α – коэффициент скорости образования вещества u; $\mathrm{I}_{\mathrm{Y}}^{i}$ – индексная функция, равная единице для клеток, находящихся на границе клеточного ансамбля, и нулю для остальных клеток. Суммирование производится по всем клеткам j, которые являются соседними с клеткой i ($j \in \varepsilon(i)$).

Регуляция экспрессии генов y, c, w и z описывается в модели сигмоидной функцией (Mjolsness *et al.*, 1991):

$$\mathbf{g}(\mathrm{X}) = \frac{1}{2}\left(1 + \frac{X}{\sqrt{1+X^2}}\right).$$

Для вычислительных экспериментов с моделью был использован пакет Cellzilla (http://computableplant.caltech.edu/~bshapiro/Cellzilla/html/index.html).

По разработанному алгоритму были подобраны параметры и методом установления во времени получено стационарное решение (рис., ж, з) для распределения продуктов экспрессии генов CLV3 и WUS в АМП, которое качественно согласуется с экспериментально наблюдаемым (рис., а, б).

Предложенный С.В. Николаевым с соавт. (2013) механизм взаимной регуляции пространственно распределенной экспрессии генов *CLV1/2/3* и *WUS*, характерной для структуры ниши стволовых клеток в АМП, и гипотетического гена *Y* способен устойчиво поддерживать такой пространственный паттерн экспрессии. Ключевыми моментами предложенного механизма регуляции являются: а) петля положительной обратной связи между клетками верхушки АМП и ОЦ, представленная в виде взаимодействия генов *Y* и *WUS*, которая, как показано в статье С.В. Николаева с соавт. (2010), удерживает ОЦ на фиксированном расстоянии от верхушки АМП; б) активация экспрессии *CLV3* сигналом *Y*, распространяющимся сверху, что обеспечивает правильное пространственное распределение экспрессии *CLV3*; в) замкнутость модели,

т. е. для получения устойчивого к возмущениям нужного стационарного решения (пространственного паттерна экспрессии рассматриваемых генов) никакая из областей экспрессии генов не фиксируется принудительно.

ДИНАМИЧЕСКАЯ МОДЕЛЬ ГОМЕОСТАЗА НИШИ СТВОЛОВЫХ КЛЕТОК В МЕРИСТЕМЕ ПОБЕГА *ARABIDOPSIS THALIANA* (Hohm *et al.*, 2010)

В статье Hohm с соавт. (2010) представлена модель (схема модели – рис., и) регуляции структуры ниши стволовых клеток в апикальной меристеме побега, основанная на взаимодействии генов *CLV3/WUS*. Целью авторов было промоделировать механизм формирования паттернов экспрессии генов *CLV3* и *WUS* на продольном срезе АМП, используя для этого минимальное количество предположений об участниках этого процесса. Модель использовалась для изучения поведения системы при возмущениях, а также для нахождения допустимых значений параметров, при которых сохраняется гомеостаз ниши стволовых клеток в АМП, несмотря на изменяющиеся внешние условия.

Ядром схемы взаимодействия между компонентами модели являются два контура обратных связей, проходящих через *WUS*. Взаимодействия между компонентами модели осуществляются по механизму реакция–диффузия и представлены в виде системы дифференциальных уравнений:

$$\frac{\partial[WUS]}{\partial t} = D_{WUS}\Delta[WUS] + \xi\rho_{\text{anc}}\frac{[WUS]^2[facX]}{1+([CLV3]+[CLV3_{ext}])^3} - \mu_{WUS}[WUS] + \sigma_{WUS},$$

$$\frac{\partial[facX]}{\partial t} = D_{facX}\Delta[facX] + \xi\rho_{\text{anc}}\frac{[WUS]^2[facX]}{1+([CLV3]+[CLV3_{ext}])^3} + \frac{\sigma_{facX}}{1+\frac{[facX]}{K_{facX}}},$$

$$\frac{\partial[WUS_{sig}]}{\partial t} = D_{WUS_{sig}}\Delta[WUS_{sig}] + \rho_{WUS_{sig}}[WUS] - \mu_{WUS_{sig}}[WUS_{sig}],$$

$$\frac{\partial[st]}{\partial t} = D_{st}\Delta[st] + 1_{Id}(i)\rho_{\text{st}}\frac{\left(\frac{[WUS_{sig}]}{K_{st}}\right)^5}{1+\left(\frac{[WUS_{sig}]}{K_{st}}\right)^5} - \mu_{st}[st],$$

$$\frac{\partial[CLV3]}{\partial t} = D_{CLV3}\Delta[CLV3] + c_{k0}\rho_{CLV3}[st] - \mu_{CLV3}[CLV3].$$

В модели присутствуют следующие параметры:

ρ – коэффициент скорости реакции, σ – скорость экспрессии (не зависит от концентрации каких-либо веществ), μ – скорость деградации, *K* – кинетическая константа.

Переменные модели (концентрации веществ), а также основные модельные предположения относительно них приведены ниже:

[*st*] –уровень «стволовости» клеток меристемы контролируется WUS-зависимым сигналом, на который способны реагировать только клетки внешнего слоя. Пропорционально уровню «стволовости» в стволовых клетках синтезируется сигнальная молекула CLV3.

[*CLV3*] способен диффундировать с распадом в соседние клетки и ограничивать (в зависимости от своей концентрации) экспрессию *WUS*.

[*WUS*] диффундирует очень медленно и стимулирует образование WUS-сигнала, который является более мобильным и способен диффундировать в соседние клетки. Хотя потенциально все клетки (в рамках модели) способны синтезировать WUS, в модели присутствует пространственный параметр, который делает клетки, находящиеся ближе к верхушке меристемы, более «компетентными» к синтезу WUS. Введение этого параметра было обусловлено необходимостью корректного позиционирования паттернов экспрессии генов *WUS* и *CLV3*.

[WUS_{sig}] синтезируется в тех клетках, где есть WUS (в зависимости от его концентрации). Подобно CLV3, WUS-сигнал может быстро диффундировать и распадаться с постоянной скоростью. В зависимости от концентрации WUS-сигнала клетки становятся стволовыми.

Только внешние слои клеток меристемы являются компетентными к WUS-сигналу.

[*facX*] введен в модель для того, чтобы учесть CLV-независимую регуляцию экспрессии гена *WUS*. Изначально *facX* экспрессируется во всех клетках и свободно диффундирует. facX стимулирует экспрессию *WUS*, а WUS подавляет экспрессию *facX*, что реализовано в модели через активную деградацию (или потребление) facX с помощью WUS. Таким образом, взаимодействие между facX и WUS основано на механизме активатор–субстрат, что позволяет получать обособленные области экспрессии *WUS*.

Следует отметить, что для получения правдоподобной картины расположения паттернов экспрессии генов *CLV3* и *WUS* (решение модели для этих генов – рис., к, л) авторы использовали дополнительное пространственное ограничение на расположение клеток, в которых разрешена экспрессия этих генов (предписанное расположение компетентных клеток). Hohm с соавт. (2010) предположили, что *CLV3*-экспрессирующие (т. е. стволовые) клетки определяются сигналом «стволовости», распространяющимся сверху, в то время как WUS «модулирует» интенсивность этой экспрессии (и распространяется снизу от ОЦ).

ИЗУЧЕНИЕ РОЛИ ЦИТОКИНИНА В РЕГУЛЯЦИИ НИШИ СТВОЛОВЫХ КЛЕТОК В АМП (Gordon *et al.*, 2009; Chickarmane *et al.*, 2012)

Одномерная модель регуляции расположения зон экспрессии генов вдоль продольной оси АМП с делением клеток была рассмотрена в статье Chickarmane с соавт. (2012). Устойчивое позиционирование ОЦ на определенном расстоянии от верхушки меристемы обеспечивалось петлей регуляции между ОЦ и верхушкой АМП (схема модели – рис., м). Транскрипция *WUS* активируется цитокинин-регулируемым транскрипционным фактором, подробнее модель этого механизма исследована в более ранней работе (Gordon *et al.*, 2009). Диффундирующий сигнал – CLV3-сигнал – синтезируется в первой клетке массива и регулируется сигналом «стволовости» Sstemcells и диффундирующим сигналом, производным от *WUS*. CLV3-сигнал распространяется по клеточному массиву и делает клетки компетентными к экспрессии *CLV3*. Так как CLV3-сигнал синтезируется только в первой клетке массива, его градиент экспоненциально уменьшается и поэтому ген *CLV3* экспрессируется лишь в нескольких первых клетках. WUS активирует диффундирующий сигнал Sstemcells, поддерживающий стволовость клеток и активирующий *CLV3* и CLV3-сигнал. Цитокинин синтезируется в первой клетке массива. Таким образом, в начале одномерного массива (сверху) расположены клетки, в которых происходит экспрессия *CLV3*, затем – клетки, маркированные экспрессией *WUS* (решение модели – рис., н).

ОБСУЖДЕНИЕ

На протяжении более 20 лет считается, что гены *CLV3* и *WUS* являются важными участниками в механизме поддержания постоянной пространственной структуры ниши стволовых клеток АМП (Barton, 2010). В настоящем обзоре было рассмотрено несколько моделей, в которых теоретически изучается роль в этом механизме регуляторного контура *CLV3/WUS*.

В моделях Yadav с соавт. (2011, 2013) ставится вопрос о том, какой механизм обеспечивает взаимную регуляцию экспрессии генов в клетках, находящихся в разных пространственных компартментах, так, что сохраняется нужная пространственная структура АМП в целом, а именно: каким образом синтезирующийся в ЦЗ WUS, перемещаясь в соседние клетки, влияет на активацию генов *CLV3* и *KAN1*.

В моделях С.В. Николаева с соавт. (2007, 2010, 2013) сделан акцент на взаимодействии между разными группами клеток в пространстве АМП. Основанная на опосредованной активации экспрессии *CLV3* белком WUS модель поддержания пространственной локализации ЦЗ и ОЦ на продольном срезе АМП была рассмотрена в работе С.В. Николаева с соавт. (2007) при дополнительной фиксации зоны экспрессии гена *Y*, замкнутый вариант этой модели представлен в работе С.В. Николаева с соавт. (2013). Кроме того, в работе С.В. Николаева с соавт. (2010) на одномерном варианте модели было явно смоделировано деление клеток и показано, что предложенная регуляция обеспечивает

локализацию ОЦ на стабильном расстоянии от верхушки АМП при делении клеток.

В модели Hohm с соавт. (2010) для получения правдоподобной картины расположения ЦЗ и ОЦ авторы удерживали на верхушке АМП зону экспрессии некоторого гипотетического сигнала, определяющего «стволовость» клеток. Кроме того, дополнительно предписывали, в каких клетках АМП разрешена экспрессия генов *CLV3* и *WUS*.

В работе Chickarmane с соавт. (2012) предложена модель с учетом роста и деления клеток. Главным вопросом было позиционирование зоны экспрессии гена *WUS* относительно верхушки меристемы, поэтому модель рассчитывали на одномерном массиве клеток, расположенных вдоль центральной оси меристемы. В начале массива (сверху) расположены клетки, в которых происходит экспрессия *CLV3*, затем – клетки, маркированные экспрессией *WUS*.

Было показано (Yadav *et al.*, 2011), что белковый продукт гена *WUS* перемещается из зоны синтеза (ОЦ) в ЦЗ меристемы, где проникает в ядро и непосредственно взаимодействует с последовательностями ДНК в промоторе гена *CLV3*. Однако, согласно сложившимся представлениям о распространении сигнальных молекул в АМП, эти данные порождают много вопросов. Например, неясно, почему экспрессия *CLV3* не наблюдается непосредственно в зоне синтеза WUS, где его концентрация должна быть максимальной? Почему экспрессия гена *CLV3* имеет максимум в поверхностных клетках меристемы и падает по направлению к глубинным слоям, хотя при активации экспрессии сигналом из ОЦ ожидается обратная картина? Чтобы согласовать прямую активацию экспрессии *CLV3* продуктом гена *WUS*, авторы модели (Yadav *et al.*, 2011) постулировали, что имеется некий гипотетический сигнал, распространяющийся из верхушки АМП (из слоя L1), и дополнительно фиксировали расположение клеток ОЦ.

Таким образом, в результате моделирования выявляется некоторого рода конфликт между постулатом о непосредственной активации гена *CLV3* продуктом экспрессии гена *WUS*, основанным на экспериментальных данных, и такими постулатами моделей, как пассивное изотропное распространение молекул-сигналов и обобщенное пространство АМП. Как показывают вычислительные эксперименты, для получения правильного расположения зоны экспрессии гена *CLV3* в моделях с непосредственной активацией гена *CLV3* продуктом экспрессии гена *WUS* (Yadav *et al.*, 2011, 2013; Chickarmane *et al.*, 2012) необходимо постулировать дополнительный активирующий сигнал, распространяющийся сверху (из слоя L1), и эта активация должна превалировать над активацией со стороны ОЦ. Кроме того, вычислительные эксперименты (Николаев и др., 2010; Chickarmane *et al.*, 2012) показывают, что стабильное расположение ОЦ при делениях клеток может обеспечиваться регуляторным контуром между ОЦ и верхушкой АМП, и в предложенный в работе С.В. Николаева с соавт. (2010) положительный регуляторный контур «ОЦ–верхушка АМП» органично вписывается опосредованная активация WUS-Y-CLV3. Эти обстоятельства позволяют предполагать, что именно этот регуляторный контур является центральным звеном в системе регуляции пространственного паттерна экспрессии генов, характерного для ниши стволовых клеток в АМП, в то время как непосредственная активация экспрессии *CLV3* продуктом гена *WUS* является модулирующим сигналом.

ЗАКЛЮЧЕНИЕ

Хотя компьютерные модели пока еще не стали основным инструментом для изучения развития растений, ряд работ, рассмотренных в настоящем обзоре, являются хорошим примером, показывающим важность их использования наряду с экспериментальными методами для изучения регуляции морфогенеза.

Несмотря на кажущуюся проработанность моделей, многие детали до сих пор остаются невыясненными. Например, неизвестными остаются процессы, посредством которых осуществляются регуляторные взаимодействия генов при формировании пространственных паттернов генной активности в АМП. В рассмотренных моделях в качестве такого механизма передачи сигналов между клетками, разделенными в пространстве, рассматривается диффузия молекул-регуляторов.

Работа выполнена при частичной финансовой поддержке гранта РФФИ № 11-04-01748-а.

ЛИТЕРАТУРА

Николаев С.В., Зубаирова У.С., Пененко А.В. и др. Модель регуляции структуры ниши стволовых клеток в апикальной меристеме побега *Arabidopsis thaliana* // Докл. АН. 2013. Т. 451. № 5. С. 336–338.

Николаев С.В., Зубаирова У.С., Фадеев С.И. и др. Исследование одномерной модели регуляции размеров возобновительной зоны в биологической ткани с учетом деления клеток // Сиб. журн. индустр. математики. 2010. Т. 13. Вып. 4(44). С. 70–82.

Николаев С.В., Колчанов Н.А., Фадеев С.И. и др. Исследование одномерной модели регуляции размеров возобновительной зоны в биологической ткани // Вычисл. технол. 2006. Т. 11. Вып. 2. С. 67–81.

Николаев С.В., Пененко А.В., Лавреха В.В. и др. Модельное изучение роли белков CLV1, CLV2, CLV3 и WUS в регуляции структуры апикальной меристемы побега // Онтогенез. 2007. Т. 38. Вып. 6. С. 457–462.

Чуб В.В., Синюшин А.А. Фасциация цветка и побега: от феноменологии к построению моделей преобразования апикальной меристемы // Физиол. растений. 2012. Т. 59. Вып. 4. С. 1–17.

Barton M.K. Twenty years on: the inner workings of the shoot apical meristem, a developmental dynamo // Dev. Biol. 2010. V. 341. P. 95–113.

Brand U., Fletcher J.C., Hobe M. *et al.* Dependence of stem cell fate in Arabidopsis on a feedback loop regulated by CLV3 activity // Science. 2000. V. 289. P. 617–619.

Brand U., Grunewald M., Hobe M., Simon R. Regulation of CLV3 expression by two homeobox genes in Arabidopsis // Plant Physiol. 2002. V. 129. P. 565–575.

Brand U., Hobe M., Simon R. Functional domains in plant shoot meristems // Bioessays. 2001. V. 23. P. 134–141.

Bowman J.L., Eshed Y. Formation and maintenance of the shoot apical meristem // Trends Plant Sci. 2000. V. 5. P. 110–115.

Chickarmane V.S., Gordon S.P., Tarr P.T. *et al.* Cytokinin signaling as a positional cue for patterning the apical–basal axis of the growing Arabidopsis shoot meristem // Proc. Natl Acad. Sci. USA. 2012. V. 109. No. 10. P. 4002–4007.

Fujita H., Toyokura K., Okada K., Kawaguchi M. Reaction-diffusion pattern in shoot apical meristem of plants // PLoS ONE. 2011. V. 6. e18243.

Geier F., Lohmann J.U., Gerstung M. *et al.* A quantitative and dynamic model for plant stem cell regulation // PLoS ONE. 2008. V. 3. e3553.

Gordon S.P., Chickarmane V.S., Ohno C., Meyerowitz E.M. Multiple feedback loops through cytokinin signaling control stem cell number within the Arabidopsis shoot meristem // Proc. Natl Acad. Sci. USA. 2009. V. 106. P. 16529–16534.

Gross-Hardt R., Laux T. Stem cell regulation in the shoot meristem // J. Cell Sci. 2003. V. 116. P. 1659–1666.

Hohm T., Zitzler E., Simon R. A dynamic model for stem cell homeostasis and patterning in Arabidopsis meristems // PLoS ONE. 2010. V. 5. e9189.

Jönsson H., Gruel J., Krupinski P., Troein C. On evaluating models in computational morphodynamics // Curr. Opin. Plant Biol. 2012. V. 15. P. 103–110.

Jönsson H., Heisler M., Reddy G.V. *et al.* Modeling the organization of the WUSCHEL expression domain in the shoot apical meristem // Bioinformatics. 2005. V. 21 (Suppl. 1) P. i232–i240.

Jönsson H., Shapiro B., Meyerowitz E., Mjolsness E. Signaling in multicellular models of plant development // On Growth Form and Computers / Ed. S. Kumar, P. Bentley. L.: Acad. Press, 2003. P. 156–161.

Kerstetter R.A., Bollman K., Taylor R.A. *et al.* KANADI regulates organ polarity in Arabidopsis // Nature. 2001. V. 411. P. 706–709.

Kwiatkowska D. Structural integration at the shoot apical meristem: models, measurements, and experiments // Am. J. Bot. 2004. V. 91. P. 1277–1293.

Lenhard M., Laux T. Stem cell homeostasis in the Arabidopsis shoot meristem is regulated by intercellular movement of CLAVATA3 and its sequestration by CLAVATA1 // Development. 2003. V. 130. P. 3163–3173.

Mjolsness E., Sharp D.H., Reinitz J. A connectionist model of development // J. Theor. Biol. 1991. V. 152. P. 429–453.

Newman I.V. Pattern in the meristems of vascular plants. III. Pursuing the patterns in the apical meristem where no cell is a permanent cell // J. Linn. Soc. (Botany). 1965. V. 59. P. 185–214.

Ogawa M., Shinohara H., Sakagami Y., Matsubayashi Y. Arabidopsis CLV3 peptide directly binds CLV1 ectodomain // Science. 2008. V. 319. P. 294.

Reddy V.G., Meyerowitz E.M. Stem-cell homeostasis and growth dynamics can be uncoupled in the Arabidopsis shoot apex // Science. 2005. V. 310. No. 5748. P. 663–667.

Rojo E., Sharma V.K., Kovaleva V. *et al.* CLV3 is localized to the extracellular space, where it activates the Arabidopsis CLAVATA stem cell signaling pathway // Plant Cell. 2002. V. 14. P. 969–977.

Sablowski R. Plant stem cell niches: from signalling to execution // Curr. Opin. Plant Biol. 2011. V. 14. P. 4–9.

Sahlin P., Melke P., Jцnsson H. Models of sequestration and receptor cross-talk for explaining multiple mutants in plant stem cell regulation // BMC Syst. Biol. 2011. V. 5. 2.

Schoof H., Lenhard M., Haecker A. *et al.* The stem cell population of Arabidopsis shoot meristems in maintained by a regulatory loop between the CLAVATA and WUSCHEL genes // Cell. 2000. V. 100 P. 635–644.

Sharma V.K., Carles C., Fletcher J.C. Maintenance of stem cell populations in plants // Proc. Natl Acad. Sci. USA. 2003. V. 100 (Suppl 1) P. 11823–11829.

Traas J., Doonan J.H. Cellular basis of shoot apical meristem development // Int. Rev. Cytol. 2001. V. 208. P. 161–206.

Williams L., Fletcher J.C. Stem cell regulation in the Arabidopsis shoot apical meristem // Curr. Opin. Plant Biol. 2005. V. 8. P. 582–586.

Yadav R.K., Girke T., Pasala S. *et al.* Gene expression map of the Arabidopsis shoot apical meristem stem cell niche // Proc. Natl Acad. Sci. USA. 2009. V. 106. P. 4941–4946.

Yadav R.K., Perales M., Gruel J. *et al.* WUSCHEL protein movement mediates stem cell homeostasis in the Arabidopsis shoot apex // Genes Dev. 2011. V. 25. P. 2025–2030.

Yadav R.K., Perales M., Gruel J. *et al.* Plant stem cell maintenance involves direct transcriptional repression of differentiation program // Mol. Syst. Biol. 2013. V. 9. P. 654.

MODELS OF STEM CELL NICHE STRUCTURE REGULATION IN SHOOT APICAL MERISTEM

U.S. Zubairova, S.V. Nikolaev

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: ulyanochka@bionet.nsc.ru

Summary

The experimental data obtained to date provided grounds for certain concepts of the stem cell niche regulation in shoot apical meristem. Mathematical modeling is used for checking their consistency and coherence with experimental evidence. In this paper, we summarize mathematical models of stem cell niche regulation offered by different authors, analyze the experimental base and working hypotheses formalized in these models, and identify methodological differences in the approaches to the construction of these models.

Key words: shoot apical meristem, stem cell niche, CLAVATA3, WUSCHEL, mathematical model.

УДК 579.8.06: 546.161

ИССЛЕДОВАНИЕ ВОСПРОИЗВОДИМОСТИ РЕЗУЛЬТАТОВ ИДЕНТИФИКАЦИИ МИКРООРГАНИЗМОВ С ПОМОЩЬЮ МЕТОДА МАЛДИ ВРЕМЯПРОЛЕТНОЙ МАСС-СПЕКТРОМЕТРИИ В ЗАВИСИМОСТИ ОТ УСЛОВИЙ КУЛЬТИВИРОВАНИЯ НА ПРИМЕРЕ *GEOBACILLUS STEAROTHERMOPHILUS*

© 2013 г. **К.В. Старостин, Е.А. Демидов, А.С. Розанов, А.В. Брянская, С.Е. Пельтек**

Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия,
e-mail: starostin@bionet.nsc.ru

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

Потребность в быстрых и точных методах идентификации микроорганизмов является насущной проблемой для самых различных сфер человеческой деятельности. Прежде всего это, конечно, касается клинической диагностики, но данные методы востребованы также в экологическом мониторинге, фармацевтической и пищевой промышленности, научных исследованиях и т. д. На сегодняшний день существует множество методов идентификации микроорганизмов (фенотипические, генотипические, хемотаксономические методы, прямое белковое профилирование и др.). В данной работе авторы оценивают влияние различных условий культивирования, таких как температура, время роста и тип питательной среды, на воспроизводимость результатов идентификации микроорганизмов методом МАЛДИ времяпролетной масс-спектрометрии на примере трех штаммов *Geobacillus stearothermophilus*.

Ключевые слова: МАЛДИ, прямое белковое профилирование, идентификация микроорганизмов, *Geobacillus stearothermophilus*.

ВВЕДЕНИЕ

В настоящее время идентификация микроорганизмов методом МАЛДИ (матрично ассоциированная лазерная дисорбция/ионизация) времяпролетной масс-спектрометрии или прямое белковое профилирование является достойным конкурентом большинству существующих методов идентификации микроорганизмов, так как обеспечивает высокую скорость и низкую трудозатратность анализов, не уступая при этом в точности идентификации (Dickinson *et al*., 2004; Mellmann *et al*., 2008; Böhme *et al*., 2013). На данный момент метод хорошо себя зарекомендовал и широко применяется в клинической диагностике. Существующие коммерчески доступные базы белковых профилей основных патогенных микроорганизмов содержат несколько тысяч белковых профилей, что позволяет идентифицировать большинство патогенов человека. Однако в отличие от клинической диагностики, в остальных областях идентификация микроорганизмов методом МАЛДИ времяпролетной масс-спектрометрии не нашла столь широкого распространения. Прежде всего, это связано с отсутствием единого протокола для выращивания конкретных микроорганизмов.

Метод идентификации с помощью МАЛДИ времяпролетной масс-спектрометрии основан на получении спектра белкового профиля клеток микроорганизмов и сравнении его с эталонным спектром в базе данных. Спектр обычно

содержит значения масс белков в диапазоне от 2 до 20 кДа. В работе Ryzhov, Fenselau (2001) была проведена идентификация белков *E. coli*, присутствующих в полученных спектрах. Более половины белков в этом диапазоне оказались рибосомальными белками, остальная часть белков относилась к ДНК-связывающим белкам и белкам холодового шока. Рибосомальные белки очень консервативны, что обеспечивает их таксономическую специфичность, а их набор не меняется в зависимости от условий и фазы роста. Так как в спектрах присутствуют белки, трансляция которых может зависеть от внешних условий, возникает вопрос: насколько данные белки могут повлиять на воспроизводимость спектров и точность идентификации на видовом и субвидовом уровнях при варьировании условий роста микробиологической культуры. В нескольких работах было показано, что при изменении условий роста, таких как среда и время роста, меняется состав белковых профилей, при этом указывается, что полученная изменчивость не оказывает серьезного влияния на результаты идентификации (Ruelle *et al*., 2004; Valentine *et al*., 2005). В приведенных работах так же, как и в большинстве других публикуемых исследований, объектом идентификации служат патогены и другие клинически значимые виды и штаммы микроорганизмов, выращенные в оптимальных условиях. Вместе с тем большой интерес представляет использование данного метода для определения микроорганизмов в природных сообществах, что может быть востребовано при экологических исследованиях и поиске биотехнологически значимых штаммов, среди которых особый интерес представляют экстремофильные микроорганизмы. При исследовании неизвестных организмов, как это происходит в случае природных изолятов, невозможно заранее подобрать оптимальные условия для роста, поэтому для эффективного применения метода необходима уверенность в том, что отклонения от оптимума не окажут значительного влияния на результаты идентификации. В данной работе мы решили изучить вопрос влияния условий культивации на точность идентификации на примере трех штаммов *Geobacillus stearothermophilus*. Выбор данного вида обусловлен его широкой представленностью среди различных термальных экосистем и его перспективностью для использования в биотехнологических схемах и способностью представителей рода *Geobacillus* ферментировать гексозы и пентозы, синтезируя этанол и молочную кислоту (Cripps *et al*., 2009). В работе использовались имеющиеся в нашей коллекции штаммы G1w1, 18(x) и 20. Штаммы 18(x) и 20 выделены из проб, полученных из термальных источников Баргузинской долины в районе о. Байкал. Штамм G1w1 выделен из проб, полученных из термальных источников полуострова Камчатка.

МАТЕРИАЛЫ И МЕТОДЫ

Культивирование штаммов

Исследуемые штаммы выращивали на агаризованной среде Luria Bertani Medium (LB), обедненной среде LB (LB/5) и мясо-пептонном агаре (МПА) в течение 6–72 ч при температурах 60–70 °С.

Масс-спектрометрический анализ

Подготовку образцов и масс-спектрометрический анализ проводили согласно стандартной методике, разработанной фирмой «Bruker Daltonics» для идентификации микроорганизмов посредством программного обеспечения Biotyper (Freiwald, Sauer, 2009).

Подготовка образцов к масс-спектрометрическому анализу

Образец исследуемой культуры массой 10–15 мг ресуспендировали в 300 мкл деионизированной воды и инактивировали добавлением 900 мкл этанола. Полученную смесь тщательно перемешивали и центрифугировали 2 мин 16000 g. После удаления супернатанта осадок сушили 5 мин на вакуумном концентраторе. Высушенный осадок ресуспендировали в 50 мкл 70 % муравьиной кислоты для разрушения клеточных стенок, после чего добавляли 50 мкл ацетонитрила для экстракции белковой фракции. Полученную смесь тщательно перемешивали и центрифугировали 2 мин 16000 g. Для проведения масс-спектрометрического анализа использовали аликвоту из верхнего слоя супернатанта.

Получение масс-спектров белковых экстрактов

Аликвоту подготовленного белкового экстракта объемом 0,7 мкл наносили на стальную масс-спектрометрическую мишень и высушивали на воздухе. После высыхания на образец наслаивали 0,7 мкл раствора матрицы HCCA (6 мг/мл раствор α-циано-4-гидроксикоричной кислоты в 50 %-м ацетонитриле и 2 %-й трифторуксусной кислоте) и снова высушивали на воздухе.

В работе использовался масс-спектрометр Ultraflex III фирмы «Bruker Daltonics». Спектры снимали в линейном позитивном режиме с частотой лазера 100 Гц в диапазоне масс 2000–20000 Да. Напряжение на ускоряющих электродах – 25 кВ и 23,45 кВ, напряжение на линзе – 6 кВ без задержки экстракции.

Для получения каждого спектра суммировали данные от 500 лазерных импульсов. Внешнюю калибровку проводили с использованием точных значений масс известных белков *Escherichia coli*: RL36 4365,3 Да, RS22 5096,8 Да, RL34 5381,4 Да, RL32 6315,0 Да, RL29 7274,5 Да, RS19 10300,1 Да.

Идентификация образцов

Спектры белковых профилей импортировали в программу Biotyper и идентифицировали по стандартным настройкам «Biotyper MSP Identification Standart method». Результатом идентификации является присвоение исследуемому спектру таксономического идентификатора род/вид и численного рейтинга точности идентификации, представленного в виде логарифмической шкалы от 0 до 3. Значение 3 соответствует абсолютному совпадению, значения от 2,300 до 3,000 – достоверное определение до вида; от 2,000 до 2,299 – достоверное определение до рода и надежное определение до вида, от 1,700 до 1,999 – надежное определение до рода.

Создание эталонных суперспектров

Для создания характеристичных (эталонных) суперспектров отбирали по 12 проб для каждой бактериальной культуры. Из каждой пробы получали белковый экстракт, согласно описанной выше методике, и снимали по 3 масс-спектра для каждой пробы. Полученные серии по 36 масс-спектров импортировали в программу «Biotyper» и использовали для создания характеристичных суперспектров с помощью стандартных настроек программы «BioTyper MSP Creation Standart Method». Полученные суперспектры сохраняли в базе данных программы «Biotyper».

РЕЗУЛЬТАТЫ

Geobacillus ssp. – термофильные аэробные или факультативно анаэробные микроорганизмы с диапазоном роста в районе 40–70 °C (Nazina *et al*., 2001). Используемые в работе штаммы растут в диапазоне температур 60–70 °C с оптимумом при температуре 65 °C. Оптимальной средой для роста является LB, и в течение ночи штаммы обычно образуют плотный газон на поверхности агаризованной среды. Данные условия культивации были приняты за стандартные.

На первом этапе проводилось исследование спектров белковых экстрактов в зависимости от времени культивации. Культуры исследуемых штаммов переходили в экспоненциальную фазу роста после 6–9 ч инкубации, в связи с этим время инкубации 9 ч было выбрано как первая точка анализа, далее спектры снимали через 24, 48 и 72 ч от начала инкубации. На рис. 1 приведены спектры для штамма G1w1. Видно, что с ростом времени инкубации ухудшается качество спектров – интенсивность пиков и их разрешенность. Например, при времени инкубации 9 ч хорошо видны пики с массами 8836, 9400, 10134 и 10595 Да. После 24 ч их относительная интенсивность снижается. После двух суток инкубации данные пики еще можно определить визуально, но из-за отношения сигнал/шум автоматический алгоритм программы flexAnalysis не определяет их. После 72 ч инкубации данные пики пропадали из спектра.

При времени инкубации 72 ч удалось получить спектры только для штамма G1w1. Для штаммов 18(x) и 20 значительное снижение качества спектров наблюдалось уже при времени инкубации 48 ч. На рис. 2 представлены спектры для штамма 20. При длительном времени инкубации из спектра пропадают пики с

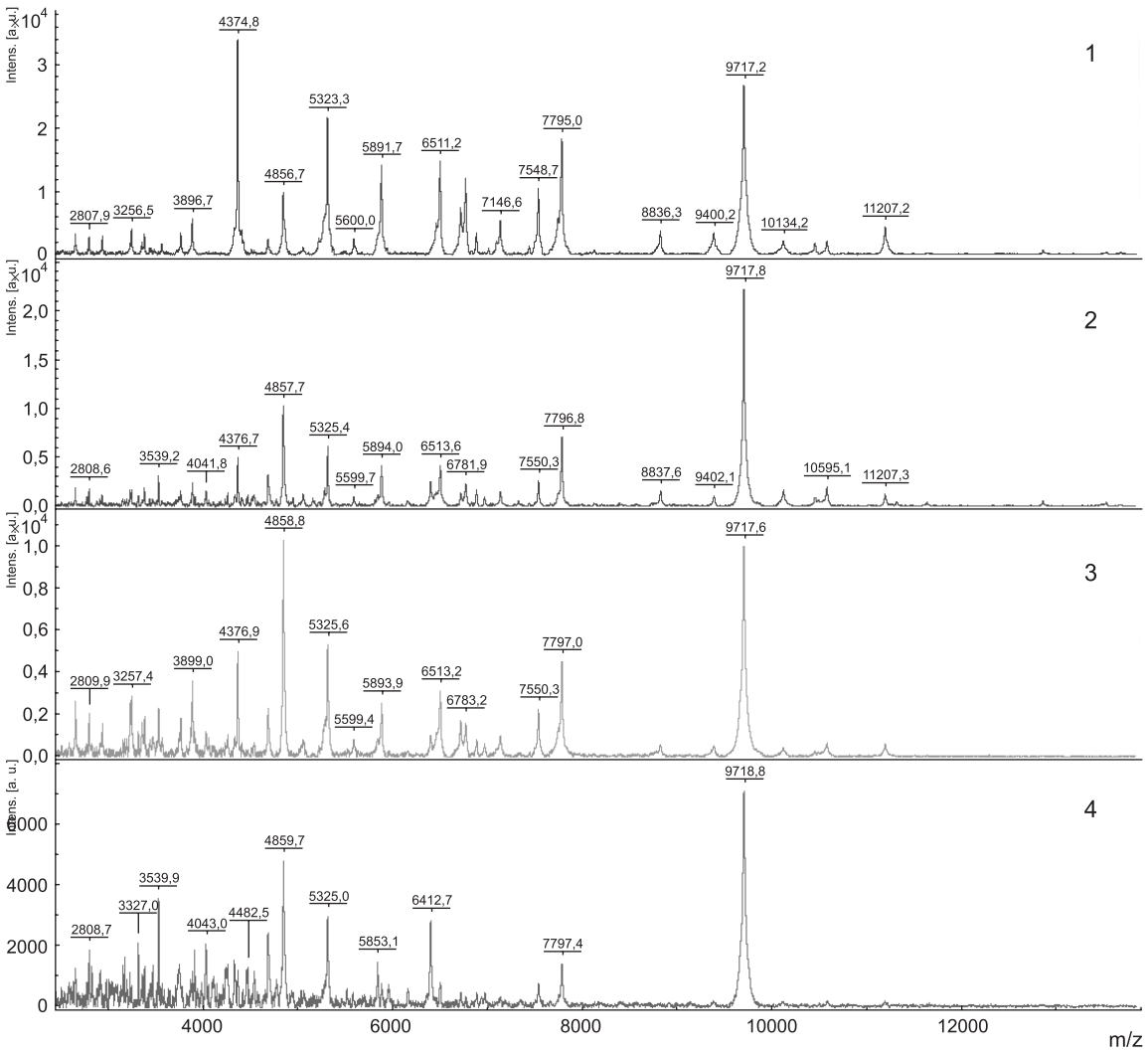


Рис. 1. Спектры штамма G1w1 на среде LB при температуре роста 65 °С и времени культивации: 9 ч (1), 24 ч (2), 48 ч (3), 72 ч (4).

массами 4362, 7550, 8837, 9402, 10561, 11223 Да (рис. 2). В спектрах с инкубацией 9 и 24 ч также наблюдаются отличия как в представленности пиков, так и в их относительной интенсивности. Например, пик с массой 6691 Да представлен как ярко выраженный мажорный пик в спектре с инкубацией 24 ч, при этом он отсутствует в спектре с 48-часовой инкубацией, а при 9-часовой инкубации представлен в виде минорного пика, слабо разрешенного относительно соседних пиков.

Ухудшение качества спектров, очевидно, связано с выходом роста культур на стационарную фазу с последующей гибелью клеток и спорообразованием. Это подтверждается данными, полученными при анализе масс-спектров культур, выращенных на средах с разным содержанием питательных веществ. Для проведения исследования использовались три среды: стандартная среда LB, обедненная среда LB и МПА – наиболее богатая питательными веществами. При времени инкубации 9 ч спектры для LB и МПА практически идентичны и по представленности пиков, и по их относительной интенсивности (рис. 3). Спектры для бедной среды отличаются от спектров для LB и МПА более низкими показателями сигнал/шум и разрешенность пиков, представленностью и относительной интенсивностью пиков.

При времени инкубации 24 ч (рис. 4) для штамма 18(х) на всех трех средах наблюдаются отличия в относительных интенсивностях

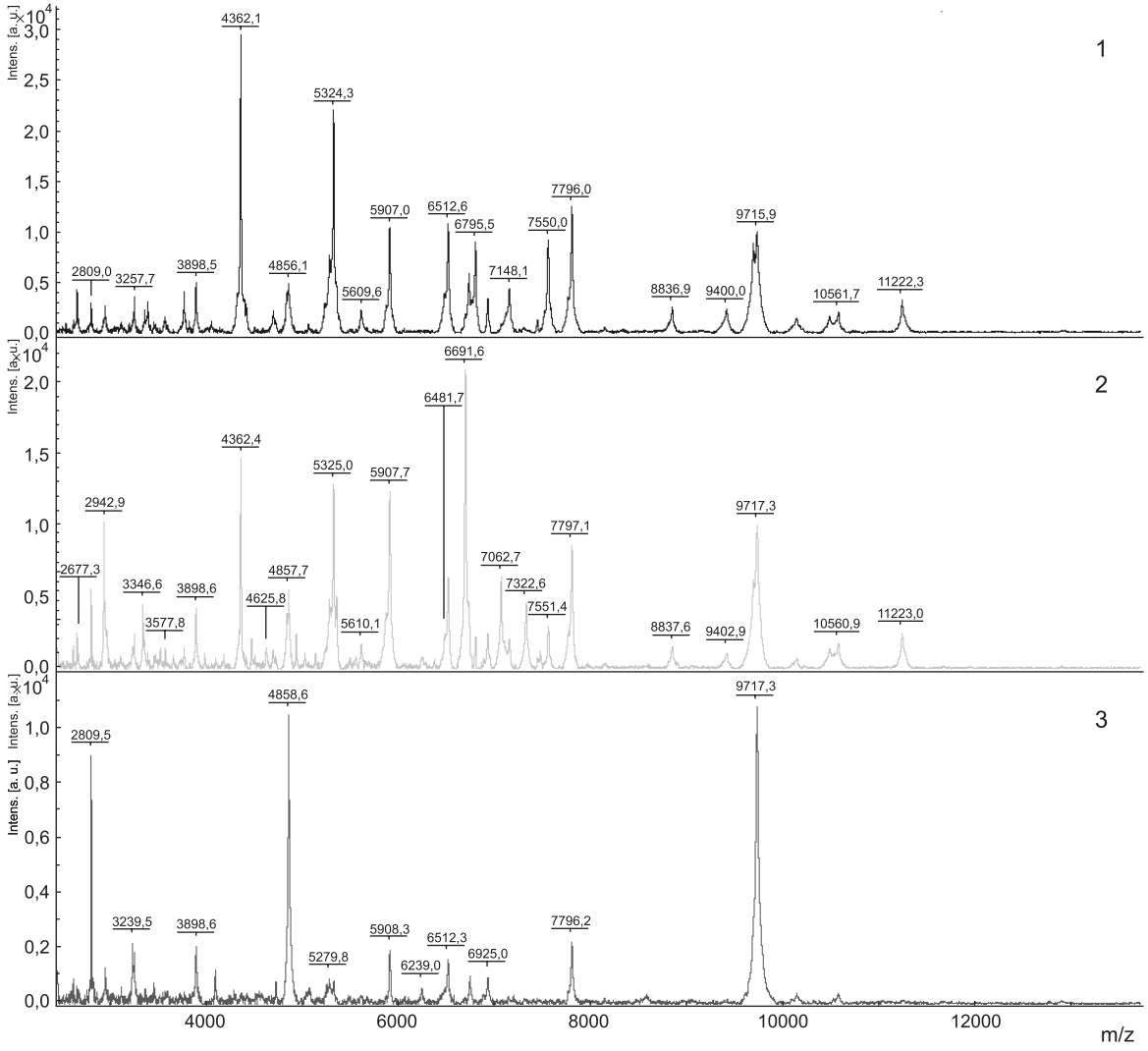


Рис. 2. Спектры штамма 20 на среде LB при температуре роста 65 °С и времени культивации: 9 ч (1), 24 ч (2), 48 ч (3).

пиков, при этом наиболее близки друг другу по представленности пиков стандартная и обедненная среда LB. Спектр для среды МПА имеет очень низкое качество и заметные отличия в представленности пиков: в его спектре отсутствуют мажорные пики с массами 6691, 7061 и 7321, в то время как при времени инкубации 9 ч данные пики присутствуют только в спектре для обедненной среды LB (рис. 3).

Сопоставимые данные были получены для штаммов 20 и G1w1. При времени инкубации 9 ч, когда фаза роста близка к экспоненциальной, спектры обладают наилучшим качеством и максимальным сходством в представленности пиков и их относительной интенсивности. С выходом на стационарную фазу, когда начинаются процессы гибели клеток и спорообразования, происходят ухудшение качества спектров и снижение количества пиков.

Для изучения влияния температуры культивирования на масс-спектрометрические характеристики белковых экстрактов исследуемые штаммы выращивали на среде LB в течение ночи при температурах инкубации 60, 65 и 70 °С. Как видно на примере штамма G1w1 (рис. 5), температура не оказывает заметного влияния на представленность пиков и их относительную интенсивность.

Для определения точности идентификации при разных условиях роста для всех трех штаммов были получены характеристичные суперспектры при стандартных условиях куль-

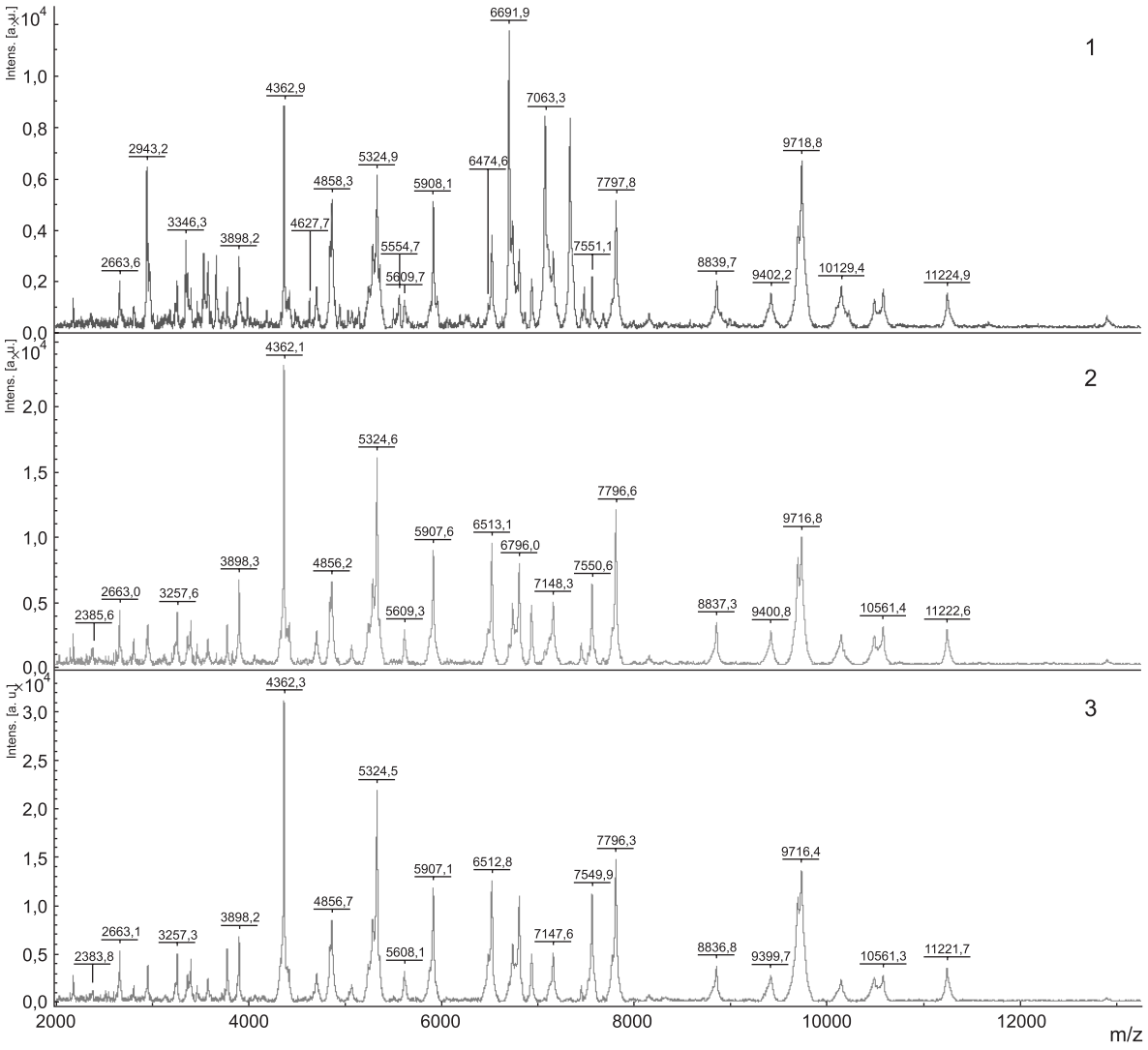


Рис. 3. Спектры штамма 18(x) при температуре культивации 65 °C и времени инкубации 9 ч на средах:

1 – обедненная среда LB/5; 2 – среда LB; 3 – среда МПА.

тивирования (среда LB, инкубация в течение ночи при температуре 65 °C). Полученные суперспектры (G1w1_65_LB, 18(x)_65_LB, 20_65_LB) внесли в базу данных, состоящую из 165 штаммов, относящихся к 24 родам и 41 виду, в том числе 20 представителей рода *Geobacillus*. С использованием имеющейся базы данных была проведена идентификация исследуемых штаммов по спектрам, полученным при вариации различных условий роста. Данные приведены в таблице.

Как видно из полученных данных, большинство образцов были идентифицированы как *Geobacillus stearothermophilus* с рейтингом идентификации выше 2, один штамм был идентифицирован с рейтингом 1,95, что дает точность идентификации только до рода, и у трех штаммов рейтинг был ниже 1,7, что не позволят их надежно идентифицировать. Штаммы 18(x) и 20 идентифицировались как *G. stearothermophilus*, но ни в одном случае не были идентифицированы по созданным для них в стандартных условиях суперспектрам – 18(x)_65_LB и 20_65_LB. Это можно объяснить высокой степенью филопротеомного родства данных штаммов со штаммами, внесенными в нашу базу данных, что не позволяет методу различить их. Штамм G1w1 в большинстве случаев был идентифицирован как G1w1_65_LB, что говорит о его заметной филопротеомной

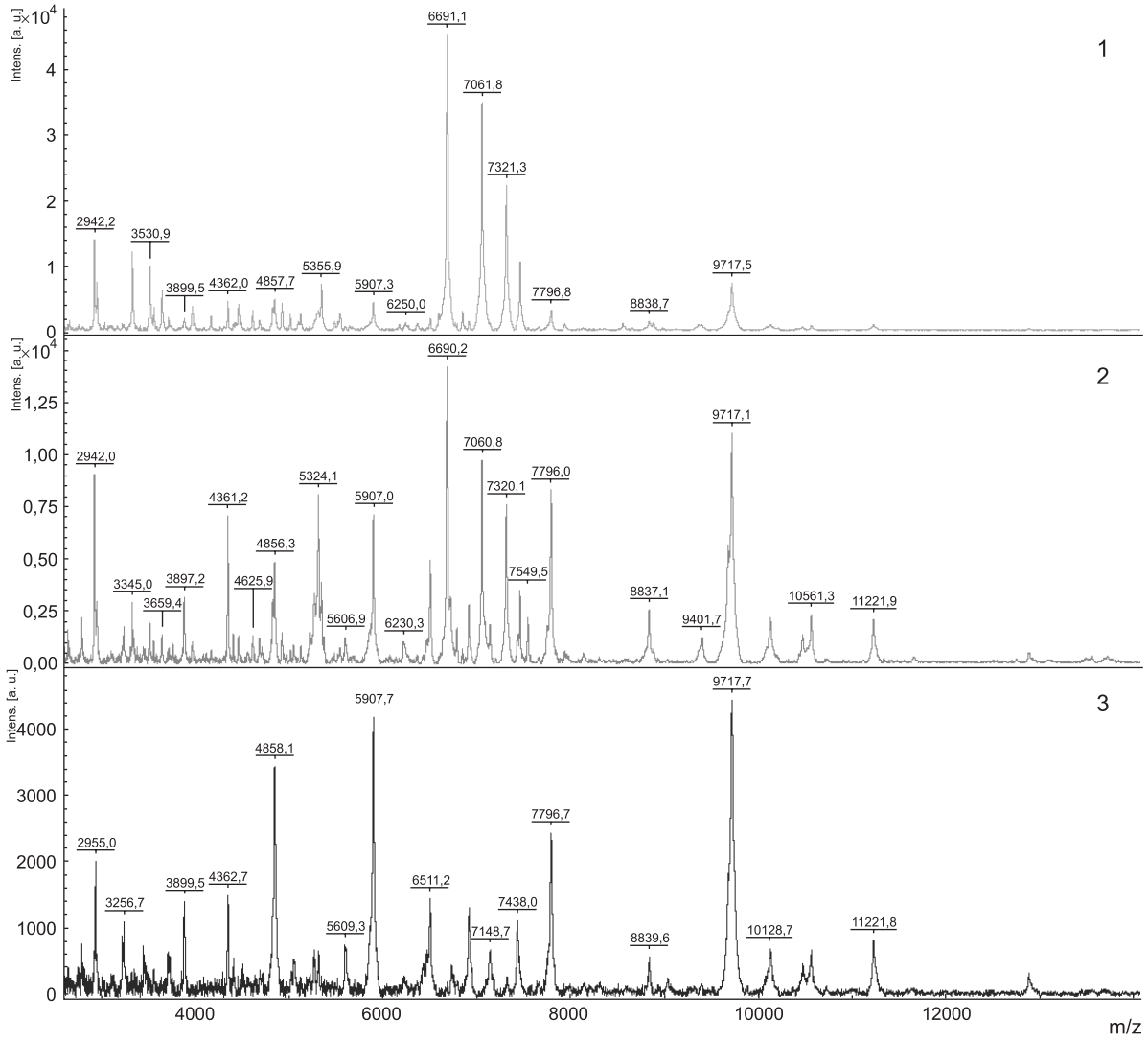


Рис. 4. Спектры штамма 18(х) при температуре культивации 65 °С, времени инкубации 24 ч на средах:

1 – обедненная среда LB/5; 2 – среда LB; 3 – среда МПА.

дистанции от остальных штаммов и принципиальной возможности метода идентифицировать микроорганизмы на субвидовом уровне.

При анализе зависимости точности идентификации от условий роста можно сделать вывод, что с течением времени культивации точность идентификации падает, что связано с уменьшением количества биологического материала в процессе спорообразования и гибели клеток при переходе в стационарную фазу и ухудшением в связи с этим качества масс-спектров (рис. 1, 2). При анализе культур, отобранных на экспоненциальной фазе роста (6–24 ч), наблюдается высокая точность идентификации с рейтингом, превышающим 2,3. Питательность среды не влияет на точность идентификации в рамках экспоненциального роста, а влияет лишь на скорость роста культуры и время выхода в стационарную фазу. Для штаммов, выращенных на среде МПА, заметное снижение рейтинга точности идентификации начинается уже после 24 ч. При времени культивации 48 ч для штамма 18(х) спектры не удалось получить; для штаммов 20 и G1w1 качество полученных спектров было неудовлетворительным, что и проявилось в соответствующих рейтингах идентификации: 1,129 и 1,438. При этом для штамма G1w1 на среде LB были получены спектры в диапазоне времени культивации 6–72 ч (рис. 1) с рейтин-

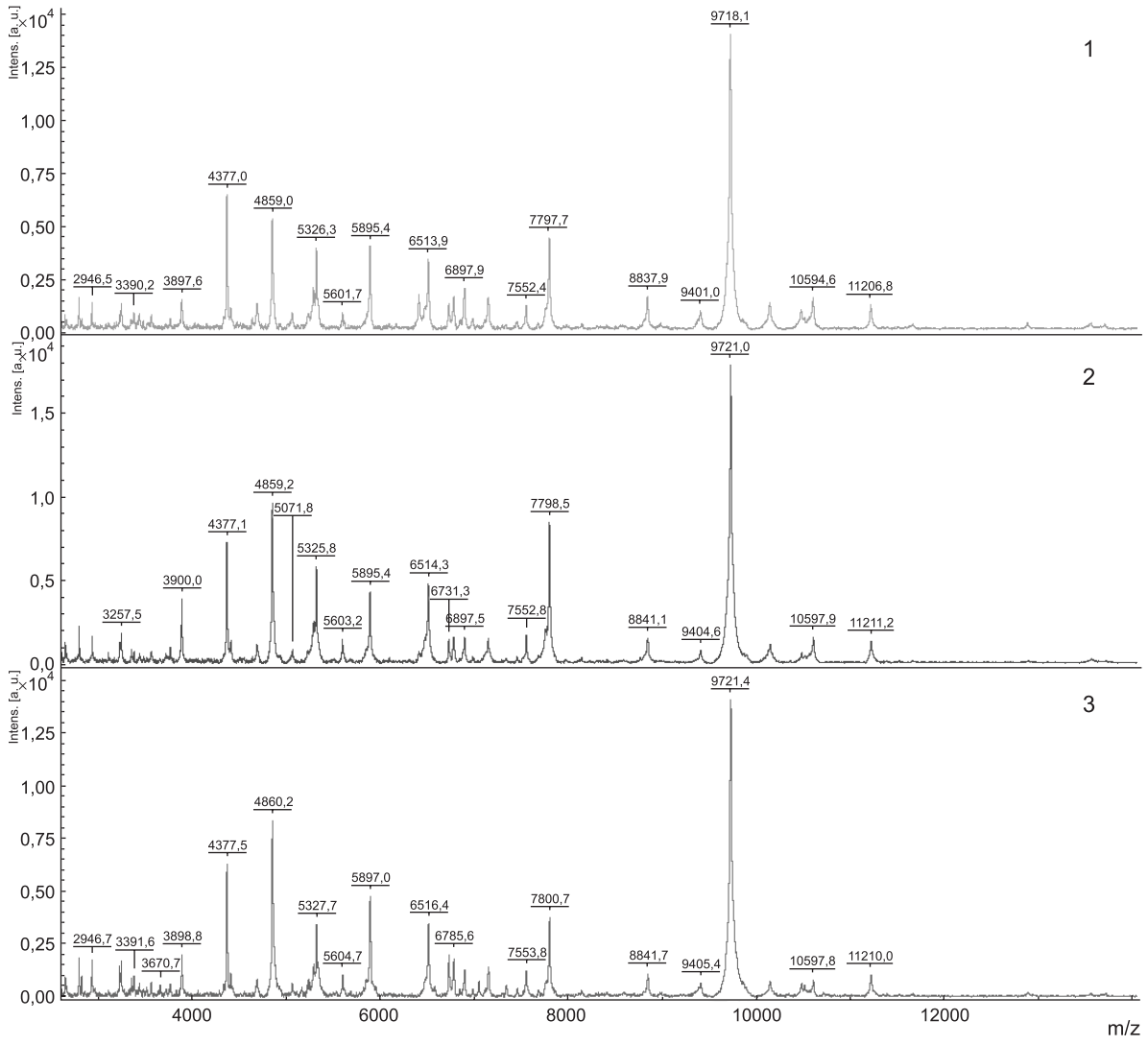


Рис. 5. Спектры штамма G1w1, выращенного на среде LB при температурах культивации: 60 °C (1), 65 °C (2), 70 °C (3).

гами идентификации 2,521–2,063. Температура культивации в анализируемом нами диапазоне 60–70 °C не влияла на качество спектров (рис. 5) и результаты идентификации (табл.). При температурах 55 и 75 °C данные штаммы не росли.

ЗАКЛЮЧЕНИЕ

Основным фактором, влияющим на точность идентификации штаммов, является время инкубации. Для получения наилучшего результата необходимо отбирать пробы культуры во время экспоненциального роста, в противном случае начинаются гибель биологического материала и снижение качества масс-спектров, что снижает точность идентификации. При удовлетворении этого условия такие факторы, как питательность среды и температура роста, в рамках физиологических требований данного вида не играют существенной роли. Помимо снижения качества спектров в некоторых случаях наблюдалось изменение состава пиков масс-спектра для разных сред и времен инкубации, что может повлиять на возможность дифференцировать исследуемые культуры на внутривидовом уровне, особенно для «родственных» штаммов, имеющих низкую филопротеомную дистанцию, но не оказывает существенное влияние на идентификацию до вида, так же, как не препятствует идентифика-

Таблица

Результаты идентификации и рейтинг точности идентификации
для штаммов G1w1, 18(x) и 20, выращенных при различных условиях культивации

Штамм	Среда	Температура, °С	Время, ч	Результат идентификации	Рейтинг
18(x)	LB/5	65	9	Geobacillus stearothermophilus – 23	2,550
18(x)	LB/5	65	24	Geobacillus stearothermophilus – 23	2,378
18(x)	LB/5	65	48	Geobacillus stearothermophilus – 53	2,232
18(x)	LB	65	9	Geobacillus stearothermophilus – 44	2,564
18(x)	LB	65	24	Geobacillus stearothermophilus – 23	2,480
18(x)	LB	65	48	Geobacillus stearothermophilus – 17	2,520
18(x)	МПА	65	9	Geobacillus stearothermophilus – 44	2,529
18(x)	МПА	65	24	Geobacillus stearothermophilus – 17	2,285
18(x)	LB	60	ночь	Geobacillus stearothermophilus – 53	2,432
18(x)	LB	65	ночь	Geobacillus stearothermophilus – 23	2,494
18(x)	LB	70	ночь	Geobacillus stearothermophilus – 44	2,484
20	LB/5	65	9	Geobacillus stearothermophilus – 44	2,443
20	LB/5	65	24	Geobacillus stearothermophilus – 47	2,458
20	LB/5	65	48	Geobacillus stearothermophilus – 22	2,129
20	LB	65	9	Geobacillus stearothermophilus – 44	2,403
20	LB	65	24	Geobacillus stearothermophilus – 23	2,440
20	LB	65	48	Geobacillus stearothermophilus – 48	2,213
20	МПА	65	9	Geobacillus stearothermophilus – 47	2,489
20	МПА	65	24	Geobacillus stearothermophilus – 47	2,378
20	МПА	65	48	Нет достоверной идентификации	1,129
20	LB	60	ночь	Geobacillus stearothermophilus – 47	2,463
20	LB	65	ночь	Geobacillus stearothermophilus – 23	2,447
20	LB	70	ночь	Geobacillus stearothermophilus – 47	2,537
G1w1	LB/5	65	9	G1w1_65_LB	2,451
G1w1	LB/5	65	24	G1w1_65_LB	1,950
G1w1	LB/5	65	48	Нет достоверной идентификации	1,162
G1w1	LB	65	6	G1w1_65_LB	2,521
G1w1	LB	65	9	G1w1_65_LB	2,479
G1w1	LB	65	12	G1w1_65_LB	2,556
G1w1	LB	65	24	G1w1_65_LB	2,545
G1w1	LB	65	48	G1w1_65_LB	2,163
G1w1	LB	65	72	G1w1_65_LB	2,063
G1w1	МПА	65	9	G1w1_65_LB	2,465
G1w1	МПА	65	24	Geobacillus stearothermophilus – Gus2(3)	2,445
G1w1	МПА	65	48	Нет достоверной идентификации	1,438
G1w1	LB	60	ночь	G1w1_65_LB	2,492
G1w1	LB	65	ночь	G1w1_65_LB	2,619
G1w1	LB	70	ночь	G1w1_65_LB	2,510

ции штаммов с высокой филопротеомной дистанцией. Минимизировать этот фактор также позволяет проведение анализа на начальной стадии роста. Как было показано при времени инкубации 9 ч, спектры для всех используемых сред были практически идентичны.

Работа выполнена при финансовой поддержке гранта № 14.512.11.0057 Министерства образования и науки Российской Федерации.

ЛИТЕРАТУРА

Böhme K., Fernández-No I.C., Pazos M. *et al.* Identification and classification of seafood-borne pathogenic and spoilage bacteria: 16S rRNA sequencing versus MALDI-TOF MS fingerprinting // Electrophoresis. 2013. V. 34. No. 6. P. 877–887.

Cripps R.E., Elay K., Leak D.J. *et al.* Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production // Metab. Eng. 2009. V. 11. No. 6. P. 398–408.

Dickinson D.N., La Duc M.T., Satomi M. *et al.* MALDI-TOFMS compared with other polyphasic taxonomy approaches for the identification and classification of *Bacillus pumilus* spores // J. Microbiol. Meth. 2004. V. 58. No. 1. P. 1–12.

Freiwald A., Sauer S. Phylogenetic classification and identification of bacteria by mass spectrometry // Nat. Protoc. 2009. V. 4. No. 5. P. 732–742.

Mellmann A., Cloud J., Maier T. *et al.* Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass spectrometry in comparison to 16S rRNA gene sequencing for species identification of nonfermenting bacteria // J. Clin. Microbiol. 2008. V. 46. No. 6. P. 1946–54.

Nazina T.N., Tourova T.P., Poltaraus A.B. *et al.* Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans* // Intern. J. Syst. Evol. Microbiol. 2001. V. 51. No. 2. P. 433–446.

Ruelle V., El Moualij B., Zorzi W. *et al.* Rapid identification of environmental bacterial strains by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry // Rapid Commun. in Mass Spectrometry: RCM. 2004. V. 18. No. 18. P. 2013–9.

Ryzhov V., Fenselau C. Characterization of the protein subset desorbed by MALDI from whole bacterial cells // Analyt. Chem. 2001. V. 73. No. 4. P. 746–50.

Valentine N., Wunschel D., Petersen C., Wahl K. Effect of culture conditions on microorganism identification by matrix-assisted laser desorption ionization mass spectrometry // Appl. Environmental Microbiol. 2005. V. 71. No. 1. P. 58–64.

REPRODUCIBILITY OF THE RESULTS OF MICROBE IDENTIFICATION BY MALDI-TOF MASS SPECTROMETRY DEPENDING ON GROWTH CONDITIONS BY THE EXAMPLE OF *GEOBACILLUS STEAROTHERMOPHILUS*

K.V. Starostin, E.A. Demidov, A.S. Rozanov, A.V. Bryanskaya, S.E. Peltek

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: starostin@bionet.nsc.ru

Summary

Rapid and accurate methods of microorganism identification are essential in various human activities. They include primarily clinical diagnostics. In addition, they are in demand in environment monitoring, pharmacology, food industry, research, etc. There are diverse approaches to microbe identification: phenotyping, genotyping, chemotaxonomy, direct protein profiling, etc. In this work, the effects of various growth conditions, such as temperature, growth time, and nutrition medium, on the reproducibility of microbe identification by MALDI-TOF mass spectrometry are considered by the example of three *Geobacillus stearothermophilus* strains.

Key words: MALDI, direct protein profiling, microorganism identification, *Geobacillus stearothermophilus*.

УДК 579.8.06

ПРИМЕНЕНИЕ МАЛДИ ВРЕМЯПРОЛЕТНОЙ МАСС-СПЕКТРОМЕТРИИ ДЛЯ ИДЕНТИФИКАЦИИ МИКРООРГАНИЗМОВ

© 2013 г. **Е.А. Демидов, К.В. Старостин, В.М. Попик, С.Е. Пельтек**

Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия,
e-mail: scratch_nsu@ngs.ru

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

В последнее десятилетие наряду с классическими и молекулярно-биологическими методами идентификации микроорганизмов, все чаще применяется метод идентификации микроорганизмов по их белковым профилям, или прямое белковое профилирование. Данный метод не уступает по таким показателям, как точность и специфичность идентификации, однако его выгодно отличают быстрота проведения и более низкая себестоимость анализов. В данном обзоре приведены современные представления о возможностях данного метода, а также дано сравнение с используемыми методами идентификации микроорганизмов.

Ключевые слова: МАЛДИ, прямое белковое профилирование, идентификация микроорганизмов.

ВВЕДЕНИЕ

Быстрая и точная идентификация микроорганизмов является востребованной задачей во многих приложениях человеческой деятельности как научного, так и прикладного характера. Наибольшее значение среди них представляет клиническая диагностика, так как точность и скорость идентификации патогена могут сыграть решающую роль в успешности лечения. Другими важными областями являются санитарный и эпидемиологический контроль, противодействие угрозам биотерроризма, экологические и микробиологические исследования.

ФЕНОТИПИЧЕСКИЕ МЕТОДЫ ИДЕНТИФИКАЦИИ

Традиционно для идентификации микроорганизмов использовались фенотипические методы, основанные на анализе физиологических и морфологических характеристик – размер и форма микроорганизмов, условия роста, способность к спорообразованию, а также биохимические тесты – окраска по Грамму, способность специфично расщеплять определенные субстраты или устойчивость к определенным компонентам среды. Исследование физиологических и морфологических характеристик не позволяет достичь высокого таксономического разрешения и в настоящее время выполняет лишь вспомогательную функцию, а для точной идентификации на видовом и внутривидовом уровнях используются тест-системы. На данный момент создано и коммерчески доступно множество таких тест-систем, позволяющих с высокой точностью идентифицировать определенные таксономические группы (O'Hara, 2005). Отдельно можно отметить иммунологические методы анализа, основанные на антителах, специфически связывающихся с мембранными белками клеток микроорганизмов, что позволяет с помощью флюоресцентных меток идентифицировать отдельные виды и так называемые серотипы. К недостаткам фенотипических методов можно отнести тот момент, что отдельные тест-системы и наборы антител специализированы для идентификации определенных таксономических групп микроорганизмов и не предназначены для широкого скрининга образцов. Помимо

этого многие биохимические процедуры могут занимать длительное время, что очень критично в клинической диагностике и многих других областях применения.

ГЕНОТИПИЧЕСКИЕ МЕТОДЫ ИДЕНТИФИКАЦИИ

Генотипические методы основаны на анализе структуры ДНК, реализованной в виде гибридизации ДНК, секвенировании ДНК, либо анализе фингерпринта, полученного в результате сайт-специфичной рестрикции ДНК. Пионерами в этой области были ДНК-ДНК гибридизация (Notermans, Wernars, 1990) и гель-электрофорез в пульсирующем поле PFGE (Heinzen *et al*., 1990), в течение долгого времени остававшиеся «золотыми стандартами» при идентификации микроорганизмов. Прорывом в области генотипических методов стала разработка эффективных способов секвенирования ДНК, основанных на полимеразно-цепной реакции. Ввиду того что секвенирование целого генома до сих пор остается дорогой и времязатратной процедурой, при идентификации микроорганизмов используются сиквенсы отдельных участков ДНК. В основном это крайне консервативные последовательности, как, например, гены домашнего хозяйства. За последние 15 лет большую популярность завоевал метод секвенирования гена 16s рРНК. В ряде работ было показано преимущество этого метода над классическими фенотипическими методами (Becker *et al.*, 2004; Cloud *et al*., 2004; Bosshard *et al*., 2006). Также значимым преимуществом методов, основанных на ПЦР, является способность работать с микроколичествами исследуемого материала, что позволяет идентифицировать некультивируемые микроорганизмы. Другим набирающим все большую популярность является метод MLST, основанный на секвенировании одновременно нескольких генов домашнего хозяйства, что позволяет увеличить точность идентификации и дискриминационную способность метода (Maiden, 2006; Turner, Feil, 2007).

ХЕМОТАКСОНОМИЧЕСКИЕ МЕТОДЫ ИДЕНТИФИКАЦИИ

Еще одним подходом являются хемотаксономические методы, основанные на исследовании биохимического состава клетки. Методы заключаются в определении биохимических маркеров, имеющих определенную таксономическую специфичность. Еще в начале 70-х годов прошлого века проводились успешные опыты по идентификации микроорганизмов с помощью пиролитической газо-жидкостной хроматографии (Reiner *et al*., 1972). В приведенном исследовании на примере *Salmonella* spp. было показано наличие в пирохроматографических спектрах характеристичных пиков, позволяющих идентифицировать отдельные серотипы. В 1975 г. К. Фенслау была предложена идея использовать масс-спектрометрию для хемотаксономических исследований (Anhalt, Fenselau, 1975). Масс-спектры, полученные из хлороформ-метанольных экстрактов лиофилизированных клеток, имели достоверные различия для представителей различных видов бактерий. Так как масс-спектрометрические методы отличаются высокой скоростью и точностью анализа, дальнейшее развитие хемотаксономии было связано преимущественно с ними.

МЕТОД МАЛДИ

Однако действительно большой прорыв в идентификации микроорганизмов сделало появление в арсенале масс-спектрометрии «мягкого» способа ионизации молекул исследуемого вещества (Despeyroux *et al*., 1996; Krishnamurthy *et al*., 1996), такого как матрично-ассоциированная десорбция/ионизация (МАЛДИ). В комплексе с времяпролетной масс-спектрометрией это дало возможность проводить анализ сложных биоорганических молекул, в частности тяжелых, труднолетучих молекул нуклеиновых кислот и белков.

Метод МАЛДИ времяпролетной масс-спектрометрии основан на десорбции и ионизации исследуемого вещества с помощью лазерного излучения в присутствии вспомогательного вещества – матрицы с последующим разделением ионов во времяпролетном масс-анализаторе. Под воздействием лазерных импульсов матрица, сокристаллизованная с исследуемым веществом, активно поглощает излучение лазера, что приводит к ее десорбции (рис. 1). Переходя в газовую фазу, матрица увлекает за собой молекулы

исследуемого вещества, а также способствует их ионизации с образованием преимущественно однозарядных ионов (Karas *et al*., 1987). Метод позволяет проводить прямой масс-спектрометрический анализ белковой фракции микробной клетки (прямое белковое профилирование), т. е. без фракционирования и очистки отдельных белков, и получать уникальные для данного вида масс-спектры (рис. 2) с высокой точностью и разрешением, характеризующие исследуемый объект по типу «отпечатков пальцев» (Ben, van Baar, 2000).

Одним из подходов для идентификации отдельных микроорганизмов методами МАЛДИ времяпролетной масс-спектрометрии является поиск по существующим базам белковых профилей микроорганизмов. Для идентификации необходима база данных характеристичных (эталонных) спектров, представляющих собой суперспектры, полученные усреднением серии

Рис. 1. Схема процесса ионизации в условиях МАЛДИ.

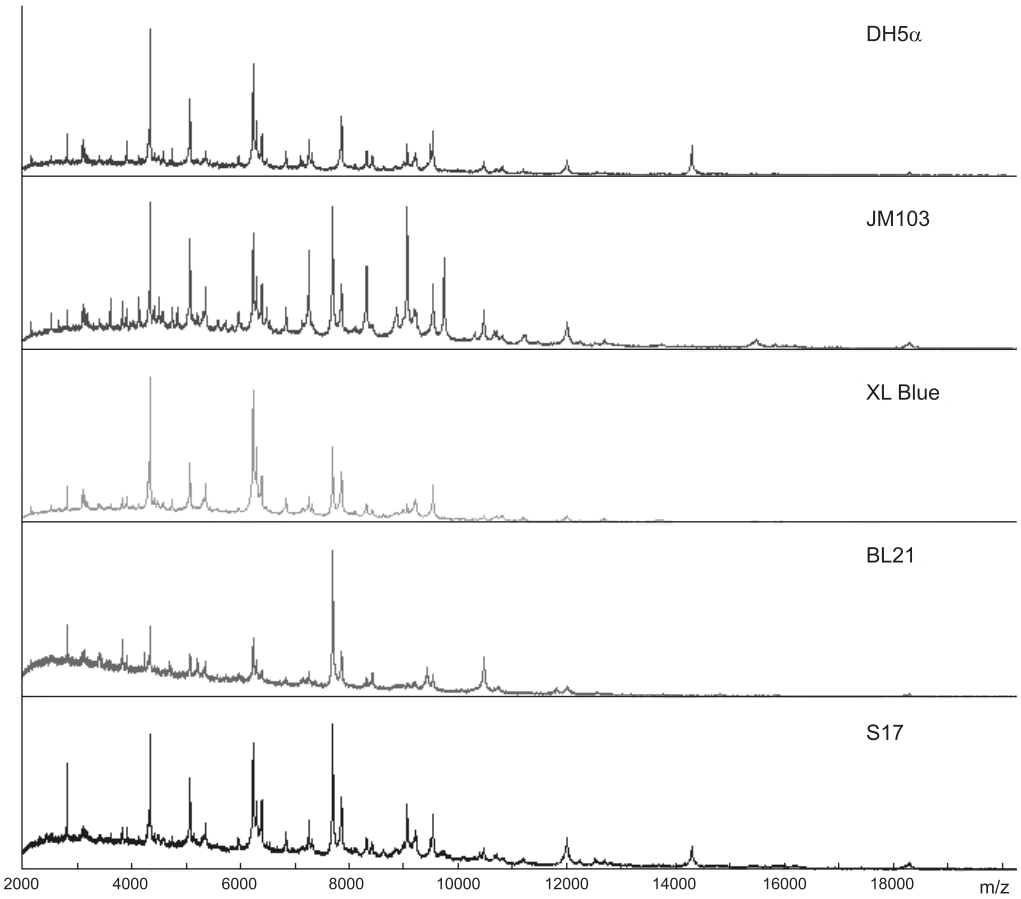


Рис. 2. Масс-спектры различных штаммов *E. coli.*

По оси абсцисс – отношение масса/заряд (m/z) .

единичных спектров, что позволяет добиться большей точности и воспроизводимости анализа. В рамках процедуры идентификации происходит попарное сравнение пиков в спектре исследуемого образца с пиками эталонных суперспектров, находящихся в базе данных. Каждому сравнению с суперспектром в базе данных присваивается численный рейтинг, вычисленный на основании количества совпадений. Идентификация микроорганизмов происходит по наилучшему совпадению (Sandrin *et al*., 2013), при этом не происходит идентификация конкретных белков.

Однако, несмотря на широкие потенциальные возможности использования этого подхода для идентификации и типирования микроорганизмов, на данный момент не существует единого протокола для пробоподготовки и идентификации. При данном подходе пробоподготовка может представлять собой простое нанесение клеточной культуры на подложку масс-спектрометра в смеси с матрицей. Однако чаще всего прибегают к различным методам лизиса клеток для получения белкового экстракта (Šedo *et al*., 2011). Также остается открытым вопрос о воспроизводимости белковых профилей отдельных микроорганизмов в зависимости от условий культивирования и стадии роста.

Данный подход к идентификации получил сейчас наибольшее распространение, поскольку позволяет быстро и надежно идентифицировать микроорганизмы, для которых получены суперспектры белковых профилей. На сегодняшний день различными фирмами выпускаются масс-спектрометры, предназначенные для решения задач идентификации микроорганизмов. Существуют также коммерчески доступные базы данных различных микроорганизмов. Все это привело к тому, что данная методика активно применяется в клинической диагностике.

Тем не менее существует биоинформационный метод бактериального профилирования. Этот метод включает в себя идентификацию отдельных пиков в белковых профилях бактерий при помощи баз данных с геномными последовательностями. Данный метод не требует создания специализированной базы данных и такой жесткой стандартизации методик культивирования и пробоподготовки (Fenselau *et al*., 2007), однако для его успешной реализации необходимы масс-спектрометры, обеспечивающие сверхвысокое разрешение, что увеличивает время анализа и стоимость оборудования.

ВОСПРОИЗВОДИМОСТЬ И СРАВНЕНИЕ С ДРУГИМИ МЕТОДАМИ ИДЕНТИФИКАЦИИ

Для идентификации микроорганизмов обычно используются спектры в диапазоне масс 2–20 кДа. Анализ масс-спектров *E. coli* в диапазоне 2–20 кДа показал, что из 2000 белков, предсказанных на основании данных секвенированного генома *E. coli*, в спектрах присутствует только 30 (Ryzhov, Fenselau, 2001). Около половины пиков были отнесены к рибосомальным белкам, оставшаяся часть – к ДНК-связывающим белкам и белкам холодового шока. Рибосомальные белки относятся к белкам домашнего хозяйства и вследствие этого являются достаточно консервативными, что обеспечивает их таксономическую специфичность. Помимо этого рибосомальные белки в большом количестве присутствуют в цитоплазме клеток – до половины массы растущей клетки, а их набор остается неизменным вне зависимости от внешних условий и стадии роста, что и обеспечивает воспроизводимость масс-спектров. Исследования внутри- и межлабораторной воспроизводимости показали высокую надежность метода. В работе S. Barbuddhe с соавт. (2008) было показано на примере видов рода *Listeria*, что при использовании трех различных приборов фирмы «Bruker Daltonics» в двух разных лабораториях не наблюдается существенных различий в полученных спектрах для представителей этого рода. В работе A. Mellmann с соавт. (2008) были исследованы 10 случайно выбранных из коллекции штаммов на трех различных приборах. Было показано, что результаты идентификации не зависят от используемого прибора и демонстрируют схожий рейтинг точности идентификации при использовании программы Biotyper. Визуальное сравнение спектров, полученных для каждого отдельного штамма на всех трех приборах, также не выявляет значительных отличий. В 2009 г. было проведено крупное исследование межлабораторной воспроизводимости с участием 8 лабораторий из различных

стран мира (Mellmann *et al*., 2009). Используя оборудование и программное обеспечение фирмы «Bruker Daltonics», каждая лаборатория проводила идентификацию 60 образцов с помощью базы данных, состоящей из 2800 штаммов. В результате идентификации 97,29 % образцов были определены до вида и только 2,5 % – до рода. Из всех образцов 98,75 % были определены верно. Данное исследование показывает, что при использовании стандартных протоколов и единой базы данных метод идентификации с помощью МАЛДИ времяпролетной масс-спектрометрии обеспечивает высокую точность и воспроизводимость.

Важным вопросом является зависимость результатов идентификации от условий культивирования микробиологических культур, таких как среда и время роста. При исследовании спектров культур *B. subtilis*, *Y. enterocolitica*, *E. coli*, каждая из которых выращена на 4 разных средах: минимальной среде M9, триптическом соевом бульоне TSB, LB и кровяном агаре, было показано что в зависимости от выбора среды состав пиков может меняться, но при этом также имеется набор пиков, присутствующих в спектрах вне зависимости от этого выбора (Valentine *et al*., 2005). В работе Ruelle с соавт. (2004) также указано на изменение состава пиков в спектре в зависимости от выбора среды. Также в данной работе было проведено исследование влияния времени инкубации на полученные спектры. Было показано на примере *E. coli*, что при длительном времени культивирования происходит исчезновение части пиков из спектра. Однако, как показано в данных работах, возникающие различия в спектрах при разных условиях культивации культур не мешают проведению достоверной идентификации.

Метод идентификации микроорганизмов с помощью МАЛДИ времяпролетной масс-спектрометрии сильно выигрывает у классических методов идентификации по время- и трудозатратности анализа и при этом не уступает им в точности. В нескольких работах было показано преимущество МАЛДИ идентификации над секвенированием гена 16S рРНК и фенотипическими тестами – одними из наиболее распространенных методов при идентификации микроорганизмов. В работе K. Böhme с соавт. (2013) провели сравнение этих двух методов на 50 штаммах, представляющих собой патогены, выделенные из морских пищевых продуктов. Идентификация с помощью секвенирования гена 16S рРНК позволила определить 50 % штаммов до видового уровня, в то время как с помощью МАЛДИ удалось идентифицировать 76 %. Из 25 штаммов, для которых не удалось получить видовую идентификацию с помощью секвенирования, 18 были идентифицированы с помощью МАЛДИ. Для сравнения, только 5 штаммов из 12 неидентифицированных до вида с помощью МАЛДИ удалось определить с помощью секвенирования гена 16S рРНК. Д. Диксоном и коллегами было проведено исследование результатов идентификации спор *B. pumilus* с помощью фенотипического метода, основанного на «Biolog identification system», секвенировании генов 16S рРНК и *gyrB*, ДНК-ДНК гибридизации и МАЛДИ времяпролетной масс-спектрометрии (Dickinson *et al*., 2004). С помощью метода «Biolog identification system» было определено только 3 штамма из 18 как *B. pumilus*. 8 штаммов были некорректно определены как *B. subtilis*. Оставшиеся 7 штаммов не имели совпадений с базой данных этой системы. Другие 4 метода идентифицировали все штаммы, при этом МАЛДИ времяпролетная масс-спектрометрия, секвенирование гена *gyrB* и ДНК-ДНК гибридизация продемонстрировали схожую дискриминационную способность, заметно превосходя метод секвенирования гена 16S рРНК. Более высокая дискриминационная способность МАЛДИ времяпролетной масс-спектрометрии по сравнению с секвенированием гена 16S рРНК была также показана в работе Mellmann с соавт. (2008) на примере комплекса *B. cepacia*, реклассифицированного на основании данных полиморфизма гена *recA*. Метод секвенирования гена 16S рРНК в отличие от МАЛДИ времяпролетной масс-спектрометрии показал неспособность дифференцировать штаммы в рамках этого комплекса. Также в данной работе было проведено сравнение результатов идентификации 80 клинически значимых штаммов с помощью масс-спектрометрии, позволившей определить 82,5 % изолятов до видового уровня и 95,2 % – до рода, с фенотипическими тестами API 20NE и Vitek 2, определившими соответственно 61 и 54 % изолятов.

Особой важностью обладает вопрос внутривидовой идентификации, позволяющей идентифицировать отдельные штаммы, имеющие клиническое и прикладное значение. Помимо приведенных выше работ (Dickinson *et al*., 2004; Mellmann *et al*., 2008) успешная дифференциация на внутривидовом уровне показана в работе Ghyselinck с соавт. (2011). Метод МАЛДИ времяпролетной масс-спектрометрии продемонстрировал высокое таксономическое разрешение, позволяющее различать на внутривидовом уровне представителей родов *Stenotrophomonas*, *Bacillus*, *Rhodococcus* и *Pseudomonas*. В работе S. Barbuddhe с соавт. (2008) масс-спектрометрический анализ корректно объединил серотипы *Listeria monocytogenes* в три линии в соответствии с данными PFGE (pulsed-field gel-electrophoresis). Seibold с соавт. (2010) удалось идентифицировать 45 изолятов *Francisella tularensis* на уровне штаммов.

ЗАКЛЮЧЕНИЕ

Приведенные работы показывают, что метод идентификации с помощью МАЛДИ времяпролетной масс-спектрометрии не уступает большинству известных методов по точности идентификации и дискриминационной способности, а некоторые из них заметно превосходит. Метод обладает большим потенциалом для идентификации на внутривидовом уровне, который будет реализовываться с усовершенствованием приборной базы, появлением более представленных библиотек эталонных суперспектров и сиквенсов бактериальных геномов.

Работа была выполнена в рамках Государственного контракта № 14.512.11.0050 «Создание методов метаболической инженерии термофильных микроорганизмов для получения штаммов-продуцентов молочной кислоты».

ЛИТЕРАТУРА

Anhalt J.P., Fenselau C. Identification of bacteria using mass spectrometry // Analyt. Chem. 1975. V. 500. No. 2. P. 219–225.

Barbuddhe S.B., Maier T., Schwarz G. *et al.* Rapid identification and typing of listeria species by matrix-assisted laser desorption ionization-time of flight mass spectrometry // Appl. Environ. Microbiol. 2008. V. 74. No. 17. P. 5402–5407.

Becker K., Harmsen D., Mellmann A. *et al.* Development and evaluation of a quality-controlled ribosomal sequence database for 16S ribosomal DNA-based identification of Staphylococcus species // J. Clin. Microbiol. 2004. V. 42. No. 11. P. 4988–4995.

Ben L.M., van Baar. Characterisation of bacteria by matrix-assisted laser desorption/ionization and electrospray mass spectrometry // FEMS Microbiol. Rev. 2000. V. 24. No. 2. P. 193–219.

Böhme K., Fernández-No I.C., Pazos M. *et al.* Identification and classification of seafood-borne pathogenic and spoilage bacteria: 16S rRNA sequencing versus MALDI-TOF MS fingerprinting // Electrophoresis. 2013. V. 34. No. 6. P. 877–887.

Bosshard P.P., Zbinden R., Abels S. *et al.* 16S rRNA gene sequencing versus the API 20 NE system and the VITEK 2 ID-GNB card for identification of nonfermenting gram-negative bacteria in the clinical laboratory // J. Clin. Microbiol. 2006. V. 44. No. 4. P. 1359–1366.

Cloud J.L., Conville P.S., Croft A. *et al.* Evaluation of partial 16S ribosomal DNA sequencing for identification of Nocardia species by using the MicroSeq 500 system with an expanded database // J. Clin. Microbiol. 2004. V. 42. No. 2. P. 578–584.

Despeyroux D., Phillpotts R., Watts P. Electrospray mass spectrometry for detection and characterization of purified cricket paralysis virus (CrPV) // Rapid Commun. Mass Spectrom. 1996. V. 10. No. 8. P. 937–941.

Dickinson D.N., La Duc M.T., Satomi M. *et al.* MALDI-TOFMS compared with other polyphasic taxonomy approaches for the identification and classification of *Bacillus pumilus* spores // J. Microbiol. Meth. 2004. V. 58. No. 1. P. 1–12.

Fenselau C., Russell S., Swatkoski S., Edwards N. Proteomic strategies for rapid characterization of micro-organisms // Eur. J. Mass Spectrom. 2007. V. 13. No. 1. P. 35–39.

Ghyselinck J., Van Hoorde K., Hoste B. *et al.* Evaluation of MALDI-TOF MS as a tool for high-throughput dereplication // J. Microbiol. Meth. 2011. V. 86. No. 3. P. 327–336.

Heinzen R., Stiegler G.L., Whiting L.L. *et al.* Use of pulsed field gel electrophoresis to differentiate Coxiella burnetii strains // Ann. N.Y. Acad. Sci. 1990. V. 590. P. 504–513.

Karas M., Bachmann D., Bahr D., Hillenkamp F. Matrix-assisted ultraviolet-laser desorption of nonvolatile compounds // Int. J. Mass Spectrom. Ion Proc. 1987. V. 78. P. 53–68.

Krishnamurthy T., Ross P.L., Rajamani U. Detection of pathogenic and non-pathogenic bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry// Rapid Commun. Mass Spectrom. 1996. V. 10. No. 8. P. 883–888.

Maiden M.C. Multilocus sequence typing of bacteria // Annu. Rev. Microbiol. 2006. V. 60. P. 561–588.

Mellmann A., Bimet F., Bizet C. *et al.* High interlaboratory reproducibility of matrix-assisted laser desorption ionization-time of flight mass spectrometry-based species identification of nonfermenting bacteria // J. Clin. Microbiol. 2009. V. 47. No. 11. P. 3732–3734.

Mellmann A., Cloud J., Maier T. *et al.* Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass spectrometry in comparison to 16S rRNA gene sequencing

for species identification of nonfermenting bacteria // J. Clin. Microbiol. 2008. V. 46. No. 6. P. 1946–1954.

Notermans S., Wernars K. Evaluation and interpretation of data obtained with immunoassays and DNA-DNA hybridization techniques // Intern. J. Food Microbiol. 1990. V. 11. No. 1. P. 35–49.

O'Hara C. Manual and automated instrumentation for identification of Enterobacteriaceae and other aerobic gram-negative bacilli // Clin. Microbiol. Rev. 2005. V. 18. No. 1. P. 147–162.

Reiner E., Hicks J.J., Ball M.M., Martin W.J. Rapid characterization of salmonella organisms by means of pyrolysis-gas-liquid chromatography // Analyt. Chem. 1972. V. 44. No. 6. P. 1058–1061.

Ruelle V., El Moualij B., Zorzi W. *et al.* Rapid identification of environmental bacterial strains by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry // Rapid Commun. Mass Spectrom. 2004. V. 18. No. 18. P. 2013–2019.

Ryzhov V., Fenselau C. Characterization of the protein subset desorbed by MALDI from whole bacterial cells // Analyt. Chem. 2001. V. 73. No. 4. P. 746–750.

Sandrin T.R., Goldstein J.E., Schumaker S. MALDI TOF MS profiling of bacteria at the strain level: A review // Mass Spectrom. Rev. 2013. V. 32. No. 3. P. 188–217.

Šedo O., Sedláček I., Zdráhal Z. Sample preparation methods for MALDI-MS profiling of bacteria // Mass Spectrom. Rev. 2011. V. 30. No. 3. P. 417–434.

Seibold E., Maier T., Kostrzewa M. *et al.* Identification of *Francisella tularensis* by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry: fast, reliable, robust, and cost-effective differentiation on species and subspecies levels // J. Clin. Microbiol. 2010. V. 48. No. 4. P. 1061–1069.

Turner K.M., Feil E.J. The secret life of the multilocus sequence type // Intern. J. Antimicrobial Agents. 2007. V. 29. P. 129–135.

Valentine N., Wunschel S., Wunschel D. *et al.* Effect of culture conditions on microorganism identification by matrix-assisted laser desorption ionization mass spectrometry // Appl. Environ. Microbiol. 2005. V. 71. No. 1. P. 58–64.

MALDI-TOF MASS SPECTROMETRY IN MICROORGANISM IDENTIFICATION

E.A. Demidov, K.V. Starostin, V.M. Popik, S.E. Peltek

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: scratch_nsu@ngs.ru

Identification of microorganisms from their protein profiles, or direct protein profiling, has been extensively used in the last decade. Being sufficiently accurate and specific, this method is fast and inexpensive. In this review, current notions of the potential of this method are considered and it is compared with other approaches to microorganism identification.

Key word: MALDI-TOF, direct protein profiling, microorganism identification.

УДК 663.15:663.51:579.66

БИОТЕХНОЛОГИЧЕСКИЙ ПОТЕНЦИАЛ НОВОЙ ТЕХНИЧЕСКОЙ КУЛЬТУРЫ – МИСКАНТУС СОРТ СОРАНОВСКИЙ

© 2013 г. **Н.М. Слынько, Т.Н. Горячковская, С.В. Шеховцов, С.В. Банникова, Н.В. Бурмакина, К.В. Старостин, А.С. Розанов, Н.Н. Нечипоренко, С.Г. Вепрев, В.К. Шумный, Н.А. Колчанов, С.Е. Пельтек**

Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия,
e-mail: peltek@bionet.nsc.ru

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

Мискантус сорт Сорановский внесен в Государственный реестр селекционных достижений, допущенных к использованию, в 2013 г. Методами фенотипирования и анализа ДНК новая техническая культура отнесена к виду *Miscanthus sacchariflorus*. На основе биомассы мискантуса ведется разработка методов получения сахаросодержащего субстрата для биотехнологий. Для выполнения этой задачи было изучено влияние механической и химической обработок в сочетании с ферментативным гидролизом на получение сахаров из биомассы мискантуса. Наилучший результат получен при использовании помола до размера частиц ~100 мкм с последующей обработкой гидроксидом кальция. После ферментативного гидролиза таких образцов были получены субстраты с концентрацией восстанавливающих сахаров от 10 до 15 г/л.

Ключевые слова: целлюлозосодержащее сырье, мискантус, ферментативный гидролиз, биомасса, генотипирование.

ВВЕДЕНИЕ

Мискантус – род многолетних травянистых растений семейства мятликовых. К роду *Miscanthus* относят более 20 видов, распространенных от тропической и Южной Африки до Восточной и Юго-Восточной Азии. В России на Дальнем Востоке встречается 3 вида: мискантус сахароцветный (*Miscanthus sacchariflorus*), мискантус краснеющий (*Miscanthus purpurascens*), мискантус китайский (*Miscanthus sinensis*) (Открытый иллюстрированный атлас …, www.plantarium.ru/page/view/item/41884.html).

В настоящее время в мире происходят увеличение площадей культивирования мискантуса (формы M. × giganteus) и разработка новых технологий на основе его биомассы. Мискантус считается одним из самых эффективных аккумуляторов солнечной энергии на планете (Dohleman, Long, 2009). Целлюлоза растительной биомассы представляет собой практически неисчерпаемый источник возобновляемого сырья, которое может быть конвертировано ферментативным путем в глюкозу. В свою очередь, глюкоза является незаменимым сырьем для микробиологических процессов получения жидких и газообразных видов топлива (этанола, бутанола, этилена и др.), органических и аминокислот, кормового белка и многих других полезных продуктов микробиологического синтеза. Основной проблемой, ограничивающей использование растительной биомассы в биотехнологии, является отсутствие технологии обработки сырья, способной превратить клеточную стенку растений в субстрат для микроорганизмов. Разработка энергоэффективных технологий получения сахаросодержащего сырья из целлюлозосодержащей биомассы является краеугольным камнем современной биотехнологии. Задача ферментирования рас-

тительной биомассы решается уже длительное время, но интенсивные исследования были начаты относительно недавно с целью разработки технологии ферментативного биокатализа для промышленного производства (Dashtban *et al*., 2009; Canilha *et al*., 2012; Jönsson *et al*., 2013; Silva *et al*., 2013).

Мискантус культивируется в сибирских условиях Институтом цитологии и генетики СО РАН. Его главные качества:

– высокая урожайность при минимальных затратах на возделывание (возможность получать до 15 т сухой массы в течение 15–20 лет после однократных затрат на его посадку);

– способность расти на почвах, непригодных для традиционного земледелия;

– хорошее соотношение содержания холоцеллюлозы и лигнина (Шумный, 2010а, б).

Мискантус сорт Сорановский внесен в Государственный реестр селекционных достижений, допущенных к использованию, в 2013 г. По морфологическим данным образцы мискантуса сорта Сорановский могут относиться или к *M. sinensis*, или к *M. sacchariflorus*. Для точной видовой идентификации нового сорта был проведен анализ последовательности ДНК фрагмента пластидного межгенного спейсера.

Для оценки биотехнологического потенциала новой технической культуры проведен сравнительный анализ гидролизатов биомассы мискантуса коммерчески доступными целлюлазами после различных предобработок. Было изучено влияние механической и химических обработок на получение сахаров из биомассы мискантуса.

МАТЕРИАЛЫ И МЕТОДЫ

Для исследований была использована биомасса мискантуса, выращенного на экспериментальных полях ИЦиГ СО РАН, урожая 2012 г.

Для выделения ДНК к 50 мг молотого мискантуса добавляли 1 мл буфера, содержащего 3 % CTAB, 1 M NaCl, 10 мМ Tris-HCl (pH 8,0) и 2 мМ ЭДТА, и инкубировали 1 ч при 60 °С при постоянном перемешивании. Затем к полученной смеси добавляли 1 мл хлороформа и центрифугировали 10 мин при 13000 об/мин. К супернатанту добавляли равный объем изопропанола, перемешивали и центрифугировали 10 мин при 13000 об/мин. Осадок промывали 75 %-м этанолом, высушивали и растворяли в бидистиллированной воде.

Праймеры для амплификации фрагмента пластидного межгенного спейсера между генами tRNA-Leu и tRNA-Phe: Misc-Leu-Fw (5′-TGGAA-GCTGT-TCTAA-CGAAT-C-3′) и Misc-Leu-Rv (5′-AATGG-GACTC-TCTCT-TTATC-CTC-3′) были синтезированы фирмой «Биосан». Амплификацию фрагмента пластидного межгенного спейсера между генами *tRNA-Leu* и *tRNA-Phe* проводили по следующей схеме: 50 мкл реакционной смеси содержали 1,5 мМ $MgCl_2$, 65 мМ Tris-HCl (pH 8,8), 16 мМ $(NH_4)_2SO_4$, 0,05 % Tween-20, по 0,2 мМ dNTP, 0,3 мМ праймеров (tRNA-Leu и tRNA-Phe: Misc-Leu-Fw (5′-TGGAA-GCTGT-TCTAA-CGAAT-C-3′) и Misc-Leu-Rv (5′-AATGG-GACTC-TCTCT-TTATC-CTC-3′)), 1 нг матричной ДНК и 1 е.а. рекомбинантной Taq-полимеразы (SibEnzyme, Новосибирск). Профиль реакции: стадия денатурации – 2 мин при 94 °С; 30 циклов: 94 °С, 20 с; 55 °С, 20 с; 72 °С, 20 с. Полученный продукт секвенировали при помощи набора BigDye 3.1 (Applied Biosystems) с двух сторон при помощи указанных выше праймеров. Капиллярный электрофорез проводили в Межинститутском центре секвенирования СО РАН.

Помол проводили измельчителем МАН-30 (производства ЗАО МВМ, РФ). Порошки смешивали с водой в соотношении жидкая фаза к твердой – Ж/Т, мл/г, равном 10. Методики анализов биомассы взяты из книги «А.В. Оболенской, З.П. Ельницкой, А.А. Леоновича. Лабораторные работы по химии древесины и целлюлозы» (М.: Экология, 1991). Определялись: растворимость в горячей воде (экстрактивность), выход холоцеллюлозы (целлюлоза + гемицеллюлоза) хлорным методом, зольность, содержание альфа-целлюлозы и лигнина.

Ферментативный гидролиз проводили в течение 24 ч коммерчески доступным препаратом ЦеллоЛюксА при 50 °С. Фермент добавляли в соотношении 1 : 100 в расчете на исходную биомассу мискантуса влажностью 8 %, pH реакционной смеси доводили соляной кислотой до 5,5, перед добавлением ферментов смесь автоклавировали при 0,5 атм 20 мин.

Для химической обработки были использованы серная кислота (концентрация 1,5 %;

температура процесса 100 °С; соотношение Ж/Т, мл/г составляло 10, продолжительность реакции 20 ч); уксусная кислота (60 %, 100 °С; Ж/Т = 6, 3 ч), гидроксид натрия (2 %; 100 °С; Ж/Т = 6, 2 ч) и гидроксид кальция (7,5 %, 60 °С; Ж/Т = 10, 240 ч).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Определение видовой принадлежности мискантуса сорта Сорановский

Мискантус сорт Сорановский внесен в Государственный реестр селекционных достижений, допущенных к использованию, в 2013 г. Исходный селекционный материал был получен в результате ботанических и ресурсных исследований Дальнего Востока (экспедиции ИЦиГ СО РАН в 1990-х годах) (Шумный, 2010а). В районах обследования род *Miscanthus* представлен тремя видами:

1. *M. sacchariflorus* (Maxim.) Benth. – мискантус сахароцветный. Растения с тонкими, ползучими корневищами. Нижняя цветковая чешуя с короткой прямой остью. Волоски хохолка белые, иногда красноватые.

2. *M. sinensis* Anderss. – мискантус китайский. Растения с укороченным толстым корневищем. Колосковые чешуи несут длинные волоски. Волоски хохолка негустые, сильно оттопыренные, грязно-белые, но нередко бывают красноватые.

3. *M. purpurascens* Anderss. – мискантус краснеющий. Растения с длинными, горизонтальными корневищами. Ветки соцветия неветвистые. Колосковые чешуи не несут волосков. Хохолки после цветения густые, красноватые.

По морфологическим данным образцы мискантуса сорта Сорановский не были отнесены к конкретному виду. Для точной идентификации видовой принадлежности был проведен анализ консервативного участка его геномной ДНК.

На основе последовательностей представителей рода *Miscanthus* из GenBank был проведен дизайн праймеров для амплификации фрагмента пластидного межгенного спейсера между генами tRNA-Leu и tRNA-Phe. Этот короткий фрагмент (202 п.н. без праймеров) содержит две полиморфные позиции, отличающие *M. sinensis* от *M. sacchariflorus*. Полученный фрагмент ДНК длиной 202 п.н. был идентичен последовательностям *M. sacchariflorus* из GenBank, а также последовательностям *M.* × *giganteus*, *M. lutarioriparius* и *M. changii*. Известно, что *M.* × *giganteus* является аллотриплоидным гибридом *M. sacchariflorus* и *M. sinensis* и имеет пластидный геном от *M. sacchariflorus* (Hodkinson *et al*., 2002a). *M. lutarioparius* – одно из названий *M. sacchariflorus* spp. *lutarioparius* – подвида *M. sacchariflorus* (Sun *et al*., 2010). *M. changii* встречается также под названием *M. longiberbis* var. *changii*. Указаний на систематические отношения этого вида с *M. sacchariflorus* и *M. sinensis* найти не удалось.

Полученная последовательность отличалась одной заменой T > G в позиции выравнивания 60 от последовательности *M. sinensis* AB622625. От всех остальных последовательностей *M. sinensis*, а также *M. oligostachyus*, *M. junceus* и *M. floridulus* она отличалась двумя заменами: T > G в позиции 60 и A > T в позиции 183 (AB622623, AB622624, AB622626-AB622628, AJ426570-AJ426573, EU434104, GQ870006, JN544253, JN544254, JN642289-JN642291).

Виды *M. oligostachyus*, *M. junceus* и *M. floridulus* – близкородственные виды к *M. sinensis* и *M. sacchariflorus*; *M. oligostachyus* встречается также под названием *Miscanthus sinensis* var. *purpurascens*. По данным молекулярно-филогенетического анализа (Hodkinson *et al*., 2002b), представители этой группы видов не выделяются в монофилетичные ветви. Кроме того, систематика этой группы осложняется случаями неправильного определения.

Так как по морфологическим данным образцы мискантуса сорта Сорановский могут относиться или к виду *M. sinensis*, или к виду *M. sacchariflorus*, то на основании анализа последовательности ДНК фрагмента пластидного межгенного спейсера tRNA-Leu-Phe мискантус сорта Сорановский был отнесен к виду *M. sacchariflorus*.

Предобработка и гидролиз биомассы

Растворимость нарезанной соломы мискантуса в горячей воде весьма невелика – всего 3,8 % от взятой массы. Помол даже без каких-либо добавок заметно увеличивает растворимость биомассы в горячей воде – до 7,7 % (экстрактивные вещества). Содержание холоцеллюлозы (70 %)

характеризует предельно возможный процент конверсии биомассы в сахара, поскольку других полисахаридов в биомассе нет. На рис. 1 приведен анализ химического состава биомассы.

Альфа-целлюлоза входит в состав холоцеллюлозы, однако ее содержание представляет самостоятельный интерес, в силу того что это наиболее трудно гидролизуемая часть растительной биомассы. Лигнин – сложное полимерное вещество, не относящееся к полисахаридам. Лигнин скрепляет целлюлозные волокна и обеспечивает твердый и жесткий матрикс, усиливающий прочность клеточных стенок на растяжение и в особенности на сжатие. Это главный опорный материал дерева. Он также предохраняет клетки от повреждения под действием физических и химических факторов. Перевод лигнина в растворимую форму является принципиальным моментом, существенно облегчающим гидролиз биомассы. Лигнин имеет самостоятельную коммерческую ценность, возможно, его отделение и реализация как побочного продукта может способствовать повышению рентабельности комплексной переработки биомассы мискантуса.

Размеры клетки листа злаковых составляют 30–40 мкм (Зверева, 2010). Получение порошков биомассы с размерами частиц порядка размеров клетки может привести к существенному увеличению выхода сахаров в процессе ферментативного гидролиза. Развитие современных технологий помола позволяет достигать тонины помола, сопоставимой с размерами клетки листа злаков. Для помола соломы мискантуса использован измельчитель МАН-30. Для повышения степени деструкции биомассы в процессе помола были добавлены хлорид натрия, поташ (K_2CO_3) или речной песок в соотношении 10 % по массе. Предобработку биомассы проводили с целью

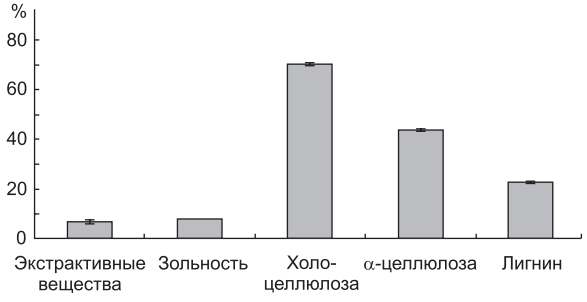


Рис. 1. Химический анализ биомассы мискантуса.

увеличения выхода сахаров при последующем ферментативном гидролизе.

После помола был проведен ферментативный гидролиз. На рис. 2 приведены результаты анализа полученных гидролизатов на содержание глюкозы и восстанавливающих сахаров.

Из рис. 2 видно, что в исходных препаратах практически отсутствует глюкоза, а концентрация общих восстанавливающих сахаров составляет чуть больше 2 г/л. Ферментативный гидролиз неодинаково эффективен после помолов с различными добавками. Интересно, что достаточно высокие концентрации сахаров, как глюкозы, так и общих восстанавливающих, получены после помола с речным песком. Возможно, достаточно твердые частицы песка во время помола увеличивают степень деструкции растительных тканей, облегчая впоследствии доступ ферментов к полисахаридам. Наибольший выход сахаров получен после помола с K_2CO_3. По-видимому, дополнительная щелочная обработка оказалась достаточно благоприятной для последующего ферментативного гидролиза.

Для той же цели – повышение выхода сахаров при ферментативном гидролизе – после помола биомассы были проведены химические

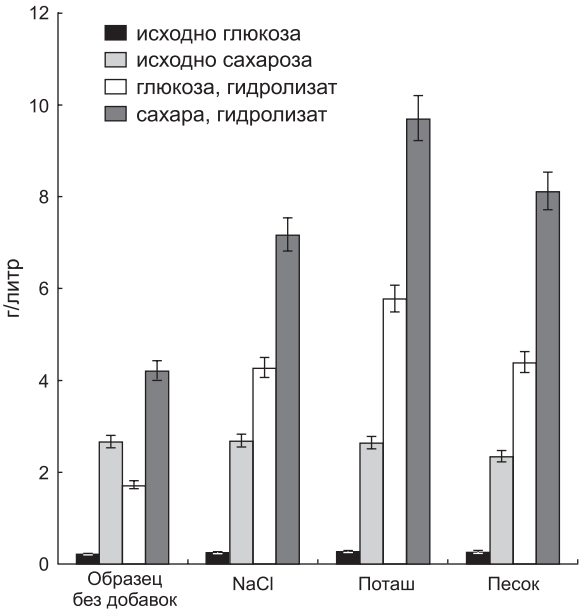


Рис. 2. Концентрация глюкозы и общих восстанавливающих сахаров в образцах биомассы до и после ферментативного гидролиза.

Образец без добавок – суспензия измельченной биомассы в воде, NaCl, поташ и песок – суспензия измельченной с соответствующей добавкой биомассы в воде.

предобработки. По окончании реакции смеси подвергались ферментативному гидролизу. Результаты анализа содержания глюкозы и общих восстанавливающих сахаров до и после ферментативного гидролиза в реакционных смесях приведены на рис. 3.

Из рис. 3 видно, что исходные (негидролизованные ферментантивно, но по-разному обработанные) образцы биомассы существенно различаются не только по количеству глюкозы и общих восстанавливающих сахаров, но и по их соотношению, вероятно, в результате частичного гидролиза гемицеллюлозы химическими агентами, например, после обработки серной кислотой. После предобработки серной кислотой наблюдается максимальный химический гидролиз. В этой реакционной смеси ферментативный гидролиз привнес лишь незначительную прибавку по выходу общих сахаров, прибавка по содержанию глюкозы была значительнее – более 1 г/л. Минимальный химический гидролиз прошел при обработке гидроксидом кальция. В этой реакционной смеси продукция сахаров получена в основном за счет ферментативного гидролиза. Оба помола и уксуснокислая делигнификация дали около 2 г/л общих восстанавливающих сахаров, однако ферментативный гидролиз эффективно прошел только после уксуснокислой обработки биомассы.

Таким образом, наибольший выход сахаров получен в результате ферментативного гидролиза образцов, полученных помолом с последующими предобработками: уксуснокислой делигнификацией, обработкой серной кислотой или гидроксидом кальция. Следует отметить, что содержание биомассы в образцах неодинаково: в образцах $Ca(OH)_2$ и H_2SO_4 Ж/Т = 10, в то время как в образце AcOH Ж/Т = 6.

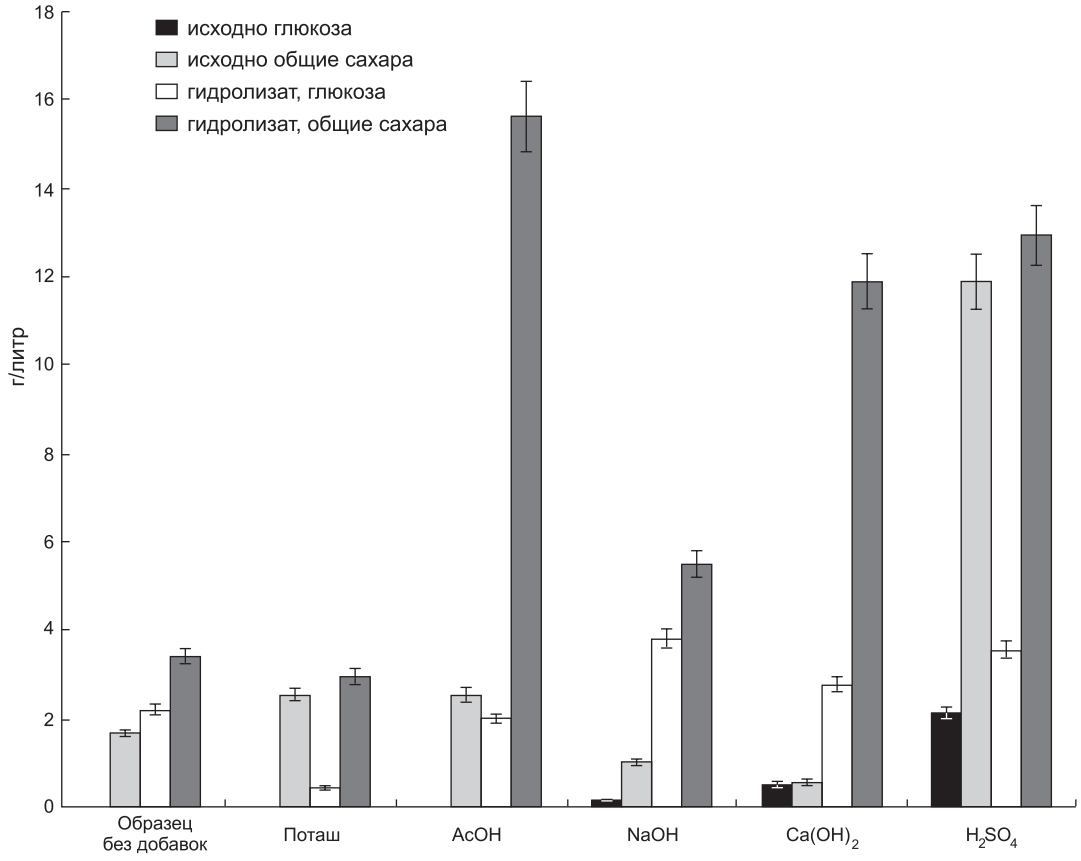


Рис. 3. Концентрация глюкозы и общих восстанавливающих сахаров в образцах биомассы до и после ферментативного гидролиза.

Образец без добавок – суспензия измельченной биомассы в воде, поташ – суспензия измельченной с поташем биомассы в воде, AcOH – обработка уксусной кислотой, NaOH, $Ca(OH)_2$ и H_2SO_4 – обработки соответствующими реагентами. Для образцов: без добавок, поташ и AcOH – исходные значения по глюкозе не определяли.

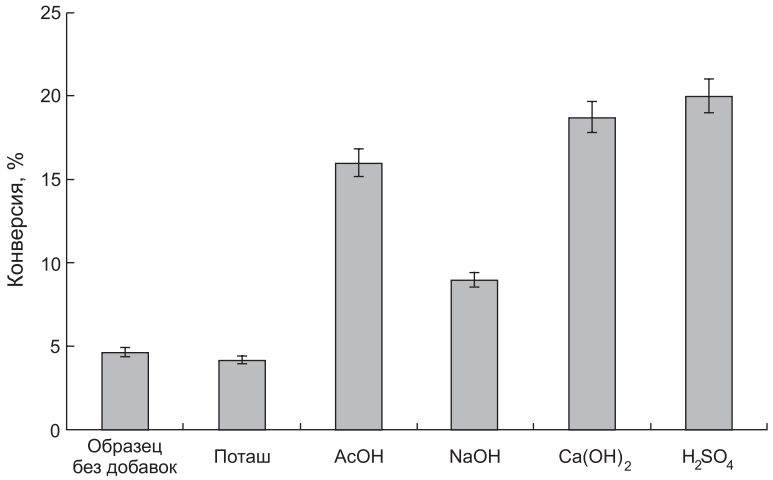


Рис. 4. Конверсия биомассы в общие восстанавливающие сахара после механической, химической и последующей ферментативной обработок.

В результате механической, химической и последующей ферментативной обработок конверсия биомассы в общие восстанавливающие сахара составляет от 4 до 20 % (рис. 4).

Расчет степени конверсии биомассы в сахара проводили с учетом соотношения жидкой и твердой фаз в реакционных смесях. В пересчете на 70 % полисахаридов в составе биомассы и пропорцию Ж/Т получаем предельно возможную концентрацию полисахаридов 64 г/л в случае Ж/Т = 10 и 99 г/литр в случае Ж/Т = 6. Для различных видов помола полученные 3 г/л общих восстанавливающих сахаров составляют 4,7 %-ю конверсию биомассы от теоретически возможного значения. Для уксусной кислоты полученные 16 г/л общих восстанавливающих сахаров составляют 16 %-ю конверсию биомассы от теоретически возможного значения. Для $Ca(OH)_2$ полученные 12 г/л составляют 18,75 % от теоретически возможного значения.

Таким образом, в результате механической, химической и последующей ферментативной обработок конверсия биомассы в общие восстанавливающие сахара составляет 16–18 % от теоретически возможного значения. Наилучший результат получен при использовании помола с последующей обработкой гидроксидом кальция.

Показано, что общее содержание полисахаридов в биомассе мискантуса составляет 70 %, а лигнина – более 20 %. Мискантус может служить перспективным источником возобновляемого сырья для получения сахаросодержащих сиропов, широко востребованных для биотехнологических нужд и получения лигнина. Улучшение ферментативного гидролиза биомассы до предельно возможного и разработка методов выделения лигнина позволят перерабатывать биомассу мискантуса более чем на 90 %.

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации (ГК 14.512.11.0072 от 19.04.2013 г.) в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 гг.»

ЛИТЕРАТУРА

Зверева Г.К. Структурная организация мезофилла листовых пластинок злаков увлажненных местообитаний // Электрон. журн. Ботан. сада-института ДВО РАН. 2010. Т. 5. Р. 51–54.

Оболенская А.В., Ельницкая З.П., Леонович А.А. Лабораторные работы по химии древесины и целлюлозы. М.: Экология, 1991.

Открытый иллюстрированный атлас сосудистых растений России и сопредельных стран http://www.plantarium.ru/page/view/item/41884.html

Шумный В.К., Колчанов Н.А., Сакович Г.В. и др. Поиск возобновляемых источников целлюлозы для многоцелевого использования // Вестн. ВОГиС. 2010а. Т. 14. № 3. С. 569–578

Шумный В.К., Вепрев С.Г., Нечипоренко Н.Н. и др. Новая форма мискантуса китайского (Веерника китайского *Miscanthus sinensis* Anders.) как перспективный источник целлюлозосодержащего сырья // Вестн. ВОГиС. 2010б. Т. 14. № 1. С. 122–126.

Canilha L., Chandel A.K., dos Santos Milessi T.S. *et al*. Bioconversion of sugarcane biomass into ethanol: an overview about composition, pretreatment methods, detoxification of hydrolysates, enzymatic saccharification, and ethanol fermentation // J. Biomed Biotechnol. 2012.

989572. Published online 2012 November 26. doi: 10.1155/2012/989572

Dashtban M., Schraft H., Qin W. Fungal bioconversion of lignocellulosic residues; opportunities & perspectives // Int. J. Biol. Sci. 2009. V. 5. No. 6. P. 578–595.

Dohleman F.G., Long S.P. More productive than maize in the midwest: how does Miscanthus do it? // Plant. Physiol. 2009. V. 150. No. 4. P. 2104–2115.

Hodkinson T.R., Chase M.W., Lledó M.D. *et al*. Phylogenetics of Miscanthus, Saccharum and related genera (Saccharinae, Andropogoneae, Poaceae) based on DNA sequences from ITS nuclear ribosomal DNA and plastid trnLintron and trnL-F intergenic spacers // J. Plant Res. 2002b. V. 115. No. 5. P. 381–392.

Hodkinson T.R., Chase M.W., Takahashi C. *et al*. The use of dna sequencing (ITS and trnL-F), AFLP, and fluorescent in situ hybridization to study allopolyploid Miscanthus (Poaceae) // Amer. J. Bot. 2002a. V. 89. No. 2. P. 279–286. doi: 10.3732/ajb.89.2.279.

Jönsson L.J., Alriksson B., Nilvebrant N.-O. Bioconversion of lignocellulose: inhibitors and detoxification // Biotechnol. Biofuels. 2013. V. 6. Issue 1. P. 16. Published online 2013 doi: 10.1186/1754-6834-6-16

Silva J.P.A., Carneiro L.M., Roberto I.C. Treatment of rice straw hemicellulosic hydrolysates with advanced oxidative processes: a new and promising detoxification method to improve the bioconversion process // Biotechnol. Biofuels. 2013. 6: 23. Published online 2013 February 15. doi: 10.1186/1754-6834-6-23

Sun Q., Lin Q., Yi Z-L. *et al*. A taxonomic revision of Miscanthus s.l. (Poaceae) from China // Bot. J. Linn. Soc. 2010. V. 164. P. 178–220.

THE BIOTECHNOLOGICAL POTENTIAL OF THE NEW CROP, MISCANTHUS CV. SORANOVSKII

N.M. Slynko, T.N. Goryachkovskaya, S.V. Shekhovtsov, S.V. Bannikova, N.V. Burmakina, K.V. Starostin, A.S. Rozanov, N.N. Nechiporenko, S.G. Veprev, V.K. Shumny, N.A. Kolchanov, S.E. Peltek

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: peltek@bionet.nsc.ru

Summary

Miscanthus cv. Soranovskii was registered as a state breeding achievement in 2013. On the base of phenotyping and DNA sequencing, the new crop was identified as *Miscanthus sacchariflorus*. The development of methods for converting its biomass to sugar-containing substrate for biotechnological use is in progress. Effects of mechanical and chemical treatments combined with enzymatic hydrolysis on sugar production from Miscanthus biomass have been studied. The best procedure includes grinding to ~100 μm particles followed by treatment with calcium hydroxide. Enzymatic hydrolysis of such materials yields substrates containing 10 to 15 g/L reducing sugars.

Key words: cellulose materials, Miscanthus, enzymatic hydrolysis, biomass, genotyping.

Федеральное государственное унитарное предприятие «ПОЧТА РОССИИ»

Бланк заказа периодических изданий Ф. СП-1

АБОНЕМЕНТ на газету / **журнал** **4 2 1 5 3**

(индекс издания)

Вавиловский журнал

(наименование издания)

генетики и селекции

Количество Комплектов

на 20___год по месяцам

1	2	3	4	5	6	7	8	9	10	11	12

Куда

(почтовый индекс) (адрес)

Кому

(фамилия, инициалы)

---------- Линия отреза ----------

ПВ | Место | Литер — **ДОСТАВОЧНАЯ КАРТОЧКА** **4 2 1 5 3**

(индекс издания)

Газету
На Журнал **Вавиловский журнал генетики и селекции**

(наименование издания)

Стоимость: подписки руб. — переадресовки руб. — Кол-во комплектов

на 20___год по месяцам

1	2	3	4	5	6	7	8	9	10	11	12

почтовый индекс

код улицы

дом корпус

город

Село

область

район

улица

квартира

Фамилия И.О.

Отредактировано и подготовлено к печати
в редакционно-издательском отделе ИЦиГ СО РАН

Редакторы: А.А. Ончукова, И.Ю. Ануфриева
Дизайн: А.В. Харкевич
Компьютерная графика: А.В. Харкевич, Т.Б. Коняхина
Компьютерная верстка: Т.Б. Коняхина, Н.С. Глазкова

Подписано в печать 27.10.2013 г.
Формат бумаги 60×84 1/8. Усл.-печ. л. 23,4. Уч.-изд. л. 22,5
Тираж 250. Заказ 301

Отпечатано в типографии ФГУП «Издательство СО РАН»
630090, Новосибирск, Морской проспект, 2