

На правах рукописи

Ковалева Галина Юрьевна

АНАЛИЗ РЕГУЛЯЦИИ ГЕНОВ
МЕТАБОЛИЗМА И ТРАНСПОРТА МЕТИОНИНА И ЛЕЙЦИНА
МЕТОДАМИ СРАВНИТЕЛЬНОЙ ГЕНОМИКИ

03.00.28 – биоинформатика

Автореферат
диссертации на соискание ученой степени
кандидата биологических наук

Москва – 2008

Работа выполнена на факультете биоинженерии и биоинформатики Московского государственного университета имени М.В. Ломоносова и в Учебно-научном центре „Биоинформатика“ Учреждения Российской академии наук Института проблем передачи информации им. А.А. Харкевича РАН.

Научный руководитель: кандидат физико-математических наук,
доктор биологических наук, профессор
Михаил Сергеевич Гельфанд

Официальные оппоненты: доктор биологических наук, профессор
Владимир Васильевич Поройков

Государственное Учреждение Научно-
Исследовательский институт
биомедицинской химии имени В.Н.
Ореховича РАН;

кандидат физико-математических наук
Олег Владимирович Демин

Институт физико-химической биологии
им. А.Н. Белозерского МГУ

Ведущая организация: Учреждение Российской академии наук
Институт теоретической и
экспериментальной биофизики РАН

Защита диссертации состоится ____ 2008 года в ____ часов на заседании диссертационного совета Д.002.077.02 при Учреждении Российской академии наук Институте проблем передачи информации им. А.А. Харкевича РАН по адресу: 127994, г. Москва, ГСП-4, Большой Каретный переулок, д. 19, стр. 1.

С диссертацией можно ознакомиться в библиотеке Учреждения Российской академии наук Института проблем передачи информации им. А.А. Харкевича РАН

Автореферат разослан ____ 2008 года

Ученый секретарь диссертационного совета
доктор биологических наук, профессор

Рожкова Г.И.

Общая характеристика работы

Актуальность темы

Аминокислоты являются структурным компонентом всех белков, и, таким образом, представляют собой важнейшие компоненты каждой живой клетки. Способностью к биосинтезу всех аминокислот, входящих в полипептидные цепи, обладают только некоторые микроорганизмы и растения, тогда как человек и животные вынуждены часть необходимых аминокислот получать с пищей, — это так называемые незаменимые аминокислоты, в которых относятся изолейцин, лейцин, лизин, метионин, фенилаланин, треонин, триптофан и валин. Для производства пищевых и кормовых добавок, содержащих незаменимые аминокислоты, необходимо создание систем биосинтеза этих аминокислот в промышленных масштабах, что, в свою очередь, требует детального и тщательного исследования устройства и регуляции метаболических путей, обеспечивающих биосинтез незаменимых аминокислот в различных микроорганизмах-продуцентах этих аминокислот.

Анализ транскрипционной регуляции экспрессии генов является важным разделом компьютерной генетики и системной биологии. Это связано с тем, что в настоящее время сравнительно легко производится определение последовательностей геномов, и темпы секвенирования значительно опережают темпы получения данных из биологических экспериментов. В настоящий момент уже более 700 геномных последовательностей определены полностью, и более 2000 геномов — частично. С ростом количества секвенированных геномных последовательностей появлялись все новые способы компьютерного анализа геномов, которые произвели переворот в области функциональной аннотации генов. Если ранее предсказание потенциальной функции белка производилось переносом экспериментально установленной функции белка и основывалось исключительно на сходстве аминокислотных последовательностей, то в настоящее время в сравнительной геномике большое значение имеют анализ сцепления генов на хромосоме, случаев слияния генов, профилей встречаемости генов в полных геномах, а также анализ общих регуляторных сигналов. Одновременное использование всех этих методик позволяет обнаружить функциональную связь между белками, участвующими в одном метаболическом пути. Метаболическая реконструкция позволяет как обнаружить новые аспекты исследуемого пути в хорошо изученных организмах, так и описать метаболический потенциал ранее не охарактеризованных организмов.

В данной работе новейшие методы сравнительной геномики были использованы для анализа регуляции транскрипции генов метаболизма и транспорта двух незаменимых аминокислот, метионина и лейцина, а также цистеина, в различных микроорганизмах.

Цель исследования

Целью исследования являлось изучение генов биосинтеза и транспорта лейцина, метионина и цистеина в геномах дрожжей и бактерий методами сравнительной геномики.

Задачи исследования

1. Оценить уровень консервативности сайтов связывания регуляторов транскрипции, контролирующих биосинтез метионина и лейцина, в геномах дрожжей и сравнить его с уровнем консервативности прилегающих некодирующих областей.
2. Выявить ген, отвечающий за транспорт промежуточного продукта пути биосинтеза лейцина, альфа-изопропилмалата, из матрикса митохондрий в цитозоль у дрожжей.
3. Описать структуру всех генетических локусов, содержащих гены метаболизма и транспорта метионина в геномах коринебактерий.
4. Выявить в геномах коринебактерий новые гены, участвующие в метаболизме и транспорте метионина.
5. Описать структуру всех генетических локусов, содержащих гены биосинтеза метионина и цистеина в геномах стрептококков.
6. Идентифицировать потенциальные регуляторные сайты связывания транскрипционных регуляторов, контролирующих путь биосинтеза метионина и цистеина.
7. С использованием методов сравнительной геномики выявить в геномах стрептококков новые гены, участвующие в биосинтезе и транспорте метионина и цистеина.

Научная новизна и практическое значение

В представленной работе впервые получены следующие результаты:

1. Впервые был проведен детальный анализ уровней консервативности некодирующих областей геномов дрожжей и уровней консервативности сайтов связывания транскрипционных регуляторов и прилегающих к ним областей.
2. Проведен систематический анализ экспрессионных данных и данных сравнительной геномики, которые вместе позволили предсказать транспортер альфа-изопропилмалата, промежуточного продукта биосинтеза лейцина.
3. В геномах коринебактерий описан регулон McbR, включающий в себя гены метаболизма и транспорта метионина, предсказано несколько новых потенциальных членов данного регулона.
4. В геномах стрептококков проведено описание структурных генов, ответственных за биосинтез аминокислот метионина и цистеина.
5. Проведен систематический анализ регуляторных областей структурных генов пути биосинтеза метионина и цистеина в стрептококках, позволивший предсказать потенциальные сайты связывания двух транскрипционных регуляторов, и установить соответствие выявленных сигналов каждому из регуляторов.

Апробация работы

Материалы исследований по теме диссертации были представлены на международных конференциях: XI Международная научная конференция студентов, аспирантов и молодых ученых “Ломоносов-2004” (Москва), BGRS’2004 (Bioinformatics of Genome Regulation and Structure, Новосибирск), BITS’2005 (Bioinformatics Italian Society, Милан, Италия), XII Международная научная конференция студентов, аспирантов и молодых ученых “Ломоносов-2005” (Москва), MCCMB’2005 (2nd Moscow Conference on Computational Molecular Biology, Москва), ECCB’2005 (4th European Conference on Computational Biology, Мадрид, Испания), BGRS’2006 (Bioinformatics of Genome Regulation and Structure, Новосибирск), ECCB’2006 (5th European Conference on Computational Biology, Эйлат, Израиль), ISMB/ECCB’07 (15th International Conference on Intelligent Systems for Molecular Biology and 6th European Conference on Computational Biology, Вена, Австрия), MCCMB’2007 (3rd Moscow Conference on Computational Molecular Biology, Москва), а также на конференции молодых ученых ИППИ РАН ИТиС’2007 (Информационные Технологии и Системы, Звенигород), межлабораторном семинаре ИППИ РАН, на научных семинарах в Учебно-научном центре «Биоинформатика» ИППИ РАН.

Объем и структура диссертации

Диссертационная работа изложена на 107 страницах машинописного текста и состоит из введения и трех разделов. Первый раздел включает обзор литературы по теме диссертации. Второй раздел содержит описание использованных в работе программ и алгоритмов. Третий раздел содержит описание проведенных в работе исследований и состоит из трех частей, каждая из которых включает изложение полученных в работе результатов и их обсуждение.

- **Первая часть** посвящена сравнительному анализу некодирующих последовательностей геномов дрожжей и анализу консервативности регуляторных сайтов связывания факторов транскрипции, обеспечивающих контроль уровня экспрессии генов метаболизма метионина и лейцина.

- **Вторая часть** посвящена применению методов сравнительной геномики при анализе регуляции генов метаболизма и транспорта метионина в геномах коринебактерий.

- **Третья часть** посвящена исследованию регуляции генов метаболизма и транспорта метионина и цистеина в геномах стрептококков.

Список цитируемой литературы, приведенный в конце диссертации, содержит 156 наименований. Работа содержит 41 рисунок и 5 таблиц.

Содержание работы

Первый раздел. Обзор литературы.

Содержит обзор литературных данных по рассмотренным в диссертации проблемам. В данной главе вводятся основные понятия и определения, использующиеся в сравнительной геномике, а также приводятся описания

метаболических путей, исследовавшихся в диссертационной работе, и известных данных о регуляции этих метаболических путей.

Второй раздел. Материалы и методы.

Данный раздел содержит описание основных методов, алгоритмов и программ, использованных в работе.

В начале раздела описывается процедура определения принадлежности гена регулону методом проверки соответствия, согласно которому ген считается членом регулона, если в его регуляторной области обнаружен потенциальный сайт, который сохраняется перед ортологичными генами в родственных геномах. Поиск потенциальных сайтов связывания транскрипционных регуляторов проводился с использованием метода построения матриц позиционных весов.

Третий раздел. Результаты и обсуждение.

Исследование консервативности сайтов связывания факторов транскрипции, контролирующих биосинтез метионина и лейцина в дрожжах.

В данном разделе описано применение методов сравнительной геномики при работе с геномами дрожжей. Основным подходом, использованным в данной части работы, был метод филогенетического футпринтинга, согласно которому функционально значимые области генома эволюционируют медленнее остальных и, соответственно, более консервативны.

На первом этапе работы с использованием программы MLAGAN были построены множественные полногеномные выравнивания 5 геномов рода *Saccharomyces*: *S.cerevisiae*, *S.paradoxus*, *S. mikatae*, *S.bayanus* и *S.kudriavzevii*. Анализ 5'-нетранслируемых областей генов в данных выравниваниях выявил непосредственно перед стартовым кодоном область относительно высокой консервативности, которая уже была описана ранее¹, и, вероятнее всего, объясняется накоплением в этом районе сайтов связывания факторов инициации трансляции. Кроме того, на большем удалении от точки старта трансляции (в 5'-направлении) была обнаружена область более высокой консервативности, начинающаяся на расстоянии 70 п.н., достигающая максимальных значений на расстоянии около 120 п.н. и возвращающаяся к фоновому уровню на расстоянии 220 п.н. относительно точки старта трансляции. Наблюдаемая область повышенной консервативности (-220 -70 п.н.), вероятнее всего, связана с расположением промоторов в этом районе.

Результаты, полученные при анализе 3'-некодирующих областей полученных множественных выравниваний, частично противоречат более ранним наблюдениям¹. Показанное ранее падение уровня консервативности, как выяснилось, сопровождается последующим заметным подъемом консервативности в районе 25-155 п.н. после стоп-кодона. Предположительно, обнаруженная область повышенной консервативности связана с обеспечением стабильности молекулы мРНК и расположением сайтов внутриклеточной локализации, сконцентрированных на расстоянии 50-100 п.н. после стоп-кодона.

¹ Shabalina S.A., Ogurtsov A.Y., Rogozin I.B., Koonin E.V., Lipman D.J. Comparative analysis of orthologous eukaryotic mRNAs: potential hidden functional signals // Nucleic Acids Res. — 2004 — vol. 32, no. 5. — Pp.1774-1782.

На следующем этапе работы проводился анализ уровней консервативности сайтов связывания факторов транскрипции в геномах дрожжей. Анализ проводился на множественных полногеномных выравниваниях пяти геномов дрожжей, сконструированных при помощи программы MLAGAN. В работу были включены сайты связывания факторов транскрипции, представленные в базе данных SCPD, длиной не менее 5 п.н.

Как и ожидалось, уровень консервативности сайтов связывания был заметно выше уровня консервативности прилегающих нефункциональных участков нетранслируемых областей. Кроме того, уровень консервативности был максимален для центральных позиций сайтов связывания факторов транскрипции и постепенно снижался по мере удаления от центра сайта. Однако, уровень консервативности остается повышенным в сравнении с фоновым не только в пределах области, описанной как функциональный сайт, но и далеко за пределами сайта связывания. В целом, область повышенной консервативности, включающая в себя сайт связывания фактора транскрипции, занимает область протяженностью около 100 п.н.

Для геномов эукариот характерны очень короткие сайты связывания факторов транскрипции, и одним из механизмов обеспечения специфичной регуляции экспрессии генов является кластеризация сайтов, которая обеспечивает кооперативное связывание белков-регуляторов с цепью ДНК. Согласно наиболее вероятной гипотезе, повышенная консервативность областей, прилегающих к сайту связывания, может быть связана с наличием рядом с исследуемым сайтом других регуляторных последовательностей —сайтов связывания факторов транскрипции или промоторов.

Обнаруженная повышенная консервативность областей, прилегающих к сайтам связывания факторов транскрипции, может объяснять наличие расхождений в наборах сигналов, предсказанных разными алгоритмами на множественных выравниваниях геномов дрожжей. Для проверки данной гипотезы на следующем этапе работы был проведен детальный анализ уровней консервативности сайтов связывания транскрипционных активаторов путей биосинтеза двух аминокислот — метионина и лейцина.

Исследование проводилось на множественных выравниваниях ортологичных регуляторных последовательностей структурных генов пути биосинтеза метионина и лейцина. Сайт связывания считался консервативным, если он выравнивался с потенциальными сайтами связывания перед ортологичными генами. Высокая степень родства последовательностей позволила применить упрощенную количественную оценку при оценке уровней консервативности сигналов. При абсолютной консервативности сайта уровень его консервативности принимался равным единице. Если же потенциальный сайт не абсолютно консервативен, но тем не менее выравнивается с сайтом в геноме *S. cerevisiae*, консервативность данного сайта принималась равной 0,5. Для неконсервативных сайтов консервативность считалась равной нулю.

Для каждого из исследуемых геномов оценочные уровни консервативности были просуммированы отдельно для каждого регулятора. Пути биосинтеза

метионина и лейцина в дрожжах находятся под контролем следующих регуляторов/регуляторных комплексов: Gcn4p — глобальный регулятор биосинтеза аминокислот в дрожжах, в контроле биосинтеза метионина также принимают участие регуляторные комплексы Met31/Met32 и Cbf1/Met4/Met28, а в контроле биосинтеза лейцина — Leu3p². Для каждого из регуляторов была сконструирована матрица позиционных весов с использованием экспериментально известных последовательностей сайтов связывания данного регулятора. Все предсказанные сайты разделялись на три группы: «известные» (экспериментально определенные сайты), «сильные потенциальные» (сайты, вес которых превышает вес большинства экспериментально определенных сайтов) и «слабые потенциальные» (сайты с весом выше порогового, но ниже, чем в предыдущей группе).

Уровни консервативности всех факторов транскрипции в исследуемых геномах показаны на рис. 1.

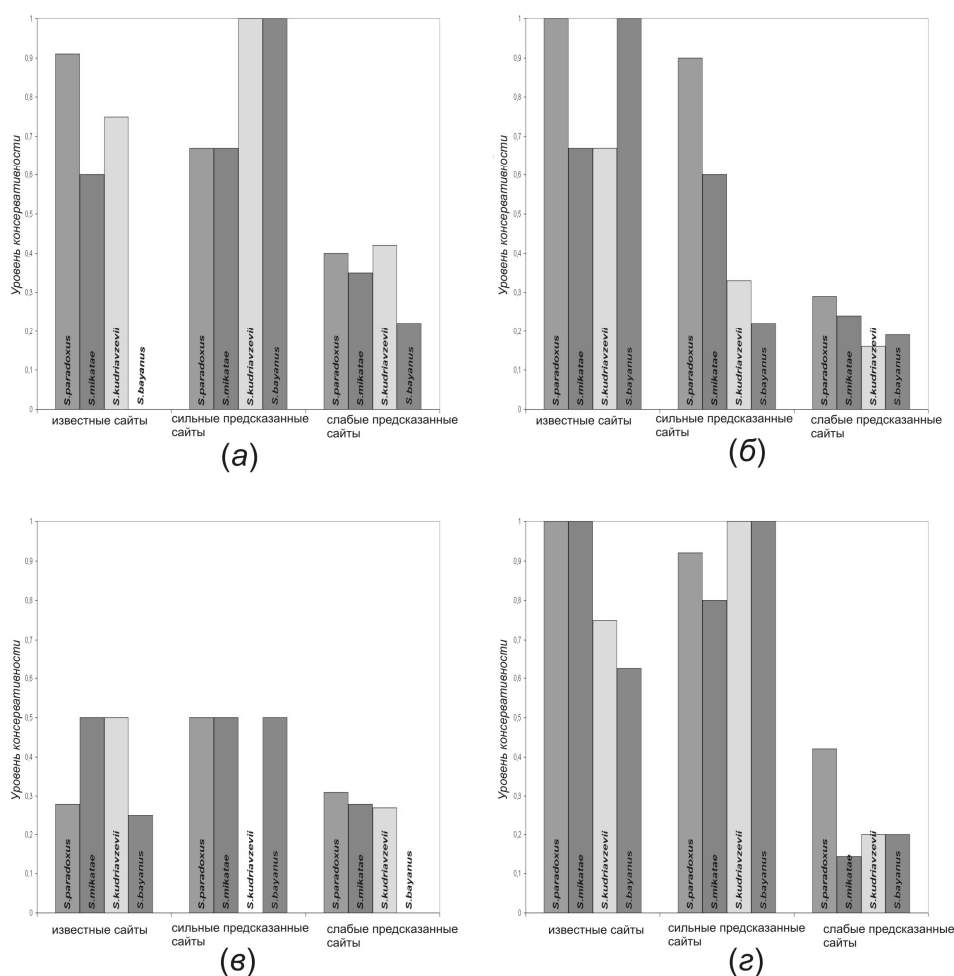


Рис. 1. Усредненные уровни консервативности сайтов связывания четырех регуляторов. Уровни консервативности показаны независимо для каждой группы сайтов для каждого регулятора. А. Сайты связывания Gcn4p. Б. Сайты связывания Cbf1-комплекса. В. Сайты связывания регулятора(ов) Met31/Met32. Г. Сайты связывания Leu3p.

² Thomas D., Surdin-Kerjan Y. Metabolism of sulfur amino acids in *Saccharomyces cerevisiae* // Microbiol. Mol. Biol. Rev., — 1997 — vol. 61, no. 4. — Pp.503-532.

Видно, что в большинстве случаев уровни консервативности для групп «известных» и «сильных потенциальных» сайтов сходны для каждого регулятора. Однако следует заметить, что даже сайты с максимальным весом не всегда консервативны во всех исследуемых геномах.

Таким образом, сайты связывания эукариотических факторов транскрипции не строго консервативны и при работе с геномами дрожжей стандартные методы филогенетического футпринтинга требуют введения дополнительных поправок. Несмотря на то, что уровни консервативности известных и сильных предсказанных сайтов сходны, даже сайты с наибольшим весом (включая и экспериментально показанные) могут быть неконсервативными даже в ближайших геномах.

В метаболическом пути биосинтеза лейцина в дрожжах были известны все структурные гены пути, за исключением переносчика альфа-изопропилмалата из матрикса митохондрий в цитозоль³. В данной работе использование метода матриц позиционных весов позволило определить продукт гена YOR271c как потенциальный переносчик альфа-изопропилмалата, промежуточного продукта биосинтеза лейцина. Предсказание функции белка основывалось на следующих наблюдениях:

- для соответствующего белка ранее была показана митохондриальная локализация⁴;
- белок YOR271c содержит несколько предсказанных трансмембранных сегментов и, следовательно, скорее всего является транспортером;
- потенциальный сайт связывания регулятора биосинтеза лейцина, Leu3p, строго консервативен во всех исследованных геномах и, кроме этого, полностью идентичен консенсусу сайта связывания;
- в эксперименте по иммунопреципитации хроматина было показано, что регулятор пути биосинтеза лейцина, Leu3p, связывается с промоторной областью гена YOR271c⁵, и, следовательно, вероятно контролирует экспрессию данного гена;
- филогенетическое дерево потенциального транспортера свидетельствует о том, что его ближайшими гомологами с известной функцией являются сидерофлексины и транспортер трикарбоксилатов из *Rattus norvegicus*.

Таким образом, сравнительно-геномный анализ показал, что продукт гена YOR271c, по всей видимости, принимает участие в пути биосинтеза лейцина, является трансмембранным белком, заякоренным в мембране митохондрий. Эти факты позволяют приписать белку YOR271c функцию потенциального

³ Kohlhaw G.B. Leucine biosynthesis in fungi: entering metabolism through the back door // Microbiol. Mol. Biol. Rev. — 2003 — vol.67, no. 1. — Pp. 1-15.

⁴ Huh W.K., Falvo J.V., Gerke L.C., Carroll A.S., Howson R.W., Weissman J.S., O'Shea E.K. Global analysis of protein localization in budding yeast // Nature — 2003 — vol.425, no. 6959. — Pp. 686-691.

⁵ Lee T.I., Rinaldi N.J., Robert F., Odom D.T., Bar-Joseph Z., Gerber G.K., Hannett N.M., Harbison C.T., Thompson C.M., Simon I., Zeitlinger J., Jennings E.G., Murray H.L., Gordon D.B., Ren B., Wyrick J.J., Tagne J.B., Volkert T.L., Fraenkel E., Gifford D.K., Young R.A. Transcriptional regulatory networks in *Saccharomyces cerevisiae* // Science — 2002 — vol.298, no. 5594. — Pp.799-804.

переносчика альфа-изопропилмалата, промежуточного продукта биосинтеза лейцина.

В целом, проведенный в данной части работы детальный анализ регуляторных областей генов и индивидуальных сайтов связывания факторов транскрипции выявил повышенные уровни консервативности некодирующих областей ДНК, принимающих участие в регуляции экспрессии генов. Как и ожидалось, степень консервативности всех областей генома падает с увеличением филогенетических расстояний между исследуемыми видами. Филогенетический футпринтинг сайтов связывания факторов транскрипции был наиболее наглядным на промежуточных филогенетических расстояниях. Было также показано, что повышенная консервативность характерна не только непосредственно для сайта связывания регуляторного белка, но характеризует и области, прилегающие к сайту связывания. Такое распространение области повышенной консервативности сайтов связывания регуляторов за пределы границ сайтов связывания может указывать на возможную кластеризацию функциональных регуляторных последовательностей, характерную для эукариот. Кроме того, был идентифицирован ген, кодирующий потенциальный митохондриальный переносчик альфа-изопропилмалата, промежуточного продукта биосинтеза лейцина.

Исследование транскрипционной регуляции генов биосинтеза метионина в коринебактериях.

В этом разделе описано применение методов современной сравнительной геномики при исследовании пути биосинтеза метионина в геномах коринебактерий.

В геноме *C. glutamicum* был известен транскрипционный регулятор пути биосинтеза метионина, McbR⁶. Сконструированное филогенетическое дерево ортологов McbR свидетельствует о том, что его ортологи имеются в геномах рода *Corynebacterium* (*C. glutamicum*, *C. diphtheriae*, *C. efficiens* и *C. jeikeium*), а также в других геномах порядка Actinomycetales, таких как *Nocardia farcinica*, *Leifsonia xyli* и *Streptomyces coelicolor*. Последующие этапы работы проводились только на этих геномах.

Гены метаболического пути и их организация в опероны у *C. glutamicum* были изучены ранее⁶, и на первом этапе работы был предпринят поиск ортологов структурных генов биосинтеза метионина в остальных исследуемых геномах. Несмотря на филогенетическую близость анализируемых геномов, анализ предсказанных оперонных структур выявил значительное количество различных эволюционных событий, которые можно разделить на следующие типы:

- *неортологичные замещения* — ген треонинсинтазы, *thrC*, в геномах *C. glutamicum*, *C. diphtheriae* и *C. efficiens*;

⁶ Rey D.A., Nentwich S.S., Koch D.J., Rückert C., Pühler A., Tauch A., Kalinowski J. The McbR repressor modulated by the effector substance S-adenosylhomocysteine controls directly the transcription of a regulon involved in sulphur metabolism of *Corynebacterium glutamicum* ATCC 13032 // Mol. Microbiol. — 2005 — vol.56, no. 4. — Pp.871-887.

- *оперонные перестройки* — описаны для большей части оперонов структурных генов пути биосинтеза метионина, например, для оперонов *thrABC*, *metXY* и оперона *cg0735-cg0737* *C. glutamicum*;

- *паралоги и дубликации* — дубликация гена алкансульфонат-монооксигеназы, *ssuD*, в геноме *C. glutamicum*; дубликация гена сульфатредуктазы, *fpr*, в геномах *C. glutamicum* и *C. efficiens*;

- *горизонтальный перенос* — появление двух пар гомологов для пары генов *cysN-cysD* в геноме *Nocardia farcinica* в результате горизонтального переноса от других бактерий порядка Actinomycetales (вероятнее всего, от бактерий рода *Mycobacterium*) и в результате горизонтального переноса из геномов протеобактерий;

- *утрата ортологов* во всех исследованных организмах была описана для следующих генов *C. glutamicum*: для регулятора генов утилизации сульфонов и сульфатных эфиров, *ssuR*, для гена флавин-редуктазы *ssuI*, для генов *seuBC* и *ssuABC*, отвечающих за утилизацию сульфатных эфиров и сульфонов, соответственно, а также для гена, кодирующего белок с неизвестной функцией, *cg3372*; кроме того, описаны случаи утраты ортологов генов *C. glutamicum* в более филогенетически далеких геномах и более частные случаи утраты ортологов лишь в некоторых из исследованных геномов;

- *полная консервативность оперона* во всех изученных геномах была показана только для двух моноцистронных оперонов.

На следующем этапе работы было подтверждено известное предсказание потенциального мотива связывания регулятора McbR⁶. На основании предсказанных и экспериментально подтвержденных сигналов была составлена обучающая выборка и построена позиционная весовая матрица для поиска новых сайтов связывания регулятора McbR. Лого всех найденных сайтов McbR в геномах рода *Corynebacterium* представлено на рис. 2.

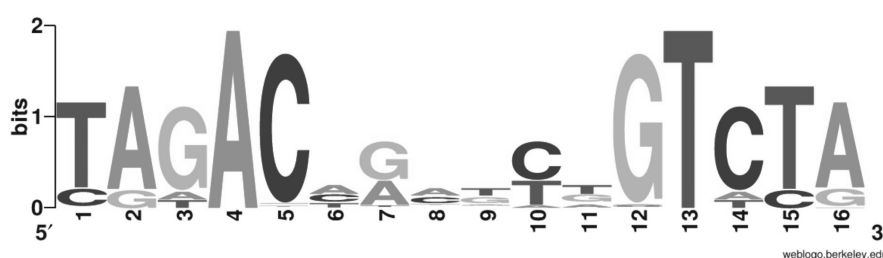


Рис. 2. Лого всех потенциальных сайтов связывания регулятора McbR.

Исследование регуляторных сигналов связывания McbR позволило предсказать потенциальные сайты связывания перед ортологами большей части регулируемых генов *C. glutamicum* в геномах рода *Corynebacterium*. Так, в частности, сайт сохраняется перед всеми найденными в этих геномах ортологами генов *metK*, *metH*, *metE*, *metB*, *cysK*, *cg3132* и перед генами дивергона *cg2674-cg2680*. Однако иногда даже в близких геномах рода *Corynebacterium* регуляторные сигналы отсутствовали перед ортологичными генами. Были также описаны ситуации разделения единого регулируемого оперона на два независимо регулирующихся

ортологичных оперона (оперон *metXY* в геноме *C. glutamicum* и дивергон утилизации сульфата *cysJIXHDNZY* в геноме *C. jeikeium*).

В геноме *C. glutamicum*, помимо уже известных сигналов, был описан новый потенциальный сайт связывания McbR перед геном *cg1739*, кодирующим белок с консервативным доменом глутамин-амидотрансферазы. Ортологи данного белка представлены во всех геномах рода *Corynebacterium* и отсутствуют в филогенетически более далеких организмах. Перед всеми ортологами этого белка имеется потенциальный сайт связывания McbR. Таким образом, согласно методу процедуры проверки соответствия, ген *cg1739* является новым потенциальным членом регулона McbR. Косвенным экспериментальным подтверждением этого предсказания является то, что в делеционном мутанте *C. glutamicum* по гену *mcbR* происходит активация экспрессии гена *cg1739*⁶.

В целом, исследование потенциальных сигналов связывания McbR в геномах рода *Corynebacterium* позволило заключить, что большинство генов регулона McbR сохраняет регуляцию со стороны этого репрессора. Применение современных методов сравнительной геномики позволило выявить два новых потенциальных члена регулона McbR. Одним из них является ген *cg1739* *C. glutamicum*, роль и потенциальную функцию которого в пути биосинтеза метионина достоверно предсказать было невозможно. Вторым новым потенциальным членом регулона McbR является ген *CE2636p* *C. efficiens* и его ортологи в *C. glutamicum* и *C. diphtheriae*. Кодированный им белок имеет значительную степень сходства с 5'-метилтиоаденозин-нуклеозидазами, которые обеспечивают восстановление пула метионина из побочного продукта биосинтеза полиаминов. Таким образом, в геномах рода *Corynebacterium* увеличение пула метионина за счет расщепления 5'-метилтиоаденозина также регулируется фактором McbR, по крайней мере, в геномах *C. diphtheriae* и *C. efficiens*.

Исследование транскрипционной регуляции биосинтеза метионина в стрептококках.

Филогенетическое дерево регуляторов LysR семейства (рис. 3) свидетельствует о том, что ортологи обоих регуляторов, MtaR и CmbR, принимающих участие в регуляции биосинтеза метионина и цистеина в геномах стрептококков, формируют две независимые ветви, каждая из которых содержит белки из всех доступных геномов стрептококков.

На первом этапе работы для всех генов, принимающих участие в биосинтезе метионина и цистеина, были определены ортологи во всех геномах стрептококков. В большинстве исследованных геномов сохраняется большая часть основных генов биосинтеза метионина и остается консервативной оперонная структура генов биосинтеза метионина и цистеина, что объясняется филогенетической близостью этих геномов. Отдельного внимания заслуживает только оперон *metEF*, детальное исследование которого показало, что, вероятнее всего, оперон *metEF* в геномах *S. agalactiae*, *S. mutans* и *S. sobrinus* появился в результате неортологичного замещения.

При поиске потенциальных сигналов связывания регуляторов пути биосинтеза метионина и цистеина использовались следующие известные характеристики сайтов связывания LysR семейства⁷:

- общий мотив связывания вида T-N₁₁-A;
- центр сайта связывания располагается на расстоянии -65 п.н. относительно точки старта транскрипции.

Так как точка старта транскрипции определяется расположением промотора в регуляторной области, на основании известных последовательностей промоторов *Bacillus subtilis* была сконструирована матрица позиционных весов для предсказания потенциальных промоторов в регуляторных областях генов пути биосинтеза метионина и цистеина в стрептококках.

Предсказанные промоторы в большей части случаев имели стандартную структуру в виде двух боксов (ТАТААТ и ТТGАСА), разделенных спейсером длиной 17 п.н. В более редких случаях наблюдалась частичная, а иногда и полная утрата -35 бокса, который в таких случаях всегда был скомпенсирован так называемым удлинненным -10 боксом (TG-ТАТААТ).

Сайты связывания факторов транскрипции являются функционально значимыми последовательностями и, следовательно, консервативны. По этой причине выявление потенциальных сайтов связывания транскрипционных регуляторов проводилось посредством анализа множественных выравниваний ортологичных регуляторных областей структурных генов биосинтеза метионина.

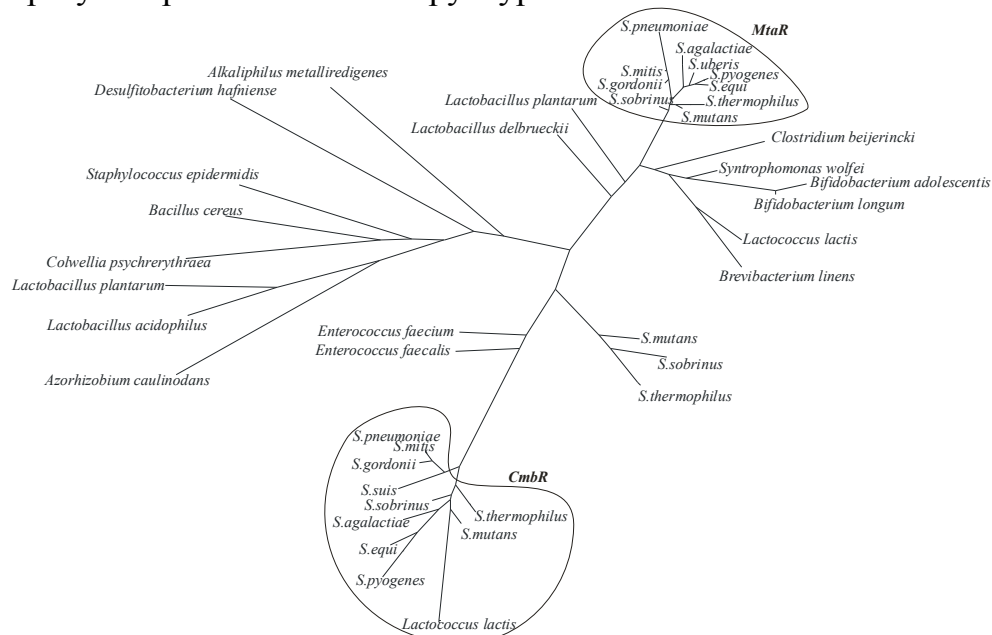


Рис.3. Филогенетическое дерево гомологов белка MtaR *S. agalactiae*. Выделены ветви ортологов MtaR и CmbR.

Выравнивания регуляторных областей позволили идентифицировать перед большей частью ортологичных генов несколько областей повышенной консервативности, одна из которых соответствует предсказанному промотору.

⁷ Schell M.A. Molecular biology of the LysR family of transcriptional regulators // Annu. Rev. Microbiol. — 1993 — vol.47. — Pp. 597-626.

Помимо потенциального промотора, в выравниваниях часто обнаруживался второй островок консервативности, вероятнее всего соответствующий сайту связывания транскрипционного регулятора семейства LysR, так как в большинстве случаев эти островки консервативности располагались на расстоянии -65 п.н. от точки старта транскрипции и имели структуру вида T-N₁₁-A.

В целом такой анализ, проведенный на множественных выравниваниях регуляторных областей генов биосинтеза метионина и цистеина, позволил выявить два различных потенциальных мотива связывания регуляторов семейства LysR. Консенсусные записи обнаруженных мотивов выглядят как TATAG-(n)₇-CTATA и TGATA-N₉-TATCA-N₂₋₄-TGATA. Лого найденных мотивов представлено на рис. 4. Нужно отметить, что в некоторых случаях сигнал располагался не на характерном для сигналов регуляторов LysR семействе расстоянии от точки старта транскрипции, однако во всех случаях смещение сигнала всегда происходило на расстояние, кратное шагу спирали ДНК. Таким образом, молекула регулятора, связанная с регуляторным сигналом, остается на той же стороне спирали ДНК относительно комплекса РНК-полимеразы, который формируется в промоторной области.

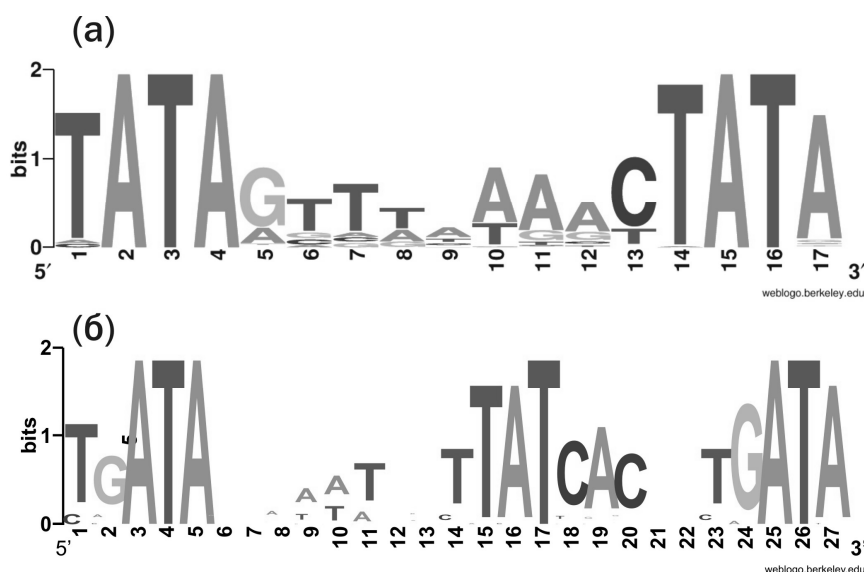


Рис. 4. Лого потенциальных сайтов связывания регуляторов MtaR (a) и CmbR (б) в геномах стрептококков.

Сигнал вида TATAG-(n)₇-CTATA был найден перед большей частью генов метаболического пути биосинтеза метионина и цистеина, тогда как сигнал структуры TGATA-N₉-TATCA-N₂₋₄-TGATA был найден перед генами, непосредственно связанными с метаболизмом цистеина. На этом основании было сделано предположение о том, что сигнал структуры TGATA-N₉-TATCA-N₂₋₄-TGATA является сайтом связывания ортологов регулятора CmbR, который чувствителен к концентрации цистеина и контролирует его биосинтез в *L. Lactis*⁸. Соответственно, сигнал вида TATAG-(n)₇-CTATA является сайтом связывания

⁸ Sperandio B., Polard P., Ehrlich D.S., Renault P., Guédon E. Sulfur amino acid metabolism and its control in *Lactococcus lactis* IL1403 // J. Bacteriol. — 2005 — vol.187, no. 11. — Pp. 3762-3778.

ортологов регулятора MtaR, контролирующего гены метаболизма и транспорта метионина^{9,10}. В дальнейшем, сигнал вида ТАТАG-(n)₇-СТАТА будет называться потенциальным сигналом связывания MtaR, а мотив структуры TGATA-N₉-ТАТСА-N₂₋₄-TGATA будет называться мотивом связывания регулятора CmbR.

На основании выборки сайтов связывания CmbR была сконструирована матрица позиционных весов для предсказания новых потенциальных сигналов связывания данного регулятора. При обнаружении потенциальных предсказанных сайтов CmbR перед несколькими ортологичными генами, для соответствующих регуляторных областей предсказывались потенциальные промоторы и строилось множественное выравнивание. Если сайт связывания перед ортологичными генами был консервативен, то, согласно методу проверки соответствия, эти гены являются членами исследуемого регулона. Такой анализ позволил выявить несколько новых потенциальных членов регулона CmbR:

- ген, кодирующий пермеазу аминокислот неизвестной специфичности. Этому гену было предложено тривиальное имя *cctY* (от Candidate Cysteine Transporter Y). Этот ген, по всей видимости, является первым геном бицистронного оперона, включающего также ген *mtuM*, кодирующий гомоцистеин S-метилтрансферазу, также принимающую участие в биосинтезе метионина и цистеина. Гены *cctY* и *mtuM* образуют единый оперон, поскольку длина межгенного расстояния между ними во всех геномах не превышает 13 п.н.

- ген, потенциально кодирующий лиганд-связывающий компонент ABC транспортера аминокислот неизвестной специфичности. Этому гену было предложено тривиальное имя *cctW* (от Candidate Cysteine Transporter W). Этот ген расположен перед опероном ABC транспортера *metNPQ*, отвечающего за транспорт метионина внутрь клетки, но, по всей видимости, транскрибируется моноцистронно, так как межгенное расстояние между *cctW* и *metNPQ* во всех геномах превышает 140 п.н. Нужно отметить, что в геноме *S. suis* перед опероном *metNPQ* располагается не ортолог *cctW*, а ген, потенциально кодирующий глутамин амидотрансферазу. Тем не менее, перед этим геном был идентифицирован сигнал связывания CmbR, расположенный на правильном расстоянии относительно точки старта транскрипции.

Кроме того, во время детального анализа множественных выравниваний регуляторных областей было отмечено, что в некоторых случаях предсказанные потенциальные сайты связывания регуляторов MtaR и CmbR могут пересекаться (это, в частности, наблюдается в случае генов *metY* и *metQ2*). Такие случаи, когда потенциальные сайты связывания обоих регуляторов накладываются друг на друга, вероятно отражают ситуации потенциальной смены регулятора, контролирующего экспрессию данного гена.

⁹ Rodionov D.A., Vitreschak A.G., Mironov A.A., Gelfand M.S. Comparative genomics of the methionine metabolism in Gram-positive bacteria: a variety of regulatory systems // Nucleic Acids Res. — 2004 — vol.32. no. 11. — Pp. 3340-3353.

¹⁰ Shelper D., Rajagopal L., Harris T.O., Rubens C.E. MtaR, a regulator of methionine transport, is critical for survival of group B streptococcus in vivo. // J. Bacteriol. — 2003 — vol.185, no. 22. — Pp. 6592-6599.

Таким образом, в данной части работы было показано, что структурные гены метаболического пути биосинтеза метионина и цистеина в геномах стрептококков контролируются двумя транскрипционными регуляторами семейства LysR. Для каждого из регуляторов был обнаружен потенциальный сигнал связывания. Были предсказаны новые члены регулона CmbR, принимающие участие в метаболизме цистеина в стрептококках. Кроме того, были обнаружены и случаи потенциальной смены транскрипционного регулятора ортологичных генов.

По всей видимости, в предковых геномах Грам-положительных бактерий регуляция пути биосинтеза метионина осуществлялась при помощи S-боксов, регуляторных вторичных структур РНК^{9,11}. Затем, в процессе эволюции, S-боксы были утрачены в последнем общем предковом геноме для групп *Lactobacillales* и *Streptococcales*. В лактобациллах произошла экспансия регулона Т-боксов⁹, в то время как у стрептококков регуляция осуществляется на этапе инициации транскрипции. Эта регуляция осуществляется двумя гомологичными транскрипционными регуляторами, которые были найдены во всех исследованных полных геномах стрептококков. Мотивы связывания и потенциальные сайты предсказаны для обоих регуляторов, а их распределение позволило приписать определенную функцию каждому из регуляторов. Таким образом, CmbR был определен как регулятор метаболизма цистеина, а MtaR — как регулятор метаболизма метионина.

Выводы

1. Детальный анализ уровней консервативности нетранслируемых областей геномов дрожжей выявил заметное повышение уровня консервативности некодирующих областей после стоп-кодона, что, вероятнее всего, связано с обеспечением стабильности молекулы мРНК.

2. Анализ уровней консервативности сайтов связывания регуляторов транскрипции дрожжей выявил неабсолютную консервативность сайтов связывания регуляторных белков даже в пределах близкородственных геномов; кроме того, область повышенной консервативности включает не только непосредственно сайт связывания регулятора, но и простирается дальше за его пределы.

3. В дрожжах предсказан потенциальный митохондриальный транспортер альфа-изопропилмалата, промежуточного продукта пути биосинтеза лейцина, потенциально кодируемый геном *YOR271c*.

4. Предсказаны новые потенциальные члены регулона McbR, регулятора метаболизма метионина, в коринебактериях: ортологи гена *cg1739 C. glutamicum*, функция которого не ясна, и ортологи гена *CE2636 C. efficiens*, принимающие участие в увеличении пула метионина за счет расщепления 5'-метилтиоаденозина;

5. Показано разделение генов биосинтеза метионина и цистеина в стрептококках на два независимых регулона, и каждому из двух регуляторов,

¹¹ Grundy F.J., Henkin T.M. The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria // Mol. Microbiol. — 1998 — vol.30, no. 4. — Pp. 737-749.

MtaR и CmbR, приписана определенная функция в регуляции путей биосинтеза метионина и цистеина;

6. Для регулятора биосинтеза метионина, MtaR, в стрептококках были обнаружены новые потенциальные сайты связывания перед генами метаболического пути биосинтеза метионина.

7. Для регулятора биосинтеза цистеина, CmbR, в стрептококках был впервые идентифицирован мотив связывания и предсказаны новые потенциальные члены регулона CmbR, которым были присвоены тривиальные имена *cctW* и *cctY*. Белки CctW и CctY, по-видимому, принимают участие в транспорте цистеина и/или его предшественников внутрь клетки.

Список работ, опубликованных по теме диссертации

Статьи в рецензируемых научных журналах

[1] Kovaleva G.Yu., Makeev V.Yu., Gelfand M.S. Computer analysis of the GCN4 regulon of yeast *Saccharomyces cerevisiae* // Biophysics (Moscow). 2003. V. 48. Suppl. 1. P. S26-S29. (Номер издан только по-английски).

[2] Kovaleva G., Bazykin G., Brudno M., Gelfand M. Comparative genomics of transcriptional regulation in yeasts and its application to identification of a candidate alpha-isopropylmalate transporter // Journal of Bioinformatics and Computational Biology. 2006. V. 4. P. 981-998.

[3] Ковалева Г.Ю., Гельфанд М.С. Регуляция биосинтеза метионина/цистеина в *Corynebacterium glutamicum* и родственных организмах // Молекулярная биология. 2007. V. 41. P. 139-150.

[4] Kovaleva G.Y., Gelfand M.S. Transcriptional regulation of the methionine and cysteine transport and metabolism in streptococci // FEMS Microbiol. Lett. 2007. V. 76. P. 207-215.

Тезисы международных конференций

[1] Ковалева Г.Ю., Макеев В.Ю., Гельфанд М.С. Компьютерный анализ CGN4-регулона дрожжей *Saccharomyces cerevisiae* и *Candida albicans* // XI Международная научная конференция студентов, аспирантов и молодых ученых "Ломоносов-2004". Москва, Россия, 2004.

[2] Kovaleva G.Yu., Mironov A.A., Gelfand M.S. The conservation of transcription factor-binding sites in *Saccharomyces* genomes // The 4th International Conference on Bioinformatics of Genome Regulation and Structure. Novosibirsk, Russia, 2004.

[3] Kovaleva G., Bazykin G., Brudno M., Gelfand M. Conservation rate of transcription factor binding sites in *Saccharomyces* genomes // Bioinformatics Italian Society. Milan, Italy, 2005.

[4] Ковалева Г.Ю., Базыкин Е.А., Брудно М., Гельфанд М. Исследование дрожжевых геномов с использованием множественных геномных выравниваний // XII Международная научная конференция студентов, аспирантов и молодых ученых "Ломоносов-2005". Москва, Россия, 2005.

[5] *Kovaleva G., Gelfand M.* Conservation of transcription factor binding sites in fungal genomes // The Moscow Conference on Computational Molecular Biology. Moscow, Russia, 2005.

[6] *Kovaleva G., Gelfand M.* Identification of a candidate alpha-isopropylmalate transporter // The IVth European Conference on Computational Biology. Madrid, Spain, 2005.

[7] *Kovaleva G., Gelfand M.* Transcriptional regulation of the methionine biosynthesis in actinobacteria and streptococci // The 5th International Conference on Bioinformatics of Genome Regulation and Structure. Novosibirsk, Russia, 2006.

[8] *Kovaleva G., Gelfand M.* Evolution of the methionine biosynthesis in streptococci // The Vth European Conference on Computational biology. Eilat, Israel, 2007.

[9] *Kovaleva G., Gelfand M.* Regulation of methionine and cysteine biosynthesis in Streptococci // The 15th International Conference on Intelligent Systems for Molecular Biology and 6th European Conference on Computational Biology. Vienna, Austria, 2007.

[10] *Kovaleva G. Yu.* Regulation of the methionine and cysteine biosynthesis in streptococci // The Moscow conference on Computational Molecular biology. Moscow, Russia, 2007.

Патент

Гельфанд М.С., Ковалева Г.Ю. Способ исследования и прогнозирования пути биосинтеза метионина в родственных геномах коринебактерий // Регистрационный № 2006127265, 2006.

Ковалева Галина Юрьевна

**АНАЛИЗ РЕГУЛЯЦИИ ГЕНОВ МЕТАБОЛИЗМА И ТРАНСПОРТА
МЕТИОНИНА И ЛЕЙЦИНА МЕТОДАМИ СРАВНИТЕЛЬНОЙ ГЕНОМИКИ**

Транскрипционная регуляция генов метаболизма и транспорта метионина и лейцина была исследована в трех таксономических группах. Анализ, проведенный на геномах дрожжей, выявил заметное повышение уровня консервативности нетранскрибирующихся областей геномов после стоп-кодона, что связано с обеспечением стабильности молекулы РНК. В регуляторных областях геномов дрожжей обнаружены неабсолютная консервативность сайтов связывания регуляторов транскрипции и распространение области повышенной консервативности значительно за пределы сайта связывания регулятора. Применение методов сравнительной геномики позволило предсказать ген транспортера альфа-изопропилмалата, промежуточного продукта биосинтеза лейцина в геномах дрожжей, и идентифицировать два новых потенциальных члена пути метаболизма метионина в геномах коринебактерий, один из которых обеспечивает увеличение пула метионина за счет побочного продукта пути биосинтеза полиаминов. Анализ, проведенный на геномах стрептококков, выявил разделение генов пути биосинтеза метионина на два независимых регулона. Для каждого из двух регуляторов предсказан мотив связывания и приписана определенная функция в метаболическом пути. Кроме того, предсказаны два новых потенциальных члена регулона CmbR, принимающих участие в транспорте цистеина и/или его предшественников внутрь клетки.

Kovaleva Galina Yurjevna

**ANALYSIS OF METHIONINE AND LEUCINE METABOLISM AND TRANSPORT
GENES BY COMPARATIVE GENOMICS METHODS**

Regulation of the methionine and leucine metabolism and transport genes was studied for three taxonomic groups. Analysis of yeast genomes described a long-tailed conservation peak in the downstream intergenic regions, linked to mRNA stability. Analysis of conservation of individual transcription factor binding sites revealed that the region of increased conservation is not limited to individual binding sites, but extends much further into the surrounding region. The analysis of individual binding sites demonstrated that regulatory motives are not absolutely conserved even in closest genomes. The comparative genomic techniques allowed for the identification of a candidate alpha-isopropylmalate transporter, an intermediate of the leucine biosynthesis, in the yeast genomes. Two new potential members of the methionine metabolism pathway of corynebacteria were identified, one of them maintaining the methionine pool maintenance at the expense of a polyamine biosynthesis by-product. Investigation of the methionine metabolism pathway in the streptococcal genomes revealed subdivision of all structural genes into two independent regulons. For each regulator, the binding motive was predicted and the potential function was ascribed. The comparative genomics methods allowed for the identification of two new members of CmbR regulon, both of them are predicted to transport cysteine and/or its precursors into the cell.