

Равчеев Дмитрий Андреевич

ИЗУЧЕНИЕ ЭВОЛЮЦИИ РЕГУЛЯТОРНЫХ СИСТЕМ ПРОКАРИОТ
МЕТОДАМИ СРАВНИТЕЛЬНО-ГЕНОМНОГО АНАЛИЗА

03.00.28 – биоинформатика

Автореферат
диссертации на соискание ученой степени
кандидата биологических наук

Москва, 2009

Работа выполнена в Учреждении Российской академии наук Институте проблем передачи информации им. А.А. Харкевича РАН.

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

Официальные оппоненты: доктор биологических наук
Северинов Константин Викторович
Институт молекулярной генетики
Российской академии наук

кандидат физико-математических наук
Демин Олег Владимирович
Институт физико-химической биологии
им. А.Н. Белозерского МГУ

Ведущая организация: Федеральное государственное унитарное предприятие
Государственный научно-исследовательский институт
генетики и селекции промышленных микроорганизмов

Защита диссертации состоится 11 июня 2008 года в 15 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 _____ 2009 года

Ученый секретарь диссертационного совета
доктор биологических наук, профессор

Рожкова Г.И.

Общая характеристика работы

Актуальность темы:

Бактерии известны своей способностью приспосабливаться к различным условиям и занимать самые разнообразные экологические ниши. Подобная приспособляемость достигается за счет способности достаточно быстро отвечать на изменение окружающих условий и физиологического состояния клетки, чему микроорганизмы обязаны системе регуляции экспрессии собственных генов. В настоящее время общепринятым является мнение, что для эффективного приспособления к окружающим условиям важен не набор имеющихся генов, но и профили их регуляции. Именно регуляция позволяет эффективно использовать имеющиеся гены в зависимости от потребностей клетки.

Долгое время изучение регуляции транскрипции осуществлялось лишь экспериментально. Однако, экспериментальное исследование регуляции транскрипции является крайне трудоемким процессом. Методы же массового анализа, как правило, характеризуются высоким уровнем шума.

В последний десяток лет для изучения регуляции активно применяются методы сравнительного анализа последовательностей геномов. Этому способствуют высокие темпы роста количества геномов с известной полной последовательностью. Так, к настоящему моменту времени известны полные последовательности более чем 800 бактериальных геномов. При этом возрастает количество геномов, относящихся к одному порядку или семейству.

Сравнительный анализ геномных последовательностей позволяет предсказывать функции новых генов, осуществлять метаболическую реконструкцию, предсказывать регуляторные взаимодействия. В случае исследования регуляции методами сравнительной геномики основной задачей является выявление последовательностей, ответственных за регуляцию генов: промоторов и терминаторов транскрипции, сайтов связывания регуляторных белков, последовательностей потенциальных белков-регуляторов. При этом сравнительный анализ позволяет не только предсказывать регуляторные последовательности, но и проследивать их эволюцию, например, скоординированные изменения в последовательности белка-регулятора и сайта его связывания, исчезновение или появление регуляторных последовательностей перед определенными генами.

Исследование эволюции регуляции не является самоцелью. Подобные исследования позволяют получить новую информацию о биохимии и физиологии микроорганизмов, выявить критические различия между паразитическими и свободно живущими организмами, а также между патогенными и непатогенными бактериями.

Цели и задачи работы:

Целью настоящей работы был анализ эволюции регуляторных взаимоотношений в геномах гамма-протеобактерий. В качестве объектов исследования были выбраны регулятор метаболизма сахаров FruR, гомологичные репрессоры PurR и RbsR и глобальные регуляторы дыхания Fnr, ArcA и NarP.

В ходе работы были поставлены следующие задачи:

1. Изучение эволюции регуляторных систем на основании сравнения аминокислотных последовательностей регуляторных белков.
2. Построение распознающих правил для предсказания потенциальных сайтов связывания соответствующих факторов транскрипции.
3. Предсказание потенциальных регулонов методами сравнительной геномики.
4. Анализ регулонов в разных группах гамма-протеобактерий, поиск таксон-специфических особенностей регуляции, построение потенциальных сценариев эволюции.

Научная новизна и практическая ценность:

Впервые была подробно исследована регуляция белком FruR в четырех порядках гамма-протеобактерий. Предсказана регуляция 30 новых генов. На основании состава соответствующего регулона в различных таксонах построена эволюционная модель, в соответствии с которой FruR эволюционировал от локального регулятора к глобальному путем расширения регулона.

Подробно исследована эволюция гомологичных регуляторов PurR и RbsR. Приведены доводы в пользу гипотезы о происхождении регуляторов в результате дупликации предкового гена, предложен сценарий эволюции для этих белков. Для PurR предсказана регуляция 27 новых генов, показана консервативность регуляции генов биосинтеза пуриновых нуклеотидов и метаболизма одноуглеродных фрагментов.

Впервые проведен анализ комплексной регуляции дыхания гамма-протеобактерий, исследована регуляция посредством глобальных регуляторов Fnr, ArcA и NarP. Предсказана регуляция 153 новых генов. Кроме того, предсказан ряд таксон-специфических регуляторных каскадов, показаны корреляции между относительной ролью регулятора в каскаде и составом соответствующего регулона.

Апробация работы:

Основные положения диссертации были представлены на следующих конференциях:

1. The Third International Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2002), 14–20 July 2002, Novosibirsk, Russia.

2. XI международная научная конференция студентов, аспирантов и молодых ученых «Ломоносов-2004», 12–15 апреля 2004, Москва, Россия.
3. The Fourth International Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2004), 25–30 July 2004, Novosibirsk, Russia.
4. XII международная научная конференция студентов, аспирантов и молодых ученых «Ломоносов-2005», 12–16 апреля 2005, Москва, Россия.
5. 2nd Moscow Conference on Computational Molecular Biology (MCCMB'05), 18–21 July 2005, Moscow, Russia.
6. Международная школа «Биоинформатика, геномика, протеомика», 11–18 апреля 2006, Алма-Ата, Казахстан.
7. XIII международная научная конференция студентов, аспирантов и молодых ученых «Ломоносов-2006» 12–15 апреля 2006, Москва, Россия.
8. 4th Bertinoro Computational Biology Meeting “Evolution of and Comparative Approaches to Gene Regulation”, 24–30 June 2006, Bertinoro, Italy.
9. Межлабораторный семинар ИППИ РАН, 27 февраля 2007, Москва, Россия.
10. 3rd Moscow Conference on Computational Molecular Biology (MCCMB'07), 27–31 July 2007, Moscow, Russia.
11. 30-я конференция молодых ученых и специалистов ИППИ РАН «Информационные технологии и системы» (ИТиС'07), 18–21 сентября 2007, Звенигород, Россия.
12. 31-я конференция молодых ученых и специалистов ИППИ РАН «Информационные технологии и системы» (ИТиС'08), 29 сентября – 3 октября 2008, Геленджик, Россия.

Структура и объем диссертации:

Диссертационная работа изложена на 219 страниц машинописного текста и включает в себя введение и пять разделов. Первый раздел представляет собой обзор литературы по теме диссертации. Второй раздел описывает применявшиеся в настоящей работе методы, источники данных и использованное программное обеспечение. В разделах 3–5 содержатся оригинальные результаты и их обсуждение. Работа содержит 59 рисунков и 18 таблиц. Список цитируемой литературы, приводимый в конце диссертации, содержит 396 наименований.

Содержание работы

Исследование эволюции обобщенного FruR (Cra)-регулона

Белок FruR (Cra) представляет собой фактор транскрипции, координирующий потоки метаболизма углеводов в *Escherichia coli*. Этот белок обеспечивает переключение между

анаболизмом и катаболизмом сахаров, в случае недостатка сахаров активируя анаболические пути, а в случае их избытка – катаболические¹.

В настоящей работе ортологи белка FruR *E. coli* были найдены в геномах 18 гамма-протеобактерий, относящихся к четырём порядкам: Enterobacteriales, Pasteurellales, Vibrionales и Pseudomonadales. Поскольку все найденные ортологи FruR достаточно близки эволюционно, для поиска потенциальных сайтов связывания этого белка возможно использование единой матрицы позиционных весов.

Для построения такой матрицы были рассмотрены 5'-некодирующие области генов *E. coli*, для регуляция которых была показана экспериментально. В результате с помощью программы SignalX была построена матрица для распознавания потенциальных сайтов связывания белка FruR. Сайт связывания белка FruR представляет собой нестрогий палиндром длиной 16 нуклеотидов (Рисунок 1), консенсус сайта при этом согласуется с ранее предсказанным².



Рисунок 1. Диаграмма Лого для сайтов связывания FruR

На следующем этапе был проведен поиск потенциальных сайтов связывания FruR в геномах гамма-протеобактерий. Для определения состава обобщенного регулона был применен метод проверки соответствия: ген считался относящимся к обобщенному регулону, если сайт перед опероном, содержащим его, был обнаружен в нескольких геномах.

В соответствии с описанной процедурой, к обобщенному регулону было отнесено в целом 57 генов. Обобщенный FruR-регулон включает в себя все гены, входящие в регулон в *E. coli*, и, кроме того, по результатам настоящего исследования к регулону было отнесено еще 30 генов. Таковыми являются гены фосфотрансферазных систем (*manXYZ*), центрального метаболизма (*tpiA*, *gapA*, *gpmA*, *pdhR-aceEF-lpdA*), белков дыхательных комплексов (*nuoABCDEFGHIJKLMN*), ферментов ассимиляции азота (*nirBDC-cysG*), транспортных белков (*copA*) и регуляторных белков (*fruR*, *crp*).

В геномах бактерий из порядка Pseudomonadales были обнаружены доменные перестройки для генов фруктозо-специфичной фосфотрансферазной системы, кодируемых опероном *fruBKA*.

¹ Saier M.H., Ramseier T.M. The catabolite repressor/activator (Cra) protein of enteric bacteria // J Bacteriol – 1996 – Vol. 178 (12) – Pp. 3411-3417.

² Negre D., Bonod-Bidaud C., Geourjon C., Deleage G., Cozzzone A.J., Cortay J.C. Definition of a consensus DNA-binding site for the *Escherichia coli* pleiotropic regulatory protein, FruR // Mol Microbiol – 1996 – Vol. 21 (2) – Pp. 257-266.

В случае Pseudomonadales первый ген из этого оперона кодирует химерный белок, N- и C-концевые участки которого ортологичны соответственно белкам FruB и PtsI *E. coli*. То есть, у Pseudomonadales *fru*-оперон кодирует не только специфичные для транспорта фруктозы белки, но и белки, необходимые для функционирования всех фосфотрансферазных систем. Возможно, это связано с тем, что в геномах бактерий этой группы гены для других фосфотрансферазных систем отсутствуют.

В ходе исследования FruR-регулона в различных гамма-протеобактериях была замечена следующая особенность: состав регулона крайне сильно отличается в различных порядках микроорганизмов. Так, в Pseudomonadales в состав регулона входит лишь один оперон, *fruB/ptsI-fruKA*. Аналогичная ситуация наблюдается и в Vibrionales, где единственным членом регулона является оперон *fruBKA*. В группе Pasteurellales регулон полностью отсутствует. У бактерий порядка Enterobacteriales, по сравнению с другими таксонами, наблюдается необычайное расширение регулона. Так, опероны *fruBKA*, *ptsHI-crr*, *manXYZ*, *glk*, *epd-pgk-fbaA*, *ppsA*, *pfkA*, *pykF*, *icdA*, *adhE*, *tpiA*, *gapA*, *gpmA*, *pdhR-aceEF-lpdA*, *nuoABCDEFGHIJKLMN*, *nirBDC-cysG*, *copA* и *crp* входят в состав FruR-регулона в большинстве геномов из этой группы. Другие же опероны, а именно *mtlADR*, *fbp*, *edd-eda*, *pckA* и *aceBAK*, регулируются лишь в *E. coli* и близких к ней *Salmonella* spp.

Таким образом, можно условно разделить все члены обобщенного FruR-регулона на три группы, в соответствии с таксономическим распространением их регуляции:

- гены, регулируемые во всех исследованных геномах, за исключением Pasteurellales;
- гены, регулируемые только в геномах Enterobacteriales;
- гены, регулируемые только в геномах *E. coli* и *Salmonella* spp.

На основании подобного распределения членов регулона по таксономическим группам, был предсказан наиболее вероятный сценарий эволюции FruR-регулона. По всей видимости, первоначально FruR-регуляция возникла у общего предка еще до отделения группы Pseudomonadales, причем первоначально FruR существовал как регулятор единственного *fru*-оперона. Такая ситуация сохранилась в геномах представителей Pseudomonadales и Vibrionales.

В дальнейшем в группе Pasteurellales регулон разрушился в процессе эволюции, причем в разных геномах можно наблюдать разные стадии этого процесса. У части организмов, таких как *Haemophilus influenzae* и *Haemophilus dukreyi* ген *fruR* оказался полностью утрачен. В геноме *Pasteurella multocida*, хотя и присутствует полноразмерный ген *fruR*, не было найдено потенциальных сайтов связывания белка ни перед одним геном, входящим в регулон в других исследованных микроорганизмах.

У Enterobacteriales произошло значительное расширение регулона путем появления новых сайтов связывания FruR перед генами центрального метаболизма, дыхания, ассимиляции азота, транспортных и регуляторных белков. В процессе дальнейшего эволюционного

расхождения разных групп Enterobacteriales происходили изменения состава регулона путем исчезновения сайтов перед теми или иными генами. Дальнейшее расширение FruR-регулона произошло в узкой группе, входящей в состав Enterobacteriales, и включающей в себя *E. coli* и *Salmonella* spp. (Рисунок 2).

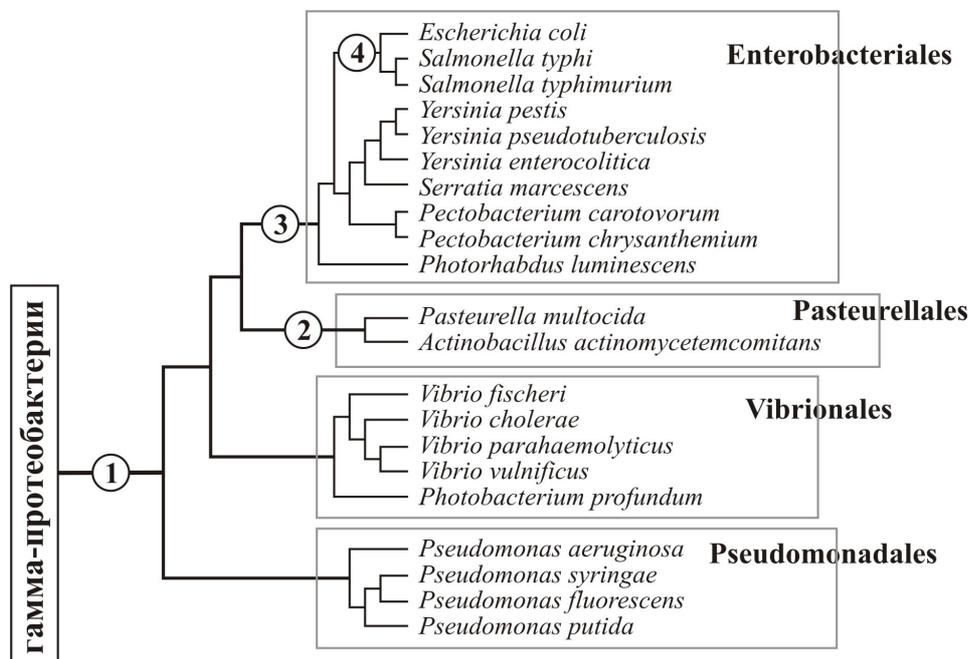


Рисунок 2. Предполагаемый сценарий эволюции FruR-регулона гамма-протеобактерий. Цифрами отмечены предполагаемые эволюционные события: 1 – возникновение FruR-регуляции оперона *fru*; 2 – разрушение регулона в Pasteurellales; 3 – расширение регулона в Enterobacteriales; 4 – дальнейшее расширение регулона в группе, включающей *E. coli* и *Salmonella* spp.

Исследование эволюции обобщенных PurR- и RbsR-регулонов

Транскрипционные факторы PurR и RbsR – это очень похожие белки. Так, последовательности указанных белков из *E. coli* тождественны на 47,0%. При этом в *E. coli* эти белки являются ближайшими гомологами, то есть сходство между ними значительно выше, чем сходство каждого из них с любым другим белком, закодированном в этом геноме, что позволяет предположить происхождение этих белков в результате относительно недавней дупликации одиночного предкового гена. Поэтому в настоящей работе было проведено исследование эволюции белков PurR и RbsR и регулируемых ими генов.

Ортологи PurR *E. coli* были найдены в 15 геномах бактерий из порядков Enterobacteriales (*E. coli*, *Salmonella typhi*, *S. typhimurium*, *Yersinia pestis*, *Y. pseudotuberculosis*, *Pectobacterium carotovorum*, *Photobacterium luminescens*), Pasteurellales (*Pasteurella multocida*, *Haemophilus influenzae*, *H. ducreyi*) и Vibrionales (*Vibrio vulnificus*, *V. parahaemolyticus*, *V. cholerae*, *V. fischeri*, *Photobacterium profundum*). Во всех перечисленных геномах, за исключением *Y. pestis*, *Y.*

pseudotuberculosis и *H. ducreyi*, были также обнаружены ортологи белка RbsR. В геномах из порядка Pseudomonadales (*Pseudomonas aeruginosa*, *P. putida*, *P. fluorescens*, *P. syringae*) был обнаружен белок, одинаково схожий с PurR и RbsR из *E. coli*. Можно предположить, что белок из Pseudomonadales происходит напрямую от предкового белка, а PurR и RbsR из Enterobacteriales, Pasteurellales и Vibrionales являются потомками копий, образовавшимися в результате дупликации исходного гена.

Ранее было показано, что в *E. coli* белок RbsR репрессирует экспрессию оперона *rbsDACBKR*, связываясь с сайтом, расположенном в промоторной области³. Кроме того, сайты перед *rbs*-опероном были предсказаны в геномах *S. typhi*, *P. multocida*, *H. influenzae* и *V. cholerae*⁴. На основе вышеперечисленных сайтов была построена матрица для распознавания потенциальных сайтов связывания белка RbsR. Сайт связывания белка RbsR представляет собой палиндром длиной 20 нуклеотидов (Рисунок 3).



Рисунок 3. Диаграмма Лого для сайтов связывания RbsR

В результате поиска потенциальные сайты связывания RbsR были обнаружены перед опероном *rbsDACBKR* во всех геномах, где присутствует ген регуляторного белка. Обнаружить консервативные потенциальные сайты перед какими-либо другими оперонами не удалось.

В геноме *H. ducreyi* не было найдено ортологов для ни одного из генов рибозного оперона, тогда как в геномах *Y. pestis* и *Y. pseudotuberculosis*, где ген белка RbsR отсутствует, были обнаружены ортологи для генов *rbsD* и *rbsK*, лежащие в одном потенциальном опероне, причем перед опероном был найден потенциальный сайт с весьма высоким весом (Рисунок 4). Подобные данные указывают на то, что утрата гена регулятора произошла относительно недавно. При этом были утеряны и все гены транспортера рибозы, кроме *rbsD*, но сохранился ген *rbsK*, кодирующий киназу рибозы.

³ Mauzy C.A., Hermodson M.A. Structural and functional analyses of the repressor, RbsR, of the ribose operon of *Escherichia coli* // Protein Sci – 1992 – Vol. 1 (7) – Pp. 831-842.

⁴ Laikova O.N., Mironov A.A., Gelfand M.S. Computational analysis of the transcriptional regulation of pentose utilization systems in the gamma subdivision of Proteobacteria // FEMS Microbiol Lett – 2001 – Vol. 205 (2) – Pp. 315-322.

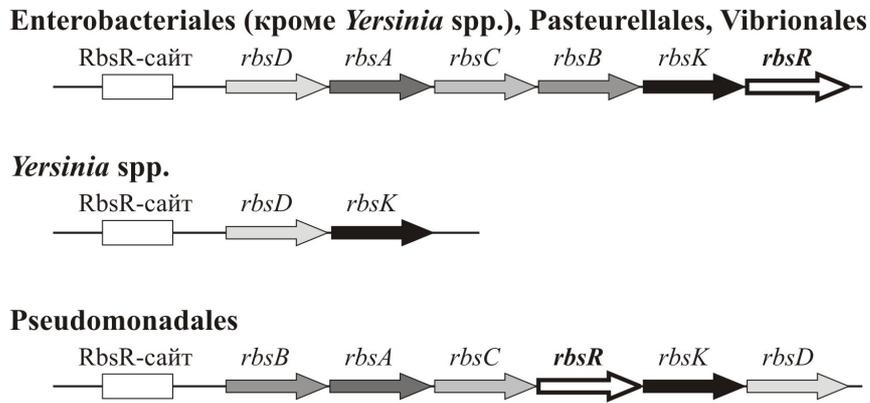


Рисунок 4. Структура рибозного оперона в различных гамма-протеобактериях.

Ввиду эволюционной близости всех найденных ортологов PurR, PurR-регуляция во всех геномах была исследована с использованием единой матрицы. В геноме *E. coli* ранее были экспериментально установлены сайты связывания пуринового репрессора перед 19 оперонами⁵. Эти сайты были использованы в качестве обучающей выборки для построения соответствующей матрицы. Сайт связывания белка PurR представляет собой нестрогий палиндром длиной 16 нуклеотидов (Рисунок 5).

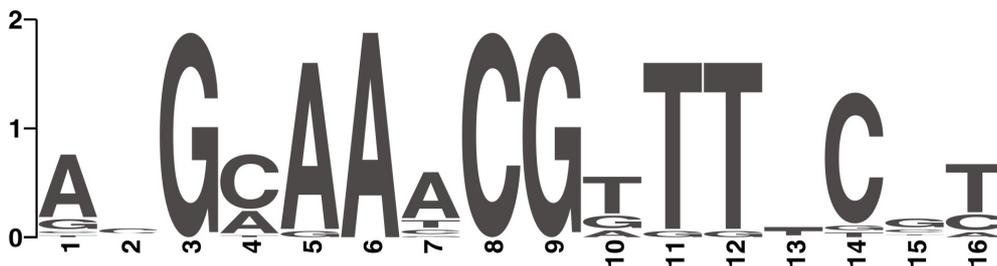


Рисунок 5. Диаграмма Лого для сайтов связывания PurR

К обобщенному PurR-регулону было отнесено в общей сложности 62 гена, объединенных как минимум в 49 оперонов. Помимо экспериментально известных членов регулона и *E. coli* и генов с ранее предсказанной регуляцией, к обобщенному регулону были добавлены 27 новых генов. К новым членам регулона относятся гены модификации пуриновых нуклеотидов (*guaC*, *gsk*, *ushA*), синтеза пиримидиновых нуклеотидов (*upp-uraA*, *carAB*, *pyrLBI*), метаболизма одноуглеродных фрагментов (*folD*, *rpiA*, *serA*, *fhs*), транспортных белков (*uraA*, *gltS*), утилизации нуклеотидов и нуклеиновых кислот (*xseA*, *cytR*), центрального метаболизма (*pckA*, *ppsA*, *glpX*) и гены с неизвестной функцией (*ydiA*, *yiiU*, *yhhQ*, *ydiJ*, *HD1120*, *VC2168*).

В ходе исследования эволюции пуриновой регуляции гамма-протеобактерий было замечено, что состав регулона достаточно сильно различается в разных видах бактерий. Все

⁵ Zalkin H., Neidhardt P. Biosynthesis of purine nucleotides // *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology – ASM Press – Washington D.C. – Pp. 1325-1333.

входящие в регулон гены можно условно разделить на три группы. Первая группа – так называемое ядро регулона, то есть гены, регулирующие во всех геномах или в подавляющем большинстве. К этой группе относятся большая часть генов биосинтеза инозинмонофосфата (ИМФ) из рибозо-5-фосфата (*purT*, *purL*, *purMN*, *purHD*, *prsA*, *purC*, *purB*) и гены метаболизма одноуглеродных фрагментов, ассоциированных с фолатами (*gcvTHP*, *folD*, *glyA*, *fhs*). Вторая группа включает гены, регуляция которых слабо консервативна, но при этом встречается в различных таксонах. Это гены азотного метаболизма (*speAB*), метаболизма одноуглеродных фрагментов (*serA*), трансмембранных белков (*gltS*, *yicE*, *yjcD*), утилизации нуклеотидов (*xseA*), а также ген с неизвестной функцией *yhhQ*. К третьей группе были отнесены гены с таксон-специфической регуляцией. Для Enterobacteriales характерна регуляция собственно гена белка регулятора PurR, генов, участвующих в модификации нуклеотидов (*purA*, *guaBA*), генов биосинтеза пиримидиновых нуклеотидов (*codBA*, *upp-uraA*, *pyrC*, *pyrD*, *carAB*, *pyrLBI*), генов трансмембранных белков (*tsx*, *yieG*) и гена с неизвестной функцией *ydiJ*. Что касается порядка Pasteurellales, то для него характерна регуляция гораздо меньшего числа генов, таких как *rpiA*, ответственного за синтез рибозо-5-фосфата, генов центрального метаболизма *pckA* и *glpX*, а также генов *yiiU* *ydiJ* и *HD1120*, функция которых неизвестна. Картина, наблюдающаяся в Vibrionales, похожа на таковую для бактерий из порядка Enterobacteriales с некоторыми отличиями. Так, для Vibrionales была установлена регуляция генов модификации пуриновых нуклеотидов (*guaBA*, *guaC*, *gsk*, *ushA*), биосинтеза пиримидиновых нуклеотидов (*upp-uraA*, *pyrC*, *pyrD*, *pyrLBI*), трансмембранных белков (*yieG*), центрального метаболизма (*pckA*, *glpX*, *ppsA*) и ряда генов с неизвестными функциями (*ydiA*, *VC2168*, *yiiU*).

В целом охарактеризовать консервативность PurR-зависимой регуляции можно следующим образом: наиболее консервативной является регуляция генов, участвующих в биосинтезе ИМФ из рибозо-5-фосфата, а также генов, ответственных за синтез необходимых для этого производных фолатов. Регуляция генов, необходимых для модификаций ИМФ и синтеза пиримидиновых нуклеотидов, так же как и генов центрального метаболизма консервативна на уровне определенных таксонов. Для остальных генов регуляция хоть и встречается в различных таксонах, но консервативной не является.

Бактерии порядка Pseudomonadales содержат белок, примерно одинаково сходный с белками PurR и RbsR из других порядков гамма-протеобактерий. В ходе анализа локализации данного гена на хромосоме было обнаружено, что он расположен совместно с *rbs*-генами, и, по всей видимости, составляет с ними один оперон *rbsBACRKD* (Рисунок 4). При сравнении структуры *rbs*-оперона в Pseudomonadales и бактериях из других таксонов было замечено, что в эволюции сохраняется состав оперона, хотя и меняется порядок генов в нем. На основании кластеризации белка-регулятора из Pseudomonadales с генами *rbs*-оперона естественно предположить, что данный белок выполняет функцию RbsR, то есть регулирует экспрессию

rhs-оперона. Исходя из этого предположения, были исследованы промоторные области оперона и с помощью программы программы SignalX были обнаружены высоко консервативные участки длиной 14 нуклеотидов. Как и в случаях PurR и RbsR других гаммапротеобактерий, сайт имеет структуру палиндрома (Рисунок 6). В результате поиска консервативные потенциальные сайты связывания RbsR Pseudomonadales были обнаружены во всех исследованных геномах из данного порядка только перед собственно *rhs*-опероном. Ни перед одним из других генов не удалось выявить значимых консервативных сайтов. Таким образом, можно утверждать, что в Pseudomonadales RbsR представляет собой регулятор одного оперона.



Рисунок 6. Диаграмма Лого для сайтов связывания RbsR Pseudomonadales

С целью более детально проследить эволюцию регуляторных систем PurR и RbsR, было проведено сравнение сайтов связывания трех регуляторных белков (Рисунок 7). Сайты всех исследуемых белков имеют палиндромную структуру. При этом центральная часть сайтов длиной 8 нуклеотидов весьма консервативна и имеет консенсус AAACGTTT, а информационное содержание крайних позиций невелико. Наиболее значительные различия наблюдаются в тройках нуклеотидов, ближайших к центральной консервативной части. Именно в этой области и наблюдаются основные различия между сайтами PurR и RbsR. В то же время, последовательность сайта RbsR Pseudomonadales более похожа на таковую для PurR, чем для RbsR. По всей видимости, сайт связывания PurR сохранил исходную структуру, существовавшую в предковом геноме.

На основании данных о филогении регуляторных белков, структуре сайтов и составе обобщенных регулонов была выдвинута следующая модель эволюции регуляторных систем PurR и RbsR (Рисунок 8). Судя по всему, первоначально белок RbsR существовал как локальный регулятор, контролирующей экспрессию одного *rhs*-оперона. Однако затем, после отделения предковых форм Pseudomonadales, произошла дупликация гена белка-регулятора. Далее в к происходили следующие события. В одной ветви белок сохранила свою функцию локального регулятора, рибозного репрессора, но при этом изменилась структура сайта связывания. Так образовался регулятор RbsR Enterobacteriales, Pasteurellales и Vibrionales. В другой же ветви структура сайта в целом сохранилась, однако, изменилась функция белка. В отличие от FruR, для PurR не наблюдается постепенного расширения регулона. По-видимому, образование регулона происходило довольно быстро в эволюционном масштабе.

Enterobacteriales, Pasteurellales, Vibrionales



Pseudomonadales



Рисунок 7. Сравнение сайтов связывания RbsR, PurR и RbsR Pseudomonadales.

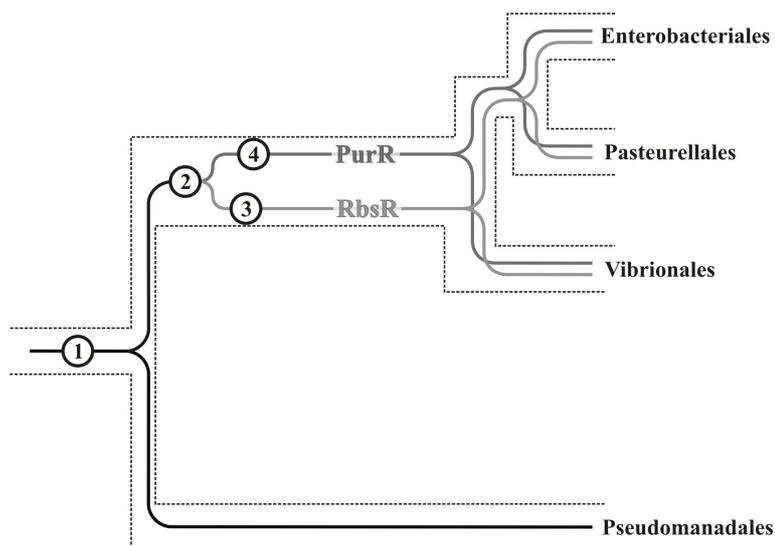


Рисунок 8. Предполагаемый сценарий эволюции регулонов RbsR, PurR и RbsR Pseudomonadales. Цифрами отмечены предполагаемые эволюционные события: 1 – RbsR – локальный регулятор *rbs*-оперона; 2 – удвоение гена белка-регулятора; 3 – изменение структуры сайта связывания RbsR с сохранением функции; 4 – изменение специфичности связывания с лигандом PurR при сохранении сайта, возникновение PurR-регулона.

Эволюция глобальной регуляции дыхания

Предшествующие исследования Fnr- и ArgC-регулонов гамма-протеобактерий^{6,7} показали, что глобальная регуляция дыхания слабо консервативна. По всей видимости, это связано с

⁶ Герасимова А.В., Родионов Д.А., Миронов А.А., Гельфанд М.С. Компьютерный анализ регуляторных сигналов в бактериальных геномах. Участки связывания Fnr // Мол Биол – 2001 – Т. 35 – №6 – С. 1001-1010.

наличием множественной регуляции у генов, задействованных в дыхании и смежных процессах. В связи с этим было принято решение о комплексном изучении глобальной регуляции дыхания, то есть, об одновременном исследовании Fnr-, ArcA- и NarP-зависимой регуляции.

Ортологи белка Fnr из *E. coli* были найдены в 12 организмах, относящихся к трем порядкам: Enterobacteriales (*Salmonella typhi*, *Yersinia pestis*, *Y. enterocolitica*, *Pectobacterium carotovorum*), Pasteurellales (*Pasteurella multocida*, *Actinobacillus actinomycetemcomitans*, *Haemophilus influenzae*, *H. ducreyi*), и Vibrionales (*Vibrio vulnificus*, *V. parahaemolyticus*, *V. cholerae*, *V. fischeri*).

Во всех перечисленных геномах были найдены и ортологи регуляторного белка ArcA. Ортологи же сенсорного белка ArcB, активирующего ArcA, были обнаружены во всех перечисленных выше геномах, кроме *H. influenzae* и *H. ducreyi*. Для ортологов ArcB была обнаружена еще одна особенность. Известно, что ключевую роль в димеризации и последующей активации ArcB *E. coli* играют цистеиновые остатки Cys180 и Cys241, расположенные в трансмембранном PAS-домене, причем основную роль в димеризации играет Cys180, а замена Cys241 является менее критичной⁸. Оба цистеиновых остатка консервативны лишь в белках Enterobacteriales и *V. fischeri*, тогда как в других Vibrionales консервативен лишь Cys180, что дает повод полагать, что у Vibrionales ArcB, скорее всего, функционален. У представителей же Pasteurellales полностью отсутствует PAS-домен. Не исключено, что у этих бактерий активация ArcA осуществляется другим сенсорным белком. Подобная ситуация была описана для другого представителя гамма-протеобактерий, *Shewanella oneidensis* MR-1⁹. Тем не менее, наличие ортологов регуляторного белка ArcA во всех вышеупомянутых организмах позволяет провести исследование соответствующей регуляции.

Поиск ортологов регуляторных белков показал, что оба регулятора, NarL и NarP, присутствуют лишь в *E. coli*, *S. typhi* и *P. carotovorum*, тогда как в геномах *Yersinia* spp., Pasteurellales и Vibrionales был обнаружен только одиночный регулятор NarP. Похожая, но не идентичная ситуация наблюдалась для сенсорных белков NarX и NarQ. В геномах, где присутствуют гены для обоих регуляторов, также обнаружены гены двух сенсорных белков. В геномах же с одиночной регуляторной системой ген *narP* существует совместно с *narQ*, за исключением *Y. pestis* и *Y. enterocolitica*, в которых обнаружена нестандартная пара NarP-NarX.

⁷ Герасимова А.В., Гельфанд М.С., Макеев В.Ю., Миронов А.А., Фаворов А.В. Регулятор ArcA в гамма-протеобактериях. Определение сайтов связывания и описание регулона // Биофизика – 2003 – Т. 48 – №1 – С. 21-25.

⁸ Malpica R., Franco B., Rodriguez C., Kwon O., Georgellis D. Identification of a quinone-sensitive redox switch in the ArcB sensor kinase // Proc Natl Acad Sci USA – 2004 – Vol. 101 (36) – Pp. 11318-11323.

⁹ Gralnick J.A., Brown C.T., Newman D.K. Anaerobic regulation by an atypical Arc system in *Shewanella oneidensis* // Mol Microbiol – 2005 – Vol. 56 (5) – 1347-1357.

Поскольку NarX взаимодействует с белком NarL, но не с NarP¹⁰ (Rabin and Stewart, 1993), последовательности сенсорных белков были рассмотрены более подробно.

Важной особенностью модуля NarX *E. coli*. является кластер цистеинов Cys265, Cys302, Cys308, Cys313 и Cys316¹¹. Было обнаружено, что в последовательностях белков из *Y. pestis* и *Y. enterocolitica* отсутствует характерные для NarX цистеиновые остатки. Следовательно, хотя сенсорные белки из бактерий рода *Yersinia* spp. близки по последовательности к NarX, они, по всей вероятности, выполняют функцию NarQ.

Были также обнаружены корреляции между строением регуляторной системы и системы нитрат-нитритного дыхания. Так, в геномах, содержащих обе системы NarX-NarL и NarQ-NarP, были обнаружены гены как для цитоплазматической, так и для периплазматической нитрат-редуктаз. В геномах, где присутствует лишь одна регуляторная двукомпонентная система, были найдены гены только для периплазматической нитрат-редуктазы.

Поскольку данные о структуре сайтов связывания NarL противоречивы, а наличие удвоенной двукомпонентной системы усложняет исследование, было принято решение ограничиться изучением регуляции в геномах, содержащих одну систему регуляции нитрат-нитритного дыхания. Во всех таких геномах была исследована регуляция белками Fnr, ArcA и NarP.

Матрица для поиска потенциальных сайтов связывания Fnr была составлена на основе экспериментально определенных сайтов из *E. coli*. Fnr-связывающий мотив представляет собой инвертированный повтор длиной 14 нуклеотидов (Рисунок 9).

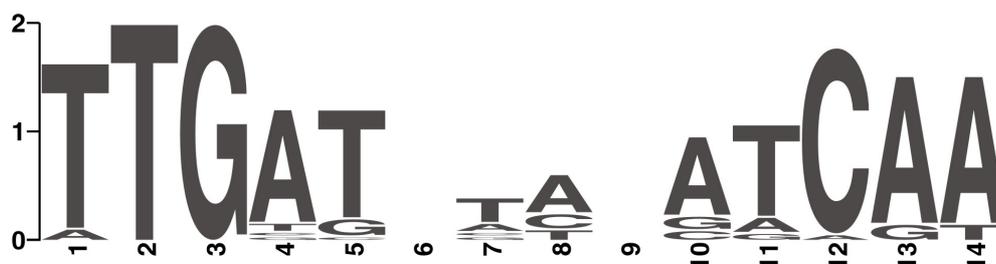


Рисунок 9. Диаграмма Лого для сайтов связывания Fnr

В случае ArcA-зависимой регуляции существующих данных о структуре сайтов связывания было недостаточно для построения чувствительного распознающего правила. Поэтому регуляторные области генов, экспрессия которых регулируется ArcA в *E. coli* были

¹⁰ Rabin R.S., Stewart V. Dual response regulators (NarL and NarP) interact with dual sensors (NarX and NarQ) to control nitrate- and nitrite-regulated gene expression in *Escherichia coli* K-12 // J Bacteriol –1993 – Vol. 175 (11) – Pp. 3259-3268.

¹¹ Stewart V. Biochemical Society Special Lecture. Nitrate- and nitrite-responsive sensors NarX and NarQ of proteobacteria // Biochem Soc Trans – 2003 – Vol. 31 (1) – Pp. 1-10.

проанализированы с помощью программы SeSiMCMC¹². Всего перед 14 генами было найдено 18 потенциальных сайтов связывания регулятора, на основании которых была построена матрица позиционных весов. Потенциальный мотив связывания ArcA представляет собой несовершенный повтор длиной 15 нуклеотидов (Рисунок 10).

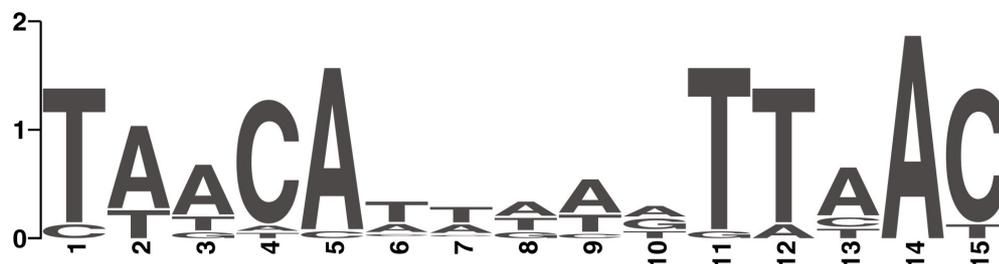


Рисунок 10. Диаграмма Лого для сайтов связывания ArcA

В отличие от других регуляторов дыхания, Fnr и ArcA, в случае NarP использование матрицы, построенной по сайтам связывания данного белка в *E. coli*, было сочтено неправомерным в силу следующих причин. В *E. coli* регуляция генов нитрат-нитритного дыхания осуществляется сразу двумя гомологичными белками, NarL и NarP, причем наблюдается различная специфичность белков-регуляторов к сайтам связывания¹³. Во всех исследованных геномах, где был обнаружен единственный ген *narP*, также были найдены гены периплазматической нитрат-редуктазы *nar* и гены *cst*-оперона, отвечающие за перенос гема в периплазму. В некоторых из этих геномов были найдены гены периплазматической нитрит-редуктазы *nrf*. Для построения матрицы для поиска сайтов связывания NarP были исследованы регуляторные области оперонов, содержащих *nar*-, *cst*- и *nrf*-гены в исследуемых геномах. Перед большинством данных оперонов были найдены палиндромные последовательности длиной 16 п.н. (Рисунок 11).



Рисунок 11. Диаграмма Лого для сайтов связывания NarP

¹² Favorov A.V., Gelfand M.S., Gerasimova A.V., Ravcheev D.A., Mironov A.A., Makeev V.J. A Gibbs sampler for identification of symmetrically structured, spaced DNA motifs with improved estimation of the signal length. *Bioinformatics* – 2005 – V. 21 (10) – Pp. 2240-2245.

¹³ Darwin A.J., Tyson K.L., Busby S.J., Stewart V. Differential regulation by the homologous response regulators NarL and NarP of *Escherichia coli* K-12 depends on DNA binding site arrangement // *Mol Microbiol* – 1997 – Vol. 25 (3) – 583-595.

В десяти геномах гамма-протеобактерий были исследованы обобщенные регулоны для факторов транскрипции Fnr, ArcA и NarP. Поскольку, как уже отмечалось ранее, регуляция каждой отдельной регуляторной системой слабо консервативна, то имеет смысл рассматривать регуляцию тремя системами в целом. В общей сложности к трем обобщенным регулонам было отнесено 153 гена, в том числе и 68 новых членов регулонов. К новым членам регулонов относятся гены дыхательных систем (*atpIBEFHAGCD*, *nqrABCDEF*, *torYZ*, *fdoGHI*, *fdxGHI*, *fdhD*, *dadAX*), центрального метаболизма (*sfcA*, *pckA*, *eno*, *pgk*, *talB*, *aldB*), метаболизма углеводов (*ptsHI-crr*, *mtlADR*, *deoCABD*, *nagB*, *glgBXCAP*), метаболизма жирных кислот (*fabBA*, *fadIJ*, *fadD*, *acpP-fabF*), ответа на кислородный стресс (*sodA*), нуклеотидредуктаз (*nrdAB*), транспортных белков (*gltP*, *gntXY*, *fadL*) и регуляторных белков (*dusB-fis*, *cpxRA*, *oxyR*, *fur*).

В ходе исследования были выявлены существенные различия в регуляции между порядками гамма-протеобактерий. Так, в *Yersinia* spp. обобщенный регулон NarP оказался намного меньше, чем в других таксонах. Обратная ситуация наблюдается для Pasteurellales, где обобщенный регулон NarP увеличен по сравнению с другими таксонами. Другой отличительной чертой этого порядка является высокая степень перекрытия между обобщенными регулонами для разных регуляторов. Характерной чертой Vibrionales является увеличение обобщенного ArcA по сравнению с другими группами.

На основе наличия консервативных предсказанных сайтов перед генами *fnr*, *arcA*, и *narP* был предсказан ряд таксон-специфических регуляторных каскадов (Рисунок 12). Для *E. coli* регуляторные каскады были известны из экспериментальных работ. В этом организме Fnr представляет собой главный регулятор дыхания, контролирующей экспрессию генов других регуляторов: Fnr репрессирует транскрипцию оперона *narXL* и собственного гена, а также регулирует транскрипцию гена *arcA*¹⁴. Перед геном *arcA* был ранее предсказан собственный сайт, и на основании положения последнего относительно экспериментально известных промоторов был сделан вывод о позитивной авторегуляции¹⁵. Регуляторы нитрат-нитритного дыхания формируют сложную сеть регуляторных взаимодействий¹⁶.

Однако, в геномах других гамма-протеобактерий наблюдается иная ситуация. В *Yersinia* spp. Fnr, по всей видимости, регулирует экспрессию *arcA*, и оба гена *fnr* и *arcA* авторегулируются. При этом система регуляции нитрат-нитритного дыхания не принимает участия в регуляторных каскадах.

Более значительные изменения в системе регуляторных каскадов были обнаружены в геномах Pasteurellales. В геномах большинства представителей этого порядка потенциальные

¹⁴ Takahashi K., Hattori T., Nakanishi T., Nohno T., Fujita N., Ishihama A., Taniguchi S. Repression of in vitro transcription of the *Escherichia coli* *fnr* and *narX* genes by FNR protein // FEBS Lett – 1994 – Vol. 340 (1-2) – Pp. 59-64.

¹⁵ Герасимова А.В., Гельфанд М.С., Макеев В.Ю., Миронов А.А., Фаворов А.В. Регулятор ArcA в гамма-протеобактериях. Определение сайтов связывания и описание регулона // Биофизика – 2003 – Т. 48 – № 1 – С. 21-25.

¹⁶ Darwin A.J., Stewart V. Expression of the *narX*, *narL*, *narP*, and *narQ* genes of *Escherichia coli* K-12: regulation of the regulators // J Bacteriol – 1995 – V. 177 (13) – Pp. 3865-3869.

сайты связывания Fnr, ArcA и NarP были обнаружены перед геном *fnr*, тогда как перед генами *arcA* и *narP* не было найдено консервативных сайтов.

В случае Vibrionales также была предсказана авторегуляция для генов *fnr*, *arcA* и *narP*. Белок ArcA, по всей видимости, регулирует экспрессию гена *narP*, тогда как Fnr контролирует экспрессию генов *arcA* и *narP*.

Таким образом, структура регуляторных каскадов, характерная для *E. coli*, не сохраняется в других гамма-протеобактериях, и относительная роль каждого регулятора в каскадах может меняться от таксона к таксону.

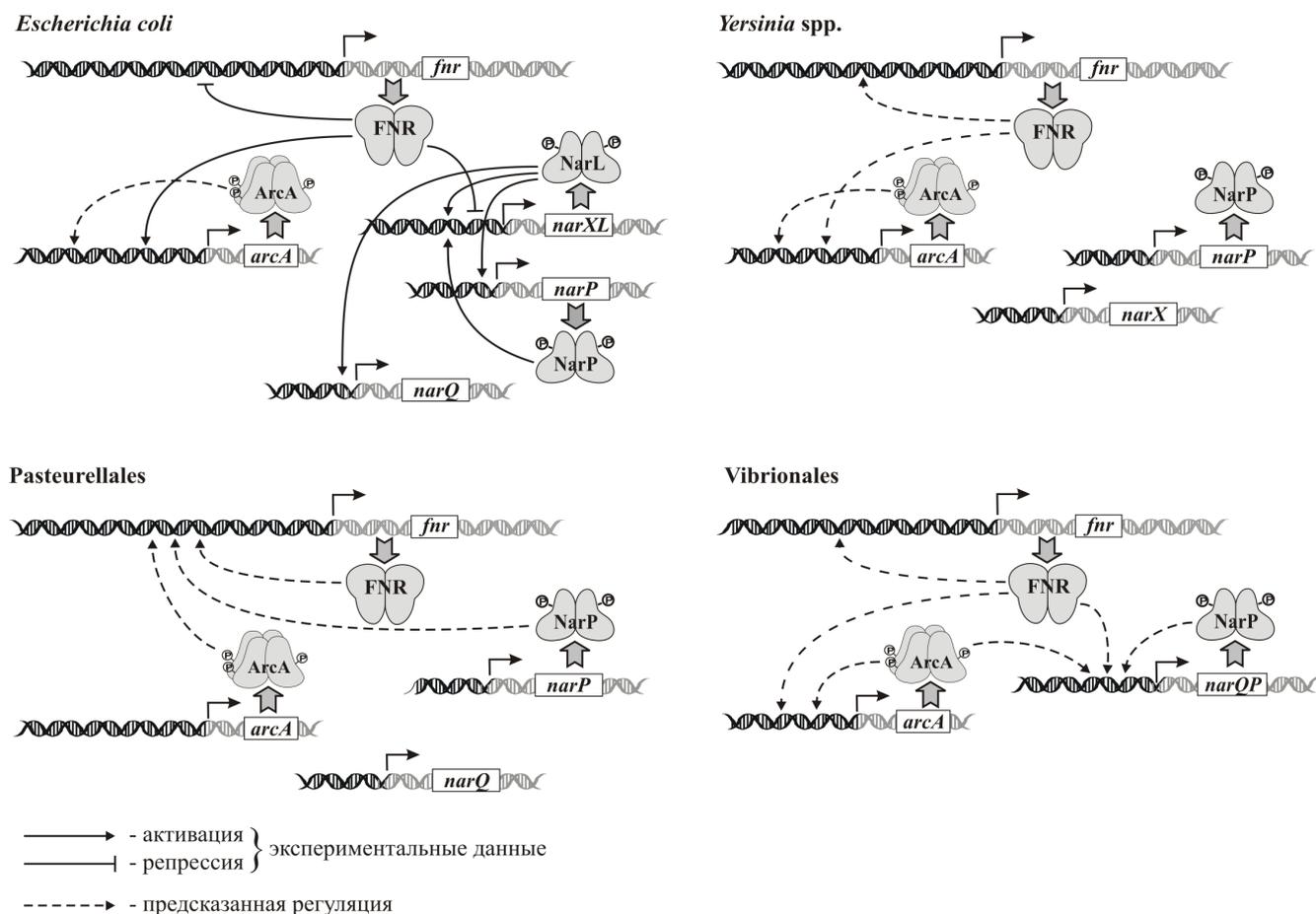


Рисунок 12. Регуляторные каскады в различных таксонах

Выводы

1. Построено распознающее правило для поиска потенциальных сайтов связывания белка фруктозного репрессора FruR и предсказано 26 новых членов обобщенного FruR-регулона.
2. Построена модель эволюции обобщенного FruR-регулона, показана эволюция от локального регулятора оперона *fruBKA* до глобального регулятора метаболизма сахаров.

3. Построены распознающие правила для поиска потенциальных сайтов белков PurR, RbsR и RbsR Pseudomonadales и предсказано 25 новых генов, входящих в обобщенный регулон PurR.
4. Построена модель эволюции, в соответствии с которой PurR и RbsR являются результатом дупликации предкового белка, репрессора рибозного оперона, с последующим изменением функции PurR и структуры мотива связывания RbsR.
5. Построены распознающие правила для поиска потенциальных сайтов связывания факторов транскрипции Fnr, ArcA и NarP, разработана и применена методика анализа комплексной регуляции в родственных геномах. Предсказано 67 новых членов обобщенных регулонов.
6. Предсказаны таксон-специфичные регуляторные каскады, показаны корреляции между перестройкой структуры регуляторных каскадов и размером обобщенного регулона: регуляторы, занимающие относительно более высокое положение в каскаде, имеют большие регулоны.
7. Для глобальной регуляции дыхания показано наличие консервативных ядер регулонов и таксон-специфичной периферической регуляции.

Список публикаций по теме диссертации

Публикации в научных журналах

1. *Равчев Д.А., Гельфанд М.С., Миронов А.А., Рахманинова А.Б.* Пуриновый регулон гамма-протеобактерий. Детальное описание // Генетика. 2002. Т. 38. №9, С. 1203-1214.
2. *Favorov A.V., Gelfand M.S., Gerasimova A.V., Ravcheev D.A., Mironov A.A., Makeev V.J.* A Gibbs sampler for identification of symmetrically structured, spaced DNA motifs with improved estimation of the signal length // Bioinformatics. 2005. Vol. 21. P. 2240-2245.
3. *Равчев Д.А., Рахманинова А.Б., Миронов А.А., Гельфанд М.С.* Регуляция нитрат-нитритного дыхания гамма-протеобактерий. исследование методами сравнительной геномики // Молекулярная биология. 2005. Т. 39. №5. С. 832-846.
4. *Ravcheev D.A., Gerasimova A.V., Mironov A.A., Gelfand M.S.* Comparative genomic analysis of regulation of anaerobic respiration in ten genomes from three families of gamma-proteobacteria (Enterobacteriaceae, Pasteurellaceae, Vibrionaceae) // BMC Genomics. 2007. Vol. 8. Suppl. 1. P. 54.
5. *Цыганова М.О., Гельфанд М.С., Равчев Д.А.* Регуляция дыхания у энтеробактерий: сопоставление данных по экспрессии на биочипах и сравнительно-геномного анализа // Молекулярная биология, 2007. Т. 41. №3. С. 556-571.

Публикации в сборниках трудов конференций

1. *Ravcheev D.A., Gelfand M.S., Mironov A.A., Rakhmaninova A.B.* Purine regulon of gamma-proteobacteria // Proc. 3rd International Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2002). 2002. Vol. 2. P. 38-39.
2. *Равчев Д.А., Рахманинова А.Б.* NsrP-зависимая регуляция нитрат-нитритного дыхания у гамма-протеобактерий. Исследование эволюции дыхательной системы методами сравнительной геномики // Сборник тезисов XI Международной научной конференции студентов, аспирантов и молодых ученых «Ломоносов-2004». 2004. Т. 1. С. 27.
3. *Gerasimova A.V., Ravcheyev D.A., Gelfand M.S., Rakhmaninova A.B.* Comparative genomic analysis of respiration switch in gamma-proteobacteria // Proc. 4th International Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2004). 2004. V.2. P. 195-198.
4. *Равчев Д.А.* Исследование регуляции нитрат-нитритного дыхания гамма-протеобактерий методами сравнительной геномики // Сборник тезисов XII Международной конференции студентов, аспирантов и молодых ученых «Ломоносов». 2005. Т. 2. С. 34-35.
5. *Ravcheev D.A., Gerasimova A.V.* How gamma-proteobacteria switch the mode of respiration: A comparative genomic analysis // Proceedings of the International Moscow Conference on Computational Molecular Biology (MCCMB'05). 2005. P. 321-322.
6. *Tsiganova M., Ravcheev D.A.* Regulation of respiration in *Escherichia coli*: the comparison of microarray and genomics data // Proceedings of the International Moscow Conference on Computational Molecular Biology (MCCMB'05) 2005. P. 401-402.
7. *Zakirzianova V., Ravcheev D.A.* A comparative genomic analysis of an evolving regulatory system: how FruR became Cra // Proceedings of the International Moscow Conference on Computational Molecular Biology (MCCMB'05) 2005. P. 423-424.
8. *Равчев Д.А.* Исследование регуляции нитрат-нитритного дыхания гамма-протеобактерий методами сравнительной геномики // Материалы международной школы «Биоинформатика, геномика, протеомика». 2006. С. 48-51.
9. *Закирзянова В.В., Равчев Д.А.* Исследование обобщенного FruR-регулона гамма-протеобактерий методами сравнительной геномики // Материалы XIII международной конференции студентов, аспирантов и молодых ученых «Ломоносов». 2006. Т. 4. С. 45-46.
10. *Цыганова М.О., Равчев Д.А.* Регуляция дыхания у энтеробактерий: сопоставление данных по анализу экспрессии на биочипах и сравнительно-геномного анализа // Материалы XIII международной конференции студентов, аспирантов и молодых ученых «Ломоносов». 2006. Т. 4. С. 39-40.
11. *Ravcheev D.A.* Regulation of respiration in gamma-proteobacteria: evolution of multiple regulatory systems // Proc. 4th Bertinoro Computational Biology Meeting (BCB 2006).

12. *Tsoy O.V., Zakirzianova W., Ravcheev D.A.* Stories about the evolution of regulators: how FruR became CRA and RbsR became PurR // Proc. 3rd Moscow Conference on Computational Molecular Biology (MCCMB'07). 2007. P. 300.
13. *Цой О.В., Закирзянова В., Равчев Д.А.* Как PurR стал Cra, а RbsR стал PurR: истории из жизни бактериальных факторов транскрипции // Сборник трудов конференции молодых ученых «ИТиС'07». 2007. С. 303-306.
14. *Фаворов А., Гельфанд М., Герасимова А., Равчев Д., Кулаковский И., Миронов А., Макеев В.* Алгоритм SeSiMCMC для поиска участков специфического связывания белков – регуляторов транскрипции // Сборник трудов конференции молодых ученых «ИТиС'07». 2007. С. 334-337.
15. *Равчев Д.А.* Комплексная регуляция дыхания бактерий: исследование методами сравнительной геномики // Сборник трудов конференции молодых ученых и специалистов ИППИ РАН “Информационные технологии и системы-2008”. 2008. С. 273-276.
16. *Цыганова М.О., Равчев Д.А.* Регуляция дыхания бактерий: транскриптомика и сравнительная геномика // Сборник трудов конференции молодых ученых и специалистов ИППИ РАН “Информационные технологии и системы-2008”. 2008, С. 332-336.

Благодарности: Автор выражает искреннюю благодарность Михаилу Сергеевичу Гельфанду за чуткое научное руководство, Андрею Александровичу Миронову и Александру Владимировичу Фаворову за любезно предоставленное программное обеспечение, Анне Викторовне Герасимовой, Ольге Цой, Марине Цыгановой и Виоланте Закирзяновой за помощь в выполнении работы и Александре Борисовне Рахманиновой, Дмитрию Александровичу Родионову и Ольге Николаевне Лайковой за ценные советы и полезное обсуждение.

– ДЛЯ ЗАМЕТОК –

Равчев Дмитрий Андреевич

ИЗУЧЕНИЕ ЭВОЛЮЦИИ РЕГУЛЯТОРНЫХ СИСТЕМ ПРОКАРИОТ МЕТОДАМИ СРАВНИТЕЛЬНО-ГЕНОМНОГО АНАЛИЗА

Работа посвящена исследованию эволюции регуляторных взаимодействий в геномах гамма-протеобактерий. В качестве объектов исследования выбраны регуляция метаболизма сахаров белком FruR, регуляция биосинтеза пуриновых нуклеотидов и утилизации рибозы гомологичными белками PurR и RbsR, соответственно, и глобальная регуляция дыхания белками Fnr, ArcA и NarP.

Показано изменение состава FruR-регулона в различных таксонах, приведены результаты, свидетельствующие об эволюции этого белка от локального регулятора оперона *fruBKA* до глобального регулятора биосинтеза сахаров. Предсказана регуляция 30 новых генов, входящих в FruR-регулон.

Исследована эволюция белков PurR и RbsR, приведены данные, указывающие на их происхождение от общего предка, представлявшего собой регулятор утилизации рибозы, и дальнейшие видоизменения каждой из копий. Предсказаны 27 новых генов, входящие в пуриновый регулон.

Исследована глобальная регуляция дыхания, предсказана регуляция более 150 новых генов. Установлена корреляция между относительной ролью гена в структуре регуляторных каскадов и составом соответствующего регулона. Показано наличие консервативных ядер регулонов и таксон-специфичной периферической регуляции.

Ravcheev Dmitry Andreevich

INVESTIGATION OF EVOLUTION OF THE PROKARYOTIC REGULATORY SYSTEMS USING COMPARATIVE GENOMICS APPROACHES

Evolution of regulatory interactions in gamma-proteobacteria genomes was analyzed. The analysis was performed for several regulatory systems: FruR, the regulator of sugar metabolism flow; homologous proteins PurR and RbsR, the regulators of purine biosynthesis and ribose utilization, respectively; and global regulators of respiration, Fnr, ArcA and NarP.

For FruR regulation, taxon-specific changes in regulon structure were demonstrated. The predicted evolution scenario is that FruR evolved from the repressor of a single *fruBKA* operon to the global regulator of the sugar metabolism. FruR regulation was predicted for 30 new regulon members.

Evolution of homologous regulators PurR and RbsR was studied in detail. These proteins evolved by duplication of a single ancestral gene. PurR regulation was predicted for 27 genes.

Analysis of complex global regulation of respiration resulted in prediction of regulation for more than 150 genes. A set of taxon-specific regulatory cascades was predicted and correlations of the cascade structure and the regulon composition were demonstrated. The regulon members were classified into two groups: the conserved regulon core and taxon-specific periphery.