

На правах рукописи

Суворова Инна Андреевна

**Кoeволюция транскрипционных факторов семейства GNTR
и их сайтов связывания**

03.01.09 - математическая биология, биоинформатика

Автореферат

диссертации на соискание ученой степени
кандидата биологических наук

Москва – 2016

Работа выполнена в Учебно-научном центре «Биоинформатика» Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук.

Научный руководитель:

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

Официальные оппоненты:

доктор биологических наук
Карягина-Жулина Анна Станиславовна
Федеральное государственное бюджетное учреждение
«Федеральный научно-исследовательский центр эпидемиологии и микробиологии
имени почетного академика Н.Ф. Гамалеи» Министерства здравоохранения
Российской Федерации, главный научный сотрудник

кандидат физико-математических наук
Кулаковский Иван Владимирович
Федеральное государственное бюджетное учреждение науки
Институт молекулярной биологии им. В.А. Энгельгардта
Российской академии наук, старший научный сотрудник

Ведущая организация:

Федеральное государственное бюджетное учреждение науки
Институт общей генетики им. Н.И. Вавилова
Российской академии наук

Защита диссертации состоится «___» _____ 2016 года в ___ на заседании Диссертационного Совета Д.002.077.04 при Федеральном государственном бюджетном учреждении науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук по адресу: 127051, г. Москва, Большой Каретный переулок, д. 19, стр. 1.

С диссертацией можно ознакомиться в библиотеке Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук, а также на сайте <http://www.iitp.ru>.

Автореферат диссертации разослан «__» _____ 2016 года.

Ученый секретарь диссертационного совета Д.002.077.04
доктор биологических наук, профессор
Рожкова Г.И.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы

Бактерии способны приспосабливаться к самым разным, меняющимся условиям окружающей среды. Подобная адаптация осуществляется за счет изменения экспрессии генов, что позволяет клетке эффективно использовать имеющиеся ресурсы. Такая стратегия требует сложной системы регуляции, обеспечивающей адекватный ответ на внешние или внутриклеточные стимулы. Регуляция экспрессии генов осуществляется на разных уровнях: транскрипции, трансляции, посттрансляционной модификации, однако наиболее эффективным и распространенным вариантом является регуляция на стадии инициации транскрипции. Ключевым элементом такой регуляции – факторы транскрипции, специальные белки-регуляторы.

До недавнего времени исследование транскрипции проводилось исключительно экспериментальными методами, но в настоящее время развитие методов секвенирования и экспоненциальный рост количества данных о нуклеотидных и аминокислотных последовательностях привели к широкому и успешному использованию биоинформатических методов. Подобные исследования часто применяются в качестве дополнения к эксперименту, однако изучение регуляции может осуществляться и исключительно методами сравнительной геномики. Основной задачей биоинформатических исследований является выявление разнообразных регуляторных последовательностей, например, промоторов, сайтов связывания транскрипционных факторов и т.д.

Роль регуляторных взаимодействий весьма велика, и сравнительный анализ регуляции экспрессии генов у различных бактерий позволяет делать выводы об эволюции отдельных функциональных систем и самих микроорганизмов, а также особенностях их взаимодействия с окружающей средой. Таким образом, исследование ДНК-белковых взаимодействий и регуляции транскрипции является актуальной задачей современной молекулярной биологии и сравнительной геномики.

Цели и задачи исследования

Целью данной работы было исследование одного из наиболее распространенных среди бактерий семейств транскрипционных факторов, GNTR, методами сравнительной геномики.

В работе решаются следующие общие и частные задачи:

1. Реконструкция регулонов транскрипционных факторов семейства GNTR методами сравнительной геномики, построение распознающих правил для поиска их потенциальных сайтов связывания на основании результатов исследования 5'-регуляторных областей.

2. Исследование коэволюции мотивов связывания и аминокислотных последовательностей регуляторов транскрипции подсемейств FADR, HUTC и YTRA семейства GNTR путем анализа

корреляций аминокислот ДНК-связывающих НТН-доменов транскрипционных факторов и нуклеотидов соответствующих сайтов связывания, предсказание вероятных ДНК-белковых взаимодействий.

3. Анализ особенностей структуры и расположения сайтов связывания регуляторов семейства GNTR – исследование дивергонов, а также дополнительных боксов (полусайтов, симметричных элементов палиндромного мотива) у сайтов связывания.

4. Исследование регуляции метаболизма гексуронатов у *Gamma*proteobacteria родственными транскрипционными факторами UxuR и ExuR методами сравнительной геномики, разделение их мотивов связывания и построение распознающих правил для предсказания сайтов связывания, реконструкция гексуронатных регулонов, исследование оперонной структуры и идентификация новых членов регулонов, построение вероятных сценариев эволюции этой метаболической системы.

5. Исследование регуляции метаболизма малоната и пропионата у *Proteobacteria* транскрипционными факторами MatR/MdcY, MdcR, PrpR методами сравнительной геномики, выявление новых регуляторов метаболизма малоната и пропионата, построение распознающих правил для предсказания сайтов связывания и реконструкция соответствующих регулонов, исследование оперонной структуры и идентификация новых членов регулонов, построение возможной модели эволюции этих метаболических систем.

Научная новизна и практическое значение работы

В работе впервые исследован целый ряд транскрипционных факторов семейства GNTR в различных таксономических группах, предсказаны их потенциальные сайты связывания и ДНК-белковые взаимодействия, реконструированы регулоны. Кроме того, обобщены сведения о расположении и структуре сайтов связывания. Также было проведено детальное исследование регуляции метаболизма гексуронатов у *Gamma*proteobacteria и метаболизма малоната и пропионата у *Proteobacteria*. С помощью методов сравнительной геномики были обнаружены ранее неизвестные члены регулонов этих метаболических систем, показана вариабельность организации регулонов и их регуляции, в частности, были идентифицированы новые регуляторы метаболизма малоната и пропионата и выявлены их потенциальные сайты связывания. Кроме того, в работе были предложены потенциальные сценарии эволюции регулонов метаболизма гексуронатов, а также малоната и пропионата. Работа имеет теоретический характер, однако полученные данные потенциально могут найти применение в области биотехнологии и генной инженерии.

Апробация работы

Основные положения диссертации были представлены на российских и международных конференциях: Информационные технологии и системы ИТиС'09 (Бекасово, декабрь 2009), Ломоносов-2010 (Москва, апрель 2010), Информационные технологии и системы ИТиС'10 (Геленджик, сентябрь 2010), Постгеномные методы анализа в биологии, лабораторной и клинической медицине (Москва, ноябрь 2010), Информационные технологии и системы ИТиС'11 (Геленджик, октябрь 2011), Молекулярная и клеточная биология: прикладные аспекты (Москва, апрель 2012), Информационные технологии и системы ИТиС'12 (Петрозаводск, август 2012), Moscow Conference on Computational Molecular Biology MCCMB'13 (Москва, июль 2013), и на научных встречах международной учебно-научной группы «Regulation and Evolution of Cellular Systems (RECESS)» (Мюнхен, Германия, май 2011; Москва, июнь 2012; Венеция, Италия, май 2013).

Объем и структура диссертации

Диссертационная работа изложена на 132 страницах и состоит из введения, пяти глав, выводов и списка цитированной литературы. Глава 1 содержит обзор литературы по теме диссертации. Глава 2 содержит описание используемых методов и программ. Главы с 3 по 5 содержат описание собственных исследований. Список литературы включает 259 наименований. Работа содержит 18 рисунков, 10 таблиц и 4 приложения.

СОДЕРЖАНИЕ РАБОТЫ

В первой главе приведен обзор литературы по теме диссертации. Обзор содержит рассмотрение общих принципов регуляции транскрипции у прокариот: приводятся данные о строении РНК-полимеразы и промоторов, стадиях транскрипции, оперонной организации генов, регуляторных структурах РНК, основных семействах факторов транскрипции и механизмах репрессии и активации экспрессии генов. Также рассмотрены основные методы сравнительной геномики, способы биоинформатического предсказания функций генов, поиска потенциальных регуляторных сайтов и изучения регуляции транскрипции. Дальнейшая часть обзора литературы посвящена ДНК-белковым взаимодействиям, общей характеристике транскрипционных факторов семейства GNTR и их мотивов связывания, а также описанию некоторых метаболических систем, регулируемых транскрипционными факторами семейства GNTR. В качестве примера были выбраны системы утилизации гексуронатов у Gammaproteobacteria, а также малоната и пропионата у Proteobacteria.

Во второй главе описаны программы и методы, использованные для выявления регуляторных взаимодействий и реконструкции регулонов, анализа корреляций и предсказания ДНК-белковых контактов, а также визуализации и статистического анализа полученных результатов.

Последовательности геномов 307 исследованных бактерий были взяты из базы данных GenBank [Benson et al., 2012]. Гомологи исследованных факторов транскрипции были идентифицированы при помощи программы PSI-BLAST [Altschul et al., 1997], ортологи определялись с помощью построения филогенетических деревьев, а также с учетом геномного контекста. Для выравнивания нуклеотидных и аминокислотных последовательностей использовалась программа MUSCLE [Edgar, 2004]. Построение и визуализация филогенетических деревьев осуществлялась с помощью программ пакета PHYLIP [Felsenstein, 1996] и программы Dendroscope [Huson et al., 2007], соответственно.

Для каждого из исследованных транскрипционных факторов была проведена реконструкция регулонов: были идентифицированы потенциальные регулируемые гены и сайты связывания. Потенциальные сайты связывания идентифицировались методом филогенетического футпринтинга [Rodionov, 2007]. Поиск сайтов связывания в геномах осуществлялся при помощи матриц позиционных весов нуклеотидов (PWM) [Gelfand et al., 2000]. Построение PWM для мотивов связывания всех исследованных транскрипционных факторов проводилось с помощью программы SignalX, для поиска ортологов и потенциальных сайтов использовался пакет программ GenomeExplorer [Mironov et al., 2000], а также веб-сервер RegPredict [Novichkov et al., 2010]. Поиск сайтов связывания транскрипционных факторов проводился в области от -400 до +50 нуклеотидов относительно старта трансляции. Диаграммы Logo мотивов связывания были построены при помощи программы WebLogo [Crooks et al., 2004]. Для подтверждения принадлежности определенного гена к регулону применялся метод проверки соответствия [Mironov et al., 1999; Gelfand et al., 2000]. Проводилось также предсказание оперонной структуры генов, перед которыми найдены сайты связывания: гены относили к одному оперону, если они транскрибировались в одном направлении, межгенное расстояние не превышало 200 нуклеотидов, и подобная организация сохранялась в ряде геномов. Реконструированные регулоны размещены в базе данных RegPrecise [Novichkov et al., 2010].

Анализ корреляций аминокислот ДНК-связывающих НТН-доменов транскрипционных факторов семейства GNTR и нуклеотидов соответствующих сайтов связывания осуществлялся с помощью программы Prot-DNA-Korr (<http://bioinf.fbb.msu.ru/Prot-DNA-Korr/main.html>) отдельно для каждого из исследованных подсемейств (FADR, HUTC и YTRA).

Статистический анализ данных для дивергонов и дополнительных боксов мотивов связывания проводился при помощи программы STATISTICA [Hill et al., 2007].

Третья глава включает общие сведения об исследованных транскрипционных факторах подсемейств FADR, HUTC и YTRA семейства GNTR (таксономическое распределение, число регулируемых оперонов и сайтов связывания и т.п.), данные о корреляциях аминокислот НТН-доменов транскрипционных факторов и нуклеотидов соответствующих сайтов связывания, а также исследование особенностей структуры мотивов связывания и расположения сайтов в дивергонах.

GNTR – широко распространенное среди различных таксономических групп бактерий семейство транскрипционных факторов, регулирующих самые разные биологические процессы [Rigali et al., 2002]. Регуляторы семейства GNTR содержат консервативный N-концевой ДНК-связывающий НТН-домен, однако различаются структурой С-концевых доменов, осуществляющих олигомеризацию и связывание эффекторов, в соответствии с чем в семействе выделяют шесть подсемейств (FADR, HUTC, MOCR и YTRA, а также минорные подсемейства ARAR и PLMA) [Rigali et al., 2004]. В настоящей работе были исследованы транскрипционные факторы подсемейств FADR, HUTC и YTRA. Сайты связывания были предсказаны для 1252 транскрипционных факторов семейства GNTR (64 ортологические группы) из 307 бактериальных геномов. Содержание исследованных транскрипционных факторов в индивидуальных геномах и разных таксонах варьирует (Таблица 1). Например, регуляторы подсемейства YTRA распространены преимущественно среди представителей Firmicutes, тогда как регуляторы подсемейства FADR типичны для Proteobacteria.

Таблица 1. Исследованные факторы транскрипции семейства GNTR. Общие данные

Количество/Подсемейство		FADR	HUTC	YTRA
Ортологические группы		36	16	12
Исследованные транскрипционные факторы		634	389	229
Регулируемые опероны		1740	975	283
Сайты (включая дивергентные и множественные)		2396	1341	294
Таксономическое распределение исследованных транскрипционных факторов				
Proteobacteria	Alpha	76	39	3
	Beta	151	64	0
	Gamma	308	112	25
	Delta	10	1	0
Firmicutes	Bacilli	18	97	89
	Clostridia	1	14	53
Actinobacteria		64	60	43
Thermotogae		0	0	14
Chloroflexi		6	0	1
Bacteroidetes		0	1	0
Cyanobacteria		0	1	0
Archaea		0	0	1

С целью предсказания вероятных ДНК-белковых взаимодействий, для транскрипционных факторов подсемейств FADR, HUTC и YTRA был проведен анализ корреляций аминокислот ДНК-связывающих НТН-доменов и нуклеотидов соответствующих сайтов связывания.

Анализ корреляций показал (Рисунок 1), что для подсемейства FADR в целом скоррелированные нуклеотидные и аминокислотные позиции, вероятно, определяющие специфичность связывания, хорошо соответствуют парам контактирующих позиций для известной ДНК-белковой структуры FadR *E.coli* и AraR *B.subtilis*.

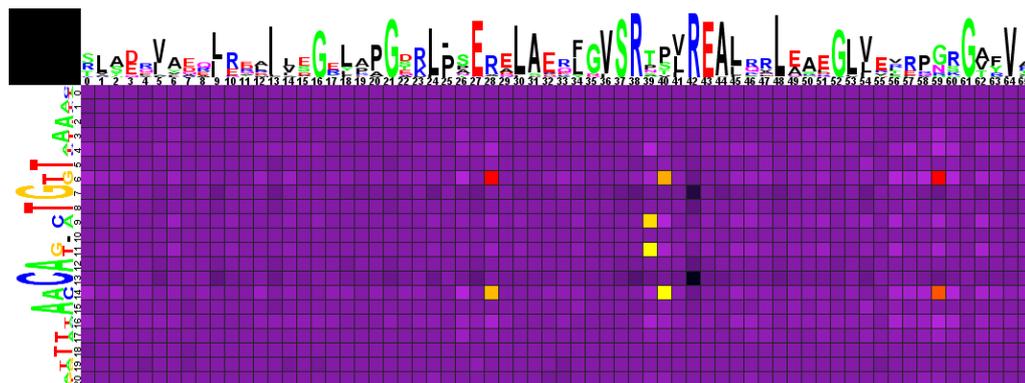


Рисунок 1. Карта интенсивности корреляций аминокислот НТН-доменов транскрипционных факторов подсемейства FADR и нуклеотидов соответствующих сайтов связывания

Диаграммы Logo ДНК-связывающих НТН-доменов и сайтов связывания показаны, соответственно, сверху и слева от карты интенсивности. Общая высота символов в каждой позиции соответствует информационному содержанию, тогда как высота конкретного символа пропорциональна частоте встречаемости аминокислоты/нуклеотида в данной позиции. Уровень корреляции показан цветом и изменяется по градиенту от желтого до красного для статистически значимо (выше автоматически определяемого порога) коррелирующих пар аминокислот и нуклеотидов; прочие пары показаны фиолетово-черным.

Для аминокислот в позиции 28 НТН-домена, формирующих один из контактов FadR *E.coli* с ДНК, показана корреляция с нуклеотидами в позициях 6/14. Вероятными контактами в этой позиции являются Arg-G и Asp-C. С нуклеотидами 6/14 коррелируют также аминокислоты в позициях 40 и 59, которые являются важными для взаимодействия с ДНК у FadR *E. coli* и AraR *B. subtilis*. Наиболее часто встречающиеся в позиции 40 аминокислоты – пролин и серин. Серин ассоциирован с наличием G/C пары (предположительно, образуя контакт с G), тогда как в случае пролина G/C пара достоверно избегается. Глицин в позиции 59 коррелирует с наличием G/C пары, при этом A/T пара достоверно избегается, однако эта корреляция может быть обусловлена стерическими причинами и не отражать непосредственное ДНК-белковое взаимодействие. Кроме

того, аминокислоты в позиции 39 НТН-домена, для которых показано участие во взаимодействии FadR *E. coli* и AraR *B. subtilis* с ДНК, коррелируют с центральными нуклеотидами 9/11. Аспарагин в данной позиции достоверно коррелирует с А/Т парой, вероятно, взаимодействуя с А.

Анализ корреляций (Рисунок 2) показывает, что в подсемействе НУТС позиции, определяющие специфичность связывания, сходны с таковыми для FadR *E. coli* и для подсемейства FADR в целом. Как и в подсемействе FADR, для аминокислот в позиции 28 показана корреляция с нуклеотидами 8/17 – аргинин значимо коррелирует с G/C парой (вероятный контакт Arg-G), тогда как А/Т пара достоверно избегается. С нуклеотидами 8/17 также коррелируют и аминокислоты в позициях 43 и 62, участвующие в связывании с ДНК у FadR *E. coli* (вероятный контакт Trp-C). Кроме того, аминокислоты в позиции 39, как это было показано и для подсемейства FADR, коррелируют с центральными нуклеотидами 12/13, вероятным контактом здесь является Asp-C.

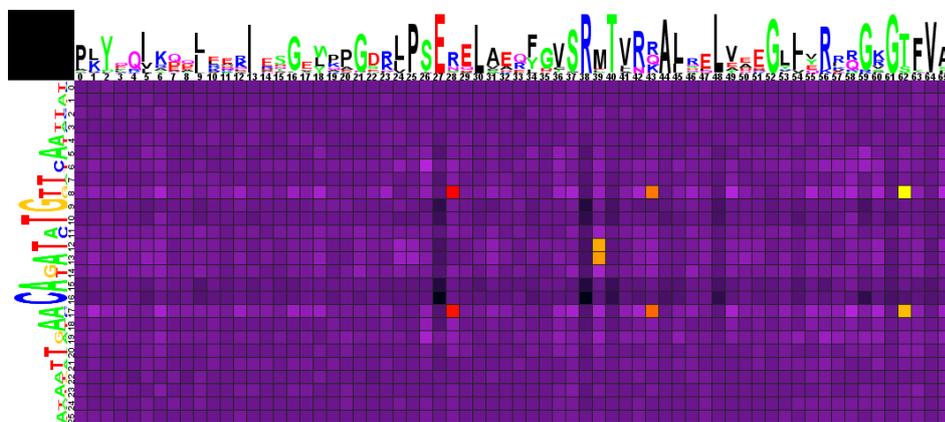


Рисунок 2. Карта интенсивности корреляций аминокислот НТН-доменов транскрипционных факторов подсемейства НУТС и нуклеотидов соответствующих сайтов связывания

Обозначения как на Рисунке 1.

Подсемейство УТРА имеет ряд отличий от остальных исследованных подсемейств семейства GNTR: крайне редко характерная для подсемейств FADR и НУТС дивергентная организация регулируемых оперонов; типичны одиночные сайты связывания, а двойные и тройные сайты, распространенные среди других подсемейств, идентифицированы всего в нескольких случаях. Мотивы связывания транскрипционных факторов подсемейства УТРА существенно длиннее, чем мотивы регуляторов остальных подсемейств семейства GNTR. Несмотря на вышеупомянутые отличия, высокая консервативность НТН-доменов внутри всего семейства позволяет точно сопоставить их для транскрипционных факторов различных подсемейств, и проведенный корреляционный анализ показывает, что позиции аминокислот в составе НТН-домена, определяющие специфичность связывания с ДНК, сходны у всех исследованных подсемейств, в

том числе и УТРА (Рисунок 3). Так, нуклеотиды 12-13/29-30, вероятно, участвуют в специфических взаимодействиях с аминокислотными остатками в позициях 27 и 28. Аспарагин и тирозин в позиции 28 достоверно ассоциированы с наличием пары А/Т в позициях 12/30 (вероятные контакты – Asn-A и Tyr-A). В позиции 27 треонин достоверно коррелирует с А/Т парой в позициях 12/30 и 13, предположительно формируя контакт Thr-A. Корреляции показаны для пар нуклеотидов 16-17/25-26 и аминокислотных остатков в позициях 37 и 39, важных для взаимодействия с ДНК у FadR *E.coli*. Серин в позиции 37 и изолейцин в позиции 39 достоверно коррелируют с А/Т в позициях 16-17/25-26, тогда как гистидин в позиции 39 достоверно ассоциирован с G/C парой в позициях 25 и 26. В последнем случае, контактом, вероятно, является His-G. Кроме того, корреляции отмечены для аланина в позиции 39 с А/Т в позициях 12,13 и G/C в позиции 24, а также для глицина в позиции 44 с А/Т в позиции 13.

Результаты анализа корреляций показывают, что предсказанные ДНК-белковые взаимодействия для всех трех исследованных подсемейств регуляторов семейства GNTR хорошо соотносятся с ДНК-белковыми контактами, известными для FadR *E. coli* и AraR *B. subtilis* [van Aalten et al., 2001; Xu et al., 2001; Jain et al., 2013]. Кроме того, большая часть предсказанных контактов (Arg-G, Asn-A, Asp-C, Gly-G, His-G, Trp-C) согласуется с ранее описанными закономерностями взаимодействия [Marabotti et al., 2008].

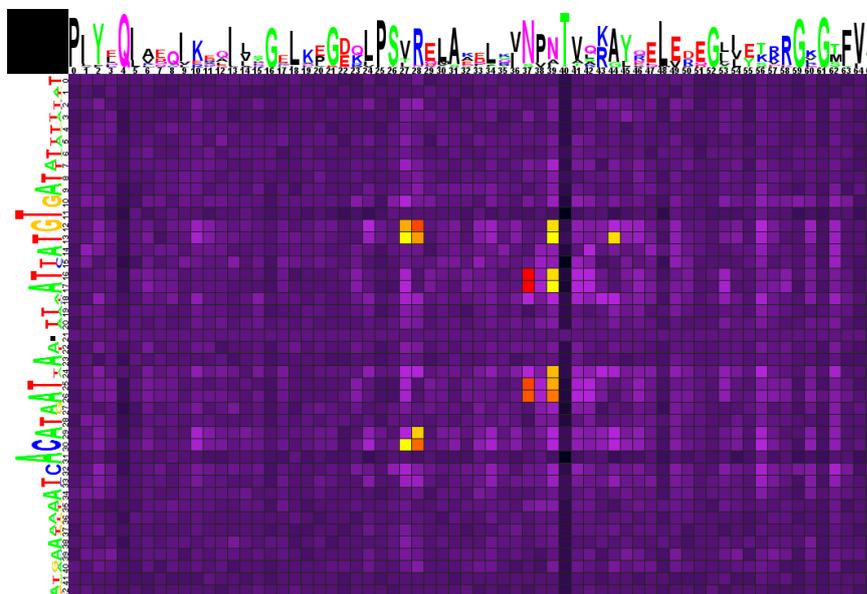


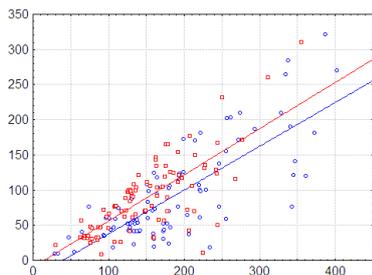
Рисунок 3. Карта интенсивности корреляций аминокислот НТН-доменов транскрипционных факторов подсемейства УТРА и нуклеотидов соответствующих сайтов связывания

Обозначения как на Рисунке 1.

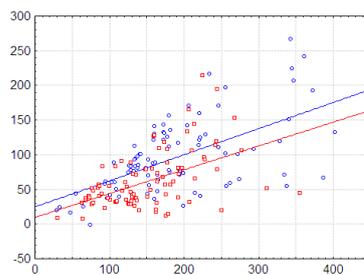
В настоящей работе были также исследованы регулируемые дивергоны двух подсемейств семейства GNTR (FADR и HUTC, подсемейство YTRA не представлено в связи с практически полным отсутствием в регулонах данной группы дивергентно регулируемых оперонов). Найденные дивергоны были разделены на две группы: дивергоны, имеющие в составе ген транскрипционного фактора, и дивергоны, полностью состоящие из структурных генов (контрольная группа). Дивергоны были также разделены по числу сайтов связывания в межгенной области (с 1 или 2 сайтами).

В случае дивергонов с единичным сайтом связывания (с геном транскрипционного фактора в составе) как FADR ($n = 96$), так и HUTC-подсемейств ($n = 94$) для обоих оперонов дивергона наблюдается тенденция к линейному росту расстояния от старта гена до середины сайта с увеличением размера межгенной области (Рисунок 4А, 4Б). Аналогичная тенденция наблюдается также и для дивергонов контрольной группы (FADR, $n = 33$; HUTC, $n = 23$) (Рисунок 4В).

А. Опероны с геном фактора транскрипции



Б. Структурные опероны



В. Контроль

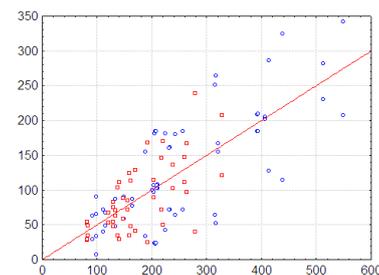


Рисунок 4. Зависимость расстояния между стартом гена и сайтом связывания от размера межгенной области в дивергонах с единичным сайтом

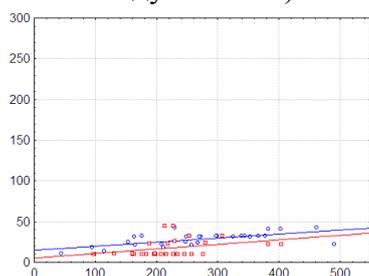
Межгенное расстояние показано на оси X. Расстояние между центром сайта и старт-кодоном гена – на оси Y. Показаны линии регрессии. Синим цветом обозначены данные для подсемейства FADR, красным – для подсемейства HUTC.

Таким образом, единичные сайты обычно локализируются в центре межгенной области и, вероятно, участвуют в регуляции обоих оперонов дивергона. При этом в дивергонах, содержащих ген фактора транскрипции, с ростом размера межгенной области расстояние между сайтом связывания и структурным опероном возрастает медленнее, чем расстояние между сайтом и опероном с геном регулятора в составе (Рисунок 4А, 4Б). Это может быть следствием того, что авторегуляция генов транскрипционных факторов выражена слабее, чем регуляция структурных генов.

В обоих подсемействах FADR ($n = 100$) и HUTC ($n = 60$) наблюдаются две группы дивергонов с двойными сайтами связывания, содержащих ген транскрипционного фактора (Рисунок 5).

Первая группа включает дивергоны (FADR, $n = 29$; HUTC, $n = 32$), в которых расстояние между двойными сайтами относительно неизменно вне зависимости от размеров дивергона (Рисунок 5А). В этой группе расстояние от старта гена до проксимального сайта возрастает с увеличением размера межгенной области как для структурных оперонов, так и для оперонов с геном фактора транскрипции в составе (Рисунок 6). Таким образом, двойные сайты в данной группе дивергонов обычно локализованы в центре межгенной области, и можно предполагать, что они образуют пару, кооперативно регулирующую транскрипцию обоих дивергентных оперонов.

А. Первая группа (постоянное расстояние между сайтами)



Б. Вторая группа (возрастающее расстояние между сайтами)

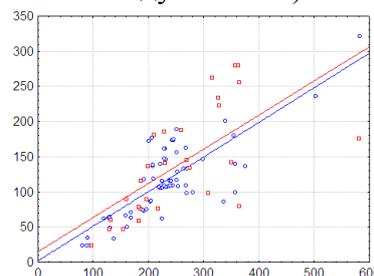
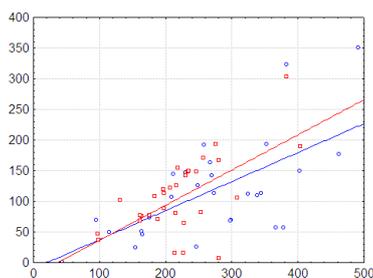


Рисунок 5. Две группы дивергонов с двойными сайтами, зависимость расстояния между сайтами связывания от размера межгенной области

Межгенное расстояние показано на оси X. Расстояние между центрами сайтов – на оси Y.

Обозначения как на Рисунке 4.

А. Опероны с геном фактора транскрипции



Б. Структурные опероны

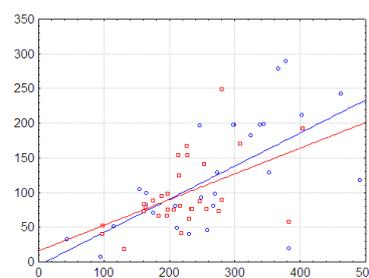


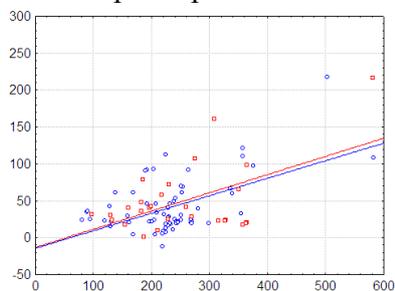
Рисунок 6. Зависимость расстояния между стартом гена и проксимальным сайтом связывания от размера межгенной области в дивергонах с общими двойными сайтами

Межгенное расстояние показано на оси X. Расстояние между центром проксимального сайта и старт-кодоном гена – на оси Y. Обозначения как на Рисунке 4.

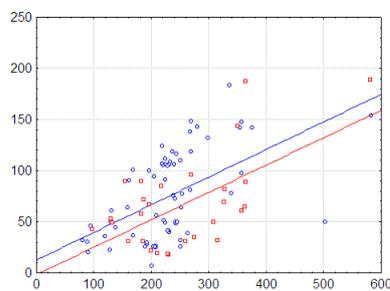
Вторая группа включает дивергоны (FADR, $n = 71$; HUTC, $n = 28$), где расстояние между парой сайтов линейно возрастает с увеличением размеров межгенной области дивергона (Рисунок 5Б). Для обоих оперонов дивергона наблюдается также тенденция к увеличению расстояния от старта гена до проксимального сайта с ростом межгенного расстояния, однако эта зависимость

существенно менее выражена, нежели в случае дивергонов первой группы (Рисунок 7А, 7Б). Следовательно, в данном случае мы не наблюдаем кооперативной регуляции, эти сайты предположительно независимо регулируют каждую из частей дивергона. Аналогичная тенденция была отмечена и для контрольной группы дивергонов (FADR, n = 46; HUTC, n = 19) в обоих подсемействах (Рисунок 7В). В контроле наблюдается только один тип дивергонов, где сайты независимы и принадлежат ближайшему из оперонов дивергона, осуществляя регуляцию его транскрипции. Таким образом, двойные сайты в дивергонах могут быть классифицированы на кооперативные и оперон-специфические, однако функциональное отличие дивергонов обоих типов не было определено, так как между ними не наблюдается существенного различия генного состава.

А. Опероны с геном фактора транскрипции



Б. Структурные опероны



В. Контроль

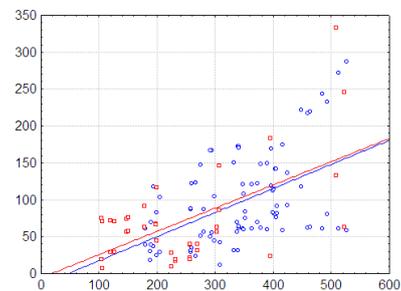


Рисунок 7. Зависимость расстояния между стартом гена и проксимальным сайтом связывания от размера межгенной области в дивергонах с независимыми двойными сайтами

Межгенное расстояние показано на оси X. Расстояние между центром проксимального сайта и старт-кодом гена – на оси Y. Обозначения как на Рисунке 4.

Для всех исследованных в настоящей работе факторов транскрипции семейства GNTB был также проведен анализ областей, фланкирующих идентифицированные сайты связывания. В 23 ортологических группах (13 групп, 170 регуляторов и 450 сайтов связывания в подсемействе FADR; 4 группы, 186 регуляторов и 514 сайтов в подсемействе HUTC, а также 6 групп, 120 регуляторов и 167 сайтов в подсемействе YTRA) было отмечено наличие слабых дополнительных полусайтов (боксов) с одной или с обеих сторон от основного сайта, на расстоянии 7-12 нуклеотидов (нт) от центра сайта.

С целью оценить статистическую значимость этого наблюдения, было проведено сравнение дополнительных боксов с истинными сайтами, а также, в качестве контроля, со случайными последовательностями соответствующей длины (псевдобоксы). Каждый из сайтов сравнивался с двумя случайными последовательностями в 5'-регулируемой области анализируемых генов, на расстоянии 20 и 21 нт от начала истинного сайта, соответственно. Вес для каждой половины

истинного палиндромного сайта был рассчитан при помощи соответствующих половин матриц позиционных весов нуклеотидов (PWM) для данного транскрипционного фактора (полученные значения – $W_{true\ left}$ и $W_{true\ right}$, соответственно). Соответствующие PWM были также использованы для расчета весов дополнительных боксов ($W_{near\ left}$ и $W_{near\ right}$, соответственно, слева и справа) и псевдобоксов, при этом вес для каждого из псевдобоксов был рассчитан дважды, с помощью каждой из половин PWM ($W_{random\ left1,2}$ и $W_{random\ right1,2}$).

После расчета весов осуществлялся отбор данных: из каждой пары дополнительных боксов для дальнейшего анализа выбирался один с наибольшим весом ($W_{near\ left}$ или $W_{near\ right}$). Каждый дополнительный бокс сравнивался с одним из контрольных псевдобоксов: из двух псевдобоксов, вес которых был рассчитан по той же части PWM, что и вес выбранного дополнительного бокса ($W_{random\ left1,2}$ или $W_{random\ right1,2}$), выбирали один с наибольшим весом (например, $W_{random\ left1}$ или $W_{random\ left2}$).

Мотивы связывания транскрипционных факторов из разных ортологических групп различались по длине и структуре, следовательно, различались также и соответствующие PWM. Поэтому для сравнения данных по каждому из подсемейств в целом рассчитанные веса дополнительных боксов и псевдобоксов были нормированы на значения весов соответствующих истинных полусайтов ($W_{true\ left}$ или $W_{true\ right}$), ориентированных так же, как и выбранный для анализа дополнительный бокс.

Данные для каждого из подсемейств GNTR были нормированы по следующим формулам:

$$S_{near} = \frac{W_{true} - W_{near}}{W_{true}} \quad (1); \quad S_{random} = \frac{W_{true} - W_{random}}{W_{true}} \quad (2),$$

– где S_{near} и S_{random} – нормированные веса для дополнительных боксов и псевдобоксов, соответственно; W_{true} , W_{near} , W_{random} – рассчитанные по PWM веса выбранных для анализа истинных полусайтов, дополнительных боксов и псевдобоксов, соответственно.

Распределение значений S_{near} и S_{random} (Рисунок 8) статистически значимо различалось для всех трех подсемейств FADR, HUTC и YTRA (парный критерий Вилкоксона, $p < 0,001$). При этом среднее значение веса для дополнительных боксов W_{near} приблизительно соответствует половине веса истинного полусайта W_{true} , тогда как среднее значение W_{random} близко к нулю, что свидетельствует о правильности выбранного контроля. Возможное участие данных дополнительных боксов в процессе регуляции транскрипции представляется интересным объектом дальнейших экспериментальных исследований. Эти дополнительные боксы, предположительно, могут участвовать в альтернативной димеризации транскрипционных факторов или же в связывании дополнительных субъединиц, позволяя кооперативную регуляцию и более тонкий контроль экспрессии.

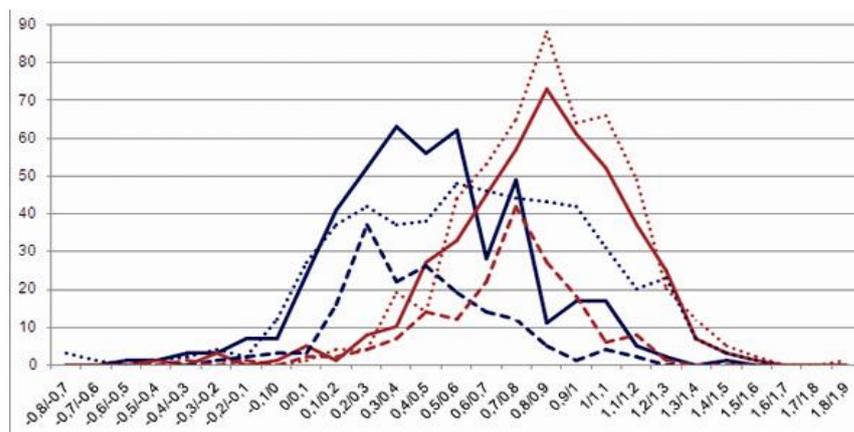


Рисунок 8. Распределение значений S_{near} и S_{random} для подсемейств FADR, HUTC и YTRA

Интервалы значений S показаны на оси X. Количество значений S в соответствующем интервале – на оси Y. Синим цветом показаны распределения значений S_{near} , красным – S_{random} . Данные для подсемейства FADR обозначены непрерывными линиями, для подсемейства HUTC – точечным пунктиром, для подсемейства YTRA – пунктирными линиями.

Четвертая глава посвящена исследованию регуляции метаболизма гексуронатов у Gammaproteobacteria и включает выявление мотивов связывания соответствующих регуляторов, реконструкцию гексуронатных регулонов, а также построение вероятных сценариев эволюции этой метаболической системы.

E. coli может использовать в качестве источника углерода гексуронаты, которые поступают в клетку при помощи двух транспортеров – GntP и ExuT [Bates Utz et al., 2004]. Кроме того, у *E. coli* имеется транспортер β -D-глюкуронидов UidB и глюкуронидаза UidA, также обеспечивающие поступление глюкуроната в клетку [Blanco et al., 1986; Rodionov et al., 2000]. И D-глюкуронат, и D-галактуронат метаболизируются до 2-кето-3-деокси-D-глюконата (KDG), соответствующие метаболические пути включают D-глюкуронат/D-галактуронат изомеразу (UxaC), а также две пары аналогичных по функциям ферментов – D-маннонат и D-альтронат гидролазы UxuA и UxaA и оксидоредуктазы UxuB и UxaB, соответственно (Рисунок 9). Метаболизм гексуронатов у *E. coli* регулируется двумя родственными белками подсемейства FADR семейства GNTR – UxuR и ExuR. UxuR осуществляет контроль метаболизма D-глюкуроната, репрессируя опероны *uxuAB*, *uidABC*, *gntP* и *uxuR*, тогда как ExuR осуществляет репрессию генов, участвующих в метаболизме как D-галактуроната, так и D-глюкуроната, включая *exuT*, *uxaCA*, *uxaB*, *uxuAB* и *exuR* [Mata-Gilsinger et al., 1983; Portalier et al., 1980; Robert-Baudouy et al., 1981; Rodionov et al., 2000].

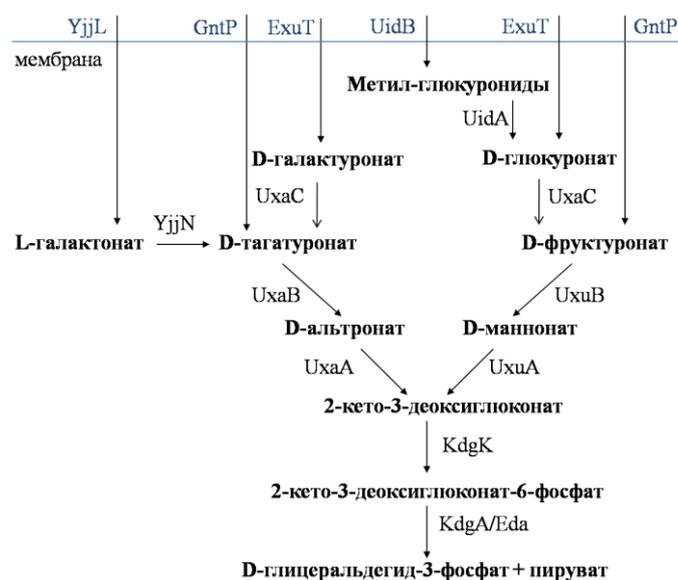


Рисунок 9. Схема метаболизма гексуронатов у *E. coli*

Транспортные белки показаны голубым цветом, ферменты – черным.

Ортологи UxuR и ExuR были обнаружены только у представителей Gammaproteobacteria. При этом ExuR присутствует только у представителей Enterobacteriales (а также *Photobacterium profundum* среди Vibrionales), тогда как UxuR характерен для целого ряда порядков Gammaproteobacteria (Alteromonadales, Enterobacteriales, Oceanospirillales, Pasteurellales и Vibrionales). Одним из возможных эволюционных сценариев является наличие у предковых форм регулонов с обоими транскрипционными факторами, с последующей потерей ExuR во всех группах, кроме Enterobacteriales и *P. profundum*. Альтернативный вариант предполагает дубликацию исходного регуляторного гена у общего предка Enterobacteriales, с последующей быстрой эволюцией ExuR и горизонтальным переносом этого регулятора к *P. profundum*.

Специфические профили (PWM) для UxuR и ExuR были построены по сайтам связывания известных членов соответствующих регулонов *E. coli* и их ортологов в геномах родственных Gammaproteobacteria. В данной работе было показано, что мотивы связывания UxuR и ExuR отличаются структурой 3'-участка, а также длиной центрального спейсера между группами GT/AC двух половин палиндромного сайта. Мотив связывания ExuR не точно палиндромный в позициях 6/14, 7/13 и 9/11 (Рисунок 10), что дает возможность кросс-узнавания одной и той же области обоими регуляторами ExuR и UxuR (показано пунктиром). Таким образом, в результате дубликации и последующей эволюции ExuR, вероятно, приобрел новую специфичность взаимодействия с ДНК.

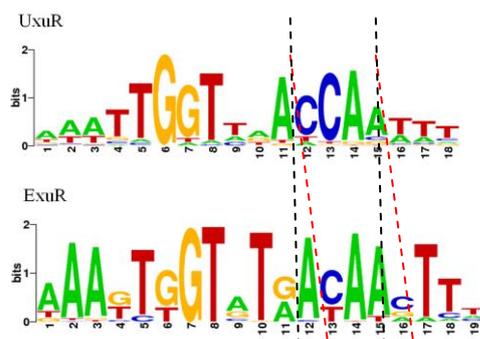


Рисунок 10. Диаграмма Logo мотивов связывания UxuR и ExuR

По горизонтальной оси – позиция нуклеотида в сайте; по вертикальной оси – информационное содержание в битах. Высота каждого столбца пропорциональна информационному содержанию данной позиции; высота каждого символа отражает частоту его встречаемости в данной позиции.

В целом организация регулонов UxuR и ExuR у Gammaproteobacteria разнообразна и лабильна. У большинства исследованных Enterobacteriales в геноме присутствуют как *ихи*, так и *иха* гены, таким образом, они способны метаболизировать и D-глюкуронат, и D-галактуронат. При этом у *Escherichia* spp., *Pectobacterium atrosepticum*, *Shigella* spp. и *Yersinia* spp. обнаружены как UxuR, так и ExuR, контролирующие метаболические пути утилизации D-глюкуроната и D-галактуроната, соответственно. В то же время, *Citrobacter koseri*, *Enterobacter* sp. 638, *Klebsiella pneumoniae* и *Serratia proteamaculans* имеют только транскрипционный фактор ExuR, регулирующий оба гексуронатных катаболических пути.

Кроме того, в этой таксономической группе *Photobacterium luminescens* и *Salmonella* spp. имеют только путь утилизации D-глюкуроната, находящийся под негативным контролем UxuR. Структура регулона UxuR у указанных бактерий отличается от таковой у остальных исследованных Enterobacteriales: у *P. luminescens* гексуронатный регулон включает гены *ихиAB*, *ихиPQM*, *ихиR*, *ихаC*, *kdgK* и *eda* (аналогичный состав регулон UxuR имеет у Vibrionales, см. далее), а у *Salmonella* spp. – гены *ихиAB*, *ихиR*, *ихаC* и *ехиТ* (при этом ген регулятора UxuR располагается в отдельном локусе).

Для большинства представителей Enterobacteriaceae характерно наличие множества паралогов генов D-маннонат оксидоредуктазы (*ихиВ*, *yeiQ* и *ydfI*), при этом в 5'-регуляторной области *yeiQ* у *E. coli*, *Salmonella enterica* Typhimurium LT2 и *Shigella* spp. были идентифицированы предполагаемые сайты связывания UxuR.

В отличие от бактерий порядка Enterobacteriales, для прочих представителей Gammaproteobacteria из порядков Alteromonadales, Oceanospirillales, Pasteurellales и Vibrionales характерно наличие только репрессора UxuR, контролирующего метаболизм D-глюкуроната. Исключением является *P. profundum*, у которого, вероятно, в результате горизонтального переноса

генов от представителей Pasteurellales, в геноме присутствуют гены галактуронатного метаболизма (*ихаСВА*), а также как UxuR, так и ExuR, осуществляющие регуляцию катаболизма D-глюкуроната и D-галактуроната, соответственно. Характерной чертой гексуронатных регулонов у бактерий этих групп (а также *P. luminescens* среди Enterobacteriales, предположительно, получившего данные гены от общих предков с Vibrionales и Alteromonadales) является наличие в их составе генов TRAP транспортеров, названных нами UxuPQM. Так как встречаемость генов *ихуPQM* среди Gammaproteobacteria коррелирует с отсутствием в геноме гена транспортера гексуронатов *ехuT*, эти транспортеры, вероятно, обеспечивают импорт гексуронатов в клетку. У большинства исследованных в этой работе бактерий (*Haemophilus* spp., *Marinomonas* sp. MWYL1, *P. luminescens*, *Psychromonas ingrahamii*, *Vibrio* spp.) в геноме обнаруживается только один кластер *ихуPQM* в составе гексуронатных регулонов, колокализованный с генами пути утилизации глюкуроната. Следовательно, эти транспортеры вероятнее всего вовлечены в транспорт D-глюкуроната/D-фруктуроната. В то же время у некоторых исследованных нами бактерий (*Actinobacillus succinogenes*, *Mannheimia succiniciproducens*, *P. profundum*) в геноме присутствует несколько копий генов *ихуPQM* в составе гексуронатных регулонов. В этих геномах один из наборов генов *ихуPQM* колокализуется с генами, отвечающими за утилизацию D-галактуроната. Это дает возможность предположить, что некоторые из паралогов TRAP транспортеров в составе гексуронатных регулонов обеспечивают импорт D-галактуроната.

Общий интермедиат параллельных путей утилизации D-глюкуроната и D-галактуроната, KDG, далее последовательно подвергается действию киназы KdgK и альдолазы Eda/KdgA. У *E. coli* эта часть метаболического пути находится под контролем репрессора KdgR из семейства ICLR [Murray et al., 2005]. Нами было показано, что гены *kdgK* и *eda* у представителей Pasteurellales, Vibrionales, а также у *P. luminescens* и *P. ingrahamii* колокализованы с генами регулона UxuR и входят в его состав. Все эти бактерии, за исключением *Vibrio* spp., не имеют регулятора KdgR, таким образом, их регулон UxuR расширяется, включая гены дальнейшего метаболического пути гексуронатов. У *Vibrio* spp. идентифицированы две паралогичные копии генов *kdgK* и *eda*, одна из которых находится под контролем UxuR, а вторая – под регуляцией KdgR [Rodionov et al., 2004]. В состав регулона UxuR у *P. profundum*, *P. ingrahamii* и *Vibrio* spp. также предположительно входит ген транспортера *ygjV*, который колокализован с генами *kdgK* и *eda* в этих геномах и, вероятно, участвует в транспорте каких-либо интермедиатов гексуронатного катаболизма.

Кластер генов *yjj* кодирует транспортер L-галактоната (YjjL), L-галактонат оксидоредуктазу (YjjN) и транскрипционный фактор семейства GNTR (YjjM) [Reed, 2006]. Ортологи генов *yjj* найдены у представителей Enterobacteriales (*E. coli*, *P. atrosepticum*, *Salmonella enterica* Typhimurium LT2, *S. proteamaculans*, *Shigella* spp.), а также двух Pasteurellales (*A. succinogenes* и *M. succiniciproducens*). Слабые сайты связывания UxuR были предсказаны для дивергона *yjjM(L)-yjjN*

у *E. coli* и *Shigella* spp. По результатам наших предсказаний лабораторией О. Озолинь Института биофизики клетки РАН экспериментально подтверждена UxuR-зависимая регуляция транскрипции *ujjM* и *ujjN* у *E. coli*.

Кластер дивергентных генов *lfaR-lfaTA* (соответственно, репрессор семейства LACI, MFS-транспортер, гомологичный EхuT, и α -глюкозидаза), вероятно, участвует в утилизации углеводов растительного происхождения и проявлении фитопатогенных свойств [Van Gijsegem et al., 2008]. Гены *lfa* были идентифицированы у патогенных и условно-патогенных Enterobacteriales, таких как *C. koseri*, *Enterobacter* sp. 638, *K. pneumonia*, *P. atrosepticum*, *S. proteamaculans*, *Shigella sonnei* и *Yersinia* spp., у которых этот кластер генов всегда находится под регуляцией UxuR или EхuR.

Пятая глава посвящена исследованию регуляции метаболизма малоната и пропионата у Proteobacteria и включает реконструкцию соответствующих регулонов, выявление новых регуляторов и их мотивов связывания, построение возможной модели эволюции этих метаболических систем.

Малонат может использоваться как источник углерода многими бактериями. Описаны две группы структурных генов, участвующих в метаболизме малоната (Рисунок 11). Первая метаболическая система была охарактеризована у *Rhizobium leguminosarum* [Lee et al., 2000] и включает гены *matA*, *matB* и *matC*, кодирующие малонил-СоА декарбоксилазу, малонил-СоА синтазу и транспортер малоната, соответственно, а также ген *matR* малонатного регулятора семейства GNTR. Другая система метаболизма малоната описана у *Acinetobacter calcoaceticus* [Koo et al., 2000]. Она включает оперон *mdcLMACDEGBH*, кодирующий субъединицы транспортера малоната (MdcLM), малонат декарбоксилазы и вспомогательных белков (MdcABCDEFGHI), а также ген регулятора *mdcY* семейства GNTR. У некоторых Gammaproteobacteria в составе малонатного оперона присутствует другой ген транспортера малоната – *mdcF*. Кроме того, у ряда представителей Gammaproteobacteria регулятором метаболизма малоната является транскрипционный фактор MdcR из семейства LYSR – авторепрессор и активатор экспрессии генов *mdc* [Koo et al., 2000; Peng et al., 1999].

Пропионат также может служить источником углерода для многих бактерий. Метаболизм этого соединения тесно связан с путем утилизации малоната, а также с центральным метаболизмом, например, циклом трикарбоновых кислот (ЦТК) (Рисунок 11). Ферменты метилцитратного пути, кодируемые генами *prpBCDEF* и *acnB(acnM)/acnD*, осуществляют превращение пропионата в пируват и сукцинат [Brämer et al., 2001]. Пропионат может также включаться в цитрамалатный цикл, в состав которого входят ферменты, кодируемые генами *pccBA*, *epi*, *mutB* и *meaB*, и осуществляющие превращение пропионил-СоА в сукцинил-СоА [Hubbard et al., 2007; Meister et al., 2005]. Известно, что транскрипционным активатором генов

пропионатного метаболизма *prp* у *E. coli* и *Ralstonia eutropha* является σ^{54} -зависимый транскрипционный фактор PrpR из семейства Fis. Кроме того, у *Pseudomonas* spp. и *Vibrio cholerae* был описан иной регулятор метаболизма пропионата – транскрипционный фактор семейства GNTR, ген которого колокализован с кластером генов *prp* [Brämer et al., 2001].

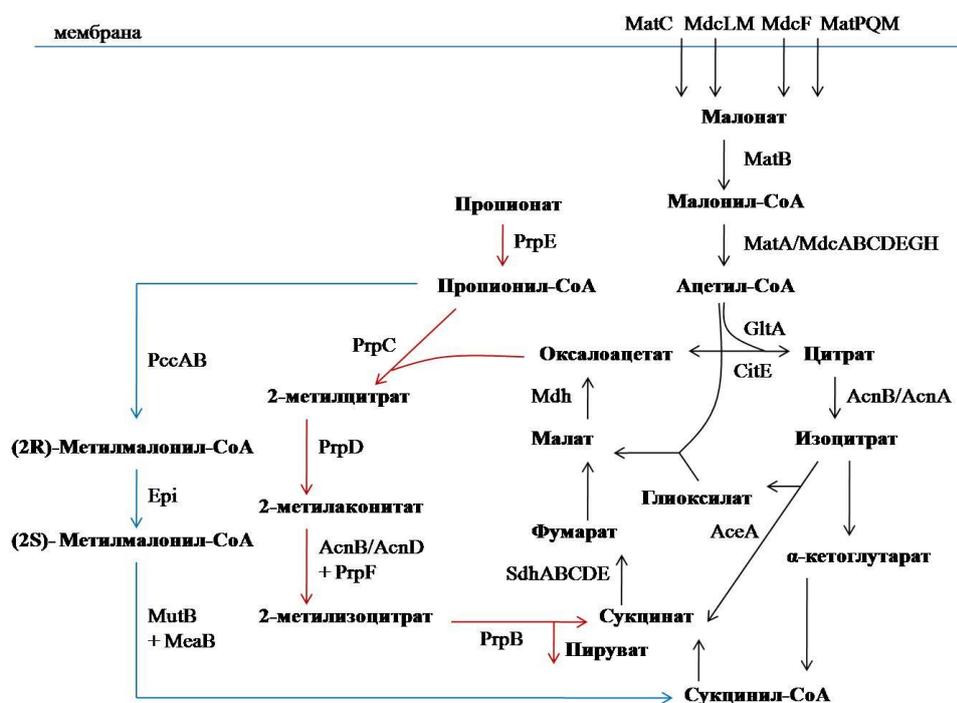


Рисунок 11. Схема метаболизма малоната и пропионата

Метилцитратный путь отмечен красными стрелками, часть цитрамалатного цикла – синими.

Регуляторы MatR и MdcY имеют высокую степень сходства (49%), кроме того, ранее предсказанные сайты связывания этих транскрипционных факторов идентичны [Lee et al., 2000; Koo et al., 2000]. Филогенетический анализ также не выявил явного разделения между регуляторами, гены которых колокализованы с генами *mat* или *mdc*. В связи с этим, в данной работе этот регулятор обозначен как MdcY, по названию, использованному в более ранней публикации. Наличие в геноме и организация генов *mdc* и *mat* у разных бактерий сильно варьируют. В ряде геномов присутствуют как гены *mat*, так и *mdc*, тогда как другие бактерии имеют только оперон *mdc* или *mat*. У многих Alphaproteobacteria, а также у ряда Beta- и Gammaproteobacteria гены метаболизма малоната колокализованы с генами, кодирующими TRAP транспортеры (названные *matPQM* у *Sinorhizobium meliloti*), которые образуют с ними оперон или же имеют собственные сайты связывания MdcY. Нуль-мутанты по генам *matPQM* не способны расти на среде, содержащей малонат в качестве единственного источника углерода [Chen et al., 2010]. В нашей работе было показано, что в большинстве случаев наличие генов этих TRAP транспортеров в составе малонатного регулона коррелирует с отсутствием генов иных известных

транспортеров малоната – *mdcLM*, *mdcF* и *matC*. Среди полностью секвенированных геномов, ортологи *matPQM* были идентифицированы только среди бактерий, в геноме которых присутствуют гены метаболизма малоната. Все это позволяет уверенно предполагать, что эти TRAP транспортеры участвуют в транспорте малоната и входят в состав регулона MdcY.

Кроме выявления мотивов связывания и реконструкции регулонов известных регуляторов MdcY, MdcR и PrpR у Proteobacteria, в настоящей работе был также описан ряд новых факторов транскрипции, участвующих в регуляции метаболизма малоната и пропионата, идентифицированы их сайты связывания и реконструированы соответствующие регулоны (Таблица 2).

Наиболее однородная регуляция наблюдается среди Alphaproteobacteria: гены метаболизма малоната находятся под регуляцией MdcY, а утилизация пропионата контролируется PrpQ* (Таблица 3). Оба транскрипционных фактора типичны для Alphaproteobacteria, тогда как у Beta- и Gammaproteobacteria они встречаются редко, вероятно, как результат горизонтального переноса.

Среди Gammaproteobacteria возможный перенос гена *mdcY* произошел у Alteromonadales и Pseudomonadales. В то же время, у представителей Chromatiales и Xanthomonadales присутствует иной регулятор метаболизма малоната, MlnR* из семейства GNTR, типичный для Betaproteobacteria и, вероятно, унаследованный от общего с ними предка. У прочих Gammaproteobacteria метаболизм малоната находится под контролем MdcR, и данный фактор транскрипции представляется исходным и основным малонатным регулятором в этой таксономической группе (Таблица 3).

Метаболизм пропионата у Gammaproteobacteria регулируется преимущественно транскрипционным фактором PrpR* из семейства GNTR, однако у некоторых представителей Enterobacteriales и всех исследованных Xanthomonadales пропионатным регулятором служит белок PrpR из семейства FIS. PrpR также обнаруживается у Betaproteobacteria и может быть унаследован Beta- и Gammaproteobacteria от их общего предка, или же получен общим предком Enterobacteriales и Xanthomonadales от Betaproteobacteria в результате горизонтального переноса генов. Возможно, мы наблюдаем промежуточную эволюционную стадию замещения транскрипционным фактором PrpR* регулятора PrpR у Gammaproteobacteria. Транскрипционный фактор PrpR* также присутствует у ряда Betaproteobacteria и *Geobacter metallireducens* (Deltaproteobacteria), который, вероятно, получил PrpR* в результате горизонтального переноса генов от общего предка Alteromonadales. Часть Betaproteobacteria получила PrpR* в результате горизонтального переноса от *Pseudomonadales* (например, *Bordetella petrii*), тогда как ряд других представителей этого таксона (например, *Bordetella bronchiseptica*, *Verminephrobacter eiseniae*) унаследовали данный транскрипционный фактор от общего предка с Gammaproteobacteria.

Таблица 2. Регуляторы метаболизма малоната и пропионата у Proteobacteria

Регулятор	Представлен среди таксонов	Регулируемые гены	Мотив связывания
MdcY (семейство GNTR, подсемейство FADR)	Alphaproteobacteria (Rhizobiales, Rhodobacterales, Rhodospirillales, Sphingomonadales), Betaproteobacteria (Burkholderiales, Rhodocyclales), Gammaproteobacteria (Alteromonadales, Pseudomonadales)	<i>matABC, matPQM, mdcABCDEGH, mdcF, mdcLM</i>	TTGTATACAA, что согласуется с полученными ранее данными сравнительного анализа и ДНК-футпринтинга [Lee H.Y., 2000; Коо JH., 2000]
MdcR (семейство LYSR)	Betaproteobacteria (Burkholderiales) и Gammaproteobacteria (Alteromonadales, Enterobacteriales, Pseudomonadales)	<i>mdcABCDEGH, mdcF, mdcLM, matC</i>	ATCrTTAcyyTgArsgTAAyGAT
PrpR (семейство FIS)	Betaproteobacteria (Burkholderiales) и Gammaproteobacteria (Enterobacteriales и Xanthomonadales)	<i>prpBCDEF, acnD</i>	rTTTCAwwwTGAAAy
MInR* (семейство GNTR, подсемейство FADR), родственник MdcY	Betaproteobacteria (Burkholderiales, Rhodocyclales), Gammaproteobacteria (Chromatiales, Xanthomonadales)	<i>mdcABCDEGH, matABC, matPQM, mutB, meab, pccBA, epi, prpE</i>	Тип 1. Betaproteobacteria – кТАуТурТААТТАурmwkwm Тип 2. Gammaproteobacteria - rТААТТАуGwyGТААТТАm
PrpR* (семейство GNTR, подсемейство FADR)	Betaproteobacteria (Burkholderiales), Gammaproteobacteria (Alteromonadales, Oceanospirillales, Pseudomonadales, Vibrionales), <i>Geobacter metallireducens</i>	<i>prpBCDEF, acnD, acnB, acnA</i>	ATTGTCGACAAT
PrpQ* (семейство XRE)	Alphaproteobacteria (Caulobacterales, Rhizobiales, Rhodobacterales, Rhodospirillales, Sphingomonadales), Betaproteobacteria (Burkholderiales)	<i>prpBCDF, acnD, pccBA, mutB</i>	TTTGCRAA (часто в нескольких копиях)
SdhR* (семейство GNTR, подсемейство HUTC)	Betaproteobacteria (Burkholderiales)	<i>sdhABCDE, gltA, mdh, citE, tam, aceA, acnA, acnB, acnD, prpBCDF</i>	TCTTATGTCTTATATAAGACATAAGA
Регулятор* семейства GNTR, подсемейства FADR, родственник MdcY	<i>Burkholderia</i> spp.	<i>mdcABCDEGH</i>	Потенциальных мотивов не выявлено в связи с высоким уровнем консервативности межгенной области
Регулятор* семейства LYSR	<i>Burkholderia</i> spp.	<i>mdcABCDEFGHI LM</i>	

Новые регуляторы метаболизма малоната и пропионата, описанные в данной работе, отмечены знаком *

Интересный вариант регуляции обнаруживается у *Bordetella avium*, имеющей две копии гена *prpR** в геноме, один из которых предположительно получен от представителей Pseudomonadales, тогда как второй унаследован от общего предка с Gammaproteobacteria.

Наиболее разнообразная регуляция наблюдается у Betaproteobacteria (Таблица 3). Среди бактерий этой группы идентифицировано пять регуляторов метаболизма малоната: MdcY, предположительно полученный ими от Alphaproteobacteria в результате горизонтального переноса генов; MdcR, полученный от Gammaproteobacteria или же унаследованный от общего предка Beta- и Gammaproteobacteria; специфические для бактерий рода *Burkholderia* spp. транскрипционные факторы из семейств LYSR и GNTR. Однако у большинства Betaproteobacteria метаболизм малоната находится под контролем регулятора MlnR* из семейства GNTR.

Регуляция утилизации пропионата у Betaproteobacteria также разнообразна (Таблица 3). PrpQ* является регулятором пропионатного метаболизма только у трех близкородственных представителей Comamonadaceae, получивших данный транскрипционный фактор, вероятно, в результате горизонтального переноса генов от Alphaproteobacteria. У большинства Betaproteobacteria основным регулятором утилизации пропионата (либо единственным, либо совместно с PrpR или иногда PrpR*) служит SdhR*. Следует отметить, что у многих Betaproteobacteria, имеющих MlnR*, отсутствует регулятор PrpR из семейства FIS, что может быть обусловлено тем, что эти транскрипционные факторы контролируют альтернативные пути метаболизма пропионата.

Таблица 3. Распределение регуляторов метаболизма малоната и пропионата среди различных групп Proteobacteria

Таксон	Количество родов бактерий, имеющих данный транскрипционный фактор						
	MdcY	MlnR*	MdcR	PrpR	PrpR*	PrpQ*	SdhR*
Alpha	12	0	0	0	0	17	0
Beta	4	11	5	3	2	3	12
Gamma	2	3	7	6	16	0	0
Delta	0	0	0	0	1	0	0

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

1. Исследованы 1252 транскрипционных фактора семейства GNTR из 64 ортологических групп, для которых идентифицированы мотивы связывания и реконструированы соответствующие регулоны.

2. Для подсемейств FADR, HUTC и YTRA семейства GNTR исследована коэволюция мотивов связывания ДНК и аминокислотных последовательностей регуляторов транскрипции, предсказаны вероятные ДНК-белковые контакты. Показано, что большая часть предсказанных контактов сходна для всех трех подсемейств и хорошо соотносится с известными данными о ДНК-белковых взаимодействиях.

3. Проведен анализ структуры дивергонов, регулируемых транскрипционными факторами подсемейств FADR и HUTC семейства GNTR, выявлены основные тенденции расположения сайтов связывания в дивергонах с единичными и двойными сайтами.

4. Для 23 ортологических групп регуляторов семейства GNTR показано наличие слабых дополнительных боксов рядом с основным палиндромным сайтом связывания, которые, предположительно, могут участвовать в альтернативной димеризации транскрипционных факторов или же в связывании дополнительных субъединиц.

5. Детально исследована регуляция метаболизма гексуронатов у *Gamma*proteobacteria, построена модель эволюции родственных регуляторов UxuR и ExuR, разделены их мотивы связывания, реконструированы соответствующие регулоны, в том числе, идентифицирован ряд новых членов гексуронатных регулонов. По результатам предсказаний лабораторией Института биофизики клетки РАН экспериментально подтверждена UxuR-зависимая регуляция транскрипции *yjjM* и *yjjN* у *E. coli*.

6. Детально исследована регуляция метаболизма малоната и пропионата у *Proteobacteria*, показана ее варибельность и разнообразие в различных таксономических группах, построена модель эволюции данных метаболических систем. Идентифицирован ряд новых транскрипционных регуляторов метаболизма малоната и пропионата. Для каждого из исследованных транскрипционных регуляторов метаболизма малоната и пропионата определены мотивы связывания и реконструированы соответствующие регулоны, в том числе описаны их новые члены.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в научных журналах:

1. Suvorova, I.A., Tutukina, M.N., Ravcheev, D.A., Rodionov, D.A., Ozoline, O.N., Gelfand, M.S. (2011). Comparative genomic analysis of the hexuronate metabolism genes and their regulation in gammaproteobacteria. *J Bacteriol.* **193(15)**:3956-63.
2. Suvorova, I.A., Ravcheev, D.A., Gelfand, M.S. (2012). Regulation and evolution of malonate and propionate catabolism in proteobacteria. *J Bacteriol.* **194(12)**:3234-40.
3. Suvorova, I.A., Korostelev, Y.D., Gelfand, M.S. (2015). GntR Family of Bacterial Transcription Factors and Their DNA Binding Motifs: Structure, Positioning and Co-Evolution. *PLoS One.* **10(7)**:e0132618.

Тезисы конференций:

1. Суворова И.А., Равчеев Д.А. Коэволюция белков семейства FadR и их сайтов связывания. Труды 32-й конференции молодых ученых и специалистов ИППИ РАН (ИТиС'09), Бекасово, 2009, с. 297-299.
2. Суворова И.А. Утилизация гексуронатов в гамма-протеобактериях. Исследование методами сравнительной геномики. Материалы XVII международной конференции студентов, аспирантов и молодых ученых «Ломоносов-2010», Москва, 2010.
3. Суворова И.А. Белок Prp2 подсемейства FadR – новый вариант регуляции метаболизма пропионата. Труды 33-й конференции молодых ученых и специалистов ИППИ РАН (ИТиС'10), Геленджик, 2010, с. 350-351.
4. Суворова И.А. Факторы транскрипции семейства GntR: эволюция регуляторных систем. Материалы международной научно-практической конференции «Постгеномные методы анализа в биологии, лабораторной и клинической медицине», Москва, 2010.
5. Suvorova I.A. Comparative genomic analysis of the hexuronate metabolism genes and their regulation in Gamma-Proteobacteria. Proceedings of Russian-German seminar «Regulation and Evolution of Cellular Systems» (RECESS, Регуляция и эволюция клеточных процессов), Мюнхен, 2011.
6. Суворова И.А. Регуляция метаболизма малоната и пропионата у Протеобактерий. Труды 34-й конференции молодых ученых и специалистов ИППИ РАН (ИТиС'11), Геленджик, 2011, с. 28-30.
7. Суворова И.А. Гельфанд М.С. Исследование сайтов связывания и их окружения для транскрипционных факторов семейства GntR. Конференция молодых ученых «Молекулярная и клеточная биология: прикладные аспекты» в рамках отчетной конференции по программе фундаментальных исследований ПРАН «Молекулярная и клеточная биология», Москва, 2012, с. 29.

8. Suvorova I.A. Comparative analysis of the GntR family transcription factors and their binding sites. Proceedings of Russian-German seminar «Regulation and Evolution of Cellular Systems» (RECESS, Регуляция и эволюция клеточных процессов), Москва, 2012.

9. Суворова И.А. Исследование сайтов связывания и их окружения для транскрипционных факторов семейства GntR. Труды 35-й конференции молодых ученых и специалистов ИППИ РАН (ИТиС'12), Петрозаводск, 2012, с. 338-340.

10. Suvorova I.A. GntR family of bacterial transcription factors and their binding motifs: structure, positioning and co-evolution. Proceedings of Russian-German seminar «Regulation and Evolution of Cellular Systems» (RECESS, Регуляция и эволюция клеточных процессов), Венеция, 2013.

11. Suvorova, IA. GntR family of bacterial transcription factors and their binding motifs: structure and co-evolution. Proceedings of Moscow Conference on Computational Molecular Biology (MCCMB'13), Москва, 2013.