

**Факультет биоинженерии и биоинформатики Федерального государственного
образовательного учреждения высшего профессионального образования
Московского государственного университета
имени М.В. Ломоносова,**

**Учебно-научный центр «Биоинформатика»
Федерального государственного бюджетного учреждения науки
Институт проблем передачи информации им. А.А. Харкевича
Российской академии наук**

На правах рукописи

Цой Ольга Владиславовна

**ЭВОЛЮЦИЯ СИСТЕМ РЕГУЛЯЦИИ ТРАНСКРИПЦИИ
В ГЕНОМАХ БАКТЕРИЙ**

03.01.09 - математическая биология, биоинформатика

Диссертация на соискание ученой степени
кандидата биологических наук

Научный руководитель:
доктор биологических наук, профессор
Михаил Сергеевич Гельфанд

Москва, 2014

Содержание

Введение	3
Глава 1 - Обзор литературы	8
1.1 Сеть регуляции транскрипции и ее уровни.....	8
1.2 Идентификация регуляторных элементов транскрипционной сети.....	9
1.3 Путь утилизации этаноламина	12
1.4 Регуляторные блоки в сетях регуляции транскрипции	17
Глава 2 - Материалы и методы	23
2.1 Исследованные геномы.....	23
2.2 Программное обеспечение и базы данных.....	29
Глава 3 - Предсказание регуляторных элементов транскрипционной сети на примере пути деградации этаноламина	31
3.1 Поиск ортологов генов утилизации этаноламина	31
3.2 Эволюция системы утилизации этаноламина.....	40
3.3 Регуляция утилизации этаноламина	48
Глава 4 - Эволюция регуляторных блоков на примере блока треугольник	58
4.1 Определение локальных и глобальных транскрипционных факторов ..	58
4.2 Функциональная аннотация треугольников	61
4.3 Эволюция регуляторного блока типа треугольник в близкородственных организмах на примере штаммов <i>Escherichia coli</i>	61
4.4 Эволюция регуляторного блока типа треугольник в Enterobacteriales ..	67
Обсуждение	76
Выводы	79
Список публикаций по теме диссертации	81
Список литературы	82

Введение

Актуальность темы исследования

Многие системы в живых организмах образуют биологические сети. Примером такой сети является и система регуляции транскрипции. Самый нижний уровень этой сети составляют наборы взаимодействующих базовых элементов, состоящие из транскрипционного фактора, участка ДНК, с которым он связывается, и регулируемого гена. Объединяясь, на следующем уровне эти наборы образуют регуляторные блоки. Один или несколько регуляторных блоков образуют функциональный модуль, а несколько функциональных модулей составляют всю транскрипционную сеть. Растущее количество данных и экспериментальных методик позволяют изучать организмы в контексте целых сетей, модулей и блоков, а не отдельных функциональных систем. В настоящей работе мы анализировали эту систему на уровне блоков и наборов базовых элементов методами сравнительной геномики.

Методы сравнительной геномики позволяют реконструировать транскрипционную сеть в наборе родственных геномов. Поиск новых регуляторных элементов в бактериальных геномах является одним из наиболее сложных этапов ее изучения. В самом простом случае существуют экспериментальные данные хотя бы для одного генома, опираясь на которые становится возможным изучение других, родственных организмов. Но более частыми бывают ситуации, когда ни локализация, ни структура регуляторных элементов не известна.

Постановка и решение вопросов, связанных с транскрипционной сетью бактериальных генов, необходима для детальной функциональной аннотации генов; реконструкции метаболических путей – не только универсальных, но, в большей степени, таксон-специфичных; изучения роли этих путей в организме; а также исследования эволюции регуляторных систем и организмов в целом. Изучение принципов, лежащих в основе организации регуляторных сетей – неотъемлемая

часть понимания молекулярных механизмов жизнедеятельности, в том числе, патогенных бактерий, биотехнологических штаммов, организмов, применяемых в системах биоочистки и т.п.

Степень разработанности темы

В настоящий момент изучению систем регуляции транскрипции посвящено много работ, но они ограничены изучением нескольких модельных организмов (например, *Escherichia coli* [1] и *Bacillus subtilis* [2]) или проведены только на уровне отдельных функциональных систем. В то же время в базе данных Genbank [3] содержится несколько тысяч полных последовательностей бактериальных геномов разной степени родства, и еще больше находится в процессе определения последовательности или аннотации [3]. Рост числа геномов делает изучение индивидуальных геномов слишком трудоемким, но, с другой стороны, такое количество информации позволяет реконструировать биологические системы (в том числе, транскрипционные сети) в немодельных организмах, и их изучение представляется интересным уже не просто в рамках отдельных организмов, а в ряду близко родственных геномов, что позволяет проследить за особенностями их эволюции. Для подобного анализа требуется применение компьютерных, а не экспериментальных, методов.

Цели и задачи исследования

Целью настоящего исследования была реконструкция и анализ сетей транскрипционных регуляторных элементов в близкородственных геномах с помощью методов компьютерной биологии для немодельных организмов. В рамках поставленной цели решались следующие общие и частные задачи:

Предсказание регуляторных элементов транскрипционной сети *de novo* на примере пути утилизации этаноламина.

Изучение эволюции регуляторных блоков на примере блока треугольник.

Научная новизна

В настоящей работе были обобщены сведения о закономерностях эволюции транскрипционной регуляторной сети, а также проведено детальное исследование на материале обширной группы близкородственных геномов. Мы впервые показали, что различные регуляторные взаимодействия изменяются по-разному: как в зависимости от положения в транскрипционной сети, так и от свойств транскрипционного фактора.

С помощью методов сравнительной геномики мы обнаружили ранее неизвестные регуляторные элементы в пути утилизации этаноламина, а именно мотив участка связывания транскрипционного фактора AraC-семейства EutR. Кроме того, нам удалось показать регуляторную связь этого пути с метаболизмом кобаламина – кофактора основного фермента пути (этанолмин лиазы), а также описать эволюцию генов этого пути.

Теоретическая и практическая значимость

На сегодняшний день доступно несколько тысяч полных бактериальных геномов, включая штаммы, и еще больше находится в процессе определения последовательности или функциональной аннотации. Но, несмотря на экспоненциально растущее количество геномов, экспериментальное изучение регуляторных сетей, вследствие его трудоемкости, ограничено несколькими модельными организмами либо отдельными функциональными системами. Таким образом, следствием большого количества данных является необходимость применения методов компьютерной биологии и статистики для их анализа, позволяющих описывать свойства немодельных организмов, основываясь на экспериментально полученных данных по родственным модельным организмам. Более того, применение сравнительно-геномных методов позволяет описывать регуляторные взаимодействия в таксономических группах, не содержащих хорошо изученных модельных организмов.

Методология и методы исследования

Для решения поставленных задач использовались разные методы сравнительной геномики. Применялись методы поиска гомологов на основе критерия двустороннего лучшего сходства, поддержанного анализом филогенетических деревьев и анализом структуры оперона, методы анализа аминокислотных, нуклеотидных последовательностей, а также построения множественных выравниваний и филогенетических деревьев методом наилучших соседей с применением статистического размножения выборки. Для проверки статистической значимости использовался критерий χ^2 и гипергеометрический тест.

Положения, выносимые на защиту

1. Анализ таксономического распределения генов утилизации этаноламина показал, что существует два возможных пути катаболизма использования этаноламина, связанных с типом оперона. Первый, короткий, позволяет использовать этаноламин только в качестве источника азота. Второй, длинный, дает возможность использовать его и как источник азота, и как источник углерода.
2. Исследование эволюции генов утилизации этаноламина показало, что короткий оперон является предковым, из которого путем добавления новых генов образовался длинный тип. В ходе эволюции генов утилизации этаноламина произошло минимум три события горизонтального переноса.
3. В *Enterobacteriales* и *Burkholderiales* предсказан мотив участка связывания транскрипционного фактора EutR. В *Enterobacteriales* участки связывания EutR обнаружены в промоторной области семи оперонов, среди которых есть непосредственно гены утилизации этаноламина, а также гены синтеза кофактора основного фермента пути этаноламинлиазы – кобаламина. Это наблюдение устанавливает связь между путями утилизации этаноламина и синтезом кобаламина за счет EutR-зависимой регуляции. В *Burkholderiales*

участки связывания EutR обнаружены только непосредственно в промоторной области генов утилизации этаноламина.

4. Локальная регуляция является эволюционно подвижной, что согласуется с необходимостью быстрой адаптации к условиям окружающей среды. В штаммах *Escherichia coli* взаимодействия типа Л→ген (локальный транскрипционный фактор →ген) чаще сохраняются в регуляторных блоках, чем вне их. На уровне порядка Enterobacteriales эти взаимодействия консервативнее в парных взаимодействиях, по сравнению с регуляторными блоками, что может быть результатом изменчивости транскрипционной сети.
5. Разные типы треугольников на уровне порядка Enterobacteriales эволюционируют по-разному. Регуляция как глобальными, так и локальными транскрипционными факторами в несогласованном треугольнике оказывается консервативнее, чем в согласованном.

Степень достоверности и апробация результатов

По материалам диссертации опубликовано 2 статьи в международных рецензируемых научных журналах. Результаты работы были представлены на международной конференции МССМВ'09, российских конференциях ИТИС'08, ИТИС'10, а также на международных семинарах RECESS'10, RECESS'11 и Chemical, Synthetic And Systems Biology: New Directions Of Biochemistry In The 21st Century'11.

Структура и объем работы

Диссертация изложена на 100 страницах машинописного текста и содержит следующие разделы: обзор литературы, материалы и методы, результаты в двух главах, обсуждение и выводы. В конце приведен список литературы. Материал включает 21 рисунок, 14 таблиц и список литературы, содержащий 166 ссылок.

Глава 1 - Обзор литературы

1.1 Сеть регуляции транскрипции и ее уровни

Многие биологические системы в живых организмах или сообществах образуют сети: пищевая цепь, нейронные сети, сигнальные каскады. Основная функция этих и подобных биологических систем состоит в адаптации клеток к изменяющимся условиям среды: изменениям температуры, осмотического давления, концентрации питательных веществ. Одним из основных путей адаптации является регуляция экспрессии генов, первым этапом которой является транскрипция – процесс синтеза РНК по матрице ДНК, осуществляемый специальным ферментом, ДНК-зависимой РНК-полимеразой. Процесс транскрипции регулируют, в частности, специальные белки: транскрипционные факторы, которые связываются с особыми участками ДНК перед регулируемым геном. Вся совокупность транскрипционных факторов и регулируемых ими генов образует транскрипционную сеть.

Существуют два основных типа регуляции – позитивная и негативная. Если при связывании транскрипционного фактора с узнаваемым им участком ДНК уровень экспрессии повышается, то эта регуляция называется позитивной. В таком случае транскрипционный фактор называется активатором. Если уровень экспрессии уменьшается, то регуляция называется негативной, а транскрипционный фактор – репрессором.

В некоторых случаях для того, чтобы транскрипционный фактор связался с участком ДНК, необходим сигнал – изменение физиологических условий (температуры, рН), малая молекула, пептид или специальный белок, так называемый индуктором. Регулируемый таким образом ген называется индуцибельным, т.е. его экспрессия изменяется под действием индуктора. Гены, которые экспрессируются на постоянном уровне, называются конститутивными.

Как любую сложную систему, транскрипционную сеть можно изучать на разных уровнях. Самый нижний уровень этой сети представляют отдельные базовые

регуляторные элементы (транскрипционные факторы, участки связывания в ДНК и регулируемые гены). На следующем уровне эти элементы, объединяясь, образуют регуляторные блоки. Далее блоки объединяются в функциональные модули, которые все вместе и составляют транскрипционную сеть. Изучение транскрипционной сети интересно на всех этих уровнях. В настоящей работе мы анализировали первый и второй уровень системы регуляции транскрипции методами сравнительной геномики.

1.2 Идентификация регуляторных элементов транскрипционной сети

В настоящий момент детальное описание систем регуляции транскрипции существует только для нескольких модельных видов на уровне организма (например, *Escherichia coli* [1] и *Bacillus subtilis* [2]) или на уровне отдельных функциональных систем. В то же время в базе данных Genbank [3] содержится несколько тысяч полных последовательностей бактериальных геномов разной степени родства, и еще больше находится в процессе определения последовательности или аннотации [3]. Такое количество информации позволяет реконструировать биологические системы (в том числе, системы регуляции транскрипции) в немодельных организмах, но для анализа требуется применение компьютерных, а не экспериментальных, методов.

Существуют различные подходы к реконструкции транскрипционных сетей в зависимости от того, насколько хорошо она изучена [4, 5]. Ряд методов использует существующие экспериментальные данные о транскрипционных факторах и их участках связывания, на основе которых в изучаемом организме предсказываются новые регулируемые гены [6, 7, 8, 9, 10, 11]. Другие подходы опираются на идентификацию транскрипционных факторов, участков связывания и регулируемых генов *de novo* [12,13,14,15,16,17,18]. Подобные исследования можно проводить как для отдельных транскрипционных факторов [5] и их семейств [19, 20, 21], так и для функциональных систем и даже для организма в целом [22, 23, 24].

Методы сравнительной геномики основаны на предположении о том, что группы совместно регулируемых генов, так называемые регулоны, сохраняют в ходе эволюции свой основной состав. Если ген имеет участок связывания транскрипционного фактора в одном геноме, его ортологи в геномах родственных организмов, как правило, будут иметь схожие участки связывания ортолога этого фактора [25]. Таким образом, общий подход к реконструкции транскрипционной сети можно описать следующим образом [26]. На предварительном этапе анализа проводится поиск ортологов изучаемого гена в организмах разной степени родства. Далее в 5'-областях всех найденных ортологов проводится поиск участков связывания рассматриваемого транскрипционного фактора. Эта задача сильно упрощается, если существуют экспериментальные данные о последовательности, структуре и локализации каких-либо участков связывания. Если такой информации нет, используются различные методы предсказания (см. ниже). На основе выравнивания экспериментальных или предсказанных участков связывания составляется распознающее правило.

Вид распознающего правила может быть разным. Одним из наиболее эффективных является матрица позиционных весов. Числа в ячейках такой матрицы – это позиционные веса (вес каждого нуклеотида в каждой позиции участка связывания), которые вычисляются по формуле [25]:

$$W(b, k) = \log(N(b, k) + 0,5) - 0,25 \sum_{i=A,C,G,T} \log(N(i, k) + 0,5),$$

где $N(b, k)$ – количество нуклеотидов b в позиции k в обучающей выборке. Первый член в формуле зависит от того, сколько раз данный нуклеотид встретился в данной позиции, второй – от консервативности самой позиции. Вес участка связывания определяется суммой соответствующих позиционных весов, и измеряется в единицах стандартного отклонения распределения весов на случайных последовательностях. Для поиска участков связывания с помощью матрицы

позиционных весов необходимо установить следующие параметры: пороговые значения весов участков связывания и область поиска.

Далее для детального описания регулона и предсказания регулируемых генов может использоваться так называемый метод проверки соответствия: ген считается относящимся к регулону, если потенциальный участок связывания был обнаружен в его 5'-области в нескольких геномах [25].

Часто при изучении транскрипционной сети последовательность участка связывания неизвестна, но ее можно предсказать, используя различные подходы сравнительной геномики. Выбор алгоритма предсказания зависит от количества и качества исходных данных (величины выборки, степени родства геномов). Одним из возможных подходов является метод филогенетического футпринтинга. Он основывается на предположении о консервативности регуляторных элементов в 5'-областях анализируемых генов в родственных геномах. Сначала строится дерево по аминокислотным последовательностям белков, кодируемых анализируемыми генами. Затем на его основании производятся множественные выравнивания 5'-областей, соответствующие внутренним узлам дерева, до тех пор, пока выравнивание имеет смысл. Таким образом, определяется локализация родственных консервативных областей и то, насколько они сохраняются в ряду геномов. Эти консервативные области и являются потенциальными участками связывания, используемыми для дальнейшего построения распознающего правила [27]. Этот подход оптимален при наличии большой выборки геномов разной степени родства, но неэффективен для очень близких геномов, так как их 5'-области консервативны на всем своем протяжении, и отдельные более консервативные области выделить невозможно. Филогенетический футпринтинг также не дает положительных результатов для очень далеких геномов, 5'-области которых практически неконсервативны и не могут быть выравнены. В этих случаях возможно применение алгоритмов предсказания без предварительного выравнивания.

Участки связывания транскрипционных факторов часто имеют нетривиальную внутреннюю структуру, представляя собой палиндромы или прямые повторы. В каждой 5'-области производится поиск всех последовательностей, имеющих подобную, заранее заданную структуру, а далее они сравниваются между собой. Из выборки наиболее похожих последовательностей строится матрица позиционных весов, которая используется вновь для поиска структурированных последовательностей. Процедура повторяется до схождения [26].

Все описанные подходы имеют определенные ограничения. Во-первых, требуется достаточно большая выборка геномов; во-вторых, поскольку распознающее правило и метод проверки соответствия опираются на сопоставление геномов нескольких организмов, они не могут использоваться для изучения видоспецифичной регуляции.

В настоящей работе изучается функциональная система – путь утилизации этаноламина, для которой нет никаких экспериментальных данных о последовательности участка связывания транскрипционных факторов. Только для некоторых бактериальных геномов есть информация, какой транскрипционный фактор регулирует экспрессию генов этого пути. Нами была поставлена задача, используя методы сравнительной геномики, установить для этой системы участки связывания транскрипционных факторов и регуляторные связи с другими генами и функциональными системами.

1.3 Путь утилизации этаноламина

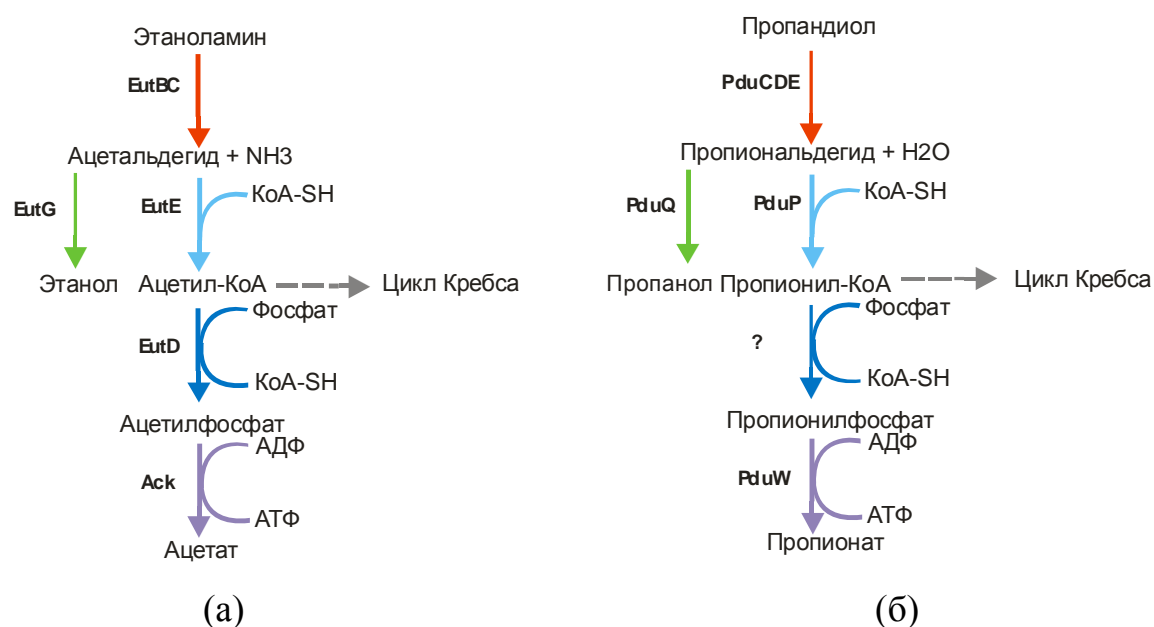
Одним из источников углерода и азота для бактерий может служить этаноламин. В процессе расщепления этаноламина возможно также получение энергии в виде АТФ за счет субстратного фосфорилирования или получения восстановленных форм восстановительных элементов в цикле Кребса. Путь утилизации этаноламина является частным случаем реакций деградации диолов с

образованием альдегидов (Рисунок 1). Все известные подобные пути требуют наличия фермента лиазы и его кофактора – кобаламина (витамина В12).

Путь утилизации этаноламина экспериментально изучен у *Salmonella typhimurium* LT2. Анализ различных штаммов *S. typhimurium*, мутантных по разным регионам *eut*-оперона, позволил определить состав оперона и роль продуктов отдельных генов в пути утилизации этаноламина *S. typhimurium* [28, 29].

Основным ферментом пути является этаноламин лиаза (ЕС 4.3.1.7) – белковый комплекс, состоящий из двух субъединиц: большой EutB и малой EutC. Этанолмин лиаза осуществляет главную реакцию деградации этаноламина – расщепление до ацетальдегида и аммиака, который используется как источник азота.

Дальнейшее превращение ацетальдегида осуществляется оксидоредуктазой EutE до ацетил-КоА или алкогольдегидрогеназой EutG до этанола. Ацетил-КоА в свою очередь может быть превращен в ацетат с получением одной молекулы АТФ с помощью ферментов ацетат киназы Ask и фосфоацетилтрансферазы EutD либо отправлен в цикл Кребса [30](Рисунок 1).



Одинаковым цветом показаны химически сходные реакции.

Рисунок 1 - (а) Путь утилизации этаноламина; (б) Путь утилизации пропандиола.

Кофактором этаноламин лиаза является аденозилкобаламин, который может быть получен извне или из цианокобаламина или гидроксикобаламина, синтезируемых в клетке. Превращение в аденозилкобаламин происходит с помощью аденозилкобаламинтрансферазы EutT [31]. Этот фермент, в отличие от общей аденозилкобаламинтрансферазы CobA, связанной с путем синтеза кобаламина, присутствует только в организмах, имеющих ферменты утилизации этаноламина. Цианокобаламин и гидроксикобаламин являются ингибиторами этаноламин лиаза. EutA, компонент ферментативного комплекса, который обнаружен у многих бактерий, растущих на этаноламине, защищает этаноламин лиазу от их воздействия.

У некоторых бактерий, способных к деградации диолов, обнаружен особый компартмент, так называемая метаболосома [30, 32] (Рисунок 2). У таких организмов превращение этаноламина до ацетил-коА происходит внутри этого компартмента. Структурные белки метаболосомы – EutS, EutM, EutN, EutL, EutK – являются гомологами структурных белков карбоксисомы – органеллы цианобактерий, ответственной за накопление CO_2 для фиксации ферментом рибулозобисфосфаткарбоксилазой [33]. Предположительная роль метаболосом в клетках бактерий – предотвращение потерь альдегидов во время утилизации диолов [34].



Рисунок 2 - Метаболосома в клетке *S. typhimurium* (из [32]).

У *S. typhimurium* гены, кодирующие многие вышеупомянутые ферменты, образуют единый оперон – *eutSPQTDMMNEJGHABCLKR*. EutH осуществляет транспорт этаноламина в клетку, а роль EutP, EutQ, EutJ до настоящего момента остается неизученной.

Ранее предпринимались попытки экспериментального изучения регуляции транскрипции *eut*-оперона [35]. Для *S. thyphimurium* была экспериментально показана роль EutR, ген которого также находится в составе *eut*-оперона, в регуляции экспрессии генов утилизации этаноламина. Штаммы *S. thyphimurium*, мутантные по гену *eutR*, утрачивают способность расти на этанолаmine. Эта способность появляется вновь при трансформации мутантов плазмидой, содержащей ген *eutR*. Также экспериментально показано, что EutR является активатором экспрессии генов утилизации этаноламина [35].

Было сделано предположение, что для активации *eut*-оперона требуются два сигнала – наличие этаноламина в среде и доступ к кофактору – кобаламину. При этом внутри *eut*-оперона существуют два промотора – конститутивный PII и

индуцибельный P_I. В отсутствие одного из сигналов происходит слабая экспрессия EutR за счет транскрипции с промотора P_{II}. Как только в среде появляются оба сигнала – и этаноламин, и кобаламин - EutR связывается с промотором P_I и вызывает экспрессию всех генов утилизации этаноламина, в том числе и самого регулятора *eutR* [35] (Рисунок 3). Участки связывания EutR с ДНК не установлены.

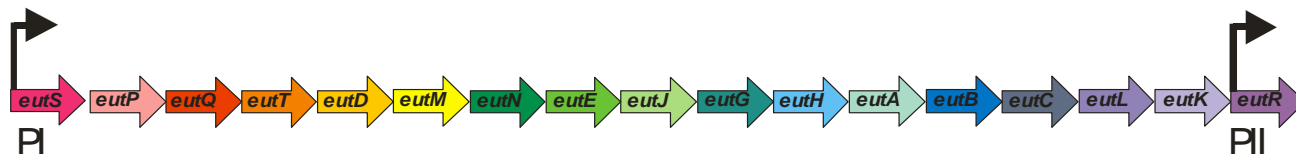


Рисунок 3 - Конститутивный (P_{II}) и индуцибельный (P_I) промоторы *eut*-оперона.

Транскрипционный фактор EutR обнаружен далеко не у всех бактерий, содержащих ферменты утилизации этаноламина. Для таких организмов возникает отдельный вопрос о том, как происходит регуляция этого пути. Для таксономической группы Firmicutes ранее замечено, что рядом с генами утилизации этаноламина часто обнаруживаются гены двухкомпонентной регуляторной системы, например, гены *lin1136* и *lin1137* в *Listeria innocua* (А. Мушегян, частное сообщение).

Происхождение этой сложной метаболической системы неизвестно.

Путь утилизации этаноламина интересен еще и тем, что была показана его связь с пищевыми отравлениями [36]. Так, обнаружено, что большинство бактерий, вызывающих пищевые отравления, имеют ферменты утилизации этаноламина, а многие также и пропандиола. Интересно, что путь утилизации этаноламина биохимически похож на путь утилизации пропандиола (Рисунок 1), регуляция которого экспериментально изучена в той же *S. typhimurium*. Гены утилизации пропандиола, так же как и в случае утилизации этаноламина, образуют единый оперон – *pduABCDEFGHIJKLMNPOQSTUVWX*. Установлено, что регуляция экспрессии генов утилизации пропандиола осуществляется транскрипционным

фактором *PocR* из семейства *AraC*, а также прямо или опосредованно глобальными транскрипционными факторами *Crp* и *ArcA* [37]. *PocR* также контролирует синтез кофактора основного фермента – кобаламина [38]. Участки связывания *PocR* в регуляторных областях генов утилизации пропандиола и синтеза кобаламина экспериментально установлены [39]. Таким образом, для пути утилизации пропандиола показана его связь с синтезом кофактора, что позволяет на основании сходства двух путей предположить такую же связь и для системы утилизации этаноламина. Предсказание участков связывания *EutR* с промоторной областью *eut*-оперона и поиск новых регулируемых им генов, вероятно, позволит найти связь между путями утилизации этаноламина и синтеза кобаламина.

1.4 Регуляторные блоки в сетях регуляции транскрипции

Если на первом уровне система регуляции транскрипции состоит из отдельных регуляторных элементов, то на следующем уровне ее можно рассматривать как направленный граф, в вершинах которого находятся транскрипционные факторы и регулируемые гены, а ребра отображают связь факторов с регулируемыми генами. Особенностью графа на основе биологических сетей является наличие внутренних структур, повторяющихся чаще, чем в случайном графе. Такие подграфы, частота которых в биологическом графе значимо больше, чем в случайном графе, получили название «структурных мотивов» («network motif») [40]. Здесь мы будем называть такие подграфы «регуляторные блоки», чтобы отличать этот объект от мотива участков связывания транскрипционных факторов.

Для различных биологических сетей характерны разные регуляторные блоки. Было показано, что в транскрипционных сетях самыми частыми регуляторными блоками являются «треугольник» («feed-forward loop») и «веер» («bi-fan») [40].

В настоящей работе мы рассматривали регуляторный блок из трех элементов – «треугольник» [40, 41]. Это структура из трех генов, два из которых являются транскрипционными факторами. Один из этих транскрипционных факторов (*X*, так

называемый общий) регулирует другой (Y, специальный), а вместе они регулируют ген (Z) (Рисунок 4). Несмотря на то, что всего существует 13 возможных способов связать три вершины в направленном графе (Рисунок 5), показано, что в транскрипционных сетях среди блоков с тремя вершинами перепредставлен только «треугольник» [40, 41].

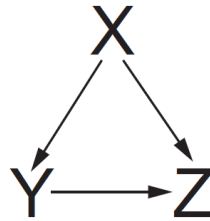
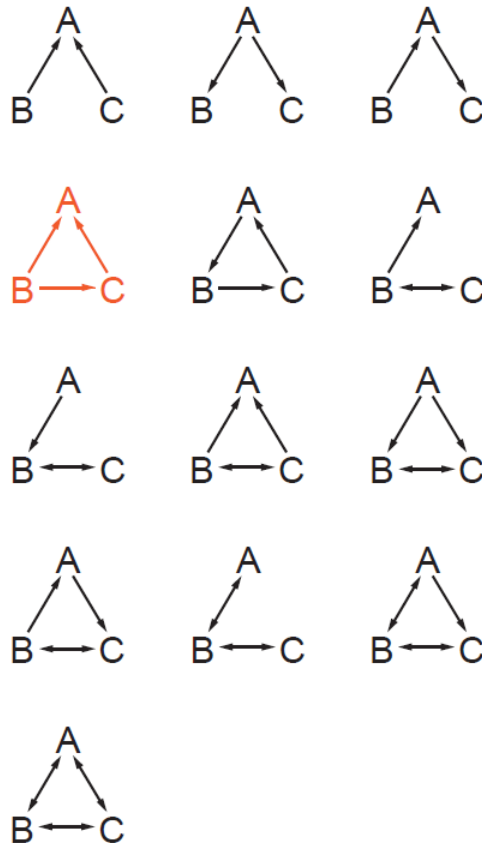


Рисунок 4 - Общая схема регуляторного блока типа «треугольник».



Оранжевым выделен блок, соответствующий «треугольнику».

Рисунок 5 – Возможные регуляторные блоки с тремя элементами.

Так как транскрипционный фактор может быть активатором или репрессором, существует восемь различных подтипов регуляторного блока треугольник [42]. В зависимости от того, какие транскрипционные факторы находятся в вершинах треугольника, различают согласованные и несогласованные треугольники (Рисунок 6). Мотив является согласованным, если прямой эффект общего транскрипционного фактора имеет тот же знак (положительный или отрицательный), что и не прямой эффект, опосредованный специальным фактором. Если знаки не совпадают, треугольник называется несогласованным. Например, если X и Y активируют экспрессию гена Z, и X активирует экспрессию гена фактора Y, то треугольник будет согласованным. Было показано, что подавляющее большинство треугольников является согласованными типа C1 (Рисунок 6) [40].

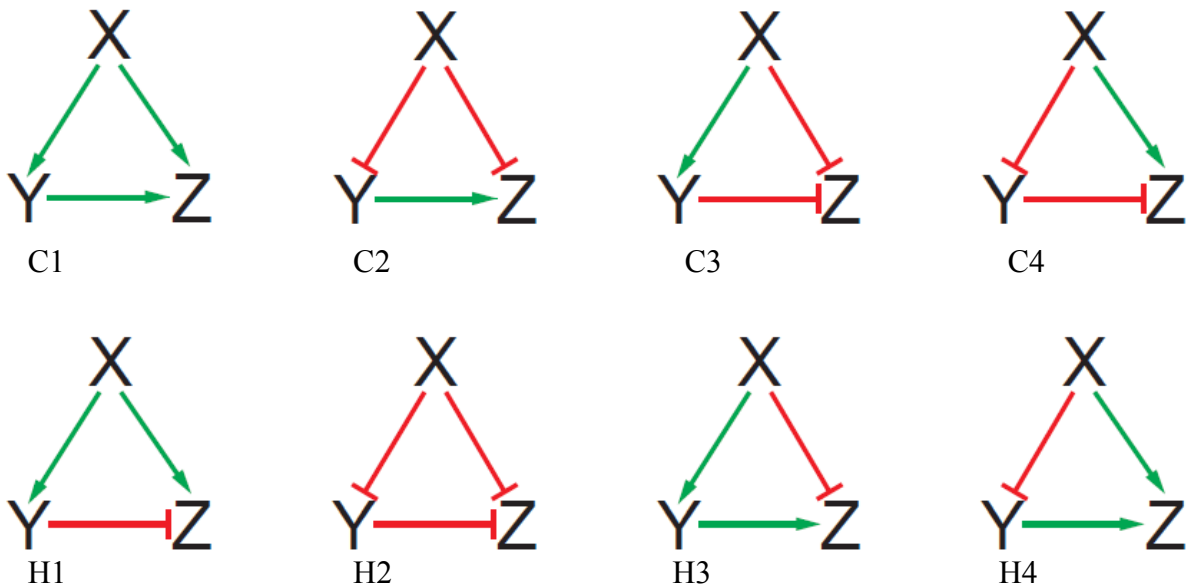


Рисунок 6 - Типы согласованных (С) и несогласованных (Н) треугольников.

Взаимодействия внутри треугольников можно разделить по природе образующего его транскрипционного фактора. В научной литературе принято разделять транскрипционные факторы на так называемые локальные и глобальные, хотя единого формального критерия разделения не существует. Одно из первых определений: глобальный транскрипционный фактор – это такой фактор, который

участвует в регуляции нескольких метаболических путей [43, 44]. Другой подход основан на количестве регулируемых генов или оперонов: в ряде работ устанавливался порог равный 15 генам или 10 оперонам [40, 45]. Транскрипционные факторы, контролирующие большее количество генов или оперонов, считались глобальными. Недавно был предложен структурный подход, в котором транскрипционная регуляция рассматривалась как сеть, а глобальным считался фактор, который регулирует несколько модулей в этой сети [46]. В *E. coli* только семь транскрипционных факторов – CRP, IHF, FNR, Fis, ArcA, Lrp, H-NS – подходят под большинство вышеперечисленных критериев. Все глобальные факторы перечислены в Таблице 1.

Таблица 1. Глобальные транскрипционные факторы в *E. coli*

Подход	Транскрипционные факторы	Ссылка
По количеству регулируемых метаболических путей	LexA, HtpR, GlnG, Crp, PhoB, NifA, IHF, FNR, Fis, ArcA, Lrp, H-NS	[44, 43]
По количеству регулируемых генов или оперонов	ArcA, Fnr, IHF, Crp, RpoS, Rob, Lrp, HNS, Fis, RpoE, Mlc, NarL, CspA	[40, 45, 43]
По количеству регулируемых модулей	IHF, CspA, Crp, Fnr, HNS, OmpR, RpoN, RpoS, ArcA, NarL	[46]

Тип транскрипционных факторов может быть важен потому, что глобальные и локальные транскрипционные факторы по-разному сохраняются в ходе эволюции. Глобальные транскрипционные факторы (такие, как CRP) могут сохраняться во всей таксономической группе и даже за ее пределами, в то время как локальные (например, XylR) часто обнаруживаются только в нескольких близкородственных организмах [47]. Таким образом, тип транскрипционного фактора может влиять на эволюцию структурного мотива, в формировании которого он принимает участие.

Большая часть треугольников принадлежит к одному из двух типов: либо оба транскрипционных фактора глобальные, либо один глобальный, а другой локальный [47], в то время как треугольники с двумя локальными транскрипционными факторами практически не встречаются.

Так как регуляторные блоки часто встречаются в биологических сетях, возникает вопрос об их биологическом значении. В ряде работ были изучены динамические свойства самых распространенных типов регуляторных блоков – С1 и Н1. Согласованный блок типа С1 действует таким образом, чтобы предотвратить экспрессию генов в ответ на кратковременный сигнал, т.е. только долгосрочный стимул способен включить оба транскрипционных фактора для изменения экспрессии регулируемого гена [48]. Примером экспериментально изученной системы на основе согласованного треугольника типа С1 может служить система регуляции оперона *ara*, кодирующего гены метаболизма арабинозы, транскрипционными факторами Cpr и AraC [48]. Несогласованный блок типа Н1, напротив, вызывает кратковременную экспрессию гена в ответ на кратковременный сигнал [48].

Другой интересной научной задачей, связанной с треугольниками, является эволюция этого регуляторного блока в целом и эволюция его частей. Ранее было показано, что треугольники консервативны в организмах, обитающих в похожих условиях и ведущих схожий образ жизни [49]. Также, сохранение трех взаимодействий внутри треугольника скоррелировано, в отличие от сохранения трех

случайных взаимодействий: в треугольнике исчезновение одного взаимодействия приводит к исчезновению остальных [50]. Интересно, что на консервативность треугольника также влияет тип транскрипционного фактора: активаторы в ходе эволюции теряются чаще, чем репрессоры, что может объясняться тем, что для безболезненного исчезновения репрессора необходимо исчезновение всех регулируемых им генов [51]. Однако все упомянутые исследования эволюции треугольников рассматривали их как систему из двух транскрипционных факторов и регулируемого гена, но не принимали во внимание эволюцию других важных элементов, таких как участки связывания транскрипционных факторов.

В настоящей работе мы изучили эволюцию всех элементов треугольников в близкородственных бактериальных организмах на уровне штаммов *Escherichia coli* K12 и порядка Enterobacteriales. Мы предположили, что высокая частота встречаемости и биологическая значимость треугольника также найдут отражение и в эволюции этого блока: взаимодействия внутри треугольника будут консервативнее по сравнению с изолированными взаимодействиями. Отдельно мы изучили консервативность взаимодействия в зависимости от природы транскрипционного фактора и места в сети транскрипции.

Глава 2 - Материалы и методы

2.1 Исследованные геномы

Полные последовательности геномов были получены из базы данных GenBank [3]. Список исследованных геномов отсортирован по степени многочисленности таксономических групп (Таблица 2). Сокращения взяты из базы данных KEGG [52].

Таблица 2. Исследованные геномы

Геном	Ссылка	Сокращение
Proteobacteria: Alphaproteobacteria		
<i>Agrobacterium tumefaciens</i> str. C58	[53]	atu
<i>Bradyrhizobium japonicum</i> USDA 110	[54]	bjj
<i>Bradyrhizobium</i> sp. BTAi1	[55]	bbt
<i>Bradyrhizobium</i> sp. ORS278	[55]	bra
<i>Brucella suis</i> 1330	[56]	bsu
<i>Mesorhizobium loti</i> MAFF303099	[57]	mlo
<i>Mesorhizobium</i> sp. BNC1 (<i>Chelativorans</i> sp. BNC1)	(Нет публикации)	mes
<i>Nitrobacter hamburgensis</i> X14	[58]	nha
<i>Nitrobacter winogradskyi</i> Nb-255	[59]	nwi
<i>Ochrobactrum anthropi</i> ATCC 49188	[60]	oan
<i>Rhizobium etli</i> CFN 42	[61]	ret
<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> 3841	[62]	rle
<i>Rhodopseudomonas palustris</i> BisA53	[63]	rpe
<i>Rhodopseudomonas palustris</i> BisB5	[63]	rpd
<i>Rhodopseudomonas palustris</i> BisB18	[63]	rpc

<i>Rhodopseudomonas palustris</i> CGA009	[64]	rpa
<i>Rhodopseudomonas palustris</i> HaA2	[63]	rpb
<i>Sinorhizobium meliloti</i> 1021	[65]	sme
<i>Xanthobacter autotrophicus</i> Py2	Нет публикации	xau
<i>Azorhizobium caulinodans</i> ORS 571	[66]	azc
<i>Jannaschia</i> sp. CCS1	[67]	jan
<i>Granulibacter bethesdensis</i> CGDNIH1	[68]	gbe
<i>Rhodospirillum rubrum</i> ATCC 11170	[69]	rru
Proteobacteria: Betaproteobacteria		
<i>Azoarcus</i> sp. BH72	[70]	azo
<i>Ralstonia eutropha</i> H16	[71]	reh
<i>Cupriavidus metallidurans</i> CH34	Нет публикации	rme
<i>Ralstonia solanacearum</i> GMI1000	[72]	rso
<i>Acidovorax avenae</i> subsp. <i>citrulli</i> AAC00-1	Нет публикации	aav
<i>Burkholderia cenocepacia</i> AU 1054	Нет публикации	bcn
<i>Burkholderia ambifaria</i> AMMD	[73]	bam
<i>Burkholderia mallei</i> ATCC 23344	[74]	bma
<i>Burkholderia pseudomallei</i> K96243	[75]	bps
<i>Burkholderia</i> sp. 383	(Нет публикации)	bur
<i>Burkholderia thailandensis</i> E264	[76]	bte
<i>Burkholderia vietnamiensis</i> G4	Нет публикации	bvi
<i>Burkholderia xenovorans</i> LB400 [76]	[77]	bxe
<i>Chromobacterium violaceum</i> ATCC 12472 [77]	[78]	cvi
<i>Methylibium petroleiphilum</i> PM1 [78]	[79]	mpe

<i>Rhodoferax ferrireducens</i> DSM 15236	Нет публикации	rfr
<i>Polaromonas naphthalenivorans</i> CJ2	[80]	pna
Proteobacteria: Gammaproteobacteria		
<i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i> A449	[81]	asa
<i>Photorhabdus luminescens</i> subsp. <i>laumondii</i> TTO1	[82]	plu
<i>Serratia proteamaculans</i> 568	(Нет публикации)	spe
<i>Shigella dysenteriae</i> Sd197	[83]	sdv
<i>Sodalis glossinidius</i> str. 'morsitans'	[84]	sgl
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> MAFF 311018	(Нет публикации)	xoo
<i>Citrobacter koseri</i> ATCC BAA-895	(Нет публикации)	cko
<i>Escherichia coli</i> str. K-12 substr. MG1655	[85]	eco
<i>Escherichia coli</i> 536	[86]	ecp
<i>Escherichia coli</i> 55989	[87]	eck
<i>Escherichia coli</i> APEC O1	[88]	ecv
<i>Escherichia coli</i> ATCC 8739	Нет публикации	ecl
<i>Escherichia coli</i> B str. REL606	[89]	ebr
<i>Escherichia coli</i> BL21(DE3)	[89]	ebd
<i>Escherichia coli</i> BW2952	[90]	ebw
<i>Escherichia coli</i> CFT073	[91]	ecc
<i>Escherichia coli</i> E24377A	[92]	ecw
<i>Escherichia coli</i> ED1a	[87]	ecq
<i>Escherichia coli</i> HS	[92]	ecx
<i>Escherichia coli</i> IAI1	[87]	ecr

<i>Escherichia coli</i> IAI39	[87]	ect
<i>Escherichia coli</i> O127:H6 str. E2348/69	[93]	ecg
<i>Escherichia coli</i> O157:H7 EDL933	[94]	ece
<i>Escherichia coli</i> O157:H7 str. EC4115	[95]	ecf
<i>Escherichia coli</i> O157:H7 str. Sakai	[96]	ecs
<i>Escherichia coli</i> O157:H7 str. TW14359	[97]	etw
<i>Escherichia coli</i> S88	[87]	ecz
<i>Escherichia coli</i> SE11	[98]	ecy
<i>Escherichia coli</i> SMS-3-5	[99]	ecm
<i>Escherichia coli</i> UMN026	[87]	eum
<i>Escherichia coli</i> UTI89	[100]	eci
<i>Escherichia coli</i> str. K-12 substr. DH10B	[101]	ecd
<i>Citrobacter rodentium</i> ICC168	[102]	cro
<i>Enterobacter</i> 638	Нет публикации	ent
<i>Enterobacter cloacae</i> subsp. <i>cloacae</i> ATCC 13047	[103]	enc
<i>Cronobacter turicensis</i> z3032	[104]	ctu
<i>Cronobacter sakazakii</i> ATCC BAA-894	[105]	esa
<i>Klebsiella pneumoniae</i> 342	[106]	kpe
<i>Erwinia billingiae</i> Eb661	[107]	ebi
<i>Erwinia amylovora</i> ATCC 49946	[108]	eay
<i>Pectobacterium atrosepticum</i> SCRI1043	[109]	eca
<i>Erwinia tasmaniensis</i> strain ET1/99	[110]	eta
<i>Erwinia pyrifoliae</i> DSM 12163	[111]	epy
<i>Pectobacterium carotovorum</i> subsp. <i>carotovorum</i> PC1	Нет публикации	pct

<i>Pectobacterium wasabiae</i> WPP163	(Нет публикации)	pwa
<i>Pantoea ananatis</i> LMG 20103	[112]	pan
<i>Pantoea vagans</i> C9-1	[113]	pva
<i>Dickeya dadantii</i> Ech586	(Нет публикации)	ddc
<i>Dickeya zeaе</i> Ech1591	(Нет публикации)	dze
<i>Marinobacter aquaeolei</i> VT8	(Нет публикации)	maq
<i>Salmonella enterica subsp. enterica</i> serovar <i>Typhi</i> str. CT18	[114]	sty
<i>Salmonella enterica</i> LT2	[115]	stm
<i>Shigella boydii</i> Sb227	[116]	sbo
<i>Shigella sonnei</i> Ss046	[116]	ssn
<i>Acinetobacter baumannii</i> ATCC 17978	[117]	aba
<i>Marinomonas</i> sp. MWYL1	Нет публикации	mmw
<i>Xanthomonas axonopodis</i> pv. <i>citri</i> str. 306	[118]	xac
<i>Xanthomonas campestris</i> pv. <i>campestris</i> str. 8004	[119]	xcc
<i>Klebsiella pneumoniae subsp.</i> <i>pneumoniae</i> MGH 78578	Нет публикации	kpn
<i>Pseudoalteromonas atlantica</i> T6c	Нет публикации	pat
<i>Pseudomonas mendocina</i> ymp	(Нет публикации)	pmu
<i>Pseudomonas aeruginosa</i> PAO1	[120]	paе
<i>Pseudomonas fluorescens</i> Pf-5	[121]	pfl
<i>Pseudomonas fluorescens</i> PfO-1	[122]	pfo
<i>Pseudomonas stutzeri</i> A1501	[123]	psa
<i>Pseudomonas putida</i> KT2440	[124]	ppu
<i>Pseudomonas syringae</i> pv. <i>tomato</i> str.	[125]	pst

DC3000		
<i>Psychromonas ingrahamii</i> 37	Нет публикации	pin
Proteobacteria: Deltaproteobacteria		
<i>Mycococcus xanthus</i> DK 1622	[126]	mxa
<i>Pelobacter carbinolicus</i> DSM 2380	Нет публикации	pca
<i>Pelobacter propionicus</i> DSM 2379	Нет публикации	ppd
Firmicutes		
<i>Listeria welshimeri</i> serovar 6b str. SLCC5334	[127]	lwe
<i>Listeria innocua</i> Clip11262	[128]	lin
<i>Listeria monocytogenes</i> EGD-e	[128]	lmo
<i>Alkaliphilus metalliredigens</i> QYMF	Нет публикации	amt
<i>Alkaliphilus oremlandii</i> OhILAs	Нет публикации	aoe
<i>Clostridium difficile</i> 630	[129]	cde
<i>Clostridium perfringens</i> ATCC 13124	[130]	cpe
<i>Clostridium phytofermentans</i> ISDg	Нет публикации	cph
<i>Clostridium tetani</i> E88	[131]	ctc
<i>Clostridium acetobutylicum</i> ATCC 824	[132]	cac
<i>Desulfitobacterium hafniense</i> Y51	[133]	dsy
<i>Desulfotomaculum reducens</i> MI-1	(Нет публикации)	drm
<i>Symbiobacterium thermophilum</i> IAM 14863	[134]	sth
<i>Enterococcus faecalis</i> V583	[135]	efa
<i>Streptococcus sanguinis</i> SK36	[136]	ssa
Actinobacteria		
<i>Frankia alni</i> ACN14a	[137]	fal

<i>Streptomyces avermitilis</i> MA-4680	[138]	sma
<i>Mycobacterium ulcerans</i> Agy99	[139]	mul
<i>Mycobacterium smegmatis</i> str. MC2 155	[140]	msm
<i>Mycobacterium vanbaalenii</i> PYR-1	Нет публикации	mva
<i>Nocardia farcinica</i> IFM 10152	[141]	nfa
<i>Rhodococcus</i> sp. RHA1 (<i>Rhodococcus</i> <i>jostii</i> RHA1)	[142]	rha
<i>Nocardioides</i> sp. JS614	[143]	nca
Acidobacteria		
<i>Solibacter usitatus</i> Ellin6076	(Нет публикации)	sus
Bacterioidetes		
<i>Cytophaga hutchinsonii</i> ATCC 33406	[144]	chu
Fusobacteria		
<i>Fusobacterium nucleatum</i> subsp. <i>nucleatum</i> ATCC 25586	[145]	fnu
Chlorophlexi		
<i>Herpetosiphon aurantiacus</i> ATCC 23779	(Нет публикации)	hau

2.2 Программное обеспечение и базы данных

Список гомологичных белков был получен с помощью программы PSI-BLAST [146], ортологичные белки определены с использованием критерия «лучшего двухстороннего поиска» (bidirectional best hit, ВВН). Для выравнивания аминокислотных и нуклеотидных последовательностей использованы программы ClustalX [147], Muscle [148] и MACAW [149]. Для построения филогенетических деревьев и оценки достоверности их топологии («бутстреп-анализ») использован пакет программ PHYLIP [150]. Визуализация филогенетических деревьев

проводилась с помощью программ GeneMaster (А. А. Миронов, неопублик.) и ITOL [151]. Для поиска участков связывания и анализа ортологов использовался пакет программ Genome Explorer [152]. Экспериментально определенные участки связывания транскрипционных факторов были получены из базы данных RegulonDB [1]. Треугольники были определены на основе этой базы данных с помощью скриптов на языке программирования Perl. Построение матриц позиционных весов проводилось с помощью программы Signal X [152] или они были получены из баз данных RegPrecise [153] и RegTransBase [154]. Для построения диаграмм Лого, отображающих структуру участка связывания транскрипционного фактора, использована программа Weblogo (<http://weblogo.berkeley.edu/>). Реконструкция предкового состояния оперонов получена с помощью программы MESQUITE (mesquiteproject.org). Разбиение на функциональные категории проведено на основе данных базы данных NCBI COG [155]. Для оценки статистической значимости использовался критерий χ^2 и гипергеометрический тест, реализованные в программном пакете GNU R.

Глава 3 - Предсказание регуляторных элементов транскрипционной сети на примере пути деградации этаноламина

3.1 Поиск ортологов генов утилизации этаноламина

Гены, кодирующие отдельные ферменты утилизации этаноламина, транспортеры, транскрипционные факторы и структурные белки метаболосомы, обнаружены во многих бактериальных организмах, но без основного фермента – этаноламин лиазы – функциональность пути в целом оказывается под вопросом. Поэтому для изучения транскрипционной регуляции отобраны только такие организмы, в которых есть ортологи двух субъединиц этаноламин лиазы EutB и EutC, и которые, следовательно, могут катаболизировать этаноламин.

Ортологи EutB и EutC были обнаружены в более чем 100 видах бактерий из различных таксономических групп: Proteobacteria, Firmicutes, Actinobacteria, Acidobacteria, Bacteroidetes, Chloroflexi, Flavobacteria, Lentisphaerae, Planctomycetacia и Fusobacteria.

Структура *eut*-оперона оказалась различной в разных группах бактерий. Мы выделили два основных типа подобных оперонов – *короткий*, содержащий минимальный набор генов утилизации этаноламина, и *длинный* (Таблица 3) (Рисунок 7).

Таблица 3. Типы оперонов генов утилизации этаноламина в исследованных геномах

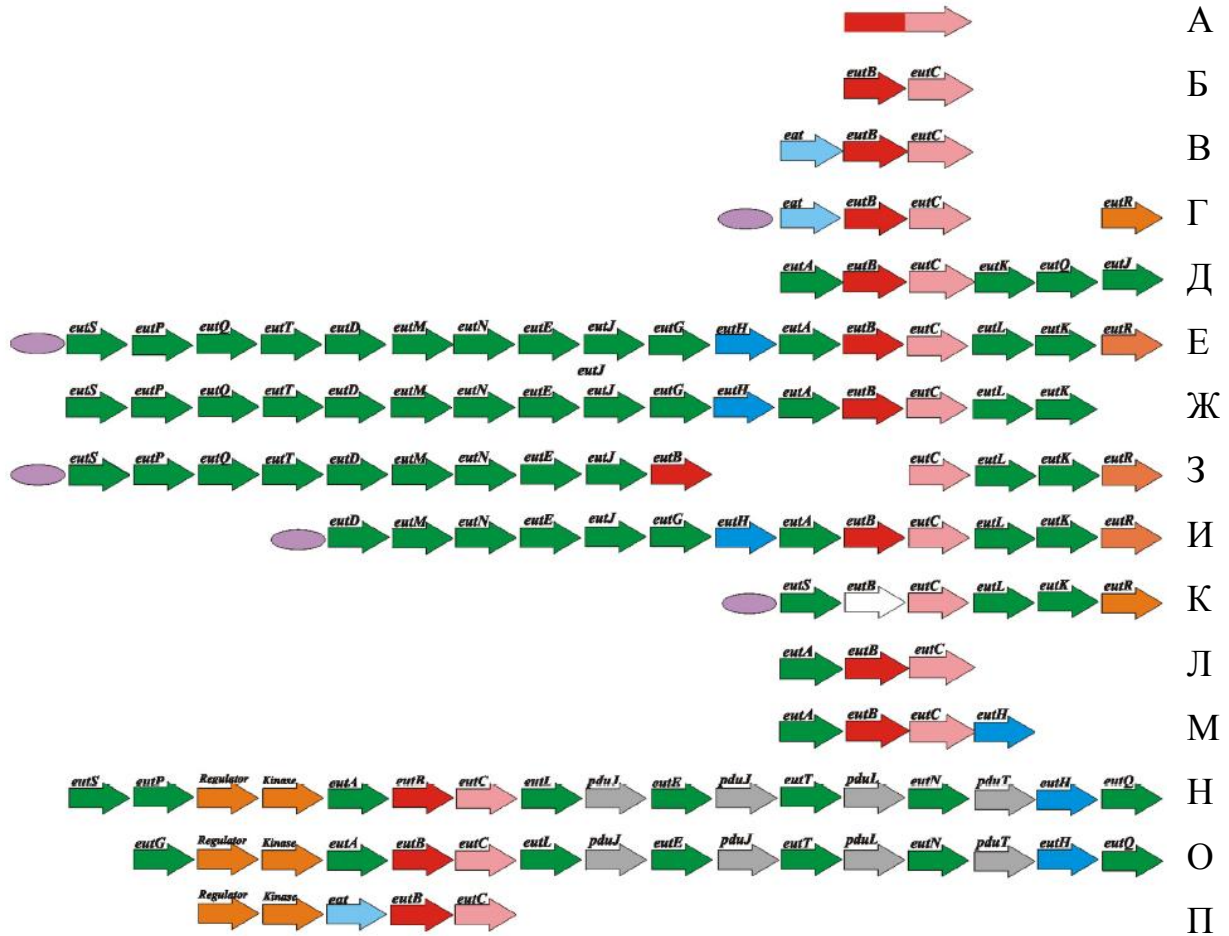
Оперонная структура	Геномы
<i>eutBC</i> (химерный белок)	<i>Mucococcus xanthus</i> DK 1622 <i>Pelobacter carbinolicus</i> DSM 2380 <i>Pelobacter propionicus</i> DSM 2379
<i>eutBC</i>	<i>Bradyrhizobium japonicum</i> USDA 110 <i>Bradyrhizobium</i> sp. BTAi1

	<p><i>Bradyrhizobium sp.</i> ORS278</p> <p><i>Xanthobacter autotrophicus</i> Py2</p> <p><i>Frankia alni</i> ACN14a</p> <p><i>Cytophaga hutchinsonii</i> ATCC 33406</p> <p><i>Herpetosiphon aurantiacus</i></p> <p><i>Azorhizobium caulinodans</i> ORS 571</p> <p><i>Jannaschia sp.</i> CCS1</p> <p><i>Granulibacter bethesdensis</i> CGDNIH1</p> <p><i>Azoarcus sp.</i> BH72</p> <p><i>Ralstonia eutropha</i> H16</p> <p><i>Ralstonia metallidurans</i> CH34</p> <p><i>Ralstonia solanacearum</i> GMI1000</p> <p><i>Acidovorax avenae subsp. citrulli</i> AAC00-1</p> <p><i>Aeromonas salmonicida subsp. salmonicida</i> A449</p> <p><i>Marinomonas sp.</i> MWYL1</p> <p><i>Pseudoalteromonas atlantica</i> T6c</p> <p><i>Pseudomonas aeruginosa</i> PAO1</p> <p><i>Pseudomonas fluorescens</i> Pf-5</p> <p><i>Pseudomonas mendocina</i> ymp</p> <p><i>Pseudomonas syringae pv. tomato str.</i> DC3000</p>
<i>eutABC</i>	<i>Symbiobacterium thermophilum</i> IAM 14863
<i>eat-eutBC</i>	<i>Rhizobium leguminosarum bv. viciae</i> 3841

	<p><i>Solibacter usitatus</i> Ellin6076</p> <p><i>Streptomyces avermitilis</i> MA-4680</p> <p><i>Mycobacterium ulcerans</i> Agy99</p> <p><i>Mycobacterium smegmatis</i> str. MC2 155</p> <p><i>Mycobacterium vanbaalenii</i> PYR-1</p> <p><i>Rhodococcus</i> sp. RHA1</p> <p><i>Clostridium acetobutylicum</i> ATCC 824</p> <p><i>Chromobacterium</i> ATCC 12472</p> <p><i>Sodalis glossinidius</i> str. 'morsitans'</p> <p><i>Xanthomonas oryzae</i> pv. <i>oryzae</i> MAFF 311018</p> <p><i>Marinobacter aquaeolei</i> VT8</p> <p><i>Acinetobacter baumannii</i> ATCC 17978</p> <p><i>Xanthomonas axonopodis</i> pv. <i>citri</i> str. 306</p> <p><i>Xanthomonas campestris</i> pv. <i>campestris</i> str. 8004</p> <p><i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i> MGH 78578</p> <p><i>Pseudomonas fluorescens</i> PfO-1</p> <p><i>Pseudomonas stutzeri</i> A1501</p> <p><i>Pseudomonas putida</i> KT2440</p> <p><i>Psychromonas ingrahamii</i> 37</p>
<i>eutR+eat-eutBC</i>	<p><i>Burkholderia cenocepacia</i> AU 1054</p> <p><i>Burkholderia cepacia</i> AMMD</p> <p><i>Burkholderia mallei</i> ATCC 23344</p>

	<p><i>Burkholderia pseudomallei</i> K96243</p> <p><i>Burkholderia</i> sp. 383</p> <p><i>Burkholderia thailandensis</i> E264</p> <p><i>Burkholderia vietnamiensis</i> G4</p> <p><i>Burkholderia xenovorans</i> LB400</p> <p><i>Methylibium petroleiphilum</i> PM1</p> <p><i>Rhodoferax ferrireducens</i> DSM 15236</p> <p><i>Polaromonas naphthalenivorans</i> CJ2</p>
<i>eutSPQTMNEJGHABCLK</i>	<p><i>Serratia proteamaculans</i> 568</p> <p><i>Shigella dysenteriae</i> Sd197</p> <p><i>Citrobacter koseri</i> ATCC BAA-895</p> <p><i>Escherichia coli</i> str. K-12 substr. MG1655</p> <p><i>Marinobacter aquaeolei</i> VT8</p> <p><i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Typhi</i> str. CT18</p> <p><i>Salmonella enterica</i> LT2</p> <p><i>Shigella boydii</i> Sb227</p> <p><i>Shigella sonnei</i> Ss046</p> <p><i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i> MGH 78578</p> <p><i>Pseudomonas fluorescens</i> Pf-5</p> <p><i>Photorhabdus luminescens</i> subsp. <i>laumondii</i> TTO1</p>
<i>eutSP-response regulator-histidine kinase-eutABCLK(K)EKT(pduAA)pduL-unknown-(eutN)pduN(pduS)pduT-</i>	<p><i>Nocardia farcinica</i> IFM 10152</p> <p><i>Listeria welshimeri</i> serovar <i>6b</i> str. SLCC5334</p>

<i>eutHQ*</i>	<p><i>Listeria innocua</i> Clip11262</p> <p><i>Listeria monocytogenes</i> EGD-e</p> <p><i>Alkaliphilus metalliredigens</i> QYMF</p> <p><i>Alkaliphilus oremlandii</i> <i>OhILAs</i></p> <p><i>Clostridium difficile</i> 630</p> <p><i>Clostridium perfringens</i> ATCC 13124</p> <p><i>Clostridium phytofermentans</i> ISDg</p> <p><i>Clostridium tetani</i> E88</p> <p><i>Desulfitobacterium hafniense</i> Y5</p> <p><i>Desulfotomaculum reducens</i> MI-1</p> <p><i>Enterococcus faecalis</i> V583</p> <p><i>Streptococcus sanguinis</i> SK36</p> <p><i>Fusobacterium nucleatum</i> <i>subsp. nucleatum</i> ATCC 25586</p>
---------------	--



Короткие опероны: А, *Deltaproteobacteria*; Б, *Proteobacteria*, *Chlorophlexi* и *Bacteroidetes*; В, некоторые *Proteobacteria* и *Acidobacteria*; Г, *Betaproteobacteria* (*eutR* обнаружен отдельно от *eutBC* и *eat*). Длинные опероны: Д, *Nocardioidea* sp.; Е, *Enterobacteriaceae*; Ж, *M. aquaeolei*; З, *S. boydii* Sb227; И, *S. sonnei* Ss046; К, *S. dysenteriae* Sd197; Л, *Symbiobacterium thermophilum* and *P. luminescens*; М, *P. fluorescens* Pf-5; Н, *Clostridiaceae* и *F. nucleatum*; О, *Listeriaceae* и *Enterococcaceae*; П, *C. acetobutylicum*.

Белым цветом обозначен псевдоген *eutB*. Предсказанные участки связывания EutR обозначены фиолетовыми овалами (см. «Регуляция утилизации этаноламина»).

Рисунок 7 – Разнообразие оперонных структур, содержащих гены *eutBC*.

Опероны короткого типа состоят из генов основного фермента – *eutBC*, а также могут содержать гены транспортера этаноламина *eat* и транскрипционного фактора *eutR*. Подобная оперонная структура встречается почти во всех *Proteobacteria* и *Actinobacteria*. Представители этих таксономических групп, вероятно, могут расщеплять этаноламин до ацетальдегида и аммиака, и использовать последний в метаболизме азота.

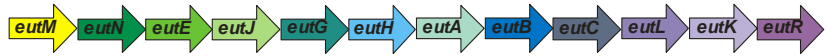
Для *Enterobacteriales* и *Firmicutes* характерен длинный тип оперона, содержащий от 5 до 22 генов (некоторые из этих генов могут и не относиться к метаболизму этаноламина). Подобные организмы, вероятно, приобретают способность использовать этаноламин не только как источник азота, а также как источник углерода и энергии.

Внутри порядка *Enterobacteriales* наблюдается существенное разнообразие состава оперона (Рисунок 8). Так, если *S.enterica*, *C.koseri*, *K.pneumoniae* и *E.coli* содержат полный набор генов, то оперон у *S.sonnei* имеет вид *eutMNEJGHABCLKR*, а у *S.boydii* – *eutPQTMNEBCLKR*. В *S.dysenteriae* сохранились только гены *eutSCLKR*, а в *P.luminescens* – *eutBC*. Оперон *P.luminescens* является сокращенным вариантом длинного типа оперона, а не короткого типа, что явно следует из анализа филогенетических деревьев, о котором подробнее будет сказано ниже.

S. enterica, *C. koseri*, *K. pneumoniae*, *E. coli*



S. sonnei



S. boydii



S. dysenteriae



P. luminescens



Для каждого организма показан один оперон. Ортологичные гены расположены друг над другом.

Рисунок 8 - Полиморфизм оперонных структур у представителей порядка Enterobacteriales.

Мы обнаружили интересные случаи сохранения двух типов оперона независимо внутри одного генома в *K. pneumoniae*, *M. aquaeolei*, *P. fluorescens*.

Некоторые представители Fusobacteria, Chlorobi, Acidobacteria также содержат ортологи отдельных генов утилизации этаноламина. Fusobacteria обладает длинным типом оперона, а Chlorobi и Acidobacteria – коротким.

Самая сложная и разнородная оперонная структура наблюдается у представителей Firmicutes. В опероне содержится не только полный набор генов утилизации этаноламина с замещением *eutR* на гены двухкомпонентной

регуляторной системы, но также встречаются гены пути утилизации пропандиола (Таблица 4).

Таблица 4. Примеры оперонных структур генов утилизации этаноламина у представителей Firmicutes. Условные обозначения: ГК – гистидинкиназа, РО – регулятор ответа, ФН – функция неизвестна.

Геном	Оперонная структура
<i>Alkaliphilus metalliredigens</i>	<i>eutSP-ГК-РО- eutABCLKEKT-pduL- eutJ-ФН-pduNST-eutHG</i>
<i>Alkaliphilus oremlandii</i>	<i>eutSP-pduO-eutGHABCLKEKT-pduL- ФН-pduN-eutHQ</i>
<i>Clostridium difficile</i>	<i>eutSP-ГК-РО-eutABCLKEKT-pduL- ФН-pduNT-eutHQ</i>
<i>Clostridium perfringens</i>	<i>eutSP-ГК-РО-eutABCLKEKT-pduL- ФН-eutNT-eutHQ</i>
<i>Clostridium phytofermentans</i>	<i>eutGSP- eutABCLKEKT-pduL- ФН- pduNT-eutHQ</i>
<i>Clostridium tetani</i>	<i>eutSP-ГК-РО-eutABCLKEKT-pduL- eutJ-ФН-pduN-eutHQ</i>
<i>Desulfotomaculum reducens</i>	<i>eutSP-ГК-РО-eutABCLKT-pduNST- eutQE-pduL</i>
<i>Desulfitobacterium hafniense</i>	<i>eutPABCLT-pduN</i>
<i>Listeria innocua</i>	<i>ГК-РО- eutABCLKEKT-pduL-ФН- pduNT-eutHQ</i>
<i>Listeria monocytogenes</i>	<i>ГК-РО- eutABCLKEKT-pduL-ФН- pduNT-eutHQ</i>
<i>Listeria welshimeri</i>	<i>ГК-РО- eutABCLKEKT-pduL-ФН-</i>

	<i>pduNT-eutHQ</i>
<i>Symbiobacterium thermophilum</i>	<i>eutGBC</i>
<i>Enterococcus faecalis</i>	<i>eutS-ГК-PO-eutBCLKE</i>
<i>Streptococcus sanguinis</i>	<i>eutS- ГК-PO- eutABCLKEKKNQ</i>

3.2 Эволюция системы утилизации этаноламина

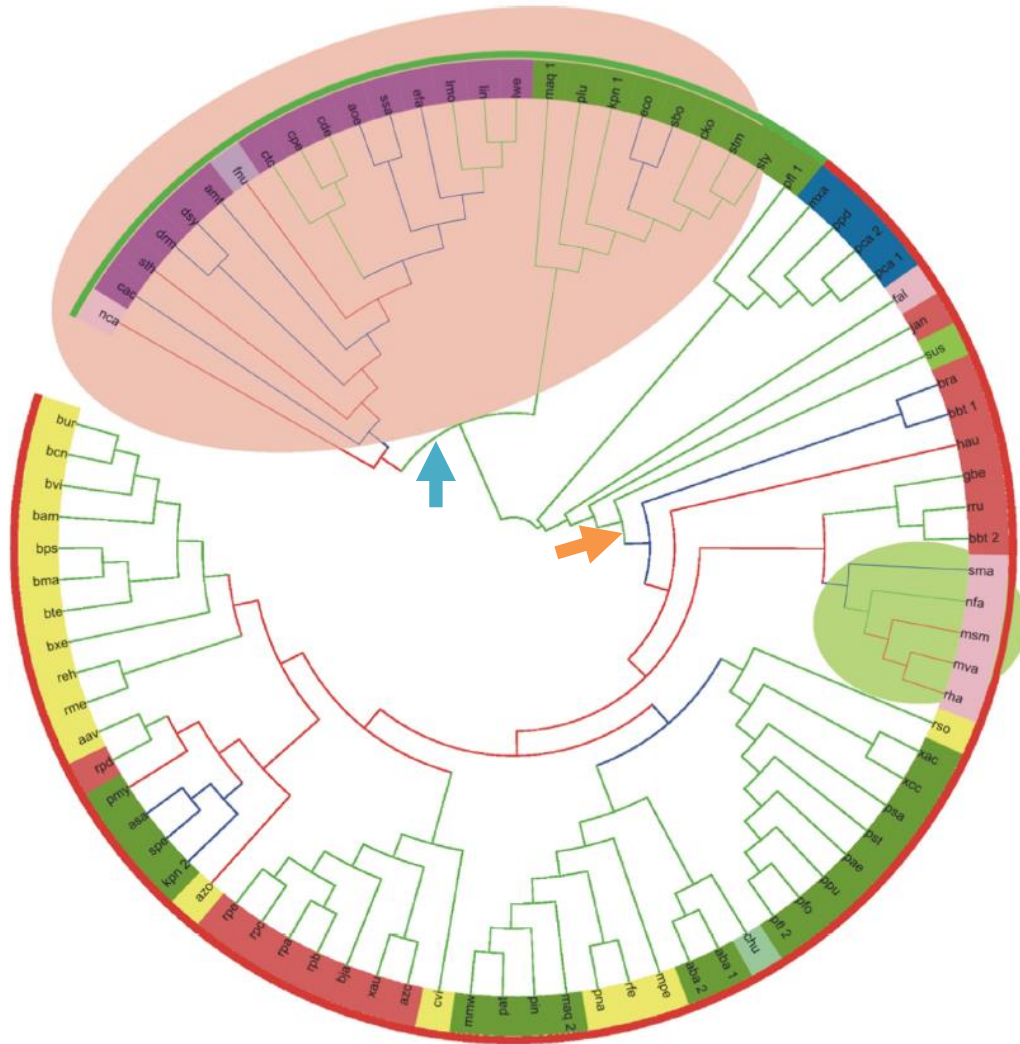
Подробное изучение состава оперонов в различных таксонах бактерий естественным образом поставило вопрос об эволюции системы утилизации этаноламина. На основе аминокислотных последовательности ферментов EutB и EutC из полных геномов были построены филогенетические деревья методом объединения соседних пар («neighbour-joining», см. Материалы и Методы). Топология обоих деревьев оказалась сходной, и потому для последующего анализа можно было использовать любое из деревьев. Было использовано дерево EutB, поскольку для него существует возможность укоренения – можно использовать внешнюю группу, состоящую из белковых последовательностей большой субъединицы пропандиол дегидратазы PduC (Рисунок 9). Для оценки достоверности наблюдаемого расположения ветвей применялось статистическое размножение выборки (“бутстреп-анализ”, см. Материалы и Методы). Каждому расщеплению ветвей внутри дерева соответствует значение (“бутстреп-значение”), которое равно вероятности наблюдать такое же расщепление ветвей внутри филогенетических деревьев, построенных на основе выравниваний, многократно созданных случайным выбором позиций исходного выравнивания. Высокие бутстреп-значения ветвей дерева говорят о высокой достоверности его топологии.

Мы обнаружили, что гены *eutB* из разных типов оперонов, короткого и длинного, представляют отдельные ветви. Первая ветвь содержит белки из Proteobacteria и Actinobacteria, вторая – из Enterobacteriales, относящиеся к Proteobacteria, и Firmicutes (Рисунок 10). В случае геномов, содержащих два *eut-*

оперона (*K. pneumoniae*, *M. aquaeolei*, *P. fluorescens*), *eutB* из оперонов разного типа лежат на разных ветвях.

Случаи, когда расположение ветвей на филогенетическом дереве не согласуется с таксономией, представляют особый интерес (Рисунки 9, 10). Такая топология дерева свидетельствует о возможном событии горизонтального переноса генов. *Горизонтальный* перенос генов – процесс передачи генетического материала от одного организма к другому, не являющегося его потомком. В противном случае перенос генов называется *вертикальным*.

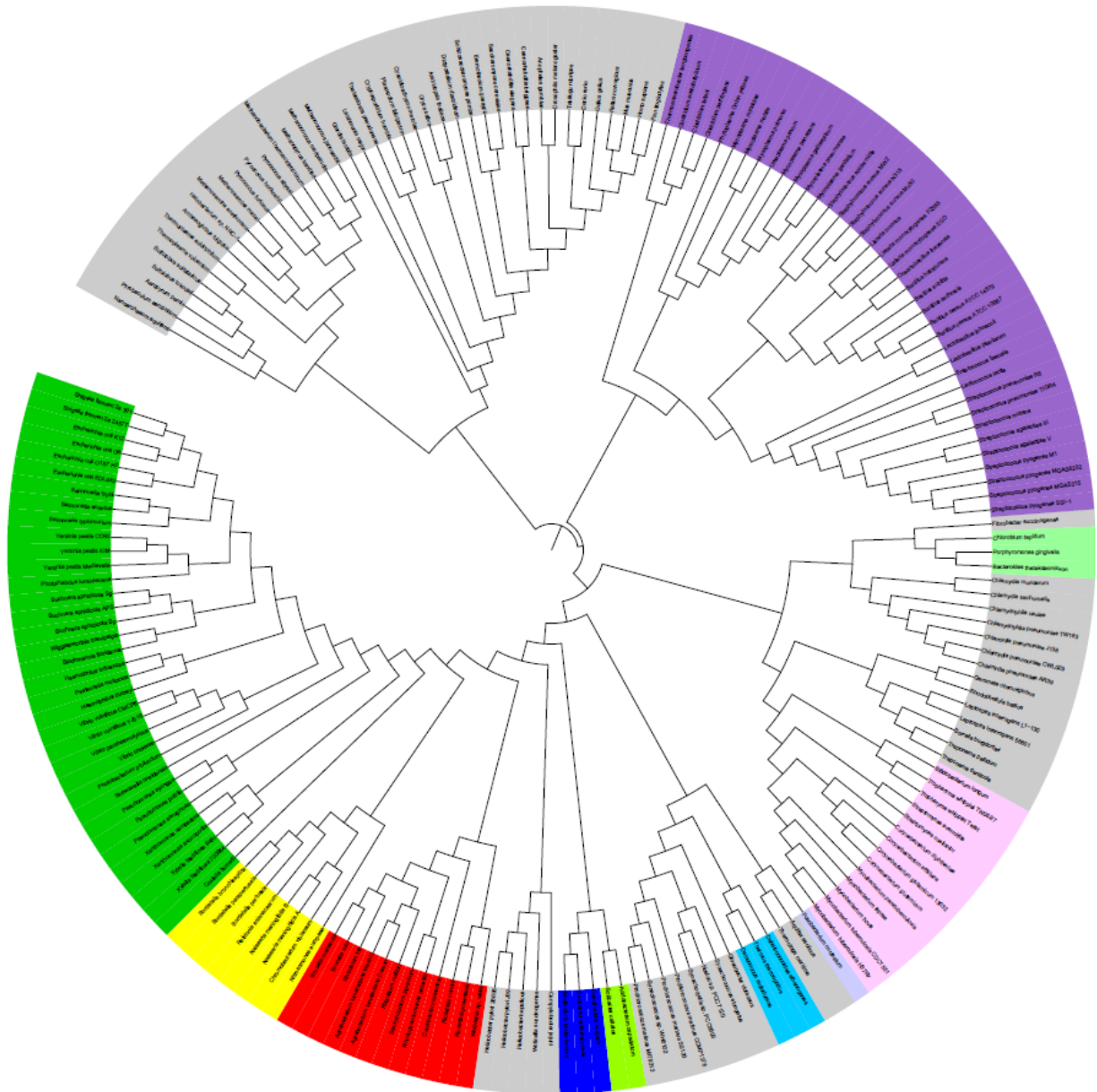
Первый случай отклонения от таксономии был обнаружен для белков бактерий типа Actinobacteria. Ветвь Actinobacteria оказывается внутри ветви, соответствующей белкам из типа Proteobacteria (бутстреп-значение_{cutB}1 > 70%, Рисунки 9, 10), в то время как на дереве видов (Рисунок 11) ветви Actinobacteria и Proteobacteria являются равнозначными. Кроме того, и Actinobacteria, и Proteobacteria имеют короткий тип оперона. Такую топологию можно объяснить следующим образом: вначале произошел горизонтальный перенос короткого оперона генов утилизации этаноламина из Proteobacteria к общему предку Actinobacteria, а далее он распространился вертикальным наследованием по некоторым организмам этого таксона. Таким образом, гены утилизации этаноламина из Actinobacteria являются потомками генов из Proteobacteria.



Цвет ветвей соответствует бутстреп-значениям: зеленый - $> 70\%$, синий – от 50 до 70% , красный – $< 50\%$. Цвет внешнего круга соответствует длине оперона: зеленый – длинный оперон, красный – короткий оперон. Цвет внутреннего круга соответствует таксономической принадлежности: красный, Alphaproteobacteria; желтый, Betaproteobacteria; зеленый, Gammaproteobacteria; синий, Deltaproteobacteria; фиолетовый, Firmicutes; розовый, Actinobacteria; зелено-желтый, Acidobacteria; светло-фиолетовый, Fusobacteria; голубой, Chlorophlexi. Овалами обозначены гены, расположение которых на дереве не согласуется с таксономией, т.е. вероятные события горизонтального переноса генов. Стрелками обозначены ветви, бутстреп-значения которых поддерживают гипотезу горизонтального переноса генов:

оранжевая стрелка – бутстреп-значение $_{\text{eutB}1}$, голубая стрелка – бутстреп-значение $_{\text{eutB}2}$ (см. текст). Обозначения геномов взяты из базы данных KEGG [52]. Детальное описание обозначений геномов: см. Материалы и методы. Длины ветвей не отражают расстояния между видами.

Рисунок 10 – Сводное филогенетическое дерево EutB, построенное методом объединения соседних пар с бутстреп-анализом.



Цвет круга соответствует таксономической принадлежности: красный, Alphaproteobacteria; желтый, Betaproteobacteria; зеленый, Gammaproteobacteria;

синий, Deltaproteobacteria; фиолетовый, Firmicutes; розовый, Actinobacteria; зелено-желтый, Acidobacteria; светло-фиолетовый, Fusobacteria; голубой, Chlorophlexi; и светло-зеленый, Bacteroidetes. Серым обозначены таксоны, в которых не найдены ортологи генов деградации этаноламина. Длины ветвей не отражают расстояния между видами.

Рисунок 11 – Таксономическое дерево видов, полученное из ITOL [150].

Второй случай отклонения от таксономии обнаружен в *Fusobacterium nucleatum*. Ветвь, соответствующая белкам из *F. nucleatum* (тип Fusobacteria), находится внутри ветви Firmicutes (бустреп-значение_{cutB2}>70%, Рисунки 9, 10). Fusobacteria и Firmicutes являются равнозначными таксонами (Рисунок 11), поэтому и здесь, вероятно, также имел место относительно недавний горизонтальный перенос из Firmicutes в Fusobacteria. Дополнительным свидетельством недавнего переноса является то, что в остальных четырех представителях Fusobacteria, для которых доступна полногеномная последовательность (*Ilyobacter polytropus*, *Leptotrichia buccalis*, *Sebaldella termitidis*, *Streptobacillus moniliformis*), генов утилизации этаноламина не найдено.

Самый яркий пример отличия топологии дерева EutB от дерева таксонов – расщепление ветви Gammaproteobacteria, относящейся к Proteobacteria, на две (Рисунок 9, 10). Расположение одной из ветвей соответствует таксономии (Рисунок 9-11), а другая ветвь (порядок Enterobacteriales) становится сестринской к Firmicutes. Интересно, что для организмов из первой ветви характерен короткий тип оперона, в то время как организмы второй ветви, как и Firmicutes, содержат длинный тип оперона. Можно предположить две возможные причины возникновения такой топологии. Во-первых, стоит отметить, что только для этих организмов характерно наличие специального компартмента для ферментов утилизации этаноламина – метаболосомы. По этой причине можно предположить, что эволюция ферментов в

условиях макромолекулярного комплекса с метаболосомой могла проходить конвергентно в двух таксономически далеких ветвях. Но более вероятной является возможность еще одного события горизонтального перенос генов. В таком случае, остается открытым вопрос, происходил ли горизонтальный перенос генов от *Proteobacteria* к *Firmicutes*, или наоборот.

Для исследования поставленного вопроса мы реконструировали вероятное строение оперона *eut* у предка *Firmicutes*, предка *Proteobacteria* и предка *Enterobacteriales* (Таблица 5), используя метод максимальной экономии (*maximal parsimony*). В качестве исходных данных методы реконструкции предкового состояния признака рассматривают наблюдаемое распределение признака в организмах (в данном случае, состав оперона генов утилизации этаноламина) и филогенетическое дерево организмов. Используемый далее метод максимальной экономии в каждом узле дерева минимизирует количество событий, которые привели к наблюдаемому распределению признака.

Таблица 5. Реконструкция состава *eut*-оперонов у предков порядка *Enterobacteriales* и типов *Firmicutes* и *Proteobacteria*. Условные обозначения – «0» – высокая вероятность отсутствия ортолога гена в предковом опероне; «1» – высокая вероятность наличия ортолога гена в предковом опероне; «0,1» – наличие или отсутствие гена достоверно не предсказывается.

Ген	Предок		
	<i>Enterobacteriales</i>	<i>Firmicutes</i>	<i>Proteobacteria</i>
<i>eat</i>	0	0	1
<i>eutP</i>	1	0,1	0
<i>eutQ</i>	1	0,1	0

<i>eutT</i>	1	0,1	0
<i>eutD</i>	1	0	0
<i>eutJ</i>	1	0	0
<i>eutG</i>	1	0,1	0
<i>eutH</i>	1	0,1	0
<i>eutA</i>	1	1	0
<i>eutB</i>	1	1	1
<i>eutC</i>	1	1	1
<i>eutR</i>	0	0	0,1
<i>Two-component system</i>	0	1	0

На основании реконструкции составов предковых оперонов можно выдвинуть следующую гипотезу об эволюции генов утилизации этаноламина. Из всех генов утилизации этаноламина только *eutBC* были у предков всех трех исследуемых таксономических групп. Таким образом, предковый оперон с высокой вероятностью состоял только из генов *eutBC*. Опираясь на тот факт, что похожий тип оперона у ныне живущих бактерий масштабно представлен только у Proteobacteria, мы выдвинули первую гипотезу. Она состоит в том, что гены утилизации этаноламина *eutBC* были у общего предка всех Proteobacteria. Позднее к ним добавился транспортер *eat*. Далее этот основной набор генов был дополнен геном, кодирующим транскрипционный фактор EutR. В таком виде – *eat-eutBC* и смежный ген *eutR* – состав генов утилизации этаноламина сохранился у ныне живущих представителей Proteobacteria.

В ветви Enterobacteriales предковый оперон был дополнен структурными генами метаболосомы, геном синтеза аденозилкобаламина *eutT* и генами расщепления этаноламина до этанола и ацетата с получением АТФ (Рисунок 1).

Что касается Firmicutes, то их *eut*-опероны могли возникнуть в результате горизонтальных переносов от Enterobacteriales с заменой транскрипционного фактора EutR на двухкомпонентную регуляторную систему (как, например, гены *lin1136* и *lin1137* в *Listeria innocua*).

В результате горизонтальных переносов от Enterobacteriales к другим Proteobacteria в геномах *M. aquaeolei* VT8, *K. pneumoniae subsp. pneumoniae* MGH 78578, *P. fluorescens* Pf-5 возникло два независимых *eut*-оперона.

Вторая гипотеза состоит в том, что путь утилизации этаноламина был еще раньше – у гипотетического общего предка Firmicutes и Proteobacteria. Дальнейшая эволюция *eut*-оперонов в этих ветвях привела к независимому образованию короткого и длинного типов. Короткий тип развился в Proteobacteria, а длинный – в Firmicutes. Горизонтальный перенос генов от Firmicutes к древней бактерии, предку порядка Entrobacteriales, привел к тому, что у большинства Enterobacteriales возник длинный *eut*-оперон. Геномы Proteobacteria с двумя *eut*-оперонами (*M. aquaeolei* VT8, *K. pneumoniae subsp. pneumoniae* MGH 78578, *P. fluorescens* Pf-5) являются результатами горизонтальных переносов либо от Firmicutes, либо от Enterobacteriales, при которых собственный оперон утилизации этаноламина еще не успел деградировать.

3.3 Регуляция утилизации этаноламина

EutR – транскрипционный фактор семейства AraC. Он обнаруживается у Enterobacteriales в составе *eut*-оперона и у Burkholderiales в непосредственной близости от *eut*-оперона. У Firmicutes и *F. nucleatum* (Fusobacteria) внутри *eut*-оперона находятся гены двухкомпонентной регуляторной системы, которые, предположительно, и контролируют утилизацию этаноламина в этих организмах. В остальных группах бактерий не обнаружено ни ортологов EutR, ни ортологов генов белков двухкомпонентной системы. В этих организмах транскрипционный фактор еще не описан (либо он расположен на большом расстоянии от генов деградации

Зеленым отмечены ветви, у представителей которых обнаружен ортолог транскрипционного фактора EutR. Обозначения геномов взяты из базы данных KEGG [52]. Детальное описание условных обозначений: см. Материалы и методы.

Рисунок 12 – Филогенетическое дерево для белка EutB, построенное методом объединения соседних пар.

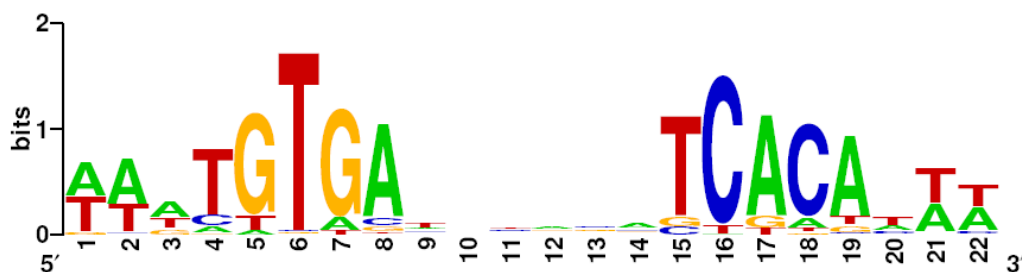
У Enterobacteriales *eut*-оперон содержит 17 генов, первым из которых является ген *eutS*. Методом филогенетического футпринтинга в промоторной области гена *eutS* найдены два консервативных участка. Один из них при детальном рассмотрении обнаружил сходство с участком связывания CRP – wwwTGTGAtyurgwTCACTtWt (Рисунок 13, для сравнения Рисунок 14 и Рисунок 15). Известно, что глобальный транскрипционный фактор Crp прямо или опосредованно контролирует экспрессию генов утилизации пропандиола [37]. Пути утилизации этаноламина и пропандиола биохимически очень похожи, более того этаноламин является производным диолов, в котором вторая ОН-группа заменена на аминогруппу. Таким образом, можно предположить, что гены утилизации этаноламина также регулируются транскрипционным фактором Crp.

cko_eutS	tttcactttaattcattccggcgcggggttccccgcgcttttttatttctctct--TTCT
sty_eutS	gcaccacgatatttatcgttaacaccttcgacagaagtcccttttattttcttt--CTCT
eco_eutS	tttatattcatgcggttgcatatgaaagtttatgcaccacagcgaatatctctcca--TTCT
kpn_eutS	ggtgagtaaaaaacgctgagcctgcctttttaattccgcccgaatattattttctcTCCT
sd_y_eutS	tttatattcatgcggttgcatatgaaagtttatgcaccacagcgaatatctctcca--TTCT
cko_eutS	TAGTGATCTACTTCACCTTTTAAAtctgCTTGCCGAATTTTGTTATTTACtctgacaaaa
sty_eutS	TAGTGATCTACCTCACCTTTTACAatcacCTTGCCGAATTTTGTTATTTACtccgacaaaa
eco_eutS	TAGTGATCTACCTCACCTTTTAAAcgcgCTTGCCGAATTTTGTTATTTACtctgacgaaa
kpn_eutS	GAGTGATCCACTTCACCTTTTAAAcccaCTTGCCGAATTTTGTTATTTGCcgcgctgaaa
sd_y_eutS	TAGTGATCTACCTCACCTTTTAAAcgcgCTTGCCGAATTTTGTTATTTACtctgacgaaa

Участок связывания Crp Участок связывания EutR

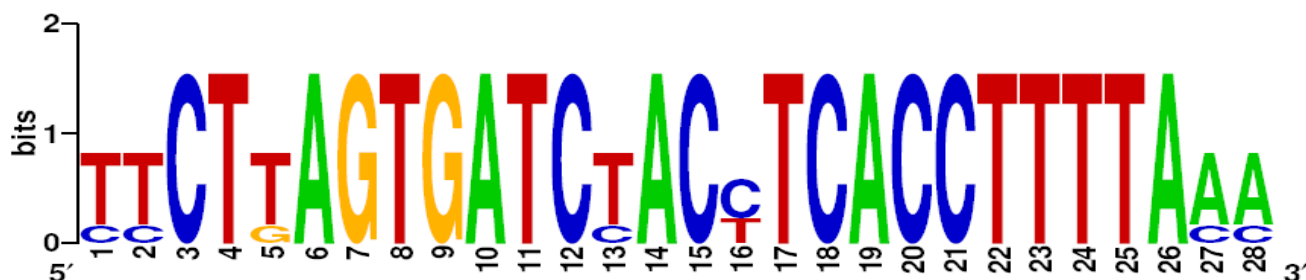
Рисунок 13 – Множественное выравнивание регуляторных областей *eut*-оперона геномов бактерий порядка Enterobacteriales.

Обозначения геномов взяты из базы данных KEGG. Детальное описание условных обозначений: см. Материалы и методы.



По горизонтальной оси указан номер позиции нуклеотида, по вертикальной – информационное содержание позиции в битах. Относительная высота каждой буквы соответствует частоте нуклеотида в данной позиции.

Рисунок 14 – Диаграмма logo для участка связывания CpG.



Условные обозначения – см. Рисунок 14.

Рисунок 15 – Диаграмма logo для предполагаемого участка связывания CpG в регуляторной области *eut*-оперона.

Мы предположили, что вторая консервативная область является участком связывания EutR (Рисунок 13). Косвенным подтверждением этого предположения является тот факт, что наличие транскрипционного фактора EutR коррелирует с наличием найденного участка связывания в промоторной области оперона *eut* (в геномах, не содержащих гена транскрипционного фактора EutR, подобного участка найдено не было).

У представителей Burkholderiales оперон содержит ген транспортера этаноламина *eat* и гены этаноламин лиазы *eutBC*. В промоторной области гена *eat*

была обнаружена только одна консервативная область (Рисунок 16). При сравнении обнаруженных участков (Рисунок 17) выявилось их сходство, что также подтверждает наши предположения о том, что это участок связывания именно транскрипционного фактора EutR.

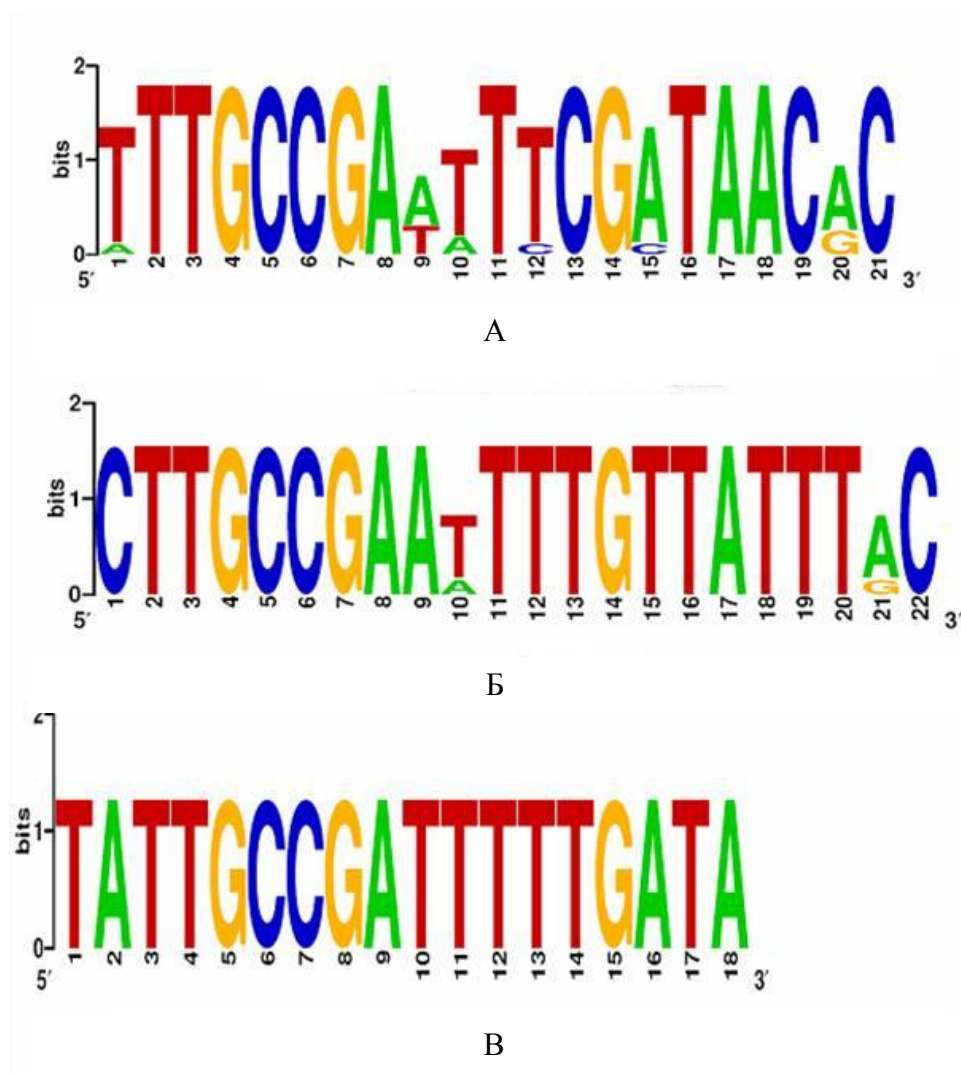
```

mpe_eat   ttcaa CATTGCCGATTTCCGCTAACGC tccctctgcgct
pna_eat   tgaaa CTTTGCCGAAATTCGATAACGC ccaaaaaccagg
rfe_eat   aaaaa CTTTGCCGAAATTCGATAACGC aaaaaagctgcc
buc_eat   ctcat CTTTGCCGAAATTCGATAACAC gtcctttaatga
bur_eat   gcgat CTTTGCCGAAATTCGATAACAC gtcgtttaatga
bvi_eat   gccat CTTTGCCGAAATTCGATAACAC gtcgttttatga
bce_eat   gcgat CTTTGCCGAAATTCGATAACAC gtcgtttaatga
bup_eat   cggac GTTTGCCGATTTTCGATAACAC gtcgattaatga
Bum_eat   cggac GTTTGCCGATTTTCGATAACAC gtcgattaatga
bth_eat   cggac GTTTGCCGATTTTCGATAACAC gtcgattaatga
bxе_eat   ctgaa CTTTGCCGAAATTCGATAACAC ctсgaataatgg

```

Обозначения геномов взяты из базы данных KEGG [52]. Детальное описание условных обозначений: см. Материалы и методы.

Рисунок 16 – Множественное выравнивание регуляторных областей *eut*-операона геномов бактерий порядка Burkholderiales.



Условные обозначения – см. Рисунок 14.

Рисунок 17 – Сравнение диаграмм Лого для предполагаемых участков связывания EutR в регуляторных областях *eut*- и *cob*- оперонов: А) Участок связывания EutR в регуляторной области *eut*-оперона у Burkholderiales; Б) Участок связывания EutR в регуляторной области *eut*-оперона у Enterobacteriales; В) Участок связывания EutR в регуляторной области *cob*-оперона у Enterobacteriales.

На основании полученных последовательностей участков связывания EutR была построена матрица позиционных весов. Эту матрицу использовали для поиска

генов, экспрессия которых может регулироваться EutR. Поиск охватывал область -400...+100 от старта трансляции.

Сканирование при помощи матрицы позиционных весов обнаружило около 60 генов из 7 регулонов, относящихся к разным метаболическим путям (Таблица 6). Потенциальный участок связывания EutR консервативен во всех изученных геномах Enterobacteriales в промоторной области гена *stpA*, кодирующего репрессорный транскрипционный фактор [156]. Связь этого транскрипционного фактора с утилизацией этаноламина неясна, но он может опосредованно принимать участие в ее регуляции.

В некоторых геномах потенциальный участок связывания EutR обнаружен перед генами синтеза клеточной стенки, метаболизма ксилозы и мембранного белка. Функциональная связь продуктов этих генов с системой утилизации этаноламина также неизвестна.

Похожая консервативная область обнаружена также перед геном *cbiA* – первым геном оперона синтеза кобаламина. Кобаламин является кофактором основного фермента утилизации этаноламина [157], поэтому этот случай представляет особый интерес (Таблица 6).

Кобаламин (витамин В12) имеет сложную химическую структуру, состоящую из корринового кольца, в центре которого располагается ион кобальта. Он является кофактором для ряда ферментативных реакций, в ходе которых протекает расщепление ковалентной связи углерод-кобальт самого кобаламина [157]. Гены синтеза кобаламина представляют единый *cob*-оперон – *cbiABCDEFGHIJKLMNQOP-cobUST*. Известно, что в ряде геномов *cob*-оперон контролируется транскрипционным фактором утилизации пропандиола PoxR [38]. Кроме того, синтез кобаламина часто ингибируется с помощью В12-рибопереключателем [158].

Таблица 6. Участки связывания EutR перед потенциальными членами регулона. Условные обозначения: «+» – обнаружен потенциальный участок связывания непосредственно перед геном; «*» – потенциальный участок связывания обнаружен перед опероном, в который входит данный ген; «-» – потенциальный участок связывания перед опероном не найден;

«0» – ортологов данных генов не обнаружено. Условные обозначения геномов взяты из базы данных KEGG [52]. Детальное описание обозначений геномов см. Материалы и методы.

Гены	Геномы					Функция
	kpn	sty	cko	eco	sdv	
<i>yabB</i>	+	+	-	+	+	Неизвестный белок
<i>mraZW-ftsLI- murEF-mraY- murD-ftsW-murGC- ddlB-ftsQAZ-lpxC</i>	*	*	*	*	*	Деление клетки/синтез клеточной стенки
<i>cbiA</i>	+	+	+	0	0	Синтетаза а,с- диамида кобириновой кислоты
<i>cbiBCDETFGHJKL MNQOP-cobUST</i>	*	*	*	0	0	Синтез кобаламина
<i>yciW</i>	-	+	+	+	+	Неизвестный белок
<i>yhhM</i>	-	+	+	+	+	Белок внутренней мембраны
<i>xylF</i>	+	0	+	+	-	Связывающий компонент периплазмы

<i>xulGHR</i>	*	0	*	*	*	АВС транспортер ксилозы
<i>eutS</i>	+	+	+	+	+	Белок карбоксисомы
<i>eutPQTDMEJGH</i> <i>ABCKLR</i>	*	*	*	*	*	Утилизация этанолamina
<i>stpA</i>	+	+	+	+	+	ДНК-связывающий белок

Обнаруженный участок связывания EutR перекрывается с одним из участков связывания RosR. Видимо, как и в случае утилизации пропандиола, транскрипционный фактор EutR также может одновременно контролировать утилизацию этаноламина и синтез необходимого для этого кофактора – кобаламина (Рисунок 18).

```
cko TGATAATTCGTCCGCCGAGTCCGCA TAAATAGCAATATTTTGC TATTGCCGATTTTGGAT
sty -----TGAATAGCAATATTTTGC TATTGCCGATTTTGGAT
kpn --CCGGATCACAAACTGAGTCTTTT TAAATGACAATATTTTGC TATTGCCGATTTTGGAT
```

Участок связывания RosR Участок связывания EutR

```
cko ATCGTATTGTATGTCAGTCATTATAAAATGAAAAAACT ATAGCAGAGCAATAAATTGTC
sty AGTCTTTTGTATCTTAATCATTTTCAGAAAGAAAATTTAT CTGGTGTAACAATAAATTGTC
kpn ATCCCGAGTATCTTTCATCAATATAAAATGAAAAAACG ACAGTAGAGCAATAAATTGTC
```

Участок связывания RosR

```
cko ATAGCGCAACAAAATAATGTTATCTGAAGGATAAACTGAGCCAGTTTTTATTTTTGTGAT
sty ATAGCGCAACAAAATAATAAAATTTAGGGCATAAACTGCACCAGTTTTTATTTTTGTGAT
kpn ATAGCGCAACAAAATAGTGATATTCGAGGGGTAAACTACACCAGTTTTTATTTTTGTGAT
```

Участок связывания RosR (продолжение)

```
cko GAGATACAGTGTCATTATGAATAGGTTTCATTATTTCTTTTGCAGTGTTCAT--TCAA
sty GAAGTACAGTGTCAGGAAAGATAAGTTTTTATTACGCTTTGTGATACGTTTTTATCGTCAA
kpn GAGATAGAGTGTCG-----TTAATA-----CGAGACAGCTGATTATCAC
```

```
cko AACA-----GGCGTCATTACACCATTG-----CGTTGGCATAATGGCGC--
```


sty TATACCGGTAAGGATGAGTAGATTAACGTCAGATGAGCAACCGCTCAGGCTATTGCTCAC
kpn AACACC-----GCTTTCCTTCACTGACGT-----CATGTGACGAGTGATGA-

cko -----GTAACAGGTAATTAAAATTTAATTGTCGTTATGAGCGAGAAAGTGATGCTCATA
sty AGAAATGTAAAACGTAGCACATTATTAATTGTCGTTATGGGTGTGCCGT--ACAGCCATA
kpn --AACGGTAAAAGAAAACATAAAGT-AATACCCGTCG-----ATGCAGGGCGCATCGGCG

cko ACATCAGCCGACAGGTTTGGCACA---TTTGTGTCAGGGAAGGGGGTGAGAATCCCCCGC
sty ACGTAAACCAACAGGTTTGCCACA---TTTGTGGTAGGGAAGGGGGTGAAAATCCCCCGC
kpn AAACAAACCGACAGGTTTCGACACCAGTTTGGTGTGTCAGGGAAGGGGGTGTAATCCCCCGC

B12-рибопереключатель

cko AGCCCCGCTGCTGTGATGCTGACGACCCCGTAA-GACCACTGATCGCAAGATTGGGAAG
sty AGCCCCGCTGCTGTGATGCTGACGACCCCGTAAAGACCACTGATCGCAAGATTGGGAAG
kpn AGCCCCGCTGCTGTGATGCTGACGACCCCGCAAAAACCACTGATCGAAAGATTGGGAAG

B12-рибопереключатель (продолжение)

cko GACGGGCGAGGAGGACGСТАAGCCAGAAGACCTGCCTGTCGGTAT-ТТАССААСААСТТС
sty GACGGGCGAGGAGGACGСТАAGCCAGAAGACCTGCCTGTCGGTGA-ТАССААСААСТТС
kpn GGCGGGCGAGAACGAGGСТАAGCCAGAAGACCTGCCTGCCGGТААСТГАССААСААСТТС

B12-рибопереключатель (продолжение)

Условные обозначения геномов: см. Материалы и методы.

Рисунок 18 – Множественное выравнивание регуляторных областей *cob*-оперона геномов бактерий порядка Enterobacteriales.

Глава 4 - Эволюция регуляторных блоков на примере блока треугольник

4.1 Определение локальных и глобальных транскрипционных факторов

Как было сказано выше, тип транскрипционного фактора может влиять на эволюцию транскрипционной сети в целом.

Мы объединили несколько подходов для определения глобальных транскрипционных факторов. Транскрипционный фактор считался глобальным, если он, во-первых, принимает участие в регуляции более 20 оперонов, и, во-вторых, эти опероны принадлежат разным метаболическим путям. Информация о метаболических путях взята из базы данных KEGG [52], данные об оперонах – из базы данных RegulonDB [1] (Таблица 7). К списку семи универсальных глобальных факторов (CRP, IHF, FNR, Fis, ArcA, Lrp, H-NS, Таблица 1) был добавлен FUR. FUR регулирует более 40 оперонов из различных метаболических путей: ответ на стресс, гликолиз, глюконеогенез, устойчивость к ионам кобальта и никеля, цикл Кребса, образование поринов, дыхание, метаболизм пуринов, хемотаксис, биосинтез метионина, биосинтез 2,3-дигидроксибензоата [159, 160].

Таблица 7. Количество оперонов, регулируемых различными транскрипционными факторами

Транскрипционный фактор	Количество регулируемых оперонов
CRP	200
FNR	109
Fis	98
IHF	80
ArcA	63

H-NS	59
Fur	44
Lrp	38
NsrR	36
LexA	35
CpxR	33
NarL	31
FruR	24
SoxS	22
PhoP	21
MarA, FlhDC	20
PurR	19
GadE	17
PhoB, GadX, ArgR	16
OxyR, NtrC	15
OmpR	13
Rob	12
NarP, RcsAB	11
Nac, MetJ	10
CysB, GadW, FadR, CytR, TyrR, EvgA	9
NagC, NagC, RutR,	8
PdhR, DnaA, IscR	7
MetR, ExuR, AraC, FhlA, FhlA, GntR	6
TrpR, TorR, LeuO, GalS, GalR, HU, CsgD, MalT	5
ArgP, UxuR, BaeR, AsnC, AscG, CdaR, RstA,	4

GlpR, Zur	
Ada, GcvA, SgrR, MprA, RhaS, NanR, DsdC, PspF, DeoR, LrhA, NhaR, IdnR, NrdR, AgaR, NadR, SoxR, AcrR, CueR, DcuR, CynR, AllR, SdiA	3
EnvY, GlcC, HcaR, BetI, IlvY, CadC, BirA, BolA, PaaX, AlsR, PutA, PrpR, FabR, Cbl, NorR, DhaR, CspA, PepA, MngR, AppY, UidR, Mall, CusR, RcnR, FucR, StpA, UlaR, YdeO, IclR, QseB, ZraR, LysR, LsrR, CaiF, MelR, XylR	2
AlaS, GatR, MqsR-YgiT, KdpE, KdgR, ZntR, RbsR, MarR, GlrR, AlpA, RtcR, MhpR, HdfR, ArsR, AdiY, LacI, HyfR, GutR, TdcR, UhpA, BglI, FeaR, EnvR, AtoC, MurR, CreB, DicA, EbgR, TdcA, CsiR, YeiL, TreR, NikR, RhaR, ChbR, RelEB, AllS, SlyA, MtlR, NemR, LldR, XapR, MntR, YoeB-YefM, YiaJ, GutM, HipB	1

Ряд транскрипционных факторов не был определен как глобальные, несмотря на то, что они принимают участие в регуляции большого числа оперонов. Это NsrR, LexA, CpxR, NarL. Каждый из этих транскрипционных факторов регулирует один метаболический путь или функциональную систему: NsrR принимает участие в защите клетки от оксида азота [161, 162], LexA связан с SOS-репарацией [163], CpxR – с конъюгацией [164], а NarL регулирует гены анаэробного дыхания и защиты клетки от нитритов и нитратов [165]. Таким образом, они подходят только под один из двух критериев глобального транскрипционного фактора, принятых в этой работе, и далее рассматривались как локальные.

4.2 Функциональная аннотация треугольников

Регуляторные блоки могут быть равномерно представлены во всех метаболических путях (функциональных системах) или сосредоточены только в некоторых. Анализ функциональных категорий по базе данных COG [154] генов, входящих в состав треугольников, показал, что в треугольниках перепредставлены гены, участвующие в производстве энергии (р-значение 0), транспорте и метаболизме сахаров (р-значение $7,3 \times 10^{-7}$). Для оценки статистической значимости использован гипергеометрический тест. Этот результат объясняется тем, что большинство треугольников образовано глобальными транскрипционными факторами дыхания Fnr и ArcA или катаболизма сахаров CRP [47].

4.3 Эволюция регуляторного блока типа треугольник в близкородственных организмах на примере штаммов *Escherichia coli*

Транскрипционное взаимодействие может исчезнуть по разным причинам – может исчезнуть транскрипционный фактор, регулируемый ген или участок связывания транскрипционного фактора. В отличие от транскрипционного фактора или регулируемого гена, участок связывания не эволюционирует по принципу «все или ничего», мутации в его последовательности могут приводить как к его полному исчезновению, так и к небольшим изменениям в силе связывания с транскрипционным фактором. Анализ близкородственных организмов на уровне штаммов позволяет изучать влияние изменения участка связывания транскрипционного фактора на эволюцию транскрипционной сети, так как на этом уровне родства большинство транскрипционных факторов и регулируемых генов сохраняется.

Эволюция мотива треугольник была изучена в близкородственных организмах на примере 25 штаммов *E. coli* (Таблица 8). Для анализа мы отобрали такие межгенные участки, которые сохраняются по крайней мере в 10 штаммах. Мы считали, что межгенный участок сохраняется, если, во-первых, сохраняются оба

гена, между которыми он расположен, во-вторых, ориентация этих генов остается неизменной. Кроме того, здесь мы полагали, что участок связывания исчезает, если в нем возникает хотя бы одна замена. Для очень близких организмов со степенью схожести последовательности более 95% (как, например, штаммы *E. coli*), событие замены происходит очень редко, поэтому необходимо использовать такие жесткие критерии, чтобы таким образом фиксировать редкие события изменения последовательности участка связывания.

Таблица 8. Список исследованных геномов штаммов *E. coli*

Название организма	Сокращение
<i>Escherichia coli</i> K12 MG1655	eco
<i>Escherichia coli</i> 536	esp
<i>Escherichia coli</i> 55989	eck
<i>Escherichia coli</i> APEC O1	ecv
<i>Escherichia coli</i> ATCC 8739	ecl
<i>Escherichia coli</i> B str. REL606	ebr
<i>Escherichia coli</i> BL21(DE3)	ebd
<i>Escherichia coli</i> BW2952	ebw
<i>Escherichia coli</i> CFT073	ecc
<i>Escherichia coli</i> E24377A	ecw
<i>Escherichia coli</i> ED1a	ecq
<i>Escherichia coli</i> HS	ecx
<i>Escherichia coli</i> IAI1	ecr
<i>Escherichia coli</i> IAI39	ect

<i>Escherichia coli</i> O127:H6 str. E2348/69	ecg
<i>Escherichia coli</i> O157:H7 EDL933	ece
<i>Escherichia coli</i> O157:H7 str. EC4115	ecf
<i>Escherichia coli</i> O157:H7 str. Sakai	ecs
<i>Escherichia coli</i> O157:H7 str. TW14359	etw
<i>Escherichia coli</i> S88	ecz
<i>Escherichia coli</i> SE11	ecy
<i>Escherichia coli</i> SMS-3-5	ecm
<i>Escherichia coli</i> UMN026	eum
<i>Escherichia coli</i> UTI89	eci
<i>Escherichia coli</i> str. K-12 substr. DH10B	ecd

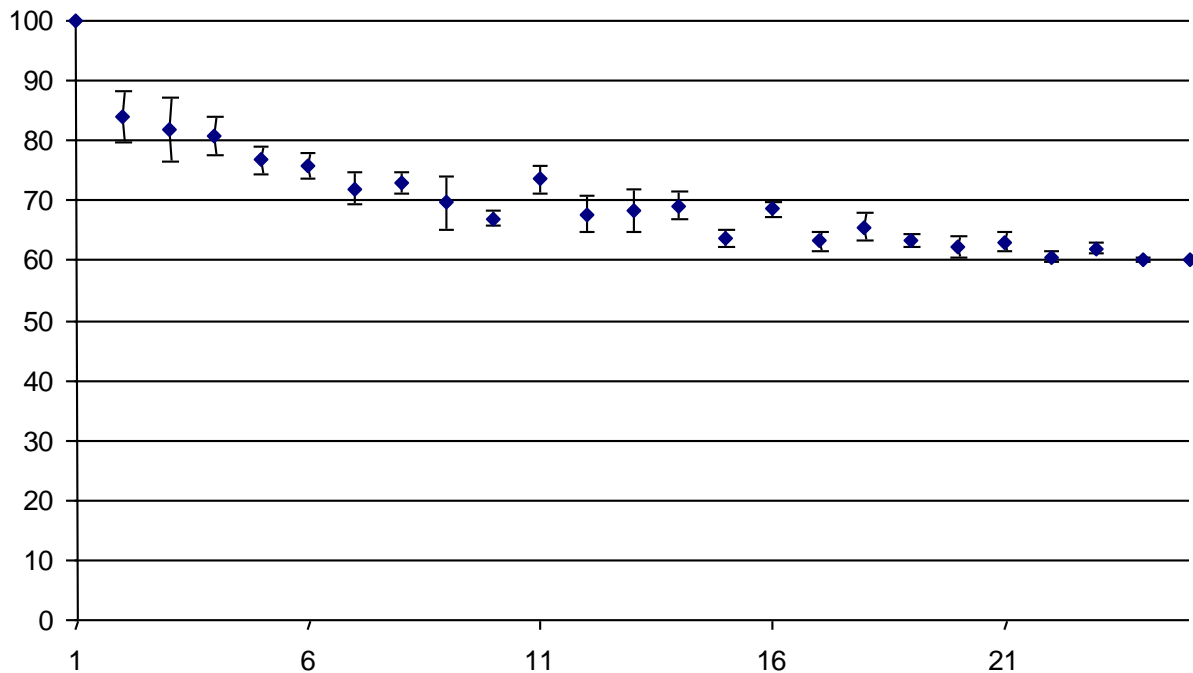
Информация о регуляторных взаимодействиях была получена из базы данных RegulonDB [1]. В 25 штаммах *E. coli* обнаружено 335 регуляторных взаимодействий, принимающих участие в образовании треугольников, и 367 прочих взаимодействий, которые мы назвали *парными*. В 335 взаимодействиях из треугольников у 194 в вершине находится глобальный транскрипционный фактор, а у 161 – локальный. В парных взаимодействиях в вершинах 105 взаимодействий находится глобальный транскрипционный фактор, а в вершинах 262 – локальный (Таблица 9).

Анализ консервативности участков связывания транскрипционных факторов показал, что из 199 глобальных взаимодействий сохранилось 165: 63 парных и 102 из треугольников. Из 421 локального взаимодействия сохранилось 261 взаимодействие: 152 парных и 109 из треугольников. Оказалось, что локальные взаимодействия более консервативны в треугольниках, чем в парных взаимодействиях (статистическая значимость 0.047 согласно тесту χ^2) (Таблица 9).

Таблица 9. Количество различных типов регуляторных взаимодействий в 25 штаммах *E.coli*.

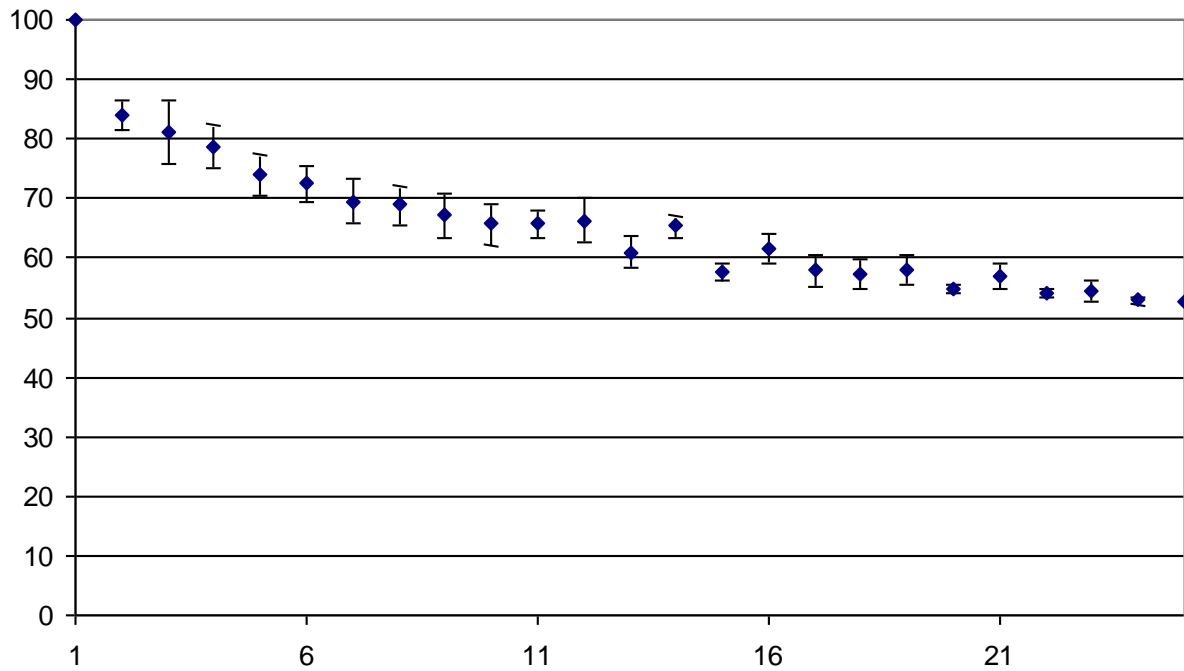
	Треугольники (всего)	Треугольники (консервативные)	Парные (всего)	Парные (консервативные)
Глобальные	194	102	105	63
Локальные	161	109	262	152
Всего	335	211	367	215

Анализ был сделан на выборке из 25 штаммах *E. coli*. Так как их число постоянно растет, возникает вопрос, насколько может измениться полученный результат, если анализировать большее число организмов. Мы проанализировали, как меняется доля консервативных регуляторных взаимодействий в зависимости от числа штаммов и обнаружили, что она практически не меняется, начиная с 15 ± 2 штаммов, в зависимости от типа взаимодействия (Рисунки 19-22).



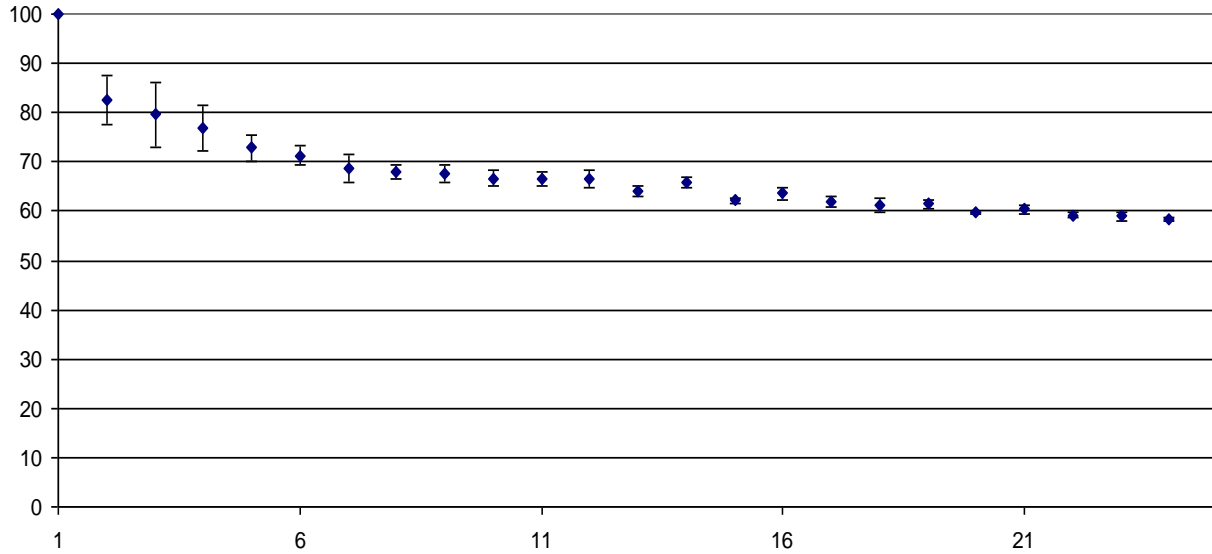
По оси x – количество штаммов, по оси y – доля консервативных участков связывания транскрипционных факторов (в %).

Рисунок 19 – Изменение доли консервативных участков связывания глобальных транскрипционных факторов в зависимости от количества анализируемых штаммов (для парных взаимодействий).



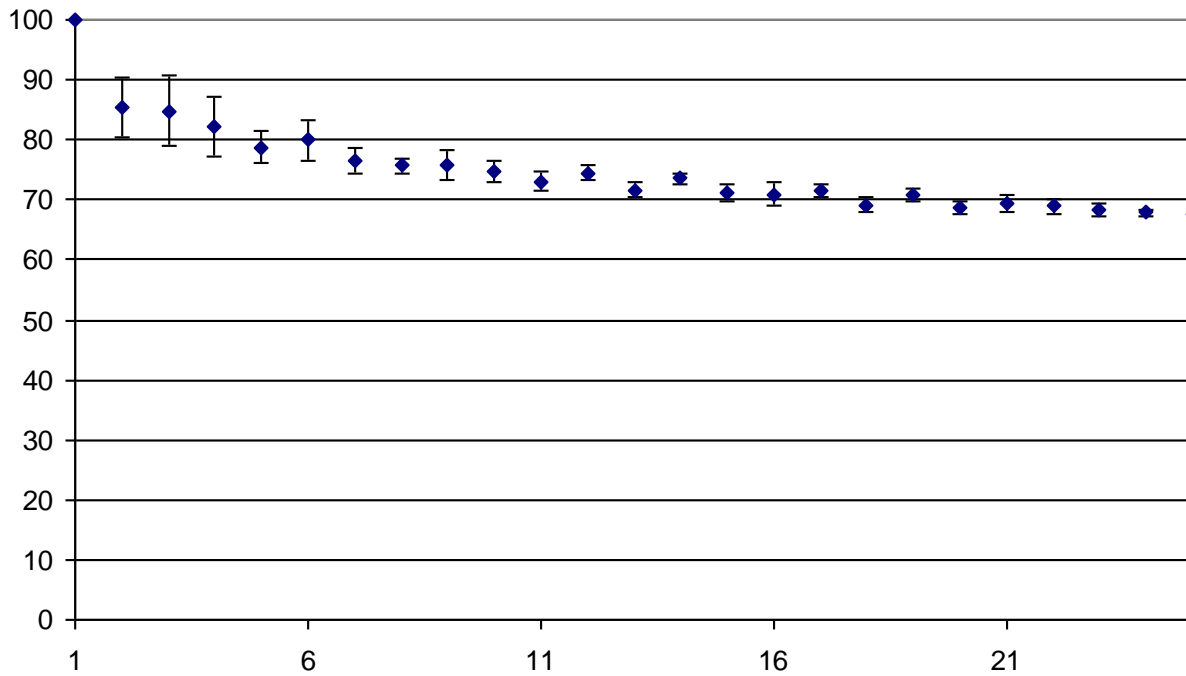
Оси как на рисунке 19.

Рисунок 20 – Изменение доли консервативных участков связывания глобальных транскрипционных факторов в зависимости от количества анализируемых штаммов (для треугольников).



Оси как на рисунке 19.

Рисунок 21 – Изменение доли консервативных участков связывания локальных транскрипционных факторов в зависимости от количества анализируемых штаммов (для парных взаимодействий).



Оси как на рисунке 19.

Рисунок 22 – Изменение доли консервативных участков связывания локальных транскрипционных факторов в зависимости от количества анализируемых штаммов (для треугольников).

Таким образом, добавление новых штаммов не должно существенным образом влиять на полученные оценки.

4.4 Эволюция регуляторного блока типа треугольник в *Enterobacteriales*

В *Enterobacteriales* мы проанализировали частоту всех возможных механизмов изменения регуляторного взаимодействия:

- сохранение всех трех элементов – транскрипционного фактора, гена и участка связывания (обозначено в таблице как «Консервативная регуляция»)
- исчезновение транскрипционного фактора («Нет транскрипционного фактора»)
- исчезновение регуляторного гена («Нет регулируемого гена»)
- исчезновение участка связывания: хотя бы одного или всех в случае множественного регуляции («Нет участка связывания»); исчезновение участка связывания определялось как сильное уменьшение его веса, определяемого с помощью матрицы позиционных весов.

Информация по регуляторным взаимодействиям, как и при анализе штаммов *E. coli*, была использована из базы данных RegulonDB [1]. В базе данных RegulonDB для 175 транскрипционных факторов известен по крайней мере один регулируемый ген.

Как было сказано выше, мутации в последовательности участка связывания приводят к изменению силы связывания с транскрипционным фактором, а не к его полному исчезновению. Сила связывания с транскрипционным фактором может быть примерно оценена при помощи матрицы позиционных весов – чем выше вес участка связывания, тем выше может быть его сила связывания с транскрипционным фактором. Чтобы принять во внимание изменения в силе связывания (т.е. веса)

участка связывания, мы анализировали только такие транскрипционные факторы, для которых матрица позиционных весов есть в базах данных RegPrecise [153] или ее можно создать на основе последовательностей участков связывания из RegulonDB [1]. Таким образом, для анализа было использовано 96 транскрипционных факторов (Таблица 10).

Участки связывания искали перед ортологами известных регулируемых генов *E. coli*. Веса участков связывания, рассчитанные по позиционной матрице весов, находятся в диапазоне от 3,02 (участок связывания глобального транскрипционного фактора CRP перед опероном *glpFKX*) до 9,73 (участок связывания локального транскрипционного фактора CaiF перед опероном *caiTABCDE*), с медианой около 4,6. Вес одного участка связывания перед одним геном (опероном) внутри порядка Enterobacteriales в среднем изменяется на 0,9. Участок связывания считался консервативным в данном организме, если его вес уменьшился не более чем на две единицы по сравнению с весом аналогичного участка в *Escherichia coli* K12.

Таблица 10. Список изученных транскрипционных факторов

Транскрипционный фактор	Таксон, по которому создана матрица	Источник матрицы позиционных весов
AcrR	Enterobacteriales	RegPrecise
Ada	Enterobacteriales	RegPrecise
AgaR	<i>Escherichia coli</i>	В данной работе
AllR	Enterobacteriales	RegPrecise
AllS	Enterobacteriales	RegPrecise
ArcA	Enterobacteriales	RegPrecise
ArgR	Enterobacteriales	RegPrecise
AscG	Enterobacteriales	RegPrecise
AsnC	<i>Escherichia coli</i>	В данной работе

BetI	<i>Shewanella</i> spp	RegPrecise
BirA	Proteobacteria	RegPrecise
CaiF	<i>Escherichia coli</i>	В данной работе
ChbR	Enterobacteriales	В данной работе
CpxR	<i>Escherichia coli</i>	В данной работе
CreB	Enterobacteriales	В данной работе
CRP	Enterobacteriales	RegPrecise
CsgD	<i>Escherichia coli</i>	В данной работе
CueR	<i>Escherichia coli</i>	В данной работе
CusR	Enterobacteriales	В данной работе
DcuR	Enterobacteriales	RegPrecise
DeoR	Enterobacteriales	RegPrecise
DgsA	<i>Escherichia coli</i>	В данной работе
DnaA	<i>Escherichia coli</i>	В данной работе
EbgR	Enterobacteriales	RegPrecise
EvgA	<i>Escherichia coli</i>	В данной работе
ExuR	<i>Escherichia coli</i>	В данной работе
FabR	Enterobacteriales	В данной работе
FadR	Enterobacteriales	RegPrecise
FNR	Enterobacteriales	RegPrecise
FruR	<i>Escherichia coli</i>	В данной работе
FUR	Enterobacteriales	RegPrecise
GadE	Enterobacteriales	RegPrecise
GalR	<i>Escherichia coli</i>	В данной работе
GalS	<i>Escherichia coli</i>	В данной работе
GlpR	Enterobacteriales	RegPrecise

GlrR	<i>Escherichia coli</i>	В данной работе
GntR	<i>Escherichia coli</i>	В данной работе
HipB	<i>Escherichia coli</i>	В данной работе
IclR	Enterobacteriales	RegPrecise
IdnR	<i>Escherichia coli</i>	В данной работе
IlyY	Gamma proteobacteria	RegPrecise
IscR	Gamma proteobacteria	RegPrecise
KdgR	Enterobacteriales	RegPrecise
KdpE	Enterobacteriales	В данной работе
LacI	Enterobacteriales	RegPrecise
LexA	Enterobacteriales	RegPrecise
LldR	<i>Escherichia coli</i>	RegPrecise
Mall	Enterobacteriales	RegPrecise
MalT	<i>Escherichia coli</i>	В данной работе
MarA	Enterobacteriales	RegPrecise
MarR	Enterobacteriales	RegPrecise
MelR	<i>Escherichia coli</i>	В данной работе
MetR	Enterobacteriales	RegPrecise
MntR	Enterobacteriales	RegPrecise
ModE	<i>Escherichia coli</i>	В данной работе
MurR	Enterobacteriales	В данной работе
NadR	Enterobacteriales	RegPrecise
NagC	Gamma proteobacteria	RegPrecise
NanR	Gamma proteobacteria	RegPrecise
NarP	Enterobacteriales	RegPrecise
NemR	Enterobacteriales	В данной работе

NhaR	Gammaproteobacteria	RegPrecise
NikR	Enterobacteriales	RegPrecise
NorR	Gammaproteobacteria	RegPrecise
NrdR	<i>Escherichia coli</i>	В данной работе
NsrR	Enterobacteriales	RegPrecise
NtrC	Gammaproteobacteria	RegPrecise
OmpR	<i>Escherichia coli</i>	В данной работе
PdhR	Enterobacteriales	RegPrecise
PhoB	<i>Escherichia coli</i>	В данной работе
PhoP	<i>Escherichia coli</i>	В данной работе
PrpR	Enterobacteriales	В данной работе
PspF	<i>Escherichia coli</i>	В данной работе
PurR	<i>Escherichia coli</i>	В данной работе
PutA	<i>Escherichia coli</i>	В данной работе
RbsR	Enterobacteriales	RegPrecise
RcnR	<i>Escherichia coli</i>	В данной работе
RcsAB	<i>Escherichia coli</i>	В данной работе
RhaR	Enterobacteriales	В данной работе
RhaS	Enterobacteriales	RegPrecise
RstA	Enterobacteriales	В данной работе
RutR	Gammaproteobacteria	RegPrecise
SlyA	<i>Escherichia coli</i>	В данной работе
SoxR	<i>Shewanella</i> spp	RegPrecise
TdcA	Enterobacteriales	В данной работе
TreR	Enterobacteriales	RegPrecise
TrpR	<i>Escherichia coli</i>	В данной работе

TyrR	Enterobacteriales	RegPrecise
UlaR	<i>Escherichia coli</i>	В данной работе
UxuR	<i>Escherichia coli</i>	В данной работе
XylR	<i>Escherichia coli</i>	В данной работе
YiaJ	Enterobacteriales	RegPrecise
ZntR	<i>Shewanella</i> spp	RegPrecise
Zur	Gamma proteobacteria	RegPrecise

На основе данных из RegulonDB для Enterobacteriales мы получили 473 парных взаимодействия и 418 взаимодействий, образующих треугольники. Разделив их в зависимости от типа транскрипционного фактора, мы получили 6 видов взаимодействий: Г (глобальный транскрипционный фактор)→регулируемый ген, Л (локальный транскрипционный фактор)→регулируемый ген, Г→Г, Г→Л, Л→Л, Л→Г. Оказалось, что взаимодействия типа Л→Г встречаются очень редко, а взаимодействия типа Г→Г характерны только для треугольников и не встречаются в парах. Таким образом, мы изучили оставшиеся четыре типа регуляторных взаимодействий.

В таблицах указано количество определенных событий, произошедших с каждым типом взаимодействия (Таблицы 11-12).

Таблица 11. Количество событий, произошедших с парными взаимодействиями в порядке Enterobacteriales

	Г→Л	Л→Л	Г→ген	Л→ген	Л→ген (в <i>S.enterica</i>)
Нет транскрипционного фактора	0	73	0	240	45

Нет регулируемого гена	31	8	607	674	26
Нет участка связывания (хотя бы одного)	7	62	708	1201	14
Нет участка связывания (всех)	1	25	596	746	11
Консервативная регуляция	51	321	840	2522	250
Всего	89	452	2155	4637	335

Таблица 12. Количество событий, произошедших с треугольниками в порядке Enterobacteriales

	Г→Л	Л→Л	Г→ген	Л→ген	Л→ген (в <i>S.enterica</i>)
Нет транскрипционного фактора	0	189	0	208	44
Нет регулируемого гена	281	82	572	193	15
Нет участка связывания (хотя бы одного)	186	117	828	395	17
Нет участка связывания (всех)	105	44	536	186	8
Консервативная регуляция	290	483	1310	596	110
Всего	757	871	2710	1371	186

Для всех проанализированных организмов взаимодействие с локальным транскрипционным фактором оказалось более консервативным в парных

взаимодействиях. Статистическая значимость теста χ^2 для *Salmonella enterica* равна 0.006. Для других организмов результат не значим, но наблюдаемое отклонение сохраняется. Консервативность регуляторных взаимодействий, образованных глобальными транскрипционными факторами, не различалась между треугольниками и парами. Таким образом, для подобных регуляторных взаимодействий наличие структурного мотива не влияет на их консервативность.

Далее мы изучили поведение самых частых треугольников согласованного (С1) и несогласованного (Н1) типа. Согласованный треугольник оказался менее консервативным, чем несогласованный: участки связывания в таком треугольнике исчезают быстрее (Таблицы 13-14). Статистическая значимость теста χ^2 равна 0,021 для глобальных взаимодействий, и $6,7 \times 10^{-7}$ для локальных.

Таблица 13 Количество событий, произошедших с согласованными треугольниками типа С1

	Г→Л	Л→Л	Г→ген	Л→ген
Нет транскрипционного фактора	0	72	0	138
Нет регулируемого гена	157	15	131	32
Нет участка связывания (хотя бы одного)	50	40	282	184
Нет участка связывания (всех)	10	14	178	77
Консервативная регуляция	87	107	483	147
Всего	294	234	896	501

Таблица 14. Количество событий, произошедших с несогласованными треугольниками типа H1

	Г→Л	Л→Л	Г→ген	Л→ген
Нет транскрипционного фактора	0	193	0	100
Нет регулируемого гена	135	50	245	92
Нет участка связывания (хотя бы одного)	154	58	367	131
Нет участка связывания (всех)	57	30	166	26
Консервативная регуляция	271	258	712	276
Всего	560	559	1324	599

На уровне порядка транскрипционная глобальная регуляция не зависит от позиции в транскрипционной сети (находится ли взаимодействие в составе регуляторного блока или нет). Кроме того, глобальные транскрипционные факторы сохраняются во всех организмах. Напротив, элементы локальной транскрипционной регуляции оказываются зависимыми от места в транскрипционной сети. Вне регуляторных блоков в среднем более половины взаимодействий остаются неизменными (Таблица 9), в то время как в треугольниках сохраняется половина или меньше (Таблица 10).

Так как существует восемь разных типов треугольников, каждый тип может вносить свой вклад в консервативность/неконсервативность блока в целом. Анализ показал, что в неконсервативность локальных взаимодействий внутри треугольника существенный вклад вносят взаимодействия внутри согласованного треугольника C1.

Обсуждение

Мы рассмотрели эволюцию регуляторных блоков в транскрипционной сети на уровне штаммов *E. coli* и на уровне порядка Enterobacteriales на примере треугольников. Отдельно был рассмотрен вопрос о предсказании элементов транскрипционной сети на примере пути утилизации этаноламина.

Задача предсказания участков связывания транскрипционных факторов *de novo*, то есть в отсутствие экспериментально полученных сведений, является интересной и трудной. Путь утилизации этаноламина является примером того, когда известен транскрипционный фактор, но не известны ни его участок связывания, ни полный набор регулируемых им генов. Ранее экспериментально было показано, что в *S. typhimurium* транскрипционный фактор пути утилизации этаноламина EutR индуцирует экспрессию генов пути только в присутствии двух эффекторов – кобаламина и этаноламина [29]. В двух таксонах Enterobacteriales и Betaproteobacteria перед опероном, содержащим ген *eutR*, мы обнаружили консервативную последовательность. Похожая последовательность присутствует также перед генами синтеза кобаламина в Enterobacteriales. Таким образом, в этом таксоне EutR регулирует и синтез основного фермента пути, и синтез и транспорт кофактора кобаламина.

При изучении эволюции транскрипционной сети в целом первоначальная гипотеза, что регуляторные блоки, в силу их биологической значимости, будут относительно высоко консервативны, оказалась не вполне точной. Ранее эволюция регуляторных блоков рассматривалась только с точки зрения сохранения транскрипционного фактора и регулируемого гена [49, 51]. Изменения в самой пластичной части этой структуры – участках связывания – не принимались во внимание, в то время как часто достаточно одной мутации в последовательности ДНК, чтобы уничтожить или, наоборот, создать участок связывания [166]. Таким

образом, именно мутации являются основным инструментом эволюции регуляторных взаимодействий.

Мы показали, что консервативность регуляторного взаимодействия зависит от его положения в транскрипционной сети. Часть сети, находящаяся под локальной регуляцией, оказалась самой подвижной, что согласуется с необходимостью быстрой адаптации организмов к изменениям в окружающей среде. В штаммах *E. coli* регуляторные взаимодействия локального транскрипционного фактора оказались более консервативными в треугольниках, в то время как на уровне порядка Enterobacteriales более консервативными оказались взаимодействия в составе пар. Возможное объяснение этого противоречия может состоять в том, что транскрипционная сеть Enterobacteriales проанализирована неполностью. Во-первых, она до сих пор остается неполной – нам известны не все транскрипционные факторы и, тем более, не все регулируемые ими гены. Таким образом, в наше поле зрения не попадает часть событий в транскрипционной сети, которая может быть значительной и, вероятно, вносить существенный вклад в консервативность/неконсервативность взаимодействий. Во-вторых, анализ на уровне порядка был изначально ограничен такими транскрипционными факторами, для которых возможно создание матрицы позиционных весов, в то время как на уровне штаммов такого ограничения не было. Кроме того, на уровне порядка участки связывания некоторых транскрипционных факторов и транскрипционная сеть в целом более изменчивы, чем на уровне штаммов, где изменчивость невелика, и транскрипционная сеть здесь более устойчива – доля консервативных взаимодействий остается неизменной после определенного количества анализируемых штаммов и не изменяется с их добавлением. Неконсервативность треугольников может быть результатом этой изменчивости.

Для эволюции регуляторного блока также важна функциональность – если блок перестает быть организму полезным, то он легче теряется в эволюции.

Локальные регуляторные взаимодействия в согласованных треугольниках типа С1 менее консервативны по сравнению с несогласованными типа Н1. Наше объяснение состоит в том, что в случае, когда экспрессия гена регулируется двумя транскрипционными факторами в одном направлении, один из них может стать избыточным, и такая регуляторная связь быстро исчезает. Напротив, в несогласованном треугольнике два транскрипционных фактора регулируют экспрессию гена в разных направлениях, и исчезновение одного из регуляторных взаимодействий повлечет за собой общее изменение режима экспрессии.

Рост числа бактериальных организмов с полностью определенной геномной последовательностью, а также экспериментальных данных по транскрипционной регуляции приведет к более полной реконструкции транскрипционной сети, и, значит, увеличит точность подобного анализа. Исследование на других группах бактерий сможет продемонстрировать, являются ли обнаруженные нами закономерности универсальными или таксон-специфичными.

Выводы

1. Анализ таксономического распределения генов утилизации этаноламина показал, что существует два возможных пути катаболизма использования этаноламина, связанных с типом оперона. Первый, короткий, позволяет использовать этаноламин только в качестве источника азота. Второй, длинный, дает возможность использовать его и как источник азота, и как источник углерода.
2. Исследование эволюции генов утилизации этаноламина показало, что короткий оперон является предковым, из которого путем добавления новых генов образовался длинный тип. В ходе эволюции генов утилизации этаноламина произошло минимум три события горизонтального переноса.
3. В *Enterobacteriales* и *Burkholderiales* предсказан мотив участка связывания транскрипционного фактора EutR. В *Enterobacteriales* участки связывания EutR обнаружены в промоторной области семи оперонов, среди которых есть непосредственно гены утилизации этаноламина, а также гены синтеза кофактора основного фермента пути этаноламинлиазы – кобаламина. Это наблюдение устанавливает связь между путями утилизации этаноламина и синтезом кобаламина за счет EutR-зависимой регуляции. В *Burkholderiales* участки связывания EutR обнаружены только непосредственно в промоторной области генов утилизации этаноламина.
4. Локальная регуляция является эволюционно подвижной, что согласуется с необходимостью быстрой адаптации к условиям окружающей среды. В штаммах *Escherichia coli* взаимодействия типа Л→ген (локальный транскрипционный фактор →ген) чаще сохраняются в регуляторных блоках, чем вне их. На уровне порядка *Enterobacteriales* эти взаимодействия консервативнее в парных взаимодействиях, по сравнению с регуляторными блоками, что может быть результатом изменчивости транскрипционной сети.

5. Разные типы треугольников на уровне порядка Enterobacteriales эволюционируют по-разному. Регуляция как глобальными, так и локальными транскрипционными факторами в несогласованном треугольнике оказывается консервативнее, чем в согласованном.

Список публикаций по теме диссертации

Статьи в научных журналах

1. Tsoy O., Ravcheev D., Mushegian A. «Comparative genomics of ethanolamine utilization». // *Journal of Bacteriology*. - 2009 – V. 191(23) – P. 7157-7164
2. Tsoy O.V., Pyatnitskiy M.A., Kazanov M.D., Gelfand M.S. «Evolution of transcriptional regulation in closely related bacteria». // *BMC Evol Biol.* – 2012 – V. 12(1) – P. 200

Тезисы конференций

1. Tsoy O., Ravcheyev D., Mushegian A. «Comparative genomics of the ethanolamine utilization pathway». // 4-th Moscow Conference on Computational Molecular Biology'09. Book of abstracts. – 2008 – P.352
2. Tsoy O., Mushegian A. «Ethanolamine utilization: comparative genomics of spoiled food». // *Regulation and Evolution of Cellular Systems (RECESS, 2010)*, Germany, June 20-23.
3. Цой О., Остерман И. Эволюция регуляторных взаимодействий в бактериях. // *Информационные технологии и системы (ИТиС'10)*. Сборник тезисов. – 2010 – с.343-344.
4. Tsoy O. «The evolution of transcriptional regulation in bacteria». // *Regulation and Evolution of Cellular Systems (RECESS, 2010)*, Germany, May 10-13.
5. Tsoy O., Gelfand M. «Evolution of transcriptional regulation in Enterobacteriales». // *ASBMB Symposia. Chemical, Synthetic and Systems Biology: New Directions of Biochemistry in the 21st Century*, Utah, USA, October 12-16.

Список литературы

1. RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more/H. Salgado, M. Peralta-Gil, S. Gama-Castro, et al.// *Nucleic Acids Res.* – 2013 – 41 (Database issue) - D203-213.
2. DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information/N. Sierro, Y. Makita, M. de Hoon, K. Nakai//*Nucleic Acids Res.* - 2008 - 36(Database issue) - D93-96.
3. GenBank/D.A. Benson, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, D.L. Wheeler//*Nucleic Acids Res.* - 2005 - 33(Database issue) - D34-38.
4. Gelfand M.S. Recognition of regulatory sites by genomic comparison/M.S. Gelfand//*Res Microbiol.* - 1999 - V.150, №9-10 - PP.755-771.
5. Rodionov D.A. Comparative genomic reconstruction of transcriptional regulatory networks in bacteria/D.A. Rodionov//*Chem Rev.* - 2007 - V.107, №8 - PP.3467-3497
6. Integrative inference of gene-regulatory networks in *Escherichia coli* using information theoretic concepts and sequence analysis/C. Kaleta, A. Göhler, S. Schuster, et al.//*BMC Syst Biol.* - 2010 - V.4 - P.116.
7. Babu M.M., Lang B., Aravind L. Methods to reconstruct and compare transcriptional regulatory networks/M.M. Babu, B. Lang, L. Aravind//*Methods Mol Biol.* - 2009 - V.541 - PP.163-180.
8. Genome-scale reconstruction of the Lrp regulatory network in *Escherichia coli*/B.K. Cho, C.L. Barrett, E.M. Knight, et al.//*Proc Natl Acad Sci U S A.* - 2008 - V.105, №49 - PP.19462-19467.
9. Wu W.S., Li W.H., Chen B.S. Computational reconstruction of transcriptional regulatory modules of the yeast cell cycle/W.S. Wu, W.H. Li, B.S. Chen//*BMC Bioinformatics.* - 2006 - V.7 - P.421.

10. Li H., Wang W. Dissecting the transcription networks of a cell using computational genomics/H. Li, W. Wang//*Curr Opin Genet Dev.* - 2003 - V.13, №6 - PP.611-616.
11. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation/F.P. Roth, J.D. Hughes, P.W. Estep, G.M. Church//*Nat Biotechnol.* - 1998 - V.16, №10 - PP.939-945.
12. Comparative genomic analysis of the hexuronate metabolism genes and their regulation in gammaproteobacteria/I.A. Suvorova, M.N. Tutukina, D.A. Ravcheev, et al./*J Bacteriol.* - 2011 - V.193, №15 - PP.3956-3963.
13. Reconstruction of xylose utilization pathway and regulons in Firmicutes/Y. Gu, Y. Ding, C. Ren, et al.//*BMC Genomics.* - 2010 - V.11 - P.255.
14. Transcriptional regulation of NAD metabolism in bacteria: genomic reconstruction of NiaR (YrxA) regulon/D.A. Rodionov, X. Li, I.A. Rodionova, C. Yang, et al.//*Nucleic Acids Res.* - 2008 - V.36, №6 - PP.2032-2046.
15. Computational reconstruction of iron- and manganese-responsive transcriptional networks in alpha-proteobacteria/D.A. Rodionov, M.S. Gelfand, J.D. Todd, et al.//*PLoS Comput Biol.* - 2006 - V.2, №12 - e163.
16. Duret L., Bucher P. Searching for regulatory elements in human noncoding sequences/L. Duret, P. Bucher//*Curr Opin Struct Biol.* - 1997 - V.7, №3 - PP.399-406.
17. Galas D.J., Eggert M., Waterman M.S. Rigorous pattern-recognition methods for DNA sequences. Analysis of promoter sequences from *Escherichia coli*/D.J. Galas, M. Eggert, M.S. Waterman//*J Mol Biol.* - 1985 - V.186, №1 - PP.117-128.
18. Staden R. Computer methods to locate signals in nucleic acid sequences/R. Staden//*Nucleic Acids Res.* - 1984 - V.12, №1-2 - PP.505-519.
19. Изучение связывания ДНК факторами транскрипции LacI семейства методами машинного обучения/Г.Г. Федонин, А.Б. Рахманинова, Ю.Д. Коростелев, и др.//*Молекулярная биология* - 2011 - Т.45 №4 - СС. 724-737.

20. Жаров И.А., Гельфанд М.С., Казаков А.Е.//Регуляция генов множественной лекарственной устойчивости транскрипционными факторами подсемейства BhlR/И.А. Жаров, М.С. Гельфанд, А.Е. Казаков//Молекулярная биология - 2011 - Т.45,№4 - СС. 715-723.
21. Camas F.M., Alm E.J., Poyatos J.F. Local gene regulation details a recognition code within the LacI transcriptional factor family/F.M. Camas, E.J. Alm, J.F. Poyatos//PLoS Comput Biol. - 2-10 - V.6,№11 - e1000989.
22. Inference of the transcriptional regulatory network in *Staphylococcus aureus* by integration of experimental and genomics-based evidence/D.A. Ravcheev, A.A. Best, N. Tintle, et al.//J Bacteriol. - 2011 - V.193,№13 - PP.3228-3240.
23. Comparative genomic reconstruction of transcriptional networks controlling central metabolism in the *Shewanella* genus/D.A. Rodionov, P.S. Novichkov, E.D. Stavrovskaya, et al.//BMC Genomics. - 2011 - V.12 Suppl 1 - S3.
24. Reconstruction of regulatory and metabolic pathways in metal-reducing delta-proteobacteria/D.A. Rodionov, I. Dubchak, A. Arkin, et al.//Genome Biol. - 2004 - V.5,№11 - R90.
25. Computer analysis of transcription regulatory patterns in completely sequenced bacterial genomes/A.A. Mironov, E.V. Koonin, M.A. Roytberg, M.S. Gelfand//Nucleic Acids Res. - 1999 - V.27,№14 - PP.2981-2989.
26. Comparative analysis of regulatory patterns in bacterial genomes/M.S. Gelfand, P.S. Novichkov, E.S. Novichkova, A.A. Mironov/Brief Bioinform. - 2000 - V.1,№4 - PP.357-371.
27. Embryonic epsilon and gamma globin genes of a prosimian primate (*Galago crassicaudatus*). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints/D.A. Tagle, B.F. Koop, M. Goodman, et al.//J Mol Biol. - 1988 - V.203,№2 - 439-455.

28. Roof D.M., Roth J.R. Ethanolamine utilization in *Salmonella typhimurium*/D.M. Roof, J.R. Roth//J Bacteriol. - 1988 - V.170,№9 - PP.3855-3863.
29. Roof D.M., Roth J.R. Functions required for vitamin B12-dependent ethanolamine utilization in *Salmonella typhimurium*/D.M. Roof, J.R. Roth//J Bacteriol. - 1989 - V.171,№6 - 3316-3323.
30. The 17-gene ethanolamine (eut) operon of *Salmonella typhimurium* encodes five homologues of carboxysome shell proteins/E. Kofoed, C. Rappleye, I. Stojiljkovic, J. Roth//J Bacteriol. - 1999 - V.181,№17 - PP.5317-5329.
31. Evidence that a B12-adenosyl transferase is encoded within the ethanolamine operon of *Salmonella enterica*/D.E. Sheppard, J.T. Penrod, T. Bobik, et al.//J Bacteriol. - 2004 - V.186,№22 - PP.7635-7644.
32. Brinsmade S.R., Paldon T., Escalante-Semerena J.C. Minimal functions and physiological conditions required for growth of *salmonella enterica* on ethanolamine in the absence of the metabolosome/S.R. Brinsmade, T. Paldon, J.C. Escalante-Semerena//J Bacteriol. - 2005 - V.187,№23 – PP.8039-8046.
33. Biogenesis and Ultrastructure of Carboxysomes from Wild Type and Mutants of *Synechococcus* sp. Strain PCC 7942/M.I. Orus, M.L. Rodriguez, F. Martinez, E. Marco//Plant Physiol. - 1995 - V.107,№4 - PP.1159-1166.
34. Penrod J.T., Roth J.R. Conserving a volatile metabolite: a role for carboxysome-like organelles in *Salmonella enterica*/J.T. Penrod, J.R. Roth//J Bacteriol. - 2006 - V.188,№8 – PP.2865-2874.
35. Roof D.M., Roth J.R. Autogenous regulation of ethanolamine utilization by a transcriptional activator of the eut operon in *Salmonella typhimurium*/D.M. Roof, J.R. Roth//J Bacteriol. - 1992 - V.174,№20 – PP.6634-6643.
36. Systematic association of genes to phenotypes by genome and literature mining/J.O. Korbel, T. Doerks, L.J. Jensen, et al.//PLoS Biol. - 2005 - V.3,№5 - e134.

37. Ailion M., Bobik T.A., Roth J.R. Two global regulatory systems (Crp and Arc) control the cobalamin/propanediol regulon of *Salmonella typhimurium*/M. Ailion, T.A. Bobik, J.R. Roth//*J Bacteriol.* - 1993 - V.175, №22 – PP.7200-7208.
38. Bobik T.A., Ailion M., Roth J.R. A single regulatory gene integrates control of vitamin B12 synthesis and propanediol degradation/T.A. Bobik, M. Ailion, J.R. Roth//*J Bacteriol.* - 1992 - V.174, №7 – PP.2253-2266.
39. Rondon M.R., Escalante-Semerena J.C. In vitro analysis of the interactions between the PoxR regulatory protein and the promoter region of the cobalamin biosynthetic (cob) operon of *Salmonella typhimurium* LT2/M.R. Rondon, J.C. Escalante-Semerena//*J Bacteriol.* - 1996 - V.178, №8 – PP.2196-2203.
40. Network motifs in the transcriptional regulation network of *Escherichia coli*/S.S. Shen-Orr, R. Milo, S. Mangan, U. Alon//*Nat Genet.* - 2002 - V.31, №1 - PP.64-68.
41. Network motifs: simple building blocks of complex networks/R. Milo, S. Shen-Orr, S. Itzkovitz, et al.//*Science.* - 2002 - V.298, №5594 - PP.824-827.
42. Mangan S., Alon U. Structure and function of the feed-forward loop network motif/S. Mangan, U. Alon//*Proc Natl Acad Sci U S A.* - 2003 - V.100, №21 - 11980-11985.
43. Gottesman S. Bacterial regulation: global regulatory networks/S. Gottesman//*Annu Rev Genet.* - 1984 - V.18 - PP.415-441.
44. Martínez-Antonio A., Collado-Vides J. Identifying global regulators in transcriptional regulatory networks in bacteria/A. Martínez-Antonio, J. Collado-Vides//*Curr Opin Microbiol.* - 2003 - V.6, №5 - PP.482-489.
45. Madan Babu M., Teichmann S.A. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*/M. Madan Babu, S.A. Teichmann//*Nucleic Acids Res.* - 2003 - V.31, №4 - PP.1234-1244.

46. Ma H.W., Buer J., Zeng A.P. Hierarchical structure and modules in the Escherichia coli transcriptional regulatory network revealed by a new top-down approach/H.W. Ma, J. Buer, A.P. Zeng//BMC Bioinformatics. - 2004 - V.5 - P.199.
47. Gelfand M.S. Evolution of transcriptional regulatory networks in microbial genomes/M.S. Gelfand//Curr Opin Struct Biol. - 2006 - V.16,№3 – PP.420-429.
48. Mangan S., Zaslaver A., Alon U. The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks/S. Mangan, A. Zaslaver, U. Alon//J Mol Biol. - 2003- V.334,№2 - PP.197-204.
49. Madan Babu M., Teichmann S.A., Aravind L. Evolutionary dynamics of prokaryotic transcriptional regulatory networks/M. Madan Babu, S.A. Teichmann, L. Aravind//J Mol Biol. - 2006 - V.358,№2 - PP.614-633.
50. Impact of Transcription Units rearrangement on the evolution of the regulatory network of gamma-proteobacteria/A.D. González Pérez, E. González González, V. Espinosa Angarica, et al.//BMC Genomics. - 2008 - V.9 - P.128.
51. Hershberg R., Margalit H. Co-evolution of transcription factors and their targets depends on mode of regulation/R. Hershberg, H. Margalit//Genome Biol. - 2006 - V.7,№7 - R62.
52. KEGG for integration and interpretation of large-scale molecular data sets/M. Kanehisa, S. Goto, Y. Sato, et al.//Nucleic Acids Res. - 2012 - 40(Database issue) - D109-D114.
53. The genome of the natural genetic engineer Agrobacterium tumefaciens C58/D.W. Wood, J.C. Setubal, R. Kaul, et al.//Science. - 2001 - V.294,№5550 - PP.2317-2323.
54. Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110/T. Kaneko, Y. Nakamura, S. Sato, et al.//DNA Res. - 2002 - V9,№6 - PP.189-197.

55. Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia/E. Giraud, L. Moulin, D. Vallenet, et al.//*Science*. - 2007 - V.316,№5829 - PP.1307-1312.
56. The *Brucella suis* genome reveals fundamental similarities between animal and plant pathogens and symbionts/I.T. Paulsen, R. Seshadri, K.E. Nelson, et al.//*Proc Natl Acad Sci U S A*. - 2002 - V.99,№20 - PP.13148-13153.
57. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*/T. Kaneko, Y. Nakamura, S. Sato, et al.//*DNA Res*. - 2000 - V.7,№6 - 331-338.
58. Complete genome sequence of *Nitrobacter hamburgensis* X14 and comparative genomic analysis of species within the genus *Nitrobacter*/S.R. Starkenburg, F.W. Larimer, L.Y. Stein, et al.//*Appl Environ Microbiol*. - V.74,№9 - PP.2852-2863.
59. Genome sequence of the chemolithoautotrophic nitrite-oxidizing bacterium *Nitrobacter winogradskyi* Nb-255/Starkenburg S.R., Chain P.S., Sayavedra-Soto L.A., et al.//*Appl Environ Microbiol*. - 2006 - V.72,№3 - PP.2050-2063.
60. Genome of *Ochrobactrum anthropi* ATCC 49188 T, a versatile opportunistic pathogen and symbiont of several eukaryotic hosts/P.S. Chain, D.M. Lang, D.J. Comerci, et al.//*J Bacteriol*. - 2011 - V.193,№16 - PP.4274-4275.
61. The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons/V. González, R.I. Santamaría, P. Bustos, et al.//*Proc Natl Acad Sci U S A*. - 2006 - V.103,№10 - 3834-3839.
62. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components/Young J.P., Crossman L.C., Johnston A.W., et al.//*Genome Biol*. - 2006 - V.7,№4 - R34.
63. Multiple genome sequences reveal adaptations of a phototrophic bacterium to sediment microenvironments/Y. Oda, F.W. Larimer, P.S. Chain, et al.//*Proc Natl Acad Sci U S A*. - 2008 - V.105,№47 - 18543-18548.

64. Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodospseudomonas palustris*/F.W. Larimer, P. Chain, L. Hauser, et al.//*Nat Biotechnol.* - 2004 - V.22,№1 - PP.55-61.
65. The composite genome of the legume symbiont *Sinorhizobium meliloti*/F. Galibert, T.M. Finan, S.R. Long, et al.//*Science.* - 2001 - V.293,№5530 - PP.668-672.
66. The genome of the versatile nitrogen fixer *Azorhizobium caulinodans* ORS571/K.B. Lee, P. De Backer, T. Aono, et al.//*BMC Genomics.* - 2008 - V.9 - P.271.
67. Ecological genomics of marine Roseobacters/M.A. Moran, R. Belas, M.A. Schell, et al.//*Appl Environ Microbiol.* - 2007 - V.73,№14 - PP.4559-4569.
68. Genome sequence analysis of the emerging human pathogenic acetic acid bacterium *Granulibacter bethesdensis*/D.E. Greenberg, S.F. Porcella, A.M. Zelazny, et al.//*J Bacteriol.* - 2007 - V.189,№23 - PP.8727-8736.
69. Complete genome sequence of *Rhodospirillum rubrum* type strain (S1)/A.C. Munk, A. Copeland, S. Lucas, et al.//*Stand Genomic Sci.* - 2011 - V.4,№3 - PP.293-302.
70. Complete genome of the mutualistic, N₂-fixing grass endophyte *Azoarcus* sp. strain BH72/A. Krause, A. Ramakumar, D. Bartels, et al.//*Nat Biotechnol.* - 2006 - V.24,№11 - 1385-1391.
71. Genome sequence of the bioplastic-producing "Knallgas" bacterium *Ralstonia eutropha* H16/A. Pohlmann, W.F. Fricke, F. Reinecke, et al.//*Nat Biotechnol.* - 2006 - V.24,№10 - 1257-1262.
72. Genome sequence of the plant pathogen *Ralstonia solanacearum*/M. Salanoubat, S. Genin, F. Artiguenave, et al.//*Nature.* - 2002 - V.415,№6871 - PP.497-502.
73. *Burkholderia ambifaria* sp. nov., a novel member of the *Burkholderia cepacia* complex including biocontrol and cystic fibrosis-related isolates/T. Coenye, E. Mahenthiralingam, D. Henry, et al.//*Int J Syst Evol Microbiol.* - 2001 - V.51(Pt 4) - PP.1481-1490.

74. Structural flexibility in the *Burkholderia mallei* genome/W.C. Nierman, D. DeShazer, H.S. Kim, et al.//Proc Natl Acad Sci U S A. - 2004 - V.101,№39 - PP.14246-14251.
75. Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*/M.T. Holden, R.W. Titball, S.J. Peacock, et al.//Proc Natl Acad Sci U S A. - 2004 - V.101,№39 - 14240-14245.
76. Bacterial genome adaptation to niches: divergence of the potential virulence genes in three *Burkholderia* species of different survival strategies/H.S. Kim, M.A. Schell, Y. Yu, et al.//BMC Genomics. - 2005 - V.6 - P.174.
77. *Burkholderia xenovorans* LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility/P.S. Chain, V.J. Denef, K.T. Konstantinidis, et al.//Proc Natl Acad Sci U S A. - 2006 - V.103,№42 - 15280-15287.
78. Brazilian National Genome Project Consortium. The complete genome sequence of *Chromobacterium violaceum* reveals remarkable and exploitable bacterial adaptability/Brazilian National Genome Project Consortium//Proc Natl Acad Sci U S A. - 2003 - V.100,№20 - PP.11660-11665.
79. Whole-genome analysis of the methyl tert-butyl ether-degrading beta-proteobacterium *Methylobium petroleiphilum* PM1/S.R. Kane, A.Y. Chakicherla, P.S. Chain//J Bacteriol. - 2007 - V.189,№5 - 1931-45.
80. The genome of *Polaromonas naphthalenivorans* strain CJ2, isolated from coal tar-contaminated sediment, reveals physiological and metabolic versatility and evolution through extensive horizontal gene transfer/J.M. Yagi, D. Sims, T. Brettin, D. Bruce, E.L. Madsen//Environ Microbiol. - 2009 - V.11,№9 - 2253-2270.
81. The genome of *Aeromonas salmonicida* subsp. *salmonicida* A449: insights into the evolution of a fish pathogen/M.E. Reith, R.K. Singh, B. Curtis, et al.//BMC Genomics. - 2008 - V.19 - P.427.

82. The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*/E. Duchaud, C. Rusniok, L. Frangeul, et al.//*Nat Biotechnol.* - 2003 - V.21,№11 - 1307-1313.
83. Genome dynamics and diversity of *Shigella* species, the etiologic agents of bacillary dysentery/F. Yang, J. Yang, X. Zhang, et al.//*Nucleic Acids Res.* - 2005 - V.33,№19 - PP.6445-6458.
84. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host/H. Toh, B.L. Weiss, S.A. Perkin, et al.//*Genome Res.* - 2006 - V.16,№2 - PP,149-156.
85. The complete genome sequence of *Escherichia coli* K-12/F.R. Blattner, G. 3rd Plunkett, C.A. Bloch, et al.//*Science.* - 1997 - V.277,№5331 - PP.1453-1462.
86. Role of pathogenicity island-associated integrases in the genome plasticity of uropathogenic *Escherichia coli* strain 536/B. Hochhut, C. Wilde, G. Balling, et al.//*Mol Microbiol.* - 2006 - V.61,№3 - PP.584-595.
87. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths/M. Touchon, C. Hoede, O. Tenaillon, et al.//*PLoS Genet.* - 2009 - V.5,№1 - e1000344.
88. The genome sequence of avian pathogenic *Escherichia coli* strain O1:K1:H7 shares strong similarities with human extraintestinal pathogenic *E. coli* genomes/Johnson T.J., Kariyawasam S., Wannemuehler Y., et al.//*J Bacteriol.* - 2007 - V.189,№8, PP.3228-3236.
89. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). H. Jeong, V. Barbe, C.H. Lee, et al.//*J Mol Biol.* - 2009 - V.394,№4 - PP.644-652.
90. Genomic sequencing reveals regulatory mutations and recombinational events in the widely used MC4100 lineage of *Escherichia coli* K-12./T. Ferenci, Z. Zhou, T. Betteridge, et al.//*J Bacteriol.* - 2009 - V.191,№12 - PP.4025-4029.

91. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*/R.A. Welch, V. Burland, G. 3rd Plunkett, et al.//Proc Natl Acad Sci U S A. - 2002 - V.99,№26 - 17020-17024.
92. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates/D.A. Rasko, M.J. Rosovitz, G.S. Myers, et al.//J Bacteriol. - 2008 - V.190,№20 - 6881-6893.
93. Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69/A. Iguchi, N.R. Thomson, Y. Ogura, et al.//J Bacteriol. - 2009 - V.191,№1 - PP.347-354.
94. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7/N.T. Perna, G. 3rd Plunkett, V. Burland, et al.//Nature. - 2001 - V.409,№6819 - PP.529-533.
95. Genomic anatomy of *Escherichia coli* O157:H7 outbreaks/M. Eppinger, M.K. Mammel, J.E. Leclerc, J. Ravel, T.A. Cebula//Proc Natl Acad Sci U S A. - 2011 - V.108,№50 - PP.20142-20147.
96. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12/T. Hayashi, K. Makino, M. Ohnishi, et al.//DNA Res. - 2001 - V.8,№11 - PP.11-22.
97. Analysis of the genome of the *Escherichia coli* O157:H7 2006 spinach-associated outbreak isolate indicates candidate genes that may enhance virulence/B.R. Kulasekara, M. Jacobs, Y. Zhou, et al.//Infect Immun. - 2009 - V.77,№9 - PP.3713-3721.
98. Complete genome sequence and comparative analysis of the wild-type commensal *Escherichia coli* strain SE11 isolated from a healthy adult/K. Oshima, H. Toh, Y. Ogura, et al.//DNA Res. - 2008 - V.15,№6 - PP.375-386.
99. Insights into the environmental resistance gene pool from the genome sequence of the multidrug-resistant environmental isolate *Escherichia coli* SMS-3-5/W.F. Fricke, M.S. Wright, A.H. Lindell, et al.//J Bacteriol. - 2008 - V.190,№20 - PP.6779-6794.

100. Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach/S.L. Chen, C.S. Hung, J. Xu, et al.//Proc Natl Acad Sci U S A. - 2006 - V.103,№15 - 5977-5982.
101. The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse/T. Durfee, R. Nelson, S. Baldwin, et al.//J Bacteriol. - 2008 - V.190,№7 - PP.2597-2606.
102. The *Citrobacter rodentium* genome sequence reveals convergent evolution with human pathogenic *Escherichia coli*/N.K. Petty, R. Bulgin, V.F. Crepin, et al.//J Bacteriol. - 2010 - V.192,№2 - PP.525-538.
103. Complete genome sequence of *Enterobacter cloacae* subsp. *cloacae* type strain ATCC 13047/Y. Ren, Y. Ren, Z. Zhou, et al.//J Bacteriol. - 2010 - V.192,№9 - PP.2463-2464.
104. Complete genome sequence of *Cronobacter turicensis* LMG 23827, a food-borne pathogen causing deaths in neonates/R. Stephan, A. Lehner, P. Tischler, T. Rattei//J Bacteriol. - 2011 - V.193,№1 - PP.309-310.
105. Genome sequence of *Cronobacter sakazakii* BAA-894 and comparative genomic hybridization analysis with other *Cronobacter* species/E. Kucerova, S.W. Clifton, X.Q. Xia, et al.//PLoS One. - 2010 - V.5,№3 - e9556.
106. Complete genome sequence of the N₂-fixing broad host range endophyte *Klebsiella pneumoniae* 342 and virulence predictions verified in mice/D.E. Fouts, H.L. Tyler, R.T. DeBoy, et al.//PLoS Genet. - 2008 - V.4,№7 - e1000141.
107. Genome comparison of the epiphytic bacteria *Erwinia billingiae* and *E. tasmaniensis* with the pear pathogen *E. pyrifoliae*/M. Kube, A.M. Migdoll, I. Gehring, et al.//BMC Genomics. - 2010 - V.11 - P.393.
108. Complete genome sequence of the plant pathogen *Erwinia amylovora* strain ATCC 49946/M. Sebahia, A.M. Bocsanczy, B.S. Biehl, et al.//J Bacteriol. - 2010 - V.192,№7 - PP.2020-2021.

109. Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors/K.S. Bell, M. Sebaihia, L. Pritchard, et al.//Proc Natl Acad Sci U S A. - 2004 - V.101,№30 - 11105-11110.
110. The genome of *Erwinia tasmaniensis* strain Et1/99, a non-pathogenic bacterium in the genus *Erwinia*/M. Kube, A.M. Migdoll, I. Müller, et al.//Environ Microbiol. - 2008 - V.10,№9 - PP.2211-2222.
111. Complete genome sequence of the fire blight pathogen *Erwinia pyrifoliae* DSM 12163T and comparative genomic insights into plant pathogenicity/T.H. Smits, S. Jaenicke, F. Rezzonico, et al.//BMC Genomics. - 2010 - V.11 - P.2.
112. Genome sequence of *Pantoea ananatis* LMG20103, the causative agent of Eucalyptus blight and dieback/P. De Maayer, W.Y. Chan, S.N. Venter, et al.//J Bacteriol. - 2010 - V.192,№11 - PP.2936-2937.
113. Genome sequence of the biocontrol agent *Pantoea vagans* strain C9-1/T.H. Smits, F. Rezzonico, T. Kamber, et al.//J Bacteriol. - 2010 - V.192,№24 - PP.6486-6487.
114. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18/J. Parkhill, G. Dougan, K.D. James, et al.//Nature. - 2001 - V.413,№6858 - PP.848-852.
115. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2/M. McClelland, K.E. Sanderson, J. Spieth, et al.//Nature. - 2001 - V.413,№6858 - PP.852-856.
116. Genome dynamics and diversity of *Shigella* species, the etiologic agents of bacillary dysentery/F. Yang, J. Yang, X. Zhang, et al.//Nucleic Acids Res. - 2005 - V.33,№19 - PP.6445-6558.
117. New insights into *Acinetobacter baumannii* pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis/M.G. Smith, T.A. Gianoulis, S. Pukatzki, et al.//Genes Dev. - 2007 - V.21,№5 - 601-614.

118. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities/A.C. da Silva, J.A. Ferro, F.C. Reinach, et al.//*Nature*. - 2002 - V.417,№6887 - PP.459-463.
119. Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*/W. Qian, Y. Jia, S.X. Ren, et al.//*Genome Res*. - 2005 - V.15,№6 - PP.757-767.
120. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen/C.K. Stover, X.Q. Pham, A.L. Erwin, et al.//*Nature*. - 2000 - V.406,№6799 - PP.959-964.
121. Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5/I.T. Paulsen, C.M. Press, J. Ravel, et al.//*Nat Biotechnol*. - 2005 - V.23,№7 - PP.873-878.
122. Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*/M.W.Silby, A.M. Cerdeño-Tárraga, G.S. Vernikos, et al.//*Genome Biol*. - 2009 - V.10,№5 - R51.
123. Nitrogen fixation island and rhizosphere competence traits in the genome of root-associated *Pseudomonas stutzeri* A1501/Yan Y., Yang J., Dou Y., et al.//*Proc Natl Acad Sci U S A*. - 2008 - V.105,№21 - PP.7564-7569.
124. Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440/K.E. Nelson, C. Weinel, I.T. Paulsen, et al.//*Environ Microbiol*. - 2002 - V.4,№12 - PP.799-808.
125. The complete genome sequence of the Arabidopsis and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000/C.R. Buell, V. Joardar, M. Lindeberg, et al.//*Proc Natl Acad Sci U S A*. - 2003 - V.100,№18 – PP.10181-10186.
126. Evolution of sensory complexity recorded in a myxobacterial genome/B.S. Goldman, W.C. Nierman, D. Kaiser, et al.//*Proc Natl Acad Sci U S A*. - 2003 - V.103,№41 – PP.15200-15205.

127. Whole-genome sequence of *Listeria welshimeri* reveals common steps in genome reduction with *Listeria innocua* as compared to *Listeria monocytogenes*/T. Hain, C. Steinweg, C.T. Kuenne, et al.//*J Bacteriol.* - 2006 - V.188,№21 – PP.7405-7415.
128. Comparative genomics of *Listeria* species/P. Glaser, L. Frangeul, C. Buchrieser, et al.//*Science.* - 2001 - V.294,№5543 – PP.849-852.
129. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome/M. Sebaihia, B.W. Wren, P. Mullany, et al.//*Nat Genet.* - 2006 - V.38,№7 – PP.779-786.
130. Skewed genomic variability in strains of the toxigenic bacterial pathogen, *Clostridium perfringens*/G.S. Myers, D.A. Rasko, J.K. Cheung, et al.//*Genome Res.* - 2006 - V.16,№6 - PP.1031-1040.
131. The genome sequence of *Clostridium tetani*, the causative agent of tetanus disease./H. Bruggemann, S. Baumer, W.F. Fricke, et al.//*Proc Natl Acad Sci U S A.* - 2003 - V.100,№3 – PP.1316-1321.
132. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*/J. Nölling, G. Breton, M.V. Omelchenko, et al.//*J Bacteriol.* - 2001 - V.183,№16 – PP.4823-4838.
133. Complete genome sequence of the dehalorespiring bacterium *Desulfitobacterium hafniense* Y51 and comparison with *Dehalococcoides ethenogenes* 195/H. Nonaka, G. Keresztes, Y. Shinoda, et al.//*J Bacteriol.* - 2006 - V.188,№6 – PP.2262-2274.
134. Genome sequence of *Symbiobacterium thermophilum*, an uncultivable bacterium that depends on microbial commensalism/K. Ueda, A. Yamashita, J. Ishikawa, et al.//*Nucleic Acids Res.* - 2004 - V.32,№16 - PP.4937-4944.

135. Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*/I.T. Paulsen, L. Banerjee, G.S. Myers, et al.//*Science*. - 2003 - V.299,№5615 – PP.2071-2074.
136. Genome of the opportunistic pathogen *Streptococcus sanguinis*/P. Xu, J.M. Alves, T. Kitten, et al.//*J Bacteriol*. - 2007 - V.189,№8 – PP.3166-3175.
137. Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography/P. Normand, P. Lapierre, L.S. Tisa, et al.//*Genome Res*. - 2007 - V.17,№1 - PP.7-15.
138. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*/H. Ikeda, J. Ishikawa, A. Hanamoto, et al.//*Nat Biotechnol*. - 2003 - V.21,№5 - PP.526-531.
139. Reductive evolution and niche adaptation inferred from the genome of *Mycobacterium ulcerans*, the causative agent of Buruli ulcer/T.P. Stinear, T. Seemann, S. Pidot, et al.//*Genome Res*. - 2007 - V.17,№2 - PP.192-200.
140. ICDS database: interrupted CoDing sequences in prokaryotic genomes/E. Perrodou, C. Deshayes, J. Muller, et al.//*Nucleic Acids Res*. - 2006 - 34(Database issue) - D338-D343.
141. The complete genomic sequence of *Nocardia farcinica* IFM 10152/J. Ishikawa, A. Yamashita, Y. Mikami, et al.//*Proc Natl Acad Sci U S A*. - 2004 - V.101,№41 – PP.14925-31490.
142. The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse/M.P. McLeod, R.L. Warren, W.W. Hsiao, et al.//*Proc Natl Acad Sci U S A*. - 2006 - V.103,№42 – PP.15582-15587.
143. Genome Sequence of the ethene- and vinyl chloride-oxidizing actinomycete *Nocardioides* sp. strain JS614/N.V. Coleman, N.L. Wilson, K. Barry, et al.//*J Bacteriol*. - 2011 - V.193,№13 – PP.3399-3400.

144. Genome sequence of the cellulolytic gliding bacterium *Cytophaga hutchinsonii*/G. Xie, D.C. Bruce, J.F. Challacombe, et al.//*Appl Environ Microbiol.* - 2007 - V.73,№11 - PP.3536-3546.
145. Genome sequence and analysis of the oral bacterium *Fusobacterium nucleatum* strain ATCC 25586/V. Kapatral, I. Anderson, N. Ivanova, et al.//*J Bacteriol.* - 2002 - V.184,№7 - PP.2005-2018.
146. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs/S.F. Altschul, T.L. Madden, A.A. Schäffer, et al.//*Nucleic Acids Res.* - 1997 - V.25,№17 - PP.3389-3402.
147. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools/J.D. Thompson, T.J. Gibson, F. Plewniak, et al.//*Nucleic Acids Res.* - 1997 - V.25,№24 - PP.4876-4882.
148. Edgar R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput/R.C. Edgar//*Nucleic Acids Res.* - 2004 - V.32,№5 - 1792-1797.
149. Schuler G.D., Altschul S.F., Lipman D.J. A workbench for multiple alignment construction and analysis/G.D. Schuler, S.F. Altschul, D.J. Lipman//*Proteins.* - 1991 - V.9,№3 - PP.180-190.
150. Felsenstein J. PHYLIP - Phylogeny Inference Package (Version 3.2)/J. Felsenstein//*Cladistics* - 1989 - V5 - PP.164-166.
151. Letunic I., Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation/I. Letunic, P. Bork//*Bioinformatics.* - 2007 - V.23,№1 - PP.127-128.
152. Миронов А.А., Винокурова Н.П., Гельфанд М.С. Программное обеспечение анализа бактериальных геномов/А.А. Миронов, Н.П. Винокурова, М.С. Гельфанд//*Молекулярная биология* - 2000 - т.34,№2 - СС. 253-262.

153. RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes/P.S. Novichkov, O.N. Laikova, E.S. Novichkova, et al.//Nucleic Acids Res. - 2010 - 38(Database issue) - D111-D118.
154. RegTransBase--a database of regulatory sequences and interactions in a wide range of prokaryotic genomes/A.E. Kazakov, M.J. Cipriano, P.S. Novichkov, et al.//Nucleic Acids Res. - 2007 - 35(Database issue) - D407-D412.
155. The COG database: an updated version includes eukaryotes/R.L. Tatusov, N.D. Fedorova, J.D. Jackson, et al.//BMC Bioinformatics. - 2003 - V.4 - P.41.
156. LeuO antagonizes H-NS and StpA-dependent repression in *Salmonella enterica* ompS1/M.A. De la Cruz, M. Fernández-Mora, C. Guadarrama, et al.//Mol Microbiol. - 2007 - V.66,№3 - PP.727-743.
157. Toraya T. Radical catalysis of B12 enzymes: structure, mechanism, inactivation, and reactivation of diol and glycerol dehydratases/T. Toraya//Cell Mol Life Sci. - 2000 - V.57,№1 - PP.106-127.
158. Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes/D.A. Rodionov, A.G. Vitreschak, A.A. Mironov, M.S. Gelfand//J Biol Chem. - 2003 - V.278,№42 - PP.41148-41159.
159. Vassinova N., Kozyrev D. A method for direct cloning of fur-regulated genes: identification of seven new fur-regulated loci in *Escherichia coli*/N. Vassinova, D. Kozyrev//Microbiology. - 2000 - 146 Pt 12 - PP.3171-3182.
160. Stojiljkovic I., Bäumlér A.J., Hantke K. Fur regulon in gram-negative bacteria. Identification and characterization of new iron-regulated *Escherichia coli* genes by a fur titration assay/I. Stojiljkovic, A.J. Bäumlér, K. Hantke//J Mol Biol. - 1994 - V.236,№2 - PP.531-545.
161. *Escherichia coli* NsrR regulates a pathway for the oxidation of 3-nitrotyramine to 4-hydroxy-3-nitrophenylacetate/L.D. Rankin, D.M. Bodenmiller, J.D. Partridge, et al.//J Bacteriol. - 2008 - V.190,№18 - PP.6170-6177.

162. Bodenmiller D.M., Spiro S. The yjeB (nsrR) gene of Escherichia coli encodes a nitric oxide-sensitive transcriptional regulator/D.M. Bodenmiller, S. Spiro//J Bacteriol. - 2006 - V.188,№3 - PP.874-881.
163. Identification of additional genes belonging to the LexA regulon in Escherichia coli/A.R. Fernández De Henestrosa, T. Ogi, S. Aoyagi, et al.//Mol Microbiol. - 2000 - V.35,№6 - PP.1560-1572.
164. Activation of the Cpx regulon destabilizes the F plasmid transfer activator, TraJ, via the HslVU protease in Escherichia coli/I.C. Lau-Wong, T. Locke, M.J. Ellison, et al.//Mol Microbiol. - 2008 - V/67,№3 - PP.516-527.
165. Uden G., Bongaerts J. Alternative respiratory pathways of Escherichia coli: energetics and transcriptional regulation in response to electron acceptors/G. Uden, J. Bongaerts//Biochim Biophys Acta. - 1997 - V.1320,№3 - PP.217-234.
166. Berg J., Willmann S., Lässig M. Adaptive evolution of transcription factor binding sites/J. Berg, S. Willmann, M. Lässig//BMC Evol Biol. - 2004 - V.4 - P.42.