

**Российская академия наук**  
**Федеральное государственное бюджетное учреждение науки**  
**Институт проблем передачи информации им. А.А. Харкевича**

На правах рукописи

**Лейн Семен Александрович**

**Эволюция транскрипционной регуляции метаболизма углеводов  
в бактериях**

Диссертация на соискание ученой степени  
кандидата биологических наук

Специальность 03.01.09 - математическая биология, биоинформатика

Научный руководитель:  
к.б.н. Родионов Дмитрий Александрович

Москва – 2014

## Оглавление

Оглавление.....	2
Введение.....	5
Актуальность темы.....	5
Цели и задачи исследования .....	7
Научная новизна и практическое значение работы .....	8
Основные положения, выносимые на защиту.....	9
Глава 1. Обзор литературы.....	11
1.1 Регуляция инициации транскрипции у бактерий .....	11
1.1.1 Структура РНК-полимеразы и механизм взаимодействия с промотором .....	12
1.1.2 Факторы транскрипции.....	14
1.2 Методы сравнительной геномики для реконструкции регуляторных сетей у бактерий .....	19
1.2.1 Методы сравнительной геномики для анализа сайтов связывания факторов транскрипции .....	21
1.2.2 Методы предсказания функции гена на основании сходства аминокислотных последовательностей.....	24
1.2.3 Методы анализа геномного контекста.....	25
1.3 AraR-регулон утилизации арабинозы.....	26
1.4 Регуляция утилизации N-ацетилгалактозамина в протеобактериях .	29
1.5 HexR – регулятор центрального метаболизма углерода .....	31
1.6 Механизмы регуляции путей утилизации сахаров в <i>B. subtilis</i> .....	34
1.6.1 Регуляция метаболизма углеводов с помощью фосфотрансферазных транспортных систем.....	35
1.6.2 Регуляция метаболизма углеводов субстрат-связывающими факторами транскрипции .....	38
1.6.3. СсрА-зависимая катаболитная репрессия оперонов метаболизма сахаров.....	38
1.6.4. СсрА-независимая катаболитная репрессия оперонов метаболизма сахаров.....	41
Глава 2. Материалы и методы .....	47
2.1 Принципы применения методов сравнительной геномики к анализу регуляции транскрипции .....	47
2.2 Геномы .....	49
2.3 Программное обеспечение .....	53
Глава 3. Исследование эволюции AraR регулона .....	55
3.1 Исследование эволюции регуляторной системы AraR.....	55
3.2 Построение распознающего правила для поиска потенциальных сайтов связывания AraR .....	56
3.3 Структура AraR регулона в изучаемых геномах.....	61
3.4 Эволюция AraR регулона .....	65
3.5 Обсуждение.....	68
Глава 4. Исследование регуляции AraR регулона в протеобактериях.....	70
4.1 Исследование эволюции регуляторной системы AraR .....	70

4.2 Построение распознающего правила для поиска потенциальных сайтов связывания AgaR .....	70
4.3 Структура AgaR регулона и путей утилизации N-ацетилгалактозамина и галактозамина .....	72
4.4 Эволюция AgaR регулона .....	78
4.5 Обсуждение .....	80
<b>Глава 5. HexR – регулятор центрального метаболизма углеводов.....</b>	<b>83</b>
5.1 Исследование эволюции регуляторной системы HexR .....	83
5.2 Построение распознающего правила для поиска потенциальных сайтов связывания HexR .....	85
5.3 Ядро HexR регулона.....	86
5.4. Таксон-специфическая регуляция генов HexR регулона.....	87
5.5 Обсуждение .....	90
<b>Глава 6. Эволюция регуляции катаболизма сахаров в бактериях семейства <i>Bacillaceae</i> .....</b>	<b>92</b>
6.1 Поиск потенциальных регуляторов метаболизма углеводов среди ортологов факторов транскрипции <i>B. subtilis</i> в бактериях семейства <i>Bacillaceae</i> .....	92
6.3 Реконструкция регулонов утилизации сахаров и их производных в семействе бактерий <i>Bacillaceae</i> .....	101
6.4 Обсуждение .....	105
<b>Выводы .....</b>	<b>109</b>
<b>Список публикаций по теме диссертации .....</b>	<b>111</b>
<b>Благодарности .....</b>	<b>114</b>
<b>Список литературы .....</b>	<b>115</b>
<b>Приложения .....</b>	<b>141</b>

## Список используемых сокращений и обозначений

пн – пары нуклеотидов

НТН – спираль-поворот-спираль

PTS – фосфотрансферазная система

PRD – домен регуляции фосфотрансферазной системой

НАГА – N-ацетилгалактозамин

ГА – галактозамин

КоА – кофермент А

цАМФ – циклический аденозинмонофосфат

АТФ – аденозинтрифосфат

НАД – никотинамидадениндинуклеотид

НАДФ – никотинамидадениндинуклеотидфосфат

## **Введение**

### **Актуальность темы**

Углеводы и их производные являются основными источниками углерода и энергии для большинства гетеротрофных бактерий, а также важным строительным материалом в живой природе. Они образуют различные по составу и структуре моно-, олиго- и полисахариды. В ходе эволюции бактерии выработали множество путей утилизации углеводсодержащих соединений, подстраиваясь под это многообразие субстратов. Большинство бактерий в геноме обнаруживают до нескольких десятков подобных путей. Для эффективной работы такой системы необходима слаженная регуляция активности нужных генов в зависимости от доступности тех или иных субстратов. Ключевым элементом здесь выступают факторы транскрипции – специальные белки-регуляторы, которые активируют или репрессируют инициацию транскрипции определенных генов. Большой интерес представляет эволюция систем утилизации различных сахаров, что помимо регуляции включает в себя возникновение альтернативных биохимических путей, неортологичное замещение генов, образование функционально гетерогенных семейств паралога. Причем, этот интерес исходит как от фундаментальной науки, так и от промышленной биотехнологии, перед которой стоит задача эффективной переработки органических отходов, большинство из которых является растительной биомассой и, следовательно, содержит большое разнообразие углеводсодержащих соединений.

До недавнего времени основным средством получения новых знаний о регуляции являлись кропотливые эксперименты с модельными микроорганизмами. Однако, несмотря на достаточно большую надежность данных методов, они не дают представления о полной картине регуляторных взаимодействий, сосредоточившись лишь на небольшом фрагменте системы. Значительно улучшили ситуацию методы высокопроизводительных экспериментов. И хотя в них прослеживается изменение экспрессии тысяч

генов, они характеризуются высоким уровнем шума и сложностью разделения прямых регуляторных эффектов от косвенных, например таких как регуляторные каскады или ко-регуляция генов.

В связи с резким увеличением числа полностью секвенированных геномов в последнее время у исследователей появился мощный инструмент для биоинформатического анализа бактериальных геномов – методы сравнительной геномики. К настоящему моменту по данным базы KEGG (1) полностью секвенировано почти 2000 бактериальных геномов, которые широко покрывают разнообразные таксономические группы. При этом удешевление технологий секвенирования позволяет получать геномные данные с нарастающей скоростью. Изучение регуляции транскрипции генов методами биоинформатики ставит своей задачей выявление регуляторных участков ДНК (промоторов, терминаторов, РНК-переключателей, сайтов связывания транскрипционных факторов), регулируемых генов, а также предсказание новых ДНК-связывающих белков для обнаруженных сайтов ДНК. Т.е. основная цель – получить полное описание регулона – совокупности оперонов, находящихся под контролем одного фактора транскрипции.

Быстрое увеличение числа геномных последовательностей делает необходимым решение другой важной проблемы геномики – получение достоверной функциональной аннотации генов. Основным способом аннотации генов до недавнего времени являлся перенос функции экспериментально охарактеризованных белков на другие с помощью поиска сходства последовательностей. Для осуществления поиска по базам данных последовательностей были разработаны специализированные программы, наибольшую популярность среди которых получил алгоритм BLAST и его варианты (2).

Несмотря на большой успех, методы, основанные на поиске сходства в последовательностях, не позволяют аннотировать многие гены и нередко производят неточные или ошибочные аннотации (3). Существенно ситуацию меняют подходы сравнительной геномики, одним из которых является анализ

эволюции структуры регулона в различных геномах (3,4). Данный подход предполагает, что ко-регулируемые гены могут принадлежать одному пути. Это дает существенную помощь в предсказании аннотации.

Целый ряд исследований был посвящен реконструкции регуляции различных метаболических путей в бактериях методами сравнительной геномики, например, регуляция биосинтеза аргинина (5), пуринов (6), метаболизма пентоз (7,8), энергетического метаболизма (9), утилизации ксилозы (10), утилизации N-ацетилглюкозамина (11), метаболизма НАД (12). Результаты данных биоинформатических работ показывают, что сравнительная геномика является эффективным методом для исследования регуляторных взаимодействий в бактериях.

### **Цели и задачи исследования**

Целью данной работы был анализ эволюции регулонов метаболизма углеводов в геномах бактерий с помощью методов сравнительной геномики. В качестве объектов исследования были выбраны регулятор катаболизма арабинозы в *Bacillus subtilis* AraR, репрессор катаболизма N-ацетилгалактозамина в протеобактериях AgaR, регулятор центрального метаболизма углеводов в протеобактериях HexR и 43 регулятора метаболизма сахаров в семействе бактерий *Bacillaceae*.

В работе решаются следующие задачи:

1. Реконструкция регулонов метаболизма сахаров в экспериментально неизученных бактериях методами сравнительной геномики, построение распознающих правил для предсказания сайтов связывания факторов транскрипции, описание состава регулона и оперонной структуры.
2. Предсказание функций новых генов и уточнение существующих аннотаций генов, вовлеченных в катаболизм различных углеводов.

3. Анализ таксон-специфических особенностей регуляции транскрипции генов катаболизма углеводов и построение вероятных сценариев эволюции регуляторных взаимодействий.

### **Научная новизна и практическое значение работы**

В работе впервые полностью описаны локальные регулоны метаболических путей утилизации арабинозы и N-ацетилгалактозамина (AraR и AgaR), а также глобальный регулон HexR, контролирующий центральный метаболизм углерода в гамма- и бета-протеобактериях. Для семейства *Bacillaceae* реконструировано 11 новых регуляторных систем для генов катаболизма арил-бета-глюкозидов, инозитола, рамногалактуронана, фруктозы, глюкозамина, хитина, мальтодекстрина, N-ацетилмурамата и альфа-галактозидов. При этом, для каждого нового белка-регулятора были не только описаны наборы регулируемых генов, но и предсказаны их потенциальные ДНК сайты связывания и молекулы-эффекторы. Для 11 ранее известных факторов транскрипции, контролирующих пути катаболизма углеводов, были впервые обнаружены их мотивы ДНК сайтов.

Метаболическая реконструкция путей утилизации арабинозы позволила обнаружить новые ферменты: L-рибулокиназу AgaB-II и L-арабиноизомеразу AgaA-II, а также большое разнообразие систем транспорта и деградации арабинозосодержащих полисахаридов. Также было показано, что обнаруженная ранее в биоинформатическом анализе *Clostridium acetobutylicum* L-рибулокиназа AgaK является наиболее распространенным вариантом в пути утилизации арабинозы в бактериях типа *Firmicutes*.

Обнаружено множество вариаций в начальных стадиях метаболических путей катаболизма и транспорта N-ацетилгалактозамина у протеобактерий. Анализ геномного контекста *aga* генов позволил предположить специфичность новых PTS транспортных систем к различным аminosахарам. Было впервые показано, что функция галактозамин-6-фосфат диаминазы/изомеразы, ранее приписываемая ферменту AgaI, принадлежит белку AgaS.

Также в работе были предложены потенциальные сценарии эволюции изученных регулонов метаболизма сахаров. В результате анализа филогении регуляторных систем было выдвинуто предположение о сильном влиянии горизонтальных переносов в периферических путях метаболизма сахаров.

Работа имеет преимущественно теоретический характер, однако полученные данные могут применяться и в генной инженерии. Показанные варианты метаболических путей могут в будущем использоваться в качестве строительных блоков для получения высокопродуктивных генно-инженерных штаммов микроорганизмов, например, для производства биотоплива.

### **Основные положения, выносимые на защиту**

- 1) Реконструкция AraR регулонов в четырех порядках бактерий *Bacillales*, *Lactobacillales*, *Clostridiales* и *Thermotogales* показала широкое разнообразие вариантов метаболических путей утилизации арабинозы. Помимо новых транспортеров и гидролитических ферментов были обнаружены новые неортологичные варианты L-рибулокиназы и L-арабинозаизомеразы (AraB-II и AraA-II). Предложена модель эволюции AraR регулона путем таксон-специфического расширения изначального регулона, содержащего гены *araEKDA*.
- 2) В результате реконструкции AgaR регулона в протеобактериях было показано, что наибольшее разнообразие вариантов метаболических путей приходится на их первые стадии, включающие: (1) внеклеточный гидролиз полисахаридов и транспорт внутрь клетки сахарных остатков – продуктов гидролиза; (2) фосфорилирование и деацетилирование составляющих их аминсахаров. Показана функция белка AgaS в качестве основной галактозамин-6-фосфат деаминазы/изомеразы в катаболизме N-ацетилгалактозамина. Предложена вероятная модель эволюции генов, кодирующих начальные стадии пути утилизации N-

ацетилгалактозамина, из генов схожей биохимической функции, участвующих в пути утилизации N-ацетилглюкозамина в бактериях рода *Shewanella*.

- 3) Сравнительный анализ HexR регулонов показал значительные изменения в их размере и составе среди различных групп протеобактерий. Обнаружено, что в бактериях порядков *Vibrionales*, *Aeromonadales*, *Psychromonadales* и *Alteromonadales* регулятор HexR является глобальным регулятором центрального метаболизма углерода.
- 4) Реконструированы 43 регулона, отвечающих за метаболизм различных углеводов в 10 бактериях семейства *Bacillaceae*. Для 22 регуляторов были впервые предсказаны ДНК мотивы сайтов связывания. Из них для 11 факторов транскрипции впервые предсказаны регулируемые гены. Показана высокая степень разнообразия регуляторных систем катаболизма сахаров у исследованных бактерий.

# Глава 1

## Обзор литературы

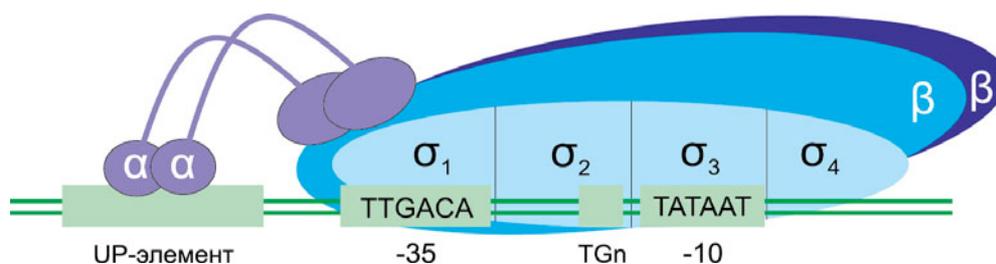
### 1.1 Регуляция инициации транскрипции у бактерий

Бактерии известны своей способностью быстрой адаптации к различным условиям окружающей среды. Основной причиной такой эффективности является экономное использование своего генетического материала для экспрессии необходимых генов в правильное время.

Несмотря на то, что регуляция может происходить на всех стадиях образования функциональных продуктов, ключевым шагом регуляции экспрессии генов является инициация транскрипции. Регуляция на уровне транскрипции позволяет обходиться без затрат большого количества ресурсов клетки на образование ненужных в данный момент макромолекул. Регуляция инициации транскрипции основана на взаимодействии *транс* факторов – РНК-полимеразы и специальных белков, называемых факторами транскрипции – *цис*-регуляторными элементами – участками ДНК в 5'-некодирующей области гена. *Цис*-регуляторные элементы часто также называются сайтами связывания. На бактериальной хромосоме транскрипционная единица образуется из последовательного расположения следующих генетических элементов: регуляторной области (сайтов связывания белков-регуляторов и промотера – сайта посадки РНК-полимеразы), одного или нескольких генов и терминатора транскрипции. Транскрипционная единица, содержащая только один ген называется моноцистронной, в ином случае – полицистронной. Часто один и тот же ген может быть транскрибирован с нескольких промоторов. Таким образом транскрипционные единицы могут перекрываться. Набор всех перекрывающихся транскрипционных единиц называется опероном (13).

### 1.1.1 Структура РНК-полимеразы и механизм взаимодействия с промотором

Главная роль в транскрипции у бактерий принадлежит белковому комплексу ДНК-зависимой РНК-полимеразы (Рис. 1.1). Состав РНК-полимеразы насчитывает пять субъединиц кор-фермента  $\beta\beta'\alpha_2\omega$ , который служит для элонгации транскрипции, а также  $\sigma$ -фактор, отвечающий за распознавание промотора (14). Активный центр, способный наращивать молекулу РНК комплементарно ДНК матрице РНК-полимеразы, сформирован комплексом из  $\beta$  и  $\beta'$  субъединиц (15). Каждая из двух идентичных  $\alpha$  субъединиц состоит из двух доменов, соединенных короткой линкерной областью. N-концевой домен  $\alpha$  субъединиц отвечает за связывание с комплексом  $\beta$  и  $\beta'$  субъединиц. С-концевой домен способен связываться с ДНК, что необходимо для функционирования некоторых промоторов (16), а также может участвовать в белок-белковых взаимодействиях с активаторами и репрессорами (17). Предполагается, что  $\omega$  субъединица, образующая комплекс с  $\beta'$  субъединицей, не участвует напрямую в транскрипции, но выполняет структурную функцию (18).



**Рисунок 1.1.** Модель присоединения РНК полимеразы к промотору.

Для инициации транскрипции РНК-полимераза должна образовать холофермент с  $\sigma$ -фактором, который выполняет функции распознавания элементов промотора, позиционирования кор-фермента РНК-полимеразы на промоторе и расплетания двойной спирали ДНК.  $\sigma$ -факторы содержат до четырех доменов. Домены 2, 3 и 4 участвуют в распознавании элементов

промотора (19,20). Функция домена 1 до сих пор не ясна. К тому же обнаружено, что домен 1 часто отсутствует.

Большинство бактерий содержит несколько альтернативных  $\sigma$ -факторов, отвечающих за регуляцию наборов генов путем активации транскрипции в ответ на различные стимулы. Каждый  $\sigma$ -фактор распознает свой набор промоторных последовательностей. Однако, большинство генов при нормальных условиях регулируется одним основным  $\sigma$ -фактором «домашнего хозяйства». В случае *E. coli* им является  $\sigma^{70}$ , в *B. subtilis* –  $\sigma^A$  (19). Эти  $\sigma$ -факторы узнают несимметричные промоторные последовательности состоящие из четырех элементов. Главными среди них являются два ДНК элемента, характеризующиеся консенсусными последовательностями ТАТААТ и ТТGАСА и расположенные на расстоянии приблизительно -10 и -35 нуклеотидов от старта транскрипции соответственно. Промоторные -10 и -35 элементы взаимодействуют с вторым и четвертым доменами  $\sigma$ -фактора РНК-полимеразы соответственно. Два дополнительных промоторных элемента называются расширенным -10 элементом и UP-элементом. Расширенный -10 элемент размером 3-4 нуклеотида находится непосредственно перед -10 элементом и взаимодействует с доменом 3  $\sigma$ -фактора (20). UP-элемент находится перед -35 сайтом и взаимодействует с С-концевым доменом  $\alpha$ -субъединицы РНК-полимеразы (16). Вместе все четыре элемента промотора отвечают за специфическое связывание РНК-полимеразы, однако, найденные в природе последовательности промоторов всегда отличаются от консенсуса. Предполагается, что промотор, идеально совпадающий с консенсусом, связывал бы РНК-полимеразу слишком сильно и это препятствовало бы инициации транскрипции. Также, слабые промоторы позволяют более точно регулировать экспрессию генов с помощью активации транскрипции *транс* факторами (17).

После связывания холофермента РНК-полимеразы с промотором происходит расплетание примерно 14 нуклеотидов двойной спирали ДНК в районе старта транскрипции (21). Данная структура называется открытым

комплексом. После формирования открытого комплекса  $\sigma$ -фактор отсоединяется от РНК-полимеразы и дальнейшую элонгацию транскрипции осуществляет только кор-фермент (17).

Инициация транскрипции – сложный многостадийный процесс, на который влияет множество факторов, таких как промоторные последовательности ДНК, наличие  $\sigma$ -факторов и факторов транскрипции, структура хромосомы и наличие молекул эффекторов. Таким образом, регуляция транскрипции может эффективно осуществляться с помощью различных механизмов.

### ***1.1.2 Факторы транскрипции***

Факторами транскрипции называются белки, способные специфично связываться с определенными ДНК последовательностями в промоторной области гена в ответ на наличие внутриклеточного или внешнего сигнала, и тем самым регулировать инициацию транскрипции. Эти последовательности называются сайтами связывания.

Количество факторов транскрипции зависит от среды обитания и образа жизни данного организма. Исследования зависимости количества факторов транскрипции от размера генома показало близкую к квадратичной пропорцию (22-24). Более того, ускоренный рост числа регуляторов транскрипции при линейном увеличении числа генов считается одним из основных факторов, сдерживающих рост геномов прокариот. Увеличение количества факторов транскрипции создает необходимость более точного различения сайтов связывания, а, следовательно, проблему различения сигнала от шума. Также сильно возрастает нагрузка на метаболические пути для поддержания нужных концентраций регуляторов в клетке (24,25). Свободноживущие организмы, такие как *E. coli* и *B. subtilis* обладают большими геномами, которые кодируют 5-7% факторов транскрипции от общего числа генов (26,27). Образованная данными факторами транскрипции регуляторная сеть необходима для координации экспрессии специализированных наборов генов, позволяющих

бактериям выживать в меняющихся внешних условиях. В случае стабильных внешних условий, многие регуляторные взаимодействия оказываются не нужны. Примером может служить паразитическая бактерия *Rickettsia prowazekii* (28), геном которой кодирует только восемь факторов транскрипции (1% от числа генов).

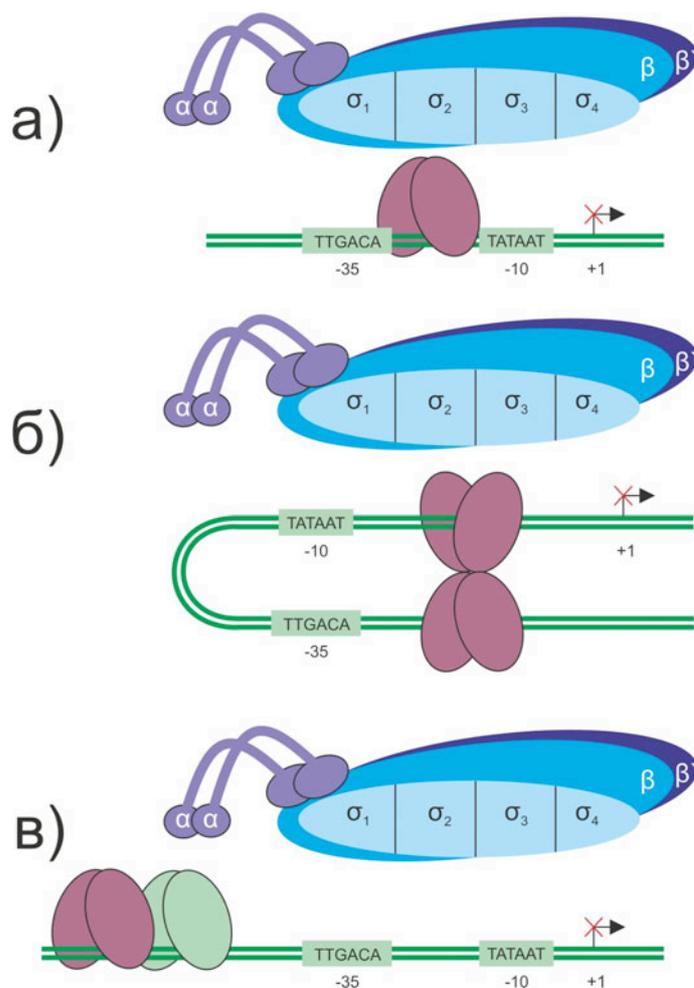
Регуляторная сеть объединяет в себя все регулоны данной бактерии и подчиняется степенному закону распределения. Регуляторные сети в бактериях характеризуются малым числом факторов транскрипции, отвечающих за большое число регуляторных взаимодействий, и малым числом факторов транскрипции, контролирующих несколько генов (29). Так, в *E. coli* всего семь факторов транскрипции (CRP, FNR, IHF, FIS, ArcA, NarL and Lrp) напрямую регулируют экспрессию 51% генов (30). Такая топология сети ставит вопрос о разделении факторов транскрипции на «глобальные» и «локальные». Наиболее проработанная концепция, позволяющая отличить локальные регуляторы от глобальных, дает следующие критерии: 1) число регулируемых генов, 2) частота случаев ко-регуляции гена совместно с другими факторами транскрипции, 3) способность регулировать гены, принадлежащим к различным функциональным категориям, 4) регуляция транскрипционных единиц с промоторами специфичными к различным  $\sigma$ -факторам, 5) способность чувствовать сигнал, отвечающий за большой спектр внешних условий (30). К сожалению, использование данных критериев требует тщательной реконструкции сети транскрипционной регуляции у бактерий, и на данный момент они применимы только к *E. coli*. В *E. coli* этим критериям соответствуют CRP, IHF, FNR, FIS, ArcA, Lrp и Hns (30). На практике применяются упрощенные критерии, включающие один или несколько вышеперечисленных пунктов.

Факторы транскрипции классифицируются на семейства исходя из двух доменов, позволяющих им функционировать в качестве регуляторов. Первый домен, называемый ДНК связывающий, отвечает за непосредственное связывание фактора транскрипции с ДНК. В бактериях большинство таких

доменов имеет структуру спираль-поворот-спираль (helix-turn-helix, НТН-домен) (22,31). Несмотря на то, что были найдены и другие структуры, такие как цинковые пальцы, антипараллельные  $\beta$ -листы и спираль-петля-спираль, они составляют лишь малую фракцию среди известных прокариотических факторов транскрипции (27). Второй домен факторов транскрипции, называемый эффекторным, принимает информацию о состоянии среды и обеспечивает изменение уровня экспрессии генов. Во многих случаях эффекторный домен также служит для олигомеризации молекул фактора транскрипции в виде гомодимеров. В следствие этого сайты связывания регуляторов представляют собой симметричные структуры в виде палиндромов или прямых повторов (32).

Существует четыре механизма регуляции активности факторов транскрипции. Первый – непосредственное взаимодействие фактора с лигандами, которыми могут выступать малые молекулы или физико-химические сигналы, отражающие информацию о состоянии клетки или внешней среды (33). Классическим примером регуляции активности транскрипционного фактора с помощью изменения концентрации вещества является LacI репрессор оперона катаболизма лактозы *lacZYA* в *E. coli*. Присутствие в среде лактозы вызывает повышение в клетке концентрации аллолактозы – эффектора LacI репрессора. При связывании аллолактозы с репрессором происходит аллостерическое изменение конформации белка, что уменьшает его сродство к сайту связывания, а, следовательно, позволяет РНК полимеразе подойти к промотору (34). Вторым механизмом является ковалентная модификация фактора транскрипции. Подобным образом функционируют двухкомпонентные системы, например, система регуляции ответа на анаэробные условия ResD-ResE в *B. subtilis* (35). Они состоят из гистидиновой киназы, которая является сенсором внешних сигналов и обычно локализована в плазматической мембране клетки. При поступлении сигнала гистидиновая киназа фосфорилирует себя и затем передает фосфорную группу на соответствующий регулятор ответа (36). Следующий способ – это секвестрация регулятора с помощью специального белка, часто заякоренного

на мембране. Этот механизм иллюстрируется работой токсин-антитоксин системы SdpR-SdpI-SdpC из *B. subtilis* (37). SdpI – мембранный белок ответа на белковый токсин SdpC. SdpI может связывать SdpC и в этом виде приобретает сродство к SdpR – репрессору оперона *sdpRI*. Таким образом, в присутствии SdpC репрессор SdpR захватывается на мембране и экспрессия оперона *sdpRI* увеличивается. Наконец, существует каскадная регуляция факторов транскрипции. При этой системе фактор транскрипции всегда активен, но его экспрессия, а следовательно концентрация в клетке контролируется другими регуляторами. Примером служит система SoxS-SoxR в *E. coli*. Экспрессия регулятора генов ответа на окислительный стресс SoxS в свою очередь контролируется фактором транскрипции SoxR, который напрямую чувствует окислительно-восстановительное состояние клетки (38).



**Рисунок 1.2.** Механизмы работы белков-репрессоров. Фиолетовым цветом обозначен репрессор. Зеленым цветом – активатор.

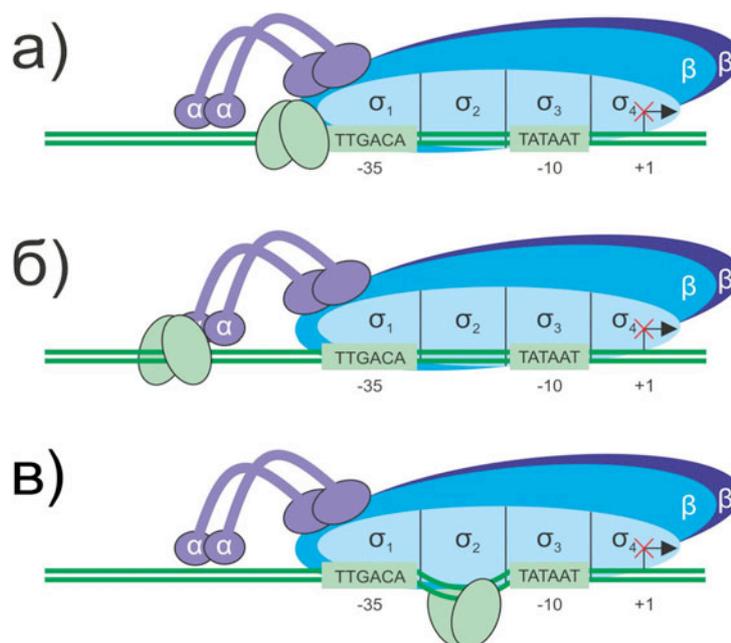
Факторы транскрипции могут исполнять функцию активатора или репрессора в зависимости от позиции связывания с ДНК относительно старта транскрипции и механизма взаимодействия с РНК полимеразой. Однако известны и белки выполняющие обе роли в зависимости от промотора (17).

Репрессоры подавляют инициацию транскрипции у регулируемых генов. Часто репрессия происходит по простому механизму, когда фактор транскрипции закрывает промотор от связывания с РНК полимеразой. Сайт связывания репрессора в таком случае часто перекрывается или находится в непосредственной близости от промотора (Рис. 1.2а). По этому механизму работает LacI репрессор (34). Иной принцип действия демонстрирует GalR репрессор из *E. coli* (39). В этом случае репрессор не препятствует связыванию РНК полимеразы с промотором, но путем связывания с несколькими сайтами связывания вокруг промотора создает петлю на ДНК, что делает невозможным дальнейшее продвижение РНК полимеразы (Рис. 1.2б). Третий механизм представлен белками анти-активаторами, которые взаимодействуют напрямую с активаторами транскрипции (Рис. 1.2в). Так, CytR в *E. coli* связывается со своим сайтом и активатором CRP, блокируя действие последнего (40).

Для белков-активаторов также выделяют несколько механизмов работы. Активаторы класса I связываются с сайтом в 5' области -35 элемента промотора и взаимодействуют с С-концевым участком  $\alpha$  субъединицы РНК полимеразы (Рис. 1.3б). Так функционирует регулятор CRP при активации транскрипции оперона *lacZYA* в *E. coli* (41). Сайты связывания активаторов класса II перекрываются с -35 элементом промотора, и регулятор взаимодействует непосредственно с доменом 4  $\sigma$  субъединицы РНК-полимеразы. А активаторы AraC семейства SoxS, MarA и Rob в *E. coli* действуют по такому принципу на части регулируемых промоторов (42). Активаторы I и II классов помогают РНК-полимеразе присоединиться к ее промотеру. Еще одним механизмом является изменение конформации промотора, что позволяет ему взаимодействовать с РНК-полимеразой. В этом случае активатор

присоединяется к ДНК в непосредственной близости от промотора или к самому промотору (Рис. 1.3в). Таким действием обладают активаторы MerR семейства, сайт посадки которых находится между -35 и -10 элементами промотора (43).

Таким образом, действие факторов транскрипции тесно связано со структурой промоторных областей регулируемых генов, что позволяет изучать регуляцию транскрипции с помощью анализа геномных последовательностей.



**Рисунок 1.3.** Механизмы действия белков-активаторов.

## 1.2 Методы сравнительной геномики для реконструкции регуляторных сетей у бактерий

К настоящему моменту разработано множество способов для экспериментального изучения регуляции транскрипции. Их можно разбить на две большие группы, каждая из которых имеет свои преимущества и недостатки. К первой группе относятся такие методы как направленный мутагенез, использование химерных конструкций (fusion construction), замедление ДНК в геле (gel shift assay) и определение защищенных от расщепления ДНК-азами и химическими реагентами участков (footprinting). Хотя эти методы позволяют успешно изучать регуляцию отдельных генов, но

они достаточно трудоемкие и имеют серьезные ограничения в производительности (44). Другой, относительно новый подход использует наборы высокопроизводительных экспериментов, таких как комбинирование анализа микрочипов с иммунопреципитацией хроматина (ChIP-on-chip) или скрининга библиотеки геномных последовательностей (Genomic SELEX). И хотя эти методы дают картину регуляторных взаимодействий на уровне целой клетки, возникает ряд сложностей с их применением. Во-первых, для обнаружения взаимодействия необходимо подобрать условия, при которых изучаемые факторы транскрипции активируются. Во-вторых, регуляторные каскады, ко-регуляция генов несколькими факторами транскрипции и другие не прямые эффекты создают сильный шум, что делает прямой анализ данных крайне затруднительным (45,46). Современное развитие методов биоинформатики позволяет не только во многих случаях преодолевать вышеперечисленные трудности в изучение регуляции генов, но и применять эти методы в качестве самостоятельного инструмента.

С большой скоростью растет количество полных геномов бактерий. С 1995 года, когда был полностью секвенирован геном паразитической бактерии *Haemophilus influenzae* Rd (47), к настоящему времени полностью прочитано почти 2000 бактериальных геномов. Такое количество позволяет эффективно применять методы сравнительной геномики, в частности для реконструкции известных регулонов в еще не изученных организмах или предсказания новых регулонов. Объединение биоинформатических методов поиска сайтов связывания факторов транскрипции с другими методами анализа генома, объединенными под общим названием анализ геномного контекста (genome context analysis) позволяют также существенно улучшить качество функциональной аннотации генов, предсказывать структуру оперонов и проводить метаболическую реконструкцию.

### ***1.2.1 Методы сравнительной геномики для анализа сайтов связывания факторов транскрипции***

Регуляторные сайты в разных местах ДНК, распознаваемые одним и тем же регуляторным белком, часто имеют различия в своих последовательностях. Так, например, глобальный регулятор CsrA в *B. subtilis* практически не имеет консервативных позиций в сайтах связывания. При этом все сайты проявляют общие свойства последовательности и все похожи на обобщенную последовательность, называемую консенсусом. Поэтому задачу распознавания сайтов связывания можно сформулировать следующим образом: как исходя из набора последовательностей, в которой мы имеем сильные основания подразумевать присутствие сайтов связывания одного фактора транскрипции, можно извлечь эти сайты.

Наиболее распространенным методом для поиска регуляторных сайтов является филогенетический футпринтинг (phylogenetic footprinting), в основе которого лежит поиск консервативных участков на выравнивании ортологичных последовательностей (48). Ортологичными называются гены в разных организмах, произошедшие путем наследования одного гена общего предка. Гены, произошедшие путем дубликации гена-предшественника в одном организме, называются паралогами (49). Однако паралоги часто меняют как свою функцию, так и регуляторные сайты (50,51), поэтому применение данного метода при изучении регуляции таких генов часто не оправдано. К тому же для филогенетического футпринтинга необходим достаточно высокий уровень сходства регуляторных последовательностей. Пользоваться данным методом можно как напрямую, анализируя множественные выравнивания, так и с помощью специализированных программ, например, FootPrinter (52).

Из-за ограниченной области применения филогенетического футпринтинга часто используются специализированные алгоритмы анализа регуляторных последовательностей, которые можно разделить на три категории (53): методы перечисления, детерминистической оптимизации и вероятностной оптимизации.

В основе методов перечисления лежит определение всех возможных сайтов, которые встречаются в обучающей выборке последовательностей и соответствуют некоторым граничным условиям для построения распознающего правила. Так, например, из обучающей выборки создаются группы схожих олигонуклеотидных последовательностей, которые затем ранжируются по вероятности встретить данную группу в случайной последовательности данного организма. Считается, что группы с лучшими статистическими оценками, возможно, представляют сайты связывания факторов транскрипции (54). В этой группе методов распознающее правило, отражающее регуляторный сигнал, обычно описано как консенсусная последовательность, позволяющая некоторое число ошибочных совпадений. Так, для регулятора *SsrA* консенсус выглядит, как WTGNNARCGNWWCAW (где W означает A или T, R – A или G и N – любое основание) (55). Однако, построить оптимальный консенсус и использовать его в качестве правила распознавания бактериальных сайтов крайне затруднительно (56).

Следующие две группы методов основаны на использовании матриц позиционных весов в качестве распознающих правил. Среди детерминистических алгоритмов наиболее популярным является метод максимизации ожидания (expectation maximization), состоящих из двух шагов. Изначальная матрица строится по одному найденному сайту. На первом шаге для каждого следующего сайта той же длины из представленной последовательности рассчитывается ожидаемая вероятность, что он является сайтом связывания фактора транскрипции, а не фоновым шумом. Затем производится сравнение вычисленных вероятностей для всех сайтов и происходит оптимизация распознающего правила (57). Варианты данного алгоритма реализованы в таких программах, как MEME (58), SignalX (59) и интернет сервере RegPredict (60).

В качестве примера алгоритма вероятностной оптимизации можно привести метод выборки по Гиббсу (Gibbs sampling). Алгоритм извлекает начальную случайную выборку сайтов из исходных последовательностей и

строит на этой основе начальное распознающее правило. На каждой итерации производится вероятностная оценка: нужно ли удалить один из исходных сайтов или добавить новый сайт из обучающей выборки, чтобы улучшить матрицу позиционных весов, отражающую вероятность связывания белка с сайтом (61). Алгоритм реализован в таких программах, как AlignACE (62), Gibbs Motif Sampler (63) и SeSiMCMC (64).

Каждый из этих алгоритмов определяет, является ли найденная последовательность сайтом связывания или нет по некоторой характеристике, выведенной из распознающего правила. Соответственно, возникает проблема выбора порогового значения. Опыт биоинформатической реконструкции даже хорошо изученных регулонов показывает, что выбрать порог, отделяющий истинные сайты от ложных, весьма трудоемкая задача. При завышении порога теряется значительное число экспериментально подтвержденных сайтов, при понижении возрастает число ложноположительных предсказаний. Здесь существенную помощь оказывают методы сравнительной геномики.

Обязательным условием применения методов сравнительной геномики является присутствие ортологичных факторов транскрипции во всех анализируемых геномах. Более того, выбор геномов сильно зависит от ожидаемой консервативности регуляторного сигнала между организмами. При слишком высоком родстве геномов регуляторные области обычно практически идентичны, что не позволяет делать выводы о ложности найденного сайта. В далеких бактериях регуляторные сигналы могут существенно отличаться и построение единого распознающего правила может быть невозможно (65).

Для этой цели широко используется метод проверки соответствия (consistency check), который основан на предположении о консервативности регулонов в близких видах, содержащих ортологичные факторы транскрипции (66). В случае, если сайт сохраняется перед оперонами, содержащими ортологичные гены в нескольких родственных организмах, то регуляция с большой вероятностью предсказана правильно. Если сайт обнаруживается лишь в одном геноме, то скорее всего он является перепредсказанием.

### ***1.2.2 Методы предсказания функции гена на основании сходства аминокислотных последовательностей***

Биологическую значимость в реконструкции регулонов несет не только описание ДНК-белковых взаимодействий, но и предсказание функций регулируемых генов. Стандартным способом определения функциональной аннотации для белковой последовательности является ее сравнение с экспериментально изученными белками. Аннотация переносится на изучаемый белок, если доказана его ортологичность экспериментально изученному. Для выявления сходства обычно пользуются поиском по белковым базам данных GenBank (67) или UniProt (68) с помощью семейства программ BLAST (2). Было замечено, что аминокислоты в белках меняются не случайным образом, поэтому поиск в базах данных ведется с помощью матриц подстановки аминокислот, таких как PAM (69) и BLOSSUM (70), которые отражают вероятности переходов одних аминокислот в другие. Для поиска отдаленных гомологий широко применяются методы основанные на позиционно-специфичных матрицах весов (position-specific score matrices), которые отражают вероятность встретить определенную аминокислоту в определенном месте в выравниваниях исходной последовательности и последовательностей из баз данных. Наиболее популярной программой, где реализован данный метод, является PSI-BLAST (2).

Дополнительную информацию о белке можно узнать из анализа доменов, сравнивая имеющийся белок с известными структурными мотивами, хранящимися в таких базах данных, как Pfam (71), SUPERFAMILY (72), SMART (73). Уточнить функцию генов можно путем идентификации специфических мотивов, характерных для определенных функциональных групп. Сигнальные пептиды секретлируемых белков можно обнаружить с помощью программы SignalP (74), трансмембранные сегменты – программой TMPred (75), ДНК-связывающие сегменты – программой NPS@ (76). Несмотря на успешность этого подхода во многих случаях, он испытывает большие затруднения в аннотации ранее не изученных генов и часто дает ошибки в

аннотации функционально гетерогенных семейств паралогов. Так методы основанные на сходстве аминокислотных последовательностей позволяют предсказать от 40 до 80% функций белок-кодирующих генов, в то время как остальные аннотируются как предполагаемые (hypothetical) (3).

### **1.2.3 Методы анализа геномного контекста**

Как отмечалось ранее, анализом геномного контекста называется группа методов сравнительной геномики, предназначенная для уточнения функциональной аннотации генов. Эти методы основываются на таких данных, как кластеризация генов на хромосоме, слияние генов и профили встречаемости генов (3). И хотя данные методы более трудоемкие и гораздо хуже поддаются автоматизации, чем методы, основанные на поиске гомологий, они позволяют существенно повысить эффективность аннотации на основании лишь геномной последовательности и нередко предсказывать функции генов, не имеющих гомологов.

Было замечено, что гены из одного метаболического пути часто встречаются в непосредственной близости на хромосомах бактерий (77,78). Это позволяет сделать утверждение, что гены, ортологи которых кластеризуются на хромосоме у различных видов, вероятно, имеют функциональное родство. Такие предсказания позволяют характеризовать гены, не имеющие экспериментально изученных гомологов, и реконструировать недостающие фрагменты метаболических путей. Примером может служить биоинформатическое описание биосинтеза жирных кислот в *Streptococcus pneumoniae*, где за счет кластеризации генов на хромосоме были открыты ферменты FabK и FabM (79).

Метод исследования профилей встречаемости генов исходит из положения, что гены входящие в один метаболический путь имеют тенденцию наследоваться либо пропадать из генома совместно (80). Поэтому группы генов, которые в месте встречаются в одних геномах и отсутствуют в других, могут иметь функциональную связь. Так, благодаря данному методу были

предсказаны функции генов, отвечающих за конечные стадии деоксиксилулоза фосфатного пути биосинтеза изопреноидов в *E. coli* – *ispG* и *ispH* (81,82).

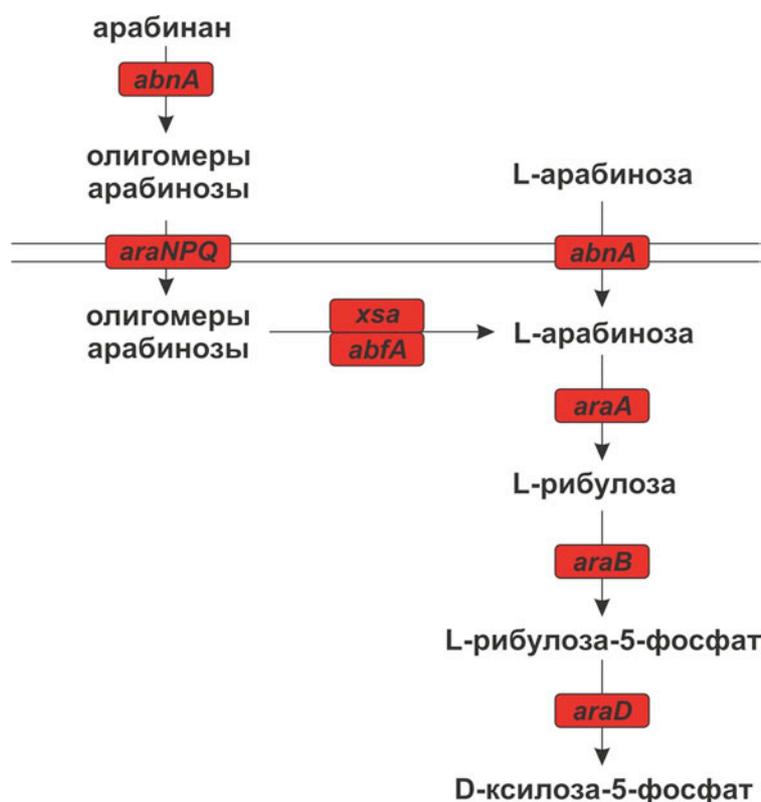
Метод поиска слияния генов основан на поиске различных генов из одного генома, которые образуют одну открытую рамку считывания в другом геноме, где, соответственно, кодируют один мультидоменный белок (83,84). Такие гены с большой вероятностью могут быть функционально связаны, входить в один метаболический путь или составлять один белковый комплекс. Поиск слияния генов использовался для предсказания функций при уточнении аннотаций таких геномов как *Methanobacterium thermoautotrophicum* (85) и *Thermotoga maritima* (86). Также с помощью этого метода в *Homo sapiens* был открыт фермент фосфопантетеин аденилилтрансфераза из пути биосинтеза кофермента А (КоА). В *H. sapiens* обнаружили гомолога известного в бактериях фермента дефосфо-КоА киназы из того же пути, слитого с доменом неизвестной функции, который при экспериментальной проверке обладал искомой активностью (87).

Методы анализа геномного контекста реализованы как в программах STRING (88) и PhydBac (89), так и в составе веб-серверов SEED (90) и MicrobesOnline (91).

### **1.3 AraR-регулон утилизации арабинозы**

В *B. subtilis* фактор транскрипции AraR регулирует экспрессию набора генов, кодирующих ферменты и транспортеры, необходимые для утилизации арабиноза-содержащих полисахаридов (Рис. 1.4). AraR – белок длиной 362 аминокислоты, который обладает химерной организацией, где два функциональных домена имеют разное филогенетическое происхождение (92). N-концевой ДНК-связывающий домен содержит мотив «крылатая спираль-поворот-спираль (winged helix-turn-helix, wHTH)», и принадлежит к GntR семейству факторов транскрипции (93). Большой C-концевой домен имеет гомологию с белками LacI семейства бактериальных регуляторов (94).

AraR регулон насчитывает пять оперонов. По два сайта было обнаружено в промоторных областях *araABDLMNPQ-abfA* (95,96), *abnA* (97) и *araE* (98). В случае генов *araR* (96) и *xsa* (97) было обнаружено по одному сайту. В отсутствие L-арабинозы, являющейся эффектором, AraR связывается с ДНК и репрессирует экспрессию генов (99). Консенсусом для сайтов связывания AraR является последовательность длиной 16 ппн – 5'-АТТТГТАСГТАСАААТ-3' (100). Было предложено два механизма AraR-зависимой регуляции. В случае присутствия двух сайтов в промоторной области происходит кооперативное связывание четырех молекул белка AraR в виде тетрамера, в результате чего в регуляторной области образуется петля на ДНК, что эффективно останавливает транскрипцию (99). При одиночном сайте транскрипция репрессируется слабее по механизму простого перекрытия фактором транскрипции регуляторной области.



**Рисунок 1.4.** AraR регулон в *B. subtilis* на основе экспериментальных данных.

Красным отмечены гены, находящиеся под репрессией.

Гены *araA*, *araB* и *araD* кодируют L-арабиноза изомеразу, L-рибулокиназу и L-рибулозо-5-фосфат-4-эпимеразу соответственно. Эти ферменты необходимы для превращения L-арабинозы в D-ксилулозу-5-фосфат, которая в дальнейшем входит в пентозофосфатный путь (101). Функция гена *araL* до сих пор остается неизвестной. Недавние исследования показали активность AraM в качестве глицерол-1-фосфат дегидрогеназы (102). И хотя связь данного фермента с путем утилизации арабинозы не ясна, было предположено, что он необходим для биосинтеза фосфоглицеролипидов.

**Таблица 1.1.** AraR регулоны в бактериях типа *Firmicutes*, полученные с помощью методов сравнительной геномики (8). Знак «/» обозначает дивергентное положение генов.

<i>B. subtilis</i>	<ul style="list-style-type: none"> <li>• <i>araE/araR</i></li> <li>• <i>araABDLMNPQ-abfA</i></li> <li>• <i>ydjK</i></li> <li>• <i>xsa</i></li> <li>• <i>abnA</i></li> </ul>
<i>Bacillus halodurans</i>	<ul style="list-style-type: none"> <li>• <i>abfA-araM</i></li> <li>• <i>araDBA-xsa</i></li> <li>• <i>araR</i></li> <li>• <i>BH1061</i></li> </ul>
<i>Bacillus stearothermophilus</i>	<ul style="list-style-type: none"> <li>• <i>araR</i></li> <li>• <i>araDBA</i></li> </ul>
<i>C. acetobutylicum</i>	<ul style="list-style-type: none"> <li>• <i>ptk</i></li> <li>• <i>abf2-CAC1530</i></li> <li>• <i>araDA/araR</i></li> <li>• <i>araK-araE2-araA2</i></li> <li>• <i>araE1</i></li> </ul>
<i>E. faecium</i>	<ul style="list-style-type: none"> <li>• <i>araK-araDA/araE</i></li> <li>• <i>araR</i></li> <li>• <i>abfA</i></li> </ul>

В AraR регулоне присутствуют две транспортные системы. Гены *araNPQ* кодируют ABC транспортер, участвующий в переносе олигомеров арабинозы (95). Другой ген, *araE* кодирует пермеазу широкой специфичности, которая

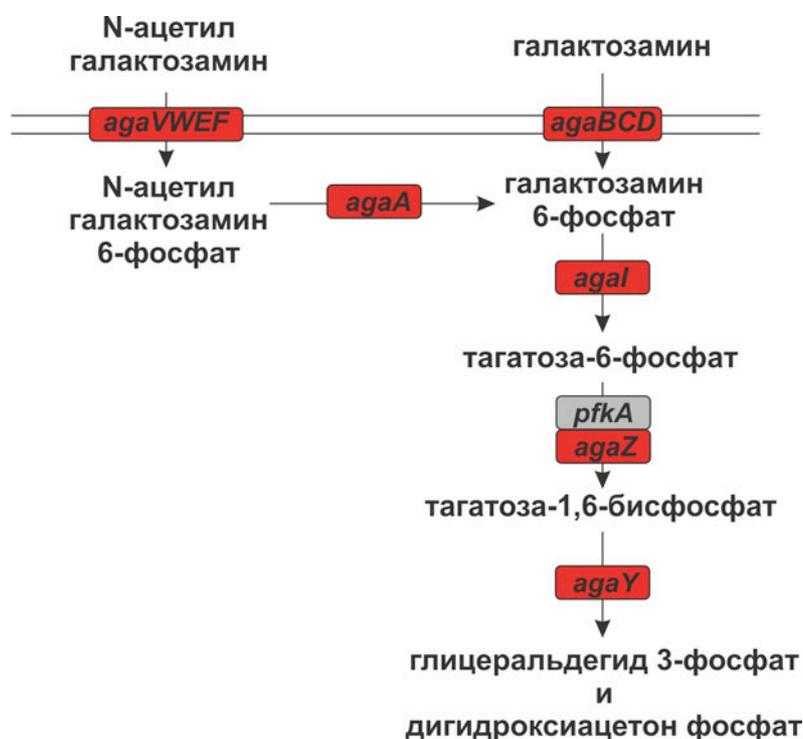
кроме L-арабинозы также способна к транспорту ксилозы и галактозы (103). Наконец, в составе регулона присутствуют три гидролазы. Арабинофуранозидазы AbfA и Xsa предположительно расщепляют арабинозосодержащие олигомеры внутри клетки (97), тогда как AbnA является секретлируемой эндо-арабинаназой, которая действует на полисахарид арабинан (104).

В 2001 году был осуществлен анализ AraR регулона методами сравнительной геномики в группах бактерий типа *Firmicutes* (8) (Табл. 1.1). Наиболее значимым результатом этой работы стало предсказание новой рибулокиназы AraK в качестве неортогичного замещения отсутствующей рибулокиназы AraB в *Clostridium acetobutylicum* и *Enterococcus faecium*. Ген *araK*, кодирующий сахарную киназу, в этих геномах кластеризуется с генами пути утилизации арабинозы. При этом, ортологи *araB* отсутствуют в изучаемых геномах, что доказывает функцию *araK*.

#### **1.4 Регуляция утилизации N-ацетилгалактозамина в протеобактериях**

Гены необходимые для утилизации N-ацетилгалактозамина первоначально были описаны в ходе биоинформатического анализа генома *E. coli* K-12 (105). Хромосомный кластер генов, отвечающих за утилизацию N-ацетилгалактозамина, включает в себя 12 генов, организованных в два оперона, *agaSYBCDI* и *agaZVWA*. Дивергентно с геном *agaZ* расположен ген фактора транскрипции *agaR*. Позднее было обнаружено, что в *E. coli* K-12 произошла делеция генов *agaEF*, кодирующих две компоненты фосфотрансферной системы (PTS – phosphotransferase system). Предсказанный путь включает транспорт и последующее фосфорилирование N-ацетилгалактозамина или галактозамина, деацетилирование N-ацетилгалактозамин-6-фосфата, деаминизацию/ изомеризацию галактозамин-6-фосфата, фосфорилирование тагатоза-6-фосфата и расщепление тагатоза-1,6-бисфосфата на глицеральдегид 3-фосфат и дигидроксиацетон фосфат (Рис. 1.5). Для PTS-систем кодируемых

генами *agaVWEF* и *agaBCD* была подтверждена специфичность к N-ацетилгалактозамину или галактозамину, соответственно (106). Два белка, кодируемые в *aga* кластере, AgaA и AgaI, оказались гомологичны N-ацетилглюкозамин-6-фосфат деацетилазе NagA и глюкозамин-6-фосфат деаминазе NagB из *E. coli* (105). На основании этих данных было выдвинуто предположение о функции AgaA и AgaI в качестве N-ацетилгалактозамин-6-фосфат деацетилазы и галактозамин-6-фосфат деаминазы, соответственно. Фермент, отвечающий за функцию тагатоза-6-фосфат киназы до сих пор не известен. Были предложены две альтернативные гипотезы: 1) фруктоза-6-фосфат киназа PfkA также обладает активностью на тагатоза-6-фосфате (107), 2) гипотетический белок AgaZ, кодируемый в *aga* кластере, является новым семейством тагатоза-6-фосфат киназ (105). Конечный шаг пути утилизации N-ацетилгалактозамина катализируется тагатоза-1,6-бисфосфат альдозазой AgaY (108). Функция белка AgaS до сих пор остается неизвестной.



**Рисунок 1.5.** AgaR регулон в *E. coli* на основе экспериментальных данных. Репрессируемые гены помечены красным. Гены, не контролируемые AgaR, отмечены серым.

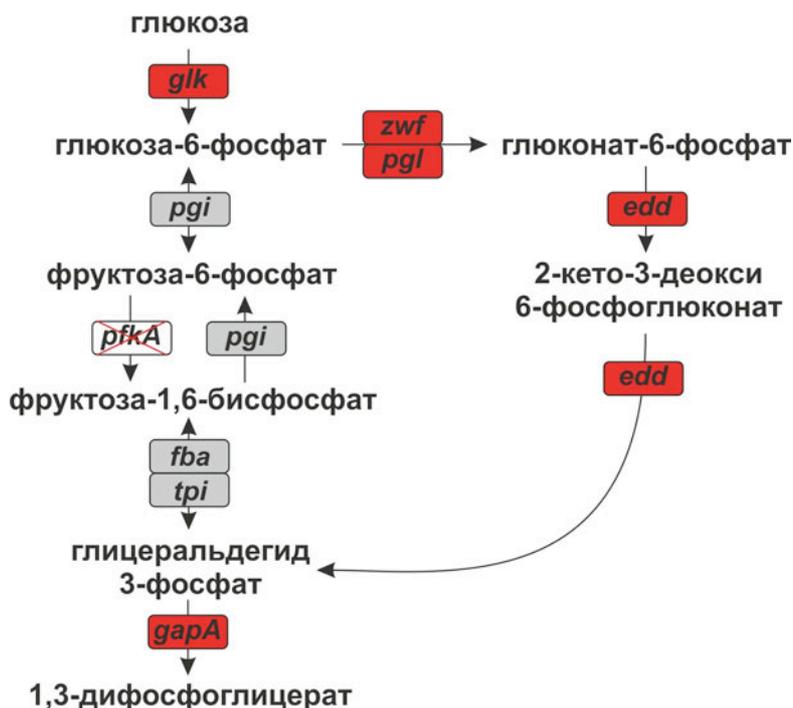
В *E. coli* транскрипция кластера генов утилизации N-ацетилгалактозамина регулируется AgaR репрессором из семейства факторов транскрипции DeoR (109) (Рис. 1.5). Белок AgaR распознает последовательности с консенсусом WRMMTTTCRTTTYRTTTYNYTTKK (где W – А или Т, Y – С или Т, R – А или G, M – А или С, K – G или Т), расположенные в промоторных областях генов *agaZ*, *agaS* и *agaR*. Эффектор AgaR неизвестен, однако было предположено, что им могут быть интермедиаты пути – N-ацетилгалактозамин-6-фосфат и/или галактозамин-6-фосфат (109).

Недавно была проведена геномная реконструкция нового варианта пути утилизации N-ацетилгалактозамина в четырех бактериях рода *Shewanella* (110), который содержит два фермента, гомологичные тем, что содержатся в *E. coli* (AgaS и AgaZ), и пять новых белков: неортологичная ацетилгалактозамин-6-фосфат деацетилаза AgaA-II, предсказанная N-ацетилгалактозамин киназа AgaK, и N-ацетилгалактозамин пермеаза AgaP, а также предполагаемая гидролаза AgaO и TonB-зависимый транспортер внешней мембраны Omp<sup>(aga)</sup>. Экспериментальная проверка показала возможность роста бактерий *Shewanella amazonensis* SB2B, *Shewanella* MR-4, MR-7, и ANA-3 на N-ацетилгалактозамине в качестве единственного источника углерода и энергии (110).

### 1.5 HexR – регулятор центрального метаболизма углерода

*E. coli* и родственные бактерии используют два глобальных фактора транскрипции для регуляции центральных и периферических путей утилизации углеводов – Crp и Cra (раннее называемый FruR) (111). Crp – рецептор циклического аденозин монофосфата (цАМФ) – осуществляет катаболитную репрессию генов при высоких концентрациях глюкозы – наиболее предпочтительного источника углерода у бактерий порядка *Enterobacteriales*. В присутствии цАМФ, Crp активирует гены путем связывания с сайтами в промоторной области ДНК. цАМФ генерируется аденилатциклазой, которая активируется компонентами глюкозо-специфичной PTS системы в отсутствие

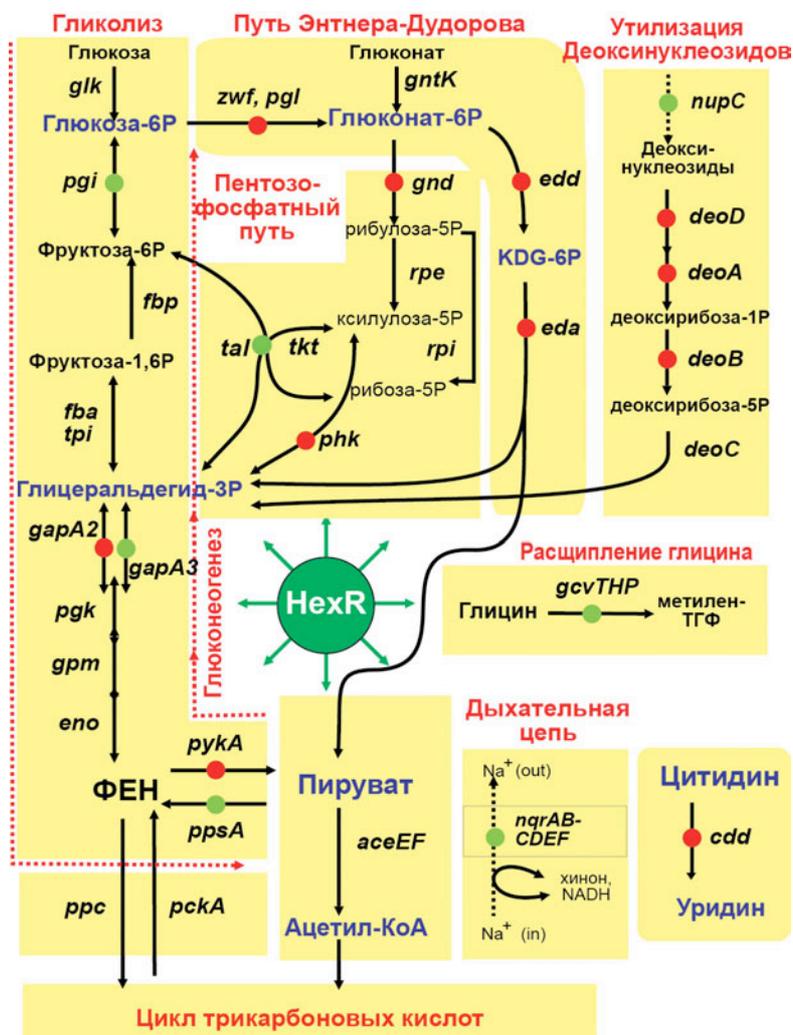
глюкозы в среде (112). Другой фактор транскрипции Csa из белкового семейства LacI был изначально охарактеризован как репрессор оперона утилизации фруктозы (113). Позднее был выяснен плейотропный характер регуляции, где репрессируются гены центральных гликолитических путей, таких как гликолиз и путь Энтнера-Дудорова, и активируются гены участвующие в глюконеогенезе и окислительном фосфорилировании (114).



**Рисунок 1.6.** HexR регулон в *P. putida* на основе экспериментальных данных. Репрессируемые гены помечены красным. Гены, не контролируемые HexR отмечены серым. Перечеркнутый ген *pfkA* отсутствует в геноме *P. putida*.

Иная стратегия регуляции генов центрального метаболизма углерода используется в бактериях рода *Pseudomonas*, для которых основным источником углерода являются органические кислоты и аминокислоты (115). В бактериях рода *Pseudomonas* отсутствует один из ключевых ферментов гликолиза – 6-фосфофруктокиназа (116). Поэтому катаболизм глюкозы происходит исключительно через путь Энтнера-Дудорова. Экспрессия всех генов, кодирующих ферменты пути Энтнера-Дудорова в *Pseudomonas putida*, таких как , глюкокиназа (*glk*), глюкоза-6-фосфат дегидрогеназа (*zwf*), 6-

фосфоглюконолактоназа (*pgl*), 6-фосфоглюконат дегидрогеназа (*edd*), 2-кето-3-деокси-6-фосфоглюконат альдозаза (*eda*) и глицеральдегид-3-фосфат дегидрогеназа (*gap-1*), репрессируются фактором транскрипции HexR (Рис. 1.6). Два мономера HexR связываются с несовершенным палиндромным сайтом с консенсусом вида TTGTN<sub>7,8</sub>ACAA в промоторных областях генов *zwf*, *edd* и *gap-1*. Связывание белка HexR с 2-кето-3-деокси-6-фосфоглюконатом, интермедиатом пути Энтнера-Дудорова, приводит к уменьшению сродства HexR к сайту связывания и освобождению промотора для инициации транскрипции (116-119).



**Рисунок 1.7.** HexR регулон в *Shewanella oneidensis* MR1 на основе данных сравнительной геномики. Красными точками гены с предполагаемой репрессией. Зелеными точками отмечены гены с предполагаемой активацией.

Недавно методами сравнительной геномики был реконструирован HexR регулон в группе бактерий рода *Shewanella* (120) (Рис. 1.7). По сравнению с *Pseudomonas putida*, регулон заметно расширился и стал контролировать множество генов основных путей центрального метаболизма углерода: гликолиза, пентозо-фосфатного пути и пути Энтнера-Дудорова. Также по положению сайтов связывания HexR относительно промотора была предсказана возможная двойная роль HexR регулятора в качестве активатора и репрессора.

### **1.6 Механизмы регуляции путей утилизации сахаров в *B. subtilis***

*B. subtilis* способна использовать по крайней мере два десятка различных моно- или дисахаридов в качестве единственных источников углерода и энергии. Общая схема утилизации углеводов включает в себя транспорт веществ в клетку, фосфорилирование и последующий катаболизм через гликолиз или пентозофосфатный путь (121,122). Между тем, несмотря на то, что множество работ было посвящено изучению катаболизма углерода в *B. subtilis*, они в большинстве затрагивают гликолиз, и информация же о работе пентозо-фосфатного пути ограничена в основном сравнением с *E. coli* (123).

Для индукции генов, кодирующих ферменты периферических путей утилизации сахаров, необходимо присутствие в среде соответствующего субстрата и отсутствие при этом предпочтительных источников углерода и энергии. Обычно ферменты катаболизма сахаров находятся под двумя типами регуляции. Первый механизм осуществляется локальными факторами транскрипции, для которых эффекторами служат различные субстраты или интермедиаты их катаболизма, а также белок-белковые взаимодействия. Факторы транскрипции, участвующие в этом механизме, подразделяются по типу получаемого сигнала на PTS-зависимые и субстрат-связывающие. Второй механизм называется катаболитной репрессией и контролируется небольшим числом глобальных регуляторов, которые опережают концентрацию наиболее предпочтительных субстратов в среде (124).

### **1.6.1 Регуляция метаболизма углеводов с помощью фосфотрансферазных транспортных систем**

PTS системы служат как для транспорта веществ в клетку, так и для передачи сигналов внешней среды (125). Большинство транспортеров, переносящих сахара в *B. subtilis*, являются PTS системами. Они состоят из трех субъединиц: фермент I (EI), HPr белок и субстрат-специфичный фермент II (EII). В *B. subtilis*, также как и во многих бактериях, гены, кодирующие субъединицу EI (*ptsI*) и HPr (*ptsH*), присутствуют в одной копии и организованы в оперон *ptsHI* (126). Оперон *ptsHI* конститутивно экспрессируется в *B. subtilis*, что отражает как необходимость двух данных белков для работы систем транспорта множества сахаров в клетку, так и включенность их в различные регуляторные процессы (127). Субъединица HPr у грамположительных бактерий несет два сайта фосфорилирования: His-15, сайт фосфорилирования белком EI, и Ser-46, который фосфорилируется специальной АТФ-зависимой HPr киназой (128). Субъединица EII может состоять из трех или четырех доменов (EIIA-EIID), которые могут быть организованы как в одном белке, так и отдельно. Транспорт углевода осуществляется связанными с мембраной доменом EIIC и, при наличии, EIID. Фосфат переносится системой EI, HPr-His15, EIIA и EIIB (129).

Из всех типов транспортеров, PTS системы оказывают наибольший вклад в регуляцию транскрипции путем фосфорилирования факторов транскрипции HPr и EIIB субъединицами. Основываясь на характеристиках сайта фосфорилирования, различают два механизма регуляции с помощью PTS систем.

В первом случае EIIA домен, акцептор фосфорной группы от HPr-His-P, соединен с белком-регулятором. Такой домен не участвует в транспорте сахаров, но способен контролировать активность своего регуляторного белка в ответ на фосфорилирование соответствующими PTS системами. Примером могут служить факторы транскрипции LicR (130) из *B. subtilis*, а также MtlR из *B. stearothermophilus* (131). Однако стоит заметить, что регуляция активности

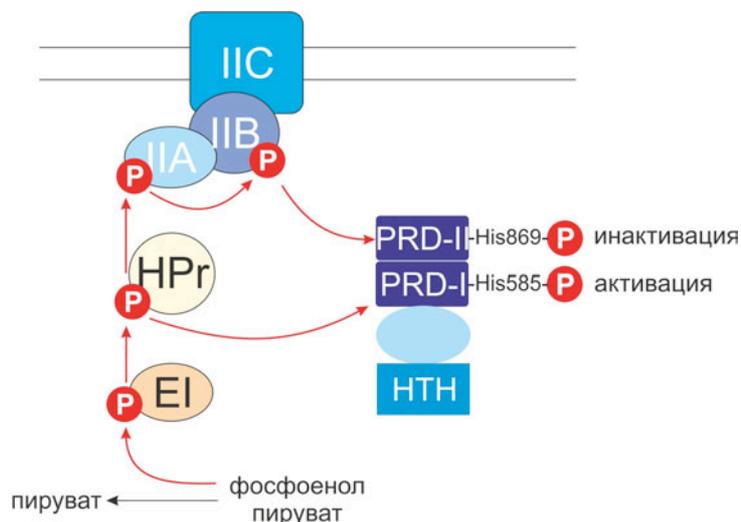
этих белков также происходит и по второму механизму, о котором говорится далее.

Белки, работающие по второму механизму, содержат в своем составе домены регуляции фосфотрансферной системой (PRD – PTS regulatory domain) (132). PRD домены в ответ на фосфорилирование компонентами PTS систем участвуют в регуляции РНК-связывающей активности транскрипционных антитерминаторов, таких как SacT, SacY, LicT и GlcT (132) или ДНК-связывающей функции активаторов транскрипции. PRD домены не обладают сходством с компонентами PTS систем, тем не менее, они способны фосфорилироваться белками HPr-His-P или EIIВ-P. Большинство PRD-содержащих белков имеют два PRD домена, организованных последовательно либо разделенных EIIА или EIIВ доменами. При этом фосфорилирование каждого из PRD доменов может независимо как усиливать, так и ингибировать способность к связыванию регулятора с ДНК или РНК (129,133).

В *B. subtilis* большинство PTS-зависимых активаторов работают именно по описанным выше механизмам. Так, активатор LicR контролирует экспрессию генов утилизации лишая (134). Он состоит из N-концевого НТН домена, двух PRD доменов и C-концевого домена сходного с EIIА субъединицей PTS систем. Было показано, что фосфорилирование двух PRD доменов стимулирует активаторную функцию LicR. При этом фосфорилирование EIIА домена лишая PTS системой приводит к инактивации LicR независимо от состояния фосфорилирования PRD доменов (130).

Гены утилизации левана контролируются регулятором LevR, который содержит N-концевой НТН домен, домен А для гидролиза АТФ и активации РНК-полимеразы, два PRD домена и домен сходный с EIIА (133,135) (Рис. 1.8). Было показано, что N-концевая часть LevR, содержащая только НТН домен и домен А достаточна для связывания со своими сайтами и конститутивной активации транскрипции (136). Модуляцию активности LevR осуществляют два PRD домена, причем в разные стороны. Инактивация LevR происходит

благодаря фосфорилированию His869 в С-концевом PRD домене EIB субъединицей леваназной PTS системы. Активация же LevR происходит при фосфорилировании His585 в другом PRD домене белком HPr (137).



**Рисунок 1.8.** Модель механизма регуляции фактора транскрипции с помощью PTS системы на примере белка LevR.

Еще одним типом регуляторов транскрипции, работающих совместно с компонентами PTS систем, являются антитерминаторы BglG семейства, которые меняют свою РНК-связывающую активность в зависимости от фосфорилирования двух PRD доменов (132). Транскрипция соответствующих оперонов конститутивна, но останавливается благодаря присутствию альтернативных структур, которые могут образовывать терминаторные шпильки в 5`-некодирующей области мРНК (127,138,139). При этом 5` регион самих терминаторов перекрывается с инвертированным повтором. Эти последовательности называются рибонуклеиновыми антитерминаторами (RAT – ribonucleic antiterminator) и образуют альтернативную терминаторную шпильку, с которой связываются белки антитерминаторы (140). В *B. subtilis* известно четыре подобных антитерминатора, участвующие в метаболизме сахаров – два регулятора утилизации сахарозы SacT и SacY (141), регулятор утилизации глюкозы GlcT (127) и регулятор утилизации  $\beta$ -глюкозидов LicT (142) (Табл. 1.2).

### **1.6.2 Регуляция метаболизма углеводов субстрат-связывающими факторами транскрипции**

Если большинство активаторов и антитерминаторов, участвующих в регуляции транскрипции генов утилизации сахаров в *B. subtilis*, в качестве сигнала используют фосфорилирование компонентами PTS систем, то репрессоры в большинстве случаев обладают доменами, способными связывать малые молекулы. Обычно эти молекулы являются интермедиатами путей утилизации соответствующих сахаров. Среди семейств, к которым в *B. subtilis* принадлежат регуляторы сахарных оперонов, работающие по данному механизму, самыми многочисленными являются LacI и GntR (Табл. 1.2). В сумме, около половины известных факторов транскрипции принадлежат этим двум семействам, регулирующих работу сахарных оперонов (26,122). Наиболее изученными факторами транскрипции являются описанный выше регулятор утилизации арабинозы AraR (92,99,100), глюконата GntR (143,144) и ксилозы XylR (145,146).

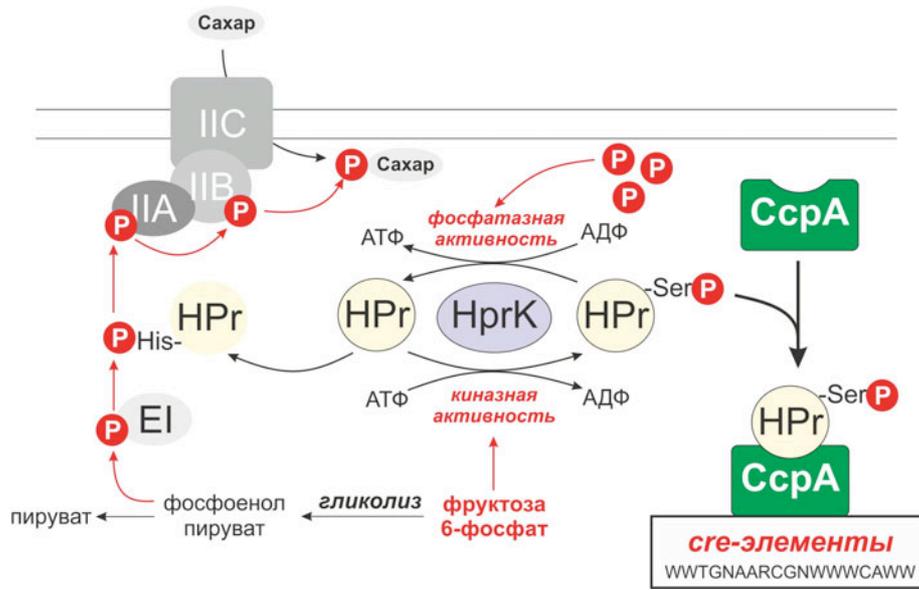
Также интересен единственный охарактеризованный в *B. subtilis* случай регуляции оперона утилизации сахаров двухкомпонентной системой. Данная система контролирует экспрессию генов катаболизма малата – *ywkAB*, *maeN* и *yflS* и состоит из регулятора ответа MalR и гистидиновой киназы MalK (147).

Единственный известный антитерминатор, контролирующей работу генов сахарного катаболизма в результате связывания с субстратом – это регулятор генов утилизации глицерола GlpP. В присутствие глицерол-3-фосфата GlpP связывается со шпилькой на мРНК и позволяет продолжить элонгацию транскрипции (148).

### **1.6.3. CsrA-зависимая катаболитная репрессия оперонов метаболизма сахаров**

Катаболитной репрессией называется регуляторный механизм, при котором клетка меняет метаболизм для максимально эффективной утилизации источников углерода и энергии (149). Гены катаболизма, подверженные

катаболитной репрессии, не экспрессируются до тех пор, пока в среде присутствуют более предпочтительные источники углерода и энергии. Этот механизм в *B. subtilis* осуществляется с помощью набора факторов транскрипции, главным из которых является глобальный регулятор – белок регуляции катаболизма А (CcpA – catabolite control protein A) (Рис. 1.9).



**Рисунок 1.9.** Модель механизма CcpA-зависимой катаболитной репрессии в *B. subtilis*.

Изучение катаболитной репрессии в *B. subtilis* началось с исследования регуляции транскрипции гена  $\alpha$ -амилазы *amyE*, в результате чего были открыты два основных компонента данного механизма – палиндромная последовательность длиной 14пн, названная *cre* (catabolite responsive element) и белок CcpA, принадлежащий к LacI семейству факторов транскрипции (150,151). К настоящему времени найдено около 50 *cre* последовательностей для различных оперонов в геноме *B. subtilis* (149). В результате геномного анализа был предложен расширенный консенсус длиной 18пн – WWTGNAARCGNWWCAWW (где W – А или Т, R – А или G, N – любой нуклеотид) (152).

Еще одним важным элементом механизма катаболитной репрессии является HPr киназа, кодируемая геном *hprK*, задачей которой является

фосфорилирование белка HPr в позиции Ser-46 (153,154). HPr киназа обладает как киназной, так и фосфатазной активностью (155). Фосфорилирующая активность HPr киназы возрастает в присутствии фруктозы-1,6-бисфосфата, фосфатазная – при повышении концентрации неорганического фосфата (154,155), что позволяет HPr киназе осуществлять функции передачи сигнала о наличии быстро усваиваемых субстратов, например глюкозы или фруктозы.

Экспрессия CsrA белка в клетке происходит конститутивно (156), однако для правильного функционирования регулятора необходим ко-репрессор. Было показано, что белок HPr, фосфорилированный по остатку Ser46, способен взаимодействовать с CsrA (157), и данный белковый комплекс связывается с множественными *cre* сайтами на ДНК (158). Другим ко-репрессором для CsrA является белок Crh, который обладает сходством с HPr (159). Crh также содержит консервативный остаток Ser-46, который может быть фосфорилирован с помощью HprK киназы, однако в нем отсутствует His-15 (159). Недавно было показано, что дефосфорилированный Crh ингибирует метилглиоксаль синтазу MglA путем образования неактивного комплекса. Данный фермент производит реакцию превращения дегидроксиацетон фосфата в метилглиоксаль, что позволяет избежать накопления фосфорилированных интермедиатов гликолиза при избытке углеводных субстратов. Однако сам метилглиоксаль является токсичным и его синтез нуждается в тщательном контроле, который и осуществляет Crh (160). Crh-Ser46-P также, как и HPr-Ser46-P, в комплексе с CsrA усиливает его сродство к *cre* сайтам (145,159,161). В *B. subtilis* Crh содержится в гораздо меньшей концентрации, чем HPr (162). К тому же делеция *crh* не влияет на катаболитную репрессию в присутствии неповрежденной копии *hpr*. Однако Crh-Ser46-P способен лишь частично замещать HPr-Ser46-P в HPr-Ser46Ala мутанте (159,163). Таким образом основная функция Crh в катаболитной репрессии до сих пор не ясна.

В зависимости от расположения *cre* сайтов относительно промотора регуляция комплексом CsrA/HPr-Ser46-P может происходить по различным механизмам – активации или репрессии. Так, для активации *cre* сайт должен

располагаться в 5`-области промотора, как, например для генов *ackA* (164) или *ilvB* (165). Недавнее исследование показало, что для активации транскрипции *ackA* в промоторной области происходит образование комплекса CsrA/HPr-Ser46-P, фактора транскрипции CodY и РНК полимеразы (166). Вероятно, в других случаях активация происходит также путем прямого взаимодействия CsrA/HPr-Ser46-P и РНК полимеразы.

На промоторе *amyE*, где *cre* сайт перекрывается с промотором, было показано, что комплекс CsrA/HPr-Ser46-P не мешает посадке РНК полимеразы. Из этого был сделан вывод об ингибировании транскрипции путем взаимодействия CsrA и РНК полимеразы (167). Если *cre* сайт находится в 3`-области промотора, то также предполагается связывание CsrA с РНК полимеразой за счет образования петли на ДНК, как в случае с *hxl* опероном (167).

#### **1.6.4. CsrA-независимая катаболитная репрессия оперонов метаболизма сахаров**

Кроме CsrA, в *B. subtilis* были обнаружены еще три фактора транскрипции, участвующие в катаболитной репрессии – CsrB (168), CsrC (169) и CsrN (170). Среди этих белков паралогом CsrA является только CsrB, который участвует в катаболитной репрессии генов утилизации глюконата *gnt* и ксилозы *hxl* в клетках, растущих на твердой среде или при низкой скорости перемешивания жидкой среды. Таким образом, можно предположить, что CsrB дополнительно контролирует катаболитную реPRESSION в зависимости от некоторых дополнительных условий, таких как плотность клеток, аэрация, фаза роста или концентрация субстратов. Однако, механизм, с помощью которого CsrB осуществляет реPRESSION, до сих пор остается неизвестным (168).

Белок CsrC принадлежит к LysR семейству и реPRESSION гены, кодирующие ферменты начала цикла трикарбоновых кислот – *citB*, *citZ* и *citC* (169). Гены *citB* и *citZ* также регулируются CsrA (171). Транскрипция гена *csrC* находится как под авторегуляцией, так и контролируется CsrA (172).

Исследование просмотра гена *citB* показало, что в связывании CsrC с ДНК участвуют два *cis*-элемента – несовершенный палиндром АТААГТСГААСТТАТ с центром в позиции -66 от старта начала транскрипции и половина палиндрама АСТТАТ в позиции -27, перекрывающаяся с -35 элементом промотора (173).

Наконец, CsrN при гликолитических условиях репрессирует гены глюконеогенеза *pckA* и *gapB*, а также некодирующую РНК SR1, влияющую на трансляцию регулятора катаболизма аргинина *AhrC* и стабилизирующую мРНК, транскрибируемую с *gapA* промотора (174). Позднее была найдена активация гена *thyB* (175). Эксперименты по определению защищенных фрагментов ДНК показали, что CsrN в каждом из промоторов закрывает два сайта и был предложен консенсус для сайтов распознавания TGTGNYATAC, где Y – Т или С, N – А, С или Т (176).

Механизм регуляции *gapA-pgk-tpi-pgm-eno* оперона фактором транскрипции CggR также тесно связан с катаболитной репрессией (177). Гены данного оперона кодируют центральные ферменты гликолиза. Эффектором, снижающим сродство CggR к своим сайтам служит фруктоза-1,6-бисфосфат, центральный интермедиат гликолиза, что, таким образом, одновременно является сигналом для включения CsrA-зависимой катаболитной репрессии и активации генов гликолиза (177).

Таблица 1.2. Регулоны утилизации сахаров в *B. subtilis*.

Имя регулятора	Функция регулируемых генов	Семейство	Механизм действия	Регулируемые гены	Ссылки*
ManR	Утилизация маннозы	BglG	активатор	<i>manR</i> <i>manPA</i>	(178) P
LicR	Утилизация бета-глюкозидов	BglG	активатор	<i>licBCAH</i>	(130) C
MtlR	Утилизация маннитола	BglG	активатор	<i>mtlAFD</i>	(179,180) C(др.), P
CspN	Глюконеогенез	CspN	репрессор	<i>gapB</i> <i>pskA</i>	(176) C
CitT	Утилизация цитрата	CitB	активатор	<i>citM</i>	(181) C
MalR	Утилизация малата	CitB	активатор	<i>ywkAB</i> <i>maeN</i> <i>yflS</i>	(147,182) C
IolR	Утилизация инозитола	DeoR	репрессор	<i>iolABCDEFGHIJ</i> <i>iolRS</i> <i>iolT</i>	(183,184) C
AcoR	Утилизация ацетона	Fis	активатор	<i>acoABCL</i>	(185) C
FrIR	Утилизация фруктозолизина	GntR	репрессор	<i>frIR</i> <i>frIB</i> <i>frIONMD</i>	(186) P
LutR	Утилизация лактата	GntR	репрессор	<i>lutABC</i>	(187) P
GmuR	Утилизация глюкоманнана	GntR	репрессор	<i>gmuBACDREFG</i>	(188) P
GudR	Утилизация глюкората Утилизация галактората	GntR	репрессор	<i>gudR</i> <i>garD</i> <i>ycbCD-gudD</i>	(189) P

NagR	Утилизация N-ацетилглюкозамина	GntR	репрессор	<i>nagABR</i> <i>nagP</i>	(190) C
TreR	Утилизация трехалозы	GntR	репрессор	<i>trePAR</i>	(191) C
GntR	Утилизация глюкозата	GntR	репрессор	<i>gntRKPZ</i>	(144) C
AraR	Утилизация арабинозы	GntR	репрессор	<i>abnA</i> <i>araE</i> <i>araR</i> <i>ksa</i> <i>araABDLMNPRQ-abfA</i>	(97,99) C
GutR	Утилизация сорбитола	GutR	активатор	<i>gutBP</i>	(192) C
HxIR	Рибулоза монофосфатный путь	HxIR	активатор	<i>hxlAB</i>	(193) C
NtdR	Утилизация неотрехалозамина	LacI	активатор	<i>ntdABC</i>	(194) P
GanR	Утилизация галактана	LacI	репрессор	<i>сусB-ganPQAB</i>	(195) P
CcpB	Метаболизм углерода	LacI	репрессор	<i>gntRKPZ</i> <i>хylAB</i>	(168) P
CcpA	Метаболизм углерода	LacI	репрессор/ активатор	<i>cimH; amyE; gntRKPZ; kdgRКАТ;</i> <i>manR; citST; levDEFG-sacC;</i> <i>срС; араЕ; ухjC-scoAB-ухjF;</i> <i>abnA; dctA; phoPR; trePAR;</i> <i>msmX; galT; асоABCCL;</i> <i>hutPHUIGM; cccA; glvARC;</i> <i>bgIPH; citM; licBCAH; acuABC;</i> <i>bgIS; хynPB; ackA; citZ-icd-mdh;</i> <i>ухаC-ехuM-ухmCD-ухuA-ухmF-</i> <i>ехuTR-ухaBA; gmuBACDREFG;</i>	(142,145,152,167, ,172,178,185,196 -216) C

*pta*; *xyxAB*; *acoR*; *lcfA*; *sigL*;  
*rbsRKDACB*; *iolABCDEFGHIJ*;  
*araABDLMNPQ-abfA*; *ilvBHC-*  
*leuABCD*; *acsA*; *glpFK*;  
*mngABCDE-prpB*; *alsSB*;  
*cydABCD*; *malA-yfiA-malP*; *ydhO*;  
*yobO*; *yxkJ*

KdgR	Утилизация пектина	LacI	репрессор	<i>kdgRKAT</i> <i>kduID</i>	(217) C
ExuR	Утилизация глюкоуроната Утилизация галактуроната	LacI	репрессор	<i>ixaC--exuM-yjmCD-ixaA-yjmF-</i> <i>exuTR-ixaBA</i>	(218) C
RbsR	Утилизация рибозы	LacI	репрессор	<i>rbsRKDACB</i>	(219) C(др.)
LevR	Утилизация левана	LevR	активатор	<i>levDEFG-sacC</i>	(136) C
AlsR	Биосинтез ацетоина	LysR	репрессор/ активатор	<i>alsR</i> <i>alsSD</i>	(220) P
CitR	Метаболизм цитрата	LysR	репрессор	<i>citA</i>	(221) P
CcpC	Метаболизм цитрата	LysR	репрессор	<i>ccpC</i> <i>citB</i> <i>citZ-icd-mdh</i>	(169,172) C
XyIR	Утилизация ксилозы	ROK	репрессор	<i>xyLAB</i> <i>xy/R</i> <i>хynPB</i>	(145,146) C
GlvR	Утилизация мальтозы	RpiR	активатор	<i>glvARC</i>	(203) P
CggR	Гликолиз	SorC	репрессор	<i>cggR-gapA-pgk-triA-pgm-eno</i>	(177) C
GlcT	Утилизация глюкозы	BglG	антитерми натор	<i>ptsGHI</i>	(222)
SacT/	Утилизация сахарозы	BglG	антитерми	<i>sacB-yveBA</i>	(140,141,223,224)

SacY			натор	<i>sacPA-ywdA</i> <i>sacXY</i>	)
LicT	Утилизация бета- глюкозидов	BglG	антитерми натор	<i>bgIPH-yxiE</i> <i>licT-bgIS</i>	(142,225)
GlpR	Утилизация глицерола	GlpR	антитерми натор	<i>glpD</i> <i>glpFK</i> <i>glpTQ</i>	(148,226,227)

\* - P – экспериментально была показана регуляция генов; C – экспериментально был показан сайт связывания фактора транскрипции у регулируемых генов; C(др.) – сайт был экспериментально показан для ортологичного фактора транскрипции в другом организме

## Глава 2

### Материалы и методы

#### 2.1 Принципы применения методов сравнительной геномики к анализу регуляции транскрипции

Геномы для проведения реконструкции определялись по следующей процедуре. С помощью поиска гомологов в базе данных «nr» интернет ресурса NCBI (<http://ncbi.nlm.nih.gov>) были получены факторы транскрипции гомологичные исходному – экспериментально подтвержденному или предполагаемому регулятору. Из полученного списка геномов, содержащих гомологи анализируемого фактора транскрипции, исключались близкие штаммы. Белковые последовательности факторов транскрипции из оставшихся геномов проверялись на ортологичность, а у соответствующих генов производился анализ геномного окружения. Эта технология позволяет определить регуляторы, которые поменяли свою функцию. По белковым последовательностям оставшихся факторов транскрипции строилось филогенетическое дерево, которое позволяет определить группы белков для построения общего распознающего правила поиска сайтов связывания.

Для поиска потенциальных сайтов связывания регуляторных белков применялись два метода: филогенетический футпринтинг и метод матриц позиционных весов (59,65). Для применения филогенетического футпринтинга строится множественное выравнивание промоторных областей ортологичных генов, в которых предполагается сохранение структуры сайтов связывания. Найденная последовательность предполагаемых сайтов связывания использовалась как в качестве обучающей выборки для построения матриц позиционных весов (анализ регулонов AraR, HexR, регулонов сахарного метаболизма в отряде *Bacillales* и частично при реконструкции AgaR регулонов), так и в качестве самостоятельного инструмента (широко применялось при анализе AgaR регулонов). С помощью филогенетического футпринтинга сложно определить таксон-специфическую регуляцию, однако

этот метод хорошо работает для определения слабых сайтов связывания в локальных регулонах.

Суть метода матриц позиционных весов заключается в следующем: на основе выравнивания регуляторных сайтов, каждый из которых имеет длину  $L$ , так называемой обучающей выборки, вычисляется вес  $W(b,i)$  каждого нуклеотида  $b$  в позиции  $i$ . Позиционные веса вычисляются по формуле:

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

где  $N(b,i)$  – частота нуклеотида  $b$  в позиции  $i$ . С помощью полученной матрицы можно для любой последовательности длины  $L$  вычислить вес  $S$ , равный

$$S = \sum_{i=1...L} W(b_i,i)$$

где  $b_i$  – нуклеотид в позиции  $i$ . Для отбора потенциальных сайтов связывания используют пороговые значения, как нижнюю границу для веса сайтов. В большинстве случаев пороговым значением выбирается минимальный вес из обучающей выборки. Поиск потенциальных сайтов производился в некодирующих областях на расстоянии от -400 до +50 нуклеотидов относительно старта трансляции.

Основным методом определения достоверности регуляции гена в данной работе является метод проверки соответствия. Согласно этому методу, ген включается в регулон при выполнении двух условий. Во-первых, если в промоторной области гена или промоторной области оперона, которому он принадлежит, был обнаружен потенциальный сайт связывания. Во-вторых, если сайт консервативен перед ортологичными генами в родственных геномах, содержащих также и ортологи регулятора (65). Дополнительным критерием включения гена в регулон служило сочетание высокого веса сайта с функциональной аннотацией гена, соответствующей функции регулона.

## 2.2 Геномы

Поскольку основой для выбора изучаемых геномов служило распределение изучаемых факторов транскрипции, в данной работе были реконструированы регулоны в большом количестве геномов. Только для реконструкции сетей регуляции метаболизма сахаров в отряде *Bacillales* геномы были отобраны заранее, исходя из филогенетического дерева бактериальных видов, предоставляемого порталом MicrobesOnline (91). В работе наибольшее число геномов принадлежит организмам, относящимся к двум типам с наибольшим количеством полностью секвенированных геномов – грамположительные бактерии относящиеся к типу *Firmicutes* и грамотрицательные бактерии из типа *Proteobacteria*. Согласно базе данных KEGG (1) на 6 сентября 2012 года известно 202 полные последовательности для различных видов *Firmicutes* и 473 – для *Proteobacteria*. Все геномы взяты из базы MicrobesOnline (91). Стоит отметить, что большинство изученных геномов имеет важное хозяйственное значение. Среди них содержатся симбионты и патогены растений и животных. Также множество штаммов применяется в биотехнологическом производстве.

Всего было исследовано 110 геномов, из которых 103 полных генома принадлежат бактериям из трех типов: *Firmicutes*, *Proteobacteria* и *Thermotogae* (Табл. 2.1).

Таблица 2.1 Список изучаемых геномов с условными обозначениями

Тип	Порядок	Организм	Обозначение	
<i>Firmicutes</i>	<i>Bacillales</i>	<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168	BSU	
		<i>Bacillus amyloliquefaciens</i> FZB42	BAY	
		<i>Bacillus cereus</i> ATCC 14579	BCE	
		<i>Bacillus clausii</i> KSM-K16	BCL	
		<i>Bacillus halodurans</i> C-125	BHA	
		<i>Bacillus licheniformis</i> DSM 13	BLD	
		<i>Bacillus pumilus</i> SAFR-032	BPU	
		<i>Anoxybacillus flavithermus</i> WK1	AFL	
		<i>Geobacillus kaustophilus</i> HTA426	GKA	
		<i>Oceanobacillus iheyensis</i> HTE831	OIH	
		<i>Paenibacillus</i> sp. JDR-2	PJD	
	<i>Clostridiales</i>	<i>Clostridium acetobutylicum</i> ATCC 824	CAC	
		<i>Clostridium beijerincki</i> NCIMB 8052	CBE	
		<i>Clostridium cellulolyticum</i> H10	CCE	
		<i>Clostridium</i> sp. SS2/1 (*)	CSS	
	<i>Lactobacillales</i>	<i>Lactobacillus brevis</i> ATCC 367	LBR	
		<i>Lactobacillus fermentum</i> IFO 3956	LFE	
		<i>Lactobacillus plantarum</i> WCFS1	LPL	
		<i>Lactobacillus reuteri</i> JCM 1112	LRF	
		<i>Lactobacillus sakei</i> subsp. <i>sakei</i> 23K	LSA	
		<i>Leuconostoc citreum</i> KM20	LCI	
		<i>Leuconostoc mesenteroides</i> subsp. <i>mesenteroides</i> ATCC 8293	LME	
		<i>Oenococcus oeni</i> PSU-1	OOE	
		<i>Pediococcus pentosaceus</i> ATCC 25745	PPE	
	<i>Proteobacteria</i>	<i>Pseudomonadales</i>	<i>Pseudomonas aeruginosa</i> PAO1	PAE
			<i>Pseudomonas entomophila</i> L48	PEN
			<i>Pseudomonas putida</i> KT2440	PPU
<i>Pseudomonas syringae</i> pv. <i>tomato</i> str. DC3000			PST	
<i>Pseudomonas fluorescens</i> Pf-5			PFL	
<i>Pseudomonas mendocina</i> ymp			PMY	

$\gamma$		<i>Pseudomonas stutzeri</i> A1501	PSA	
		<i>Azotobacter vinelandii</i> AvOP	AVN	
	<b><i>Alteromonadales</i></b>	<i>Pseudoalteromonas atlantica</i> T6c	PAT	
		<i>Alteromonas macleodii</i> 'Deep ecotype'	AMC	
		<i>Glaciecola</i> sp. HTCC2999 (*)	GHT	
		<i>Colwellia psychrerythraea</i> 34H	CPS	
		<i>Idiomarina baltica</i> OS145	ILO	
		<i>Marinobacter</i> sp. ELB17 (*)	MEL	
		<i>Marinobacter aqueolei</i>	MAQ	
		<i>Moritella</i> sp. PE36 (*)	MPE	
		<i>Psychromonas ingrahamii</i> 37	PIN	
		<i>Shewanella</i> sp. ANA-3	SHN	
		<i>Shewanella</i> sp. MR-4	SHE	
		<i>Shewanella</i> sp. MR-7	SHM	
		<i>Shewanella amazonensis</i> SB2B	SAZ	
		<b><i>Enterobacteriales</i></b>	<i>Escherichia coli</i> str. K-12 substr. MG1655	ECO
			<i>Escherichia coli</i> str. C ATCC 8739	ECL
			<i>Salmonella typhimurium</i> LT2	STM
			<i>Citrobacter koseri</i> ATCC BAA-895	CKO
			<i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i> MGH 78578	KPN
			<i>Enterobacter</i> sp. 638	ENT
			<i>Yersinia pestis</i> KIM	YPK
			<i>Serratia proteamaculans</i> 568	SSZ
			<i>Erwinia carotovora</i> subsp. <i>atroseptica</i> SCRI1043	PCT
			<i>Edwardsiella tarda</i> EIB202	ETR
			<i>Proteus mirabilis</i> HI4320	PMR
			<i>Photobacterium luminescens</i> subsp. <i>laumondii</i> TTO1	PLU
		<b><i>Oceanospirillales</i></b>	<i>Hahella chejuensis</i> KCTC 2396	HCH
			<i>Chromohalobacter salexigens</i> DSM 3043	CSA
		<b><i>Aeromonadales</i></b>	<i>Aeromonas hydrophila</i> ATCC 7966	AHA
			<i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i> A449	ASA
			<i>Tolumonas auensis</i> DSM 9187	TAU

$\gamma$	<b>Vibrionales</b>	<i>Vibrio cholerae</i> O1 biovar eltor str. N16961	VCH
		<i>Vibrio vulnificus</i> CMCP6	VVU
		<i>Vibrio harveyi</i> ATCC BAA-1116	VHA
		<i>Vibrio parahaemolyticus</i> RIMD 2210633	VPA
		<i>Vibrio shilonii</i> AK1 (*)	VSH
		<i>Vibrio splendidus</i> LGP32	VSP
		<i>Vibrio fischeri</i> ES114	VFI
		<i>Vibrio salmonicida</i> LFI1238	VSA
		<i>Vibrio angustum</i> S14 (*)	VAN
		<i>Photobacterium profundum</i> SS9	PPR
	<b>Pasteurellales</b>	<i>Haemophilus parasuis</i> SH0165	HAP
	<b>Xanthomonadales</b>	<i>Stenotrophomonas maltophilia</i> K279a	SML
	<b>Без классификации</b>	<i>Reinekea</i> sp. MED297 (*)	RMD
	$\beta$	<b>Burkholderiales</b>	<i>Burkholderia pseudomallei</i> K96243
<i>Burkholderia mallei</i> ATCC 23344			BMA
<i>Burkholderia</i> sp. 383			BUR
<i>Burkholderia cepacia</i> AMMD			BAM
<i>Burkholderia vietnamiensis</i> G4			BVI
<i>Burkholderia glumae</i> BGR1			BGL
<i>Burkholderia xenovorans</i> LB400			BXE
<i>Burkholderia phymatum</i> STM815			BPH
<i>Burkholderia cenocepacia</i> J2315			BCJ
<i>Ralstonia solanacearum</i> GMI1000			RSO
<i>Ralstonia pickettii</i> 12J			RPI
<i>Ralstonia metallidurans</i> CH34			RME
<i>Ralstonia eutropha</i> JMP134			REU
<i>Ralstonia eutropha</i> H16			REH
<i>Cupriavidus taiwanensis</i>			CTI
<i>Acidovorax avenae</i> subsp. <i>citrulli</i> AAC00-1			AAV
<i>Acidovorax</i> sp. JS42			AJS
<i>Comamonas testosteroni</i> KF-1			CTT
<i>Delftia acidovorans</i> SPH-1			DAC
<i>Polaromonas naphthalenivorans</i> CJ2			PNA

			<i>Polaromonas</i> sp. JS666	POL
			<i>Rhodoferax ferrireducens</i> DSM 15236	RFR
			<i>Variovorax paradoxus</i> S110	VAP
			<i>Verminephrobacter eiseniae</i> EF01-2	VEI
			<i>Methylibium petroleiphilum</i> PM1	MPT
			<i>Leptothrix cholodnii</i> SP-6	LCH
	<i>Neisseriales</i>	<i>Chromobacterium violaceum</i> ATCC 12472	CVI	
		<i>Neisseria meningitidis</i> MC58	NME	
	$\alpha$	<i>Caulobacteriales</i>	<i>Caulobacter</i> sp. K31	CAK
	<i>Thermotogae</i>	<i>Thermotogales</i>	<i>Thermotoga maritima</i> MSB8	TMA
<i>Thermotoga</i> sp. RQ2			TRQ	
<i>Thermotoga neapolitana</i> DSM 4359			TNA	
<i>Thermotoga petrophila</i> RKU-1			TPT	
<i>Thermotoga naphthophila</i> RKU-10			TNP	
<i>Thermotoga lettingae</i> TMO			TLE	

(\*) – неполный геном

### 2.3 Программное обеспечение

Для поиска ортологов использовался пакет программ Genome Explorer (228). В качестве критерия ортологичности здесь выступает наибольшее сходство белковых последовательностей из двух геномов при двустороннем поиске. Для функциональной аннотации с помощью оценки гомологии с экспериментально охарактеризованными белками использовался поиск с помощью алгоритма BLAST (2) по базе данных «UniprotKB/Swiss-Prot» (68). Для анализа геномного контекста использовались интернет-ресурсы MicrobesOnline (91) и SEED (90). Для определения специфичности транспортеров и гликозил гидролаз использовались базы данных TCDB (229) и CAZy (230), соответственно. Белковые домены были определены с помощью инструментов, включенных в базу данных Pfam (71). Сигнальные последовательности белков, служащие маркером секретиремости, определены с помощью программы SignalP (231). Множественные выравнивания, как

белковых, так и нуклеотидных последовательностей выполнены в программе MUSCLE (232). Филогенетические деревья были построены с помощью метода максимального правдоподобия, реализованного в программе proml из пакета PHYLIP (233). Визуализация филогенетических деревьев сделана в программе Dendroscope (234). Построение матриц позиционных весов осуществлялось в программе SignalX (228). Для предсказания потенциальных сайтов связывания регуляторов и использования метода проверки соответствия использовались инструменты веб-сервера RegPredict (<http://regpredict.lbl.gov/regpredict/>) (60). Диаграммы LOGO были построены с помощью программы WebLogo (235).

Список потенциальных и известных факторов транскрипции для *B. subtilis* был получен путем анализа соответствующих коллекций из баз данных DBD (236), MiST2 (237), DBTBS (238) и предыдущей работы по реконструкции набора регуляторов в *B. subtilis* (26).

Все реконструированные в этой работе регулоны доступны на веб-сервере RegPrecise (<http://regprecise.lbl.gov/RegPrecise>) (239).

## Глава 3

### Исследование эволюции AraR регулона

#### 3.1 Исследование эволюции регуляторной системы AraR

Фактор транскрипции AraR выделяется среди других хорошо изученных регуляторов во многом благодаря своей химерной доменной структуре, поскольку ДНК-связывающий домен белка AraR принадлежит GntR семейству, а субстрат-связывающий домен относится к LacI семейству. С целью выбора геномов, в которых целесообразно изучение AraR регулонов, был проведен поиск гомологов белка AraR из *B. subtilis* в базе данных “nr” ресурса NCBI. Для повышения точности определения гомологов поиск проводился по отдельности для ДНК-связывающего и субстрат-связывающего доменов, а затем были выбраны белки, попавшие в обе выборки. Гомологичные последовательности белка AraR были найдены в бактериях, относящихся к четырем порядкам: *Bacillales*, *Lactobacillales*, *Clostridiales* и *Thermotogales*. Из данных таксономических групп были отобраны 28 геномов для последующей реконструкции AraR регулонов.

Анализ филогенетического дерева ортологов AraR (Рис. 3.1) показал, что хорошо выраженные группы образуют белки из организмов, принадлежащих к порядкам *Thermotogales* и *Lactobacillales*. Белки из организмов порядка *Bacillales* разбились на дереве на три группы. Также интересно, что ортолог AraR из *Clostridium acetobutylicum* лежит отдельно от белков из других организмов порядка *Clostridiales*. В каждом из геномов обнаружилось по одному гену *araR*, однако в геноме *Bacillus licheniformis* данный регуляторный ген присутствует в двух копиях.

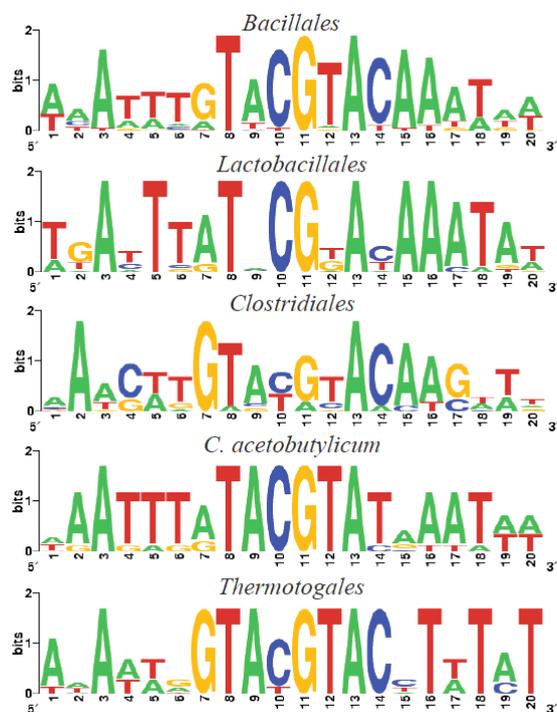


следующей итерации. Если же поиск с помощью новой матрицы повышал количество ложноположительных предсказаний, не подтверждаемых методами проверки соответствия или функциональной причастностью потенциальных регулируемых генов к пути утилизации арабинозы, то либо сохранялась старая матрица, либо новые геномы относились к другой группе и для них строилась отдельная матрица. Для повышения специфичности поиска в некоторых случаях были использованы регуляторные последовательности из дополнительных геномов, не вошедших в список исследуемых. Данные последовательности были найдены с помощью анализа геномного контекста ортологов AraR.

В результате геномы были разделены на пять групп: I) **Bacillales** (*B. subtilis*, *B. amyloliquefaciens*, *B. pumilus*, *B. licheniformis*, *A. flavithermus*, *G. kaustophilus*, *B. halodurans*, *O. iheyensis*, *Paenibacillus* sp. JDR-2), II) **Lactobacillales** (*L. brevis*, *L. fermentum*, *L. plantarum*, *L. reuteri*, *L. sakei*, *L. citreum*, *L. mesenteroides*, *O. oeni*, *P. pentosaceus*), III) **Clostridiales** (*C. cellulolyticum*, *C. beijerincki*, *Clostridium* sp. SS2/1), IV) **C. acetobutylicum**, V) **Thermotogales** (*T. maritima*, *Thermotoga* sp. RQ2, *T. neapolitana*, *T. petrophila*, *T. naphthophila*, *T. lettingae*). Все найденные мотивы сайтов связывания образуют нестрогие палиндромы длиной 20пн (Рис. 3.2).

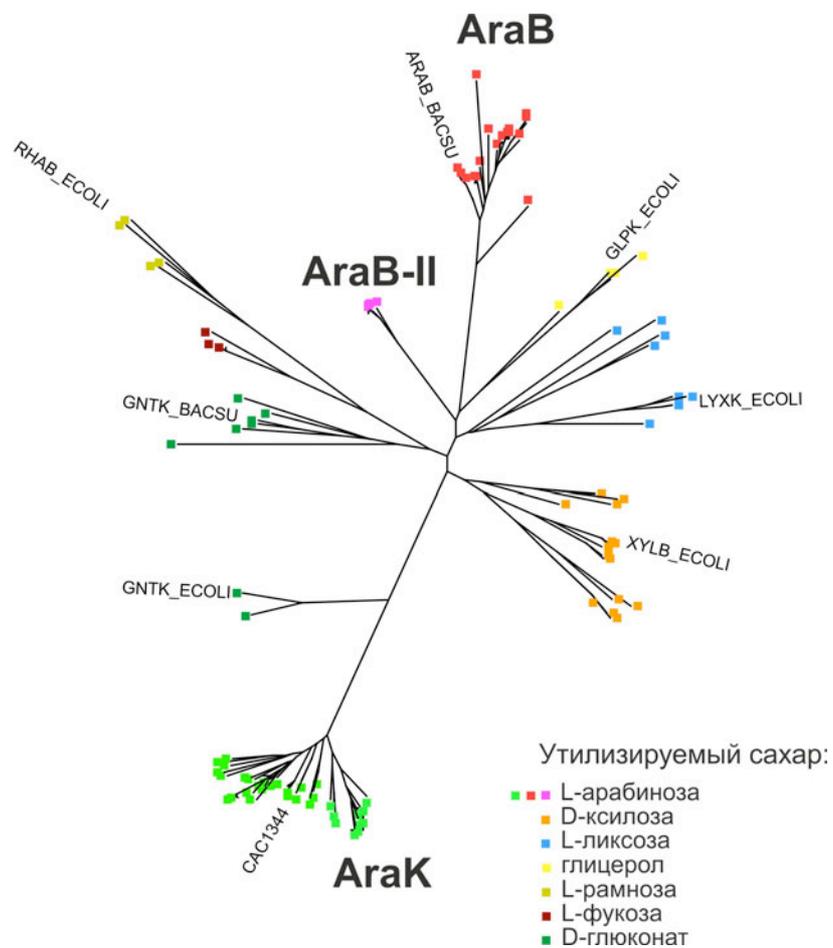
Интересно, что сайт с высоким весом 5'-ТСАТТТТТАСГТАСААТТАТ-3' находится перед геном *iolT* в *B. subtilis*, кодирующим транспортер инозитола. Однако, экспериментально было показано, что AraR не влияет на экспрессию *iolT in vivo* и не связывается с найденным сайтом *in vitro* (100). Поэтому данный сайт был искусственно исключен из AraR регулона.

Для каждой группы геномов был выбран порог немного меньше, чем минимальный вес сайта в обучающей выборке. Для всех групп, кроме *C. acetobutylicum*, был выбран порог 5,0. Для *C. acetobutylicum* использовался порог 5,1. При этом, веб сервер RegPredict позволяет просматривать сайты с порогом на 10% меньше установленного, что позволяет предсказывать слабые сайты при применении метода проверки соответствия.



**Рисунок 3.2.** Диаграммы LOGO, построенные по найденным сайтам связывания AraR для каждой из групп геномов.

По горизонтальной оси отложена позиция нуклеотида. По вертикальной – информационное содержание позиции в битах. Относительная высота каждой буквы соответствует частоте встречаемости нуклеотида в данной позиции.



**Рисунок 3.3** Филогенетическое дерево FGYY семейства киназ. Разными цветами показаны пути утилизации различных сахаров. Наиболее изученные белки отмечены идентификаторами из базы данных UniProt. (68)

**Таблица 3.1.** Наиболее консервативные члены AraR регулона. Случаи регуляции ортолога отмечены «+». Ортолог присутствует в геноме, но не регулируется – «-». Ортолог не найден в геноме – «0».

	BSU	BAY	BVD	BRU	GKA	AFL	BNA	OH	PD	CBE	CCE	CSS	CAC	LBR	LFE	LPL	LRF	LSA	LCI	LME	OOE	PPE	TMA	TRQ	TNA	TPT	TNP	TLE
<i>araR</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>araA</i>	+	+	+	+	+	+	+	+	+	+	0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0
<i>araA-//</i>	0	0	0	0	0	0	0	0	0	0	+	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+
<i>araB</i>	+	+	+	0	+	+	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>araB-//</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+	+	+	+	+
<i>araK</i>	0	0	+	+	0	0	0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0	0	0	0	0	0
<i>araD</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>araE</i>	+	+	+	+	0	0	0	+	-	-	0	0	+	+	+	+	+	+	+	+	+	+	0	0	0	0	0	0
<i>ABC*</i>	+	+	+	0	+	-	+	-	+	+	+	0	0	0	0	0	0	0	0	0	0	0	+	+	+	+	+	+
<i>abf*</i>	+	+	+	-	0	+	+	-	+	+	+	+	0	0	0	0	0	0	0	0	0	0	+	+	+	+	+	+

\* - в данных строках показано принципиальное наличие ABC транспортеров и гидролаз полимеров арабинозы, соответственно, вне зависимости от их ортологичности.

### 3.3 Структура AraR регулона в изучаемых геномах

В результате анализа регулируемых генов оказалось, что структура контролируемых AraR оперонов сильно различается в пределах даже одной группы (Рис. 3.4, Приложение 1, Табл. 3.1). Исключение составляет группа *Lactobacillales*, где во всех анализируемых геномах регулируется оперон из четырех генов *araA*, *araE*, *araK* и *araD*, составляющих минимальный путь для утилизации арабинозы.

Наиболее консервативной оказалась регуляция гена *araD*, кодирующего фермент L-рибулоза-5-фосфат-4-эпимеразу, продуктом которого является ксилулоза-5-фосфат – интермедиат пентозофосфатного пути. Ген *araD* присутствует и регулируется AraR во всех изучаемых геномах.

Для двух ферментов необходимых для утилизации арабинозы – L-арабиноза изомеразы и L-рибулокиназы – были найдены неортологичные замещения генов. Так, в геномах *C. cellulolyticum* и *T. lettingae* отсутствует ортолог AraA. При этом в опероне с генами ферментов утилизации арабинозы обнаруживается ген из семейства фукоза изомераз, названный *araA-II*. Этому гену была назначена функция альтернативной L-арабиноза изомеразы. Также интересно, что оба гена *araA* и *araA-II* находятся в одном опероне в *Clostridium* sp. SS2/1.

В геномах, где отсутствовали ортологи AraB, были обнаружены два ортологичных ряда киназ. Кроме уже предсказанной в работе Родионова Д.А. L-рибулокиназы AraK (8), в группе *Thermotogales* в опероне с *ara* генами была найдена новая киназа, названная *araB-II*. Все три киназы принадлежат FGGY семейству. Филогенетический анализ семейства FGGY (Рис. 3.3) показал крайнюю разнородность субстратов киназ из этого семейства и то, что ветви отвечающие AraB, AraB-II и AraK находятся на большом филогенетическом расстоянии друг от друга.

В геноме *B. licheniformis* были найдены сайты перед двумя оперонами утилизации арабинозы. Оба оперона содержат полный набор генов необходимых для утилизации арабинозы. Но несмотря на это в геноме

присутствует всего одна копия гена регулятора *araR*. При этом, если гены *araD* и *araA* дублированы и присутствуют в обоих оперонах, то один оперон содержит гены транспорта арабинозидов *araNPQ*, арабинофуранозидазу и L-рибулокиназу *araB*, а другой – ген транспортера мономеров L-арабинозы *araE* и L-рибулокиназу *araK*.

Среди регулируемых генов было замечено большое разнообразие гидролаз арабинозидов и транспортеров сахаров. Так, кроме предсказанного ABC транспортера арабинозидов в *B. subtilis* транспортера AraNPQ (99), в *S. beijerinckii* и *G. kaustophilus* под регуляцией был найден ABC транспортер AraFGH, имеющий сходство с транспортером ксилозы XylFGH из *E. coli* (240). Предполагается, что данный транспортер переносит мономеры арабинозы. В геномах группы *Thermotogales* под регуляцией AraR находится ABC транспортер AraN-II/AraP-II/AraQ-II, имеющий отдаленную гомологию с AraNPQ из *B. subtilis*. Также в этих геномах был найден гомологичный транспортер AraN-III/AraP-III/AraQ-III. Однако предсказать, переносят эти транспортеры арабинозиды или арабинозу, крайне затруднительно. В геноме *S. acetobutylicum* был обнаружен сайт перед опероном, содержащим транспортер *araT* из семейства сахар-протон симпортеров.

В группе *Clostridiales* был найден ABC транспортер AraT1-4. При этом в *S. cellulolyticum* AraR сайтов перед опероном найдено не было. Интересно, что в каждом случае в одном опероне с генами *araT1T2T3T4* были обнаружены гены двухкомпонентной системы – регулятора транскрипции AraI и сенсорной киназы AraJ. Анализ геномного контекста ортологов данной системы в других геномах показал кластеризацию ее генов с генами пути утилизации арабинозы в *Treponema saccharophilum* DSM 2985 и *Bryantella formatexigens* DSM 14469. При этом в указанных геномах не обнаружено ни ортологов AraR, ни иных потенциальных регуляторов генов пути утилизации арабинозы в окрестностях оперона. Это указывает, что двухкомпонентная система AraI/AraJ является дополнительным регулятором транскрипции генов утилизации арабинозы.

В AraR регулонах было обнаружено одиннадцать ортологичных рядов гликозил гидролаз. Из них альфа-арабинофуранозидаза Arb43 в *C. acetobutylicum* и альфа-глюкуронидаза BH1061 в *B. halodurans* уже были предсказаны методами сравнительной геномики (8) а Xsa, AbnA и AbfA из *B. subtilis* были проверены экспериментально (97,104). Большинство новых гликозил гидролаз, для генов которых предполагается регуляция AraR – Abf3, Abf4, Abf5 и Arb43 относятся к семействам альфа-арабинофуранозидаз. Также в оперонах с генами утилизации арабинозы в геномах группы *Thermotogales* были найдены эндо-1,4-бета-ксиланаза XynB, гидролаза из семейства ксилозидаз GH39 – XylX и гликозил гидролаза неизвестной функции GlsA. В *Thermotoga* sp. RQ2 и *T. petrophila* были найдены под AraR регуляцией идентичные опероны, состоящие из генов двух паралогичных гликозил гидролаз *abf4* семейства GH43, к которому принадлежит большинство альфа-арабинозидаз, а также два гена из семейства Ламинин G (Laminin G) и два гена из семейства Конканавалин A (Concanavalin A). Оба семейства содержат ферменты, способные связываться с углеводами, и некоторые из представителей семейств являются гликозил гидролазами. Похоже, что данные опероны кодируют системы деградации арабиноза-содержащих полисахаридов до мономеров или олигомеров.

Кроме того, в *C. acetobutylicum* и *C. beijerinckii* под регуляцией AraR были найдены гены пентозофосфатного пути – трансальдолазы *tad* и транскетолазы *tal*. В *C. acetobutylicum* сайт с высоким весом был обнаружен перед геном *ptk*, кодирующим фосфокетолазу, которая осуществляет реакцию расщепления D-ксилозы-5-фосфата до D-глицеральдегида-3-фосфата и ацетил-фосфата. Наконец, в большинстве геномов из порядков *Clostridiales* и *Thermotogales* под регуляцией обнаружен ген альдоза-1-эпимеразы *epiA*. Данный фермент, скорее всего, позволяет производить обратимую реакцию эпимеризации бета-арабинозы в альфа-арабинозу.



**Рисунок 3.4.** Оперонная структура AraR регулонов. Стрелками обозначены гены. Цветами обозначены одинаковые функции генов. Круги обозначают предполагаемые сайты связывания AraR. Цвет круга обозначает матрицу позиционных весов, с помощью которой был найден сайт: красный – *Lactobacillales*; желтый – *C. acetobutylicum*; оранжевый – *Bacillales*; зеленый – *Clostridiales*; синий – *Thermotogales*

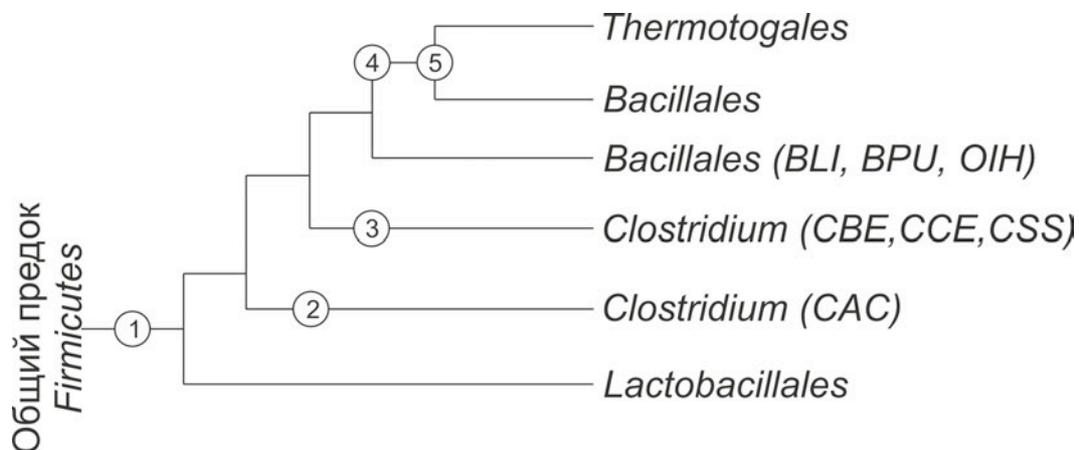
### 3.4 Эволюция AraR регулона

В ходе реконструкции AraR регулона было выявлено несколько особенностей:

- Наиболее часто под AraR регулирующей встречаются гены *araA*, *araK*, *araD* и *araE*, составляющие минимальный путь утилизации L-арабинозы (Табл. 3.1).
- Именно в этом составе регулон сохраняется в порядке *Lactobacillales*, а также в геномах *B. licheniformis*, *B. pumilus* и *O. iheyensis*.
- Структура оперонов и регулонов сильно различается даже внутри одного таксона.
- Единственным таксоном вне типа *Firmicutes*, где был найден ортолог AraM, является порядок *Thermotogales*.

Таким образом, из анализа реконструированных регулонов и филогенетических деревьев можно сделать предположение о возможном пути эволюции AraR регулона (Рис. 3.5). Вероятно, что первоначально AraR регуляция возникла в общем предке *Firmicutes* в виде простого регулона, состоящего из генов *araA*, *araB*, *araD* и *araE*, в каком виде он остался в *Lactobacillales*. Похоже, что в *C. acetobutylicum* произошел горизонтальный перенос одного из ранних вариантов регулона, где претерпел существенное расширение, захватив также гены пентозофосфатного пути *tkt*, *tal* и *ptk*. Об этом можно судить, во-первых, по относительно недалекому расположению AraR(CAC) от группы белков AraR *Lactobacillales* на филогенетическом дереве (Рис. 3.1), во-вторых, по структуре сайтов связывания. В отличие от других групп, у *Lactobacillales* и *C. acetobutylicum* обнаружено существенное снижение консервативности симметричных G и C в 7 и 14 позициях в сайтах связывания (Рис. 3.2). Интересно, что в результате мутационного анализа в *B. subtilis* оказалось, что эти позиции имеют важное значение для правильного связывания белка AraR (100), хотя и не являются критическими. Также в *C. acetobutylicum* в AraR регулоне содержатся по две копии генов *araE* и *araA*.

Причем, если уровень сходства паралогов AraA (80,5% идентичности) позволяет говорить о дупликации, то сходство паралогов AraE (35,3%) скорее указывает на горизонтальный перенос с последующей мутацией сайта.



**Рисунок 3.5.** Предполагаемый сценарий эволюции AraR регулона. Цифрами отмечены эволюционные события: 1 – регуляция одного оперона, состоящего из генов *araA*, *araK*, *araD* и *araE*; 2 – горизонтальный перенос и расширение регулона в предке *C. acetobutylicum*; 3 – транспорт арабинозы или олигоарабинозидов переходит под регуляцию двухкомпонентной системы AraI/AraJ; 4 – неортологичное замещение L-рибулокиназы AraK на AraB; 5 – горизонтальный перенос AraR регулона в бактерии порядка *Thermotogales*.

Анализ эволюции регулона указывает на то, что в общем предке *Bacillales* также присутствовал минимальный регулон в таком виде, как он сохранился в *B. licheniformis*, *B. pumilus* и *O. iheyensis*. Однако, затем произошла неортологичная замена AraK на AraB, и в состав регулона вошли системы деградации и транспорта поли- и олигоарабинозидов. Интересно, что в *B. licheniformis* под AraR регуляцией присутствуют два оперона, содержащие полный набор генов для деградации арабинозы, *araKDAE* и *araABDMNPQ-abfA*. Похоже, что второй оперон был горизонтально перенесен и сохранился для утилизации олигоарабинозидов.

В *Clostridiales* AraR регулон претерпел существенные оперонные перестройки и приобрел новый регулятор AraI/AraJ, который, судя по консервативности расположения на хромосоме, регулирует транспорт

арабинозы в клетку. Однако более точно это можно будет сказать лишь после реконструкции соответствующего регулона.

Наконец, в *Thermotogales* наблюдаются последствия возможного горизонтального переноса генов катаболизма арабинозы из *Bacillales*. Данный вывод сделан на основе анализа генов, входящих в регулон. Так, AraB-II на филогенетическом дереве стоит достаточно близко к AraB, чтобы предположить расхождение от общего предка. К тому же *Thermotogales* и *Bacillales* – это единственные таксоны, содержащие под AraR регуляцией ген *araM*.

### 3.5 Обсуждение

Регуляция транскрипции утилизации арабинозы фактором AraR была подробно изучена в грам-положительной бактерии *B. subtilis* (96,97,99,100). До настоящего времени изучение регуляции в других бактериях ограничивалось биоинформатическим анализом в четырех геномах: *B. halodurans*, *B. stearothermophilus*, *C. acetobutylicum* и *E. faecium* (8). В настоящей работе было показано, что AraR регулон широко распространен в бактериях, принадлежащих к типу *Firmicutes*, а также в бактериях порядка *Thermotogales*. Анализ сайтов связывания показал существенное изменение мотива распознавания в бактериях порядка *Lactobacillales* и *C. acetobutylicum*. Интересно, что все белки, с доменной структурой, как у AraR из *B. subtilis*, охарактеризованные биоинформатическими методами или экспериментально, являются его ортологами и регулируют утилизацию арабинозы.

Сравнительный анализ регулонов утилизации арабинозы позволил предсказать функции новых генов в различных организмах. Компоненты пути сильно варьируются при переходе от одного организма к другому. Единственным ферментом пути, консервативным во всех изученных организмах, оказалась L-рибулоза-5-фосфат-4-эпимераза AraD. Для всех остальных компонентов пути были замечены вариации. Так функциональная аннотация L-рибулокиназы, кроме как для ортологов *araB*, изученной в *B. Subtilis*, была предсказана для генов *araK* и *araB-II*. Причем, если функция AraK в качестве L-рибулокиназы была предположена из предыдущего анализа AraR регулона в *C. acetobutylicum* (8), то в настоящей работе было показано, что *araK* является наиболее распространенным типом L-рибулокиназ. Функция AraK в качестве L-рибулокиназы была проверена в совместной работе с лабораторией доктора Янг из Шанхайского Института биологических наук Китайской академии наук. Для этого были осуществлены эксперименты по анализу роста *araK* мутантов бактерии *C. acetobutylicum*, а также *in vitro* анализ ферментативной активности белка AraK и определение субстратов и продуктов

катализируемой им реакции с использованием метода высокоэффективной жидкостной хроматографии. Полученные данные показали активность AraK в качестве рибулокиназы.

Также было найдено неортологичное замещение L-арабиноза изомеразы на AraA-II в геномах *C. cellulolyticum*, *C. sp. SS2/1* и *T. lettingae*. Было показано широкое разнообразие генов транспортных систем под регуляцией AraR – как однокомпонентных пермеаз (*araE* и *araT*), так и ABC транспортеров (*araNPQ*, *araFGH*, *araN-II-araP-II-araQ-II*, *araN-III-araP-III-araQ-III* и *araT1T2T3*). К сожалению, точно определить специфичность к арабинозе или арабинозидам не представляется возможным. Также были обнаружены новые гены гидролаз, находящихся под AraR регуляцией (*abf3*, *abf4*, *abf5*, *xylX*, *xynB*, *glsA*). В *C. acetobutylicum* и *C. beijerinckii* под регуляцией AraR были обнаружены гены пентозофосфатного пути, что, вероятно, позволяет быстрее метаболизировать арабинозу.

В результате анализа филогенетических деревьев и состава регулона был предложен вероятный сценарий эволюции AraR регулона. Предположительно, в общем предке *Firmicutes* возник регулон, состоящий из генов *araA*, *araK*, *araD* и *araE*, который затем расширялся в различных таксонах, получая гены для деградации различных арабинозосодержащих полимеров. Кроме того, в ходе эволюции регулона произошел ряд горизонтальных переносов, что было убедительно доказано в настоящей работе.

## Глава 4

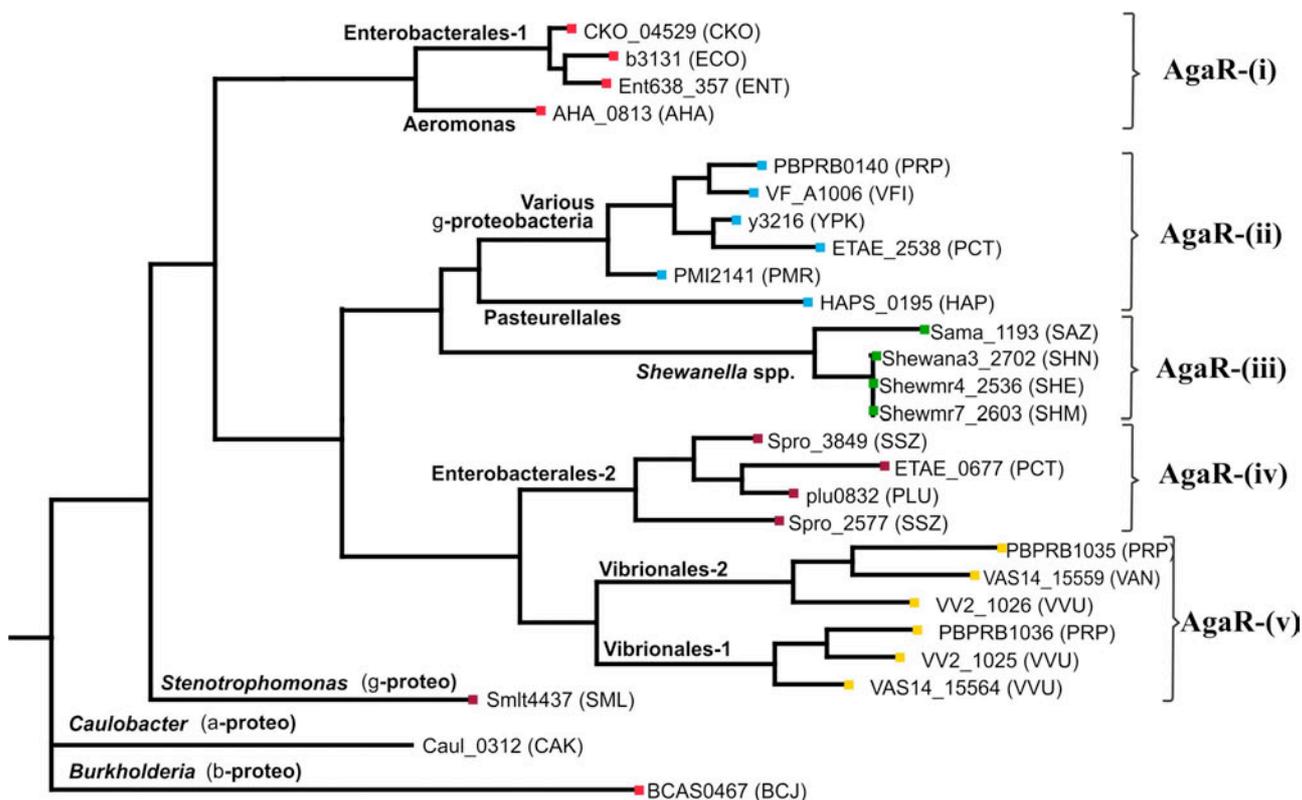
### Исследование регуляции AgaR регулона в протеобактериях

#### 4.1 Исследование эволюции регуляторной системы AgaR

Фактор транскрипции AgaR, принадлежащий к семейству ДНК-связывающих белков DeoR, был впервые охарактеризован в *E. coli* в качестве репрессора генов утилизации N-ацетилгалактозамина (НАГА) и галактозамина (ГА) (109). В протеобактериях ортологи данного белка были найдены в 21 геноме – 19 геномах гамма-протеобактерий, одном геноме бета-протеобактерий и одном геноме альфа-протеобактерий. В четырех геномах было обнаружено по две копии гена *agaR* (Табл. 4.1). Геном *P. profundum* содержит три паралога *agaR*. Все найденные ортологи гена *agaR* кластеризуются на хромосоме с генами утилизации НАГА, что доказывает консервативность функции ортологов AgaR (Рис. 4.2). Филогенетический анализ найденных в протеобактериях гомологов AgaR показал наличие пяти групп регуляторов (Рис. 4.1). Интересно, что геномы *P. profundum* и *E. tarda* содержат достаточно далекие паралоги AgaR (45% идентичности), тогда как AgaR паралоги в *S. proteamaculans* и *V. vulnificus* соответственно на 74 и 57% идентичны и принадлежат к одним и тем же кладам на дереве AgaR (Рис. 4.1).

#### 4.2 Построение распознающего правила для поиска потенциальных сайтов связывания AgaR

В каждой из пяти изучаемых таксономических групп бактерий были выделены 5'-некодирующие участки генов утилизации НАГА/ГА и использованы для проведения филогенетического футпринтинга (Приложение 2). По результатам анализа были получены консервативные фрагменты ДНК, которые использовались для построения матриц позиционных весов и последующего сканирования геномов с целью поиска дополнительных сайтов.



**Рисунок 4.1.** Филогенетическое дерево белков AgaR из протеобактерий. Пять групп регуляторов с различными типами мотивов сайтов связывания отмечены различными цветами.

Предсказанные мотивы сайтов связывания AgaR во всех пяти группах имеют сходную консенсусную последовательность СТТТС или же обратнo-комплементарную ей, GAAAG (Рис. 4.2). В группе (i), включающей в себя *E. coli* и родственные *Enterobacteriales*, предполагаемый мотив AgaR сайтов соответствует результатам определения защищенных от расщепления ДНКазой I участков (109) (Приложение 2). В группах (i), (ii) и (iii), предполагаемые мотивы AgaR сайтов имеют одинаковую структуру прямых повторов с консенсусом СТТТС-5пн-СТТТС, тогда как число таких сайтов и их ориентация для промоторной области каждого регулируемого гена может различаться. В группе (iv) предсказанный регуляторный мотив является инвертированным повтором с консенсусом СТТТС-15пн-GAAAG. Наконец, предсказанный мотив для группы (v) имеет структуру прямого повтора с консенсусом двух сайтов GAAAG, разделенных спейсером в 16-18пн.

## Геном

*Escherichia coli* C str. ATCC 8739

*Citrobacter koseri* ATCC BAA-895

*Enterobacter* sp. 638

*Serratia proteamaculans* 568

*Edwardsiella tarda* EIB202

*Yersinia pestis* KIM

*Proteus mirabilis* Hi4320

*Photorhabdus luminescens* TTO1

*Photobacterium profundum* Ss9

*Vibrio vulnificus* CMCP6

*Vibrio fischeri* Es114

*Vibrio angustum* S14

*Aeromonas hydrophila* ATCC 7966

*Shewanella amazonensis* SB2B

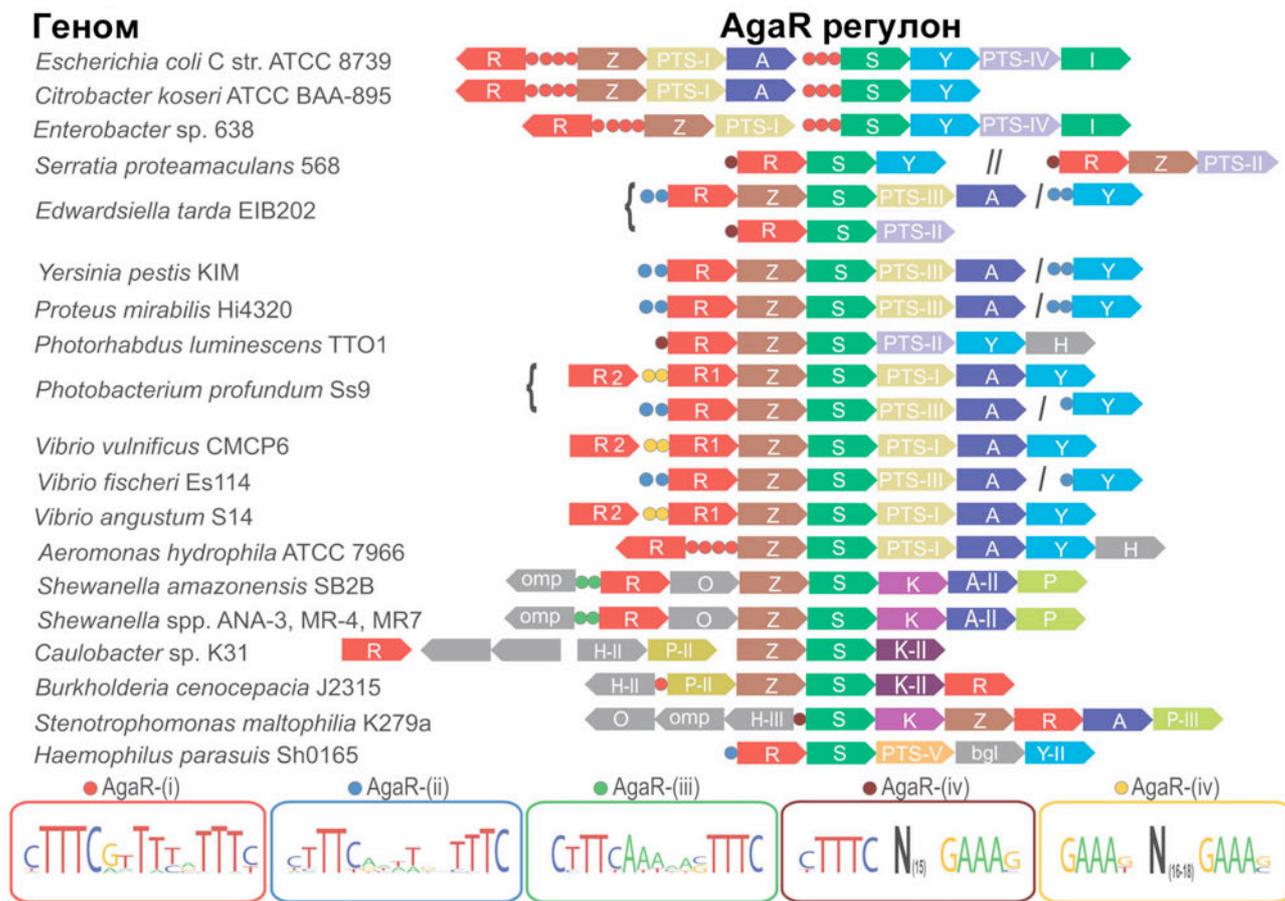
*Shewanella* spp. ANA-3, MR-4, MR7

*Caulobacter* sp. K31

*Burkholderia cenocepacia* J2315

*Stenotrophomonas maltophilia* K279a

*Haemophilus parasuis* Sh0165



**Рисунок 4.2.** Гены и потенциальные сайты связывания AgaR показаны соответственно стрелками и кругами. Гены названы по последней букве названия соответствующего белка. Гены, находящиеся в одном локусе, но не непосредственно друг за другом разделены знаком «/». Гены из разных локусов разделены знаком «//» или показаны в две строки, объединенные скобкой. Цвет потенциальных сайтов связывания AgaR соответствует одному из мотивов внизу рисунка.

### 4.3 Структура AgaR регулона и путей утилизации N-ацетилгалактозамина и галактозамина

Реконструкция AgaR регулонов в протееобактериях показала различные наборы генов, предположительно участвующих в утилизации НАГА/ГА (Табл. 4.1 и Рис 4.2). Большая доля регулируемых AgaR генов кодирует ранее неизвестные ферменты и транспортные системы. Путем анализа сходства

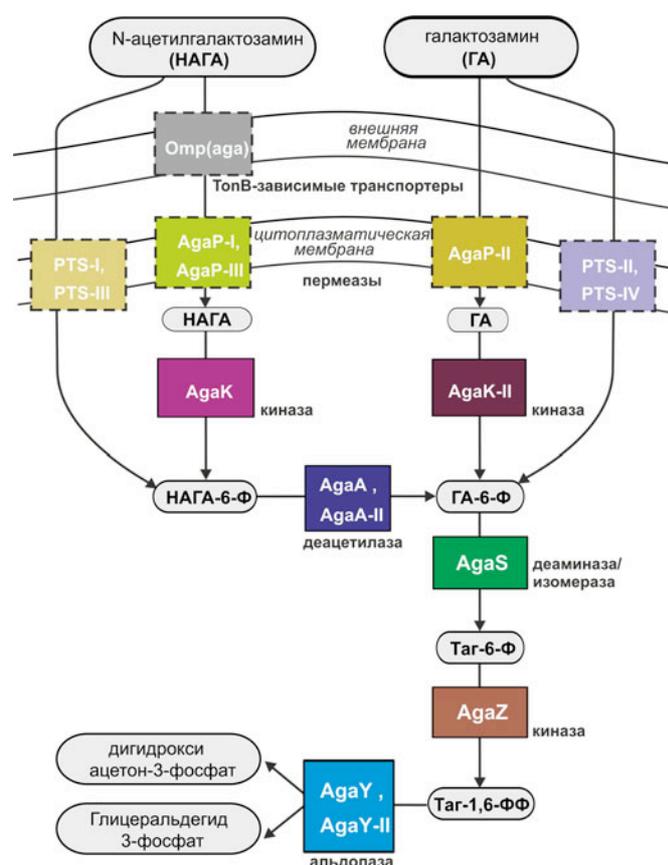
белков и геномного контекста соответствующих генов им были приписаны предполагаемые функциональные аннотации, и была проведена реконструкция метаболических путей (Рис. 4.3).

**Таблица 4.1.** Распределение компонент пути утилизации GalNAc в протеобактериях.

Таксон / Геном	Регулятор	Ферменты катаболизма <sup>2</sup>					Транспортеры <sup>2</sup>		
	AgaR <sup>1</sup>	AgaK	AgaA	AgaS	AgaZ	AgaY	Omp	AgaP	PTS
<b>Enterobacteriales</b>									
<i>Escherichia coli</i> str. C str. ATCC 8739	i	-	I	+	+	I	-	-	I,IV
<i>Citrobacter koseri</i> ATCC BAA-895	i	-	I	+	+	I	-	-	I
<i>Enterobacter</i> sp. 638	i	-	-	+	+	I	-	-	I,IV
<i>Yersinia pestis</i> KIM	ii	-	I	+	+	I	-	-	III
<i>Serratia proteamaculans</i> 568	iv,iv	-	-	+	+	I	-	-	II
<i>Edwardsiella tarda</i> EIB202	ii,iv	-	I	+	+	I	-	-	II,III
<i>Proteus mirabilis</i> HI4320	ii	-	I	+	+	I	-	-	III
<i>Photobacterium luminescens</i> TTO1	iv	-	-	+	+	I	-	-	II
<b>Vibrionales</b>									
<i>Vibrio vulnificus</i> CMCP6	v,v	-	I	+	+	I	-	-	I
<i>Vibrio fischeri</i> ES114	ii	-	I	+	+	I	-	-	III
<i>Vibrio angustum</i> S14	v,v	-	I	+	+	I	-	-	I
<i>Photobacterium profundum</i> SS9	ii,v,v	-	I	+	+	I	-	-	I,III
<b>Pasteurellales</b>									
<i>Haemophilus parasuis</i> SH0165	ii	-	-	+	-	II	-	-	V
<b>Alteromonadales</b>									
<i>Shewanella</i> sp. ANA-3, MR-4, MR-7	iii	I	II	+	+	-	+	I	-
<i>Shewanella amazonensis</i> SB2B	iii	I	II	+	+	-	+	I	-
<b>Aeromonadales</b>									
<i>Aeromonas hydrophila</i> ATCC 7966	i	-	I	+	+	I	+	-	I
<b>Xanthomonadales</b>									
<i>Stenotrophomonas maltophilia</i> K279a	iv	I	I	+	+	-	+	III	-
<b>Caulobacteriales</b>									
<i>Caulobacter</i> sp. K31	+	II	-	+	+	-	-	II	-
<b>Burkholderia</b>									
<i>Burkholderia cenocepacia</i> J2315	+	II	-	+	+	-	-	II	-

<sup>1</sup> Группы регуляторов образующих пять клад на филогенетическом древе и характеризующихся разными мотивами сайтов связывания показаны в виде римских цифр (Рис. 4.1).

<sup>2</sup> Присутствие ортолога гена, кодирующего фермент на соответствующую функциональную роль отмечено «+» или римской цифрой, которая указывает какой из вариантов неортологичного замещения присутствует в данном геноме. Отсутствие ортолога обозначено знаком «-».



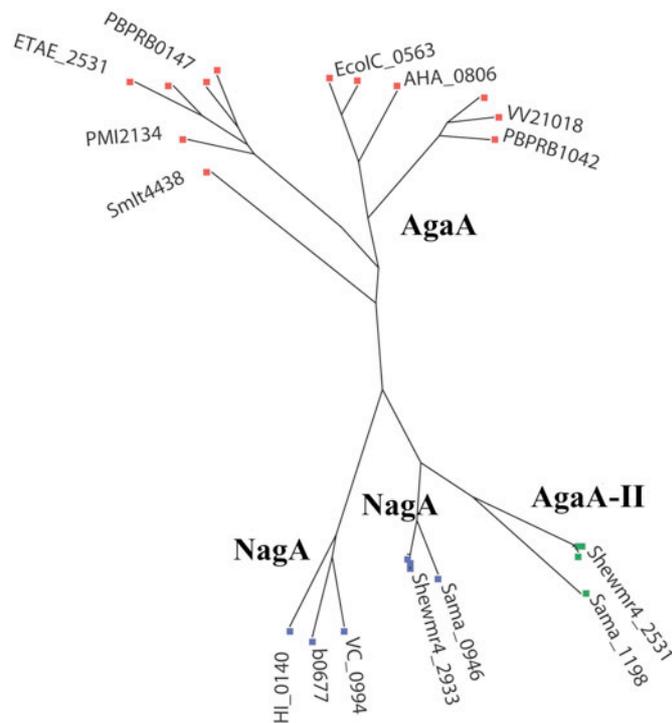
**Рисунок 4.3.** Реконструкция путей утилизации НАГА и ГА в протеобактериях. Метаболиты показаны в серых овалах. Белки показаны в прямоугольниках. Неортологичные замещения белков показаны в одном прямоугольнике под различными римскими цифрами.

Наиболее консервативным белком в пути утилизации НАГА/ГА, присутствующим во всех изучаемых организмах, является AgaS, сахарная изомераза неизвестной функции из SIS семейства. Изначально было предположено, что функцию деаминазы/изомеразы, способной преобразовывать ГА-6-фосфат в тагатоза-6-фосфат в *E. coli* выполняет AgaI. Однако, в текущей работе ортолог *agaI* был найден только в *Enterobacter* sp. 638, что показывает вспомогательную роль AgaI в исследуемом метаболическом пути. Таким образом, необходимая в пути ГА-6-фосфат деаминаза/изомераза отсутствует в большинстве геномов. Данный факт

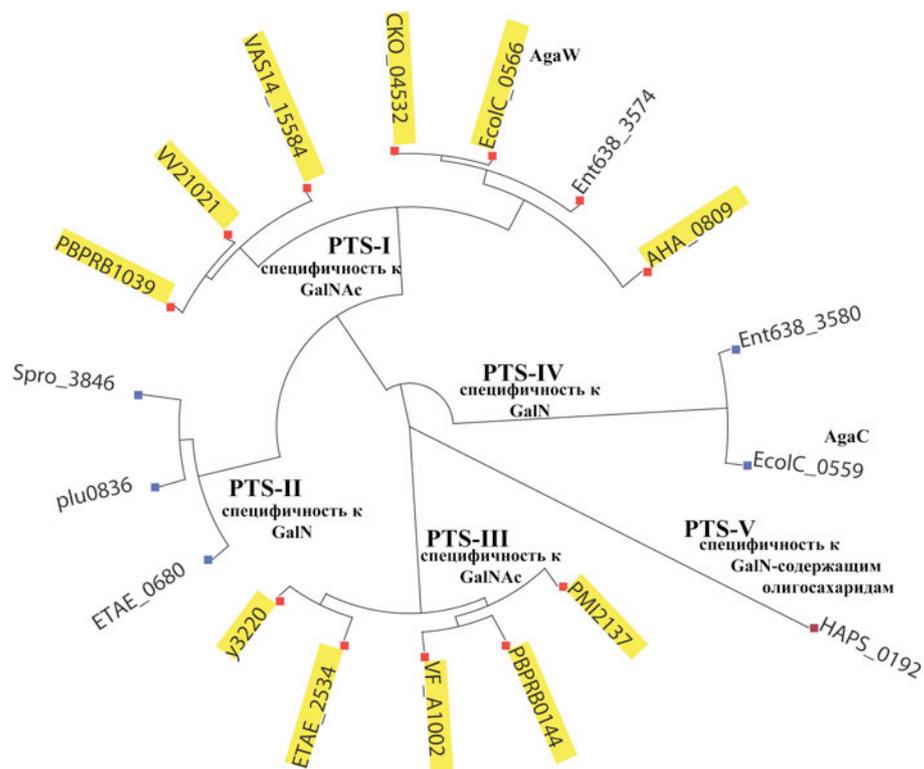
позволяет предположить, что именно AgaS выполняет роль основной ГА-6-фосфат деаминазы/изомеразы.

Ген НАГА-6-фосфат деацетилазы *agaA* присутствует в двенадцати геномах из двадцати одного и обязательно находится на хромосоме в одном кластере с остальными генами утилизации НАГА. В организмах рода *Shewanella* кластер *aga* генов содержит ген, кодирующий НАГА-6-фосфат деацетилазу, имеющую наибольшее сходство с N-ацетилглюкозамин-6-фосфат деацетилазой NagA из бактерий рода *Shewanella* (50% идентичности). Этот новый вариант НАГА-6-фосфат деацетилазы был назван AgaA-II. Все три фермента: AgaA из *E. coli*, AgaA-II и NagA из *Shewanella* принадлежат к одному семейству амидогидролаз (COG1820). Филогенетический анализ этого семейства (Рис. 4.4) подтвердил, что AgaA-II является паралогом NagA, что свидетельствует о недавней дупликации генов. Интересно, что НАГА-6-фосфат деацетилазы отсутствуют в шести исследованных протеобактериях, что показывает невозможность данных микроорганизмов утилизировать НАГА, а, следовательно, найденный в них путь является утилизацией исключительно ГА.

Транспорт и последующее фосфорилирование НАГА и ГА в *E. coli* осуществляется с помощью двух специфичных PTS систем, кодируемых в генах *agaBCD* и *agaVWEF* из AgaR регулона (106). Гены, кодирующие гомологичные PTS системы, были обнаружены под регуляцией AgaR в бактериях порядков *Enterobacteriales* и *Vibrionales*. Для определения специфичностей этих PTS систем было построено филогенетическое дерево для ПС компонент PTS, являющихся мембранными белками и определяющими специфичность транспортных систем (Рис. 4.5). На дереве хорошо различаются клады, соответствующие генам, лежащим в одном локусе с геном *agaA* и лежащим на хромосоме отдельно от последнего. Можно предположить, что связанные с *agaA* PTS системы специфичны к НАГА (PTS-I и PTS-III), тогда как PTS системы, лежащие отдельно от *agaA*, специфичны к ГА (PTS-II и PTS-IV).



**Рисунок 4.4.** Филогенетическое дерево GalNAc-6- фосфат деацетилаз и N-ацетилглюкозамин-6-фосфат деацетилаз.



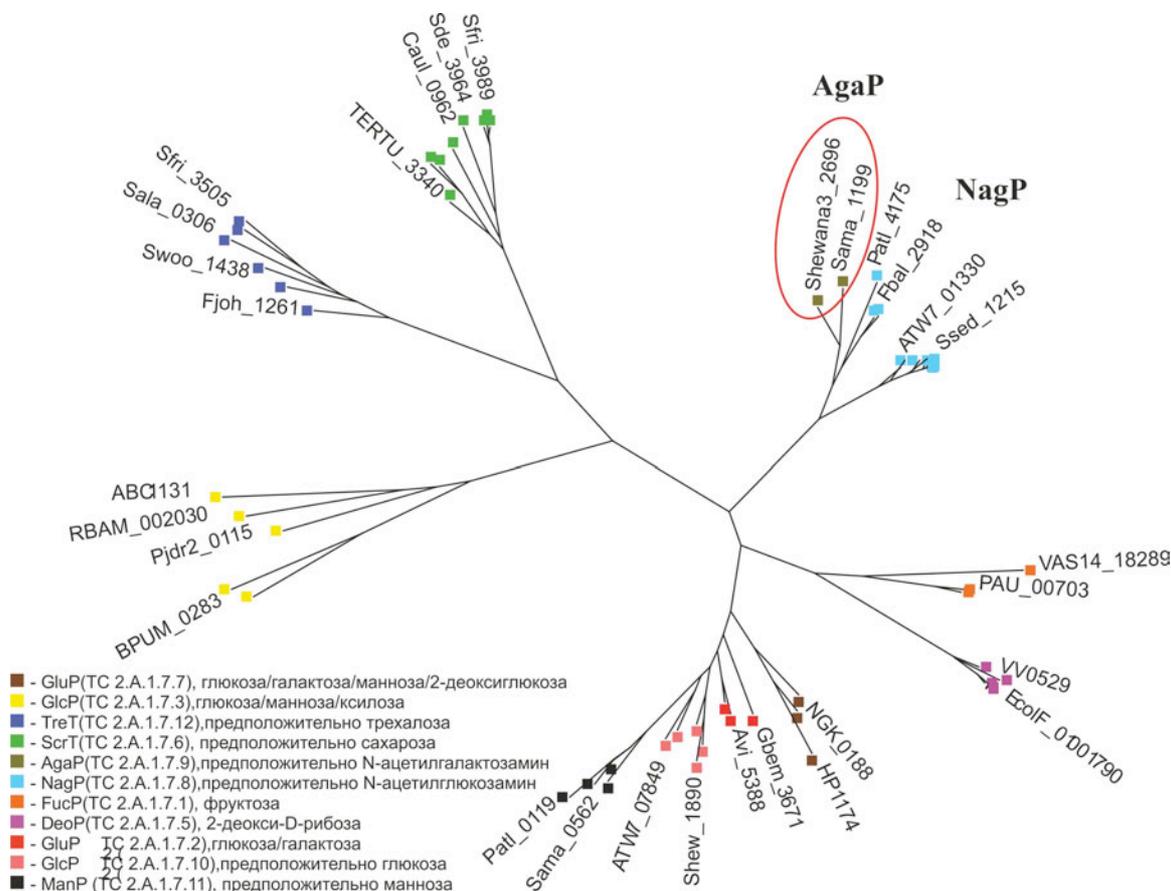
**Рисунок 4.5.** Филогенетическое дерево ПС компонент PTS систем, обнаруженных в AgaR регулонах в протеобактериях. PTS компоненты, гены которых лежат в одном локусе с генами НАГА-6-фосфат деацетилаз *agaA*, отмечены желтым цветом. Пять клад PTS систем помечены римскими цифрами.

Под регуляцией AgaR не обнаружилось PTS систем в бактериях рода *Shewanella* и в некоторых других протейобактериях. В *Shewanella* AgaR регулоны содержат новые гены, кодирующие предсказанные НАГА пермеазу AgaP, киназу AgaK и TonB-зависимый транспортер внешней мембраны Omp<sup>(aga)</sup>, который предположительно участвует в транспорте НАГА сквозь внешнюю мембрану. Предсказанная пермеаза AgaP принадлежит к GGP семейству сахарных транспортеров и является близким паралогом N-ацетилглюкозамин пермеазы из бактерий рода *Shewanella* (Рис. 4.6). Предсказанная НАГА киназа AgaK принадлежит к семейству киназ ROK и гомологична глюкокиназе Glk-II (35% идентичности) из бактерий рода *Shewanella* (110). В AgaR регулоне в *S. maltophila* также был найден ген *agaK*, однако закодированная в этом же регулоне пермеаза AgaP-III принадлежит к другому семейству, называемому Sugar\_tr.

В *Caulobacter* sp. K31 и *B. senoserpacia* в *aga* кластере закодированы сахарная киназа AgaK-II из семейства VcrAD\_BadFG и транспортер AgaP-II из семейства EamA. Поскольку в данных геномах не было обнаружено НАГА-6-фосфат деацетилазы, было предположено, что AgaP-II и AgaK-II участвуют в транспорте и фосфорилировании ГА, соответственно. Таким образом, можно предположить, что отсутствие НАГА и ГА-специфичных PTS систем в этих геномах замещено обнаруженными пермеазами и киназами.

В реконструированных AgaR регулонах были обнаружены новые гликозил гидролазы, которые, как предполагается, участвуют в метаболизме НАГА/ГА (Таблица 4.1). Так, в группе геномов *Shewanella* была найдена секретлируемая альфа-N-ацетилгалактозаминидаза AgaO, принадлежащая к семейству GH109 гликозил гидролаз. Геномы *P. luminescens* и *A. hydrophila* содержат ген гидролазы AgaH из семейства GH36, которая скорее всего также является альфа-N-ацетилгалактозаминидазой. Также ген, кодирующий секретлируемую ГА-олигосахарид гидролазу AgaH-II, был найден в *aga* кластере в *Caulobacter* sp. K31 и *B. senoserpacia*, а ген, кодирующий НАГА-олигосахарид

гидролазу AgaH-III – в *S. maltophila*. В *H. parasuis* под AgaR регуляцией находится ген, кодирующий цитоплазматическую бета-галактозидазу *bgaZ*.



**Рисунок 4.6.** Филогенетическое дерево транспортеров семейства

Глюкоза/галактоза пермеаз (GGP). Субстратные специфичности описаны в соответствии с базой данных TCDB (229).

#### 4.4 Эволюция AgaR регулона

Филогенетический анализ белков, входящих в путь утилизации НАГА, позволяет предположить наиболее вероятные эволюционные сценарии появления AgaR регулонов в различных таксонах.

Наиболее интересным представляется образование оперона утилизации НАГА в бактериях рода *Shewanella*. В этом опероне присутствуют гены, кодирующие новые компоненты пути – AgaP и AgaA-II, которые образовались, скорее всего, путем дупликации и последующего приобретения новых функций соответствующих генов из пути утилизации N-ацетилглюкозамина. В то же

время остальные компоненты – AgaR, AgaZ и AgaS были идентифицированы во всех остальных протеобактериях. Таким образом, AgaR регулон в *Shewanella* состоит как из универсальных, так и таксон-специфичных генов. Пермеаза AgaP и деацетилаза AgaA-II, скорее всего, образовались путем дупликации генов, кодирующих N-ацетилглюкозамин пермеазу NagP и N-ацетилглюкозамин деацетилазу NagA, исходя из филогенетического анализа соответствующих семейств белков (Рис. 4.4 и Рис. 4.6).

Уникальный вариант пути утилизации галактозамина был обнаружен в *H. parasuis*. Оперон *agaRS-PTS-V-bgaZ-agaY-II* также кодирует две группы белков с различным эволюционным происхождением. К первой группе относятся белки AgaR и AgaS, наиболее похожие на аналогичные белки из *Enterobacteriales*. Ко второй группе относятся компоненты транспортера PTS-V и цитоплазматическая бета-галактозидаза BgaZ, для которых наиболее близкие гомологи находятся в бактериях типа *Firmicutes*, например, кластер генов SP\_0061-64 в *Streptococcus pneumoniae*. Также в опероне был обнаружен ген *agaY-II*, кодирующий тагатоза-1,6-бисфосфат альдолазу из LacD семейства. Охарактеризованные белки данного семейства в основной массе участвуют в деградации галактозы-6-фосфат в грамположительных бактериях (241). Данные обстоятельства позволяют предположить, что часть AgaR регулона была перенесена горизонтальным переносом из *Firmicutes*.

На филогенетических деревьях различных белков обнаруживается парафилетическая группа из белков, принадлежащих бактериям двух таксонов – *Enterobacteriales* и *Vibrionales*. Сам кластер *aga* генов сохраняет свою структуру в этих видах. Из этого можно предположить, что в *Y. pestis*, *E. tarda*, *P. mirabilis*, *V. fisheri* и *P. profundum* локус генов утилизации НАГА был привнесен в результате недавних горизонтальных переносов из одного источника.

Также интересно, что в *S. proteamaculans* присутствуют два *aga* кластера. В обоих кластерах обнаружен паралог *agaR* гена, но при этом остальные гены не дублированы. Скорее всего, это является результатом дупликации

изначального *aga* кластера в предковом организме с последующей потерей дублирующихся генов.

#### 4.5 Обсуждение

Во всех изученных протеобактериях гены пути утилизации НАГА находятся в одном кластере с генами, кодирующими ортологи репрессора AgaR из *E. coli*. С помощью филогенетического анализа было обнаружено, что регуляторы образуют пять групп с различной структурой мотивов сайтов связывания. Тем не менее, все мотивы основаны на нескольких копиях последовательности СТТТС, которые встречаются в качестве прямых и инвертированных повторов (Рис. 4.2). Можно предположить, что данная последовательность является основным сайтом, который распознает мономер AgaR. Интересно, что найденные сайты связывания AgaR часто встречаются в двух или более экземплярах в промоторных областях *aga* оперонов. Это позволяет утверждать, что основным механизмом репрессии является образование петель ДНК с помощью комплекса нескольких субъединиц AgaR.

Анализ геномного контекста в дополнение к реконструкции AgaR регулонов позволил обнаружить новые гены, относящиеся к путям утилизации НАГА в различных протеобактериях (Табл. 4.1). Наиболее вариабельной частью реконструированных путей оказались т транспортные системы и ферменты, осуществляющие первые шаги пути – превращение субстрата в ГА-6-фосфат путем фосфорилирования и деацетилирования (Рис. 4.3). Последующие стадии преобразования ГА-6-фосфат в интермедиаты гликолиза консервативны практически во всех изученных бактериях.

Протеобактерии используют две основные стратегии для транспорта в цитоплазму и последующего фосфорилирования НАГА и ГА: PTS системы и комбинация сахар-специфичных пермеаз и киназ. PTS системы, схожие с ГА- и НАГА-специфичными PTS системами в *E. coli*, были обнаружены в таксонах *Enterobacteriales* и *Vibrionales*, а также в *H. parasuis*. В других таксонах эти системы замещают ГА- и НАГА-специфичные киназы и пермеазы. Данные

системы часто лежат в одном локусе с генами, кодирующими TonB-зависимые транспортеры внешней мембраны Omp<sup>(aga)</sup>.

Наиболее консервативным членом AgaR регулона является ген, кодирующий изомеразу AgaS. Также было показано, что белок AgaI, которому приписывали функцию ГА-6-фосфат деаминазы/изомеразы встречается всего в двух организмах и, тем самым, не может быть основным ферментом, выполняющим данную роль. Это позволило предположить, что именно AgaS является ГА-6-фосфат деаминазой/изомеразой. Данное предположение было экспериментально проверено на примере бактерии *Shewanella* sp. ANA-3 в совместной работе с лабораторией доктора Янг из Шанхайского института биологических наук Китайской академии наук. Была измерена ферментативная активность трех ферментов пути утилизации НАГА – AgaK, AgaA-II и AgaS, а также *in vitro* был реконструирован метаболический путь, состоящий из этих трех ферментов. Результаты показали верность предсказанных функций. Интересно, также, что фермент AgaK показал низкую активность на N-ацетилглюкозамине, при этом не обнаружив активности на иных субстратах, таких как ГА, глюкоза, галактозамин или N-ацетилманнозамин. Также низкую активность на N-ацетилглюкозамин-6-фосфате проявил и фермент AgaA-II. И хотя данные ферментативные активности крайне низки в сравнении с реакциях на основных субстратах, это дает основание считать предположение о происхождении данных ферментов путем дубликации соответствующих генов из пути утилизации N-ацетилглюкозамина правдоподобным.

Гены, кодирующие две последних стадии пути утилизации НАГА, не были обнаружены в некоторых анализируемых геномах. Предсказанная тагатоза-6-фосфат киназа AgaZ присутствует в большинстве изучаемых геномов, в то время как тагатоза-1,6-бисфосфат альдолаза AgaY обнаружена лишь в бактериях порядков *Enterobacteriales*, *Vibrionales* и *A. hydrophila*. Предыдущие работы сообщали о возможной роли AgaZ в качестве некаталитической субъединицы тагатоза-1,6-бисфосфат альдолазы (108). Однако, профили распределения генов *agaZ* и *agaY* не поддерживают эту

гипотезу, что позволяет предположить, что AgaZ имеет функцию независимую от AgaY. И хотя AgaZ не принадлежит ни к одному известному семейству с охарактеризованной функцией, единственной ролью в пути, для которой нет соответствующего фермента, является тагатоза -6-фосфат киназа. Отсутствующая в некоторых бактериях тагатоза-1,6-бисфосфат альдолаза AgaY может быть заменена альдолазой из другого пути. Так, для бактерий рода *Shewanella* было обнаружено, что гомология фруктоза-бисфосфат альдолазы Fba гораздо ближе к AgaY, чем к Fba из *E. coli* (50% и 35% сходства, соответственно).

Наконец, в работе был проанализирован сценарий эволюции AgaR регулона и было показано, что основными событиями в эволюционной истории этого регулона являются горизонтальный перенос и дубликации генов.

## Глава 5

### HexR – регулятор центрального метаболизма углеводов

#### 5.1 Исследование эволюции регуляторной системы HexR

Фактор транскрипции HexR детально исследован в бактериях рода *Pseudomonas*, где он регулирует транскрипцию генов пути Энтнера-Дудорова (117). Для поиска ортологов был выбран ген *hexR* из *P. putida*. Его ортологи были обнаружены в геномах 11 таксонов, принадлежащих к классам гамма- и бетапротееобактерий: *Pseudomonadaceae*, *Alteromonadales*, *Oceanospirillales*, *Vibrionales*, *Psychromonadaceae*, *Aeromonadales*, *Shewanellaceae*, *Enterobacteriales*, *Comamonadaceae*, *Burkholderia*, *Neisseriales* и *Ralstonia*. Среди гаммапротееобактерий, однако, *hexR* не обнаружен в таких крупных таксонах, как *Pasteurellales*, *Xanthomonadales* и *Moraxellaceae*.

Интересно, что в бактериях порядка *Pseudomonadales* были найдены два далеких паралога *hexR*. Одна группа паралогов содержит экспериментально изученный ген PP1021, который похож на остальные *hexR* гены из гаммапротееобактерий. Второй же паралог *hexR1* более близок к *hexR* из бетапротееобактерий. Бактерии рода *Burkholderia* также содержат две копии гена *hexR*, при этом одна из копий регулятора кодируется химерным геном *hexR2-glk*, образованным в результате слияния с геном глюкокиназы *glk*. Поскольку в бактериях рода *Burkholderia* не было обнаружено других гомологов *glk*, то химерный ген *hexR2-glk* – единственный кандидат на необходимую для метаболизма функцию глюкокиназы, в то время как роль HexR2 домена в качестве регулятора транскрипции не совсем ясна.

Анализ филогенетического дерева белков HexR выявил 13 групп (Рис. 5.1) и показал соответствие между филогенией ортологов HexR и таксономией исследуемых бактерий.

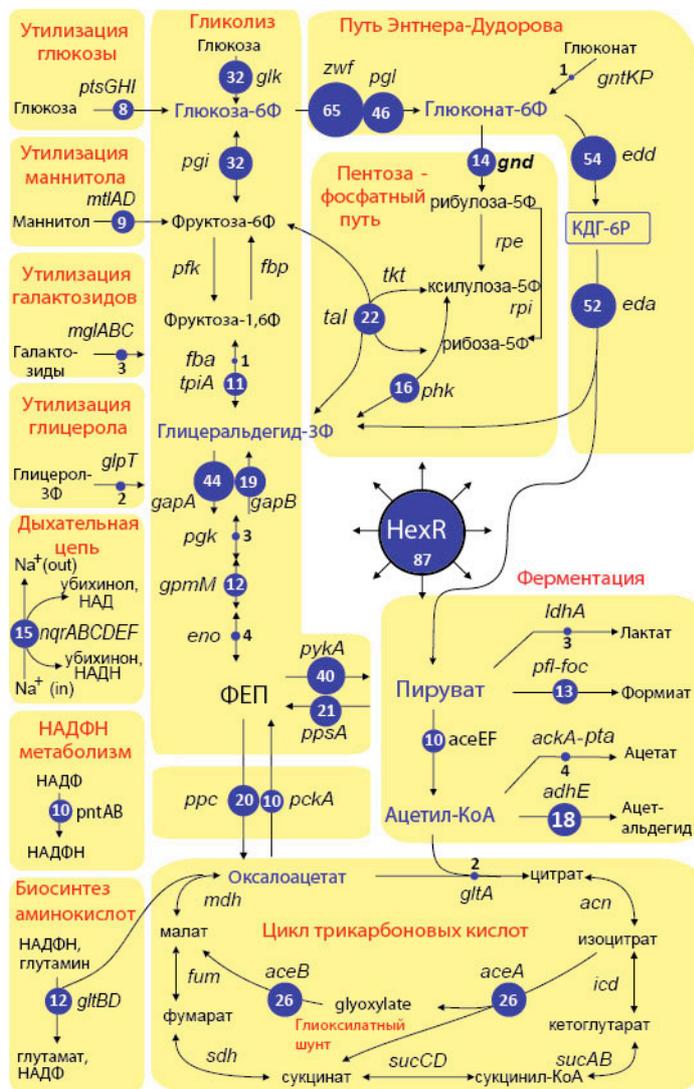


## 5.2 Построение распознающего правила для поиска потенциальных сайтов связывания HexR

Для построения профилей поиска потенциальных сайтов связывания в каждой из групп геномов, содержащих ортолог *hexR*, была набрана обучающая выборка из промоторных областей генов, предположительно входящих в HexR регулон. Эти гены были определены с помощью анализа хромосомной кластеризации *hexR* с ортологами генов из пути Энтнера-Дудорова, входящих в HexR регулон в *P. putida*. Затем этот список пополнялся по мере расширения HexR регулона новыми ортологичными рядами. По этим обучающим выборкам строилась матрица позиционных весов, которая впоследствии итеративно улучшалась, как это описано в главе Материалы и Методы. В итоге было найдено 13 различных мотивов сайтов связывания для групп: *Pseudomonadaceae* - HexR, *Pseudomonadaceae* - HexR1, *Shewanellaceae*, *Enterobacteriales*, *Oceanospirillales/Alteromonadales*, *Alteromonadales*, *Psychromonadaceae/Aeromonadales*, *Hahella/Marinobacter*, *Comamonadaceae*, *Burkholderia*, *Neisseria/Chromobacterium*, *Ralstonia* (Рис. 5.1).

Наиболее консервативный мотив, предсказанный как мотив сайтов связывания HexR в девяти изученных группах, имеет консенсус TGRAR-5-YTACA, где R – A или G, Y – C или T. Два различных мотива было предсказано для каждой из групп паралога HexR в группах бактерий семейства *Pseudomonadaceae*. Найденный путем биоинформатического анализа мотив сайта связывания HexR соответствовал экспериментально изученному мотиву для *Pseudomonadaceae* с консенсусом TGTTGT-4-8пн-ACAACAT. В то же время мотив сайтов связывания для группы *Pseudomonadaceae* HexR1 похож на мотивы сайтов связывания HexR остальных протеобактерий. Подобная разница в мотивах позволяет предположить отсутствие пересечения HexR и HexR1 регулонов. В бактериях родов *Hahella* и *Marinobacter* был обнаружен укороченный мотив с консенсусом GWAGTATACTWC, где W – A или T, который тем не менее несет признаки мотивов, свойственных HexR.

Состав реконструированных регулонов собран в приложении 3. Интересно, что самой консервативной регуляцией обладают гены центрального метаболизма углеводов, тогда как таксон-специфические регуляторные взаимодействия широко варьируются в каждой группе бактерий (Рис. 5.2).



**Рисунок 5.2.** Метаболический контекст реконструированных HexR регулонов в протеобактериях. Цифры в синих кружках обозначают количество регулонов, в которых присутствует данный ген.

### 5.3 Ядро HexR регулона

Основываясь на общем количестве случаев регуляции и распределению по таксонам, регулируемые гены можно разбить на несколько групп (Табл. 5.1). В первой группе генов присутствуют наиболее консервативные члены регулона, которые регулируются более чем в тридцати организмах из по

меньшей мере шести различных таксонов. В эту группу входят гены *zwf*, *pgl*, *edd*, *eda*, *gapA*, *pykA*, *glk* и *pgi*, которые кодируют ферменты гликолиза и пути Энтнера-Дудорова. К тому же в большинстве геномов была предсказана авторегуляция самого гена *hexR*.

Ко второй группе относятся гены, принадлежащие HexR регулону по крайней мере в десяти организмах из хотя бы двух различных таксонов. Эти гены принадлежат к таким метаболическим путям, как гликолиз (*gpmM* и *tpiA*), глюконеогенез (*ppsA*, *gapB* и *pckA*), пентозофосфатный путь (*tal*), метаболизм пирувата (*aceEF* и *ppc*), ферментация (*adhE*, *pflBA* и *grcA*), глиоксилатный шунт (*aceBA*), биосинтез аминокислот (*gltBD*) и окисление NADPH (*pntAB*).

К третьей группе принадлежат гены, которые были найдены в HexR регулоне в двух или более таксонах, но менее, чем в десяти организмах. Сюда входят гены, включенные в такие метаболические пути, как гликолиз (*pgk* и *eno*), брожение (*aldE*), ферментация формиата (*focA*), а также утилизации глюкозы и маннитола (*ptsG*, *ptsHI-crr* и *mtlADR*).

#### **5.4. Таксон-специфическая регуляция генов HexR регулона**

Оставшиеся потенциально регулируемые гены принадлежат к группе таксон-специфической регуляции (Табл. 5.2). В основном к этой группе относятся гены, перед которыми найдены потенциальные сайты связывания HexR по крайней мере в двух геномах внутри одного таксона, но не регулируемые вне этого таксона. Значительное увеличение размера HexR регулона было обнаружено в группе бактерий семейства *Shewanellaceae*, где в него вошли гены из центрального метаболизма (*gnd*, *phk* и *adhB*), утилизации нуклеозидов/дезоксинуклеозидов (*deoABD*, *nupC* и *cdd*), дыхательной цепи (*nqrA-N*) и утилизации глицина (*gcvTHP*). Второй большой HexR регулон был обнаружен в бактериях порядка *Vibrionales*. Он включает в себя гены метаболизма гликогена (*glgX* и *glgCA*), нитрит редуктазы (*nirBD*) и лактат пермеазу (*lctP*). В других организмах в HexR-регулон вошли гены, участвующие в утилизации галактозидов и глицерола (*mgl* и *glpT*), гликолизе

(*glpN*), цикле трикарбоновых кислот (*gltA*) и ферментации лактата и ацетата (*ldhA* и *ackA-pta*).

**Таблица 5.1.** Консервативное ядро HexR регулона в 11 группах протеобактерий.

Числа в таблице означают количество геномов, в которых в каждой из таксономических групп перед геном был обнаружен потенциальный сайт связывания HexR. Гены, перед которыми сайт связывания обнаружен не был, отмечены «-». Отсутствие ортолога в группе анализируемых геномов обозначено «н/о». Число анализируемых геномов в каждой группе отмечено в скобках рядом с названием таксона.

Регулируемые гены	Таксономические группы (число изученных геномов)											Функциональная роль
	Shewanellaceae (16)	Alteromonadales (5)	Oceanospirillales/ Alteromonadales (6)	Enterobacteriales (11)	Vibrionales (10)	Psychromonadaceae/ Aeromonadales (6)	Pseudomonadales (8)	Burkholderia spp. (8)	Ralstonia spp. (6)	Comamonadaceae (9)	Neisseriales (2)	
<i>hexR</i>	16	5	5	11	4	5	8	8	4	5	2	HexR регулятор, семейство RpiR
<i>zwf</i>	16	5	5	11	-	-	8	8	2	8	2	глюкозо-6-фосфат дегидрогеназа
<i>pgl</i>	16	5	4	1	н/о	-	8	8	2	н/о	2	6-фосфоглюконолактоназа
<i>edd</i>	16	5	5	1	-	-	7	8	6	4	2	6-фосфоглюконатдегидратаза
<i>eda</i>	16	5	5	3	1	-	8	8	-	4	2	2-дегидро-3-деоксиглюконат-6Ф-альдолаза
<i>glk</i>	н/о	4	5	-	н/о	4	7	8	2	-	2	глюкокиназа
<i>pgi</i>	16	1	1	-	7	-	-	-	-	5	2	глюкоза-6-Ф-изомераза
<i>pykA</i>	16	4	6	-	1	5	1	-	-	7	-	пируват киназа
<i>gapA</i>	16	5	5	-	8	4	6	-	-	-	-	глицеральдегид-3Ф-дегидрогеназа [НАД]
<i>tal</i>	16	1	-	-	-	2	-	-	-	3	-	трансальдолаза
<i>ppsA</i>	16	3	2	-	-	-	-	-	-	-	-	фосфоенолпируватсинтаза
<i>gapB</i>	16	1	2	н/о	-	-	-	н/о	н/о	н/о	н/о	глицеральдегид-3Ф-дегидрогеназа [НАДФ-Н]
<i>adhE</i>	15	-	1	-	-	2	н/о	н/о	-	н/о	-	алкогольдегидрогеназа
<i>aceBA</i>	15	н/о	1	-	8	2	-	-	-	-	-	малатсинтаза, изоцитратлиаза
<i>aceEF</i>	-	-	2	-	-	-	8	-	-	-	-	пируватдегидрогеназа
<i>ppc</i>	-	1	-	4	1-	5	-	-	-	-	-	фосфоенолпируваткарбоксилаза
<i>pckA</i>	-	1	2	4	3	-	-	н/о	н/о	н/о	н/о	фосфоенолпируваткарбокскиназа
<i>tpiA</i>	-	-	-	-	8	3	-	-	-	-	-	триозофосфатизомераза
<i>gpmM</i>	-	-	2	-	10	-	-	н/о	н/о	-	-	фосфоглицератмутата
<i>pflBA</i>	-	н/о	1	-	6	6	н/о	н/о	н/о	н/о	-	ферментация формиата
<i>grcA</i>	н/о	н/о	н/о	-	9	3	н/о	н/о	н/о	н/о	н/о	ферментация формиата
<i>gltBD</i>	-	-	-	-	9	3	-	-	-	-	-	глутаматсинтаза
<i>pntAB</i>	-	-	-	-	8	2	-	-	-	-	-	НАД(Ф)-трансгидрогеназа
<i>aldE</i>	-	3	1	-	-	2	1	-	н/о	-	-	альдоза-1-эпимераза
<i>eno</i>	-	-	1	-	-	2	1	-	-	-	-	енолаза
<i>pgk</i>	-	-	1	-	-	-	-	-	-	2	-	фосфоглицераткиназа
<i>ptsHI</i>	-	н/о	н/о	-	2	5	-	-	-	-	-	транспорт глюкозы (PTS система)
<i>mtlAD</i>	н/о	н/о	-	-	7	2	н/о	н/о	н/о	н/о	н/о	утилизация маннитола
<i>focA</i>	-	н/о	-	-	2	5	н/о	н/о	н/о	н/о	н/о	ферментация формиата
<i>ptsG</i>	-	н/о	-	-	7	-	н/о	н/о	н/о	н/о	1	утилизация глюкозы

**Таблица 5.2.** Таксон-специфические члены HexR регулона.

Таксономические группы (число изученных геномов)	Регулируемые гены	Число случаев регуляции	Функциональная роль
Shewanellaceae (16)	<i>phk</i>	16	фосфокетолаза
	<i>gnd</i>	14	6-фосфоглюконатдегидрогеназа
	<i>nqrABCDEF</i>	15	НАДН:убихиноноксидоредуктаза
	<i>deoABD</i>	14	утилизация деоксинуклеозидов
	<i>nupC</i>	15	транспорт нуклеозидов
	<i>cdd</i>	10	цитидиндеаминаза
	<i>gcvTHP</i>	12	расщепление глицина
	<i>adhB</i>	8	алкогольдегидрогеназа II
	<i>mcp</i>	9	хемотаксис
	<i>SO1118</i>	10	предполагаемый белок
Psychromonadaceae/	<i>mgIBAC</i>	3	утилизация галактозидов
Aeromonadales (6)	<i>glpT</i>	2	утилизация глицерола
	<i>ldhA</i>	3	лактатдегидрогеназа
	<i>ackA-pta</i>	4	ацетаткиназа
Oceanospirillales/	<i>gltA</i>	2	цитрат синтаза
Alteromonadales (6)	<i>mgsA</i>	2	метилглиоксальсинтаза
	<i>PF00248</i>	2	альдо/кеторедуктаза
Alteromonadales (5)	<i>fba</i>	1	фруктоза-1,6-бисфосфатальдолаза
	<i>glgP</i>	2	гликогенфосфорилаза
	<i>cpsA</i>	2	синтез капсульных полисахаридов
Enterobacteriales (11)	<i>ybfA</i>	10	предсказанный секретлируемый белок
Vibrionales (10)	<i>pepD</i>	9	аминоацил-гистидиндипептидаза
	<i>nirBD</i>	7	нитрит редуктаза [NAD(Ф)H]
	<i>glgX</i>	6	расщепление гликогена
	<i>glgCA</i>	5	гликогенсинтаза, аденилилтрансфераза
	<i>lctP</i>	2	L-лактатпермеаза
	<i>ygaW</i>	8	предсказанный мембранный белок
Pseudomonadales (8)	<i>gapN</i>	2	предсказанная глицеральдегид-3Ф-дегидрогеназа
	<i>gltRS</i>	7	регуляция транспорта глюкозы

## 5.5 Обсуждение

В настоящей работе впервые обнаружен фактор транскрипции HexR и реконструированы соответствующие регулоны в 62 геномах гаммапротеобактерий и 25 геномах бетепротеобактерий. Во всех изученных таксонах найдены высоко консервативные мотивы сайтов связывания регулятора. Вместе с тем, состав регулона значительно меняется в разных таксонах. Регулоны HexR могут быть как локальными и регулировать 1-2 оперона, как в *Enterobacteriales*, *Ralstonia* и *Burkholderia*, так и глобальным и насчитывать до 20 регулируемых оперонов, как в *Aeromonadales*, *Vibrionales* и *Shewanella*. Способность белка HexR связываться с предсказанными сайтами была экспериментально проверена для пятнадцати генов из *S. oneidensis* в лаборатории доктора Остермана из Института медицинских исследований Сэнфорд и Бернэма. Методами электрофоретического замедления ДНК в геле (Electrophoretic mobility shift assay) и флуоресцентной поляризации было показано связывание HexR со всеми предсказанными сайтами.

Филогенетический анализ совместно с реконструкцией HexR регулонов показал, что наиболее вероятным представляется расширение первоначального локального HexR регулона, контролирующего гены из пути Энтнера-Дудорова. Глобальные HexR регулоны были обнаружены только в гаммапротеобактериях. При этом, если рассматривать дерево белков HexR, видно, что белки, соответствующие глобальным регулонам, за исключением нескольких геномов *Oceanospirillales* и *Reinekea*, лежат в одной кладе. Единственная группа белков внутри этой клады, где HexR регулон является локальным, принадлежит бактериям порядка *Enterobacteriales*. Также состав глобальных регулонов позволяет предположить, что расширение происходило независимо по крайней мере три раза – в предках бактерий групп 1) *Alteromonadales*, включая *Shewanellaceae*, 2) *Aeromonadales*, *Psychromonadaceae* и *Vibrionales* и 3) *Oceanospirillales*.

Интересно, что функция HexR, как глобального регулятора метаболизма углерода во многом совпадает с ролью фактора транскрипции Cra (FruR) в *E. coli* (113). Так, если в *Enterobacteriales* сайты связывания HexR обнаружены всего перед двумя генами (*zwf* и *ybfA*), то Cra регулон насчитывает около четырех десятков оперонов по данным базы RegulonDB (242). В то же время, для *V. cholerae* HexR регулон насчитывает 15 оперонов, тогда как фактор транскрипции Cra является локальным и регулирует только *fruBKA* (Равчеев Д.А., личное сообщение). Эти наблюдения свидетельствуют о значительной пластичности сетей регуляции транскрипции центрального метаболизма углерода в протеобактериях.

## Глава 6

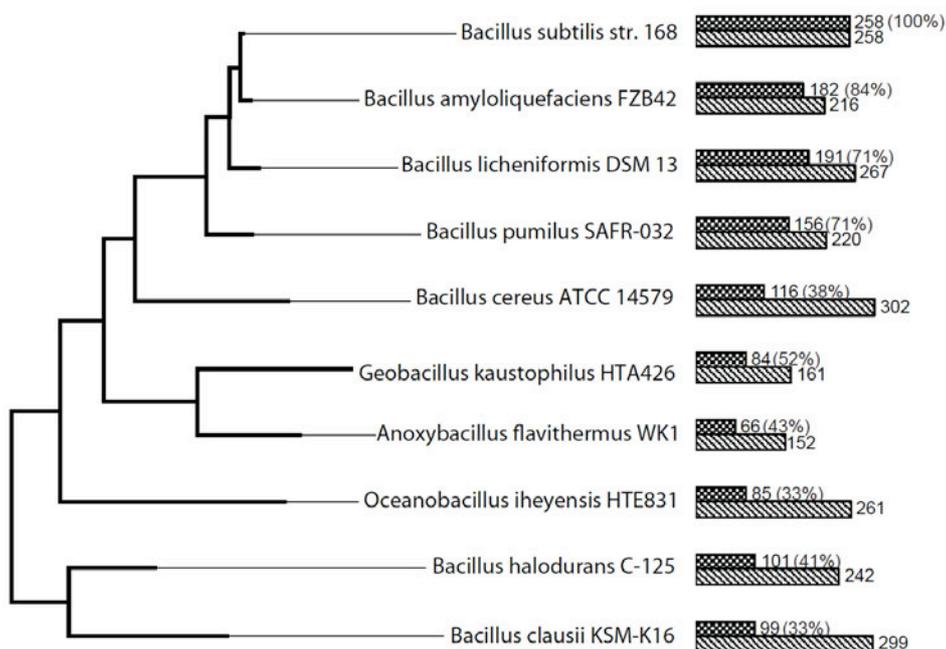
### Эволюция регуляции катаболизма сахаров в бактериях семейства *Bacillaceae*

#### 6.1 Поиск потенциальных регуляторов метаболизма углеводов среди ортологов факторов транскрипции *B. subtilis* в бактериях семейства *Bacillaceae*

Для наилучшего покрытия таксономической группы были отобраны десять геномов: *B. subtilis*, *B. amyloliquefaciens*, *B. pumilus*, *B. licheniformis*, *A. flavithermus*, *G. kaustophilus*, *B. cereus*, *B. halodurans*, *B. clausii* и *O. iheyensis*. Для изучения регуляции транскрипции необходимо было составить список факторов транскрипции. Было принято решение ограничиться теми регуляторами, чьи ортологи присутствуют в *B. subtilis*, как в модельном организме. Поэтому изначально список ДНК-связывающих факторов транскрипции был составлен для *B. subtilis*. Поиск ортологов регуляторов в остальных геномах и сравнение с общим числом регуляторов из баз данных DBD (236) и MiST2 (237) в остальных изучаемых геномах показал большую пластичность наборов факторов транскрипции даже среди близкородственных бактерий (Рис. 6.1).

Среди обнаруженных белков в *B. subtilis* было выделено 32 ранее изученных регулятора метаболизма углеводов и их производных. При этом, для 11 из них хотя и была показана регуляция генов, но не были обнаружены мотивы сайтов связывания (Табл. 6.1). Кроме того, для ранее не изученных факторов транскрипции был проведен анализ консервативности геномного контекста на предмет кластеризации гена регулятора с генами сахарных путей. И, наконец, для реконструкции были выбраны факторы транскрипции, принадлежащие к семействам, для которых известно много случаев регуляции сахарных регулонов (LacI, GntR и RpiR) и для которых не наблюдалось консервативности геномного окружения. Для них, вероятно, регулируемые гены находятся не в одном локусе с регулятором. Таким образом было выбрано

еще 11 регуляторов для дальнейшей реконструкции. Суммарно была предпринята попытка реконструировать 43 регулона.



**Рис. 6.1.** Филогенетическое дерево бактерий семейства *Bacillales*, построенное по конкатенату 16S-23S РНК. Для каждого генома представлено: ▨ - суммарное число факторов транскрипции, взятое из базы данных DBD (236), за исключением *A. flavithermus*, для которого число получено из базы данных MiST2 (237); ■ - число ортологов факторов транскрипции из *B. subtilis* и их процент от общего числа факторов транскрипции в геноме.

## 6.2 Построение распознающих правил для поиска потенциальных сайтов связывания регуляторов метаболизма углерода

Для реконструкции регулонов были использованы три подхода:

1) Данные о сайтах связывания ранее изученных факторов транскрипции были собраны в качестве обучающих выборок и на их основании построены матрицы позиционных весов, которые впоследствии были применены для поиска новых потенциально регулируемых генов и реконструкции регулонов в остальных изучаемых бактериях. Матрицы затем улучшались путем добавления в обучающие выборки сайтов, найденных в других геномах. Большинство факторов транскрипции было изучено в *B. subtilis*, кроме регулятора

утилизации рибозы RbsR, экспериментально охарактеризованного в *Corynebacterium glutamicum* (219) и регулятора утилизации маннитола MtlR, изученного ранее в *Bacillus stearothermophilus* (180). В рамках данного подхода был реконструирован 21 регулон.

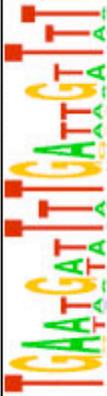
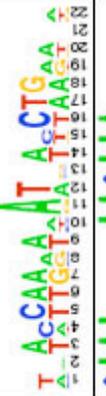
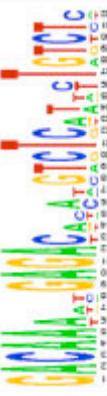
2) Для регулонов, где был показан только факт регуляции конкретных генов, методом филогенетического футпринтинга был произведен поиск потенциальных областей связывания регуляторов в 5'-некодирующих областях соответствующих регулируемых генов и их ортологов в других геномах, содержащих регулятор. Затем по полученным последовательностям была построена матрица позиционных весов и проведен поиск потенциальных сайтов. В дальнейшем было реконструировано 11 таких регулонов.

3) Регуляция факторами транскрипции, для которых не проводились экспериментальные исследования, изучалась путем анализа геномного контекста. В качестве потенциально регулируемых генов выбирались наиболее консервативные гены из геномного окружения гена регулятора. Также предполагалась возможность авторегуляции самого гена фактора транскрипции. В 5'-некодирующих областях генов производился поиск сайтов связывания методом филогенетического футпринтинга. На основании найденных консервативных областей ДНК строилась матрица позиционных весов. Для генов факторов транскрипции, где не прослеживалось консервативности кластеризации, производились попытки поиска авторегуляции. Данный подход был применен для 11 регулонов.

Для каждого ортологичного ряда исследуемых регуляторов оказалось возможным построить одну матрицу позиционных весов, что свидетельствует о достаточно высокой консервативности мотивов сайтов связывания внутри семейства *Bacillaceae* (Табл. 6.1).

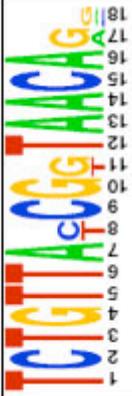
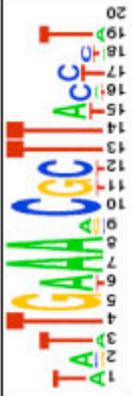
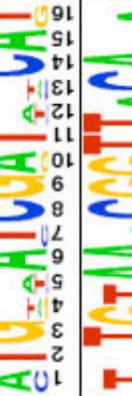
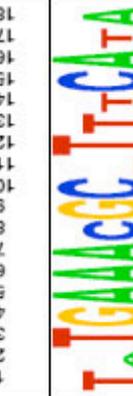
**Таблица 6.1.** Сводная таблица реконструированных регулонов. В колонке «Инф.» показана информация о литературных данных о реконструированных регулонах: «С» - были показаны сайты связывания белка-регулятора в изучаемых организмах, «С(др.)» - сайты связывания были показаны на других организмах, «Р» - были показаны только регулируемые гены, «0» - информации о регулоне в литературе не обнаружено.

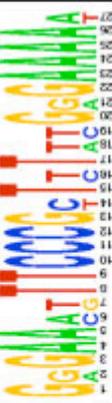
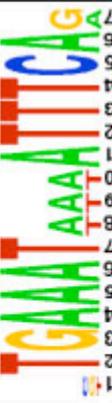
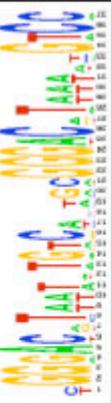
Имя регулятора	Функция регулируемых генов	Семейство	Инф.	Количество регулируемых генов									LOGO мотивов сайтов связывания регуляторов			
				BSU	BAY	BLI	BPU	BCE	AFL	GKA	BNA	BCL		OIN		
RhgR* (YesS)	Утилизация рамногалактуронана	AraC	0	17	0	19	0	0	0	0	0	0	0	0	0	
RmgR* (YtdP)	Утилизация рамногалактуронана	AraC	0	5	0	5	5	0	0	0	0	11	11	7	7	
LicR	Утилизация бета-глюкозидов	BglG	C	4	4	0	4	0	0	0	0	0	5	5	5	
ManR	Утилизация маннозы	BglG	P	3	3	3	0	0	0	0	0	0	0	0	0	
MtIR	Утилизация маннитола	BglG	C(др.), P	4	4	3	4	0	3	4	0	0	0	4	4	
CcpN	Глюконеогенез	CcpN	C	2	2	2	2	2	1	2	2	2	2	2	2	

Имя	Функция регулируемых генов	Семейств во	Инф.	BSU	BAY	B LI	BPU	BCE	AFL	GKA	BNA	BCL	OIH	LOGO мотивов сайтов связывания регуляторов
CitT	Утилизация цитрата	CitB	C	1	1	1	1	1	0	0	1	1	3	
MalR	Утилизация малата	CitB	C	4	3	3	3	0	0	0	0	0	0	
FruR	Утилизация фруктозы	DeoR	0	3	3	3	3	3	3	3	3	3	3	
IolR	Утилизация инозитола	DeoR	C	12	12	12	0	0	0	0	0	10	0	
RhaR* (YuiB)	Утилизация рамнозы	DeoR	0	5	0	5	0	0	0	0	4	0	2	
AcoR	Утилизация ацетоина	Fis	C	4	4	1	4	4	0	0	4	0	0	
AraR	Утилизация арабинозы	GntR	C	13	13	13	5	0	6	7	11	0	5	
BglR* (YudK)	Утилизация бета-глюкозидов	GntR	0	3	2	0	2	0	0	0	0	0	0	

Имя	Функция регулируемых генов	Семейство	Инф.	BSU	BAU	BLI	BPU	BCE	AFL	GKA	BNA	BCL	OIH	LOGO мотивов сайтов связывания регуляторов
FfIR	Утилизация фруктозолизина	GntR	P	6	7	0	0	0	0	0	0	0	0	
GamR* (YbgA)	Утилизация глюкозамина	GntR	0	3	0	7	0	0	0	0	6	6	6	
GmuR	Утилизация глюкоманнана	GntR	P	8	7	6	5	0	0	5	0	0	0	
GntR	Утилизация глюконата	GntR	C	4	0	4	0	0	0	0	0	0	3	
GudR	Утилизация глюкората; Утилизация галактората	GntR	P	6	0	6	0	0	0	0	0	6	7	
LutR	Утилизация лактата	GntR	P	5	5	5	3	3	5	4	4	4	5	
NagR	Утилизация N-ацетилглюкозамина	GntR	C	4	4	5	4	3	0	3	4	2	1	
TreR	Утилизация трехалозы	GntR	C	3	3	3	3	2	3	3	3	2	0	

Имя	Функция регулируемых генов	Семейство	Инф.	BSU	BAУ	BLI	BPU	BCE	AFL	GKA	BNA	BCL	OIH	LOGO мотивов сайтов связывания регуляторов
GutR	Утилизация сорбитола	GutR	C	3	3	0	0	0	0	0	0	0	0	
HxIR	Рибулоза монофосфатный путь	HxIR	C	2	2	2	2	0	0	0	0	0	2	
CsrA	Метаболизм углерода	LacI	C	217	171	195	144	44	50	44	119	119	81	
CsrB	Метаболизм углерода	LacI	P	6	2	0	0	0	0	0	0	0	0	
DegA	Утилизация инозитола	LacI	0	2	0	1	0	0	0	13	6	0	0	
ExuR	Утилизация глюкуроната; Утилизация галактуроната	LacI	C	10	5	5	4	0	0	0	0	0	0	
GanR	Утилизация галактана	LacI	P	6	0	6	6	0	0	0	5	5	0	
KdgR	Утилизация 2-кето-3-деоксимлюконата	LacI	C	6	0	6	5	0	0	0	0	0	0	

Имя	Функция регулируемых генов	Семейство	Инф.	BSU	BAУ	BLI	BPU	BCE	AFL	GKA	BNA	BCL	OIH	LOGO мотивов сайтов связывания регуляторов
MdxR	Утилизация мальтодекстрина	LacI	0	7	0	7	0	0	0	0	0	0	0	
MsmR	Утилизация альфа-галактозидов	LacI	0	5	5	5	5	0	0	0	5	0	0	
NtdR	Утилизация неотрехалозамина	LacI	P	4	0	4	4	0	0	0	0	0	0	
RbsR	Утилизация рибозы	LacI	C(др.)	7	6	6	6	5	6	6	6	6	6	
YkvZ	Утилизация бета-глюкозидов	LacI	0	2	2	1	1	2	0	0	0	0	0	
LevR	Утилизация левана	LevR	C	5	0	5	0	0	0	0	0	0	0	
AlsR	Биосинтез ацетоина	LysR	P	3	3	3	3	0	0	0	0	0	0	

Имя	Функция регулируемых генов	Семейство	Инф.	BSU	BAU	BLI	ВРУ	ВСЕ	AFL	GKA	ВНА	ВСL	OIN	LOGO мотивов сайтов связывания регуляторов								
CspC	Метаболизм цитрата	LysR	C	5	4	2	2	1	0	0	0	0	2									
CitR	Метаболизм цитрата	LysR	P	2	2	2	0	0	0	0	0	2	0									
XylR	Утилизация ксилозы	ROK	C	5	5	10	5	0	0	8	12	6	7									
GlvR	Утилизация мальтозы	RpiR	P	3	3	3	0	0	0	0	0	0	0									
MurR	Утилизация N-ацетилмурамата	RpiR	0	6	0	6	6	0	0	0	0	0	0									
CggR	Гликолиз	SorC	C	6	6	6	6	2	6	6	6	6	6									
Количество регулонов:													43	30	38	28	12	9	13	19	17	19

\* - Новое имя регулона, данное в этой работе. Старое название дано в скобках.

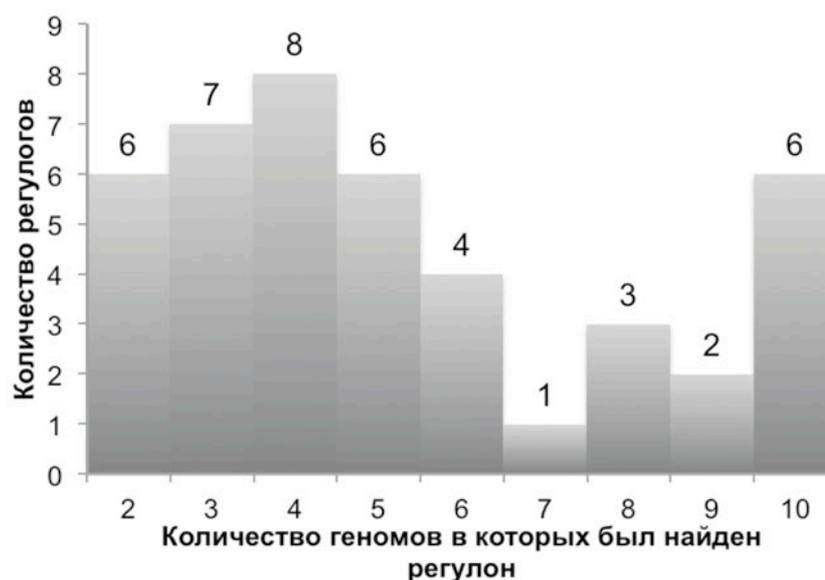
### 6.3 Реконструкция регулонов утилизации сахаров и их производных в семействе бактерий *Bacillaceae*

В данной работе удобно пользоваться термином регулог, как набором регулонов, контролируемых ортологичными факторами транскрипции в группе родственных бактерий. В результате исследования было реконструировано 43 регулога, отвечающих за метаболизм углерода в бактериях семейства *Bacillaceae*. Соответствующие факторы транскрипции принадлежат к пятнадцати семействам бактериальных регуляторов. Наибольшее число регулогов контролируется факторами транскрипции из двух семейств – LacI (11 регулогов) и GntR (10 регулогов). К остальным семействам принадлежит от одного до трех регуляторов (Табл. 6.1 отсортирована по семействам регуляторов).

Всего 6 регулогов присутствуют во всех изучаемых геномах: глобальный регулог метаболизма углеводов CspA, регулог гликолиза CggR, регулог глюконеогенеза CspN, а также регулоги утилизации фруктозы (FruR), рибозы (RbsR) и лактата (LutR). При этом больше половины реконструированных регулогов содержатся в пяти или менее геномах (Рис. 6.2). Также были выявлены неоднократные случаи исчезновения регуляторов при сохранении большинства генов регулогов. Это хорошо соотносится как с наблюдением, что факторы транскрипции эволюционируют быстрее, чем регулируемые гены (243), а также с моделью эволюции метаболического инструментария (toolbox model) бактерий, которая утверждает, что большинство метаболических путей переносится горизонтально с последующим быстрым встраиванием в существующую регуляторную сеть (244).

В сеть регуляции метаболизма углеводов в бактериях семейства *Bacillaceae* входит более 300 ортологичных рядов регулируемых генов (Приложение 4). Для *B. subtilis* предсказано более 60 новых регуляторных взаимодействий, из которых 42 сделано для CspA регулона. Среди новых членов CspA регулона в *B. subtilis* присутствуют гены утилизации различных сахаров: рамногалактуронана (*yesOPQRSTUVWXYZ-yetA-lplABCD*),

галактуроната (*kduID*), фруктозы (*fruRKA*), сахарозы (*sacPA*), мальтозы (*sucB-ganPQA*), лактата (*lutABC*) и других. Также, интересно, что сайты CsrA были найдены перед генами, кодирующими ферменты цикла трикарбоновых кислот, таких как сукцинатдегидрогеназа (*sdhCAB-ysmA*), сукцинил-КоА лигаза (*sucCD*) и 2-оксоглутарат дегидрогеназа (*odhAB*).



**Рисунок 6.2.** Распределение количества реконструированных регулонов по количеству геномов, в которых был найден регулон.

При реконструкции CsrA регулога не было найдено ни одного абсолютно консервативного случая регуляции. Наиболее консервативными являются четыре оперона, присутствующие в девяти и з десяти регулонов. Кроме уже названного оперона, кодирующего 2-оксоглутаратдегидрогеназу (*odhAB*), и оперона, кодирующего гены предсказанной лактатдегидрогеназы (*yvfVWY*), в это число входят еще два, напрямую не связанные с метаболизмом углерода: *sigL*, кодирующий сигма-фактор, необходимый для транскрипции некоторых генов метаболизма азота (211) и *mmgABCDE-prpB*, кодирующий гены, участвующие в метаболизме материнских клеток при споруляции (215). Однако если рассматривать высоко консервативную часть регулона, встречаемую в семи-восьми геномах, она сильно обогащена оперонами утилизации различных

источников углерода: мальтозы (*msmX*), глюкоманнанов (*gmuBACDREFG*), глицерола (*glpFK*), маннитола (*mtlAFD*), ксилозы (*xylAB*), рибозы (*rbsRKDACB*), инозитола (*iolABCDEFGHILJ*), арабинозы (*araABDLMNPQ-abfA*) и сахарозы (*sacPA*). В *CspA* регулоге были обнаружены гены, кодирующие 22 регулятора. Большинство из них (*GlvR*, *MsmR*, *AcoR*, *RbsR*, *LutR*, *GmuR*, *FruR*, *AraR*, *TreR*, *CspC*, *ExuR*, *GntR*, *GanR*, *RmgR*, *LicR*, *LicT*, *KdgR* и *ManR*) отвечает за контроль экспрессии генов тех или иных путей метаболизма углерода. Также сайты связывания *CspA* нашлись перед геном регулятора ответа на дефицит фосфата *phoP* и геном *ukoM*, кодирующим регулятор неизвестной функции из *MarR* семейства.

Кроме *CspA* в образовании регуляторных каскадов участвует еще один регулятор катаболитной репрессии, *CspB*. Однако, он был найден всего в двух геномах – *B. subtilis* и *B. amyloliquefaciens*. Причем, если в *B. subtilis* он регулирует опероны катаболизма глюконата (*gntRKPZ*) и ксилозы (*xylAB*), то в *B. amyloliquefaciens* в *CspB* регулон входит только оперон *xylAB*.

Для шести ранее изученных в *B. subtilis* регулонов (*MtlR*, *LutR*, *NtdR*, *GanR*, *CitR* и *XylR*) была предсказана авторегуляция гена фактора транскрипции. Кроме того, в *B. subtilis* для двух регулонов обнаружены новые гены транспортных систем: в *LutR* регулон включен ген лактат пермеазы *lutP* и в *FrlR* регулоне сильные сайты связывания предсказаны перед АТФ-связывающей компонентой неизвестного АВС транспортера.

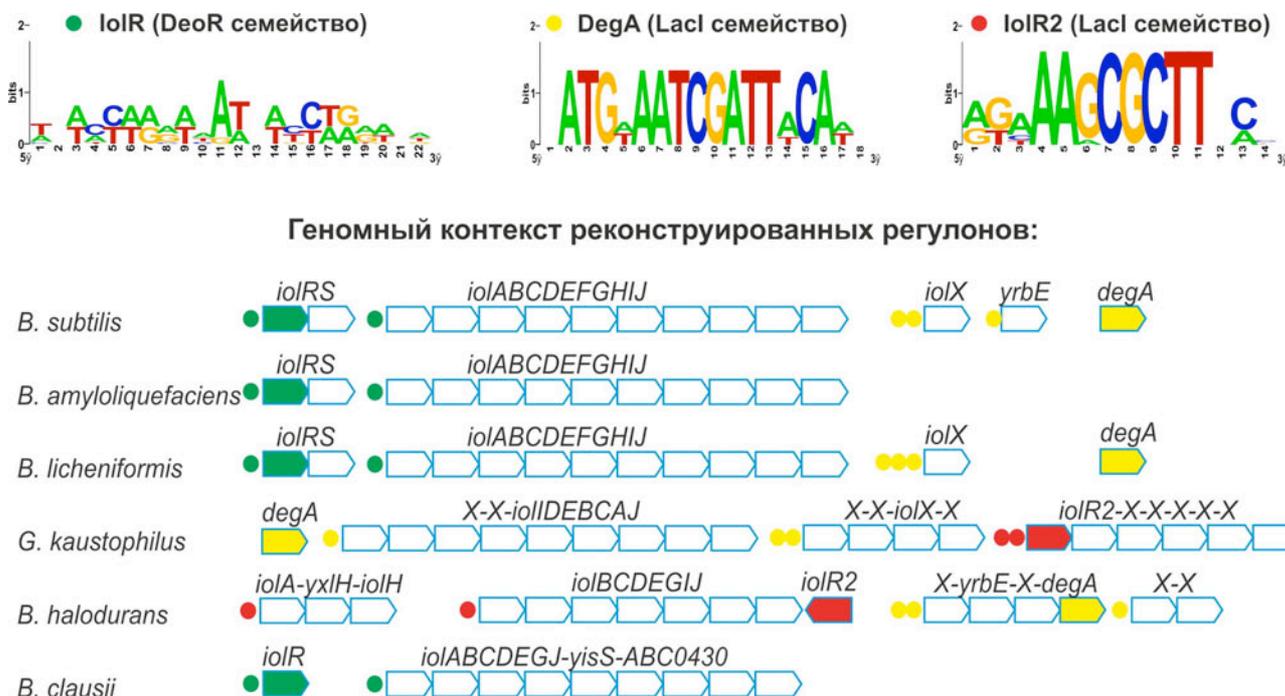
В *B. subtilis* было реконструировано 11 новых регулонов, отвечающих за метаболизм углерода. Факторы транскрипции *BglR* (раннее *YudK*) и *YkvZ* были предсказаны в качестве локальных регуляторов генов утилизации арил-бета-глюкозидов. *BglR* регулирует оперон, кодирующий ПВС компоненту предполагаемого PTS транспортера арил-бета-глюкозидов (*YuzE*) и арил-фосфо-бета-глюкозидазу (*BglA*). Регулятор транскрипции *YkvZ* контролирует также ген *bglC*, кодирующий арил-фосфо-бета-глюкозидазу. Оба фермента способны гидролизовать арил-фосфо-бета-глюкозиды растительного происхождения, такие как арбутин или салицин (245). К тому же можно

предсказать, что эффектором фактора транскрипции BglR служит салицин, поскольку известно, что экспрессия гена *bglA* индуцируется в его присутствии (245,246).

Также было обнаружено, что регулятор DegA из LacI семейства контролирует экспрессию сцилло-инозитол дегидрогеназы *iolX* (247) и предсказанной дегидрогеназы неизвестного производного инозитола *urbE* в *B. subtilis*. Состав DegA регулона сильно меняется в разных геномах. Так, функцией DegA предсказана регуляция генов катаболизма инозитола *iolIDEBCAJ* в *G. kaustophilus*. Тем самым DegA занял ту же функцию, что и фактор транскрипции IolR в *B. subtilis* и других бактериях семейства *Bacillaceae*. Для *B. halodurans* было замечено, что хотя гены пути утилизации инозитола не находятся под регуляцией ни IolR, ни DegA, рядом с *iol* опероном присутствует ген, кодирующий фактор транскрипции из LacI семейства BH2313, которому впоследствии было присвоено имя IolR2. В *B. halodurans* под IolR2 регуляцией обнаружилось два оперона генов утилизации инозитола *iolBCDEGLI* и *iolA-uxlH-iolH*. Ген *uxlH* кодирует новый транспортер, предположительно специфичный к одному из производных инозитола. Также IolR2 регулон был обнаружен в *G. kaustophilus*, где в него входит один оперон *iolR2-GK2114-11*, который кодирует две дегидрогеназы GK2114 и GK2112 и предсказанную изомеразу GK2111. Таким образом, похоже, что большинство бактерий кроме генов основного катаболизма инозитола содержат вспомогательные гены для утилизации его различных производных. При этом, хотя для других систем катаболизма сахаров подобного исследования сделано не было, путь утилизации инозитола показал высокую пластичность регуляции транскрипции (Рис.6.3).

Еще восемь регулонов отвечает за утилизацию различных сахаров и их производных: RhgR (раннее YesS) и RmgR (раннее YtdP) контролируют гены утилизации рамногалактуронана; FruR – гены утилизации фруктозы; GamR (ранее YbgA) – гены утилизации глюкозамина; MdxR – гены катабол изма

мальтодекстрина; MsmR – гены утилизации альфа-галактозидов и MurR – гены утилизации N-ацетилмурамата.



**Рисунок 6.3.** Регулоны утилизации инозитола в бактериях семейства *Bacillaceae* и соответствующие мотивы сайтов связывания. Сайты связывания показаны кругами, регулируемые гены – стрелками. Закрашенные стрелки обозначают гены регуляторов: зеленый – IoIR, желтый – DegA, красный – IoIR2. Сайты связывания раскрашены соответственно принадлежности к регулятору.

### 6.4 Обсуждение

Среди 43 реконструированных регулонов лишь 14 регулонов полностью сохраняют свой состав во всех геномах. При этом, среди них всего два регулона присутствуют во всех анализируемых организмах – FruR, контролирующий гены утилизации фруктозы и CspN, регулирующий гены глюконеогенеза. В шести и пяти геномах, соответственно, были найдены регулоны утилизации ацетона AcoR и рибулозо-монофосфатного пути HxlR. Остальные регулоны, входящие в эту группу (AlsR, CitR, GlvR, GntR, GutR, LevR, ManR, MdxR, MurR, NtdR), мало распространены среди изучаемых организмов и встречаются

в двух-четырех геномах. Также к группе абсолютно консервативных регулонов, присутствующих во всех геномах можно было бы причислить CggR регулон, однако в *B. cereus* произошел переход генов *tpi* и *pgk* в псевдогены. Еще в трех регулогах (TreR, FrlR, LicR) обнаружены изменения только в авторегуляции факторов транскрипции.

При анализе консервативности реконструированных регулогов было отмечено, что для локальных регулонов основные отличия между организмами заключаются в изменении регуляции генов, ответственных за ранние шаги метаболических путей. Так, частым изменением в составе регулона является смена регуляции транспортеров субстратов, что было замечено в 12 регулонах (KdgR, IolR, LutR, MtlR, RbsR, NagR, MalR, CitT, GudR, RmgR, XylR, RhgR). Наблюдались как потери сайтов связывания перед генами транспортеров, так и потеря самого гена или появление новых генов транспортеров. Конечно, интересны случаи замены генов транспортеров, которые позволяют предсказать их специфичность. В регулоне утилизации инозитола IolR в *B. clausii* вместо транспортера мио-инозитола IolF в появился транспортер ABC0430, и ортолог сцилло-инозитол дегидрогеназы IolX. Также отсутствие в геноме ортологов инозозаизомеразы IolII при сохранении мио-инозитолдегидрогеназы IolG позволяет предположить, что ABC0430 специфичен к сцилло-инозитолу и, вероятно, к мио-инозитолу.

В CitT регулоге встретилось три ряда транспортеров. В большинстве бактерий присутствует транспортер цитрата CitM. В *B. cereus* под регуляцией находится транспортер CitH из того же семейства субстрат/дивалентный ион симпортеров. В *O. iheyensis* сайт связывания CitT обнаружен перед генами OB3249-47, кодирующими субъединицы транспортера из семейства трехсоставных транспортеров трикарбоксилатов (tripartite tricarboxylate transporter). Соответственно, для всех этих транспортеров можно предсказать специфичность к цитрату.

В регулоге GudR в геномах *B. clausii* и *O. iheyensis* отсутствует транспортер глюкората и галактората GudP, а также галакторатдегидратаза

GarD. В геноме *B. clausii* также нет и глюкаратдегидратазы GudD, однако вместо нее в регулоне появилась дегидратаза из семейства маннонат дегидратаз ABC0466, которой была присвоена функция глюкарат дегидратазы. В обоих геномах под регуляцией GudR находится транспортер из семейства трехсоставных транспортеров трикарбоксилатов ABC0469-67, который, наиболее вероятно, специфичен исключительно к глюкарату.

Среди генов, входящих в XylR регулоны замечено большое разнообразие транспортеров и гидролаз. При этом гены *xylA* и *xylB*, кодирующие соответственно ксилозаизомеразу и ксилулозакиназу самого пути конвертации ксилозы в ксилулозу-5-фосфат – интермедиат пентозофосфатного пути – консервативны во всех регулонах. Также замещения генов гидролаз были замечены в регулогах MsmR, YkvZ, RmgR.

В регулоне утилизации N-ацетилглюкозамина NagR в *B. licheniformis* в состав регулона вошел ген, кодирующий N-ацетилмурамат-6-фосфат эстеразу, которая преобразует N-ацетилмурамат-6-фосфат в глюкозамин-6-фосфат, субстрат глюкозамин-6-фосфат деаминазы NagB. При этом, в NagR регулоне в *B. licheniformis* обнаружен только один транспортер – PTS система NagP, гомологичная специфичной к N-ацетилглюкозамину PTS системе NagP из *B. subtilis*. Возможно, основным источником N-ацетилглюкозамина для *B. licheniformis* является пептидогликан, состоящий из N-ацетилглюкозамина и N-ацетилмурамата. Таким образом, рост на данном субстрате означает одновременную активацию NagR и MurR регулонов и поступление в клетку обоих типов мономеров.

В составе GmuR регулога самой консервативной частью являются гены *gmuABC*, кодирующие PTS транспортер, и *gmuD*, кодирующий глюкозидазу. Гены, кодирующие маннокиназу GmuE и манноза-6-фосфат изомеразу GmuF, распределены мозаично и присутствуют не во всех геномах. При этом во всех геномах, где обнаружен GmuR регулон, есть манноза-6-фосфат изомераза Pmi, которая, возможно, восполняет недостающую функцию. Возможно, функцию маннокиназы выполняет также фермент из другого пути.

Одним из регулогов, где основные отличия между регулонами лежат в составе генов, кодирующих ферменты последних стадий метаболического пути, является RhaR регулог утилизации рамнозы. В *B. licheniformis* гены *rhaM* и *rhaA* кодирующие в *B. subtilis* рамнозамутаротазу и рамнозаизомеразу, соответственно, отсутствуют. Однако под регуляцией находятся гены *rhaM2* из семейства сахаросвязывающих белков ASRT и *rhaA2* из семейства ксилоза изомераз, которым присвоили функции недостающих генов. Также в *O. iheyensis* были найдены гены *rhaY* и *rhaD*, кодирующие соответственно транспортер, предположительно специфичный к рамнозе, и рамнулоза-1-фосфат альдолазу, замещающую аналогичную активность фермента RhaEW.

Регулог GamR интересен тем, что в *B. subtilis* состав регулона полностью отличается от GamR регулонов в других бактериях. В *B. subtilis* под регуляцией находятся два гена: *gamP*, кодирующий PTS систему, предположительно специфичную к глюкозамину, и *gamB*, кодирующий глюкозамин-6-фосфат деаминазу. Два этих белка составляют собой путь утилизации глюкозамина. В бактериях *B. licheniformis*, *B. clausii*, *B. halodurans* и *O. iheyensis* в состав регулонов входят гены PTS системы, секретируемой хитиназы и двух цитоплазматических глюкозидаз. При этом гидролаза VH0613 гомологична белку ChbG из *E. coli* (41% идентичности) (248), который кодирует хитоолигосахарид деацетилазу. Таким образом, можно предположить, что данный набор функций образует путь деградации хитина до олигосахаридов, их последующего транспорта в клетку, деацетилирования одного из мономером N-ацетилглюкозамина и, если предполагать аналогию с путем утилизации хитоолигосахаридов в *E. coli*, последующего расщепления до N-ацетилглюкозамин-6-фосфата и глюкозамина, которые затем активируют экспрессию генов других путей.

Исходя из наблюдаемой пластичности регуляторной сети, низкая степень консервативности глобального регулона CsrA скорее всего объясняется необходимостью клетки быстро подстраивать новые периферические пути в общий метаболизм.

## Выводы

1. Проведена реконструкция регулонов утилизации L-арабинозы, включающая в себя определение потенциальных сайтов связывания белка AraR, оперонной структуры и функциональной аннотации регулируемых генов в четырех порядках бактерий: *Bacillales*, *Lactobacillales*, *Clostridiales* и *Thermotogales*. Предсказаны неортологичные замещения ферментов L-рибулокиназы и L-арабиноизомеразы у ряда бактерий. Показано широкое разнообразие ферментов деградации арабинозосодержащих полисахаридов и систем транспорта внутрь клетки продуктов деградации. Предложена модель эволюции AraR регулона, включающая таксон-специфичное расширение регулона в определенных группах бактерий.
2. Проведена реконструкция регулонов утилизации N-ацетилгалактозамина (НАГА) у протеобактерий, включающая в себя определение потенциальных сайтов связывания белка AgaR, оперонной структуры и функциональной аннотации регулируемых генов. Показано, что в эволюции AgaR регулона основными движущими силами являлись горизонтальный перенос и дубликации генов. Реконструкция метаболических путей утилизации НАГА и производных сахаров у протеобактерий позволила обнаружить ряд новых систем транспорта данных сахаров в клетку и ферментов осуществляющих начальные стадии катаболизма аминсахаров в цитоплазме.
3. Проведена реконструкция регулонов транскрипционного фактора HexR, контролирующего центральный метаболизм углерода в геномах гамма- и бета-протеобактерий. Описание каждого регулона включает в себя определение потенциальных сайтов связывания белка HexR, оперонной структуры и функциональной аннотации

регулируемых генов. Впервые показано, что HexR является глобальным регулятором генов центрального метаболизма углерода в бактериях порядков: *Vibrionales*, *Aeromonadales*, *Psychromonadales* и *Alteromonadales*. Сравнение функционального состава HexR регулонов различных бактерий позволило определить группы консервативных и изменчивых генов регулона.

4. Проведен широкомасштабный анализ регуляторных систем, отвечающих за метаболизм углеводов и их производных в семействе бактерий *Bacillaceae*. Проведена реконструкция 43 регулонов, для каждого из которых был определен ДНК мотив сайтов связывания, оперонная структура и функциональная аннотация регулируемых генов. Для 22 регулонов мотивы сайтов связывания белка-регулятора были обнаружены впервые. Показана высокая вариабельность данной регуляторной сети у различных бактерий из семейства *Bacillaceae*.

## Список публикаций по теме диссертации

### Статьи в научных журналах:

1. Leyn, S.A., Li, X., Zheng, Q., Novichkov, P.S., Reed, S., Romine, M.F., Fredrickson, J.K., Yang, C., Osterman, A.L., Rodionov, D.A. Control of Proteobacterial Central Carbon Metabolism by the HexR Transcriptional Regulator. *Journal of Biological Chemistry* (2011); **286**(41): 35782-35794.
2. Zhang, L., Leyn, S.A., Gu, Y., Jiang, W., Rodionov, D.A., Yang, C. Ribulokinase and transcriptional regulation of arabinose metabolism in *Clostridium acetobutylicum*. *Journal of bacteriology* (2012); **194**(5): 1055-1064.
3. Leyn, S.A., Gao, F., Yang, C., Rodionov, D.A. N-Acetylgalactosamine Utilization Pathway and Regulon in Proteobacteria: Genomic Reconstruction And Experimental Characterization In *Shewanella*. *Journal of Biological Chemistry* (2012); **287**(33): 28047-28056.
4. Leyn S.A., Kazanov M.D., Sernova N.V., Ermakova E.O., Novichkov P.S., Rodionov D.A. Genomic Reconstruction of the Transcriptional Regulatory Network in *Bacillus subtilis*. *Journal of bacteriology* (2013); **195**(11):2463-73.

### Тезисы конференций:

1. Лейн С.А., Равчеев Д.А. Эволюция химерного белка: исследование AraR зависимой регуляции методами сравнительной геномики. Труды 32-й конференции молодых ученых и специалистов ИППИ РАН ИТиС'09. 2009, с. 303-304.
2. Dmitry Rodionov, Xiaoqing Li, Samantha Reed, Margaret Romine, James Fredrickson, Pavel Novichkov, Semen Leyn, Andrei Osterman. Transcription regulation of *Shewanella* Central Carbon Metabolism by

- HexR. Proceedings of 2010 Genomic Science Contractor-Grantee and Knowledgebase Workshop. 2010, pp. 139-140.
3. D.A. Rodionov, X. Li, A. Osterman, S. Leyn. Comparative and Functional Genomics of the Central Carbon Metabolism Regulon HexR in Proteobacteria. Proceedings of American Society for Microbiology. 110th General Meeting. 2010, R-2867.
  4. Semen Leyn, Dmitry Rodionov. HexR – new central carbohydrate metabolism transcription regulator. Comparative approach study. Труды 33-й конференции молодых ученых и специалистов ИППИ РАН ИТиС'10. 2010, с. 345-349.
  5. Semen Leyn, Xiaoqing Li, Andrei Osterman, Dmitry Rodionov. Comparative and functional genomics of the central carbon metabolism regulon HexR in Proteobacteria. Proceedings of 3rd Annual Joint Conference on Systems Biology, Regulatory Genomics, and Reverse Engineering Challenges 2010 – DREAM5. 2010, p. 211.
  6. Pavel Novichkov, Alexey Kazakov, Semen Leyn, Dmitry Ravcheev, Andrei Osterman, Inna Dubchak, Adam Arkin, Dmitry Rodionov. Large-Scale Genomic Reconstruction of Transcriptional Regulatory Networks in Bacteria. Genomic Science Awardee Meeting IX and USDA-DOE Plant Feedstock Genomics for Bioenergy Awardee Meeting. 2011, pp. 199-200.
  7. Semen Leyn, Marat Kazanov, Pavel Novichkov, Dmitry Rodionov. Reference collection of transcriptional regulons in Bacillales family of bacteria. Proceedings of Moscow Conference on Computational Molecular Biology 2011. 2011, pp. 202-203.
  8. Semen Leyn, Fang Gao, Chen Yang, Dmitry Rodionov. Comparative genomic reconstruction of N - acetylgalactosamine catabolic pathways and transcriptional regulons in Proteobacteria. Proceedings of Moscow Conference on Computational Molecular Biology 2011. 2011, pp. 204-205.

9. Semen Leyn, Dmitry Rodionov. Comparative genomic reconstruction of N-acetylgalactosamine catabolic pathways and transcriptional regulons in Proteobacteria. Труды 34-й конференции молодых ученых и специалистов ИППИ РАН ИТиС'11. 2011, с. 45-47.
10. Pavel Novichkov, Dmitry Ravcheev, Alexey Kazakov, Semen Leyn, Adam Arkin, Inna Dubchak, Dmitry Rodionov. Toward System Biology KnowledgeBase on Transcriptional Regulation in Bacteria. Proceedings of 2012 Genomic Science Awardee Meeting X. 2012, pp. 187-188.
11. Semen Leyn, Marat Kazanov, Pavel Novichkov, Dmitry Rodionov. Reference Collection of Transcriptional Regulons in Bacillales. Proceedings of 2012 Genomic Science Awardee Meeting X. 2012, p. 188.

## **Благодарности**

Автор выражает искреннюю благодарность Дмитрию Александровичу Родионову за научное руководство и поддержку в работе над диссертацией, Дмитрию Андреевичу Равчееву и Михаилу Сергеевичу Гельфанду за помощь и ценное обсуждение работы, Наталии Серновой за помощь в реконструкции СсрА регулона, Инне Суворовой и Ольге Цой за помощь в проверке текста диссертации.

## Список литературы

1. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. and Tanabe, M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res*, **40**, D109-114.
2. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**, 3389-3402.
3. Osterman, A. and Overbeek, R. (2003) Missing genes in metabolic pathways: a comparative genomics approach. *Curr Opin Chem Biol*, **7**, 238-251.
4. Gelfand, M.S., Novichkov, P.S., Novichkova, E.S. and Mironov, A.A. (2000) Comparative analysis of regulatory patterns in bacterial genomes. *Brief Bioinform*, **1**, 357-371.
5. Makarova, K.S., Mironov, A.A. and Gelfand, M.S. (2001) Conservation of the binding site for the arginine repressor in all bacterial lineages. *Genome Biol*, **2**, research0013.0011–0013.0018.
6. Равчеев, Д.А., Гельфанд, М.С., Миронов, А.А. and Рахмананова, А.Б. (2002) Пуриновый регулон гамма-протеобактерий. Детальное описание. *Генетика*, **38**, 1203-1214.
7. Laikova, O.N., Mironov, A.A. and Gelfand, M.S. (2001) Computational analysis of the transcriptional regulation of pentose utilization systems in the gamma subdivision of Proteobacteria. *FEMS Microbiol Lett*, **205**, 315-322.
8. Rodionov, D.A., Mironov, A.A. and Gelfand, M.S. (2001) Transcriptional regulation of pentose utilisation systems in the Bacillus/Clostridium group of bacteria. *FEMS Microbiol Lett*, **205**, 305-314.
9. Ravcheev, D.A., Li, X., Latif, H., Zengler, K., Leyn, S.A., Korostelev, Y.D., Kazakov, A.E., Novichkov, P.S., Osterman, A.L. and Rodionov, D.A. (2012) Transcriptional regulation of central carbon and energy metabolism in bacteria by redox-responsive repressor Rex. *J Bacteriol*, **194**, 1145-1157.

10. Gu, Y., Ding, Y., Ren, C., Sun, Z., Rodionov, D.A., Zhang, W., Yang, S., Yang, C. and Jiang, W. (2010) Reconstruction of xylose utilization pathway and regulons in Firmicutes. *BMC Genomics*, **11**, 255.
11. Yang, C., Rodionov, D.A., Li, X., Laikova, O.N., Gelfand, M.S., Zagnitko, O.P., Romine, M.F., Obratsova, A.Y., Neilson, K.H. and Osterman, A.L. (2006) Comparative genomics and experimental characterization of N-acetylglucosamine utilization pathway of *Shewanella oneidensis*. *J Biol Chem*, **281**, 29872-29885.
12. Rodionov, D.A., Li, X., Rodionova, I.A., Yang, C., Sorci, L., Dervyn, E., Martynowski, D., Zhang, H., Gelfand, M.S. and Osterman, A.L. (2008) Transcriptional regulation of NAD metabolism in bacteria: genomic reconstruction of NiaR (YrxA) regulon. *Nucleic Acids Res*, **36**, 2032-2046.
13. Balleza, E., Lopez-Bojorquez, L.N., Martinez-Antonio, A., Resendis-Antonio, O., Lozada-Chavez, I., Balderas-Martinez, Y.I., Encarnacion, S. and Collado-Vides, J. (2009) Regulation by transcription factors in bacteria: beyond description. *FEMS Microbiol Rev*, **33**, 133-151.
14. Ebright, R.H. (2000) RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J Mol Biol*, **304**, 687-698.
15. Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A. and Darst, S.A. (2000) A structural model of transcription elongation. *Science*, **289**, 619-625.
16. Gourse, R.L., Ross, W. and Gaal, T. (2000) UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Mol Microbiol*, **37**, 687-695.
17. Browning, D.F. and Busby, S.J. (2004) The regulation of bacterial transcription initiation. *Nat Rev Microbiol*, **2**, 57-65.
18. Hampsey, M. (2001) Omega meets its match. *Trends in Genetics*, **17**, 190-191.

19. Gross, C.A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J. and Young, B. (1998) The functional and regulatory roles of sigma factors in transcription. *Cold Spring Harb Symp Quant Biol*, **63**, 141-155.
20. Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O. and Darst, S.A. (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science*, **296**, 1285-1290.
21. Tomsic, M., Tsujikawa, L., Panaghie, G., Wang, Y., Azok, J. and deHaseth, P.L. (2001) Different roles for basic and aromatic amino acids in conserved region 2 of Escherichia coli sigma(70) in the nucleation and maintenance of the single-stranded DNA bubble in open RNA polymerase-promoter complexes. *J Biol Chem*, **276**, 31891-31896.
22. Madan Babu, M. and Teichmann, S.A. (2003) Evolution of transcription factors and the gene regulatory network in Escherichia coli. *Nucleic Acids Res*, **31**, 1234-1244.
23. van Nimwegen, E. (2003) Scaling laws in the functional content of genomes. *Trends Genet*, **19**, 479-484.
24. Ranea, J.A., Grant, A., Thornton, J.M. and Orengo, C.A. (2005) Microeconomic principles explain an optimal genome size in bacteria. *Trends Genet*, **21**, 21-25.
25. Bird, A.P. (1995) Gene number, noise reduction and biological complexity. *Trends Genet*, **11**, 94-100.
26. Moreno-Campuzano, S., Janga, S.C. and Perez-Rueda, E. (2006) Identification and analysis of DNA-binding transcription factors in Bacillus subtilis and other Firmicutes--a genomic approach. *BMC Genomics*, **7**, 147.
27. Perez-Rueda, E. and Collado-Vides, J. (2000) The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. *Nucleic Acids Res*, **28**, 1838-1847.
28. Andersson, S.G., Zomorodipour, A., Andersson, J.O., Sicheritz-Ponten, T., Alsmark, U.C., Podowski, R.M., Naslund, A.K., Eriksson, A.S., Winkler, H.H.

- and Kurland, C.G. (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature*, **396**, 133-140.
29. Barabasi, A.L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet*, **5**, 101-113.
  30. Martinez-Antonio, A. and Collado-Vides, J. (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr Opin Microbiol*, **6**, 482-489.
  31. Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M. and Iyer, L.M. (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev*, **29**, 231-262.
  32. van Hijum, S.A., Medema, M.H. and Kuipers, O.P. (2009) Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiol Mol Biol Rev*, **73**, 481-509, Table of Contents.
  33. Martinez-Antonio, A., Janga, S.C., Salgado, H. and Collado-Vides, J. (2006) Internal-sensing machinery directs the activity of the regulatory network in *Escherichia coli*. *Trends Microbiol*, **14**, 22-27.
  34. Wilson, C.J., Zhan, H., Swint-Kruse, L. and Matthews, K.S. (2007) The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. *Cell Mol Life Sci*, **64**, 3-16.
  35. Nakano, M.M. and Hulett, F.M. (1997) Adaptation of *Bacillus subtilis* to oxygen limitation. *FEMS Microbiol Lett*, **157**, 1-7.
  36. Mascher, T., Helmann, J.D. and Udden, G. (2006) Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiol Mol Biol Rev*, **70**, 910-938.
  37. Ellermeier, C.D., Hobbs, E.C., Gonzalez-Pastor, J.E. and Losick, R. (2006) A three-protein signaling pathway governing immunity to a bacterial cannibalism toxin. *Cell*, **124**, 549-559.
  38. Demple, B. (1996) Redox signaling and gene control in the *Escherichia coli* soxRS oxidative stress regulon--a review. *Gene*, **179**, 53-57.

39. Choy, H.E. and Adhya, S. (1992) Control of gal transcription through DNA looping: inhibition of the initial transcribing complex. *Proc Natl Acad Sci U S A*, **89**, 11264-11268.
40. Valentin-Hansen, P., Sogaard-Andersen, L. and Pedersen, H. (1996) A flexible partnership: the CytR anti-activator and the cAMP-CRP activator protein, comrades in transcription control. *Mol Microbiol*, **20**, 461-466.
41. Ebright, R.H. (1993) Transcription activation at Class I CAP-dependent promoters. *Mol Microbiol*, **8**, 797-802.
42. Zafar, M.A., Shah, I.M. and Wolf, R.E., Jr. (2010) Protein-protein interactions between sigma(70) region 4 of RNA polymerase and Escherichia coli SoxS, a transcription activator that functions by the prerecruitment mechanism: evidence for "off-DNA" and "on-DNA" interactions. *J Mol Biol*, **401**, 13-32.
43. Brown, N.L., Stoyanov, J.V., Kidd, S.P. and Hobman, J.L. (2003) The MerR family of transcriptional regulators. *FEMS Microbiol Rev*, **27**, 145-163.
44. Minchin, S.D. and Busby, S.J. (2009) Analysis of mechanisms of activation and repression at bacterial promoters. *Methods*, **47**, 6-12.
45. Grainger, D.C., Lee, D.J. and Busby, S.J. (2009) Direct methods for studying transcription regulatory proteins and RNA polymerase in bacteria. *Curr Opin Microbiol*, **12**, 531-535.
46. Ishihama, A. (2010) Prokaryotic genome regulation: multifactor promoters, multitarget regulators and hierarchic networks. *FEMS Microbiol Rev*, **34**, 628-645.
47. Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M. *et al.* (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science*, **269**, 496-512.
48. Blanchette, M. and Tompa, M. (2002) Discovery of regulatory elements by a computational method for phylogenetic footprinting. *Genome Res*, **12**, 739-748.

49. Fitch, W.M. (1970) Distinguishing homologous from analogous proteins. *Syst Zool*, **19**, 99-113.
50. Martinez-Nunez, M.A., Perez-Rueda, E., Gutierrez-Rios, R.M. and Merino, E. (2010) New insights into the regulatory networks of paralogous genes in bacteria. *Microbiology*, **156**, 14-22.
51. Teichmann, S.A. and Babu, M.M. (2004) Gene regulatory network growth by duplication. *Nat Genet*, **36**, 492-496.
52. Blanchette, M. and Tompa, M. (2003) FootPrinter: A program designed for phylogenetic footprinting. *Nucleic Acids Res*, **31**, 3840-3842.
53. D'Haeseleer, P. (2006) How does DNA sequence motif discovery work? *Nat Biotechnol*, **24**, 959-961.
54. Pavesi, G., Mereghetti, P., Mauri, G. and Pesole, G. (2004) Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res*, **32**, W199-203.
55. Schumacher, M.A., Sprehe, M., Bartholomae, M., Hillen, W. and Brennan, R.G. (2011) Structures of carbon catabolite protein A-(HPr-Ser46-P) bound to diverse catabolite response element sites reveal the basis for high-affinity binding to degenerate DNA operators. *Nucleic Acids Res*, **39**, 2931-2942.
56. Stormo, G.D. (2000) DNA binding sites: representation and discovery. *Bioinformatics*, **16**, 16-23.
57. MacIsaac, K.D. and Fraenkel, E. (2006) Practical strategies for discovering regulatory DNA sequence motifs. *PLoS Comput Biol*, **2**, e36.
58. Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*, **2**, 28-36.
59. Gelfand, M.S., Koonin, E.V. and Mironov, A.A. (2000) Prediction of transcription regulatory sites in Archaea by a comparative genomic approach. *Nucleic Acids Res*, **28**, 695-705.
60. Novichkov, P.S., Rodionov, D.A., Stavrovskaya, E.D., Novichkova, E.S., Kazakov, A.E., Gelfand, M.S., Arkin, A.P., Mironov, A.A. and Dubchak, I.

- (2010) RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach. *Nucleic Acids Res*, **38**, W299-307.
61. Lawrence, C.E., Altschul, S.F., Boguski, M.S., Liu, J.S., Neuwald, A.F. and Wootton, J.C. (1993) Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science*, **262**, 208-214.
62. Roth, F.P., Hughes, J.D., Estep, P.W. and Church, G.M. (1998) Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat Biotechnol*, **16**, 939-945.
63. Thompson, W.A., Newberg, L.A., Conlan, S., McCue, L.A. and Lawrence, C.E. (2007) The Gibbs Centroid Sampler. *Nucleic Acids Res*, **35**, W232-237.
64. Favorov, A.V., Gelfand, M.S., Gerasimova, A.V., Ravcheev, D.A., Mironov, A.A. and Makeev, V.J. (2005) A Gibbs sampler for identification of symmetrically structured, spaced DNA motifs with improved estimation of the signal length. *Bioinformatics*, **21**, 2240-2245.
65. Rodionov, D.A. (2007) Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem Rev*, **107**, 3467-3497.
66. Mironov, A.A., Koonin, E.V., Roytberg, M.A. and Gelfand, M.S. (1999) Computer analysis of transcription regulatory patterns in completely sequenced bacterial genomes. *Nucleic Acids Res*, **27**, 2981-2989.
67. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Sayers, E.W. (2011) GenBank. *Nucleic Acids Res*, **39**, D32-37.
68. Magrane, M. and Consortium, U. (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)*, **2011**, bar009.
69. Kosiol, C. and Goldman, N. (2005) Different versions of the Dayhoff rate matrix. *Mol Biol Evol*, **22**, 193-199.
70. Henikoff, S. and Henikoff, J.G. (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A*, **89**, 10915-10919.
71. Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J. *et al.* (2012) The Pfam protein families database. *Nucleic Acids Res*, **40**, D290-301.

72. Gough, J., Karplus, K., Hughey, R. and Chothia, C. (2001) Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol*, **313**, 903-919.
73. Letunic, I., Doerks, T. and Bork, P. (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res*, **40**, D302-305.
74. Bendtsen, J.D., Nielsen, H., von Heijne, G. and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol*, **340**, 783-795.
75. Hofmann, K. and Stoffel, W. (1993) TMbase - A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler*, **374**, 166.
76. Combet, C., Blanchet, C., Geourjon, C. and Deleage, G. (2000) NPS@: network protein sequence analysis. *Trends Biochem Sci*, **25**, 147-150.
77. Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G.D. and Maltsev, N. (1999) The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci U S A*, **96**, 2896-2901.
78. Galperin, M.Y. and Koonin, E.V. (2000) Who's your neighbor? New computational approaches for functional genomics. *Nat Biotechnol*, **18**, 609-613.
79. Marrakchi, H., Choi, K.H. and Rock, C.O. (2002) A new mechanism for anaerobic unsaturated fatty acid formation in *Streptococcus pneumoniae*. *J Biol Chem*, **277**, 44809-44816.
80. Pellegrini, M., Marcotte, E.M., Thompson, M.J., Eisenberg, D. and Yeates, T.O. (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A*, **96**, 4285-4288.
81. Hecht, S., Eisenreich, W., Adam, P., Amslinger, S., Kis, K., Bacher, A., Arigoni, D. and Rohdich, F. (2001) Studies on the nonmevalonate pathway to terpenes: the role of the GcpE (IspG) protein. *Proc Natl Acad Sci U S A*, **98**, 14837-14842.
82. Cunningham, F.X., Jr., Lafond, T.P. and Gantt, E. (2000) Evidence of a role for LytB in the nonmevalonate pathway of isoprenoid biosynthesis. *J Bacteriol*, **182**, 5841-5848.

83. Enright, A.J., Iliopoulos, I., Kyrpides, N.C. and Ouzounis, C.A. (1999) Protein interaction maps for complete genomes based on gene fusion events. *Nature*, **402**, 86-90.
84. Marcotte, E.M., Pellegrini, M., Ng, H.L., Rice, D.W., Yeates, T.O. and Eisenberg, D. (1999) Detecting protein function and protein-protein interactions from genome sequences. *Science*, **285**, 751-753.
85. Yanai, I., Derti, A. and DeLisi, C. (2001) Genes linked by fusion events are generally of the same functional category: a systematic analysis of 30 microbial genomes. *Proc Natl Acad Sci U S A*, **98**, 7940-7945.
86. Kyrpides, N.C., Ouzounis, C.A., Iliopoulos, I., Vonstein, V. and Overbeek, R. (2000) Analysis of the *Thermotoga maritima* genome combining a variety of sequence similarity and genome context tools. *Nucleic Acids Res*, **28**, 4573-4576.
87. Daugherty, M., Polanuyer, B., Farrell, M., Scholle, M., Lykidis, A., de Crecy-Lagard, V. and Osterman, A. (2002) Complete reconstitution of the human coenzyme A biosynthetic pathway via comparative genomics. *J Biol Chem*, **277**, 21431-21439.
88. Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M. *et al.* (2009) STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res*, **37**, D412-416.
89. Enault, F., Suhre, K., Poirot, O., Abergel, C. and Claverie, J.M. (2004) Phydbac2: improved inference of gene function using interactive phylogenomic profiling and chromosomal location analysis. *Nucleic Acids Res*, **32**, W336-339.
90. Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crecy-Lagard, V., Diaz, N., Disz, T., Edwards, R. *et al.* (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res*, **33**, 5691-5702.

91. Dehal, P.S., Joachimiak, M.P., Price, M.N., Bates, J.T., Baumohl, J.K., Chivian, D., Friedland, G.D., Huang, K.H., Keller, K., Novichkov, P.S. *et al.* (2010) MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res*, **38**, D396-400.
92. Franco, I.S., Mota, L.J., Soares, C.M. and de Sa-Nogueira, I. (2006) Functional domains of the *Bacillus subtilis* transcription factor AraR and identification of amino acids important for nucleoprotein complex assembly and effector binding. *J Bacteriol*, **188**, 3024-3036.
93. Haydon, D.J. and Guest, J.R. (1991) A new family of bacterial regulatory proteins. *FEMS Microbiol Lett*, **63**, 291-295.
94. Weickert, M.J. and Adhya, S. (1992) A family of bacterial regulators homologous to Gal and Lac repressors. *J Biol Chem*, **267**, 15869-15874.
95. Sa-Nogueira, I., Nogueira, T.V., Soares, S. and de Lencastre, H. (1997) The *Bacillus subtilis* L-arabinose (*ara*) operon: nucleotide sequence, genetic organization and expression. *Microbiology*, **143 ( Pt 3)**, 957-969.
96. Sa-Nogueira, I. and Mota, L.J. (1997) Negative regulation of L-arabinose metabolism in *Bacillus subtilis*: characterization of the *araR* (*araC*) gene. *J Bacteriol*, **179**, 1598-1608.
97. Raposo, M.P., Inacio, J.M., Mota, L.J. and de Sa-Nogueira, I. (2004) Transcriptional regulation of genes encoding arabinan-degrading enzymes in *Bacillus subtilis*. *J Bacteriol*, **186**, 1287-1296.
98. Sa-Nogueira, I. and Ramos, S.S. (1997) Cloning, functional analysis, and transcriptional regulation of the *Bacillus subtilis* *araE* gene involved in L-arabinose utilization. *J Bacteriol*, **179**, 7705-7711.
99. Mota, L.J., Tavares, P. and Sa-Nogueira, I. (1999) Mode of action of AraR, the key regulator of L-arabinose metabolism in *Bacillus subtilis*. *Mol Microbiol*, **33**, 476-489.
100. Franco, I.S., Mota, L.J., Soares, C.M. and de Sa-Nogueira, I. (2007) Probing key DNA contacts in AraR-mediated transcriptional repression of the *Bacillus subtilis* arabinose regulon. *Nucleic Acids Res*, **35**, 4755-4766.

101. Sa-Nogueira, I. and de Lencastre, H. (1989) Cloning and characterization of *araA*, *araB*, and *araD*, the structural genes for L-arabinose utilization in *Bacillus subtilis*. *J Bacteriol*, **171**, 4088-4091.
102. Guldan, H., Sterner, R. and Babinger, P. (2008) Identification and characterization of a bacterial glycerol-1-phosphate dehydrogenase: Ni(2+)-dependent AraM from *Bacillus subtilis*. *Biochemistry*, **47**, 7376-7384.
103. Krispin, O. and Allmansberger, R. (1998) The *Bacillus subtilis* AraE protein displays a broad substrate specificity for several different sugars. *J Bacteriol*, **180**, 3250-3252.
104. Leal, T.F. and de Sa-Nogueira, I. (2004) Purification, characterization and functional analysis of an endo-arabinanase (AbnA) from *Bacillus subtilis*. *FEMS Microbiol Lett*, **241**, 41-48.
105. Reizer, J., Ramseier, T.M., Reizer, A., Charbit, A. and Saier, M.H., Jr. (1996) Novel phosphotransferase genes revealed by bacterial genome sequencing: a gene cluster encoding a putative N-acetylgalactosamine metabolic pathway in *Escherichia coli*. *Microbiology*, **142 ( Pt 2)**, 231-250.
106. Brinkkotter, A., Kloss, H., Alpert, C. and Lengeler, J.W. (2000) Pathways for the utilization of N-acetyl-galactosamine and galactosamine in *Escherichia coli*. *Mol Microbiol*, **37**, 125-135.
107. Nobelmann, B. and Lengeler, J.W. (1996) Molecular analysis of the *gat* genes from *Escherichia coli* and of their roles in galactitol transport and metabolism. *J Bacteriol*, **178**, 6790-6795.
108. Brinkkotter, A., Shakeri-Garakani, A. and Lengeler, J.W. (2002) Two class II D-tagatose-bisphosphate aldolases from enteric bacteria. *Arch Microbiol*, **177**, 410-419.
109. Ray, W.K. and Larson, T.J. (2004) Application of AgaR repressor and dominant repressor variants for verification of a gene cluster involved in N-acetylgalactosamine metabolism in *Escherichia coli* K-12. *Mol Microbiol*, **51**, 813-826.

110. Rodionov, D.A., Yang, C., Li, X., Rodionova, I.A., Wang, Y., Obraztsova, A.Y., Zagnitko, O.P., Overbeek, R., Romine, M.F., Reed, S. *et al.* (2010) Genomic encyclopedia of sugar utilization pathways in the *Shewanella* genus. *BMC Genomics*, **11**, 494.
111. Saier, M.H., Jr. (1998) Multiple mechanisms controlling carbon metabolism in bacteria. *Biotechnol Bioeng*, **58**, 170-174.
112. Gosset, G., Zhang, Z., Nayyar, S., Cuevas, W.A. and Saier, M.H., Jr. (2004) Transcriptome analysis of Crp-dependent catabolite control of gene expression in *Escherichia coli*. *J Bacteriol*, **186**, 3516-3524.
113. Saier, M.H., Jr. and Ramseier, T.M. (1996) The catabolite repressor/activator (Cra) protein of enteric bacteria. *J Bacteriol*, **178**, 3411-3417.
114. Ramseier, T.M., Bledig, S., Michotey, V., Feghali, R. and Saier, M.H., Jr. (1995) The global regulatory protein FruR modulates the direction of carbon flow in *Escherichia coli*. *Mol Microbiol*, **16**, 1157-1169.
115. Collier, D.N., Hager, P.W. and Phibbs, P.V., Jr. (1996) Catabolite repression control in the Pseudomonads. *Res Microbiol*, **147**, 551-561.
116. Lessie, T.G. and Phibbs, P.V., Jr. (1984) Alternative pathways of carbohydrate utilization in pseudomonads. *Annu Rev Microbiol*, **38**, 359-388.
117. Daddaoua, A., Krell, T. and Ramos, J.L. (2009) Regulation of glucose metabolism in *Pseudomonas*: the phosphorylative branch and entner-doudoroff enzymes are regulated by a repressor containing a sugar isomerase domain. *J Biol Chem*, **284**, 21360-21368.
118. del Castillo, T., Duque, E. and Ramos, J.L. (2008) A set of activators and repressors control peripheral glucose pathways in *Pseudomonas putida* to yield a common central intermediate. *J Bacteriol*, **190**, 2331-2339.
119. Petruschka, L., Adolf, K., Burchhardt, G., Dervede, J., Jurgensen, J. and Herrmann, H. (2002) Analysis of the *zwf-pgl-eda*-operon in *Pseudomonas putida* strains H and KT2440. *FEMS Microbiol Lett*, **215**, 89-95.
120. Rodionov, D.A., Novichkov, P.S., Stavrovskaya, E.D., Rodionova, I.A., Li, X., Kazanov, M.D., Ravcheev, D.A., Gerasimova, A.V., Kazakov, A.E., Kovaleva,

- G.Y. *et al.* (2011) Comparative genomic reconstruction of transcriptional networks controlling central metabolism in the *Shewanella* genus. *BMC Genomics*, **12 Suppl 1**, S3.
121. Reizer, J., Bachem, S., Reizer, A., Arnaud, M., Saier, M.H., Jr. and Stulke, J. (1999) Novel phosphotransferase system genes revealed by genome analysis - the complete complement of PTS proteins encoded within the genome of *Bacillus subtilis*. *Microbiology*, **145 ( Pt 12)**, 3419-3429.
  122. Stulke, J. and Hillen, W. (2000) Regulation of carbon catabolism in *Bacillus* species. *Annu Rev Microbiol*, **54**, 849-880.
  123. Zamboni, N., Fischer, E., Laudert, D., Aymerich, S., Hohmann, H.P. and Sauer, U. (2004) The *Bacillus subtilis* yqjI gene encodes the NADP<sup>+</sup>-dependent 6-P-gluconate dehydrogenase in the pentose phosphate pathway. *J Bacteriol*, **186**, 4528-4534.
  124. Sonenshein, A.L., Hoch, J.A. and Losick, R. (2002) *Bacillus Subtilis and Its Closest Relatives*. ASM Press.
  125. Stulke, J. and Hillen, W. (1998) Coupling physiology and gene regulation in bacteria: the phosphotransferase sugar uptake system delivers the signals. *Naturwissenschaften*, **85**, 583-592.
  126. Gonzy-Treboul, G., Zagorec, M., Rain-Guion, M.C. and Steinmetz, M. (1989) Phosphoenolpyruvate:sugar phosphotransferase system of *Bacillus subtilis*: nucleotide sequence of ptsX, ptsH and the 5'-end of ptsI and evidence for a ptsHI operon. *Mol Microbiol*, **3**, 103-112.
  127. Stulke, J., Martin-Verstraete, I., Zagorec, M., Rose, M., Klier, A. and Rapoport, G. (1997) Induction of the *Bacillus subtilis* ptsGHI operon by glucose is controlled by a novel antiterminator, GlcT. *Mol Microbiol*, **25**, 65-78.
  128. Jault, J.M., Fieulaine, S., Nessler, S., Gonzalo, P., Di Pietro, A., Deutscher, J. and Galinier, A. (2000) The HPr kinase from *Bacillus subtilis* is a homo-oligomeric enzyme which exhibits strong positive cooperativity for nucleotide and fructose 1,6-bisphosphate binding. *J Biol Chem*, **275**, 1773-1780.

129. Deutscher, J., Francke, C. and Postma, P.W. (2006) How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol Mol Biol Rev*, **70**, 939-1031.
130. Tobisch, S., Stulke, J. and Hecker, M. (1999) Regulation of the lic operon of *Bacillus subtilis* and characterization of potential phosphorylation sites of the LicR regulator protein by site-directed mutagenesis. *J Bacteriol*, **181**, 4995-5003.
131. Henstra, S.A., Tolner, B., ten Hoeve Duurkens, R.H., Konings, W.N. and Robillard, G.T. (1996) Cloning, expression, and isolation of the mannitol transport protein from the thermophilic bacterium *Bacillus stearothermophilus*. *J Bacteriol*, **178**, 5586-5591.
132. Stulke, J., Arnaud, M., Rapoport, G. and Martin-Verstraete, I. (1998) PRD--a protein domain involved in PTS-dependent induction and carbon catabolite repression of catabolic operons in bacteria. *Mol Microbiol*, **28**, 865-874.
133. Greenberg, D.B., Stulke, J. and Saier, M.H., Jr. (2002) Domain analysis of transcriptional regulators bearing PTS regulatory domains. *Res Microbiol*, **153**, 519-526.
134. Tobisch, S., Glaser, P., Kruger, S. and Hecker, M. (1997) Identification and characterization of a new beta-glucoside utilization system in *Bacillus subtilis*. *J Bacteriol*, **179**, 496-506.
135. Debarbouille, M., Martin-Verstraete, I., Klier, A. and Rapoport, G. (1991) The transcriptional regulator LevR of *Bacillus subtilis* has domains homologous to both sigma 54- and phosphotransferase system-dependent regulators. *Proc Natl Acad Sci U S A*, **88**, 2212-2216.
136. Martin-Verstraete, I., Debarbouille, M., Klier, A. and Rapoport, G. (1994) Interactions of wild-type and truncated LevR of *Bacillus subtilis* with the upstream activating sequence of the levanase operon. *J Mol Biol*, **241**, 178-192.
137. Martin-Verstraete, I., Charrier, V., Stulke, J., Galinier, A., Erni, B., Rapoport, G. and Deutscher, J. (1998) Antagonistic effects of dual PTS-catalysed

- phosphorylation on the *Bacillus subtilis* transcriptional activator LevR. *Mol Microbiol*, **28**, 293-303.
138. Arnaud, M., Vary, P., Zagorec, M., Klier, A., Debarbouille, M., Postma, P. and Rapoport, G. (1992) Regulation of the *sacPA* operon of *Bacillus subtilis*: identification of phosphotransferase system components involved in SacT activity. *J Bacteriol*, **174**, 3161-3170.
  139. Kruger, S. and Hecker, M. (1995) Regulation of the putative *bglPH* operon for aryl-beta-glucoside utilization in *Bacillus subtilis*. *J Bacteriol*, **177**, 5590-5597.
  140. Aymerich, S. and Steinmetz, M. (1992) Specificity determinants and structural features in the RNA target of the bacterial antiterminator proteins of the BglG/SacY family. *Proc Natl Acad Sci U S A*, **89**, 10410-10414.
  141. Arnaud, M., Debarbouille, M., Rapoport, G., Saier, M.H., Jr. and Reizer, J. (1996) In vitro reconstitution of transcriptional antitermination by the SacT and SacY proteins of *Bacillus subtilis*. *J Biol Chem*, **271**, 18966-18972.
  142. Kruger, S., Gertz, S. and Hecker, M. (1996) Transcriptional analysis of *bglPH* expression in *Bacillus subtilis*: evidence for two distinct pathways mediating carbon catabolite repression. *J Bacteriol*, **178**, 2637-2644.
  143. Fujita, Y., Fujita, T., Miwa, Y., Nihashi, J. and Aratani, Y. (1986) Organization and transcription of the gluconate operon, *gnt*, of *Bacillus subtilis*. *J Biol Chem*, **261**, 13744-13753.
  144. Yoshida, K., Miwa, Y., Ohmori, H. and Fujita, Y. (1995) Analysis of an insertional operator mutation (*gntOi*) that affects the expression level of the *Bacillus subtilis* *gnt* operon, and characterization of *gntOi* suppressor mutations. *Mol Gen Genet*, **248**, 583-591.
  145. Galinier, A., Deutscher, J. and Martin-Verstraete, I. (1999) Phosphorylation of either *crh* or *HPr* mediates binding of CcpA to the *Bacillus subtilis* *xyn cre* and catabolite repression of the *xyn* operon. *J Mol Biol*, **286**, 307-314.
  146. Dahl, M.K., Degenkolb, J. and Hillen, W. (1994) Transcription of the *xyl* operon is controlled in *Bacillus subtilis* by tandem overlapping operators spaced by four base-pairs. *J Mol Biol*, **243**, 413-424.

147. Doan, T., Servant, P., Tojo, S., Yamaguchi, H., Lerondel, G., Yoshida, K., Fujita, Y. and Aymerich, S. (2003) The *Bacillus subtilis* ywkA gene encodes a malic enzyme and its transcription is activated by the YufL/YufM two-component system in response to malate. *Microbiology*, **149**, 2331-2343.
148. Glatz, E., Persson, M. and Rutberg, B. (1998) Antiterminator protein GlpP of *Bacillus subtilis* binds to glpD leader mRNA. *Microbiology*, **144 ( Pt 2)**, 449-456.
149. Fujita, Y. (2009) Carbon catabolite control of the metabolic network in *Bacillus subtilis*. *Biosci Biotechnol Biochem*, **73**, 245-259.
150. Nicholson, W.L., Park, Y.K., Henkin, T.M., Won, M., Weickert, M.J., Gaskell, J.A. and Chambliss, G.H. (1987) Catabolite repression-resistant mutations of the *Bacillus subtilis* alpha-amylase promoter affect transcription levels and are in an operator-like sequence. *J Mol Biol*, **198**, 609-618.
151. Weickert, M.J. and Chambliss, G.H. (1990) Site-directed mutagenesis of a catabolite repression operator sequence in *Bacillus subtilis*. *Proc Natl Acad Sci U S A*, **87**, 6238-6242.
152. Miwa, Y., Nakata, A., Ogiwara, A., Yamamoto, M. and Fujita, Y. (2000) Evaluation and characterization of catabolite-responsive elements (cre) of *Bacillus subtilis*. *Nucleic Acids Res*, **28**, 1206-1210.
153. Galinier, A., Kravanja, M., Engelmann, R., Hengstenberg, W., Kilhoffer, M.C., Deutscher, J. and Haiech, J. (1998) New protein kinase and protein phosphatase families mediate signal transduction in bacterial catabolite repression. *Proc Natl Acad Sci U S A*, **95**, 1823-1828.
154. Reizer, J., Hoischen, C., Titgemeyer, F., Rivolta, C., Rabus, R., Stulke, J., Karamata, D., Saier, M.H., Jr. and Hillen, W. (1998) A novel protein kinase that controls carbon catabolite repression in bacteria. *Mol Microbiol*, **27**, 1157-1169.
155. Kravanja, M., Engelmann, R., Dossonnet, V., Bluggel, M., Meyer, H.E., Frank, R., Galinier, A., Deutscher, J., Schnell, N. and Hengstenberg, W. (1999) The

- hprK gene of *Enterococcus faecalis* encodes a novel bifunctional enzyme: the HPr kinase/phosphatase. *Mol Microbiol*, **31**, 59-66.
156. Miwa, Y., Saikawa, M. and Fujita, Y. (1994) Possible function and some properties of the CcpA protein of *Bacillus subtilis*. *Microbiology*, **140 ( Pt 10)**, 2567-2575.
  157. Deutscher, J., Kuster, E., Bergstedt, U., Charrier, V. and Hillen, W. (1995) Protein kinase-dependent HPr/CcpA interaction links glycolytic activity to carbon catabolite repression in gram-positive bacteria. *Mol Microbiol*, **15**, 1049-1053.
  158. Fujita, Y., Miwa, Y., Galinier, A. and Deutscher, J. (1995) Specific recognition of the *Bacillus subtilis* gnt cis-acting catabolite-responsive element by a protein complex formed between CcpA and seryl-phosphorylated HPr. *Mol Microbiol*, **17**, 953-960.
  159. Galinier, A., Haiech, J., Kilhoffer, M.C., Jaquinod, M., Stulke, J., Deutscher, J. and Martin-Verstraete, I. (1997) The *Bacillus subtilis* crh gene encodes a HPr-like protein involved in carbon catabolite repression. *Proc Natl Acad Sci U S A*, **94**, 8439-8444.
  160. Landmann, J.J., Busse, R.A., Latz, J.H., Singh, K.D., Stulke, J. and Gorke, B. (2011) Crh, the paralogue of the phosphocarrier protein HPr, controls the methylglyoxal bypass of glycolysis in *Bacillus subtilis*. *Mol Microbiol*, **82**, 770-787.
  161. Martin-Verstraete, I., Deutscher, J. and Galinier, A. (1999) Phosphorylation of HPr and Crh by HprK, early steps in the catabolite repression signalling pathway for the *Bacillus subtilis* levanase operon. *J Bacteriol*, **181**, 2966-2969.
  162. Gorke, B., Fraysse, L. and Galinier, A. (2004) Drastic differences in Crh and HPr synthesis levels reflect their different impacts on catabolite repression in *Bacillus subtilis*. *J Bacteriol*, **186**, 2992-2995.
  163. Singh, K.D., Schmalisch, M.H., Stulke, J. and Gorke, B. (2008) Carbon catabolite repression in *Bacillus subtilis*: quantitative analysis of repression exerted by different carbon sources. *J Bacteriol*, **190**, 7275-7284.

164. Grundy, F.J., Waters, D.A., Allen, S.H. and Henkin, T.M. (1993) Regulation of the *Bacillus subtilis* acetate kinase gene by CcpA. *J Bacteriol*, **175**, 7348-7355.
165. Tojo, S., Satomura, T., Morisaki, K., Deutscher, J., Hirooka, K. and Fujita, Y. (2005) Elaborate transcription regulation of the *Bacillus subtilis* *ilv-leu* operon involved in the biosynthesis of branched-chain amino acids through global regulators of CcpA, CodY and TnrA. *Mol Microbiol*, **56**, 1560-1573.
166. Wunsche, A., Hammer, E., Bartholomae, M., Volker, U., Burkovski, A., Seidel, G. and Hillen, W. (2012) CcpA forms complexes with CodY and RpoA in *Bacillus subtilis*. *Febs J*, **279**, 2201-2214.
167. Kim, J.H., Yang, Y.K. and Chambliss, G.H. (2005) Evidence that *Bacillus* catabolite control protein CcpA interacts with RNA polymerase to inhibit transcription. *Mol Microbiol*, **56**, 155-162.
168. Chauvaux, S., Paulsen, I.T. and Saier, M.H., Jr. (1998) CcpB, a novel transcription factor implicated in catabolite repression in *Bacillus subtilis*. *J Bacteriol*, **180**, 491-497.
169. Jourlin-Castelli, C., Mani, N., Nakano, M.M. and Sonenshein, A.L. (2000) CcpC, a novel regulator of the LysR family required for glucose repression of the *citB* gene in *Bacillus subtilis*. *J Mol Biol*, **295**, 865-878.
170. Servant, P., Le Coq, D. and Aymerich, S. (2005) CcpN (YqzB), a novel regulator for CcpA-independent catabolite repression of *Bacillus subtilis* gluconeogenic genes. *Mol Microbiol*, **55**, 1435-1451.
171. Blencke, H.M., Reif, I., Commichau, F.M., Detsch, C., Wacker, I., Ludwig, H. and Stulke, J. (2006) Regulation of *citB* expression in *Bacillus subtilis*: integration of multiple metabolic signals in the citrate pool and by the general nitrogen regulatory system. *Arch Microbiol*, **185**, 136-146.
172. Kim, H.J., Jourlin-Castelli, C., Kim, S.I. and Sonenshein, A.L. (2002) Regulation of the *Bacillus subtilis* *ccpC* gene by *ccpA* and *ccpC*. *Mol Microbiol*, **43**, 399-410.

173. Kim, S.I., Jourlin-Castelli, C., Wellington, S.R. and Sonenshein, A.L. (2003) Mechanism of repression by *Bacillus subtilis* CcpC, a LysR family regulator. *J Mol Biol*, **334**, 609-624.
174. Gimpel, M., Heidrich, N., Mader, U., Krugel, H. and Brantl, S. (2010) A dual-function sRNA from *B. subtilis*: SR1 acts as a peptide encoding mRNA on the gapA operon. *Mol Microbiol*, **76**, 990-1009.
175. Eckart, R.A., Brantl, S. and Licht, A. (2009) Search for additional targets of the transcriptional regulator CcpN from *Bacillus subtilis*. *FEMS Microbiol Lett*, **299**, 223-231.
176. Licht, A. and Brantl, S. (2006) Transcriptional repressor CcpN from *Bacillus subtilis* compensates asymmetric contact distribution by cooperative binding. *J Mol Biol*, **364**, 434-448.
177. Doan, T. and Aymerich, S. (2003) Regulation of the central glycolytic genes in *Bacillus subtilis*: binding of the repressor CggR to its single DNA target sequence is modulated by fructose-1,6-bisphosphate. *Mol Microbiol*, **47**, 1709-1721.
178. Sun, T. and Altenbuchner, J. (2010) Characterization of a mannose utilization system in *Bacillus subtilis*. *J Bacteriol*, **192**, 2128-2139.
179. Joyet, P., Derkaoui, M., Poncet, S. and Deutscher, J. (2010) Control of *Bacillus subtilis* mtl operon expression by complex phosphorylation-dependent regulation of the transcriptional activator MtlR. *Mol Microbiol*, **76**, 1279-1294.
180. Henstra, S.A., Tuinhof, M., Duurkens, R.H. and Robillard, G.T. (1999) The *Bacillus stearothermophilus* mannitol regulator, MtlR, of the phosphotransferase system. A DNA-binding protein, regulated by HPr and iicbmtl-dependent phosphorylation. *J Biol Chem*, **274**, 4754-4763.
181. Yamamoto, H., Murata, M. and Sekiguchi, J. (2000) The CitST two-component system regulates the expression of the Mg-citrate transporter in *Bacillus subtilis*. *Mol Microbiol*, **37**, 898-912.
182. Tanaka, K., Kobayashi, K. and Ogasawara, N. (2003) The *Bacillus subtilis* YufLM two-component system regulates the expression of the malate

- transporters MaeN (YufR) and YfIS, and is essential for utilization of malate in minimal medium. *Microbiology*, **149**, 2317-2329.
183. Yoshida, K.I., Shibayama, T., Aoyama, D. and Fujita, Y. (1999) Interaction of a repressor and its binding sites for regulation of the *Bacillus subtilis* *iol* divergon. *J Mol Biol*, **285**, 917-929.
  184. Yoshida, K., Yamamoto, Y., Omae, K., Yamamoto, M. and Fujita, Y. (2002) Identification of two myo-inositol transporter genes of *Bacillus subtilis*. *J Bacteriol*, **184**, 983-991.
  185. Ali, N.O., Bignon, J., Rapoport, G. and Debarbouille, M. (2001) Regulation of the acetoin catabolic pathway is controlled by sigma L in *Bacillus subtilis*. *J Bacteriol*, **183**, 2497-2504.
  186. Deppe, V.M., Klatte, S., Bongaerts, J., Maurer, K.H., O'Connell, T. and Meinhardt, F. (2011) Genetic control of amadori product degradation in *Bacillus subtilis* via regulation of *friBONMD* expression by *FrlR*. *Appl Environ Microbiol*, **77**, 2839-2846.
  187. Chai, Y., Kolter, R. and Losick, R. (2009) A widely conserved gene cluster required for lactate utilization in *Bacillus subtilis* and its involvement in biofilm formation. *J Bacteriol*, **191**, 2423-2430.
  188. Sadaie, Y., Nakadate, H., Fukui, R., Yee, L.M. and Asai, K. (2008) Glucomanan utilization operon of *Bacillus subtilis*. *FEMS Microbiol Lett*, **279**, 103-109.
  189. Hosoya, S., Yamane, K., Takeuchi, M. and Sato, T. (2002) Identification and characterization of the *Bacillus subtilis* D-glucarate/galactarate utilization operon *ycbCDEFGHJ*. *FEMS Microbiol Lett*, **210**, 193-199.
  190. Bertram, R., Rigali, S., Wood, N., Lulko, A.T., Kuipers, O.P. and Titgemeyer, F. (2011) Regulon of the N-acetylglucosamine utilization regulator *NagR* in *Bacillus subtilis*. *J Bacteriol*, **193**, 3525-3536.
  191. Schock, F. and Dahl, M.K. (1996) Expression of the *tre* operon of *Bacillus subtilis* 168 is regulated by the repressor *TreR*. *J Bacteriol*, **178**, 4576-4581.

192. Poon, K.K., Chen, C.L. and Wong, S.L. (2001) Roles of glucitol in the GutR-mediated transcription activation process in *Bacillus subtilis*: tight binding of GutR to its binding site. *J Biol Chem*, **276**, 9620-9625.
193. Yasueda, H., Kawahara, Y. and Sugimoto, S. (1999) *Bacillus subtilis* yckG and yckF encode two key enzymes of the ribulose monophosphate pathway used by methylotrophs, and yckH is required for their expression. *J Bacteriol*, **181**, 7154-7160.
194. Inaoka, T., Takahashi, K., Yada, H., Yoshida, M. and Ochi, K. (2004) RNA polymerase mutation activates the production of a dormant antibiotic 3,3'-neotrehalosadiamine via an autoinduction mechanism in *Bacillus subtilis*. *J Biol Chem*, **279**, 3885-3892.
195. Daniel, R.A., Haiech, J., Denizot, F. and Errington, J. (1997) Isolation and characterization of the lacA gene encoding beta-galactosidase in *Bacillus subtilis* and a regulator gene, lacR. *J Bacteriol*, **179**, 5636-5638.
196. Freisenhausen, H.D., Frahm, H., Cabrijan, T. and Wiethold, G. (1976) The development of a radioimmunoassay for arginine vasopressin. *Acta Endocrinol (Copenh)*, **83**, 50-63.
197. Asai, K., Ishiwata, K., Matsuzaki, K. and Sadaie, Y. (2008) A viable *Bacillus subtilis* strain without functional extracytoplasmic function sigma genes. *J Bacteriol*, **190**, 2633-2636.
198. Inacio, J.M. and de Sa-Nogueira, I. (2007) trans-Acting factors and cis elements involved in glucose repression of arabinan degradation in *Bacillus subtilis*. *J Bacteriol*, **189**, 8371-8376.
199. Inacio, J.M., Costa, C. and de Sa-Nogueira, I. (2003) Distinct molecular mechanisms involved in carbon catabolite repression of the arabinose regulon in *Bacillus subtilis*. *Microbiology*, **149**, 2345-2355.
200. Monedero, V., Boel, G. and Deutscher, J. (2001) Catabolite regulation of the cytochrome c550-encoding *Bacillus subtilis* cccA gene. *J Mol Microbiol Biotechnol*, **3**, 433-438.

201. Puri-Taneja, A., Paul, S., Chen, Y. and Hulett, F.M. (2006) CcpA causes repression of the *phoPR* promoter through a novel transcription start site, P(A6). *J Bacteriol*, **188**, 1266-1278.
202. Kim, H.J., Roux, A. and Sonenshein, A.L. (2002) Direct and indirect roles of CcpA in regulation of *Bacillus subtilis* Krebs cycle genes. *Mol Microbiol*, **45**, 179-190.
203. Yamamoto, H., Serizawa, M., Thompson, J. and Sekiguchi, J. (2001) Regulation of the *glv* operon in *Bacillus subtilis*: YfiA (GlvR) is a positive regulator of the operon that is repressed through CcpA and *cre*. *J Bacteriol*, **183**, 5110-5121.
204. Strauch, M.A. (1995) AbrB modulates expression and catabolite repression of a *Bacillus subtilis* ribose transport operon. *J Bacteriol*, **177**, 6727-6731.
205. Asai, K., Baik, S.H., Kasahara, Y., Moriya, S. and Ogasawara, N. (2000) Regulation of the transport system for C4-dicarboxylic acids in *Bacillus subtilis*. *Microbiology*, **146 ( Pt 2)**, 263-271.
206. Wray, L.V., Jr., Pettengill, F.K. and Fisher, S.H. (1994) Catabolite repression of the *Bacillus subtilis* *hut* operon requires a cis-acting site located downstream of the transcription initiation site. *J Bacteriol*, **176**, 1894-1902.
207. Darbon, E., Servant, P., Poncet, S. and Deutscher, J. (2002) Antitermination by GlpP, catabolite repression via CcpA and inducer exclusion triggered by P-GlpK dephosphorylation control *Bacillus subtilis* *glpFK* expression. *Mol Microbiol*, **43**, 1039-1052.
208. Shivers, R.P. and Sonenshein, A.L. (2005) *Bacillus subtilis* *ilvB* operon: an intersection of global regulons. *Mol Microbiol*, **56**, 1549-1559.
209. Presecan-Siedel, E., Galinier, A., Longin, R., Deutscher, J., Danchin, A., Glaser, P. and Martin-Verstraete, I. (1999) Catabolite regulation of the *pta* gene as part of carbon flow pathways in *Bacillus subtilis*. *J Bacteriol*, **181**, 6889-6897.

210. Repizo, G.D., Blancato, V.S., Sender, P.D., Lolkema, J. and Magni, C. (2006) Catabolite repression of the citST two-component system in *Bacillus subtilis*. *FEMS Microbiol Lett*, **260**, 224-231.
211. Choi, S.K. and Saier, M.H., Jr. (2005) Regulation of sigL expression by the catabolite control protein CcpA involves a roadblock mechanism in *Bacillus subtilis*: potential connection between carbon and nitrogen metabolism. *J Bacteriol*, **187**, 6856-6861.
212. Grundy, F.J., Turinsky, A.J. and Henkin, T.M. (1994) Catabolite regulation of *Bacillus subtilis* acetate and acetoin utilization genes by CcpA. *J Bacteriol*, **176**, 4527-4533.
213. Tobisch, S., Zuhlke, D., Bernhardt, J., Stulke, J. and Hecker, M. (1999) Role of CcpA in regulation of the central pathways of carbon catabolism in *Bacillus subtilis*. *J Bacteriol*, **181**, 6996-7004.
214. Kruger, S., Stulke, J. and Hecker, M. (1993) Catabolite repression of beta-glucanase synthesis in *Bacillus subtilis*. *J Gen Microbiol*, **139**, 2047-2054.
215. Bryan, E.M., Beall, B.W. and Moran, C.P., Jr. (1996) A sigma E dependent operon subject to catabolite repression during sporulation in *Bacillus subtilis*. *J Bacteriol*, **178**, 4778-4786.
216. Martin-Verstraete, I., Stulke, J., Klier, A. and Rapoport, G. (1995) Two different mechanisms mediate catabolite repression of the *Bacillus subtilis* levanase operon. *J Bacteriol*, **177**, 6919-6927.
217. Lin, J.S. and Shaw, G.C. (2007) Regulation of the kduID operon of *Bacillus subtilis* by the KdgR repressor and the ccpA gene: identification of two KdgR-binding sites within the kdgR-kduI intergenic region. *Microbiology*, **153**, 701-710.
218. Mekjian, K.R., Bryan, E.M., Beall, B.W. and Moran, C.P., Jr. (1999) Regulation of hexuronate utilization in *Bacillus subtilis*. *J Bacteriol*, **181**, 426-433.
219. Nentwich, S.S., Brinkrolf, K., Gaigalat, L., Huser, A.T., Rey, D.A., Mohrbach, T., Marin, K., Puhler, A., Tauch, A. and Kalinowski, J. (2009) Characterization

- of the LacI-type transcriptional repressor RbsR controlling ribose transport in *Corynebacterium glutamicum* ATCC 13032. *Microbiology*, **155**, 150-164.
220. Fradrich, C., March, A., Fiege, K., Hartmann, A., Jahn, D. and Hartig, E. (2012) The transcription factor AlsR binds and regulates the promoter of the alsSD operon responsible for acetoin formation in *Bacillus subtilis*. *J Bacteriol*, **194**, 1100-1112.
221. Jin, S. and Sonenshein, A.L. (1994) Transcriptional regulation of *Bacillus subtilis* citrate synthase genes. *J Bacteriol*, **176**, 4680-4690.
222. Bachem, S. and Stulke, J. (1998) Regulation of the *Bacillus subtilis* GlcT antiterminator protein by components of the phosphotransferase system. *J Bacteriol*, **180**, 5319-5326.
223. Steinmetz, M., Le Coq, D. and Aymerich, S. (1989) Induction of saccharolytic enzymes by sucrose in *Bacillus subtilis*: evidence for two partially interchangeable regulatory pathways. *J Bacteriol*, **171**, 1519-1523.
224. Tortosa, P. and Le Coq, D. (1995) A ribonucleic antiterminator sequence (RAT) and a distant palindrome are both involved in sucrose induction of the *Bacillus subtilis* sacXY regulatory operon. *Microbiology*, **141 ( Pt 11)**, 2921-2927.
225. Schnetz, K., Stulke, J., Gertz, S., Kruger, S., Krieg, M., Hecker, M. and Rak, B. (1996) LicT, a *Bacillus subtilis* transcriptional antiterminator protein of the BglG family. *J Bacteriol*, **178**, 1971-1979.
226. Beijer, L., Nilsson, R.P., Holmberg, C. and Rutberg, L. (1993) The glpP and glpF genes of the glycerol regulon in *Bacillus subtilis*. *J Gen Microbiol*, **139**, 349-359.
227. Nilsson, R.P., Beijer, L. and Rutberg, B. (1994) The glpT and glpQ genes of the glycerol regulon in *Bacillus subtilis*. *Microbiology*, **140 ( Pt 4)**, 723-730.
228. Миронов, А.А., Винокурова, Н.П. and Гельфанд, М.С. (2000) Программное обеспечение анализа бактериальных геномов. *Молекулярная биология*, **34**, 222-231.

229. Saier, M.H., Jr., Yen, M.R., Noto, K., Tamang, D.G. and Elkan, C. (2009) The Transporter Classification Database: recent advances. *Nucleic Acids Res*, **37**, D274-278.
230. Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res*, **37**, D233-238.
231. Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*, **8**, 785-786.
232. Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.
233. Felsenstein, J. (1996) Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol*, **266**, 418-427.
234. Huson, D.H., Richter, D.C., Rausch, C., DeZulian, T., Franz, M. and Rupp, R. (2007) Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics*, **8**, 460.
235. Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res*, **14**, 1188-1190.
236. Wilson, D., Charoensawan, V., Kummerfeld, S.K. and Teichmann, S.A. (2008) DBD--taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Res*, **36**, D88-92.
237. Ulrich, L.E. and Zhulin, I.B. (2010) The MiST2 database: a comprehensive genomics resource on microbial signal transduction. *Nucleic Acids Res*, **38**, D401-407.
238. Sierro, N., Makita, Y., de Hoon, M. and Nakai, K. (2008) DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information. *Nucleic Acids Res*, **36**, D93-96.
239. Novichkov, P.S., Laikova, O.N., Novichkova, E.S., Gelfand, M.S., Arkin, A.P., Dubchak, I. and Rodionov, D.A. (2010) RegPrecise: a database of

- curated genomic inferences of transcriptional regulatory interactions in prokaryotes. *Nucleic Acids Res*, **38**, D111-118.
240. Sumiya, M., Davis, E.O., Packman, L.C., McDonald, T.P. and Henderson, P.J. (1995) Molecular genetics of a receptor protein for D-xylose, encoded by the gene xylF, in *Escherichia coli*. *Receptors Channels*, **3**, 117-128.
241. Rosey, E.L., Oskouian, B. and Stewart, G.C. (1991) Lactose metabolism by *Staphylococcus aureus*: characterization of lacABCD, the structural genes of the tagatose 6-phosphate pathway. *J Bacteriol*, **173**, 5992-5998.
242. Gama-Castro, S., Salgado, H., Peralta-Gil, M., Santos-Zavaleta, A., Muniz-Rascado, L., Solano-Lira, H., Jimenez-Jacinto, V., Weiss, V., Garcia-Sotelo, J.S., Lopez-Fuentes, A. *et al.* (2011) RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Gensor Units). *Nucleic Acids Res*, **39**, D98-105.
243. Lozada-Chavez, I., Janga, S.C. and Collado-Vides, J. (2006) Bacterial regulatory networks are extremely flexible in evolution. *Nucleic Acids Res*, **34**, 3434-3445.
244. Maslov, S., Krishna, S., Pang, T.Y. and Sneppen, K. (2009) Toolbox model of evolution of prokaryotic metabolic networks and their regulation. *Proc Natl Acad Sci U S A*, **106**, 9743-9748.
245. Setlow, B., Cabrera-Hernandez, A., Cabrera-Martinez, R.M. and Setlow, P. (2004) Identification of aryl-phospho-beta-D-glucosidases in *Bacillus subtilis*. *Arch Microbiol*, **181**, 60-67.
246. Zhang, J. and Aronson, A. (1994) A *Bacillus subtilis* bglA gene encoding phospho-beta-glucosidase is inducible and closely linked to a NADH dehydrogenase-encoding gene. *Gene*, **140**, 85-90.
247. Morinaga, T., Ashida, H. and Yoshida, K. (2010) Identification of two scyllo-inositol dehydrogenases in *Bacillus subtilis*. *Microbiology*, **156**, 1538-1546.
248. Verma, S.C. and Mahadevan, S. (2012) The chbG gene of the chitobiose (chb) operon of *Escherichia coli* encodes a chitooligosaccharide deacetylase. *J Bacteriol*, **194**, 4959-4971.

## Приложения

### Приложение 1. Положение потенциальных сайтов связывания белка AraR.

Указано положение сайта относительно старта трансляции гена.

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>B. subtilis</i>	<i>xsa</i>	-172	5,37	AAATACATACGTACAAATAT
	<i>araABDLMNPQ-abfA</i>	-104	5,5	AAAATTGTTTCGTACAAAATA
		-60	4,33	TCATTAGTACGTATCTTTTG
	<i>abnA</i>	-158	5,09	TTTTTTGTCTGTACAAATTA
	<i>araE</i>	-189	5,16	AAATGTATACGGACAAATTT
		-108	5,85	ATATTTGTACGTACTAATTA
-65		5,37	ATATAAGTACGTACAATGA	
<i>araR</i>	-92	5,37	TCAATTGTACGTACTTATAT	
	-49	5,85	TAATTAGTACGTACAAATAT	
<i>B. amyloliquefaciens</i>	<i>xsa</i>	-168	5,45	TAATTGGTTCGTACAAATTA
	<i>araABDLMNPQ-abfA</i>	-98	5,7	AAAATTGTTTCGTACAAATAA
		-56	5,19	ATATTCGTACGTATATTTTT
	<i>abnA</i>	-104	5,07	TTTTTTGTCTGTACAAATAT
	<i>araE</i>	-151	5,25	TTATTTGTGCGTACTAATTT
		-108	5,48	TCATAAGTACGTACAAATGT
<i>araR</i>	-168	5,48	ACATTTGTACGTACTTATGA	
	-125	5,25	AAATTAGTACGCACAAATAA	
<i>B. pumilus</i>	<i>araKDAE</i>	-101	5,89	ACATATGTACGTACAAATAT
	<i>araR</i>	-192	5,89	ATATTTGTACGTACATATGT
<i>B. licheniformis</i>	<i>araR</i>	-191	5,65	TTTTTTGTACGTACATATGT
		-159	4,97	TAATATGTTTCATATAAATAT
	<i>araKDAE</i>	-139	4,97	ATATTTATATGAACATATTA
		-107	5,65	ACATATGTACGTACAAAAAA
<i>araABDMNPQ-abfA</i>	-124	5,59	TTAGTTGTACGTACATATAA	
<i>A. flavithermus</i>	<i>araR</i>	-82	5,74	AAATATATACGTACAAATAA
	<i>araDBA-abf5-abfA</i>	-55	5,61	AAAAATGTACGTACAAATTT
<i>G. kaustophilus</i>	<i>araDBA</i>	-60	5,53	AAAATTGTACGTACAATAGT
	<i>araR</i>	-74	4,9	ATAATGATACGGACAAATAA
	<i>araFGH</i>	-35	5,56	TTAAATATACGTACAAATTT
<i>B. halodurans</i>	<i>BH1061</i>	-173	5,46	TTATTCGTACGAACAAATAA
	<i>abfA-araM</i>	-199	5,95	AAATTTGTACGTACAAAAAA
		-101	5,42	TATTATGTACGTACAAGTTA
	<i>araDBA-xsa</i>	-72	5,59	AAAATTGTACGTACAAGTGT
	<i>araR</i>	-37	5,14	TCATTTGTACGTATAAGTTG
<i>araPQ-yteU-araN</i>	-49	4,40	AAACGTGTACGTACACGATA	
<i>O. iheyensis</i>	<i>araKDAE</i>	-258	5,17	ATACTTGTATGTACATATTT
		-151	5,74	TAATTTGTACGTATATATAT
		-118	5,3	ACAAATGTACGTACAATAAT

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>O. iheyensis</i>	<i>araR</i>	-204	5,3	АТТАТТГТАСГТАСАТТТГТ
		-171	5,74	АТАТАТАТАСГТАСАААТТА
		-64	5,17	АААТАТГТАСАТАСААГТАТ
<i>Paenibacillus</i> sp. JDR-2	<i>araA</i>	-84	5,5	ТААГТТГТАСГТАСААСТТА
<i>Paenibacillus</i> sp. JDR-2	<i>araD</i>	-61	5,22	АААСАТГТАСГТАСААГТГТ
	<i>araKR</i>	-99	4,73	ТСАТАТГТАСАТАСААГТТА
<i>L. brevis</i>	<i>araEKDA</i>	-182	6,06	ТГАТТТАТГСГТАСАААТАТ
<i>L. fermentum</i>	<i>araA</i>	-104	5,87	ТГААТТГТАСГГАСАААТАТ
	<i>araKDRE</i>	-76	5,86	ТГАТТТАТАСГГАСАААААА
<i>L. plantarum</i>	<i>araEKDA</i>	-165	6,4	ТГАТТТАТАСГТАТАААТАТ
	<i>araR</i>	-233	5,11	ТТАТТГАТГСГГТААААТТГ
<i>L. reuteri</i>	<i>araEKDA</i>	-104	6,26	ТГАТТТАТССГТАСАААТАТ
<i>L. sakei</i>	<i>araEKDA</i>	-249	5,62	АТААТТАТТССГААТАААТАА
<i>L. citreum</i>	<i>araAEKD</i>	-127	5,99	ТГАСТТАТАСГГАСАААТАТ
<i>L. mesenteroides</i>	<i>araAEKD</i>	-137	5,63	ТГАСТТАТАСГГАСАААТГТ
<i>O. oeni</i>	<i>araKDAE</i>	-180	4,98	АААТТСАТАСГТААААСТАА
		-150	4,9	ТГАТТТАТАТГТАСГААТГТ
<i>P. pentosaceus</i>	<i>araEKDA</i>	-90	5,99	ТГАСТТГТССГТАТАААТАТ
<i>C. cellulolyticum</i>	<i>araA-II-araKD</i>	-81	5,41	АААСТТГТАСГТААСГАТАТ
	<i>araR</i>	-36	5,64	ТААГАТГТАСГТАСААГТТГ
<i>C. beijerincki</i>	<i>Cbei_4451-araFGH</i>	-309	5,06	САТГААГТАСГТАСААСАТА
	<i>araRD-tal-akt-araK</i>	-321	5,54	ААТГТТГТАСАТАСААГАТГ
		-98	5	САТСААГТГСГТАСААСТТТ
	<i>araA</i>	-261	5,65	САТСТТГТАТГТАСААСАТТ
		-228	5,02	ААГСТТГТАСАТАСАТСААА
	<i>araT1-araT2-araT2-araJ</i>	-51	5,12	ГАТГТГГТАСГТАСАТГТТТ
	<i>araI-araT3-araT4</i>	-266	5,04	ТТТСАТГТАТГТАСААГАТАТ
<i>epiA</i>	-91	4,9	АТАСТТГТАСАТАСАТГААА	
<i>Clostridium</i> sp. SS2/1	<i>araRAKD-epiA-araJ-araT4T1T2T2A-II</i>	-56	5,07	АААГААГГАССГТАСААГТТТ
<i>C. acetobutylicum</i>	<i>araE</i>	-214	5,38	СААТТСАТАСГТАТААААТС
	<i>araR</i>	-214	5,69	АТТТТСАТАСГТАТАААТТС
		-87	6,36	АААТТТАТАСГТАТСААТАТ
	<i>araDA</i>	-304	5,69	ГААТТТАТАСГТАТГААААТ
	<i>ptk</i>	-187	6,54	АААТТТАТАСГТАСАААТТА
	<i>araKEA-tal-akt-epiA</i>	-104	5,13	АААТААГГТАСГТАССАТТТТ
	<i>arb43-araT</i>	-61	5,94	ТААТАТГТАСГТАТАТАТТА
		-202	6,76	АААТТТАТАСГТАТАААТТА
<i>abf3</i>	-301	5,52	АТАТТТАТАСГТГТТААТТТ	
<i>T. maritima</i>	<i>araA</i>	-163	4,81	АТААСГГТАСГТАСГАТГГ
	<i>araN-III-araP-III-araQ-III</i>	-61	6,48	АТААТГГТАСГТАСТТТТАТ
	<i>glsA-abfA-epiA-araD-araB-II-araM</i>	-45	5,2	ААТААТГТАСГТАССТАААТ
<i>Thermotoga</i> sp. RQ2	<i>araN-II-araP-II-araQ-II-abf3-xylX-abfA-epiA-araD-araB-II-araM</i>	-49	6,16	АААТТГГТАСГТАССТАТСТ
	<i>araN-III-araP-III-araQ-III</i>	-149	6,63	АТААААГТАСГТАССТТТТАТ
	<i>glsA</i>	-57	5,2	ААТААТГТАСГТАССТАААТ

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>Thermotoga</i> sp. RQ2	<i>araA</i>	-63	6,63	АТАААGGTACGTACTTTTAT
	<i>abf4-lamG1-abf4-conA1-conA2-lamG2</i>	-222	5,69	AGATAGGTACGTACCAATTT
		-39	5,35	АТАТАТGTACGTACAAAATG
<i>T. neapolitana</i>	<i>araN-II-&lt;tan1&lt;-araP-II-araQ-II-glsA-abfA-epiA-araD-araB-II-araE</i>	-135	6,34	АСAAAAGTACGTACCTTTAT
	<i>araA</i>	-63	6,34	АТАААGGTACGTACTTTTGT
<i>T. petrophila</i>	<i>glsA</i>	-56	5,05	GATAATATACGTACCTAAAT
	<i>araN-II-araP-II-araQ-II-abf3-xylX-abfA-epiA-araD-araB-II-araM</i>	-234	5,35	САТТТТGTACGTACATATAT
	<i>araA</i>	-51	5,69	АААТТGGTACGTACCTATCT
		-63	6,4	АТАААGGTACGTACTTTTAT
	<i>abf4-lamG1-abf4-conA1-conA2-lamG2</i>	-22	5,69	AGATAGGTACGTACCAATTT
		-39	5,34	АТАТАТGTACGTACAAAATG
<i>araN-III-araP-III-araQ-III</i>	-151	6,4	АТАААAGTACGTACCTTTAT	
<i>T. naphthophila</i>	<i>glsA-abfA-epiA-Tnap_0913-14-araD-araB-II-araM</i>	-56	5,05	GATAATATACGTACCTAAAT
	<i>araA</i>	-63	6,4	АТАААGGTACGTACTTTTAT
	<i>araN-III-araP-III-araQ-III</i>	-151	6,4	АТАААAGTACGTACCTTTAT
<i>T. lettingae</i>	<i>araRD</i>	26	4,88	АТТТАТGTACGTACGTTTAC
	<i>araN-II-araP-II-araQ-II-xynB-abf3-abfA-araA-II-araB-II</i>	-170	5,2	ТААТТТGTATGTACATATAT

## Приложение 2. Поиск потенциальных сайтов связывания AgaR методом филогенетического футпринтинга.

А. Множественное выравнивание промоторных областей оперонов *agaSYBCDI* из *Escherichia coli str. K-12*, *Escherichia coli C str. ATCC 8739*, *Citrobacter koseri ATCC BAA-895*, *Enterobacter sp. 638* и *Enterobacter cancerogenus ATCC 35316*.

```
EC_agaS      CCAGCAATCCCTTTTGCTTCCTTTATCTTTTCTTTCAACGATCACAAATTCGTTTTATTCTTTTTCTCCA
EcolC_0562  TACGCAATTCTTTAATCTTCCSTTGTCTTTTCTTTCAATGATCACAGATTCGTTTTATTCTTTTTCTCCA
SKO_04536   TTAT-ATATCAATGAGATAACTACCCSTTCGTTTCATCGATCACATTTTCGTTTTATTCTATTTTGCATCA
Ent638_3577 TAA--AAATTAATGTGTACTAAAACCGTTATCTCCCCGCTCACAGTTTCGTTTTATTCTTTTTCTTGGTA
ENTCAN_04493 CTTTCTTTTCTTTTCTTCATTTTTTGTCTTTTCTTTTTCGATCACACTTTTCGTTTTATTCTTTCTTTTCGC-A
                *      *                               * * * * * * * * * * * * * * *
```

```
EC_agaS      TTGAACCTTCAGTTTTCTTTCTATAGATTTTAATCAACGAAAGACATCACCAAGTGAAATGAAACGAAAGGCA
EcolC_0562  TTGAACCTTCGTTTTCTTTCTATAGATTTTAATCAACGAAAGACATCACAAGTGAAATGAAACGAAAGGCA
SKO_04536   TTGAACCTTCGATTTCTTTCTATAGATTTTAATTTAAGTGAACAGCAGAACAAGTGAAACAAAACGAAAG-TC
Ent638_3577 TTGATCTTTCACCTTCTTTCTATAGATTTTAATTAATC-----GAACAAGTGAAACGAAACGAAAGACA
ENTCAN_04493 TTGAACCTTCGATTTCTTTCTATAAATTTTATTTCAACTGCTAAAGTGAAACAAGTGAAACGAAACGAAAGATA
                * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

->agaS

```
EC_agaS      AGTGAAAGC--GACAACGCCCGACGTCAAGTTCATCAGACTAAGGATTGAGTTATGCCAGAAAAT
EcolC_0562  AGTGAAAGC--GACAACGCCCGACGTCAAGTTCATCAGACTAAGGATTGAGTTATGCCAAAAAAT
SKO_04536   AAAGGCACAATACCAACGGCTGA-GTGGAGTTCACAAGACTAAGGACTATGTTATGCCAGAAAAT
Ent638_3577 ATGGCGATTACAATCGCGTCTGATGTGGATTTCACATTACTAAGGACAGAGTTATGCCAGAAAAC
ENTCAN_04493 GCGACACGTTTGCCAGCGTCTGATGTGGAATTCACACGACTAAGGACTGAGTTATGCCAGAAAAT
                * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

Потенциальные сайты связывания AgaR (i) типа выделены красным текстом. Регионы, экспериментально определенные, как защищенные от действия ДНКазы I в *E. coli* K12 выделены желтым. Экспериментально определенные элементы промоторов «-35» и «-10» в *E. coli* K12 подчеркнуты. Старт трансляции выделен жирным шрифтом.

Б. Множественное выравнивание межгенных областей дивергонов *agaR<>agaZVWEFA* из *Escherichia coli str. K-12*, *Citrobacter koseri ATCC BAA-895*, *Enterobacter sp. 638*, и *Aeromonas hydrophila ATCC 7966*.

```
b3132      AAACCTTCGTTTTCATTTTCGTTTTGCTTATTAACGCCTTTCTATTAAAGCAAATGCAAGCCACCTTGCCCATGACGCAAGCTAC
SKO_04530  AATCTTTCGTTTTCATTTTCGATTTTCTATTAACGCCTTTCTTTTACTGGAATGCAAGCCCC-----ATACCTGCAAGCCAT
Ent638_3572 AATCTTTCGTTTTCATTTTCGATTTATCTATTAACCTCTTTCTTTTAGCGGAATGCAAGCCCCACTGGCGTGTCTCTGGCGGAGGT
ANA_0812   GAGCTTTCACCTTCGTTACGTTT-CCCTATTAACACTCAGCCTCGATGGCAAATCAAGCTTTCATTTGAATTGAAAAGTGCCGCTCA
                * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

```
b3132      TCTCGTTTCA--G-----TGACTTTTCATTATGTTTTCTTTTGTGAATCAGATCAGAAAACCATTATCTTTCGTTTTATTTTT
SKO_04530  AACGACGTATCGGACTG--AAACCTTTCATTATGTTTTCTTTTGTGAAACAGATCCGAAAACCAATATCTTTCGTTTTATTTTT
Ent638_3572 ----TGCCG--GTGCTG-AAACCTTTCGTTATGTTTTCTTTTGTGAAACAGATCCGAAAACATATATCTTTCGTTTTATTTTT
ANA_0812   GACTCCSTA AAAAGGCCACAACACTTTCGATTCATTTCTTTTGTGATCAGATGTGAACACGCTTTTATTTTCGTTTCG-----
                * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

->agaZ

```
b3132      ATCTCACCATGACGC-AGTATCAACTGAAACAAAACGAAAGATTAAATATCGCAGTAA-TCTGAACTGGAGAGGAAAAGTG
SKO_04530  ATAGCGTCTTGCACC-AGTATCAGATGAAACAAAACGAAAGACACTTGTCTTAAGCA-TTCAACCTGGAGAGGAAAAGTG
Ent638_3572 ACAGCACCATGCAGC-AGTATAAAGTGAAATAAAAACGAAAGGTGAATACCATAACCA-TCCGAAGTGGAGAGGAAAAGTG
ANA_0812   --AACGCCGTACGCCTAGGGTAA--AGAAAGAAATCGAAAGCAAAGCATCTTGTTCAGGCCAAGAGGAAAGGGAAAAGTG
                * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

Потенциальные сайты связывания AgaR (i) типа выделены красным текстом. Регионы, экспериментально определенные, как защищенные от действия ДНКазы I в *E. coli* K12 выделены желтым. Старт трансляции выделен жирным шрифтом.





3. Множественное выравнивание промоторных областей оперонов *agaR1-agaZS-PTS-II-agaAY* из *Photobacterium profundum SS9*, *Vibrio angustum S14*, *Vibrio vulnificus CMCP6*, *Vibrio sp. MED222* и *Photobacterium sp. SKA34*

```

PBPRB1036   AAAAAAGAAAGGCAACACCTTCAGTTTAAGCAAACCGAAAGGGATTTACACTTAAAAACGAAAGAACAACCA
VV21025     TTGAGTGAAAGTAAAACCTTCACTTTGCGTGAACGAAAGGTTTTTTGACCTAAAAACGAAAGGTTACCCA
MED222_10308 ATAAGTGAAAGTAAATACTTTCACTTTATCCTAAAGTGAAAGGTTTTTTAGCCTGAAATGAAAGCTAATTGG
VAS14_15564 CAAAGTGAAAG-AAAAGCCTTCACTCTCACCCAAACGAAAGAACAAAAACCAACATACCGAAAGAAAAACAC
SKA34_07923 CAAAGTGAAAG-AAAAGCCTTCACTTTACCCAAAGTGAAAGACCAAACCAAAACAATCGAAAGGAAAACCTT
          *   * * * * *   * *   * * * * *   * *   * * * * *   * *   * * * * *   *
                                     -10
PBPRB1036   CTGTGATCTAGCTTTAACTTTTAT---TTACATTGCCCTGCTTTAATACAAATCAAAAGAAA-GCGAAACATTTT
VV21025     CTGTGATCTCCCTTTCACTTTT---CATTTTTTAGCCCTGTTTAAATAACAACGAAAGGAGA-GCGAAAGCTTTC
MED222_10308 CTGTGATCTAGGTTTCACTTTTTT--ATTATTTG-TCCTGCTTTAATACCAACCAAAAGGAAAGTCGAAACATTTT
VAS14_15564 TTGTGATCCACCTTTCAATTTTTATCCATTATTTA---CTGCTTTAATCCAAATCAAAAGAAA-TCGAAACACTTC
SKA34_07923 TTGTGATCCACCTTTCAATTTTTATCCATTATTTA---CTGCTTTAATCCACACCAAAAGAAA-TCGAAACACTTC
          * * * * * *   * * *   * * * *   * *   *   * * * * * * *   *   * * * * * * *   * * * * * * *
                                     ->agaR1
PBPRB1036   G-ACTCAATAACGAAATGAAACAAAAGATGAGGTAAAGCATGACT
VV21025     T-GATGGAAAGTGAAAGGTAACGAAAGGCGA-GTAAGTG-ATGACC
MED222_10308 A-TGCTTAAAGTGAAATGAAACGTAAGAGGT-TTC-GTATATGAGC
VAS14_15564 ACTCCTTGAATTGAAATGTAATGAAAGGGAA-ATAAGTAAATGACC
SKA34_07923 ACTCCTTGAATTGAAATGTAACGAAAGGGGA-ATAGGTAAATGACC
          *   * * * * *   * *   * * *   *   *   *   *   *   *   *   *   *   *

```

Потенциальные сайты связывания AgaR (v) типа выделены красным текстом. Дополнительные вероятные полусайты AgaR выделены фиолетовым. Потенциальный элемент промотора «-10» выделен подчеркиванием. Старт трансляции выделен жирным шрифтом.

И. Множественное выравнивание промоторных областей генов *agaS* из *Stenotrophomonas maltophilia K279a* и *Dehalococcoides sp. VS*.

```

DeVSDRAFT_0296 GTTAAAGCGGTAACGAAAGCCATCTTCGCTGCAGCTTTCGCTTTATCGCGCGTGCCCGC
Smlt4434       GTTAAAGCATTTGTGCGATA-CGGCATTGGCG-AATTCTTGCCTTTTC-CGACGTGCTC-T
          * * * * *   * * * *   * * * *   * * * *   * * * *   * * * *   *
DeVSDRAFT_0296 ACCCGAAGCCCGTTCGGC----CAATTTCTTTTGTTTTCTTTCGATAATTGACTTTCCGAAAGAAAAAC
Smlt4434       GGGCAGAGCGAAGTCGGCGTCACATTTTCTTTCTTTTCTTTCCGTATTTGACATTTCCGAAAGAAAAAC
          *   * * *   * * * *   * * * * * * * * * * * * * * * * * * * * * * * * *
                                     ->agaS
DeVSDRAFT_0296 TAAGATCGTGCCGTCGCTCAACAGGAATCTTT-CATGGACGCCACTCCGCT
Smlt4434       ACAGATGGTGCCTGCGCCCAACTGGAACCTCCGAATGGACGTCACCTGCT
          * * * * *   * * * *   * * * *   * * * *   * * * *   * * * *   * * * *

```

Потенциальные сайты связывания AgaR (iv) типа выделены красным текстом. Старт трансляции выделен жирным шрифтом.

К. Промоторная область оперона *agaP-II-agaZS-agaK-II-agaR* из *Burkholderia cenocepacia J2315*.

```

>BCAS0463
TTTCGATTTCTTTCGTTTGTGGTTTACACTAATAACGAAAGGAAACGAAAGATCATAGAT
TGGCCGCTCCCGACGGACGTTTCTGCCGATCCCTGCTGGAACGTCGTCGGGCTCGCCCGC
CGAACGCCTCCGATTTGGCCTTGACGGCATCAAACGAAAGCATTCGACGGATCAGAACC
CACTTCGAAAGACGCGACCGATGCAA

```

Потенциальный сайт связывания AgaR (i) типа выделен красным текстом. Старт трансляции выделен жирным шрифтом.

Л. Промоторная область оперона *agaRS-PTS-V-bgl-agaY-II* из *Haemophilus parasuis* Sh0165.

>HAPS\_0195

ТТАСТGAAAGTTCATCATTCATТАСТТТGATTATATCACCATAACTCAAAATGATAGTAACCATTTTGAGCACGAATACA  
ACAATCАСТТТATTCGTGGTATAGAAGTATAGTAGGAAACTTTCCAAGCCCAAATТАCTATAAACGTGACAAAACATTTTC  
TTGCTTAAAAACCCAAATGTAAATTTTGTGATCTTTATCTCA**TTTTCGAATATTTTTCGATTGTTTTTGTAAAAAGGGGT**  
ATAAACTТАСТТТCGAAAGCAATCGTAGTATGCAGTGGTTGTTTTGATGTAGGAGAGAAA**ATG**AAAAGTA

Потенциальный сайт связывания AgaR (ii) типа выделен красным текстом. Старт трансляции выделен жирным шрифтом.

**Приложение 3. Положение потенциальных сайтов связывания белка HexR.**  
Указано положение сайта относительно старта трансляции гена.

Геном	Оперон	Сайт			
		Положение	Вес	Последовательность	
<b>Enterobacteriales</b>					
<i>E. coli</i> str. K-12 substr. MG1655	<i>zwf</i>	-63	5,4	ccGTAAGAaaaATTACAA	
	<i>hexR-pykA</i>	-291	5,4	TTGTAATTTTtcTTACgg	
	<i>ybfA</i>	-165	5,5	gTGTAAcTTAATTACAg	
<i>S. typhimurium</i> LT2	<i>zwf</i>	-61	5,7	gTGTAAaAAAATTACAA	
	<i>hexR-pykA</i>	-297	5,7	TTGTAATTTtATTACAc	
	<i>ybfA</i>	-188	4,9	TTGTAAcTAAtTTACgc	
<i>K. pneumoniae</i> MGH 78578	<i>zwf</i>	-61	5,6	gTGTAAgaAAAATTACAA	
	<i>hexR-pykA</i>	-293	5,6	TTGTAATTTtctTACAc	
<i>Enterobacter</i> sp. 638	<i>zwf</i>	-62	5,8	TTGTAATaAAAATTACAA	
	<i>hexR-pykA</i>	-313	5,8	TTGTAATTTtATTACAA	
	<i>ybfA</i>	-162	5,4	TcGTAAcTTAATTACAg	
<i>Y. pestis</i> KIM	<i>zwf</i>	65	5,7	aTGTAAgaAAAATTACAA	
	<i>hexR</i>	-173	5,7	TTGTAATTTtctTACAt	
	<i>ybfA</i>	-139	5,1	acGTAAcaAAAATTACgA	
<i>S. proteamaculans</i> 568	<i>zwf-eda</i>	-40	5,7	aTGTAAgaAAAATTACAA	
	<i>hexR-pykA</i>	-329	5,7	TTGTAATTTtctTACAt	
	<i>ybfA</i>	-189	5,2	gTGTAAcaAAAATTACgt	
<i>E. carotovora</i> SCRI1043	<i>zwf-eda</i>	-39	5,7	cTGTAAaAAAATTACAg	
	<i>hexR</i>	-366	5,7	cTGTAAATTTtATTACAg	
<i>P. mirabilis</i> HI4320	<i>hexR</i>	-329	5,6	TTGTAATTTtgTTACAt	
	<i>zwf</i>	-42	5,6	aTGTAAcaAAAATTACAA	
<i>P. luminescens</i> TTO1	<i>hexR</i>	-320	5,6	TTGTAATTTtgTTACAg	
	<i>zwf</i>	-41	5,6	cTGTAAcaAAAATTACAA	
	<i>ybfA</i>	-166	5,7	aTGTAAaAAAATTACAt	
	<i>pgl</i>	-122	5,2	aTGTAAggAtATTACAt	
<i>C. koseri</i> ATCC BAA-895	<i>zwf</i>	-63	5,8	TTGTAATAAAAATTACAA	
	<i>hexR</i>	-311	5,8	TTGTAATTTTATTACAA	
	<i>ybfA</i>	-61	5,6	ATGTAAcTTAATTACAG	
	<i>pckA</i>	-258	4,9	TCGTAAATTTCTTACAT	
<i>E. tarda</i> EIB202	<i>zwf-pgl-edd-eda</i>	-40	5,4	CTGTAAAAAAATTACAA	
	<i>ppc</i>	-151	4,8	GCGTAATTAAATTTTACAG	
	<i>hexR</i>	-329	5,4	TTGTAATTTTTTTTACAG	
<b>Vibrionales</b>					
<i>V. vulnificus</i> CMCP6	<i>aceBA</i>	-309	5,4	TGTAATTAAATTACg	
	<i>grcA</i>	-248	5,0	TGTAgTaAAtTTACA	
	<i>gpml</i>	-43	5,2	TGTAgTaAAAATTACg	
	<i>ppc</i>		-174	4,7	TGTAATaAAtTTtCA
			-27	4,9	TGTAATTTtTTTACg
	<i>pgi</i>	-142	4,7	TGAAAAaAAAATTACA	
<i>focA-pflB</i>	-306	3,9	TGAAAgcAAATTACA		

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
		-253	4,3	TGTAgtTTTtcTTtCA
	<i>pflA</i>	-45	4,8	TGTAATaAAAaTaACA
	<i>gapA2-aldE</i>	-295	5,1	gGTAgTaAAATTACA
	<i>pntAB</i>	-85	5,0	TGTAatTTTAaCTACA
	<i>pckA</i>	-240	4,4	cGTAATTAAATaACc
	<i>mtlADR</i>	-338	3,9	TtTAatTTTtATTtCA
		-261	4,2	TGTtATTAaCTTACA
	<i>ptsG</i>	-320	4,4	TGTAATTTtgTTACT
	<i>nirBDC-cysG</i>	-94	4,6	gGTAgTTTAATtCg
		-81	4,7	cGTAATTgAtTTACA
	<i>pepD</i>	-258	5,3	gGTAATTAAATTACA
<i>gltBD</i>	-505	5,2	TGTAAaTTAAATTACA	
<i>V. harveyi</i> ATCC BAA-1116	<i>pgi</i>	-137	4,7	TGaAaaaAAATTACA
	<i>ppc</i>	-143	4,8	TGTAATTAAtTTtCA
		4	4,8	TGTAgtTTTttTTACg
	<i>gapA2-aldE</i>	-56	5,1	TGTAATTTtAcTACc
	<i>gpml</i>	-90	5,2	TGTAgtTaAAATTACg
	<i>pflA</i>	-137	4,8	TGTAATaAAAaTaACA
	<i>focA-pflB</i>	-297	4,6	TGTAATcgAAATTACA
		-244	4,3	TGTAgtTTTtcTTtCA
	<i>grcA</i>	-248	5,0	TGTAgtTaAAtTTACA
	<i>aceBA</i>	-306	5,4	TGTAATTTAAATTACg
	<i>pntAB</i>	-291	4,4	cGTAgTTAAtTTtCA
		-226	4,4	aGTAtTTTtATTACA
	<i>mtlADR</i>	-211	4,4	TGTtATTAAGTTACA
	<i>ptsG</i>	-321	4,4	TGTAATTTtgTTACT
		-170	4,1	TGaAAcTAAATtCg
	<i>pepD</i>	-255	5,6	TGTAATTAAATTACA
	<i>gltBD</i>	-508	5,2	TGTAAaTAAATTACA
<i>glgX</i>	-34	4,9	aGTAATTTAAATTACA	
<i>V. parahaemolyticus</i> RIMD 2210633	<i>grcA</i>	-248	5,0	TGTAgtTaAAtTTACA
	<i>aceBA</i>	-306	5,4	TGTAATTAAATTACg
	<i>pflA</i>	-44	4,8	TGTAATaAAAaTaACA
	<i>focA-pflB</i>	-296	4,9	TGTAATgAAATTACA
		-243	4,9	TGTAgtTTTtATTtCA
	<i>gapA2-aldE</i>	-56	5,1	TGTAATTTtAcTACc
	<i>pgi</i>	-137	4,7	TGaAaaaAAATTACA
	<i>ppc</i>	-175	4,8	TGTAATTAAtTTtCA
		-29	4,8	TGTAgtTTTttTTACg
	<i>gpml</i>	-90	5,2	TGTAgtTaAAATTACg
	<i>pntAB</i>	-226	4,5	aGTAtTTAAATTACA
	<i>mtlADR</i>	-210	4,4	TGTtATTAAGTTACA
	<i>ptsG</i>	-321	4,4	TGTAATTTtgTTACT
-290		4,6	TGTAatTTTAtTTACg	
-170		4,1	TGaAAcTAAATtCg	

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
	<i>nirBDC-cysG</i>	-83	4,8	TGTAATTgAtTTACA
	<i>pepD</i>	-254	5,6	TGTAATTAAATTACA
	<i>gltBD</i>	-307	5,1	TGTAAcTAAATTACA
	<i>glgX</i>	-115	4,9	aGTAATTAAATTACA
<i>V. splendidus</i> LGP32	<i>gpml</i>	-65	5,2	TGTAgTaAAATTACg
	<i>pgi</i>	-144	4,7	TGAAAAaAAATTACA
	<i>ppc</i>	-174	5,1	TGTAATTAAATTtCA
		-29	4,9	TGTAATTTttTTACg
	<i>grcA</i>	-249	5,0	TGTAgTaAAtTTACA
	<i>pflA</i>	-42	3,8	TGTAATaAAATaAac
	<i>focA-pflB</i>	-296	4,2	TGAAAacTAATTACA
		-244	3,8	TGTAgcTTtcTTtCA
	<i>gapA2-aldE</i>	54	5,1	TGTAATTTtAcTACc
	<i>pckA</i>	-241	4,3	cGTAATaAAATcACc
	<i>ptsG</i>	-323	4,4	TGTAATTTtgTTACTt
		-173	4,1	TGAAAcTAAATTtCg
	<i>nirBDC-cysG</i>	-86	4,7	cGTAATTgAtTTACA
	<i>pepD</i>	-257	5,4	cGTAATTAAATTACA
	<i>gltBD</i>	-483	4,7	TGAAAAaAAATTACA
<i>V. fischeri</i> ES114	<i>gpml</i>	-62	5,5	TGTAgTaAAATTACA
	<i>gapA2-aldE</i>	-61	5,3	TGTAATTTtAcTACc
	<i>aceBA</i>	-294	5,5	TGTAATTTAATTACg
	<i>ppc</i>	-32	5,1	TGTAACAaAAATTACA
	<i>pntAB</i>	-86	5,3	TGTcATTTtATTACA
	<i>grcA</i>	-239	4,6	TGTAgTaAAtTTACc
	<i>focA-pflB</i>	-151	4,3	TGTAgTTTtcTTtCA
		-234	3,8	TGTAgTaAAtTgAtA
	<i>pflA</i>	-45	4,2	TaTAATaAAATaACc
	<i>mtlA1A2DR</i>	-389	4,5	TGTAATaAAAAaACA
		-161	3,9	TGTAtaaAtATTACTt
	<i>gltBD</i>	-300	4,3	TGTAggaTAgTTACA
	<i>pepD</i>	-199	5,5	TGTAATTTtATTACg
	<i>ptsHI12-crr</i>	-207	5,1	TGTAAAaTAATTACA
<i>V. salmonicida</i> LFI1238	<i>gapA2-aldE</i>	-62	5,3	TGTAATTTtAcTACc
	<i>focA-pflB</i>	-356	5,2	TGaAATTAAATTACA
		-305	4,7	TGTcATTTtATTtCA
	<i>aceBA</i>	-294	5,5	TGTAATTTtATTACg
	<i>ppc</i>	-29	5,8	TGTAATaAAATTACA
	<i>gpml</i>	-44	5,5	TGTAgTaAAATTACA
	<i>pntAB</i>	-88	5,3	TGTcATTTtATTACA
	<i>grcA</i>	-236	4,3	aGTAAgTTtAcTACA
	<i>pflA</i>	-43	4,4	TGTAATaAAATggCc
	<i>mtlADR</i>	-208	4,3	TGTAgTaAAtTTtCA
	<i>nirBDC-cysG</i>	-183	5,3	TGTAATaAAAcTACg
<i>pepD</i>	-271	4,5	TGAAAcTTtATTACA	

Геном	Оперон	Сайт			
		Положение	Вес	Последовательность	
	<i>gltBD</i>	-1201	5,1	TGTAATTTtATTcCA	
	<i>ptsHI-crr</i>	-206	5,1	TGTAAaаТААТТАСА	
<i>V. angustum S14</i>	<i>gapA2-aldE</i>	-56	5,2	TGTAATTTtAcTACc	
	<i>focA-pflB</i>	-220	5,2	TGTAAcaAAAТТАСА	
	<i>pflA</i>	-47	5,0	TGTAATaAAAТaАСА	
	<i>aceBA</i>	-296	5,5	TGTAATaAAAТТАСg	
	<i>ppc</i>	-29	5,2	TGTAAcTAAATТАСg	
	<i>gpmI</i>	-109	5,5	TGTAgTaAAAТТАСА	
	<i>pntAB</i>	-86	4,9	TGTAtTTTTtAcTACA	
	<i>pckA</i>	-294	5,0	TGaATaТААТТАСА	
	<i>gltBD</i>	-496	4,4	cGTAAaTAAaTACA	
			-292	4,4	TGTAggaTAGТТАСА
<i>P. profundum SS9</i>	<i>gpmI</i>	-92	5,5	TGTAgTaAAAТТАСА	
	<i>ppc</i>	21	5,0	TGTAAcaAAAТТАСg	
	<i>grcA</i>	-134	5,5	TGTAATTTtAcTACA	
	<i>aceBA</i>	-304	5,3	gGTAATTAAATТАСg	
	<i>hexR</i>	-226	5,0	TGaATTAAAcTACA	
	<i>gapA2-aldE</i>	52	4,6	TGTAATTTtAccACc	
	<i>pgi</i>	52	4,6	TGTAATTTtAccACc	
	<i>focA-pflB</i>	-114	5,0	TGaATaТААТТАСА	
	<i>pflA</i>	-53	5,0	TGTAATaAAATgACA	
	<i>pykA</i>	-76	4,7	TGTAgTaAtATTtCA	
	<i>mtlADR</i>	-209	4,6	TGTAATTcAcTTACA	
	<i>ptsG</i>	-299	5,0	TGaATTTtATTACA	
			-177	4,5	TGaATTAAATТtCA
	<i>gltBD</i>	-294	4,4	TGTAggaTAGТТАСА	
<i>nirBDC-cysG</i>	-173	5,3	TGTAATTTAgТТАСА		
<i>V. cholerae O1</i>	<i>gapA</i>	-55	5,1	TGTAATTTTACTACC	
	<i>ppc</i>	-28	4,7	TGTAATTTTTTTTCA	
		-172	4,0	TGTAAGAAATTTTCA	
	<i>gpmM</i>	-60	5,2	TGTAGTAAAAТТАСG	
	<i>ygaW</i>	-246	5,2	TGTAATAAAAТТАСC	
	<i>aceB</i>	-256	5,4	TGTAATTAAATТАСG	
	<i>pepD</i>	-211	4,5	AGAAATTAATТАСА	
		-136	4,9	AGTAATTTAATТАСА	
	<i>mtlADR</i>	-422	4,9	TGTAGCTTAATТАСА	
	<i>pntAB</i>	-85	5,0	CGTAATTGAATТАСА	
	<i>pgi</i>	-138	4,8	TGAAAATTAATТАСА	
	<i>pflA</i>	-42	4,6	TGTAATAAAAТААСG	
	<i>glgX</i>	-99	5,0	TGAAATAAAAТТАСА	
	<i>grcA</i>	-247	5,3	TGTAGTAAAAТТАСА	
	<i>ptsG</i>	-249	4,3	TGTAATTTTGTACT	
		-98	4,1	TGAAACTTAATТАСG	
	<i>hexR</i>	-282	4,0	AGAAATAAAAТТАСА	
	<i>gltBD</i>	-487	5,6	TGTAATTTAATТАСА	

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>V. shilonii</i> AK1	<i>ppc</i>	-140	5,0	TGTAATAAAATTACA
		5	5,4	TGTAGTTAAATTACA
	<i>gpmM</i>	-63	5,2	TGTAGTAAAAATTACG
	<i>nirBD</i>	-84	4,8	TGTAATTGATTTACA
		-265	4,8	TGTAGGTAAATTACA
	<i>ygaW</i>	-157	4,9	TGTAATCAAATTACA
	<i>mtlADR</i>	-390	5,4	TGTAATAAAATTACA
		-193	4,2	TGTAATTTATCAACG
	<i>pntAB</i>	-85	4,3	TGTTTTTTTATTACA
	<i>pgi</i>	-183	4,4	TGAAAAATTATTACG
	<i>pflA</i>	-44	4,6	TGTAATAAAATGACG
	<i>glgX</i>	-112	5,1	TGAAATTAAAATTACA
	<i>glgCA</i>	-237	4,4	TGTAATCTTATTACA
		-57	5,2	TGTAATTTAACTACC
<i>ptsG</i>	-321	4,8	TGTAATTATGTTACG	
	-170	4,1	TGAAACTAAATTTCG	
<i>gltBD</i>	-483	4,7	TGAAACTAAATTACA	
<b>Shewanellaceae</b>				
<i>S. oneidensis</i> MR-1	<i>ppsA</i>	-112	5,2	TaGTAATaAAAATTACAA
	<i>phk</i>	-74	5,2	TTGTAAGaAAAATTACAA
	<i>tal-pgi</i>	-105	5,1	TTGTAATTTttTTACAt
	<i>gcvTHP</i>	-321	4,9	aTGTAaaaAAAATTACAt
	<i>nqrABCDEFG</i>	-182	4,9	TTGTtATaAAAcTACAt
	<i>gapA3</i>	-183	4,9	cTcTAATaAAAATTACAg
	<i>zwf-pgl-edd-eda</i>	-77	4,8	cTGTtgTTAtATTACAA
	<i>hexR</i>	-250	4,8	TTGTAATaTAaCaACAg
	<i>pykA</i>	-77	4,8	TTGTAAcTTtAcTACAt
	<i>gnd</i>	-83	4,8	TgGTAATTAAgTTACAA
	<i>deoABD</i>	-86	4,8	TTGTAATaTtgcTACAg
	<i>cdd</i>	-58	4,7	cTGTAaCaAAAATTACtA
	<i>mcp1</i>	-77	4,6	cTGTAaTtttTTACAg
	<i>gapA2</i>	-48	4,4	TgGTAATTTtAcTACcA
	SO1118	-251	5,4	TTGTAATTTAATTACAt
		-199	5,2	TTGTAATaAAAcTACAt
	<i>nupC</i>	-126	4,3	TTGaAATTAAAcTcCAA
<i>S. putrefaciens</i> CN-32	<i>adhE</i>	-136	5,3	TTGTAgtTTtATTACAA
	<i>phk</i>	-72	5,3	TTGTAATaAAAcTACAA
	<i>ppsA</i>	-111	5,2	TaGTAATaAAAATTACAA
	<i>tal-pgi</i>	-104	5,1	TTGTAATTTttTTACAt
	<i>gapA3</i>	-182	5,1	TTcTAATaAAAATTACAt
	<i>gcvTHP</i>	-175	4,9	aTGTAaaaAAAATTACAt
	<i>cdd</i>	-58	4,9	TTGTAaCaAAAATTACtA
	<i>nqrABCDEFG</i>	-183	4,9	TTGTtATaAAAcTACAt
	<i>zwf-pgl-edd-eda</i>	-77	4,8	cTGTtgTTAtATTACAA
	<i>hexR</i>	-250	4,8	TTGTAATaTAaCaACAg

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
	<i>pykA</i>	-76	4,8	TTGTAACtTtAcTACAt
	<i>gnd</i>	-81	4,8	TgGTAATTAAgTTACAA
	<i>deoABD</i>	-87	4,8	TTGTAATaTtgcTACAg
	<i>mcp1</i>	-79	4,6	aTGTAACtTtATTACcA
	<i>gapA2</i>	-48	4,4	TgGTAATTTtAcTACcA
	<i>adhB</i>	-58	4,3	TgGTAgtTTAgTaACAA
	<i>SO1118</i>	-177	5,0	gTGTAATTTAATTACAt
		-125	5,3	TTGTAATaAAAcTACAA
	<i>nupC</i>	-124	4,3	TTGaAATTAAAcTcCAA
<i>Shewanella</i> sp. W3-18-1	<i>phk</i>	-74	5,3	TTGTAATaAAAcTACAA
	<i>adhE</i>	-134	5,3	TTGTAgtTTtATTACAA
	<i>ppsA</i>	-111	5,2	TaGTAATaAAAATTACAA
	<i>tal-pgi</i>	-104	5,1	TTGTAATTTttTTACAt
	<i>gapA3</i>	-180	5,1	TTcTAATaAAAATTACAt
	<i>gcvTHP</i>	-175	4,9	aTGTAaaaAAAATTACAt
	<i>cdd</i>	-56	4,9	TTGTAACAaaaATTACtA
	<i>nqrABCDEF</i>	-183	4,9	TTGTtATaAAAcTACAt
	<i>hexR</i>	-248	4,8	TTGTAATaTAACAACAg
	<i>zwf-pgl-edd-eda</i>	-79	4,8	cTGTtgTTAtATTACAA
	<i>pykA</i>	-74	4,8	TTGTAACtTtAcTACAt
	<i>deoABD</i>	-85	4,8	TTGTAATaTtgcTACAg
	<i>gnd</i>	-83	4,8	TgGTAATTAAgTTACAA
	<i>mcp1</i>	-77	4,6	aTGTAACtTtATTACcA
	<i>gapA2</i>	-50	4,4	TgGTAATTTtAcTACcA
	<i>adhB</i>	-60	4,3	TgGTAgtTTAgTaACAA
	<i>SO1118</i>	-175	5,0	gTGTAATTTAATTACAt
		-123	5,3	TTGTAATaAAAcTACAA
	<i>Shewanella</i> sp. ANA-3	<i>ppsA</i>	-113	5,2
<i>phk</i>		-74	5,2	TTGTAAGaAAAATTACAA
<i>tal-pgi</i>		-103	5,1	TTGTAATTTttTTACAt
<i>gcvTHP</i>		-319	4,9	aTGTAaaaAAAATTACAt
<i>mcp1</i>		-76	4,9	cTGTAaAATTtATTACAt
<i>nqrABCDEF</i>		-182	4,9	TTGTtATaAAAcTACAt
<i>mcp2</i>		-49	4,9	gTGTAATTTAATTACAc
<i>gapA2</i>		-182	4,9	cTcTAATaAAAATTACAg
<i>hexR</i>		-249	4,8	TTGTAATaTAACAACAg
<i>zwf-pgl-edd-eda</i>		-80	4,8	cTGTtgTTAtATTACAA
<i>pykA</i>		-77	4,8	TTGTAACtTtAcTACAt
<i>gnd</i>		-83	4,8	TgGTAATTAAgTTACAA
<i>deoABD</i>		-86	4,8	TTGTAATaTtgcTACAg
<i>adhE</i>		-135	4,5	cTGTAATTTtATTACAt
<i>gapA3</i>		-50	4,4	TgGTAATTTtAcTACcA
<i>cdd</i>		-56	4,3	ccGTAACAaaaATTACtA
<i>nupC</i>		-126	4,3	TTGaAATTAAAcTcCAA

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>Shewanella</i> sp. MR-4	<i>ppsA</i>	-113	5,2	TaGTAATaAAAATTACAA
	<i>mcp2</i>	-49	5,2	aTGTAATTTAATTACAt
	<i>phk</i>	-72	5,2	TTGTAAGaAAAATTACAA
	<i>tal-pgi</i>	-103	5,1	TTGTAATTTttTTACAt
	<i>gcvTHP</i>	-321	4,9	aTGTAaaaAAAATTACAt
	<i>nqrABCDEF</i>	-182	4,9	TTGTtATaAAAcTACAt
	<i>gapA3</i>	-182	4,9	cTcTAATaAAAATTACAg
	<i>zwf-pgl-edd-eda</i>	-78	4,8	cTGTtgTTAtATTACAA
	<i>hexR</i>	-251	4,8	TTGTAATaTAACAACAg
	<i>pykA</i>	-76	4,8	TTGTAAcTTtAcTACAt
	<i>gnd</i>	-83	4,8	TgGTAATTAAGTTACAA
	<i>deoABD</i>	-86	4,8	TTGTAATaTtgcTACAg
	<i>adhE</i>	-135	4,5	cTGTAtTTTtATTACAt
	<i>gapA2</i>	-50	4,4	TgGTAATTTtAcTACcA
	<i>cdd</i>	-58	4,3	ccGTAACA AAAATTACtA
	<i>nupC</i>	-126	4,3	TTGaAATTAAAcTcCAA
<i>Shewanella</i> sp. MR-7	<i>ppsA</i>	-113	5,2	TaGTAATaAAAATTACAA
	<i>mcp2</i>	-49	5,2	aTGTAATTTAATTACAt
	<i>phk</i>	-74	5,2	TTGTAAGaAAAATTACAA
	<i>tal-pgi</i>	-105	5,1	TTGTAATTTttTTACAt
	<i>gcvTHP</i>	-320	4,9	aTGTAaaaAAAATTACAt
	<i>nqrABCDEF</i>	-180	4,9	TTGTtATaAAAcTACAt
	<i>gapA3</i>	-184	4,9	cTcTAATaAAAATTACAg
	<i>hexR</i>	-249	4,8	TTGTAATaTAACAACAg
	<i>zwf-pgl-edd-eda</i>	-80	4,8	cTGTtgTTAtATTACAA
	<i>pykA</i>	-74	4,8	TTGTAAcTTtAcTACAt
	<i>gnd</i>	-83	4,8	TgGTAATTAAGTTACAA
	<i>deoABD</i>	-86	4,8	TTGTAATaTtgcTACAg
	<i>cdd</i>	-58	4,7	cTGTAACA AAAATTACtA
	<i>adhE</i>	-133	4,5	cTGTAtTTTtATTACAt
	<i>gapA2</i>	-48	4,4	TgGTAATTTtAcTACcA
	<i>nupC</i>	-126	4,3	TTGaAATTAAAcTcCAA
<i>mcp1</i>	-73	4,1	cTGTAaATtATgACAc	
<i>S. baltica</i> OS155	<i>phk</i>	-75	5,3	TTGTAATaAAAcTACAA
	<i>ppsA</i>	-113	5,2	TaGTAATaAAAATTACAA
	<i>tal-pgi</i>	-102	5,1	TTGTAATTTttTTACAt
	<i>gapA3</i>	-182	5,0	TTcTAATaAAAATTACAg
	<i>gcvTHP</i>	-373	4,9	aTGTAaaaAAAATTACAt
	<i>cdd</i>	-56	4,9	TTGTAACA AAAATTACtA
	<i>nqrABCDEF</i>	-183	4,9	TTGTtATaAAAcTACAt
	<i>zwf-pgl-edd-eda</i>	-78	4,8	cTGTtgTTAtATTACAA
	<i>hexR</i>	-251	4,8	TTGTAATaTAACAACAg
	<i>pykA</i>	-78	4,8	TTGTAAcTTtAcTACAt
<i>adhE</i>	-134	4,8	gTGTAgtTTtATTACAt	

Геном	Оперон	Сайт			
		Положение	Вес	Последовательность	
	<i>gnd</i>	-83	4,8	TgGTAATTAAgTTACAA	
	<i>deoABD</i>	-87	4,8	TTGTAATaTtgcTACAg	
	<i>gapA2</i>	-50	4,4	TgGTAATTTtAcTACcA	
	<i>mcp1</i>	-24	4,3	TTGTtAaTAAATTAgAc	
	<i>SO1118</i>		-177	5,0	gTGAATTTAATTACAt
		-125	5,3	TTGTAATaAAAcTACAA	
<i>S. denitrificans</i> OS217	<i>ppsA</i>	-111	5,2	TaGTAATaAAATTTACAA	
	<i>adhE</i>	-137	5,2	cTGAATTTAAATTACAg	
	<i>phk</i>	-73	5,1	TcGTAATTTAAATTACAA	
	<i>gapA3</i>	-182	4,9	aTcTAATaAAATTTACAt	
	<i>hexR</i>	-248	4,8	TTGTAATaTAACAACAg	
	<i>zwf-pgl-edd-eda</i>	-89	4,8	cTGTtgTTAtATTACAA	
	<i>tal-pgi</i>	-95	4,8	TgGTAATTTtTTACAA	
	<i>pykA</i>	-162	4,8	TTGTAAcTTtAcTACAt	
	<i>gapA2</i>	-62	4,3	TgGTAATTTtAcaACTA	
<i>S. frigidimarina</i> NCIMB 400	<i>phk</i>	-72	5,4	TTGTAATTTAAATTACAg	
	<i>ppsA</i>	-58	5,2	TaGTAATaAAATTTACAA	
	<i>adhE</i>	-130	5,2	cTGAATaAAATTTACAt	
	<i>gcvTHP</i>	-153	5,1	aTGTAaaaAAATTTACAA	
	<i>tal-pgi</i>	-95	4,8	TcGTAATTTttTTACAA	
	<i>gnd</i>	-89	4,8	TgGTAATTTAgTTACAA	
	<i>gapA3</i>		-179	4,5	gTcTAATaAAATTTACAc
	<i>hexR</i>		-121	4,5	TTGaAaaaTAAcTACAg
			-53	4,5	aTGatATaAAATTTACAt
	<i>zwf-pgl-edd-eda</i>		-301	4,5	aTGAATTTtATatCAt
			-233	4,5	cTGTAgTTAttTTtCAA
	<i>gapA2</i>	-135	4,5	aTGTAaaaAAAAcTACAA	
	<i>pykA</i>	-133	4,4	TgGTAAcTTtAcTACAt	
	<i>SO1118</i>		-169	4,4	TTGTAATTTAAATcACTt
		-117	5,4	TTGTAATTTAAATTACAg	
<i>adhB</i>	-254	4,2	cgGTAATaTtATTtgAA		
<i>S. amazonensis</i> SB2B	<i>phk</i>	-76	5,3	TTGTAATaAAtTTACAA	
	<i>ppsA</i>	-113	5,2	TaGTAATaAAATTTACAA	
	<i>gapA3</i>	-182	5,0	cTcTAATaAAATTTACAA	
	<i>deoABD</i>	-86	4,9	cTGAATaTttTTACAg	
	<i>hexR</i>	-250	4,8	TTGTAATaTAACAACAg	
	<i>zwf-pgl-edd-eda</i>	-79	4,8	cTGTtgTTAtATTACAA	
	<i>nqrABCDEFG</i>	-108	4,8	TTGTAgTTAAATTTcAT	
	<i>tal-pgi</i>	-104	4,7	TTGTAATTTttTTtCAt	
	<i>pykA</i>	-76	4,5	TTGTAAcTTtccTACAt	
	<i>gapA2</i>	-47	4,4	TgGTAATTTtAcTACcA	
	<i>adhE</i>	-143	4,3	cTGTtgTTTtATTtCAg	
	<i>nupC</i>	-126	4,1	TTGaAATTTAAgTtCAA	
	<i>gnd</i>	-75	4,0	cTGAATTTcAgTTACcA	

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>S. loihica</i> PV-4	<i>phk</i>	-76	5,5	TTGTAATaAAAATTACAA
	<i>ppsA</i>	-113	5,2	TaGTAATaAAAATTACAA
	<i>pykA</i>	-76	4,9	TTGTAATTTtccTACAA
	<i>adhE</i>	-137	4,9	aTGTtATTTtATTACAt
	<i>nqrABCDEF</i>	-172	4,9	TTGTtATaAAAcTACAt
		-100	4,5	cTGTAgcTAAgTTACAA
	<i>zwf-pgl-edd-eda</i>	-173	4,6	TaGTAATTTAATcACAA
		-75	4,8	cTGTtgTTAtATTACAA
	<i>hexR</i>	-250	4,8	TTGTAATaTAACAACAg
		-152	4,6	TTGTgATTAAATTACtA
	<i>gcvTHP</i>	-396	4,7	aTGTAAaaaAAAcTACAt
	<i>gapA3</i>	-180	4,7	aTcTAATaAAAATTACAc
	<i>adhB</i>	-167	4,6	TTGTcATaAAtTTACAA
	<i>tal-pgi</i>	-106	4,5	cTGTAAATTTtTTtCAt
	<i>nupC</i>	-124	4,5	TTGaAATaAAAcTtCAA
	<i>deoABD</i>	-86	4,5	TcGTAATTAAATTACgg
	<i>gnd</i>	-171	4,4	TaGTAATTAACTTACcA
	<i>gapA2</i>	-51	4,4	TgGTAATTTtAcTACcA
SO1118	-239	5,0	aTGTAgTaAAAATTACAg	
	-187	5,2	TTGTAATaAAAcTACAt	
<i>cdd</i>	-56	4,2	TgGTAAcTTAgTTACtA	
<i>S. pealeana</i> ATCC 700345	<i>ppsA</i>	-113	5,2	TaGTAATaAAAATTACAA
	<i>mcp1</i>	-159	5,1	aTGTAAATTTAtTTACAA
	<i>phk</i>	-73	5,1	TcGTAATaAAAATTACAA
	<i>gapA3</i>	-183	4,9	aTcTAATaAAAATTACAg
	<i>zwf-pgl-edd-eda</i>	-267	4,3	aTGTAgTTTTtATTAtAt
		-53	4,8	cTGTtgTTAtATTACAA
	<i>hexR</i>	-255	4,8	TTGTAATaTAACAACAg
		-41	4,3	aTaTAATaAAAcTACAt
	<i>adhB</i>	-166	4,7	aTGTcATaAAAATTACAA
	<i>pykA</i>	-77	4,6	TcGTAAATTTtAcTACAA
	<i>deoABD</i>	-93	4,6	aTGTAAcaAtATTACgA
	<i>nqrABCDEF</i>	-109	4,6	TTGTAggTAAcTTACAA
	<i>gapA2</i>	-125	4,5	aTGTAAaaaAAATTACct
	<i>gnd</i>	-126	4,5	TaGTAATTAAGTTACcA
	<i>adhE</i>	-152	4,5	TgGTtATaAAAATTACAc
	<i>tal-pgi</i>	-108	4,3	gTGTAAATTTtTTtCAt
	SO1118	-231	4,4	TTGTtgTTAAATTACgt
		-178	5,4	TTGTAATaAAAATTACAg
<i>nupC</i>	-345	4,3	TcGTAAcTAAATTAgAt	
<i>S. halifaxens</i> HAW-EB4	<i>phk</i>	-73	5,5	TTGTAATaAAAATTACAA
	<i>ppsA</i>	-115	5,2	TaGTAATaAAAATTACAA
	<i>hexR</i>	-251	4,8	TTGTAATaTAACAACAg
		-154	4,6	TTGTgATaAAAATTACtA
	<i>zwf-pgl-edd-eda</i>	-176	4,6	TaGTAATTTtATcACAA

Геном	Оперон	Сайт			
		Положение	Вес	Последовательность	
		-79	4,8	cTGTtgTTAtATTACAA	
	<i>gapA2</i>	-183	5,1	TTcTAATaAAAATTACAt	
	<i>gapA3</i>	-243	4,2	TTGTAgtTgAATTACct	
	<i>SO1118</i>		-190	5,4	TTGTAATaAAAATTACAg
			-154	4,6	TgGTtATaAAAATTACAg
	<i>mcp1</i>	-165	4,8	gTGTAAATTTAtTTACAt	
	<i>adhE</i>	-90	4,1	gTGTAAATTTtctTcCAAt	
	<i>deoABD</i>	-93	4,4	aTGTAAcaAtATTACgg	
	<i>tal-pgi</i>	-107	4,3	gTGTAAATTTtTTtCAAt	
	<i>nqrABCDEF</i>	-109	4,3	TTGTtggtAAcTTACAA	
	<i>pykA</i>	-75	4,5	TcGTAATTTtAcTACAc	
	<i>gnd</i>	-133	4,5	TaGTAATTAAGTTACcA	
	<i>gcvT</i>	-236	4,1	aTGTAAaaaAgTTACAt	
	<i>nupC</i>	-361	4,9	cTGTAAATTAATTAgt	
<i>S. piezotolerans WP-2</i>	<i>ppsA</i>	-151	5,2	TaGTAATaAAAATTACAA	
	<i>pykA</i>	-75	4,7	TcGTAATTTtAcTACAt	
	<i>hexR</i>	-251	4,8	TTGTAATaTAAcaACAg	
	<i>zwf-pgl-edd-eda</i>	-79	4,8	cTGTtgTTAtATTACAA	
	<i>gapA2</i>	-125	4,5	aTGTAAaaaAAAATTACct	
		-50	4,4	TgGTAATTTtAcTACcA	
	<i>SO1118</i>	-181	4,9	aTcTAATTAATTTACAt	
	<i>gapA3</i>	-348	4,9	aTcTAATTAATTTACAt	
	<i>adhE</i>	-151	4,8	TaGTtATTTtATTACAg	
	<i>gnd</i>	-176	4,5	TaGTAATTTAgTTACcA	
	<i>phk</i>		-138	4,5	aTcTAgtTTtAcTACAt
			-75	5,5	TTGTAATaAAAATTACAA
	<i>deoABD</i>	-96	4,4	aTGTAAcaAtATTACgg	
	<i>nupC</i>	-126	4,5	TTGaATaAAAATtCAA	
	<i>cdd</i>	-58	4,1	aTGTAAaTAgcTTACtA	
	<i>tal-pgi</i>	-107	4,5	cTGTAAATTTtTTtCAAt	
	<i>gcvT</i>	-188	4,8	gTGTAAaTAAATTTACAt	
	<i>gcvHP</i>	-49	4,4	gTcTtATaAtATTACAg	
	<i>nqrABCDEF</i>		-177	4,4	TcGTtgTaAAAATTACAt
			-95	4,2	TTGTAAtgTAAgTTACAA
<i>adhB</i>		-164	4,3	aTGTcgtAAAATTACAg	
		-59	4,0	TgGTAcTTTAATTACcA	
<i>S. sediminis HAW-EB3</i>	<i>phk</i>	-74	5,5	TTGTAATaAAAATTACAA	
	<i>ppsA</i>	-115	5,2	TaGTAATaAAAATTACAA	
	<i>deoABD</i>	-84	5,0	TTGTAATTAATTTACgt	
	<i>pykA</i>	-75	4,9	TTGTAATTTtccTACAA	
	<i>nqrABCDEF</i>		-177	4,9	TTGTtATaAAAATTACAt
			-105	4,5	aTGTtAcTAAgTTACAA
	<i>hexR</i>	-247	4,8	TTGTAATaTAAcaACAg	
	<i>zwf-pgl-edd-eda</i>	-79	4,8	cTGTtgTTAtATTACAA	
<i>gapA3</i>	-181	4,8	TgcTAATaAAAATTACAA		

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
	<i>gcvTHP</i>	-236	4,7	aTGTAaAaAAAaAcTACaT
	<i>nupC</i>	-396	4,7	aaGTAATaAtATTtCAA
		-314	4,7	cTGTAATTTAAcaACaT
	<i>adhE</i>	-138	4,6	TcGTtATTTtATTACaT
	<i>tal-pgi</i>	-104	4,5	cTGTAATTTtTTtCAt
	SO1118 -1	-201	5,4	TTGTAATaAAAATTACaT
		-36	4,9	gaGTAATTTtATTACAA
SO1118 -2	-201	5,4	TTGTAATaAAAATTACaT	
	-37	4,9	gaGTAATTTtATTACAA	
<i>gapA2</i>	-52	4,4	TgGTAATTTtAcTACcA	
<i>S. woodyi</i> ATCC 51908	<i>phK</i>	-74	5,5	TTGTAATaAAAATTACAA
	<i>gnd</i>	-238	4,4	TaGTAATTAaAcTTACcA
	<i>pykA</i>	-77	4,9	TTGTAATTTtccTACAA
	<i>ppsA</i>	-113	5,2	TaGTAATaAAAATTACAA
	SO1118	-194	4,6	TTGTAATaAAcTTtCAg
	<i>hexR</i>	-249	4,8	TTGTAATaTAaCaACAg
		-151	4,6	TTGTgATTAAATTACtA
	<i>zwf-pgl-edd-eda</i>	-174	4,6	TaGTAATTTAAATcACAA
		-76	4,8	cTGTtgTTAtATTACAA
	<i>adhE</i>	-145	4,5	acGTtATTTAATTACaT
	<i>tal-pgi</i>	-103	4,5	cTGTAATTTtTTtCAt
	<i>nqrABCDEFG</i>	-169	4,2	TTGTcATaAAAcaACaT
	<i>deoABD</i>	-107	4,7	aTGTAgtAaAtTTACaT
	<i>nupC</i>	-123	4,3	cTGaAATTAAAcTtCAA
<i>gapA2</i>	-50	4,4	TgGTAATTTtAcTACcA	
<i>gapA3</i>	-180	4,8	TgcTAATaAAAATTACAA	
<b>Alteromonadales</b>				
<i>P. atlantica</i> T6c	<i>hexR</i>	-175	5,0	TTGTAATTTtAcaACgt
	<i>zwf-pgl-edd-glK-eda</i>	-117	5,0	acGTtgTaAAAATTACAA
	<i>pykA</i>	-189	5,7	TcGTAATTTAAATTACgA
	<i>gapA2-aldE</i>	-75	5,3	TcGTAATTAtATTtCAt
	<i>gapA3</i>	-180	5,1	TTGTAgtTTAAATTACcA
	<i>ppsA</i>	-194	5,5	TaGTAATaTAATTACaT
	<i>tal-pgi</i>	-36	5,1	TTGTAATTTtAcaACaT
	<i>wcaJ</i>	-236	4,8	TcGTAATTAtAcaACAg
	<i>gbe</i>	-92	5,4	TcGTAgtTTTtATTACAA
<i>glgP</i>	-70	5,5	TTGTAATTTAATTACgg	
<i>A. macleodii</i> 'Deep ecotype'	<i>hexR</i>	-192	5,5	aTGTAgtAaAAAATTACAA
	<i>zwf-pgl-edd-glK-eda</i>	-109	5,5	TTGTAATTTtAcTACaT
	<i>ppc</i>	-223	5,2	cTGaAATTAAATTACAA
	<i>pykA</i>	-97	5,4	cTGTAATaAAAATTACgt
	<i>gapA2-aldE</i>	-41	4,7	acGTAATTTtTTtCAt
	<i>ppsA</i>	-253	5,3	gTGTAATTTtATTACAA
	<i>pckA</i>	-151	5,2	TTGTAATTTAAATTACAA
<i>wcaJ-manC</i>	-255	5,5	cTGTAATaAAAATTACAA	

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
	<i>glgP</i>	-23	5,5	сTGTAATTAAATTACgt
	<i>MADE_00987-gtfAB</i>	-70	5,2	aaGTAATTTtATTACag
<i>C. psychrerythraea</i> 34H	<i>hexR-pykA</i>	-362	4,9	TTGTAACТАAACTACAt
		-301	5,5	TTGTAATTTtAcTACAt
		-138	5,0	aTGaAATaAAAATtCAt
	<i>zwf-pgl-edd-eda</i>	-330	5,0	aTGaAATTTtATTtCAt
		-167	5,5	aTGTAgtTaAAAATTACAA
		-106	4,9	aTGTAgtTTTAgTTACAA
	<i>gapA2</i>	-37	5,0	gTGTAATTTtAcTACAt
<i>gnd-msrB</i>	-235	5,0	aTGTAgtTaAAAATTACAc	
<i>I. baltica</i> OS145	<i>zwf-pgl-edd-eda</i>	-47	4,8	aTGTAgtTaAAgTTACAA
	<i>hexR-gapA2</i>	-98	4,8	TTGTAACtTtAcTACAt
<i>Glaciecola</i> sp. HTCC2999	<i>hexR</i>	-173	5,5	TTGTAGTAAAAATTACAA
	<i>gapA-aldE</i>	-31	4,8	ATGTAACtTAAATTCAT
		-83	5,5	ATGTAATAAACTACAT
	<i>ppsA</i>	-151	5,1	GTGTAATTTAATTACAG
	<i>zwf-pgl-edd-glK-eda</i>	-98	5,5	TTGTAATTTTACTACAA
	<i>GHTCC_010100006532</i>	-117	5,8	ATGTAATTTTATACAA
<i>pykA</i>	-69	5,3	TTGTAATATAATTACCT	
		-120	4,7	TTGAAAATTAATTACGT
<b>Psychromonadaceae/Aeromonadales</b>				
<i>A. hydrophila</i> ATCC 7966	<i>hexR</i>	-208	4,6	TTGTtgcTTtgcAACAAt
		-25	4,7	cTGaAgcTAAATTACAA
	<i>glk</i>	-332	4,7	TTGTAATTTAgcTtCag
		-149	4,6	aTGtTgcaAAgcaACAA
	<i>eno</i>	-35	4,4	TTGTtgtTTTttTTACgA
	<i>ppc</i>	-204	4,6	aTGTAAGTAAgTTACAt
	<i>pykA</i>	-110	5,2	TTGTAATTTAgTaACAA
	<i>ldhA</i>	-40	5,2	TTGTAATTTAACTACAt
	<i>ackA-pta</i>	-334	4,7	cTGTAACtTAgTtCCAA
		-218	4,7	TTGTtATTTtcaACAA
	<i>pflA</i>	-41	4,4	aTGTAATaAAATaACgg
	<i>focA-pflB</i>	-105	5,1	TTGTAgtTaAAAATTACAt
	<i>grcA</i>	-290	4,5	aTGTAgtTTAAATTACtA
	<i>aceBA</i>	-315	4,8	TTGTAATTTtATTACgt
	<i>tal</i>	-86	4,8	TTGTAATTTtATTtCag
	<i>pntAB</i>	-194	4,9	TTGTAATaAAACTcCAA
	<i>ptsHI-crr</i>	-98	4,3	TcGTtAcTTtAcaACAAt
	<i>ndh</i>	-179	4,5	TTGaAgaaTtATTACAA
	<i>glpTR</i>	-152	4,8	TTGTAATaAAACTgCAA
<i>mgIBAC</i>	-148	4,6	TTGTtgcTAtgTaACAg	
<i>gltBD</i>	-311	4,3	TgGTAAaaTtAcTACAA	
<i>A. salmonicida</i> A449	<i>hexR</i>	-214	4,6	TTGTtgcTTtgcAACAAt
		-31	4,7	cTGaAgcTAAATTACAA
	<i>glk</i>	-332	4,7	TTGTAATTTAgcTtCag

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
		-149	4,6	aTGTtgcaAAgcaACAA
	<i>eno</i>	-33	4,4	TTGTtgTTTTtTTACgA
	<i>ppc</i>	-75	4,4	TTGTAATcAcAcaACAA
	<i>pykA</i>	-109	5,0	TTGTAATTTAgTaACAg
	<i>ldhA</i>	-38	5,2	TTGTAATTTAAcTACAt
	<i>ackA-pta</i>	-336	4,7	cTGTAAcTTAgTTcCAA
		-220	4,7	TTGTtATTTtcaACAA
	<i>focA-pflB</i>	-107	5,1	TTGTAgTaAAAATTACAt
	<i>IS630-grcA</i>	-122	4,5	aTGTAgtTAAATTACtA
	<i>aceBA</i>	-313	4,8	TTGTAATTTtATTACgt
	<i>tal</i>	-96	4,8	TTGTAATTTtATTcAg
	<i>pntAB</i>	-196	4,9	TTGTAATaAAAcTcCAA
	<i>ptsHI-crr</i>	-98	4,3	TcGTtAcTTtAcaACAt
	<i>ndh</i>	-179	4,5	TTGaAgaatTATTACAA
	<i>glpTR</i>	-150	4,8	TTGTAATaAAAcTgCAA
	<i>mtlADR</i>	-220	4,4	aTGTAgccAAAcTACAg
	<i>gltBD</i>	-264	4,3	TgGTAAaaTtAcTACAA
<i>P. ingrahamii</i> 37	<i>glk</i>	-106	5,9	ATGTcATTTtATTACAA
	<i>hexR-pykA</i>	-203	5,9	TTGTAATaAAAgtACAT
	<i>ppc</i>	-153	5,4	gTGTAATTTAATTACtT
	<i>grcA</i>	-47	5,5	TTGTAAcTTtATTACAT
	<i>gapA2</i>	-204	6,0	gTGTAATTTAATTACAA
	<i>focA-pflBA</i>	-106	6,1	ATGTAATaAAAATTACAA
	<i>ackA-pta</i>	-133	6,0	TTGTcATTTAATTACAT
<i>Psychromonas</i> sp. CNPT3	<i>hexR-pykA</i>	-117	4,7	TTGTAATTTtTTtCAt
	<i>glk</i>	-180	4,7	aTGaAAaTAAATTACAA
	<i>gapA2</i>	-187	5,2	cTGTAATTTAATTACAt
	<i>ppc</i>	-32	5,2	aTGTAATTTtATTACAt
	<i>adh-adhE</i>	-159	4,9	TTGTAAcTTAcTTACAA
	<i>ackA-pta</i>	-153	5,5	TTGTAATTTtATTACAA
	<i>PCNPT3_03061-pflA</i>	-271	5,2	cTGTAATaAAAATTACAt
	<i>tpi</i>	-100	5,2	aTGTAATTTtATTACAt
<i>Moritella</i> sp. PE36	<i>ppc</i>	-142	4,6	TTGTAGTAAAAATTACAC
	<i>hexR</i>	-195	4,2	CCGTAATAAAGTCACAA
	<i>focA-pflBA</i>	-156	4,5	ACGTAACCTTAATTACAG
	<i>gltBD</i>	-247	3,9	TTGTAGTTTATTGACCA
	<i>adhE</i>	-207	5,0	TTGAAATAAAAATTACAA
		-135	4,6	ACGTAATATTACTACAA
	<i>tpiA</i>	-175	4,6	ATGTAGTTTATTACCA
	<i>ptsHI-crr</i>	-201	5,4	TTGTAATTTAACTACAA
<i>T. auensis</i> DSM 9187	<i>pykA</i>	-113	4,2	CTGTAATGTTCTTACAA
	<i>ptsHI-crr</i>	-104	4,2	TCGTTGTCATACTACAA
	<i>focA-pflBA</i>	-108	5,0	ATGTAATAAAAATTACAT
	<i>mgIBAC</i>	-272	4,7	TTGTAATTTAACATCAG
	<i>ldhA</i>	-38	4,5	GTGTAATTTTACTACAT

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
	<i>grcA</i>	-269	4,9	TTGTAGTAAATTACAA
<b>Oceanospirillales/Alteromonadales</b>				
<i>C. salexigens</i> DSM 3043	<i>aceEF</i>	-138	4,8	сTGTAAGgAaATTACAg
	<i>edd-glк-aldE-CsaI_0933-pgi-CsaI_0931</i>	-113	5,4	TTGgAAgAATACTACAA
	<i>gapA2</i>	-70	5,4	TTGTAGTATTcTTcCAA
	<i>ppsA</i>	-154	5,3	TTGTAAaAATACTcCAA
	<i>aceA</i>	-223	5,5	TTGTAGTcAaACTACAA
	<i>zwf-pgl-eda</i>	-88	4,7	TgGTAGTAAagTTACat
	<i>hexR</i>	-88	4,7	TgGTAGTAAagTTACat
	<i>pykA-CsaI_1558</i>	-67	4,3	aTGTAGTAAaATaACcA
<i>Marinomonas</i> sp. MWYL1	<i>gpmI</i>	-70	5,4	сTGTAATtAaATTACAA
	<i>glк-zwf-eda</i>	-102	5,4	TTGTAATtTTATTACat
	<i>gapA2-pykA-CsaI_1558</i>	-309	5,1	aTGTAAATTTACTACat
		-238	5,4	TTGTAATtTTtTTACAA
	<i>edd</i>	-146	5,4	TTGTAAaAaATTACAA
		-75	5,1	aTGTAGTAATATTACat
	<i>ppsA</i>	-191	4,8	TTGaAATAAaACTACat
	<i>gapA3</i>	-89	5,3	TTGTAGTgTTcTTACAA
	<i>pckA</i>	-132	4,7	TgGTAATAAaATaACat
	<i>gnd</i>	-391	4,7	aTGTtATtTTATTACcA
	<i>Mmwyl1_2513-tal</i>	-92	5,3	TTGTAATtTTATgACat
	<i>aceEF</i>	-219	4,4	TgGaAATAAaATTaAAA
<i>nqrABCDEF</i>	-278	4,5	aTGTAAATtAagcaACAA	
<i>Reinekea</i> sp. MED297	<i>oxIT</i>	-167	5,1	TTGAAATTTTACTACAA
	<i>hexR-pykA</i>	-54	4,5	TTGTGAATTTCTACAA
	<i>glк</i>	-419	4,5	TTGTAGGAAATTCACAA
		-223	4,2	TTGTAGTAAAGTTACTC
	<i>gpmM</i>	-36	5,2	TTGTAATTTTCTACAT
	<i>gapB</i>	-138	4,8	TGGTAATAAAAATACAA
	<i>ugpC</i>	-122	5,5	CTGTAATAAACTACAA
	<i>pgk</i>	-95	5,4	TTGTAATAAAAATTACAT
	<i>eno</i>	-101	5,2	TTGTAAGAAACTACAT
	<i>pflA</i>	-143	5,1	TTGTAATTTTGTACAT
	<i>pykF</i>	-346	4,9	ATGTAATTAaATTACAT
	<i>adhE</i>	-196	4,9	TTGTAGTTTTCTTACAC
	<i>gapA</i>	-136	4,4	TTGTAGTTTACTTTCAT
		-60	4,7	TTGTAATTTTACTACGT
	<i>glтA</i>	-222	4,3	TCGTAATATTCTACGA
	<i>pckA</i>	-478	4,9	TCGTAATATCACTACAA
		-223	4,2	CTGTAGTTTAgTTTCAG
<i>Marinobacter</i> sp. ELB17	<i>hexR</i>	-149	5,5	GTAGTAATTTTA
	<i>zwf-pgl-eda</i>	-13	5,5	TAAAATTACTAC
	<i>gapA-pykA</i>	-58	5,2	TACATTTACTAC
	<i>edd</i>	-125	5,2	GTAGTAAATGTA

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>M. aqueolei</i>	<i>zwf-pgl-eda</i>	-63	5,1	TAAAАСТАСТАС
	<i>hexR</i>	-156	5,1	GTAGTAGTTTTTA
	<i>pykA</i>	-175	5,1	ТААСТТТАСТАС
	<i>edd-glk</i>	-112	5,1	GTAGTAAAGTTA
<i>H. chejuensis</i> KCTC 2396	<i>zwf-pgl-eda</i>	-17	5,1	AATAATTACTAC
	<i>gapA-pykA</i>	-49	5,5	TATATTTACTAC
	<i>hexR</i>	-185	5,1	GTAGTAATTATT
	<i>edd-glk-HCH_00435</i>	-185	5,5	GTAGTAAATATA
<b>Pseudomonadaceae-HexR</b>				
<i>P. aeruginosa</i> PAO1	<i>zwf-pgl-eda</i>	-80	4,5	ATGTTGT-(6)-ACTACAT
	<i>hexR1</i>	-126	4,5	ATGTaGT-(6)-ACAACAT
	<i>edd-glk-gltRS</i>	-42	4,6	ATGTTGT-(4)-ACAACAa
	<i>gapA2</i>	-105	4,6	tGTTGT-(4)-ACAACAT
<i>P. entomophila</i> L48	<i>zwf-pgl-eda</i>	-84	4,5	ATGTTGT-(7)-ACTACAT
	<i>hexR1</i>	-139	4,5	ATGTaGT-(7)-ACAACAT
	<i>edd-glk-gltRS</i>	-203	4,4	tGTTGT-(7)-ACAAGAT
		-141	4,2	ATtTTGT-(8)-ACAACga
	<i>gapA2</i>	-147	4,2	tcGTTGT-(8)-ACAAaAT
<i>P. putida</i> KT2440	<i>hexR1</i>	-84	4,4	ATcTTGT-(7)-ACAACAa
		-146	4,5	ATGTaGT-(7)-ACAACAT
	<i>zwf-pgl-eda</i>	-86	4,5	ATGTTGT-(7)-ACTACAT
	<i>gapA2</i>	-144	4,2	tcGTTGT-(8)-ACAAaAT
		-81	4,4	ATcTTGT-(7)-ACAACAa
<i>edd-glk-gltRS</i>	-204	4,4	tGTTGT-(7)-ACAAGAT	
<i>P. mendocina</i> ymp	<i>zwf-pgl-eda-PST_3493</i>	-142	4,2	ATtTTGT-(8)-ACAACga
		-84	4,7	ATGTTGT-(7)-ACAACAT
	<i>hexR1</i>	-139	4,7	ATGTTGT-(7)-ACAACAT
	<i>gapA2</i>	-99	3,9	tgtTTGT-(7)-ACAAaAT
		-37	4,4	ATcTTGT-(7)-ACAACAa
	<i>edd</i>	-112	4,4	tGTTGT-(7)-ACAAGAT
-50		3,9	ATtTTGT-(7)-ACAAaca	
<i>P. stutzeri</i> A1501	<i>zwf-pgl-eda-gapA2-PST_3493</i>	-84	4,5	ATGTTGT-(7)-ACTACAT
	<i>hexR1</i>	-179	4,5	ATGTaGT-(7)-ACAACAT
	<i>edd-glk-gltRS</i>	-38	4,0	tgGTTGT-(6)-ACAAaAT
<i>P. fluorescens</i> Pf-5	<i>zwf-pgl-eda</i>	-84	4,5	ATGTTGT-(7)-ACTACAT
	<i>hexR1</i>	-240	4,5	ATGTaGT-(7)-ACAACAT
	<i>edd-glk-gltRS</i>	-200	4,4	tGTTGT-(7)-ACAAGAT
		-138	4,2	ATtTTGT-(7)-ACAACga
	<i>gapA2</i>	-106	4,2	tcGTTGT-(7)-ACAAaAT
<i>P. syringae</i> pv. <i>tomato</i> str. DC3000	<i>hexR1</i>	-44	4,4	ATcTTGT-(7)-ACAACAa
		-153	4,5	ATGTaGT-(7)-ACAACAT
	<i>zwf-pgl-eda</i>	-86	4,5	ATGTTGT-(7)-ACTACAT

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
	<i>gapA2</i>	-100	4,1	taGTTGT-(7)-ACAaaAT
		-38	4,4	ATcTTGT-(7)-ACAACAa
	<i>edd-glK-gltRS</i>	-202	4,4	tTGTGT-(7)-ACAagAT
		-140	4,1	ATtTTGT-(7)-ACAACta
<i>A. vinelandii AvOP</i>	<i>zwf-eda-pyKA-enj-gapN</i>	-84	5,1	TTGTTTT-(6)-ACAACAT
	<i>pgl-eda</i>	-127	5,3	TCGTAGT-(6)-ACAACAT
	<i>zwf-pgl</i>	-84	4,4	ATGTAGT-(4)-ACGCAAT
	<i>hexR</i>	-111	4,4	ATTGCGT-(6)-ACTACAT
-66		5,3	ATGTTGT-(6)-ACTACGA	
<b>Pseudomonadaceae-HexR1</b>				
<i>P. aeruginosa PAO1</i>	<i>hexR2</i>	-44	6,4	TTGTAGTATAACTACAA
	<i>zwf2</i>	-58	6,4	TTGTAGTTATAACTACAA
	<i>aceEF</i>	-182	5,7	aTGTAGTTTTACTACTa
		-130	6,2	cTGTAGTAAAACTACAA
<i>aldE</i>	-66	6,0	aTGTAGTAAAACTACAc	
<i>P. entomophila L48</i>	<i>hexR2</i>	-59	6,3	aTGTAGTATAACTACAA
	<i>zwf2</i>	-46	6,3	TTGTAGTTATAACTACAt
	<i>aceEF</i>	-178	5,6	gTGTAGTTTTACTACTa
		-126	6,2	TTGTAGTAAAACTACAc
<i>P. putida KT2440</i>	<i>hexR2</i>	-59	6,3	aTGTAGTATAACTACAA
	<i>zwf2</i>	-46	6,3	TTGTAGTTATAACTACAt
	<i>aceEF</i>	-176	5,7	aTGTAGTTTTACTACTa
		-124	6,3	aTGTAGTAAAACTACAA
<i>P. mendocina ymp</i>	<i>hexR2</i>	-104	6,3	aTGTAGTATAACTACAA
	<i>zwf2</i>	-132	6,3	TTGTAGTTATAACTACAt
	<i>aceEF</i>	-180	5,8	TTGTAGTTTTACTACTa
		-127	6,3	aTGTAGTAAAACTACAA
<i>P. stutzeri A1501</i>	<i>zwf2</i>	-58	5,6	cgGTAGTTATAACTACAA
	<i>hexR2</i>	-196	5,6	TTGTAGTATAACTACcg
	<i>aceEF</i>	-180	5,6	cTGTAGTTTTACTACTa
		-129	6,2	TTGTAGTAAAACTACAg
<i>P. fluorescens Pf-5</i>	<i>hexR2</i>	-62	6,2	cTGTAGTATAACTACAA
	<i>aceEF</i>	-177	5,4	cTGTAGTTTTACTACTc
		-125	6,2	TTGTAGTAAAACTACAc
<i>P. syringae pv. tomato str. DC3000</i>	<i>hexR2</i>	-60	6,3	aTGTAGTATAACTACAA
	<i>aceEF</i>	-180	5,4	cTGTAGTTTTACTACTg
		-128	6,2	TTGTAGTAAAACTACAc
<i>A. vinelandii AvOP</i>	<i>zwf3</i>	-42	6,1	cTGTAGTTATAACTACAt
	<i>hexR3</i>	-62	6,1	aTGTAGTATAACTACAg
	<i>aceEF</i>	-115	6,4	TTGTAGTAAAACTACAA
<b>Ralstonia</b>				
<i>R. solanacearum GMI1000</i>	<i>zwf-pgl-glK</i>	-159	5,6	aTGTAGtAATaCTACAA
	<i>edd</i>	-221	5,6	TTGTAGtATTaCTACAt
<i>R. metallidurans CH34</i>	<i>hexR</i>	-247	5,7	aTGTAGAATTTCTACAt
	<i>edd-gntK</i>	-68	5,7	aTGTAGAAATTTCTACAt

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>R. eutropha</i> JMP134	<i>hexR</i>	-165	5,7	aTG TAGATTTTCTACAt
		-24	5,8	TTGTAGAAATATCTACAt
	<i>edd-gntK</i>	-204	5,8	aTG TAGATATTCTACAA
		-63	5,7	aTG TAGAAAACTACAt
<i>R. eutropha</i> H16	<i>edd</i>	-222	5,2	ATGTAGAGATTCTACAT
		-71	5,3	CTGTAGAAAACTACAT
	<i>hexR</i>	-200	5,3	ATGTAGATTTTCTACAG
		-49	5,2	ATGTAGAATCTCTACAT
<i>R. pickettii</i> 12J	<i>edd</i>	-231	5,7	TTGTAGTATTACTACAA
		-55	4,2	TTGTAGgAAAAtTACAg
	<i>zwf-pgl-glk</i>	-269	4,2	CTGTAATTTTCCTACAA
		-93	5,7	TTGTAGTAATACTACAA
<i>C. taiwanensis</i> str. LMG19424	<i>edd</i>	-223	5,2	ATGTAGAGATTCTACAT
		-71	4,8	CTGTAGAAAACTACAG
	<i>hexR</i>	-201	4,8	CTGTAGATTTTCTACAG
		-49	5,2	ATGTAGAATCTCTACAT
<b>Burkholderia</b>				
<i>B. cepacia</i> AMMD	<i>hexR1</i>	-209	5,6	TaGTAGAAAAaCTACAt
		-66	5,8	aTG TAGATTTTCTACAA
	<i>edd-eda-gntUK</i>	-243	5,8	TTGTAGAAAACTACAt
		-100	5,6	aTG TAGtTTTCTACtA
<i>zwf-pgl-(glk/hexR2)</i>	-82	5,7	TTGTAGtTAAaCTACAA	
<i>ectC</i>	-239	5,6	aTG TAGATAAACTACAt	
<i>B. mallei</i> ATCC 23344	<i>zwf-pgl-(glk/hexR2)</i>	-86	5,7	TTGTAGtTAAaCTACAA
	<i>edd-eda-gntUK</i>	-244	5,9	TTGTAGAAAACTACAA
		-101	5,6	aTG TAGtTTTCTACtA
	<i>hexR1</i>	-213	5,6	TaGTAGAAAAaCTACAt
-70		5,9	TTGTAGATTTTCTACAA	
<i>B. vietnamiensis</i> str. G4	<i>hexR1</i>	-211	5,7	TaGTAGAAAAaCTACAA
		-68	5,8	aTG TAGATTTTCTACAA
	<i>edd-eda-gntUK</i>	-244	5,8	TTGTAGAAAACTACAt
		-101	5,7	TTGTAGtTTTCTACtA
<i>zwf-pgl-(glk/hexR2)</i>	-82	5,7	TTGTAGtTAAaCTACAA	
<i>B. xenovorans</i> LB400	<i>zwf-pgl-(glk/hexR2)</i>	-98	5,7	TTGTAGtTAAaCTACAA
	<i>edd-eda-gntUK</i>	-104	5,6	aTG TAGtTTTCTACtA
	<i>hexR1</i>	-227	5,6	TaGTAGAAAAaCTACAt
<i>B. pseudomallei</i> K96243	<i>hexR</i>	-69	5,9	TTGTAGATTTTCTACAA
		-212	5,6	TAGTAGAAAACTACAT
	<i>edd-eda</i>	-245	5,9	TTGTAGAAAACTACAA
		-102	5,6	ATGTAGTTTTTCTACTA
<i>zwf-pgl-(glk/hexR2)</i>	-85	5,7	TTGTAGTTAAACTACAA	
<i>B. sp.</i> 383	<i>hexR</i>	-200	5,6	TAGTAGAAAACTACAT
		-68	5,9	TTGTAGATTTTCTACAA
	<i>edd-eda</i>	-231	5,9	TTGTAGAAAACTACAA
		-99	5,6	ATGTAGTTTTTCTACTA

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
	<i>zwf-pgl-(glk/hexR2)</i>	-84	5,7	TTGTAGTTAAACTACAA
<i>B. glumae BGR1</i>	<i>hexR</i>	-205	5,6	TAGTAGAAAAACTACAT
		-63	5,3	CTGTAGTTTTTCTACAA
	<i>edd-eda</i>	-243	5,3	TTGTAGAAAAACTACAG
		-101	5,6	ATGTAGTTTTTCTACTA
<i>zwf-pgl-(glk/hexR2)</i>	-132	5,7	TTGTAGTTAAACTACAA	
<i>B. phymatum STM815</i>	<i>hexR</i>	-74	4,9	ATGTAGTTTTTGTACTT
		-207	5,3	TAGTAGAAAAACTACAC
	<i>edd-eda</i>	-242	4,9	AAGTACAAAAACTACAT
		-109	5,3	GTGTAGTTTTTCTACTA
	<i>zwf-pgl-(glk/hexR2)</i>	-87	5,7	TTGTAGTTAAACTACAA
<b>Comamonadaceae</b>				
<i>L. cholodni SP-6</i>	<i>edd-eda</i>	-93	5,7	ATGTAACCTTAGTTtCgg
	<i>tal-pgi-Rfer_1128</i>	-220	5,8	cTGTAACtGAGTTtCAg
	<i>zwf-hexR</i>	-48	5,8	cTGaAACTcAGTTACAg
<i>R. ferrireducens DSM 15236</i>	<i>zwf-hexR</i>	-122	6,1	ATGaAACTAAGTTACAT
	<i>tal-pgi-Rfer_1128</i>	-109	6,1	ATGTAACCTTAGTTtCAT
	<i>edd-eda</i>	-33	5,3	ATGTAAtTAAGTTACca
<i>A. avenae subsp. citrulli AAC00-1</i>	<i>pykA</i>	-47	5,6	TTGAAACCCGATTACAT
	<i>zwf</i>	-105	5,1	GTGAAACTCGGTAACAT
<i>Acidovorax sp. JS42</i>	<i>zwf-hexR</i>	-93	5,3	ATGAAACTCGATAACAT
	<i>pykA</i>	-38	5,4	TCGAAACTCGATTACAT
<i>C. testosteroni KF-1</i>	<i>pgk</i>	-157	5,2	CTGAAATCCAGTTACAG
	<i>pykA</i>	-39	5,6	TTGAAACCCGATTACAT
<i>D. acidovorans SPH-1</i>	<i>pgk</i>	-27	5,2	CTGAAACCCGATTACAG
	<i>pykA</i>	-38	5,5	TTGAAACCCGATTACAA
	<i>zwf</i>	-47	4,8	TGGTAACCAAGTTACCA
	<i>pgi</i>	-24	4,3	TGGTAATCAGATTACTT
		-109	4,8	TGGTAACCTGGTTACCA
<i>P. naphthalenivorans CJ2</i>	<i>pykA</i>	-30	5,5	CGGAAACTAAATTACAA
	<i>zwf-hexR</i>	13	4,9	ATGAAACTTTTATTACCT
	<i>edd-eda</i>	-54	4,7	CCGTAATAAAGTTACAT
<i>P. sp. JS666</i>	<i>zwf</i>	-74	5,5	GTGAAACTAAATTACAA
	<i>pykA</i>	-26	5,0	CGGGAACTAAATTACAG
	<i>edd-eda</i>	-34	4,7	CCGTAATAAAGTTACAT
	<i>tal-pgi</i>	-95	4,1	TTGTAATTTAGTTTCAC
<i>V. eiseniae EF01-2</i>	<i>hexR</i>	-44	4,6	TGCAAAACCCGTTACAT
	<i>zwf</i>	-44	5,0	ATGTAACGTGATTACAA
	<i>pgi</i>	-67	4,5	TCGTAATCAAGTTACCA
<b>Neisseriales</b>				
<i>N. meningitidis MC58</i>	<i>zwf-pgl-glk-hexR-pgi</i>	-272	5,5	TTGTAaTTTTgtTACAA
		-116	5,2	gTGTAGTATTACTACtc
	<i>edd-eda</i>	-312	5,5	TTGTAacAAAAtTACAA
	<i>gntUK</i>	-33	5,2	TTGTAcTTTTAtTACAg

<i>Геном</i>	<i>Оперон</i>	<b>Сайт</b>		
		<b>Положение</b>	<b>Вес</b>	<b>Последовательность</b>
<i>C. violaceum</i> ATCC 12472	<i>edd-eda</i>	-227	5,2	aTGTA <sup>G</sup> TTATA <sup>t</sup> TACcA
		-47	5,4	gTGTA <sup>G</sup> TAAA <sup>A</sup> CTAcA
	<i>zwf-pgl-glk-hexR-pgi</i>	-228	5,4	TaGTA <sup>G</sup> TTTT <sup>T</sup> ACTAcAc
		-48	5,2	TgGTA <sup>a</sup> TATA <sup>A</sup> CTAcAt
	<i>ptsAG</i>	-212	5,6	aTGTA <sup>G</sup> TTTT <sup>A</sup> CTAcAA

**Приложение 4.** Состав регулонов метаболизма сахаров в семействе бактерий *Vacillaceae*. «+» - сайт связывания обнаружен в 5`-некодирующей области оперона, «-» - сайт связывания обнаружен не был, «0» - ортологичные гены не были найдены в геноме. Первая строка каждого регулона (отмечена жирным шрифтом) отображает присутствие гена, кодирующего фактор транскрипции в геноме. Колонка «Эксп.» содержит ссылки на работы по экспериментальному изучению регуляции соответствующего оперона.

Регулируемые опероны	Примеры первого гена в опероне											Эксп.
		<i>BSU</i>	<i>BAY</i>	<i>BPU</i>	<i>BLI</i>	<i>AFL</i>	<i>GKA</i>	<i>BCE</i>	<i>BHA</i>	<i>BCL</i>	<i>O/H</i>	
<b>ManR</b> (утилизация маннозы)	<b>manR</b>	+	+	0	+	0	0	0	0	0	0	-
	<i>manR</i>	<i>BSU12000</i>	+	+	0	-	0	0	0	0	0	(178)
	<i>manPA</i>	<i>BSU12010</i>	+	+	0	+	0	0	0	0	0	(178)
<b>LicR</b> (утилизация бета-глюкозидов)	<b>licR</b>	+	+	+	+	0	0	+	0	+	+	-
	<i>licBCAH</i>	<i>BSU38590</i>	+	+	+	-	0	0	-	0	+	(130)
	<i>licR</i>	<i>BSU38600</i>	-	-	-	-	0	0	-	0	+	-
<b>MtlR</b> (утилизация маннитола)	<b>mtlR</b>	+	+	+	+	+	+	0	+	+	+	-
	<i>mtlAFD</i>	<i>BSU03981</i>	+	+	+	+	+	0	+	-	+	(179,180)
	<i>mtlR</i>	<i>BSU04160</i>	+	+	+	-	+	+	0	+	-	-
<b>CcpN</b> (глюконеогенез)	<b>ccpN</b>	+	+	+	+	+	+	+	+	+	+	-
	<i>gapB</i>	<i>BSU29020</i>	+	+	+	+	+	+	+	+	+	(176)
	<i>pckA</i>	<i>BSU30560</i>	+	+	+	+	+	+	+	+	+	(176)
<b>CitT</b> (транспорт цитрата)	<b>citT</b>	+	+	+	+	0	0	+	+	+	+	-
	<i>citM</i>	<i>BSU07610</i>	+	+	+	+	0	0	0	+	+	(181)
	<i>tctCBA</i>	<i>OB3249</i>	0	0	0	0	0	0	0	0	+	-
	<i>citH</i>	<i>BC0562</i>	0	0	0	0	0	0	+	0	0	-
<b>MalR</b> (утилизация малата)	<b>malR</b>	+	+	+	+	0	0	0	0	0	0	-
	<i>ywkAB</i>	<i>BSU37050</i>	+	+	+	+	0	0	0	0	0	(147)
	<i>maeN</i>	<i>BSU31580</i>	+	0	+	+	0	0	-	0	-	(182)
	<i>yflS</i>	<i>BSU07570</i>	+	+	0	-	0	0	0	0	0	(182)
<b>IoIR</b> регулон (утилизация инозитола)	<b>ioIR</b>	+	+	0	+	0	0	0	0	+	0	-
	<i>ioIABCDEFGHIJ</i>	<i>BSU39760</i>	+	+	0	+	0	-	0	-	+	(183)
	<i>ioIRS</i>	<i>BSU39770</i>	+	+	0	+	0	0	0	0	+	(183)
<b>AcoR</b> (утилизация ацетоина)	<b>acoR</b>	+	+	+	+	0	0	+	+	0	0	-
	<i>acoABCL</i>	<i>BSU08060</i>	+	+	+	+	0	0	+	+	-	(185)
<b>FrlR</b> (утилизация фруктолизина)	<b>frlR</b>	+	+	0	0	0	0	0	0	0	0	-
	<i>frlR</i>	<i>BSU32560</i>	+	+	0	0	0	0	0	0	0	(186)
	<i>frlB</i>	<i>BSU32610</i>	+	+	0	0	0	0	0	0	0	(186)
	<i>frlONMD</i>	<i>BSU32600</i>	+	+	0	0	0	0	0	0	0	(186)
	<i>yurJ</i>	<i>BSU32550</i>	+	+	0	0	0	0	0	0	0	-
<b>LutR</b> (утилизация лактата)	<b>lutR</b>	+	+	+	+	+	+	+	+	+	+	-
	<i>lutABC</i>	<i>BSU34050</i>	+	+	+	+	+	+	+	+	+	(187)
	<i>lutR</i>	<i>BSU34180</i>	+	+	-	+	-	-	-	-	+	-
	<i>lutP</i>	<i>BSU34190</i>	+	+	0	+	+	+	0	+	+	-
<b>GmuR</b> (утилизация глюкоманнана)	<b>gmuR</b>	+	+	+	+	0	+	0	+	0	0	-
	<i>gmuBACDR</i>	<i>BSU05810</i>	+	+	+	+	0	+	0	+	-	(188)
	<i>gmuE</i>	<i>BSU05860</i>	+	+	-	-	0	0	0	0	0	(188)

Регулируемые опероны	Примеры первого гена в опероне	BSU	BAY	BPU	BLI	AFL	GKA	BCE	BHA	BCL	OIH	Эксп.
<i>gmuF</i>	<i>BSU05880</i>	+	+	0	0	0	0	0	+	0	0	(188)
<b>GudR (утилизация глюкората и галактората)</b>	<b>gudR</b>	+	0	0	+	0	0	0	0	+	+	-
<i>gudR</i>	<i>BSU02500</i>	+	0	0	+	0	0	0	0	+	+	(189)
<i>garD</i>	<i>BSU02510</i>	+	0	0	+	0	0	0	0	0	0	(189)
<i>ycbCD-gudD</i>	<i>BSU02460</i>	+	0	0	+	0	0	0	0	+	+	(189)
<i>gudP</i>	<i>BSU02480</i>	+	0	0	+	0	0	0	0	0	0	(189)
<i>ABC0469-67</i>	<i>ABC0469-67</i>	0	0	0	0	0	0	0	0	+	+	-
<b>NagR (утилизация N-ацетилгалактозамина)</b>	<b>nagR</b>	+	+	+	+	0	+	+	+	+	+	-
<i>nagABR</i>	<i>BSU35010</i>	+	+	+	+	0	+	+	+	+	+	(190)
<i>nagP</i>	<i>BSU07700</i>	+	+	+	+	0	0	-	+	0	0	(190)
<i>murQ2</i>	<i>BLI04351</i>	0	0	0	+	0	0	0	0	0	0	-
<b>TreR (утилизация трехалозы)</b>	<b>treR</b>	+	+	+	+	+	+	+	+	+	0	-
<i>trePAR</i>	<i>BSU07800</i>	+	+	+	+	+	+	+	+	+	0	(191)
<b>GntR (утилизация глюконата)</b>	<b>gntR</b>	+	0	0	+	0	0	0	0	0	+	-
<i>gntRKPZ</i>	<i>BSU40050</i>	+	0	0	+	0	0	0	0	0	+	(144)
<b>AraR (утилизация арабинозы)</b>	<b>araR</b>	+	+	+	+	+	+	0	+	0	+	-
<i>araFGH</i>	<i>GK1910</i>	0	0	0	-	0	+	0	-	-	0	-
<i>abnA</i>	<i>BSU28810</i>	+	+	0	-	0	0	0	0	0	0	(97)
<i>araE</i>	<i>BSU33960</i>	+	+	+	+	0	0	0	0	0	+	(99)
<i>araR</i>	<i>BSU33970</i>	+	+	+	+	+	+	0	+	0	+	(99)
<i>xsa</i>	<i>BSU28510</i>	+	+	0	0	0	0	0	+	-	0	(97)
<i>araABDLMNPQ-abfA</i>	<i>BSU28720</i>	+	+	+	+	+	+	0	-	-	+	(99)
<i>araK</i>	<i>BPUM_2329</i>	0	0	+	+	0	0	0	0	0	0	-
<b>GutR (утилизация сорбитола)</b>	<b>gutR</b>	+	+	0	0	0	0	0	0	0	0	-
<i>gutBP</i>	<i>BSU06150</i>	+	+	0	0	0	0	0	0	0	0	(192)
<b>HxlR (рибулоза монофосфатный путь)</b>	<b>hxlR</b>	+	+	+	+	0	0	0	0	0	+	-
<i>hxlAB</i>	<i>BSU03460</i>	+	+	+	+	0	0	0	0	0	+	(193)
<b>NtdR (утилизация неотрахалозадиамина)</b>	<b>ntdR</b>	+	0	+	+	0	0	0	0	0	0	-
<i>ntdABC</i>	<i>BSU10550</i>	+	0	+	+	0	0	0	0	0	0	(194)
<i>ntdR</i>	<i>BSU10560</i>	+	0	+	+	0	0	0	0	0	0	-
<b>GanR (утилизация галактана)</b>	<b>ganR</b>	+	0	+	+	0	0	0	+	+	0	-
<i>sucB-ganPQAB</i>	<i>BSU34160</i>	+	0	+	+	-	0	0	+	+	0	(195)
<i>ganR</i>	<i>BSU34170</i>	+	0	+	+	0	0	0	-	-	0	-
<b>CcpB (катаболитная репрессия)</b>	<b>ccpB</b>	+	+	0	0	0	0	0	0	0	0	-
<i>gntRKPZ</i>	<i>BSU40050</i>	+	0	0	-	0	0	0	0	0	-	(168)
<i>xylAB</i>	<i>BSU17600</i>	+	+	-	-	-	-	-	-	-	-	(168)
<b>KdgR (утилизация 2-кето-3-деоксиглюконата)</b>	<b>kdgR</b>	+	0	+	+	0	0	0	0	0	0	-
<i>kdgRKAT</i>	<i>BSU22120</i>	+	0	+	+	0	0	0	0	0	0	(217)
<i>kduID</i>	<i>BSU22130</i>	+	0	+	+	0	0	0	-	-	-	(217)
<b>ExuR (утилизация гексоуроната)</b>	<b>exuR</b>	+	+	+	+	0	0	0	0	0	0	-
<i>uxaC--exuM-ujmCD-uxuA-ujmF-exuTR-uxaBA</i>	<i>BSU12300</i>	+	+	+	+	0	0	0	0	0	0	(218)
<b>RbsR (утилизация рибозы)</b>	<b>rbsR</b>	+	+	+	+	+	+	+	+	+	+	-
<i>rbsRKDACB</i>	<i>BSU35910</i>	+	+	+	+	+	+	+	+	+	+	-
<b>LevR (утилизация леванов)</b>	<b>levR</b>	+	0	0	+	0	0	0	0	0	0	-

Регулируемые опероны	Примеры первого гена в опероне											Эксп.
		BSU	BAY	BPU	BLI	AFL	GKA	BCE	BHA	BCL	OIH	
<i>levDEFG-sacC</i>	<i>BSU27070</i>	+	0	0	+	0	0	0	0	0	0	(136)
<b>AlsR (биосинтез ацетоацетата)</b>	<i>alsR</i>	+	+	+	+	0	0	+	0	0	0	-
	<i>alsR</i>	<i>BSU36020</i>	+	+	+	+	0	0	-	0	0	(220)
	<i>alsSD</i>	<i>BSU36010</i>	+	+	+	+	0	0	-	0	0	(220)
<b>CitR (биосинтез цитрата)</b>	<i>citR</i>	+	+	0	+	0	0	0	0	+	0	-
	<i>citA</i>	<i>BSU09940</i>	+	+	0	+	0	0	0	0	+	(221)
	<i>citR</i>	<i>BSU09930</i>	+	+	0	+	0	0	0	0	+	-
<b>CcpC (катаболитная репрессия)</b>	<i>ccpC</i>	+	+	+	+	0	0	+	0	0	+	-
	<i>ccpC</i>	<i>BSU14140</i>	+	-	+	+	0	0	-	0	+	(172)
	<i>citB</i>	<i>BSU18000</i>	+	+	+	+	-	-	+	-	+	(169)
	<i>citZ-icd-mdh</i>	<i>BSU29140</i>	+	+	-	-	-	-	-	-	-	(169)
<b>XylR (утилизация ксилозы)</b>	<i>xylR</i>	+	+	+	+	0	+	0	+	+	+	-
	<i>xylAB</i>	<i>BSU17600</i>	+	+	+	+	0	+	0	+	+	(146)
	<i>xynPB</i>	<i>BSU17570</i>	+	+	+	-	0	0	0	+	-	(145)
	<i>xylR</i>	<i>BSU17590</i>	+	+	+	+	0	+	0	+	-	-
	<i>OB3123-OB3122-OB3121</i>	<i>OB3123</i>	0	0	0	0	0	0	0	+	0	-
	<i>BH3678-BH3679</i>	<i>BH3678</i>	0	0	0	0	0	0	0	+	0	-
	<i>gunAB</i>	<i>BLi01880</i>	0	0	-	+	0	0	0	0	0	-
	<i>BLi03540-BLi03541-yurM-xylS-bgII</i>	<i>BLi03540</i>	0	0	0	+	0	0	0	0	-	-
	<i>xynA</i>	<i>BH2120</i>	0	0	0	0	0	0	0	+	0	-
<i>xylF</i>	<i>GK1881</i>	0	0	0	-	-	+	0	0	0	-	
<b>GlvR (утилизация мальтозы)</b>	<i>glvR</i>	+	+	0	+	0	0	0	0	0	0	-
	<i>glvARC</i>	<i>BSU08180</i>	+	+	0	+	0	0	0	0	0	(203)
<b>CggR (гликолиз)</b>	<i>cggR</i>	+	+	+	+	+	+	+	+	+	+	-
<i>cggR-gapA-pgk-tpiA-pgm-eno</i>	<i>BSU33950</i>	+	+	+	+	+	+	+	+	+	+	(177)
<b>FruR (утилизация фруктозы)</b>	<i>fruR</i>	+	+	+	+	+	+	+	+	+	+	-
	<i>fruRKA</i>	<i>BSU14380</i>	+	+	+	+	+	+	+	+	+	-
<b>YkvZ (утилизация фррил-фосфо-глюкозидов)</b>	<i>ykvZ</i>	+	+	+	+	0	0	+	0	0	0	-
	<i>bgIC</i>	<i>BSU03410</i>	+	+	+	+	0	-	0	0	0	-
	<i>ykvZ</i>	<i>BSU13870</i>	+	+	+	-	0	0	+	0	0	-
	<i>BC2618</i>	<i>BC2618</i>	0	0	0	0	0	0	+	0	0	-
<b>MdxR (утилизация мальтодекстрина)</b>	<i>mdxR</i>	+	0	0	+	0	0	0	0	0	0	-
	<i>yvdGHIJK-malL-pgcM</i>	<i>BSU34610</i>	+	0	0	+	0	0	0	0	0	-
<b>MurR (утилизация N-ацетилмуромата)</b>	<i>murR</i>	+	0	+	+	0	0	0	0	0	0	-
	<i>murQR-ybbF-amiE-nagZ-ybbC</i>	<i>BSU01700</i>	+	-	+	+	0	0	-	-	-	-
<b>MsmR (утилизация альфа-галактозидов)</b>	<i>msmR</i>	+	+	+	+	0	0	0	+	0	0	-
	<i>msmREFG-melA</i>	<i>BSU30260</i>	+	+	+	+	0	0	0	+	-	0
<b>BglR (утилизация арил-фосфо-глюкозидов)</b>	<i>bglR</i>	+	+	+	0	0	0	0	0	0	0	-
	<i>bglR</i>	<i>BSU40130</i>	+	+	+	0	0	0	0	0	0	-
	<i>bglA</i>	<i>BSU40110</i>	+	+	+	0	0	0	0	0	0	-
<b>RhgR (утилизация рамногалактуронана)</b>	<i>rhgR</i>	+	0	0	+	0	0	0	0	0	0	-
	<i>yesOPQ</i>	<i>BSU06970</i>	+	0	0	+	0	0	0	-	-	0
	<i>yesRSTUVWXYZ</i>	<i>BSU07000</i>	+	0	0	+	0	0	0	0	0	-
<b>GamR (утилизация глюкозамина и хитобиозы)</b>	<i>gamR</i>	+	0	0	+	0	0	0	+	+	+	-
	<i>gamBP</i>	<i>BSU02360</i>	+	0	0	0	0	0	0	0	0	-

Регулируемые опероны	Примеры первого гена в опероне	BSU	BAY	BPU	BLI	AFL	GKA	BCE	BHA	BCL	OIH	Эксп.
<i>gamR</i> <i>licH-ywbABC</i> <i>chiA</i> <i>BH0913</i>	<i>BSU02370</i>	+	0	0	+	0	0	0	-	-	+	-
	<i>BLI00335</i>	0	0	0	+	0	0	0	+	+	+	-
	<i>BH0916</i>	0	0	0	-	0	0	-	+	-	-	-
	<i>BH0913</i>	0	0	0	+	0	0	-	+	+	+	-
<b>RhaR (утилизация рамнозы)</b> <i>yuxG-rhaR-yuICDE</i> <i>rhaY</i> <i>rhaL</i>	<b>rhaR</b>	+	0	0	+	0	0	0	+	0	+	-
	<i>BSU31220</i>	+	0	0	+	0	0	0	+	-	+	-
	<i>OB0494</i>	0	0	0	-	0	0	0	0	-	+	-
	<i>BLI03559</i>	0	0	0	+	0	0	0	0	0	0	-
<b>DegA (утилизация инозитола)</b> <i>yrbE</i> <i>iolX</i> <i>GK1894-93-iolIDEBCAJ</i> <i>degA</i> <i>GK1899-96</i> <i>BH2222</i>	<b>degA</b>	+	0	0	+	0	+	0	+	0	0	-
	<i>BSU27770</i>	+	0	0	0	0	0	0	+	0	0	-
	<i>BSU10850</i>	+	0	0	+	0	0	0	+	-	0	-
	<i>GK1894</i>	-	-	0	-	0	+	0	-	-	0	-
	<i>BSU10840</i>	-	0	0	-	0	-	0	+	0	0	-
	<i>GK1899</i>	0	0	0	0	0	+	0	+	-	0	-
	<i>BH2222</i>	0	0	0	-	0	0	0	+	-	0	-
<b>RmgR (утилизация рамногалактуронана)</b> <i>ytePRSTU</i> <i>uxaAB</i> <i>BH0493-peIX</i> <i>pgIR</i> <i>xyIB</i> <i>rhgT</i>	<b>rmgR</b>	+	0	+	+	0	0	0	+	+	+	-
	<i>BSU30135</i>	+	0	+	+	0	0	0	+	+	+	-
	<i>ABC1153</i>	0	0	0	0	0	0	0	+	+	0	-
	<i>BH0493</i>	0	0	0	0	0	0	0	+	0	0	-
	<i>OB2088</i>	0	0	0	0	0	0	0	0	0	+	-
	<i>OB2087</i>	0	0	0	0	0	0	0	0	+	+	-
	<i>OB2084</i>	0	0	0	0	0	0	0	0	0	+	-
<b>CcpA (катаболитная репрессия)</b> <i>cimH</i> <i>amyE</i> <i>gntRKPZ</i> <i>kdgRKAT</i> <i>manR</i> <i>citST</i> <i>levDEFG-sacC</i> <i>ccpC</i> <i>araE</i> <i>yxjC-scoAB-yxjF</i> <i>abnA</i> <i>dctA</i> <i>phoPR</i> <i>trePAR</i> <i>msmX</i> <i>galT</i> <i>acoABCL</i>	<b>ccpA</b>	+	+	+	+	+	+	+	+	+	+	-
	<i>BSU38770</i>	+	+	0	-	0	0	0	0	0	0	(152)
	<i>BSU03040</i>	+	+	0	0	0	0	0	0	0	0	(167)
	<i>BSU40050</i>	+	0	0	+	0	0	0	0	+	+	(167)
	<i>BSU22120</i>	+	0	+	+	0	0	0	0	0	0	(152)
	<i>BSU12000</i>	+	+	0	+	0	0	0	0	0	0	(178)
	<i>BSU07580</i>	+	+	-	+	0	0	+	+	+	-	(210)
	<i>BSU27070</i>	+	0	0	+	0	0	0	0	0	+	(216)
	<i>BSU14140</i>	+	+	-	-	0	-	+	-	-	-	(172)
	<i>OB2796</i>	+	+	+	+	0	0	0	0	0	+	(199)
	<i>BSU39000</i>	+	+	0	0	0	0	-	0	+	0	(152)
	<i>BSU28810</i>	+	+	0	+	0	0	0	0	0	0	(198)
	<i>BSU04470</i>	+	+	+	0	0	0	0	0	0	0	(205)
	<i>BSU29110</i>	+	+	-	+	-	-	-	-	-	-	(201)
	<i>BSU07800</i>	+	+	+	+	-	-	+	+	-	0	(152)
	<i>ABC1514</i>	+	+	+	+	+	-	0	+	+	+	(152)
	<i>BSU38190</i>	+	+	+	+	0	-	0	-	+	0	(152)
	<i>BSU08060</i>	+	+	+	+	0	0	-	-	-	0	(152)
	<i>hutPHUIGM</i> <i>cccA</i> <i>glvARC</i> <i>bglPH</i> <i>citM</i>	<i>BSU39340</i>	+	+	0	0	0	+	+	+	-	0
<i>BSU25190</i>		+	+	+	+	-	-	-	-	-	-	(200)
<i>BSU08180</i>		+	+	-	+	0	0	0	0	+	0	(203)
<i>BSU39270</i>		+	+	+	+	0	0	0	0	0	0	(142)
<i>BSU07610</i>		+	+	+	+	0	-	-	-	-	0	(152)

Регулируемые опероны	Примеры первого гена в опероне											Эксп.
		BSU	BAY	BPU	BLI	AFL	GKA	BCE	BHA	BCL	OIH	
<i>licBCAH</i>	BSU38590	+	+	+	+	0	0	0	0	0	0	(213)
<i>acuABC</i>	BSU29690	+	+	+	+	+	-	-	-	-	-	(212)
<i>bglS</i>	BSU39070	+	+	+	0	0	0	0	+	0	0	(214)
<i>xynPB</i>	BSU17570	+	+	+	0	0	0	0	0	+	-	(145)
<i>ackA</i>	BSU29470	+	+	+	+	+	-	-	-	-	-	(164)
<i>citZ-icd-mdh</i>	BSU29140	+	+	+	+	+	-	-	-	+	-	(202)
<i>uxaC-exuM-yjmCD-uxuA-yjmF-exuTR-uxaBA</i>	BSU12300	+	+	+	+	0	0	0	0	+	0	(152)
<i>gmuBACDREFG</i>	BSU05810	+	+	+	+	0	+	0	+	+	+	(197)
<i>pta</i>	BSU37660	+	+	+	+	-	-	-	+	+	-	(209)
<i>xylAB</i>	BSU17600	+	+	+	+	0	+	0	+	+	-	(167)
<i>acoR</i>	BSU08100	+	+	+	+	0	0	+	+	0	0	(185)
<i>lcfA</i>	BSU28560	+	+	+	+	+	-	-	+	-	-	(152)
<i>sigL</i>	BSU34200	+	+	+	+	+	+	+	-	+	+	(211)
<i>rbsRKDACB</i>	BSU35910	+	+	+	-	+	-	+	-	+	+	(204)
<i>ioIABCDEFGHJI</i>	BSU39760	+	+	0	+	0	-	+	+	+	+	(152)
<i>araABDLMNPQ-abfA</i>	BSU27310	+	+	+	+	-	-	0	+	+	+	(199)
<i>ilvBHC-leuABCD</i>	BSU28310	+	+	-	-	-	-	-	-	-	+	(208)
<i>acsA</i>	BSU29680	+	+	+	+	+	-	-	+	+	+	(212)
<i>glpFK</i>	BSU09280	+	-	+	+	-	+	+	+	+	+	(207)
<i>mmgABCDE-prpB</i>	BSU24170	+	+	+	+	+	+	-	+	+	+	(215)
<i>ylbBC</i>	BSU14950	+	+	+	+	-	-	-	-	-	-	-
<i>mrp</i>	BSU01540	+	+	+	+	+	-	+	-	-	-	-
<i>yqgQ-glcK</i>	BSU24850	+	-	+	-	+	-	+	-	-	-	-
<i>yesOPQR-rhgR-yesTUVWXYZ-yetA-lplABCD</i>	BSU06970	+	0	0	+	0	0	0	0	0	0	-
<i>yhaR</i>	BSU09880	+	+	-	+	-	-	-	0	0	0	-
<i>cspD</i>	BSU21930	+	+	-	+	-	-	-	0	0	0	-
<i>nirC</i>	BSU38060	+	0	0	-	+	0	+	0	0	0	-
<i>citH</i>	BSU39060	+	+	0	0	0	0	+	0	0	0	-
<i>ganB</i>	BSU34120	+	0	0	+	0	0	0	+	0	0	-
<i>cstA</i>	BSU28710	+	+	+	+	0	0	0	0	0	0	-
<i>nupC-pdp</i>	BSU39430	+	+	+	+	-	0	0	0	0	0	-
<i>ywfl</i>	BSU37670	+	+	+	+	-	-	-	-	+	-	-
<i>glcDF</i>	BSU28680	+	+	-	+	+	-	-	-	-	-	-
<i>sucCD</i>	BSU16090	+	+	+	+	-	-	-	-	-	-	-
<i>ykoM</i>	BSU13340	+	+	+	+	0	0	0	0	0	0	-
<i>ylbP</i>	BSU15100	+	+	+	+	0	0	0	-	+	+	-
<i>araR</i>	BSU39970	+	+	+	+	-	-	0	-	0	+	-
<i>abnB</i>	BSU39330	+	+	-	+	0	0	0	+	+	0	-
<i>gatCAB</i>	BSU06670	+	+	+	+	-	-	-	-	-	-	-
<i>yqgX</i>	BSU24790	+	+	+	+	-	-	-	+	+	-	-
<i>glpTQ</i>	BSU02140	+	+	+	+	0	0	+	0	0	0	-
<i>drm-punA</i>	BSU23500	+	+	+	+	+	-	+	-	-	-	-
<i>apbA-yllA</i>	BSU15110	+	+	+	+	-	-	-	-	+	-	-
<i>BPUM_3238-36-pmi</i>	BSU35790	+	-	+	0	0	+	0	+	0	-	-
<i>cycB-ganPQA</i>	BSU34160	+	0	+	+	-	0	-	+	+	0	-

Регулируемые опероны	Примеры первого гена в опероне										Эксп.	
		BSU	BAY	BPU	BLI	AFL	GKA	BCE	BHA	BCL		OIH
<i>yticPQ</i>	BSU30170	+	0	+	+	0	0	0	+	+	-	-
<i>kduID</i>	BSU22130	+	0	+	+	0	0	0	+	+	-	-
<i>lrgAB</i>	BSU28910	+	+	+	+	0	0	+	0	0	0	-
<i>glsA1-glnT</i>	BSU02430	+	+	+	+	0	0	-	+	0	-	-
<i>sdhCAB-ysmA</i>	BSU28450	+	+	+	+	-	-	+	-	-	+	-
<i>fruRKA</i>	BSU14380	+	+	+	+	+	-	+	-	-	-	-
<i>yqgY</i>	BSU24780	+	+	+	+	-	-	-	-	+	+	-
<i>yqgW</i>	BSU24800	+	+	+	+	-	-	-	+	+	-	-
<i>msmREFG-melA</i>	BSU30260	+	+	+	+	0	0	0	+	-	0	-
<i>ndk</i>	BSU22720	+	+	+	+	+	+	-	0	+	-	-
<i>yvfVWY</i>	BSU34050	+	+	-	+	+	+	+	+	+	+	-
<i>lutR</i>	BSU34180	+	+	-	-	+	+	-	+	-	-	-
<i>odhAB</i>	BSU19370	+	+	+	+	0	+	+	+	+	+	-
<i>sacPA</i>	BSU38050	+	+	+	+	+	-	+	+	0	0	-
<i>mtlAFD</i>	BSU03981	+	+	+	+	+	+	0	+	-	+	-
<i>xsa</i>	BSU28510	+	+	-	-	0	0	0	-	+	-	-
<i>yngIHDBGFE</i>	BSU18250	+	+	+	+	-	+	-	+	+	-	-
<i>licT-bglP1</i>	BSU39080	-	-	-	-	0	0	0	+	+	0	-
<i>ywcBA</i>	BSU38230	-	0	-	-	0	+	-	0	+	+	-
<i>dagA</i>	BSU18610	-	0	0	+	0	0	0	+	0	+	-
<i>licR</i>	BSU38600	-	-	+	+	0	0	0	0	-	-	-
<i>rmgR</i>	BSU30150	-	0	-	-	0	0	0	+	+	0	-
<i>galKE</i>	BSU39410	-	+	+	+	0	-	0	-	+	0	-
<i>ganR</i>	BSU34170	-	-	0	+	0	0	0	+	-	0	-
<i>mutBA_BH2954</i>	GK2371	0	0	0	0	+	+	0	-	0	0	-
<i>ABC0302</i>	ABC0302	0	0	0	-	0	0	0	0	+	+	-
<i>lplBC-upgB</i>	ABC1135	0	0	0	0	0	0	0	+	+	0	-
<i>BH3448-46</i>	BH3448	0	0	0	+	0	-	0	+	0	0	-
<i>BH3680-82-xynB-BH3684</i>	BH3680	0	0	0	0	0	0	0	+	0	+	-
<i>sdcS</i>	ABC3994	0	0	0	+	0	0	0	0	+	+	-
<i>citMII</i>	BH0745	0	0	+	0	0	0	0	+	+	0	-
<i>ABC3348-gatABC</i>	ABC3348	0	0	0	0	0	0	0	+	+	+	-
<i>tctCBA</i>	ABC1013	0	0	0	+	0	0	0	0	+	+	-
<i>uxuA-fabG</i>	BH0701	0	0	0	+	0	0	0	+	0	+	-
<i>dctPMQ-BH0704</i>	BH0705	0	0	0	-	0	0	0	+	-	+	-
<i>dapA-eutG</i>	ABC0222	0	0	+	+	0	0	0	0	+	0	-
<i>kdgT-hop-pdxA</i>	ABC0308	0	0	+	+	0	0	0	0	+	+	-
<i>araFGH</i>	ABC0409	0	0	0	+	-	+	-	+	+	0	-
<i>lctP</i>	BH1831	0	0	0	0	+	0	0	+	+	+	-
<i>sacK</i>	BC0773	0	0	0	0	+	0	+	+	0	0	-
<i>dat</i>	BLI02962	0	0	+	+	+	+	+	0	0	0	-