

На правах рукописи

Хорошкин Матвей Сергеевич

**Реконструкция регулонов метаболических путей в бактериях  
микробиоты кишечника человека**

03.01.09 – математическая биология, биоинформатика

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени кандидата  
биологических наук

Москва – 2019

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН).

**Научный руководитель:**

**к.б.н. Родионов Дмитрий Александрович** заведующий сектором № 6 «Функциональная и сравнительная геномика прокариот» Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН)

**Оппоненты:**

**д.б.н. Озолинь Ольга Николаевна** зав. лабораторией функциональной геномики и клеточного стресса Федерального государственного бюджетного учреждения науки Института биофизики клетки Российской академии наук (ИБК РАН)

**к.б.н.Тяхт Александр Викторович** научный сотрудник лаборатории моделирования и терапии наследственных заболеваний Федерального государственного бюджетного учреждения науки Институт биологии гена Российской академии наук (ИБГ РАН)

**Ведущее учреждение:**

Федеральный исследовательский центр «Фундаментальные основы биотехнологии» Российской академии наук (ФИЦ Биотехнологии РАН)

Защита диссертации состоится «20» мая 2019 г. в 15-00 на заседании диссертационного совета Д 002.077.04 при Федеральном государственном бюджетном учреждении науки Институте проблем передачи информации им. А.А. Харкевича РАН, расположенном по адресу: 127994, г. Москва, ГСП-4, Большой Каретный пер., д. 19, стр. 1.

С текстом диссертации и автореферата можно ознакомиться в библиотеке Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук и на сайте [www.iitp.ru](http://www.iitp.ru). Автореферат разослан «\_\_» \_\_\_\_\_ 2019 г. Отзывы и замечания по автореферату в двух экземплярах, заверенные печатью, просьба высылать по вышеуказанному адресу на имя учёного секретаря диссертационного совета.

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

Рожкова Г.И.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

### **Актуальность проблемы**

Микробиом кишечника тесно взаимодействует с организмом хозяина и влияет на его жизненные системы. Патологии микробиома связывают со многими нейродегенеративными заболеваниями, иммунными заболеваниями и даже раком. Структура микробиома каждого индивидуума зависит как от генетических факторов хозяина, так и от его питания и образа жизни. Перспективы манипуляции микробиома могут существенно улучшить терапию различных заболеваний, однако на сегодняшний день методы, позволяющие менять структуру микробиома, развиты крайне слабо. Для применения таких методов на практике необходимы фундаментальные исследования, позволяющие оценить возможности и границы применения методов манипуляции микробиома кишечника.

Структура микробиома кишечника меняется в зависимости от питания организма хозяина, поэтому один из наиболее универсальных и легко-применимых способов манипуляции микробиомом кишечника является модуляция его структуры при помощи тщательного контроля питания организма хозяина, в том числе добавки определенных соединений в состав диеты. Однако, на данный момент, про метаболизм бактерий, населяющих человеческий кишечник, известно недостаточно много, чтобы с уверенностью предсказывать изменения численности различных организмов в ответ на добавление определенных соединений в диету организма хозяина. Метаболические системы были детально изучены только для отдельных модельных организмов, тогда как метаболический потенциал тысяч видов бактерий из кишечного микробиома человека остается до сих пор не известным. При этом в последние двадцать лет стремительно растет число расшифрованных бактериальных геномов. Даже при небольшом количестве доступных экспериментальных данных из геномных последовательностей можно получить важнейшую информацию о строении и регуляции метаболических систем бактерий. С этой целью была разработана группа методов сравнительной геномики.

Бактерии могут быстро перестраивать программу экспрессии своих генов в зависимости от наличия питательных субстратов в окружающей среде. Ключевым шагом в этом процессе является регуляция инициации транскрипции. Белки, называемые транскрипционными факторами, могут специфически связывать молекулу-эффе́ктор, при этом изменяется их конформация и специфичность связывания ДНК. Таким образом, при резком увеличении или уменьшении концентрации молекулы-эффе́ктора во внешней среде транскрипционные факторы могут занимать или освобождать регуляторные участки ДНК и таким образом быстро изменять экспрессию генов. Транскрипционные факторы формируют сложные регуляторные сети, которые определяют, как та или иная бактерия реагирует на изменение условий окружающей среды. Несмотря на то, что экспериментальные методы изучения таких регуляторных взаимодействий плохо масштабируемы и крайне трудоемки, методы сравнительной геномики позволяют строить точные предсказания даже для сложных регуляторных сетей.

Изучение метаболических путей и регуляторных систем у бактерий кишечного микробиома является ключевым фундаментальным шагом для последующей разработки методов манипуляции микробиома. В условиях стремительно растущих объемов данных по расшифрованным геномам бактерий, методы сравнительной геномики хорошо

подходят для решения данной задачи. Эти методы выгодно отличаются от других методик ценой, скоростью получения результатов и масштабируемостью.

### **Степень разработанности темы**

Микробиому кишечника человека посвящено много исследований, однако они покрывают далеко не все важнейшие ниши. Подавляющее большинство исследований можно отнести к одной из двух категорий: (1) описание изменений филогенетического состава микробиома в зависимости от изменения внешних условий, в т.ч описание ассоциации изменений микробиома с патологическими состояниями организма-носителя, и (2) детальное изучение крайне малой выборки бактерий-представителей кишечного микробиома. Работы из первой категории крайне важны для ассоциативного описания структуры микробиома, но не развивают понимания механизма метаболических систем, их регуляции и межорганизменных взаимодействий. С другой стороны, работы из второй категории крайне медленны, дорогостоящи и немасштабируемы. Применяемые в данной работе методы сравнительной геномики позволяют охватить нишу широкомасштабного изучения механистических аспектов биологии кишечного микробиома.

### **Цели и задачи исследования**

Целью данной работы было одновременное изучение метаболических путей и систем регуляции у бактерий кишечного микробиома. Для более глубокого покрытия многих аспектов регуляторных и метаболических систем мы совместили три ортогональных подхода: (1) изучение регуляции инициации транскрипции белками из одного семейства транскрипционных факторов (LacI) у многих бактерий, (2) изучение метаболизма углеводов и его регуляции многими факторами у бактерий их одного рода бактерий (*Bifidobacterium*) и (3) изучение одной группы метаболических путей (биосинтез витаминов группы В) у многих бактерий из микробиома кишечника.

В работе решаются следующие задачи:

1. Геномная реконструкция, структурный и функциональный анализ регулонов транскрипционных факторов из семейства LacI в геномах бактерий из различных таксономических групп.
2. Геномная реконструкция и анализ регуляции сахарного метаболизма в геномах бифидобактерий.
3. Геномная реконструкция и сравнительный анализ метаболических путей биосинтеза витаминов группы В у бактерий из человеческого микробиома.

### **Научная новизна и практическая значимость работы**

В работе впервые детально изучена регуляция метаболических путей транскрипционными факторами из семейства LacI у >270 бактерий: проведена реконструкция регулонов метаболических путей для >1300 транскрипционных факторов. Для 78% и 67% изученных транскрипционных факторов, соответственно, предсказаны функциональные роли и молекулы эффекторов. Впервые предсказано три новых глобальных регулятора, контролирующего центральный углеводный метаболизм в трех группах *Alphaproteobacteria*: GluR, GapR и PckR у *Rhizobiales*, *Rhodobacterales* и *Caulobacterales*. Для 10 представителей рода *Bifidobacterium* реконструированы регулоны катаболизма углеводов для >260 транскрипционных факторов из семейств LacI, ROK, DeoR, AgaC,

GntR и TetR. Впервые предсказан глобальный регулятор AraQ из семейства LacI, контролирующей центральный углеводный метаболизм у бифидобактерий и являющийся важной мишенью для направленной модуляции метаболизма бифидобактерий с помощью пребиотиков. Обнаружен первый пример регуляции метаболизма углеводов транскрипционными факторами из семейства TetR. Предсказана сложная сеть регуляции путей катаболизма мальтозо-содержащих полисахаридов, контролируемая множественными регуляторами у бифидобактерий.

Впервые реконструированы пути биосинтеза, сохранения и захвата из окружающей среды восьми витаминов группы В для репрезентативной выборки бактерий из кишечного микробиома человека. Предсказано несколько новых вариантов путей захвата и сохранения предшественников витаминов группы В. Сформулирована гипотеза об обмене несколькими метаболическими предшественниками витаминов группы В (тиазол, квинолилат, детиобиотин и пантоат) между различными организмами как механизм кооперации. Разработан метод для подсчета численного баланса ауксотрофных и прототрофных бактерий в образцах микробиома кишечника. Метод был применен для оценки разнообразия фенотипических профилей кишечного микробиома у людей и для оценки влияния диеты на баланс ауксотрофов и прототрофов в микробиоме кишечника человека.

#### **Выносимые на защиту положения**

- Проведена реконструкция регулонов для 268 транскрипционных факторов из семейств LacI, ROK, DeoR, AraC, GntR и TetR, контролирующей катаболизм углеводов у бифидобактерий
- AraQ является глобальным регулятором центрального углеводного метаболизма у бифидобактерий
- Транскрипционные факторы из семейства TetR регулируют катаболизм углеводов у бифидобактерий
- Проведена реконструкция регулонов для 1303 транскрипционных факторов из семейства LacI у 272 бактерий
- Предпочтительное расстояние между соседними связывающими сайтами для транскрипционных факторов из семейства LacI кратно числу нуклеотидов, соответствующему целому числу витков спирали ДНК
- Транскрипционные факторы PckR, GapR, GluR являются глобальными регуляторами центрального углеводного метаболизма в бактериях из семейств *Rhizobiales*, *Rhodobacteraceae* и *Caulobacteraceae*
- Проведена реконструкция метаболических путей биосинтеза, захвата и сохранения восьми витаминов группы В в геномах 2228 представителей кишечного микробиома
- Среди бактерий кишечного микробиома встречаются “неполные” пути биосинтеза витаминов, начинающиеся с промежуточных предшественников
- Мы предполагаем, что бактерии кишечного микробиома могут обмениваться такими веществами, как тиазол, квинолилат, детиобиотин и пантоат
- Проведена оценка того, насколько баланс между прототрофными и ауксотрофными бактериями в кишечном микробиоме варьирует у различных индивидуумов

- Мы не наблюдаем значительного изменения баланса между прототрофными и ауксотрофными бактериями в кишечном микробиоме при изменении рациона питания у организма хозяина

### **Степень достоверности и апробация результатов**

По материалам диссертации опубликовано 3 статьи в рецензируемых научных журналах. Результаты работы были представлены на международных конференциях RECOMB/ISCB'14, LAMG'14, MCCMB'15, ASM'15, и российских конференциях ИТИС'13, ИТИС'15.

### **Структура и объем диссертации**

Диссертационная работа изложена на 149 страницах и состоит из введения, четырех глав, выводов и списка цитированной литературы. Глава 1 содержит обзор литературы по теме диссертации. Главы с 2 по 4 содержат описание собственных исследований. Список литературы включает 230 наименований. Работа содержит 22 рисунков, 3 таблицы и 2 приложения.

## **СОДЕРЖАНИЕ РАБОТЫ**

### **Введение**

Во введении обоснована актуальность диссертационной работы, сформулирована цель и аргументирована научная новизна исследований, показана практическая значимость полученных результатов, представлены выносимые на защиту научные положения.

### **Глава 1. Обзор литературы**

В разделе содержится мотивировка поставленных задач, а также аналитический обзор современной литературы по проблемам, рассмотренным в диссертации.

### **Глава 2. Реконструкция метаболических путей и регулонов катаболизма углеводов в геномах *Bifidobacteria***

Бифидобактерии относятся к группе бактерий, известных как пробиотики, т.е. микроорганизмы - обитатели микробиома кишечника, которые благоприятно воздействуют на здоровье организма - хозяина. Для культивации бифидобактерий широко применяются пребиотики, т.е. вещества, не усваиваемые в верхних отделах желудочно-кишечного тракта, но усваиваются определенными группами бактерий, населяющих толстый кишечник, в том числе бифидобактериями. Наиболее распространенный класс пребиотиков - это полисахариды. Однако, несмотря на столь распространенное использование полисахаридов для стимуляции роста бифидобактерий, метаболизм углеводов и его регуляция у бифидобактерий крайне мало изучены. Поэтому в качестве одной из целей данного исследования мы выбрали изучение регуляции метаболизма углеводов у бифидобактерий. Подробное изучение широкой группы метаболических путей у одной группы бактерий позволяет наблюдать различия в строении метаболической системы у близкородственных геномов, обусловленных адаптацией к специфическим экологическим нишам.

Чтобы достичь максимального покрытия изучаемой регуляторной системы, мы идентифицировали набор семейств транскрипционных факторов (ТФ), про которые известно, что транскрипционные факторы, принадлежащие к этим семействам, часто вовлечены в регуляцию сахарных метаболических путей. Мы идентифицировали все гены, кодирующие транскрипционные факторы из этих семейств в 10 изучаемых геномах. Для семейств GntR и AraC, которые покрывают широкий репертуар метаболических путей, мы выбрали лишь те транскрипционные факторы, гены которых находились рядом с другими генами, предположительно участвующими в сахарном метаболизме. В результате, мы идентифицировали 308 транскрипционных факторов, потенциально регулирующих сахарный метаболизм. Число таких транскрипционных факторов на геном варьирует в пределах от 12 у *B. gallicum* до 54 у *B. dentium*. Большинство из них относится к двум большим семействам - LacI и ROK; транскрипционные факторы из других семейств реже встречаются в геномах бифидобактерий. Мы разделили найденные транскрипционные факторы на ортологичные группы и провели реконструкцию их регулонов, используя методы сравнительной геномики (для подробного описания, см. секцию Методы в тексте диссертации). Мы смогли найти мотивы сайтов связывания ТФ и реконструировать регулоны для 48 из 55 ортологичных групп и для 16 из 33 одиночных регуляторов. Большинство изученных транскрипционных факторов мозаично распределены среди **Таблица 1. Распределение ТФ, контролирующих метаболизм углеводов у бифидобактерий**

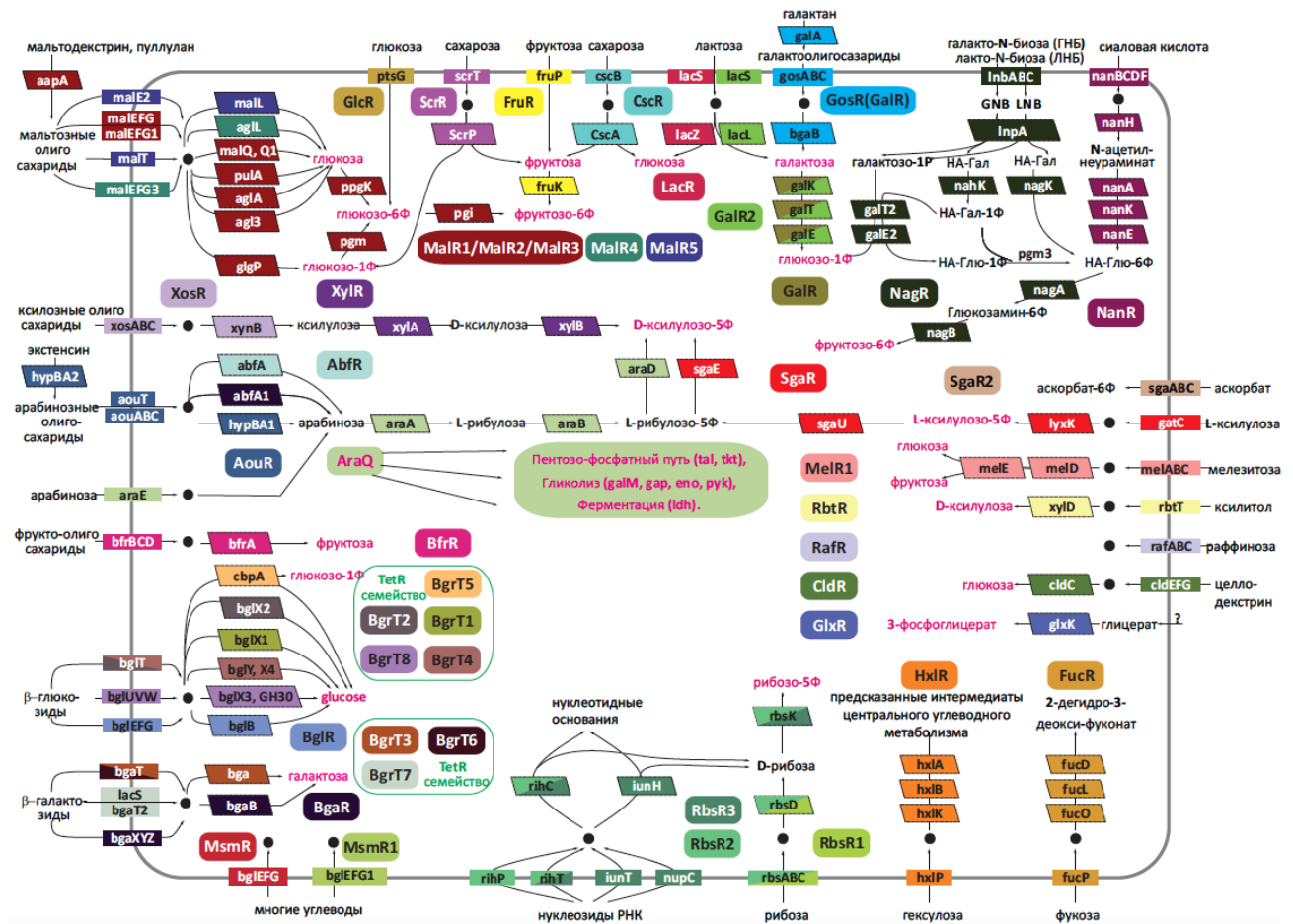
Геном	Семейства транскрипционных факторов							ТФ	Всего	ГУМ <sup>2</sup> , %
	LacI	ROK	RpiR	DeoR/ SorC	BglG	TetR <sup>1</sup>	AraC <sup>1</sup>			
<i>B. adolescentis</i>	18	4	0	1	1	6(10)	2(4)	1(10)	33	11.6
<i>B. angulatum</i>	19	4	0	1	1	2(9)	1(2)	0(8)	28	11.2
<i>B. animalis</i>	11	4	0	1	0	2(15)	0(1)	0(18)	18	10
<i>B. bifidum</i>	11	4	1	2	1	0(6)	0(2)	2(11)	21	9.3
<i>B. breve</i> DSM 20213	23	7	1	1	1	2(9)	2(3)	1(12)	38	13
<i>B. breve</i> UCC2003	30	6	1	1	1	2(7)	0(1)	1(12)	42	14.1
<i>B. dentium</i>	34	6	0	1	1	7(16)	2(6)	3(13)	54	15.5
<i>B. gallicum</i>	7	3	0	1	0	0(3)	0(0)	1(9)	12	9.2
<i>B. longum</i> NCC2705	21	6	1	1	1	1(6)	0(0)	0(9)	31	11.6
<i>B. longum</i> ATCC 15697	20	5	1	1	1	1(8)	1(2)	1(12)	31	11.8

<sup>1</sup> Общее число транскрипционных факторов из семейств TetR, AraC и GntR на геном указано в скобках

<sup>2</sup> Последний столбец показывает долю генов, участвующих в метаболизме углеводов (ГУМ), согласно базе данных IMG

изучаемых геномов; нам удалось обнаружить лишь 6 транскрипционных факторов, присутствующих во всех 10 геномах одновременно. Кроме того, мы провели метаболическую реконструкцию соответствующих биохимических путей и предсказали функции ко-регулируемых генов. Большинство реконструированных регулонов контролируют пути утилизации моно-, ди- и олигосахаридов. Ожидаемо, большинство из этих путей осуществляются при помощи лишь небольшой группы ферментативных активностей, в т.ч. оксидоредуктазы, киназы, альдолазы, гидролазы и изомеразы. Кроме того, большинство регулонов содержат в своем составе углеводные транспортеры, главным образом из семейства АТФ-связывающих кассетных транспортеров (ABC) и из суперсемейства мембранных транспортёров (MFS).

Реконструированная регуляторная сеть покрывает основные метаболические пути утилизации углеводов. Несмотря на то, что для многих углеводов регуляторная сеть устроена просто, т.е по правилу один регулятор - один метаболический путь, в некоторых случаях, разумеется, наблюдаются исключения, например такие, как (1) контролирование одного и того же пути различными регуляторами, (2) контролирование одним и тем же регулятором различных путей в различных организмах, (3) контролирование множественных альтернативных путей катаболизма одного и того же регулятора различными транскрипционными факторами и т.д. В следующих параграфах мы разберем некоторые примеры такой сложной регуляции, наблюдаемые по итогам проведенной реконструкции.

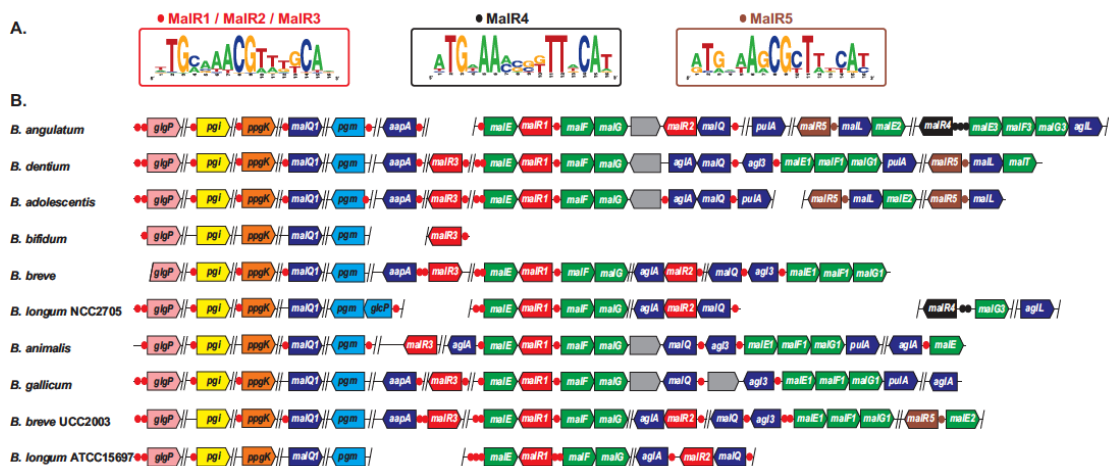


**Рис. 1. Пути утилизации углеводов, участвующие в реконструированной сети транскрипционной регуляции в геномах бифидобактерий.** Регуляторы обозначены овалами, транспортеры – прямоугольниками, а ферменты – скошенными прямоугольниками. Гены, регулируемые одним и тем же ТФ, отмечены одним и тем же цветом. Гены, регулируемые двумя разными регуляторами (как правило, в различных геномах), отмечены обоими цветами в диагональных сегментах. Общие интермедиаты центрального углеводного метаболизма показаны красным цветом. Стоит упомянуть, что ни один из изучаемых геномов не содержит все показанные здесь пути одновременно. Вещества, транспортируемые внутрь клетки, показаны черными кругами

Ген транскрипционного фактора RbsR из семейства LacI, специфичного к D-рибозе, расположен рядом с *rbs*-опероном у *Bifidobacterium breve* UCC2003, и контролирует экспрессию *rbs*-оперона. Мы идентифицировали 11 гомологов гена *rbsR*, которые можно отнести к 3 ортологичным группам регуляторов (RbsR1, RbsR2 и RbsR3) и реконструировали соответствующие регулоны в геномах 10 изученных бифидобактерий.



Две из этих трех ортологичных групп регуляторов (RbsR2 и RbsR3) характеризуются похожими ДНК-связывающим мотивами с общим консенсусом TGATAAAACGTTTTATCA. Однако, тогда как регулон RbsR2 ожидаемо содержит гены метаболизма рибозы, регулон RbsR3 содержит гены, которые можно классифицировать как гены метаболизма рибонуклеозидов: рибонуклеозид-гидролазу *rihC*, рибонуклеозидные транспортеры *rihT* и *nupC* и рибокиназу *rbsK*. Кроме того, в геноме *B. dentium* находится

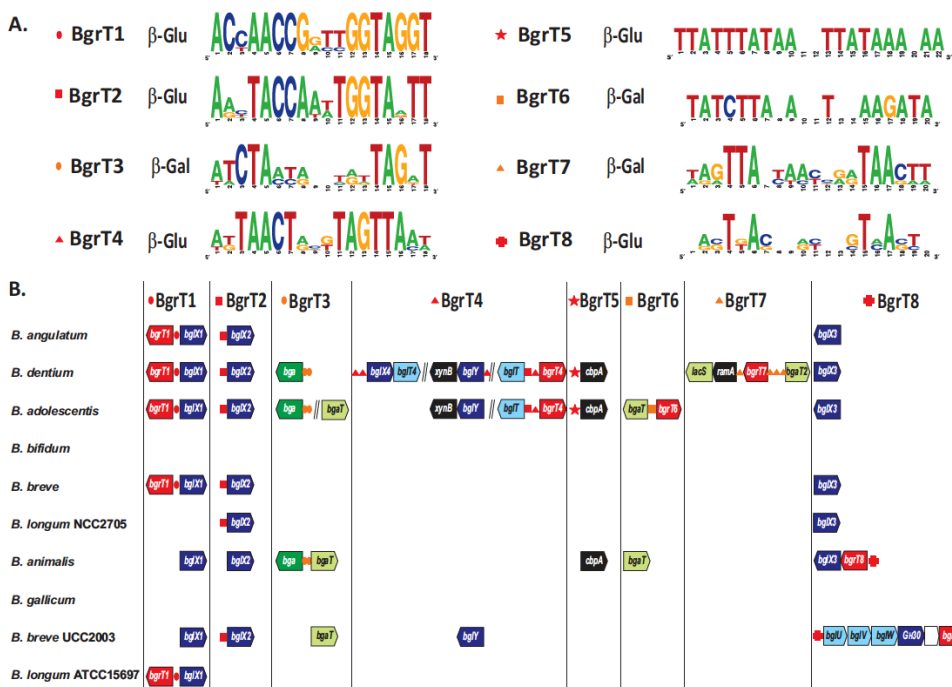


**Рисунок 2. Геномная организация регулонов утилизации мальтозы и мальтодекстрина у бифидобактерий.** (А) Лого консенсусных последовательностей предсказанных ДНК-связывающих сайтов для 5 ортологичных групп MalR регуляторов. Мотивы групп MalR1, MalR2 и MalR3 одинаковые. (Б) Геномный контекст MalR регулонов в геномах 10 бифидобактерий. Предсказанные сайты связывания показаны кружками, а гены показаны стрелками. Гены, кодирующие регуляторы MalR1/MalR2/MalR3, MalR4 и MalR5, а также соответствующие им сайты, отмечены красным, черным и коричневым цветами, соответственно. Гены ABC транспортеров мальтозы/мальтосахаридов malEFG/malEFG1/malE2/malEFG3 показаны зеленым. Гликозил гидролазы, участвующие в утилизации мальтодекстрина, показаны темно-синим. Гликоген фосфорилаза *glgP*, глюкозо-6-фосфат изомеразы *rgi*, глюкокиназа *prgK* и фосфоглюкомутаза *rgm* показаны розовым, желтым, оранжевым и светло-синим, соответственно. Гены с неизвестными функциями показаны серым. Вертикальные линии разделяют опероны, которые удалены друг от друга на хромосоме

паралог RbsR3, который, предположительно контролирует экспрессию нуклеозид-гидролазы *iunH*, специфичной к инозину и/или уридину, и предсказанного инозинового/уридинового транспортера *iunT*. Таким образом, мы наблюдаем три гомологичные группы регулонов, которые, по-видимому, разошлись в специфичности в различных геномах бифидобактерий: RbsR1 и RbsR2 контролируют метаболизм рибозы, а RbsR3 - метаболизм рибонуклеозидов. (Рис. 1)

Анализируя гены, гомологичные ранее изученным регуляторам утилизации мальтозы, мы обнаружили 5 паралогичных групп регуляторов, в состав которых входит 28 генов транскрипционных факторов. Реконструкция их регулонов показала, что три из этих пяти групп обладают похожими связывающими мотивами, и потому их регуляторные активности могут перекрываться. Мы не можем отличить их регулоны друг от друга биоинформатическими методами и потому называем их общий регулон MalR1/MalR2/MalR3. Более того, гены, кодирующие транскрипционные факторы из этих пяти групп, содержат в своих регуляторных областях сайты, похожие на мотив MalR1/MalR2/MalR3 регулона, что позволяет предположить, что они формируют

сложную регуляторную сеть, где многие регуляторы контролируют экспрессию друг друга. Найденные регулоны катаболизма мальтозы/мальтодекстрина контролируют широкий спектр генов, в т.ч. две паралогичные системы ABC-транспортеров (*malEFG*, *malEFG1*), экстрацеллюлярную  $\alpha$ -амилазу (*aapA*, ранее называемую *apuB*), различные цитоплазматические гликозил-гидролазы, такие как  $\alpha$ -глюкозидазы (*aglA*, *agl3*), амиолмальтазы (*malQ*, *malQ1*), пуллуназы (*pulA*), и другие цитоплазматические ферменты, в том числе гликоген-фосфорилазу (*glgP*), глюкокиназу (*ppgK*), фосфоглюкомугазу (*pgm*) и глюкозо-6-фосфат изомеразу (*pgi*). Три последних фермента участвуют в центральном



**Рис. 3. Геномная организация регулонов семейства TetR, контролирующих утилизацию бета-глюкозидов и бета-галактозидов в бифидобактериях.** (А) Лого консенсусных последовательностей предсказанных ДНК-связывающих сайтов для 8 ортологичных групп BgrT регуляторов. Показана специфичность регуляторов для бета-глюкозидов либо бета-галактозидов. (Б) Геномный контекст BgrT регулонов в геномах 10 бифидобактерий. Предсказанные сайты связывания показаны специальными красными либо оранжевыми символами перед соответствующими регулируемыми (показаны стрелками). Гены, кодирующие бета-глюкозидные и бета-галактозидные транспортеры, отмечены светло-синим и светло-зеленым цветами, соответственно. Бета-глюкозидазы и бета-галактозидазы отмечены темно-синим и темно-зеленым. Другие гликозил гидролазы отмечены черным. Гены регуляторов BgrT показаны красным. Вертикальные линии разделяют опероны, которые удалены друг от друга на хромосоме.

метаболизме глюкозы и глюкозо-1-фосфата, которые конвертируются в фруктозо-6-фосфат и таким образом поступают в специфичный для бифидобактерий гликолитический путь («бифидный шунт») Мы предполагаем, что ферменты и транспортеры, контролируемые регулоном MalR1/MalR2/MalR3, могут составлять полный путь катаболизма мальтозы/мальтодекстрина/пуллулана. (Рис. 2) Оставшиеся две группы ортологичных мальтозных регуляторов, (мы называем их MalR4 и MalR5), состоят из локальных регуляторов, которые контролируют паралогичные компоненты ABC транспортеров мальтозы/мальтодекстрина (*malEFG3*, *malE2*), еще один потенциальный мальтозный транспортер (*malT*), и две цитоплазматические глюкозидазы мальтодекстриновых олигосахаридов (*aglL*, *malL*; Рис. 2).

Бактериальные регуляторы семейства TetR контролируют экспрессию генов, участвующих в различных биологических процессах, в т.ч. устойчивость к антибиотикам,

межклеточные сигнальные системы и метаболизм различных веществ, однако не было найдено примеров, где они бы регулировали углеводный метаболизм. Мы обнаружили новую группу регуляторов из TetR семейства, гены которых соседствуют в хромосомных кластерах с генами катаболизма углеводов (Рис. 3). Филогенетический анализ этой группы регуляторов показал, что все регуляторы, гены которых ко-локализуются с генами катаболизма сахаров, формируют единую ветвь на дереве, которую мы называем BgrT (бета-глюкозидные регуляторы из семейства TetR). Большинство регуляторов BgrT было найдено в геномах рода *Bifidobacterium*, а также в нескольких группах Actinobacteria и в двух группах

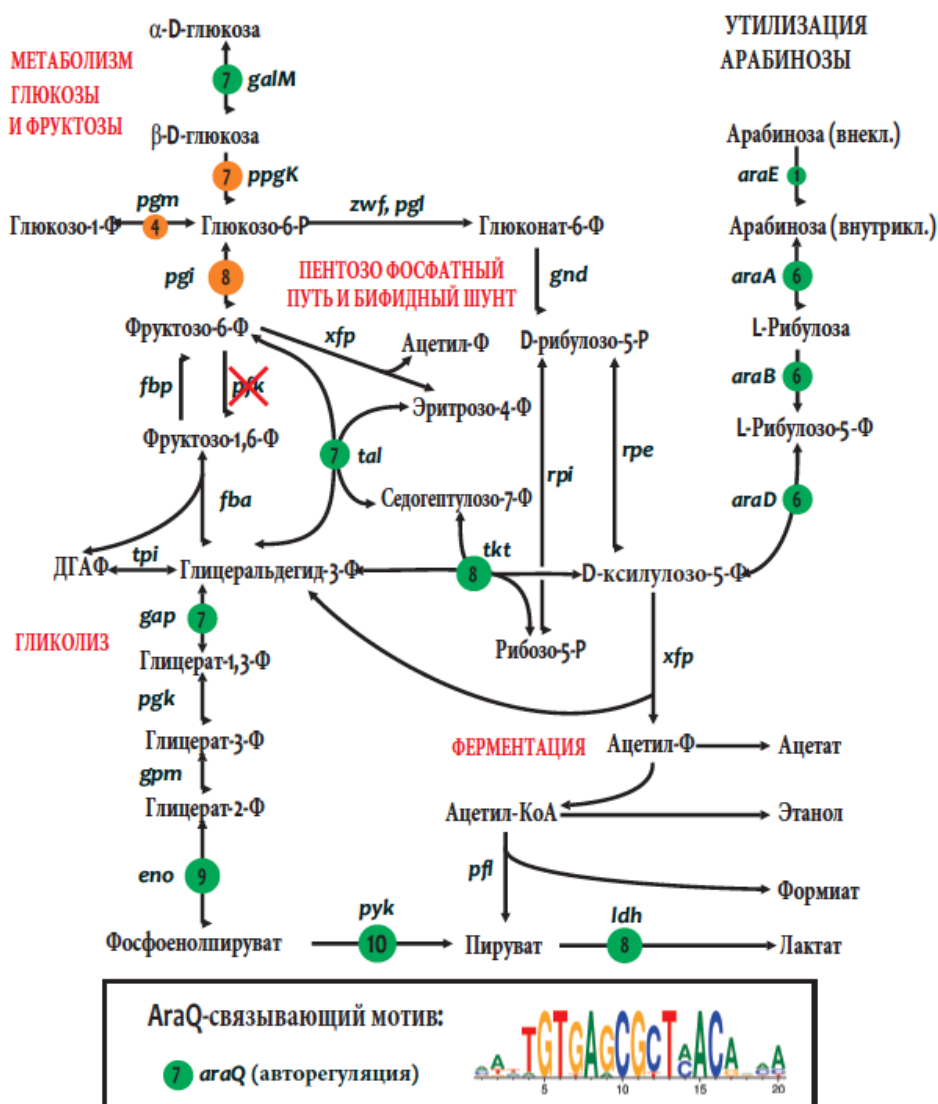


Рис. 4. Метаболический контекст и консервативное ядро реконструированных регулонов AraQ в бифидобактериях. Гены, контролируемые регулятором AraQ, показаны зелеными кружками. Числа внутри кружков обозначают число геномов, в которых регуляторная область данного гена содержит предсказанный сайт связывания AraQ. Лого консенсусных последовательностей предсказанных ДНК-связывающих сайтов AraQ показано во вставке внизу. Кроме того, гены, контролируемые регуляторами MalR, отмечены оранжевыми кружками.

протеобактерий (*Rhizobiales*, *Pseudomonadales*). Мы классифицировали 23 найденных белка BgrT из 10 бифидобактериальных геномов по 8 ортологичным группам; для всех этих групп мы смогли предсказать уникальные мотивы для сайтов связывания и

реконструировать их регулоны. 5 из 8 реконструированных регулонов содержат гены утилизации  $\beta$ -глюкозидов, тогда как оставшиеся 3 регулона контролируют гены утилизации  $\beta$ -галактозидов. Таким образом, мы впервые показали возможность регуляции углеводного метаболизма транскрипционными факторами из семейства TetR.

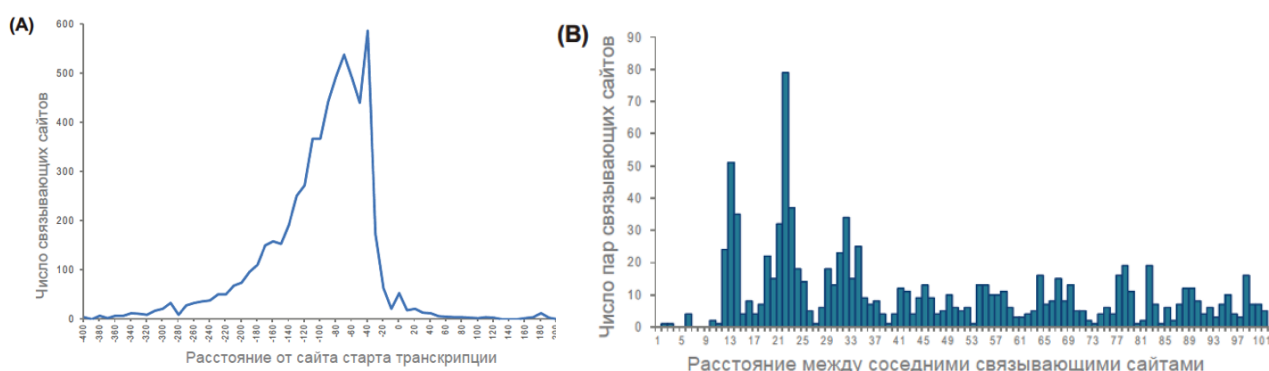
Реконструкция регулона, контролирующего метаболизм арабинозы, показала, что транскрипционный фактор *araQ* может быть глобальным регулятором, контролирующим центральный углеводный метаболизм бифидобактерий. Ортологи *araQ* присутствуют во всех 10 изучаемых геномах. Если в 6 из 10 геномах *araQ* локализован с опероном утилизации арабинозы (*araBDA*), то в оставшихся 4 геномах есть ортологи *araQ*, но нет оперона утилизации арабинозы. Мы идентифицировали возможный мотив связывания AraQ и реконструировали соответствующий глобальный регулон. Наиболее консервативными членами регулона AraQ (после оперона *ara* и гена *araQ*) являются гены центральных гликолитических путей, например, гены гликолиза (*gap*, *eno*, *pyk*), гены пентозо-фосфатного пути и бифидного шунта (*tal*, *tkl*), а также лактат-дегидрогеназа (*ldh*) и альдозо-1-эпимераза (*galM*). Любопытно, что в штаммах *B. breve* и *B. longum* мы обнаружили несколько генов из пути утилизации мальтозы и мальтозо-содержащих олигосахаридов, чьи регуляторные области содержат связывающие сайты и AraQ, и вышеупомянутых регуляторов MalR (например, мальтозный транспортер *malE* и амиломальтазу *malQ1*).

В недавнем исследовании ген *agaQ* был найден в числе ключевых генов, присутствующих у всех 45 изучаемых геномах бифидобактерий. Арабинозный оперон *agaBDA*, напротив, отсутствует у 7 штаммов бифидобактерий, а именно *B. bifidum*, *B. boum*, *B. breve*, *B. choerinum*, *B. longum* subsp. *infantis*, *B. ruminantium* и *B. thermophilum*. Мы предполагаем, что глобальная регуляция множества генов центрального углеводного метаболизма является общим свойством реконструированных регулонов AraQ у всех бифидобактерий (Рис. 4). Мы идентифицировали дальних гомологов *agaQ* в нескольких группах Actinomycetales. Большинство этих гомологов ко-локализуется в геноме с оперонами утилизации арабинозы, однако мы не смогли найти дополнительных членов этих регулонов. Мы предполагаем, что глобальные регулоны, контролирующие центральный углеводный метаболизм, развились в бифидобактериях из локального регулона катаболизма арабинозы, который присутствовал в геноме общего предка Actinomycetales. После публикации нашей работы многие наши предсказания относительно функций регулятора AraQ были экспериментально подтверждены в лаборатории Дауве ван Синдерена (статья готовится к печати).

### **Глава 3. Исследование транскрипционной регуляции и эволюции регулонов из семейства LacI**

В вышеописанном подходе центральной точкой исследования была метаболическая система утилизации углеводов и ее регуляция. Однако, не меньший интерес представляют исследования, где центральной точкой служит группа родственных регуляторов. Такой подход позволяет охватить более широкую группу геномов, и, как следствие, наблюдаемые различия в регуляторных сетях не ограничены генетическими факторами, специфичными для одного лишь таксона. В качестве ортогонального подхода, который позволил бы нам детально изучить регуляторные взаимодействия, связанные с одной группой транскрипционных факторов, мы выбрали детальное изучение регулонов для транскрипционных факторов из семейства LacI.

Для проведения данного анализа мы составили выборку из >300 геномов бактерий с целью покрыть как можно более широкий круг бактериальных таксонов. Используя поиск последовательностей, соответствующих консервативным белковым доменам, мы идентифицировали гены, кодирующие возможные транскрипционные факторы из семейства LacI. Найденные ТФ были неравно распределены среди изучаемых геномов; наибольшее среднее число регуляторов из семейства LacI на геном достигается в нескольких группах *Actinobacteria*, включая *Streptomycetaceae* и *Bifidobacteriaceae* (от 17 до 32 ТФ), в двух группах *Proteobacteria* – *Rhizobiales* и *Enterobacteriales* (15 ТФ на геном в обеих группах), и в двух группах *Firmicutes* – *Bacillales* и *Enterococcaceae* (12 ТФ на геном в обеих группах). 10% изучаемых геномов не содержат ни одного потенциального гена, кодирующего регулятор LacI. Проводя филогенетический анализ и оценивая схожесть геномных последовательностей, мы разбили все множество найденных ТФ из семейства LacI на ортологичные группы, специфичные для отдельных таксономических групп. В результате сравнительно-геномного анализа мы смогли предсказать возможные сайты связывания и их предполагаемые мотивы для 1303 (50% от всех найденных ТФ из семейства LacI) в 272 бактериальных геномах (80% изученных геномов). Базовой единицей нашего анализа является аннотированный «регулог», определяемый как набор регулонов, специфичных для отдельных геномов, и контролируемых ортологичными ТФ. Всего, мы нашли 1281 LacI ТФ-регулонов, которые составляют 332 регулога, неравно распределенных среди 39 изучаемых таксономических групп (см. в базе данных



**Рис. 5 Распределение расстояний.** (А) Расстояния между связывающим сайтом ТФ из семейства LacI и сайтом начала транскрипции. (Б) Расстояния между соседними связывающими сайтами

RegPrecise). Реконструированные регулоны содержат 7465 предсказанных сайтов связывания, 6076 оперонов и 13558 генов.

Мы разделили все реконструированные регулоны на две категории в зависимости от их размера (число регулируемых генов и оперонов) и функционального разнообразия (число регулируемых метаболических путей). Всего 125 регулонов (12 регулогов) были классифицированы как глобальные регулоны, т.к. каждый из них (1) содержал более 15 регулируемых генов, находящихся в составе хотя бы 7 оперонов, и (2) контролировал несколько разных метаболических путей. Оставшиеся регулоны (1163/1288, 90%) были классифицированы как локальные: каждый из них регулировал лишь несколько генов, как правило - входящих в состав лишь одного метаболического пути (зачастую – пути утилизации какого-либо сахара). Почти половина всех найденных глобальных регулонов (59 регулонов, 6 регулогов) – это регулоны ортологов глобального регулятора сахарного метаболизма CsrA из генома *B. subtilis*. Ортологи CsrA и соответствующие глобальные регулоны были найдены во всех изучаемых таксономических группах, относящихся

к типу *Firmicutes*. Другая большая группа глобальных регулонов (31 регулон, 3 регулога) содержит регулоны ортологов пуринового репрессора PurR из *E. coli* в трех близких таксономических группах  $\gamma$ -Proteobacteria: *Enterobacteriales*, *Pasteurellales*, и *Vibrionales*. PurR - это глобальный транскрипционный регулятор *E. coli*, контролирующий биосинтез пуринов, некоторые реакции в пути синтеза пиримидинов, метаболизм полиаминов и ассимиляцию азота. Наконец, еще 3 глобальных регулона, контролирующих центральные и периферические пути катаболизма углеводов в нескольких группах Proteobacteria - FruR в *Enterobacteriales*, PckR в *Rhizobiales*, и GapR в *Rhodobacterales* (всего 35 регулонов) – детально описаны ниже.

Для 72% изученных транскрипционных факторов (всего 943 ТФ) была присуща авторегуляция. Ожидается, авторегуляция чаще наблюдалась у локальных, чем у глобальных регулонов. Так, среди глобальных ТФ авторегуляция была найдена менее чем у половины ТФ (всего 55 ТФ), тогда как среди локальных ТФ авторегуляция была найдена более чем у трех четвертей (всего 888 ТФ).

Мотивы связывающих сайтов изучаемых ТФ из семейства LacI - это палиндромы четной длины, большинство из них содержит консервативный динуклеотид CG посередине. Более 75% найденных сайтов изучаемых ТФ расположены на участке между 140 и 30 основаниями перед старт кодоном (Рис. 5А). Менее 1% связывающих сайтов расположены внутри открытой рамки считывания, включая экспериментально подтвержденные сайты LacI в гене *lacZ*, и PurR в гене *purB* в *E. coli*. Примерно 7% связывающих сайтов расположены далеко (более чем 200 нуклеотидов) от старт-кодона.

Около 20% регулируемых оперонов регулируются двумя и более связывающими сайтами. Например, двойные сайты были найдены в 1118 оперонах. На гистограмме попарных расстояний между соседними сайтами, расположенными в одном и том же межгенном регионе, можно выделить пики на 13, 22 и 32 нуклеотидах (Рис. 5Б). Промежутки в 22 и 32 нуклеотида соответствуют 2 и 3 виткам спирали ДНК, что может указывать на кооперативное связывание 2 димеров ТФ. Некоторые регулоны ТФ из семейства LacI, например, RafR в *Enterobacteriales* и ScrR в *Burkholderiales* регулируются исключительно двойными сайтами, удаленными друг от друга на 21-22 нуклеотида. Можно предположить, что кооперативное связывание в этом случае является обязательным. Перекрывающиеся сайты, чьи центры удалены друг от друга на ~13 нуклеотидов, были найдены у 110 оперонов. Такое расположение может быть функциональным, как было показано для связывающих сайтов GntR перед геном *gntKU* в геноме *E. coli*.

Анализируя функциональное содержание реконструированных регулонов, мы предсказали возможные биологические функции и эффекторы для 190 ортологичных групп ТФ из семейства LacI. Для 182 из 190 групп мы предсказали регулируемые метаболические пути; среди них, 54 группы были отнесены к общей категории «сахарный метаболизм»; для этих групп мы не смогли установить, какие именно углеводные катаболические пути они регулируют. Мы сравнили предсказанные функции регулонов с результатами ранее опубликованных экспериментальных работ, доступными для 24 ТФ из семейства LacI; экспериментально установленные функции соответствуют функциям, предсказанным по результатам нашего анализа. Далее, основываясь на реконструкции метаболических путей и на известных метаболитах – интермедиатах метаболических путей, мы предложили возможные молекулярные эффекторы для 108 групп ТФ из семейства LacI. Для 21 из 108 групп, эффекторы известны из более ранних работ. Подавляющее большинство изученных ортологичных групп ТФ из семейства LacI



контролирует метаболизм углеводов (176 из 182, 96% групп). Кроме того, мы наблюдали, что углеводы представляют наибольший класс эффекторов для регуляторов этого семейства (103 из 107, 96% групп). Большинство предсказанных углеводных эффекторов – это моносахариды и их производные, в т.ч. гексозы (например, глюкозу, галактозу и маннозу), пентозы (например, рибозу и ксилозу), фосфаты сахаров (например, фруктозо-1-фосфат, аллозо-6-фосфат), сахарные кислоты (например, глюкуроонат и галактуронат), сахарные спирты (например, рибитол), и аминоксахара (например, ацетилгликозамин). Вторая по размеру группа углеводных эффекторов включает различные олигосахариды (14 из 26), например, целлобиозу, мальтозу, сахарозу, трегалозу, или их фосфорилированные производные (в т.ч. целлобиозо-6-фосфат или сахарозо-6-фосфат). Наконец, к числу не-углеводных эффекторы ТФ из семейства LacI относятся нуклеотидные основания (например, гуанин и гипоксантин являются ко-репрессорами PurR в *E. coli*) и нуклеозиды (например, цитидин и аденозин индуцируют CytR в *E. coli*).

В данной работе мы наблюдали несколько ортологичных групп глобальных регуляторов, в т.ч. CsrA в фирмикутах и FruR в протеобактериях. Более того, в данной работе мы идентифицировали три новых не-ортологичных регулятора (названных PckR, GarR и GluR), реконструкция регулонов которых показала, что они предположительно контролируют центральный углеводный метаболизм в трех таксономических группах альфа-протеобактерий: *Rhizobiales*, *Rhodobacterales*, и *Caulobacterales*. Мы провели функциональный анализ регулонов всех вышеуказанных ТФ у протеобактерий (кроме CsrA, т.к. подробная реконструкция регулонов CsrA была опубликована ранее).

Фруктозный репрессор FruR (также известный как репрессор и активатор катаболизма Cra) известен как глобальный регулятор центрального метаболизма в *E. coli*. FruR/Cra контролирует поток углерода, репрессируя гликолитические гены ферментов пентозо-фосфатного пути, пути Эмбдена-Мейерхофа и пути Энтнера-Дудорова, и активируя гены глюконеогенеза. Сравнительно-геномная реконструкция ортологичных регулонов FruR в гамма- протеобактериях показала, что размеры регулона напрямую связаны с таксономией изучаемых групп и с филогенией белков FruR. У *Vibrionales* и *Pseudomonadales*, регулоны FruR состоят лишь из оперона утилизации фруктозы *fruBKA* и из гена регулятора *fruR*; таким образом, FruR выступает в роли локального регулятора в этих таксономических группах. У *Enterobacteriales*, регулон FruR расширился и стал покрывать гены центральных гликолитических путей, части цикла трикарбоновых кислот и нескольких путей ферментации и дыхания. Дальнейшее распространение регулона (на гены глиоксилатного шунта) произошло в группах *Escherichia*, *Salmonella*, *Citrobacter*, *Enterobacter* и *Klebsiella*. Наконец, в близкородственных видах *Escherichia* и *Salmonella*, регулон FruR/Cra распространился на гены пути Энтнера-Дудорова. С другой стороны, регулон деградировал в *Pasteurellales*: ген регулятора отсутствует в геномах *Haemophilus* spp. и *Actinobacillus pleuropneumoniae*, тогда как в геномах *Pasteurella multocida*, *Mannheimia succiniciproducens*, *A. succinogenes* и *A. aphrophilus*, несмотря на присутствие гена регулятора, мы не смогли идентифицировать связывающие сайты для FruR.

Предполагаемый ТФ PckR (SMc02975 в *Sinorhizobium meliloti*) ранее был аннотирован как возможный регулятор карбоксиказы фосфоенолпирувата; однако, этот ген не был изучен экспериментально. По итогам проведенной геномной реконструкции мы предполагаем, что PckR работает как многофункциональный транскрипционный регулятор, одновременно репрессируя гликолитические гены из пути Эмбдена-Мейерхофа и Энтнера-Дудорова (*glk*, *fba*, *pykA*, *zwf-pgl-edd*, *eda*) и активируя гены глюконеогенеза и

цикла трикарбоновых кислот (*pckA*, *mdh-sucCDAB*, *sdhABCD*) (Рис. 6). В семействе *Xanthobacteraceae*, реконструированный регулон PckR содержит лишь несколько генов (*pckA*, *pckR*, *edd*, и *hpr*), тогда как в семействе *Rhizobiales* регулон значительно больше и содержит от 14 до 28 генов на геном. К наиболее консервативным генам реконструированного регулона относятся гены *pckA* и *edd* (в 10 и в 9 геномах, соответственно), опероны *zwf-pgl* и *mdh-sucCDAB* (в 8 геномах) и гены *fba*, *glk* и *eda* (в 7 геномах). Мы обнаружили ортологи PckR и их возможные сайты связывания в бактериях семейства *Rhizobiales*, но не обнаружили их в бактериях из других семейств, что позволяет предположить, что регулон PckR появился относительно недавно в процессе эволюции альфа-протеобактерий. После публикации

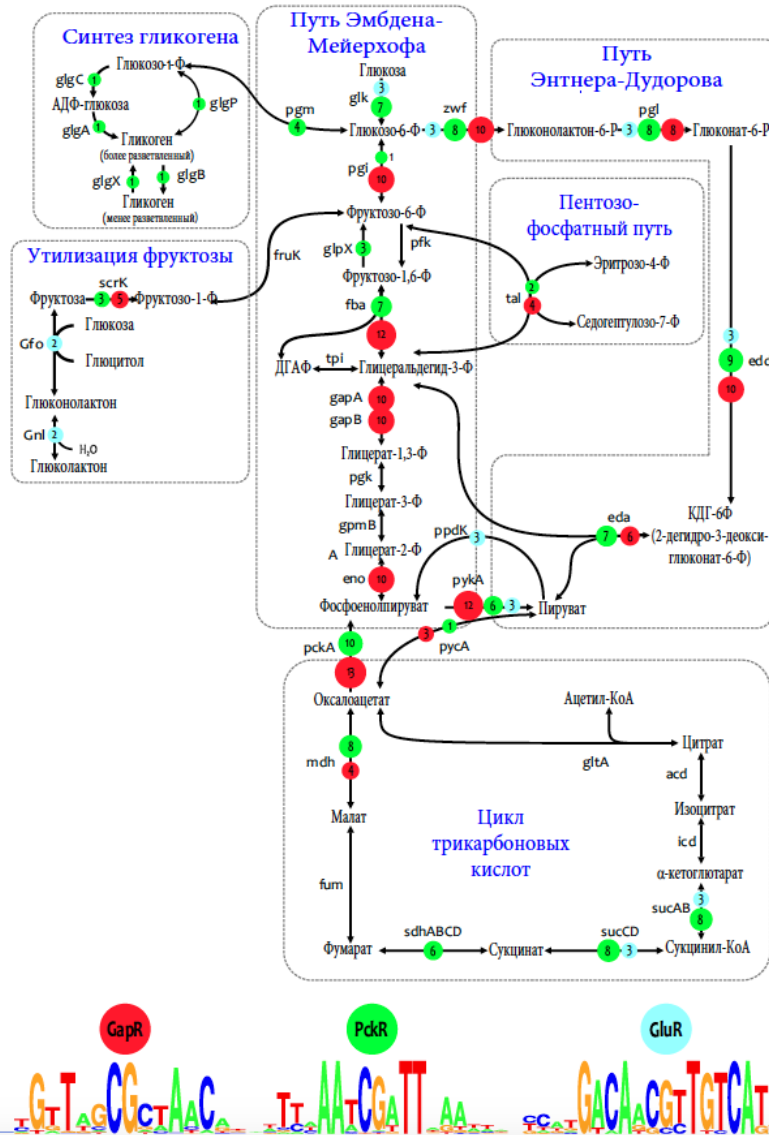


Рис. 6 Предсказанные глобальные регулоны центрального углеводного метаболизма у *Alphaproteobacteria*. (A) Регулон *GarR* у *Rhodobacteraceae*. (B) Регулон *PckR* у *Rhizobiales*. (C) Регулон *GluR* у *Caulobacter spp.* Названия регулируемых генов и оперонов показаны рядом со стрелками. Цифры в кружках показывают число геномов, в которых данный ген/оперон регулируется соответствующим ТФ

нашей работы вышла работа, в которой регулятор PckR был изучен экспериментально у *Sinorhizobium meliloti*; в этой работе многие наши предсказания были подтверждены экспериментально. В частности, было показано, что PckR связывает сайты с предсказанным мотивом, репрессирует гены гликолиза *zwf-pgl-edd*, *eda2* и *mgsA* и



активирует гены глюконеогенеза *pckA* и *fba*. Также, были проведены эксперименты, поддерживающие гипотезу о том, что эффектором PckR является фосфоенолпируват. Регулятор PckR у *Rhizobiales* можно рассматривать как частичную функциональную замену регуляторам Cra/FruR (см. выше) у *Enterobacteria* и HexR у *Shewanella*. Оба регулятора модулируют направление потока углерода через различные пути углеводного метаболизма.

Новый глобальный регулятор генов сахарного метаболизма GapR (ген RSP\_1663 в геноме *Rhodobacter capsulatus*) был найден в 13 изученных геномах семейства (Рис. 6). Реконструированные регулоны GapR содержат от 7 до 18 генов на геном, организованные в 5-13 оперонов. GapR регулирует гены путей Эмбдена-Мейерхофа и Энтнера-Дудорова (*zwf-pglpgi*, *edd-eda*, *fba*, *gapA*, *gapB*, *eno*, *pykA*), глюконеогенеза (*pckA*, *pycA*), пути утилизации фруктозы (*scrK*), пентозо-фосфатного пути (*tal*) и цикла трикарбоновых кислот (*mdh*). Мотив связывания был найден в регуляторных областях генов *gapR* в пяти геномах, что позволяет предположить наличие авторегуляции. Так же как *pckR* в *Rhizobiales*, гены *gapR* не соседствуют с регулируемым генами на хромосоме в большинстве изучаемых геномов. Роль регулятора GapR была также позднее подтверждена в независимом исследовании. Было показано, что GapR может активировать гены цикла трикарбоновых кислот (*sdhDA*, *mdh*, *fumC*, *sucCD* и *sucB*) и гены гликолиза/глюконеогенеза (*pgi*, *fba*, *pdhAB* и *pckA*) и репрессировать гены пути Энтнера-Дудорова (*zwf*, *pgl* и *eda*), связывая сайты с предсказанным нами мотивом. Авторы этого исследования предположили, что эффектором GapR может быть 6-фосфоглюконат.

Мы обнаружили еще один глобальный регулятор из семейства LacI (не-ортологичный генам PckR и GapR) в бактериях семейства *Caulobacteraceae*, и назвали его GluR (CC2053 в *Caulobacter crescentus*) (Рис. 6). Реконструированный регулон GluR в *Caulobacter* включает в себя опероны гликолитических генов *zwf-pgl-edd-glk*, *pykA* и *gnl-gfo*, а также ген глюконеогенеза *ppdK*, кодирующий пируват-фосфат ди-киназу и оперон *sucABCD*, кодирующий гены цикла трикарбоновых кислот. Мы не нашли ортологов GluR в других альфа-протеобактериях. Молекулярный эффектор GluR неизвестен; основываясь на схожести аминокислотной последовательности GluR и последовательностей предсказанных репрессоров BglR, которые, возможно, реагируют на концентрацию бета-глюкозидов и/или глюкозы, мы предполагаем, что GluR может диссоциировать от своих ДНК-сайтов связывания при связывании молекулы глюкозы. Ранее было экспериментально показано, что глюкоза активирует экспрессию генов пути Энтнера-Дудорова и что гены *edd* и *glk* необходимы для утилизации глюкозы в *C. crescentus*.

#### **Глава 4. Реконструкция метаболических путей биосинтеза витаминов группы В в бактериях микробиома кишечника человека**

В двух предыдущих главах основное внимание было уделено метаболическим путям и регуляторной сети либо для одной группы бактерий, либо для одной группы регуляторов. Другим важным вопросом является разнообразие метаболических путей среди организмов в контексте бактериального сообщества, занимающего определенную экологическую нишу. Чтобы дополнить результаты предыдущих глав и взглянуть на метаболические сети под другим углом, мы выбрали в качестве объекта третьей части нашего исследования разнообразие путей биосинтеза витаминов группы В среди бактерий, населяющих кишечный микробиом человека. Для того, чтобы сделать наш анализ как можно более полным, мы составили выборку из 2228 бактерий - представителей кишечного

микробиома, для которых были расшифрованы геномы. Выборка включает в себя 2228 геномов, представляющих 7 типов, 42 порядка, 93 семейства, 226 родов и 690 видов, а также 176 геномов, для которых не определена их таксономическая принадлежность. Больше всего геномов в данной выборке относится к типам *Firmicutes* (1046 генома), *Proteobacteria* (588 геномов), *Actinobacteria* (311 геномов) and *Bacteroidetes* (205 геномов). Основываясь на знании биохимии биосинтеза витаминов из литературы, мы установили фенотипические правила, которые описывают набор генов, чье присутствие или отсутствие в анализируемом геноме позволяет точно определить используемый вариант метаболического пути и предсказать возможные транспортируемые субстраты. Для каждого витамина, мы разделили варианты метаболических путей на две большие категории: прототрофы, которые могут синтезировать витамин сами, и ауксотрофы, которые не могут синтезировать витамин сами и которым поэтому приходится транспортировать витамины (или их метаболические предшественники) из окружающей среды. Варианты метаболических путей соответствуют альтернативным биосинтетическим путям (Рис. 7). Помимо путей биосинтеза, мы также провели анализ путей “сохранения” витаминов из окружающей среды и их последующего превращения в биологически активные кофакторы.

Одна из трудностей реконструкции метаболических путей состоит в том, чтобы отличить неполные прототрофные пути с “пропущенными” генами, которые отражают неортологичное замещение ферментативных активностей, от нетипичных ауксотрофных вариантов. Такие случаи требуют индивидуального подхода, т.к. на данный момент не существует автоматических методов для их разрешения. Например, у 94 геномов был обнаружен неполный путь синтеза В1 с недостающей синтетазой иминоглицина (ThiO или ThiH), при присутствии всех остальных генов синтеза обоих основных предшественников В1 (ThiO и ThiH) (Рис. 7). Мы предположили, что эти штаммы могут быть прототрофами, использующими неизвестный путь синтеза иминоглицина. В пути синтеза В3 фермент NadB (или альтернативный фермент NadB2), катализирующий синтез иминоаспартата из аспартата, отсутствует в геномах 34 предполагаемых прототрофов по В3, позволяя предположить использование неизвестного фермента для синтеза иминоаспартата этими бактериями. В пути синтеза В5, мы нашли 95 штаммов с отсутствующей декарбоксилазой аспартата (PanD or PanP), в которых сохранились все остальные ферменты биосинтеза В5, позволяющие предположить наличие еще неохарактеризованных альтернативных ферментов или биохимических путей синтеза бета-аланина в этих бактериях. В самом деле, существуют альтернативные пути биосинтеза бета-аланина, не связанные с метаболизмом кофакторов (например при помощи аланин-рацемазы или через метаболизм пиримидинов). Более того, в теории, бета-аланин может поступать внутрь бактерии из внешней среды. Наконец, в пути биосинтеза В7, у 74 видов из нескольких разных типов нет ферментов, выполняющих первые несколько реакций синтеза пимелоила (предшественника В7), но есть ферменты, выполняющие все последующие реакции, необходимые для сборки слитых гетероциклических колец биотина (BioF, BioA, BioD and BioB). Мы предполагаем существование еще не известных ферментов для биосинтеза пимелоила в этих видах. Основываясь на этих наблюдениях, мы классифицируем некоторые пути как неполные прототрофные пути с “отсутствующими” генами, а не как ауксотрофные пути. В то же время, мы классифицировали многих бактерий с неполными путями биосинтеза В1, В3, В5, В7 и В12 как настоящих ауксотрофов, которые

потенциально могут сохранять неканонические метаболитические предшественники соответствующих кофакторов.

Наибольшее разнообразие таких вариантов наблюдается в биосинтезе тиамин пирофосфата (ТПФ). ТПФ синтезируется из двух частей-предшественников, а именно фосфорилированного гидросиметилпиримидина (ГМП) и гидросиэтилтиазола (ГЭТ). Мы выделили несколько групп ауксотрофных штаммов, которые отличаются способностью синтезировать и/или сохранять либо предшественники (ГМП и ГЭТ), либо внеклеточный витамин В1. Биосинтез кофакторов НАД(Ф) также связан с несколькими путями сохранения и/или реутилизации либо одной, либо обеих форм витамина В3: (1) никотиновой кислоты (также известной как ниацин), и (2) никотинамида. Также, мы нашли 86 штаммов, содержащие неполный вариант пути биосинтеза НАД, состоящий из одного лишь фермента, выполняющего самый последний шаг в пути биосинтеза НАД (NadC), но не содержащий всех остальных предшествующих ферментов из этого пути; Мы предполагаем, что эти бактерии могут синтезировать НАД из квинолината. В случае

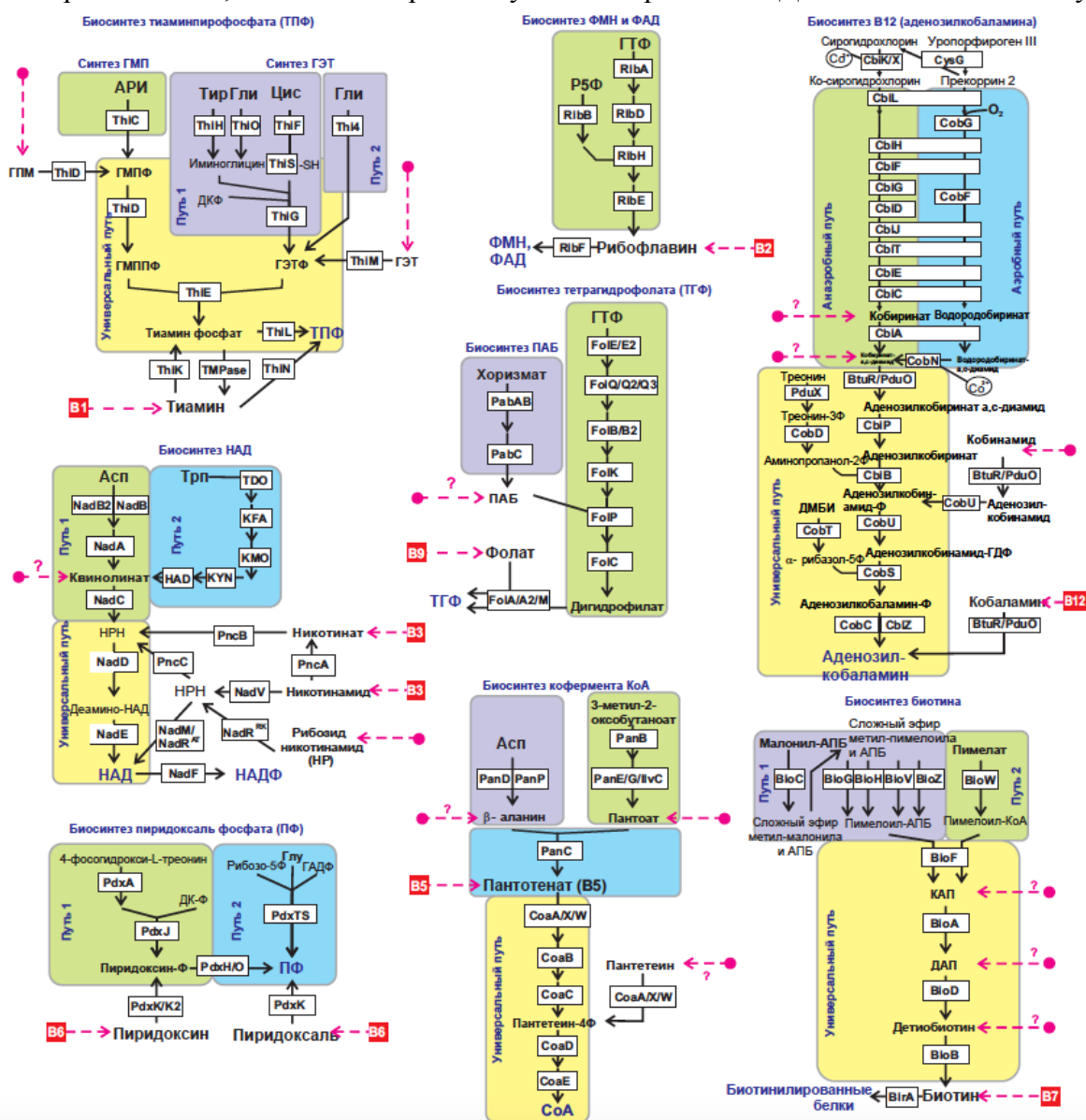


Рис. 7 Схема основных вариантов путей биосинтеза и сохранения витаминов группы В. Метаболические реакции обозначены черными стрелками, транспорт предшественников показан малиновыми стрелками; если соответствующий транспортер неизвестен, стрелка отмечена знаком

вопроса. Каждый блок метаболических реакций, отличающийся между различными вариантами путей, отмечен фоновым цветом.

витамина В5 мы выделили несколько вариантов неполных путей; мы предполагаем, что некоторые бактерии могут сохранять внеклеточный пантетеин (продукт деградации КоА), тогда как другие бактерии могут также сохранять внеклеточный пантоат. Биотин синтезируется из предшественника пимелоила через универсальный путь, состоящий из четырех последовательных реакций. Мы обнаружили геномы ауксотрофов с тремя возможными неполными путями, в которых отсутствуют гены биосинтеза трех предшественников биотина: детиобиотина, 7,8-диаминопеларгоновой кислоты и 7-кето-8-аминопеларгоновой кислоты, соответственно. Мы предполагаем, что ауксотрофы с такими неполными путями могут сохранять эти предшественники из окружающей среды. Витамин В12, в отличие от других витаминов, не является строго необходимым для всех бактерий. Тем не менее, мы обнаружили 3 варианта неполных путей биосинтеза В12. Мы предполагаем, что бактерии, несущие эти варианты путей, могут сохранять такие предшественники В12, как кобинамид, диамид кобирината и кобириновую кислоту, соответственно.

Из-за того, что определение специфичности транспортеров сугубо биоинформатическими методами представляет собой трудно решаемую задачу, мы не включали наличие/отсутствие транспортеров в качестве одного из определяющих признаков в наши фенотипические правила. Однако, с другой стороны, специфичность транспортеров можно гораздо точнее определять в контексте ауксотрофных геномов (т.к. метаболические требования ауксотрофов накладывают дополнительные ограничения на функции транспортеров) и после этого проецировать на прототрофные геномы. Мы включили все предсказанные транспортеры в соответствующие подсистемы и в схемы метаболических путей (см. приложения). Ниже, мы вкратце описываем те транспортные системы, которые, согласно нашим предположениям, могут участвовать в переносе альтернативных предшественников витаминов через клеточную мембрану.

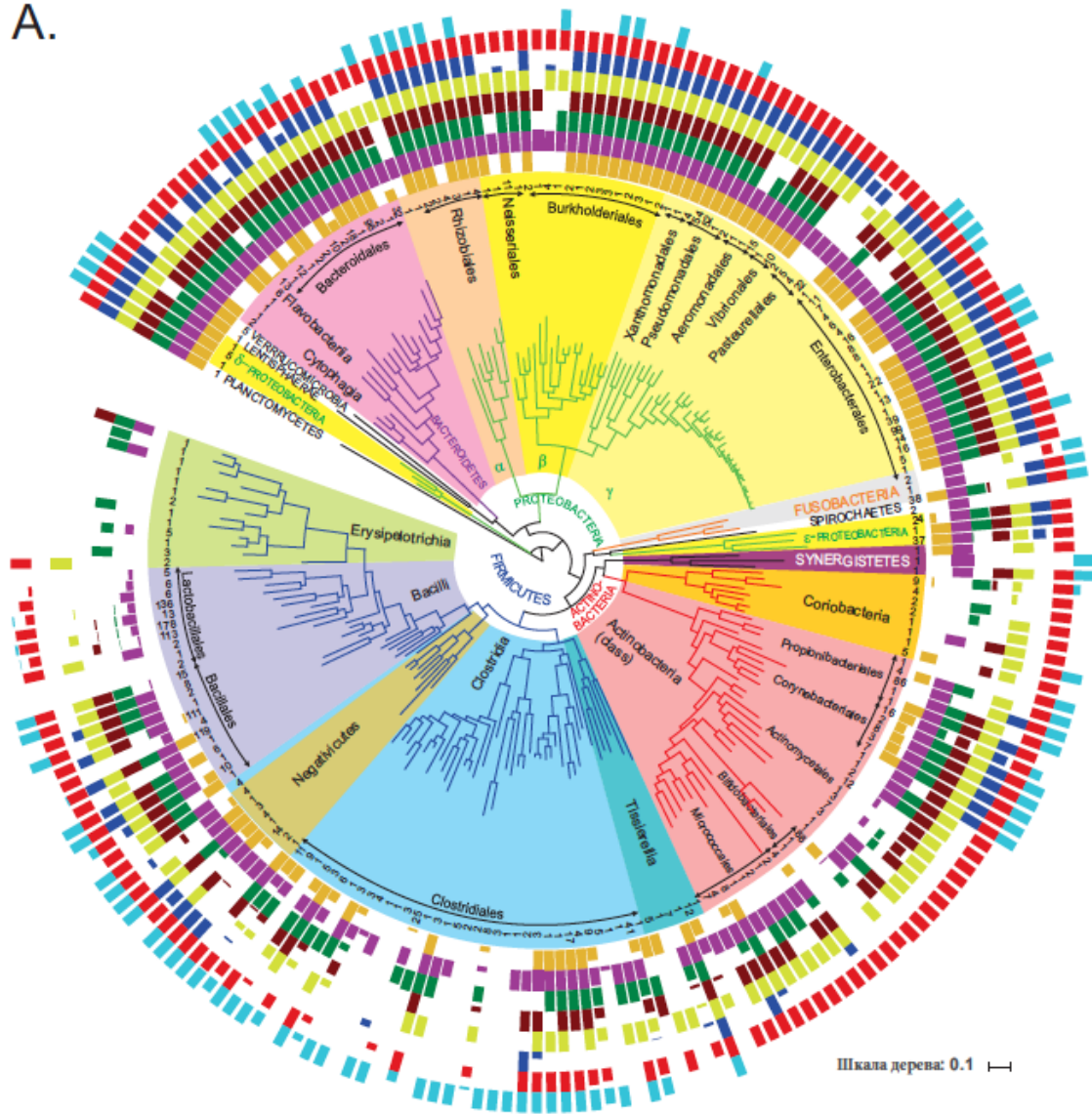
Так, анализ распределения возможных транспортеров предшественников витамина В1 позволил нам предположить, что у 97% геномов-ауксотрофов по ГМП есть гены транспортера ГМП, тогда как гены транспортера ГЭТ были найдены только у 64% ауксотрофов по ГЭТ. Транспортеры витамина В3 служат хорошим примером недостаточной изученности транспортных систем у бактерий: известные транспортеры В3, относящиеся к трем известным семействам (NiaP, NiaX и NiaY), были найдены только в 25% ауксотрофов по В3. Также, мы не смогли идентифицировать транспортную систему, ассоциированную с путем сохранения квинолината. Среди геномов-ауксотрофов по В5 мы идентифицировали ортологи известных транспортеров В5 у 75% организмов, тогда как среди ауксотрофов по пантетеину (91 геном) мы смогли найти транспортеры В5 лишь у 12%. Также, мы идентифицировали ортологи транспортера PanS в 153 геномах; среди них,

многие содержат консервативный оперон *panD-panC-panS*, кодирующий путь сохранения пантоата, что позволяет нам предположить роль транспортера PanS в импорте пантоата. В геномах большинства ауксотрофов по В7 есть ген транспортера BioY (78%), тогда как среди геномов с неполными путями синтеза биотина эта цифра достигает 87% (240 из 275 геномов). Также, в некоторых геномах с неполными путями синтеза В7 мы идентифицировали ортологи транспортера YigM. Мы предполагаем, что оба типа

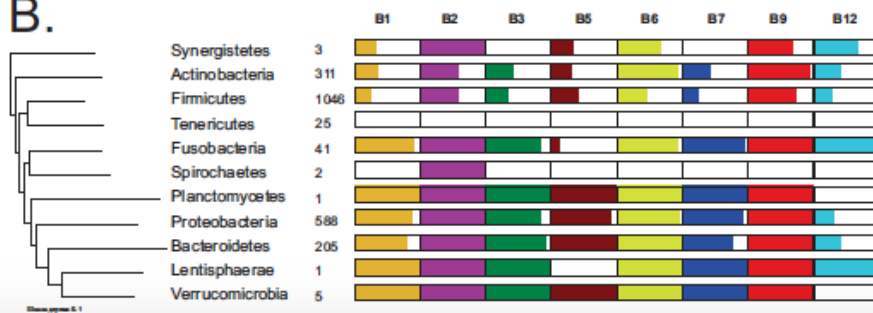
биотиновых транспортеров могут иметь более широкую специфичность к интермедиатам этого биосинтетического пути. Наконец, мы смогли идентифицировать ортологи и паралоги известных транспортеров В12 (транспортер VtuFCD(B) из семейства ABC и транспортер CbrT из семейства ECF) примерно у половины организмов - ауксотрофов по В12, в том числе у 268 организмов с неполными путями биосинтеза В12.

Из более ранних исследований известно, что метаболический обмен аминокислотами влияет на структуру бактериальных сообществ. Мы предполагаем, что метаболический обмен между бактериями распространяется и на витамины, и в т.ч, возможно, на их предшественников. В частности, мы предполагаем, что, помимо пассивного выделения витаминов при лизисе клеток, существуют активные механизмы выделения витаминов, и некоторые витамин-специфичные импортеры могут также действовать как экспортеры витаминов и их предшественников. В частности, функция экспорта может быть активна в витамин-продуцирующих штаммах, что способствует обмену витаминами и их предшественниками с ауксотрофными штаммами (акцепторами). Мы проанализировали со-встречаемость найденных витаминных транспортеров, их распределение по семействам, а также распределение фенотипов витаминного биосинтеза среди изучаемых геномов. Мы

A.



B.



**Рис. 8** Распределение прототрофных бактерий среди анализируемых геномов представителей кишечного микробиома. (A) Усредненные фенотипы по витаминам группы В среди 230 родов, представленных в изучаемой выборке геномов. Мы сконструировали филогенетическое дерево бактериальных родов на основе большего филогенетического дерева, включающего все изучаемые геномы, созданного в программе *RaxML* на основе объединенных последовательностей рибосомальных белков. Во внутреннем круге отмечено число представителей каждого рода в изучаемой выборке. Более широкие таксономические группы, такие как порядки, классы и типы, отмечены на дереве. Цветные столбики показывают усредненные витаминные фенотипы (доля прототрофов) для каждого рода. Незакрашенные столбики соответствуют ауксотрофным фенотипам. (B) Усредненные витаминные фенотипы для 11 типов. Мы получили филогенетическое дерево типов, схлопнув филогенетическое дерево родов. Число проанализированных представителей показано рядом с названием каждого типа.

наблюдали отдельные тенденции, поддерживающие нашу гипотезу. Так, рибофлавиновые пермеазы (PnuX, RibZ, ImpX, RibN, RfnT) встречаются только лишь в прототрофах по B2, тогда как активные транспортеры рибофлавина, RibU (из семейства ECF) и RibXY (из семейства ABC) представлены и в ауксотрофах, и в прототрофах. Пермеаза биотина YigM чаще встречается у бактерий - прототрофов по B7, тогда как транспортер BioY (из семейства ECF) чаще встречается у ауксотрофов. Аналогично, пермеаза пантотената PanF чаще встречается у прототрофов по B5, тогда как транспортер PanT из ECF семейства чаще встречается у ауксотрофов. Схожая тенденция наблюдается у транспортеров тиамин - пермеазы PnuT и транспортера ThiT из ECF семейства.

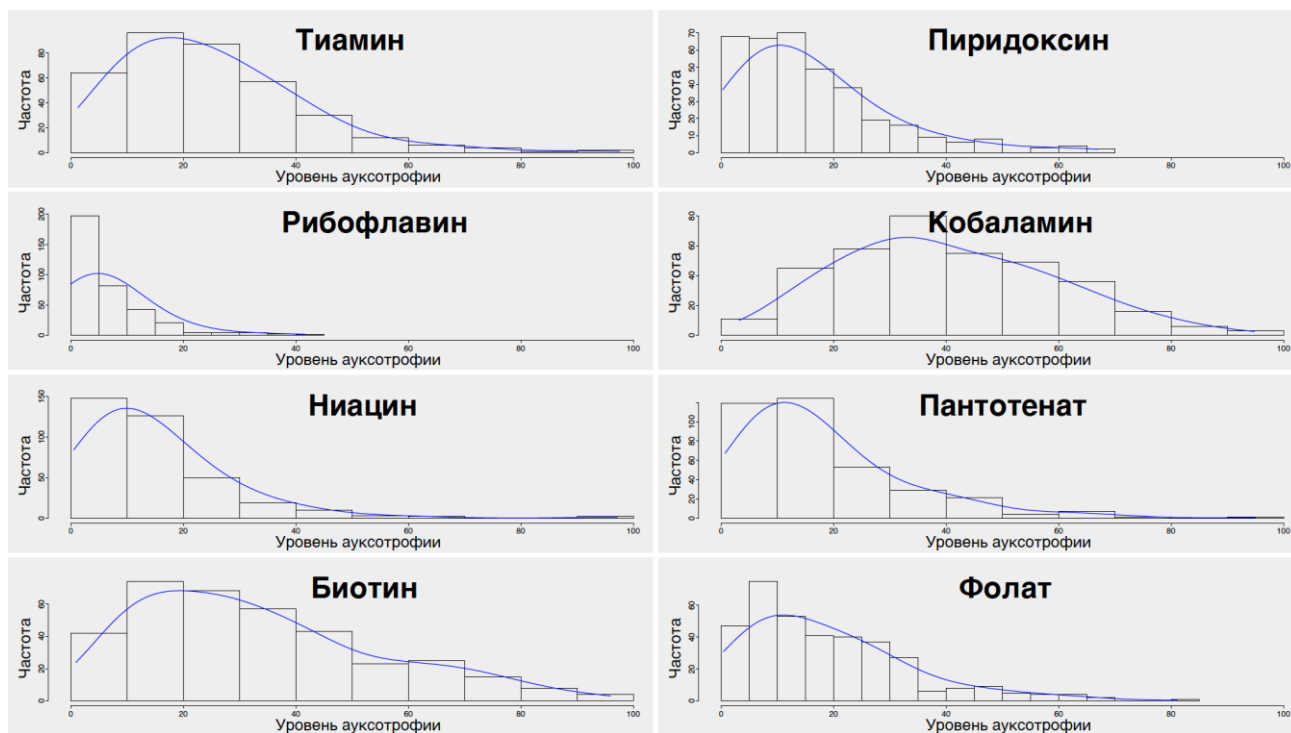
Для оценки фенотипической вариабельности на уровне видов и более широких таксономических групп мы подсчитывали число вариабельных фенотипов (варьирует от 1 до 8 для восьми витаминов) (Рис. 8). 282 из 694 изучаемых видов представлены двумя и более штаммами, в том числе 34 вида представлены 10 и более штаммами. Среди них, 45 видов (~15%) варьируют хотя бы по одному витаминному фенотипу. На уровне родов, 139 из 230 проанализированных родов были представлены двумя и более видами, и 76 из них (55%) демонстрируют вариабельные фенотипы. Рода с наиболее вариабельными фенотипами принадлежат к классам *Clostridia*, *Bacilli* и *Tissierellia* из типа *Firmicutes*. Вариабельность витаминных фенотипов растет с повышением таксономических рангов. На уровне семейств, 82 из 96 семейств представлены двумя и более родами, и 61 из них (74%) демонстрируют вариабельные фенотипы. Любопытно, что все пять семейств с наиболее вариабельными фенотипами, а именно *Eubacteriaceae*, *Ruminococcaceae*, *Clostridiaceae*, и *Lachnospiraceae*, принадлежат к классу *Clostridia*. На уровне порядков, 30 из 36 таксонов, представленных более чем одним геномом (83%) демонстрируют вариабельные фенотипы.

Вышеописанные наблюдения о распределении метаболических путей биосинтеза и сохранения витаминов интересны с точки зрения филогенетического разнообразия бактерий. Однако, в бактериальных сообществах, и, в частности, в кишечном микробиоме, численность различных бактериальных таксонов распределена неравно: несколько доминантных видов бактерий представляют до 90% микробиома численно. Поэтому с точки зрения изучения бактериального сообщества большой интерес представляют численные соотношения вариантов метаболических путей в настоящих образцах кишечного микробиома. Мы проецировали предсказанные описания метаболических путей на данные о численном соотношении бактерий в образцах кишечного микробиома.

Для каждого изучаемого образца, мы вычисляли две метрики: (1) покрытие, отражающее долю организмов в этом образце, для которых мы могли предсказать витаминные фенотипы, и (2) численные значения среднего уровня ауксотрофности, отражающие долю ауксотрофов по каждому витамину.

Мы проанализировали баланс прототрофов и ауксотрофов в образцах, собранных в процессе четырех независимых исследований. Во-первых, мы оценивали, насколько хорошо наша выборка бактериальных геномов отражает состав кишечного микробиома человека. Для этого, мы использовали данные из исследования, где у четырех здоровых индивидуумов были взяты образцы микробиома из разных участков желудочно-кишечного тракта. Мы наблюдали высокое покрытие только в образцах кишечного микробиома, но не в образцах из вышележащих участков (желудок), что говорит о том, что





**Рис.9.** Распределения средних уровней ауксотрофии для витаминов группы В в образцах кишечного микробиома, собранных из 388 индивидуумов в рамках проекта «Микробиом человека». Средний уровень ауксотрофии вычисляется как численная доля ауксотрофов среди геномов с предсказанными фенотипами.

выборка анализируемых геномов высоко специфична для кишечного микробиома, но покрывает его в достаточной мере, чтобы адекватно оценивать баланс ауксотрофов и прототрофов в сообществе.

Далее, мы использовали оценку среднего уровня ауксотрофности для того, чтобы ответить на два основных биологических вопроса: во-первых, насколько сильно варьирует средний уровень ауксотрофности среди различных индивидуумов, и, во-вторых, насколько сильно он меняется в зависимости от изменения питания организма-хозяина. Для ответа на первый вопрос, мы проанализировали 388 образцов кишечного микробиома, собранных в рамках проекта “Микробиом Человека”. Эти образцы были собраны из широкой выборки индивидуумов с различными генетическими и фенотипическими признаками и поэтому позволяют качественно оценить разнообразие кишечных бактериальных сообществ среди людей. Проанализировав эти образцы, мы получили распределения среднего уровня ауксотрофии в кишечном микробиоме. Средние уровни ауксотрофии по тиамину, ниацину, рибофлавиону, пиридоксину и пантотенату оказались довольно низкими, что указывает на то, что кишечник заселен по большей части прототрофными организмами. Это подтверждает ранее опубликованные результаты о том, что большинство витаминов всасывается клетками верхних слоев желудочно-кишечного тракта, и что самостоятельный биосинтез витаминов может быть выгоден бактериям, населяющим кишечник. Однако, средний уровень ауксотрофии по всем перечисленным витаминам все же сильно отличается от нуля: медианное значение составляет от 10% до 20%, в зависимости от витамина (Рис. 9). Многие бактерии могут существовать в кишечном микробиоме, не синтезируя витамины самостоятельно. Это указывает на возможность обмена метаболитами в этом сообществе, однако масштаб такого обмена предстоит изучить. В то же время, распределения среднего уровня ауксотрофии по кобаламину и по биотину сильно отличаются от таких распределений для других витаминов: средний



уровень ауксотрофии для этих витаминов сильно выше, чем для других (Рис. 9). Особенно высок уровень ауксотрофии для кобаламина. Это можно объяснить тем, что, как упоминалось выше, многие бактерии не нуждаются в кобаламине, т.к. в их метаболических схемах В12-зависимые ферменты замещены альтернативными В12-независимыми ферментами.

Второй интересующий нас вопрос заключается в том, сколь сильно меняется уровень ауксотрофии кишечного микробиома при изменении питания организма-хозяина. Мы проанализировали состав кишечного микробиома из двух исследований, в которых у фиксированной группы индивидуумов берут образцы кишечного микробиома через равные промежутки времени, и при этом в ходе исследования происходит смена питания. В первом исследовании были искусственно созданы две различные диеты - растительная и мясная. 10 волонтеров питались только по выбранной ими диете на протяжении 5 дней, сдавая при этом образцы кишечной микробиоты ежедневно, начиная за 4 дня до перехода на новую диету, и заканчивая 6 дней спустя после обратного перехода к нормальному питанию. Авторы данного исследования отмечают значительное изменение в составе микробиома, специфичное для мясной диеты. Мы вычислили средний уровень ауксотрофии для каждого образца в данном исследовании; мы не наблюдали значительных изменений уровня ауксотрофии при смене питания (Рис. 9). В другом исследовании были изучены изменения состава микробиома в трех различных популяциях из Монголии. Жители провинции Хэнтий ведут кочующий образ жизни, и их питание меняется в зависимости от времени года: зимой и весной основу их рациона составляют мясные продукты, а летом и осенью - молочные продукты. В то же время, жители Улан-Батора (столицы Монголии) и провинции Туве (пригородов столицы) ведут городской образ жизни, и их рацион подвержен значительно меньшим сезонным изменениям. Исследователи брали образцы у 64 индивидуумов, проживающих в одной из этих трех точек, 5 раз в течение года, а именно - в январе, в марте, в июне, в сентябре и в ноябре. Для образцов, полученных в ходе данного исследования, мы также вычислили уровень ауксотрофии; в данном случае, мы тоже не наблюдали значительных сезонных изменений среднего уровня ауксотрофии, соответствующих изменению рациона питания организма-хозяина.

## **Выводы**

1. Проведена реконструкция регулонов для 268 транскрипционных факторов из семейств LacI, ROK, DeoR, AraC, GntR и TetR, контролирующей катаболизм углеводов у бифидобактерий
2. Предсказан первый глобальный регулятор центрального углеводного метаболизма у бифидобактерий
3. Показано, что транскрипционные факторы из семейства TetR могут регулировать катаболизм углеводов
4. Проведена реконструкция регулонов для 1303 транскрипционных факторов из семейства LacI у 272 бактерий
5. Показано, что предпочтительное расстояние между соседними связывающими сайтами для транскрипционных факторов из семейства LacI кратно числу нуклеотидов, соответствующему целому числу витков спирали ДНК

6. Предсказаны 3 новых глобальных регулятора центрального углеводного метаболизма у альфа-протеобактерий
7. Проведена реконструкция метаболических путей биосинтеза, захвата и сохранения восьми витаминов группы В в геномах 2228 представителей кишечного микробиома
8. Показано, что среди бактерий кишечного микробиома встречаются “неполные” пути биосинтеза витаминов, начинающиеся с промежуточных предшественников
9. Сформулирована гипотеза о том, что бактерии кишечного микробиома могут обмениваться такими веществами, как тиазол, квинолинат, детиобиотин и пантоат
10. Проведена оценка того, насколько баланс между прототрофными и ауксотрофными бактериями в кишечном микробиоме варьирует у различных индивидуумов
11. Мы не наблюдали значительного изменения баланса между прототрофными и ауксотрофными бактериями в кишечном микробиоме при изменении рациона питания у организма хозяина

### Список публикаций по теме диссертации

#### Статьи в рецензируемых научных журналах

1. **Khoroshkin MS**, Leyn SA, Van Sinderen D, Rodionov DA. “*Transcriptional regulation of carbohydrate utilization pathways in the Bifidobacterium genus*”, **Frontiers in microbiology**, 2016; 7:120.
2. Wu M, McNulty NP, Rodionov DA, **Khoroshkin MS**, Griffin NW, Cheng J, Latreille P, Kerstetter RA, Terrapon N, Henrissat B, Osterman AL, Gordon JI. “*Genetic determinants of in vivo fitness and diet responsiveness in multiple human gut Bacteroides*”, **Science**, 2015; 350(6256):aac5992.;
3. Ravcheev DA\*, **Khoroshkin MS\***, Laikova ON, Tsoy OV, Sernova NV, Petrova SA, Rakhmaninova AB, Novichkov PS, Gelfand MS, Rodionov DA. “*Comparative genomics and evolution of regulons of the LacI-family transcription factors*”, **Frontiers in Microbiology**, 2014; 5:294.; \* ко-первые авторы

#### Тезисы конференций

1. **Khoroshkin MS**, “*Variability of B-vitamin Dependencies in Human Microbiome Genomes*”, RECOMB/ISCB Conference on Regulatory and Systems Genomics 2014, San Diego, CA, USA (устная презентация)
2. **Khoroshkin MS**, Rodionov DA, “*Transcriptional Regulation of the Carbohydrate Metabolism in Bifidobacterium Genus*”, American Society for Microbiology General Meeting 2015, New Orleans, LA, USA
3. **Khoroshkin MS**, Osterman AL, Rodionov DA, “*Variability of B-vitamin dependencies in the human microbiome genomes*”, Microbial Genomics 2014, Lake Arrowhead, CA, USA
4. **Khoroshkin MS**, Leyn SA, Rodionov DA, “*Transcriptional Regulation of the Carbohydrate Metabolism in Bifidobacterium Genus*”, Moscow Conference on Computational Molecular Biology 2015, Moscow, Russia
5. **Хорошкин МС**, Родионов ДА, “*Различные стратегии транскрипционного управления центрального углеводного метаболизма у бактерий*”, Информационные

технологии и системы – 2015, Сочи, Россия

6. **Хорошкин МС**, Родионов ДА, *"Сравнительно-геномный анализ и реконструкция метаболических регулонов для транскрипционных факторов из семейства LacI"*, Информационные технологии и системы – 2013, Калининград, Россия