

**Федеральное государственное бюджетное учреждение науки
Институт проблем передачи информации им. А.А.
Харкевича
Российской академии наук**

На правах рукописи

Хорошкин Матвей Сергеевич

*Реконструкция регулонов метаболических путей в бактериях
микробиоты кишечника человека*

*Специальность 03.01.09 – "математическая биология,
биоинформатика"*

*Диссертация на соискание ученой степени кандидата
биологических наук*

Научный руководитель
Родионов Дмитрий Александрович, к.б.н

Оглавление

Введение	4
Актуальность темы	4
Степень разработанности темы	5
Цели и задачи исследования	6
Научная новизна и практическая значимость работы	7
Выносимые на защиту положения	7
Глава 1. Обзор литературы	9
Введение	9
Регуляция инициации транскрипции	9
Изучение регуляции транскрипции.....	13
Предсказание регуляторных взаимодействий	13
Проверка предсказаний	15
Методы сравнительной геномики	16
Поиск регуляторов	17
Поиск регулируемых мишеней	17
Поиск сайтов для тренировочной выборки.....	18
Филогенетический футпринтинг	18
Поиск сайтов в группе регуляторных областей	19
Построение мотива связывающих сайтов	19
Поиск новых связывающих сайтов.....	21
Установление соответствий между мишенями и регуляторами	22
Реконструкция метаболических путей.....	24
Биология бифидобактерий.....	25
Бифидобактерии как пробиотики	25
Углеводный метаболизм бифидобактерий.....	26
Происхождение семейства транскрипционных факторов LacI.....	27
Роль метаболизма витаминов группы В в формировании структуры микробиома	27
Глава 2. Изучение транскрипционной регуляции путей утилизации углеводов в бактериях рода <i>Bifidobacterium</i>	29
Методы	29
Результаты	31
Репертуар транскрипционных факторов, участвующих в регуляции сахарного метаболизма, в геномах рода <i>Bifidobacterium</i>	31
Геномная реконструкция регулонов транскрипционных факторов	35
Анализ функций реконструированных регулонов	37
Ко-эволюция рибозных и нуклеозидных регулонов	40
Гены утилизации мальтозы и мальтодекстрина регулируются паралогичными ТФ из семейства LacI.....	41
Контроль утилизации глюкозидов и галактозидов новыми ТФ из семейства TetR	44
Обсуждение	50
Глава 3. Исследование транскрипционной регуляции и эволюции регулонов из семейства LacI	52
Методы	52
Результаты	53

Репертуар генов ТФ из семейства LacI в бактериальных геномах.....	53
Статистика реконструированных регулонов и регулогов.....	53
Глобальные и локальные регулоны.....	55
Связывающие сайты и мотивы.....	58
Регулируемые метаболические пути и эффекторы.....	60
Глобальные регуляторы, контролирующие центральный сахарный метаболизм.....	63
Обсуждение.....	70
Глава 4. Реконструкция метаболических путей биосинтеза витаминов группы В в	
бактериях микробиома кишечника человека.....	80
Методы.....	80
Составление выборки геномов-представителей микробиома кишечника человека.....	80
Геномная реконструкция подсистем метаболизма витаминов.....	81
Предсказания уровней ауксотрофии образцов человеческого микробиома.....	83
Результаты.....	84
Реконструкция путей биосинтеза и сохранения витаминов.....	84
Неполные пути биосинтеза кофакторов и сохранение витаминов.....	88
Геномное распределение транспортеров витаминов.....	93
Бинарные витаминные фенотипы: филогенетическое распределение и вариабельность.....	98
Уровни ауксотрофии в образцах человеческого кишечного микробиома.....	102
Влияние генов витаминного метаболизма на выживаемость бактерий из рода <i>Bacteroides</i> в	
кишечном микробиоме человека.....	110
Обсуждение.....	116
Выводы.....	119
Список публикаций по теме диссертации.....	120
Список используемых сокращений и обозначений.....	121
Список использованной литературы.....	121
Приложения.....	133

Введение

Актуальность темы

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 и анализ регулонов ТФ семейства LacI в 272 геномах (а также нахождение сайтов связывания, молекулярных эффекторов и функциональных характеристик регулируемых генов)
2. Геномная реконструкция и анализ регуляции сахарного метаболизма в 10 геномах бифидобактерий
3. Сравнительный анализ метаболических путей биосинтеза В-витаминов 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, контролирующих 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, AraC, GntR и TetR. Впервые предсказан глобальный регулятор AraQ из семейства LacI, контролирующий 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, квинолинат, 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

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

Глава 1. Обзор литературы

Введение

Функциональная геномика - это наука, изучающая сложные взаимоотношения между генотипами и фенотипами организмов. С практической точки зрения, ее основная цель - зная генотип (т.е совокупность всех генов) данного организма, научиться предсказывать, как будет меняться фенотип (т.е совокупность всех признаков) данного организма в зависимости от условий окружающей среды. В соответствии с центральной догмой, гены кодируют белки, а белки катализируют большинство химических реакций в клетке [1]. Даже в активно изучаемых модельных организмах (таких, как кишечная палочка, дрожжи и плодовая мушка) функции многих генов еще не изучены (например, см. [2]). Таким образом, в первом приближении функциональная геномика стремится изучить, какие роли различные гены играют в клетке. Однако, клетки с одним и тем же генотипом могут проявлять разные фенотипические признаки в зависимости от условий окружающей среды. Такая фенотипическая гибкость обеспечивается сложной системой регуляции, контролирующей количество молекул фермента, производимого на основе каждого гена (т.н. экспрессию генов), а также их активность. Изучение регуляторных систем клетки, обеспечивающих взаимодействие генотипа с условиями окружающей среды, также является областью интереса функциональной геномики. Клетка регулирует количество и активность производимых ферментов на нескольких уровнях: транскрипция, трансляция, а также посттрансляционные модификации белков. Около 80% вариативности между клетками обеспечивается регуляцией транскрипции, однако остальные уровни регуляции также очень важны [3]. Из-за доминирующей роли регуляции транскрипции, в данной работе мы решили сфокусироваться на исследовании регуляции транскрипции.

Регуляция инициации транскрипции

Уровни экспрессии различных генов в любой отдельно взятой бактерии могут отличаться на несколько порядков. Более того, экспрессия многих генов может меняться в ответ на изменения условий окружающей среды. Такая регуляция необходима для приспособляемости бактерии. Для большинства генов у большинства бактерий, ключевым регулируемым шагом, определяющим уровень экспрессии, является инициация транскрипции. Ключевую роль в этом процессе играет комплекс РНК-полимеразы. Основным фермент РНК-полимеразы состоит из 5

субъединиц (бета, бета-штрих, две альфа-субъединицы и омега-субъединицы). Основной фермент может синтезировать РНК, используя ДНК в качестве матрицы, однако не может распознавать и связывать промотеры. Таким образом, для инициации транскрипции необходимы дополнительные регуляторные белки, обеспечивающие распознавание промотеров - сигма-факторы. Связанная с сигма-фактором РНК полимеразы (такой комплекс называется холоферментом) может распознавать специфические последовательности промотеров и, таким образом, может инициировать транскрипцию [4].

Регуляторные механизмы, контролирующие инициацию транскрипции у прокариот, можно условно разделить на две большие группы: регуляция функций самой РНК-полимеразы и регуляция промотера (например, доступности промотера для РНК-полимеразы и аффинности связывания) [4].

Прежде всего, в состав РНК-полимеразы могут быть включены разные сигма-факторы (как упомянуто выше). Как правило, геномы бактерий содержат один основной сигма-фактор, который участвует в инициации транскрипции подавляющего большинства генов, и несколько альтернативных сигма-факторов. Например, в *E. coli* таким фактором является сигма-70, также известный как RpoD. Каждый альтернативный сигма-фактор распознает промотеры определенной группы генов. Как правило, эти гены входят в состав одного или нескольких метаболических путей, необходимых для ответа на определенные изменения окружающей среды или состояния клетки. Основной сигма-фактор и альтернативные сигма-факторы конкурируют за связывание основного фермента РНК-полимеразы. Изменение концентрации того или иного сигма-фактора сдвигает равновесие, соотношение РНК-полимераз, связанных с различными сигма-факторами, меняется, и, таким образом, меняется транскрипционный ландшафт клетки. Как правило, в “нормальном” состоянии клетки большинство активных РНК-полимераз содержат основной сигма-фактор, но при различных стресс-стимулах соотношение сдвигается и альтернативные сигма-факторы активируют транскрипцию генов, нужных клетке в данных условиях [4], [5].

Многие факторы напрямую взаимодействуют с РНК-полимеразой, модулируя ее активность. Инициация транскрипции - это сложный процесс, состоящий из последовательности образования различных промежуточных комплексов, переходящий в итоге в элонгацию транскрипции. Белки, которые взаимодействуют с РНК-полимеразой на разных стадиях процесса инициации, стабилизируют определенные промежуточные комплексы и таким образом регулируют инициацию транскрипции. Зачастую, такие регуляторы видо-специфичны (как, например, RbpA и CarD у актиномицет [6]). Регулятор DksA у *E. coli* может

стабилизировать или дестабилизировать взаимодействия РНК-полимеразы с различными промоторами в зависимости от концентрации петнафосфата гуанозина [7].

Некоторые молекулы могут связывать холофермент РНК-полимеразы и таким образом уменьшать количество доступных для инициации транскрипции комплексов РНК-полимеразы. Например, 6S РНК (некодирующая РНК длиной около 180 нуклеотидов) может имитировать структуру бактериального промотера и таким образом связывать холофермент РНК-полимеразы в соотношении 1:1 [8]. При этом количество несвязанных холоферментов РНК-полимеразы в клетке уменьшается и общий уровень транскрипции также понижается.

С другой стороны, инициация транскрипции также регулируется множеством факторов, действующих через связывание промотерных последовательностей, а не самой РНК-полимеразы. Несмотря на то, что существует много промотер-центричных механизмов регуляции, наиболее распространенным и наиболее важным механизмом считается регуляция при помощи так называемых транскрипционных факторов, т.е белков, специфически связывающих ДНК рядом с промоторами, и таким образом влияющих на взаимодействие РНК-полимеразы с промоторами. Транскрипционные факторы представляют собой один из наиболее важных механизмов регуляции экспрессии генов в бактериях [4].

Общая функция транскрипционных факторов - связать активность промоторов с изменениями окружающей среды. Поэтому большинство транскрипционных факторов содержит как минимум два активных сайта: один из них специфически связывает ДНК, другой может связывать лиганд - метаболит, поступающий извне. При связывании с лигандом транскрипционный фактор меняет свою конформацию, что вызывает изменения в специфичности связывания ДНК [9]. ДНК-связывающие сайты ТФ положительно заряжены, поэтому, даже если ТФ не находится в активной конформации, большую часть времени он все равно связывает ДНК. Однако, когда ТФ переходит в активную конформацию, его специфичность связывания с ДНК меняется. В активной конформации аффинность ТФ к определенным последовательностям ДНК на много порядков превышает его аффинность к любым другим последовательностям [10]. Для каждого ТФ, геном бактерии содержит некоторое число последовательностей, к которым данный ТФ обладает высокой аффинностью. Распределение аффинности по бактериальной хромосоме может меняться в зависимости от конкретного ТФ, но общий принцип схож: геном бактерии не содержит “оптимальных” последовательностей, т.е тех последовательностей, для которых аффинность данного ТФ была бы максимальной. При этом, геном содержит небольшое число “суб-оптимальных” последовательностей с высокой аффинностью [11] (также, [12] иллюстрирует похожую

концепцию для РНК-связывающих белков). Большая же часть генома обладает низкой аффинностью к ТФ. Суб-оптимальные последовательности составляют группу последовательностей, похожих, но не идентичных единой оптимальной последовательности (также именуемой “консенсусной последовательностью”). Как правило, они располагаются недалеко от промоторов, что позволяет связывающимся ТФ модулировать функции РНК-полимеразы.

ТФ, находящийся в активной или в неактивной конформации, имеет абсолютно разные распределения аффинности к последовательностям ДНК. Если ТФ в активной конформации, как сказано выше, связывает одни последовательности с намного большей аффинностью, чем другие, то ТФ в неактивной конформации связывает различные последовательности ДНК более равномерно. Вследствие этого, когда много молекул определенного ТФ переходят из активной конформации в неактивную (или наоборот) их общее расположение на хромосоме бактерии меняется. Это обеспечивает переключение транскрипционных программ в клетке [10]. В общем виде, механизм изменения экспрессии генов бактерии в ответ на резкое повышение концентрации питательного метаболита в окружающей среде устроен следующим образом: сначала, мембранные транспортеры бактерии переносят молекулы метаболита внутрь клетки. Затем метаболит связывает молекулы специфичного к нему транскрипционного фактора. При этом молекулы транскрипционного фактора меняют свою конформацию, и, вследствие этого, также меняют специфичность связывания ДНК [13]. При этом молекулы ТФ связываются с участками ДНК, находящимися рядом с промоторами генов пути катаболизма данного метаболита и активируют их транскрипцию. В случае, если данный ТФ действует как индуцируемый репрессор, молекулы ТФ могут быть связаны с регуляторными областями соответствующих генов в отсутствие метаболита, таким образом блокируя транскрипцию. При появлении молекул метаболита в цитоплазме клетки, ТФ освобождают эти промоторные области, и катаболические гены начинают экспрессироваться.

Транскрипционные факторы могут как активировать, так и репрессировать инициацию транскрипции регулируемых ими генов. При этом, существует несколько различных механизмов как активации, так и репрессии [14]. В случае репрессии, два основных известных механизма включают (1) стерическое блокирование связывания РНК-полимеразы и промотора, и (2) мультимеризацию ТФ и образование петли ДНК вокруг промотора, в результате которого транскрипция не может начаться (например, в случае регулятора GalR [9]). Петлю могут образовывать два разных ТФ [15]. Некоторые белки (так называемые “анти-активаторы”)

могут также репрессировать инициацию транскрипции, блокируя сайты связывания активаторов. Например, CytR служит таким анти-активатором для СМР [16].

В случае активации, большинство ТФ связываются с последовательностями, апстримными (т.е. находящимися ближе к 5' концу ДНК) по отношению к промотеру, и далее специфически взаимодействуют с РНК-полимеразой, таким образом повышая локальную концентрацию доступных комплексов РНК-полимеразы рядом с промотером и способствуя инициации транскрипции. В зависимости от доменной структуры, ТФ-активатор может взаимодействовать с различными субъединицами РНК-полимеразы. Например, Ли и соавторы выделяют несколько классов механизмов активации, в зависимости от расположения сайта связывания ТФ по отношению к промотеру и в зависимости от того, какие субъединицы РНК-полимеразы взаимодействуют с ТФ [17]. Также, существуют активаторы, которые не связывают РНК-полимеразу напрямую, но приспособливают промотер стерически для связывания РНК полимеразы. Например, некоторые промотеры, у которых расстояние между -35 и -10 участками не оптимально, могут связываться со специальными ТФ-активаторами, которые изгибают ДНК между этими двумя участками и тем самым располагают их более оптимальным образом пространственно [18]. Кроме того, как и в случае с репрессорами, активация может быть опосредованной - регуляторные области некоторых генов содержат перекрывающиеся сайты связывания репрессоров и активаторов, и связывание активатора предотвращает связывание репрессора и таким образом активирует инициацию транскрипции.

Изучение регуляции транскрипции

Предсказание регуляторных взаимодействий

В самом общем виде, изучение биологических регуляторных взаимодействий сводится к трем составляющим: (1) поиск регулируемых мишеней, (2) поиск регуляторов и (3) изучение взаимодействий между ними. Разумеется, эти составляющие научного процесса взаимосвязаны, но большинство исследований использует одну из этих составляющих как отправную точку. В случае транскрипционной регуляции, регуляторами являются транскрипционные факторы, а мишенями - гены.

Разработано множество экспериментальных методов для поиска регулонов. Одна группа широко используемых в современной микробиологии методов использует пертурбации изучаемых генов для установления их регуляторной роли в клетке. В частности, в последние 6 лет стали популярны методы, использующие модифицированные бактериальные CRISPR-системы для пертурбации уровней экспрессии изучаемых генов. Например, выключение или овер-экспрессия изучаемого гена позволяет наблюдать изменения в экспрессии генов-мишеней [19]. Очень информативными экспериментами являются CRISPR-скрины: эксперименты, в которых при помощи CRISPR-систем последовательно выключается экспрессия множества генов, и для дальнейших исследований выбираются те гены, при выключении которых наблюдается желаемый фенотип [20]. Однако, вышеупомянутые эксперименты являются очень дорогостоящими, и, кроме того, каждый такой эксперимент может быть применен либо лишь к одному регулятору (как в случае пертурбации индивидуального гена), либо лишь к одному фенотипу (как в случае CRISPR-скрина). Таким образом, применение таких методов к изучению множества регуляторов, контролирующих множество метаболических путей, было бы очень трудоемким и дорогостоящим.

Вышеперечисленные методы измеряют взаимодействия между регуляторами и их мишенями косвенным образом: если при искусственном изменении уровня экспрессии одного гена меняется также уровень другого гена, исследователь предполагает, что эти гены функционально связаны, однако по результатам такого эксперимента невозможно отличить прямые взаимодействия между мишенью и регулятором от опосредованных. Другая группа методов фокусируется на измерении прямых регуляторных взаимодействий [21]. Транскрипционные факторы осуществляют свои регуляторные функции, специфически связывая определенные последовательности ДНК. Методы иммунопреципитации, совмещенные с секвенированием нового поколения, позволяют определить те последовательности ДНК, с которыми связывается данный белок [22]. Однако, таким методам присущ очень высокий уровень шума: ДНК-связывающие домены транскрипционных факторов заряжены положительно, а молекула ДНК заряжена отрицательно, и, как обсуждалось выше, молекула транскрипционного фактора находится в равновесии между специфичными и неспецифичными взаимодействиями. В результате при выделении фрагментов ДНК, связанных с изучаемым белком, неизбежно выделяются также и неспецифически связанные фрагменты. Кроме того, методы иммунопреципитации также трудоемки, дорогостоящи и могут быть применены только к индивидуальным регуляторам [23].

Со всеми перечисленными методами связана общая проблема: один эксперимент позволяет изучение лишь одного регулона, либо одного фенотипа. Более того, для проведения любого такого эксперимента необходимо поддержание изучаемого микроорганизма в культуре. Для культивации многих (например, анаэробных) микроорганизмов требуется дорогостоящее оборудование, а многие другие бактерии до сих пор не были культивированы в лабораторных условиях.

В качестве комплементарного набора методов, лишенных упомянутых проблем, используются методы вычислительной биологии и, в частности, сравнительной геномики. С развитием методов секвенирования нового поколения количество отсеквенированных бактериальных геномов растет экспоненциально. На текущий момент, в базе данных NCBI [24] доступно более 180 тысяч бактериальных геномов. Если каждый индивидуальный эксперимент может быть проведен лишь в одной бактерии, то методы сравнительной геномики были разработаны специально для того, чтобы извлекать биологическую информацию из сравнительного анализа последовательности ДНК многих родственных организмов. Ключевая идея таких методов состоит в том, что на последовательности, несущие жизненно важную функцию (в том числе регуляторную), действует положительный отбор, тогда так последовательности, не играющие прямой функциональной роли, быстрее накапливают мутации. Таким образом, функционально важные последовательности чаще бывают консервативны, т.е. похожи между родственными геномами, чем последовательности, не играющие прямой функции [25]. Сравнивая последовательности ДНК родственных геномов, можно находить как цис-, так и транс-активные последовательности. В последние 20 лет был разработан широкий набор методов, построенных вокруг этого соображения и позволяющих делать масштабные качественные предсказания о регуляции бактериальных метаболических путей (подробнее см. следующую секцию). Из-за огромного объема доступных биологических данных эти методы являются очень мощным инструментом и для многих организмов позволяют делать качественные предсказания с минимальными затратами ресурсов.

Проверка предсказаний

Нужно отметить, что все перечисленные до сих пор методы изучения регуляторных взаимодействий (как экспериментальные, так и вычислительные) в конечном итоге предоставляют лишь гипотезу о возможном механизме регуляции. Вне зависимости от того,

как данные гипотезы были получены, для подтверждения или опровержения каждой гипотезы приходится применять индивидуальные эксперименты. Характер экспериментов зависит от конкретной гипотезы, однако в общем случае две основные группы экспериментов тестируют гипотезу *in vitro* и *in vivo*. *In vitro* эксперименты тестируют, связываются ли предсказанные регулятор и мишень друг с другом и насколько их взаимодействие специфично. Например, типичным экспериментом для количественного измерения взаимодействия транскрипционного фактора и фрагмента ДНК является замедление ДНК в геле (gel shift assay). *In vivo* эксперименты для проверки предсказанных регуляторных взаимодействий устроены, как правило, по одной и той же схеме. Сперва, блокируется функциональная активность гена-регулятора. Для этого применяют либо техники блокирования экспрессии гена (при помощи CRISPR, малых интерферирующих РНК и других методов), либо, если этот ген жизненно необходим для организма, ген мутируют или экспрессируют его неполную версию. Далее, измеряют экспрессию генов-мишеней; если при изменении активности гена-регулятора не наблюдается предсказанных изменений в экспрессии генов-мишеней, гипотеза отвергается. Затем следует ключевой эксперимент: в систему искусственно добавляют экзогенный регулятор (т.е. молекулы регулятора экспрессируются не с первичного гена, находящегося в геноме организма, а с гена, привнесенного в клетку организма в составе плазмиды), и измеряют, возвращается ли экспрессия генов-мишеней на прежний уровень. Если добавление экзогенного регулятора восстанавливает экспрессию генов-мишеней, гипотеза о регуляторном взаимодействии считается подтвержденной (например, см. [26]). Поскольку в биологии, в отличие от математики, нет базовых аксиом, для подтверждения гипотез используются некоторые “золотые стандарты”. Описанный эксперимент является таким “золотым стандартом” в случае изучения транскрипционной регуляции.

Методы сравнительной геномики

Как упомянуто выше, три основные составляющие процесса предсказания регуляторных взаимодействий - это (1) поиск регулируемых мишеней, (2) поиск регуляторов и (3) изучение взаимодействий между ними. В следующих секциях мы подробнее остановимся на каждом из этих разделов.

Поиск регуляторов

Происхождение большинства генов в любом ныне существующем организме объясняется дубликацией ранее существующих генов с последующим расхождением функций двух образовавшихся копий [27]. Таким образом, гены образуют большие группы, каждая из которых произошла от единого гена-предшественника. Регуляторы экспрессии генов, в том числе и транскрипционные факторы, тоже образуют лишь малое число больших семейств, каждое из которых характеризуется одним или несколькими консервативными доменами. Наличие консервативных доменов, типичных для регуляторов экспрессии генов, упрощает поиск новых регуляторов. Разработанные в начале 2000-х компьютерные программы и базы данных позволяют найти гены потенциальных транскрипционных факторов, принадлежащих к известным семействам транскрипционных факторов, в любом новом изучаемом геноме, используя сходство последовательностей в качестве основного критерия [28]. Экспериментальные методы для поиска регуляторов из новых, неизученных семейств, также существуют, но мы не будем подробно на них останавливаться в данном обзоре. При изучении нового генома в рамках сравнительно-геномного подхода типичная последовательность выбора регуляторов для более детального анализа состоит из трех шагов: (1) выбрать группы метаболических путей для подробного изучения, (2) выбрать все семейства транскрипционных факторов, про которые известно, что хоть один регулятор-член семейства контролирует схожие метаболические процессы хоть в одном геноме, и (3) использовать существующие компьютерные программы и базы данных для поиска всех генов, содержащих характеристические домены выбранных семейств [29].

Поиск регулируемых мишеней

Гены бактерий организованы в опероны - участки ДНК, транскрибируемые в единую молекулу мРНК. Экспрессия набора генов, составляющих оперон, регулируется вместе, по крайней мере на этапе инициации транскрипции. Однако по одной лишь геномной последовательности не всегда можно точно установить, какие гены находятся в одном и том же опероне. Таким образом, при реконструкции регулона задача исследователя сводится к нахождению связывающих сайтов, регулируемых оперонов и соответствующих генов, а затем к сопоставлению регулятора и регулона (обсуждено ниже). Эти задачи не могут быть решены

отдельно друг от друга. Поскольку критерии для поиска связывающих сайтов легче формализовать и автоматизировать, поиск сайтов занимает ключевую роль в большинстве методологий. Задача поиска связывающих сайтов осложняется тем, что даже связывающие сайты одного и того же регулятора не идентичны другу - одни обладают суб-оптимальной аффинностью [11]. Поэтому формальное описание связывающих сайтов не тривиально. Большинство активно используемых методов поиска сайтов можно кратко описать тремя шагами: (1) найти начальную тренировочную группу сайтов, (2) описать ее формальной конструкцией, которая позволит численно описывать, насколько похожа та или иная последовательность на сайты из тренировочной выборки, и (3) использовать эту конструкцию для поиска остальных сайтов из того же регулона [29]. Ниже мы остановимся на каждом из этих трех шагов более подробно, а затем обсудим поиск регулируемых генов и оперонов.

Поиск сайтов для тренировочной выборки

Существует две распространенные схемы поиска сайтов для тренировочной выборки. Они различаются начальной гипотезой исследователя и последующей методологией. В первом случае, исследователь предполагает, что определенный ген или оперон регулируется на транскрипционном уровне и пытается найти цис-активные сайты связывания, осуществляющие эту регуляцию, используя консервативность разных позиций в регуляторной области гена в качестве основного критерия поиска. Эта схема называется филогенетическим футпринтингом. Во втором случае, исследователь предполагает, что несколько генов регулируются вместе, и пытается найти в их регуляторных областях общие последовательности. В качестве контрольной группы используют регуляторные области других генов, которые, по предположению исследователя, не регулируются по тому же механизму.

Филогенетический футпринтинг

Как обсуждалось выше, ключевая идея сравнительной геномики состоит в том, что функциональные последовательности более консервативны, чем не функциональные. Это наблюдение лежит в основе метода филогенетического футпринтинга [30]. Метод позволяет найти цис-активные последовательности в регуляторной области даже у одного гена, при

условии, что доступны последовательности ДНК ортологичных генов из родственных организмов (ортологичными называются гомологичные гены родственных организмов, появившиеся в процессе видообразования [31]). Если у гена есть ортологи хотя бы в нескольких родственных организмах, можно выбрать такой набор геномов, что на множественном выравнивании регуляторных областей изучаемого гена хорошее качество выравнивания будет достигнуто лишь на участке, соответствующем связывающему сайту, но не на других участках. Для выделения такого участка на множественном выравнивании можно использовать автоматизированные программы, например FootPrinter [32]. Однако начальный шаг этого метода, выбор родственных геномов для анализа, сложнее автоматизировать, потому что правильный выбор геномов зависит в том числе и от специфики изучаемой группы организмов. Это требует интуиции исследователя и многие исследователи предпочитают выбирать геномы вручную. Однако, существуют интерактивные веб-серверы, упрощающую процедуру филогенетического футпринтинга для исследователя [33].

Поиск сайтов в группе регуляторных областей

Метод филогенетического футпринтинга имеет свои ограничения. Во-первых, не во всех филогенетических группах бактерий доступно необходимое разнообразие отсеквенированных геномов. Во-вторых, регуляторные взаимодействия в целом менее консервативны, чем аминокислотные последовательности белков или метаболические пути, и, вследствие этого, не для всех генов удастся найти консервативные участки в их регуляторных областях. Чтобы обойти эти проблемы, используют комплементарный набор методов, основанный на поиске мотивов, представленных в одной выборке последовательностей, но отсутствующих в другой. Данная задача легко формализуема, и потому для ее решения разработано множество алгоритмов. Одними из наиболее популярных алгоритмов являются MEME и Gibbs Motif Sampler [34], [35].

Построение мотива связывающих сайтов

В отличие от рестриктаз, транскрипционные факторы не связывают строго определенные последовательности; напротив, каждый транскрипционный фактор может связывать широкий

спектр последовательностей с разной специфичностью. Это усложняет описание и поиск сайтов связывания транскрипционных факторов. Разработано много способов репрезентации сайтов связывания; их можно разделить по способу применения на две основные группы: репрезентации сайтов связывания для визуализации и репрезентации для поиска новых сайтов. Простейшая репрезентация мотива - это так называемая консенсусная последовательность, т.е. единая последовательность, каждый элемент которой отражает один или несколько наиболее часто встречающихся в данной позиции нуклеотидов. Так, для регулятора CsrA консенсус выглядит как WTGNNARCGNWWCAW (где W означает A или T, R – A или G и N – любое основание) [36]. Консенсусная последовательность обычно используется для визуализации мотива, но может быть использована и для поиска новых связывающих сайтов [37]. Для визуализации мотивов чаще всего используют программу WebLogo [38]. Эта программа рисует мотив по набору выровненных сайтов одинаковой длины. Каждая позиция выравнивания содержит нуклеотиды в виде букв, высота которых соответствует информационному содержанию данного нуклеотида в данной позиции, что косвенно отображает встречаемость этого нуклеотида, а также разнообразие нуклеотидов в данной позиции.

Один из простейших, и, в то же время, наиболее распространенных способов репрезентации мотива с целью поиска новых сайтов - это позиционная матрица весов (ПМВ). Мотив фиксированной длины N описывается матрицей размера N на 4, где строки соответствуют позициям в мотиве, столбцы - нуклеотидам, а в клетках стоят численные значения. С помощью такой матрицы можно численно оценить, насколько близка любая последовательность длины N к изучаемому мотиву. Оценка получается при последовательном сложении клеток, соответствующих нуклеотиду в определенной позиции в изучаемой последовательности. При построении такой матрицы учитывается не только частота встречаемости каждого нуклеотида на каждой позиции, но и консервативность каждой позиции. Методы, основанные на построении ПМВ, используются повсеместно, в том числе в таких популярных программах, как PSI-BLAST [23], MEME [24], Gibbs sampler [25] и т.д. Недостатком такого способа репрезентации мотива является то, что в данной модели все позиции рассматриваются независимо друг от друга, что противоречит эмпирическим наблюдениям о структуре ДНК-связывающих белков и их мотивов [39]. Разработано несколько алгоритмов, учитывающих меж-позиционные взаимодействия [39], [40], но они применяются намного реже, чем алгоритмы, основанные на ПМВ. Это связано с тем, что в большинстве случаев ПМВ описывает связывающие мотивы достаточно точно (т.е. при существенном усложнении модели

точность предсказаний слабо возрастает), и при этом выигрывает у более сложных алгоритмов в простоте и интерпретируемости [41] (для подробного обзора см. [42]).

Поиск новых связывающих сайтов

Описанные репрезентации мотивов связывающих сайтов, такие, как ПМВ, используются для полуавтоматического поиска связывающих сайтов. Для любой определенной ПМВ длины N , все нуклеотидные последовательности в геноме можно ранжировать по сходству с изучаемым мотивом. Само по себе ранжирование не является достаточной процедурой для нахождения новых связывающих сайтов. Многие программы предлагают автоматический выбор порогового значения для определения точного набора предсказанных сайтов; в этом случае конечными предсказаниями являются все сайты связывания, находящиеся выше порогового значения в списке. Однако, выбор предсказаний только на основании сходства с ПМВ приводит к большому количеству ложных предсказаний, поэтому выбор предсказаний строится на основании дополнительных критериев. Существуют различные подходы к выбору дополнительных критериев; мы остановимся на методах, применяемых в контексте сравнительно-геномного подхода. Ниже, мы перечисляем несколько основных критериев, применяемых для того, чтобы лучше отличить настоящие функциональные сайты связывания от ложно-положительных предсказаний [43].

Во-первых, это консервативность: настоящим связывающим сайтом окажется скорее тот сайт, что встречается в регуляторной области ортологичных генов во многих родственных организмах, чем тот, что встречается в регуляторной области лишь у одного гена, но не у его ортологов.

Во-вторых, это сходство функций регулируемых генов. Гены одного метаболического пути часто регулируются вместе, и при поиске сайтов с помощью ПМВ многие сайты из верхней части ранжированного списка регулируют гены, участвующие в одном или в нескольких близких метаболических путях. В соответствии с этим критерием, среди более слабых сайтов предпочтение будет отдаваться тем, что контролируют гены из близких метаболических путей.

В-третьих, это со-встречаемость гена регулятора и найденных сайтов в геномах. Мы обсуждаем методы сопоставления связывающих сайтов и регуляторов ниже; данный критерий применим в тех случаях, когда для нескольких сайтов из начала ранжированного списка уже найден соответствующий регулятор, и задача состоит в том, чтобы выбрать настоящие

связывающие сайты среди более “слабых” (т.е. менее похожих на используемую ПМВ) сайтов. Как правило, при выборе достаточно удаленных друг от друга (филогенетически) родственных геномов для анализа, некоторые геномы не содержат гена регулятора. Если в этих геномах были найдены возможные связывающие сайты, эти сайты не могут быть функциональными из-за отсутствия регулятора, и такие сайты исключаются из списка предсказаний.

В-четвертых, это множественные сайты. Многие транскрипционные факторы связывают регуляторные области генов в форме димеров или даже тетрамеров. В свою очередь, регуляторные области многих генов содержат несколько сайтов связывания для одного и того же транскрипционного фактора. Эти сайты отличаются друг от друга: часто, один из сайтов имеет высокое сродство к регулятору, тогда как остальные сайты не достаточно специфичны к регулятору сами по себе, но из-за своего расположения способствуют кооперативному связыванию ди- или мультимеров регулятора к молекуле ДНК [44]. При попытке различить ложно-положительные предсказания и настоящие сайты среди “слабых” сайтов связывания, предпочтение будет отдано такому сайту, который находится рядом с более сильным сайтом связывания, но не такому сайту, который является единственным возможным сайтом в данной регуляторной области.

Каждый из этих критериев по отдельности легко автоматизировать, однако основная мощь сравнительно-геномного подхода исходит из возможности гибко совмещать эти критерии, приоритизируя их в зависимости от индивидуального случая. Такое гибкое манипулирование многими критериями и источниками информации требует глубокого понимания биологии бактерий, и поэтому полностью автоматизированных программ, которые могли бы делать это так же качественно, как исследователи, на данный момент не существует.

Вышеописанные методы позволяют найти новые связывающие сайты на основе начальной (тренировочной) выборки сайтов. Такой подход можно применять интерактивно, раз за разом расширяя тренировочную выборку сайтов. Однако, большинство бактериальных регулонов охватывают лишь малое число генов и оперонов, поэтому подавляющее большинство сайтов можно найти за малое количество итераций.

Установление соответствий между мишенями и регуляторами

Для большинства транскрипционных факторов характерна авторегуляция - механизм обратной связи, при котором белок, кодируемый геном регулятора, специфично связывается с цис-активным сайтом в регуляторной области собственного гена, и таким образом контролирует свою экспрессию [45]. Таким образом, регуляторные области генов транскрипционных факторов содержат свои же собственные связывающие сайты. Это позволяет легко сопоставить регулятор и мотив связывающих сайтов, таким образом связав регулятор с регулоном. Другое общее наблюдение состоит в том, что большинство транскрипционных факторов в бактериях являются “локальными” регуляторами, т.е контролируют экспрессию лишь нескольких генов и оперонов, участвующих лишь в одном-двух метаболических путях, и, что самое главное в данном контексте, ген регулятора бывает расположен близко к генам-мишеням на хромосоме [43]. С практической точки зрения при поиске регулятора, контролирующего изучаемый регулон, это наблюдение позволяет сузить список кандидатов.

Однако, оба эти метода не всегда бывают точны: существуют ситуации, в которых, руководствуясь одним лишь этим методом, исследователь может ошибиться в выборе кандидатов-регуляторов. Например, в бактериальных геномах распространены регуляторные каскады, где один транскрипционный фактор контролирует экспрессию другого [46]. В таком случае в регуляторной области гена транскрипционного фактора может находиться сайт связывания другого регулятора, что может приводить к ошибочным выводам о соответствии связывающих сайтов и регуляторов. Чтобы избежать таких ошибок, используют дополнительные критерии соответствия регулятора и связывающих сайтов. Один из главных критериев - соответствуют ли характеристики связывающего сайта ожидаемым. Поскольку каждое семейство транскрипционных факторов – это, предположительно, группа генов, произошедших от одного и того же предкового гена [47], некоторые характеристики связывающих сайтов сохраняются для всех или почти для всех регуляторов из этого семейства. Например, в случае семейства LacI общие характеристики связывающих сайтов - это (1) палиндромность и (2) динуклеотид CG в середине сайта [48]. Распространенный метод для проверки соответствия регулон - регулятор - это разбить все известные характеристики на две группы, и использовать одну группу для поиска сайтов (например - палиндромность) и другую - для подтверждения соответствия (например - консервативные позиции сайта).

Реконструкция метаболических путей

Другая задача, неразрывно связанная с реконструкцией регуляторных взаимодействий - это реконструкция метаболических путей бактерий. Детализированные экспериментальные исследования метаболических путей были проведены только в небольшом числе модельных организмов; для подавляющего большинства бактерий метаболические пути либо предсказаны *in silico*, либо не изучены вообще. При этом, как и в случае регуляторных взаимодействий, во многих группах бактерий наблюдаются консервативные филогенетические паттерны метаболических путей, что позволяет использовать сравнительно-геномные методы для построения сложных и качественных моделей метаболизма бактерий. Вкупе с экспериментальными методами проверки гипотез это мощный инструмент для изучения биологии и биохимии бактерий.

При реконструкции метаболического пути цель исследователя - построить модель метаболического пути, включающую (1) последовательность ферментов, катализирующих метаболические реакции, и кодирующих их генов, и (2) предсказания химической структуры начального и конечного химических соединений, а также интермедиатов метаболического пути. Как и для любых других исследований *in silico*, такая модель является лишь предсказанием и потому требует экспериментального подтверждения. Однако, качественные модели все равно могут быть полезны и информативны.

В зависимости от конкретной решаемой биологической задачи одни аспекты получаемых моделей могут быть важнее, чем другие, поэтому для различных задач применяются разные техники сравнительного анализа бактериальных геномов с учетом известной биологической информации. Эти техники основаны на том наблюдении, что гены ферментов из одного метаболического пути стремятся (1) со-существовать (т.е присутствовать либо отсутствовать) вместе в геномах бактерий, (2) быть регулируемыми в составе одного и того же регулона, и (3) находиться близко друг от друга на бактериальной хромосоме. В качестве примера приведем один из наиболее успешных массовых проектов по реконструкции метаболических путей, целью которого была реконструкция путей в 1000 бактериальных геномах [49]. В этом исследовании был применен подход подсистем. Реконструкция метаболических путей в рамках этого подхода состоит из следующих шагов. Сперва на базе доступной биологической информации для каждого исследуемого метаболического пути строится первоначальная модель. Такая модель (также называемая подсистемой) состоит из набора ферментативных

ролей: каждая роль - это катализ определенной реакции, но фермент, выполняющий эту реакцию, не всегда известен. Далее, анализируя доступные отсеквенированные геномы, исследователь пытается найти гены, отвечающие за выполнение ферментативных ролей в изучаемом метаболическом пути. Чаще всего, такие гены могут быть найдены по гомологии с уже известными генами из других организмов. Т.к гены из одного метаболического пути имеют тенденцию либо присутствовать, либо отсутствовать одновременно друг с другом в бактериальных геномах, в большинстве геномов удается найти гены-кандидаты либо на все регуляторные роли (если данный путь присутствует в этом геноме), либо ни для одной из них [50], [51]. Однако, в некоторых геномах поиск генов по гомологии позволяет реконструировать путь лишь частично (т.е не для всех регуляторных ролей). Причинами могут быть неортологичные замещения ферментов в метаболических путях, неточная первоначальная модель метаболического пути, или другие факторы. В таких случаях для нахождения пропущенных генов используются такие методы, как анализ геномного контекста, анализ ко-регуляции или анализ биологической информации, специфической для этого таксона [43]. Такие несоответствия первоначальной модели и наблюдаемой биологической информации являются хорошей возможностью для уточнения и расширения существующей модели метаболических путей. Применяя последовательные итерации нахождения генов, отвечающих за изучаемые функциональные роли, и уточнения существующей модели, можно разработать сложные детализированные модели метаболических путей в изучаемых геномах.

Биология бифидобактерий

Бифидобактерии как пробиотики

Пробиотики – это живые микроорганизмы, которые при наличии (в умеренных количествах) в кишечном микробиоме человека улучшают здоровье человека-носителя [52]. История термина восходит к 1908 году, когда Мечников описал позитивные эффекты некоторых ферментированных продуктов на здоровье [53]. С тех пор, множество исследований подтвердили это наблюдение. Среди наблюдаемых позитивных эффектов, вызванных штаммами пробиотиков, стоит упомянуть: снижение аллергических симптомов, защита от некоторых инфекционных заболеваний, модуляция иммунной системы, снижение уровня холестерина, снижение нетолерантности к лактозе, ослабление воспалений кишечника и

производство витаминов и ацетата [54], [55], [56], [57]. Самыми распространенными среди коммерчески доступных пробиотиков являются бифидобактерии [58] - анаэробные сахаролитические, неспорулирующие грамм-положительные бактерии из типа *Actinobacteria* [59], [60]. Большинство бактерий из рода *Bifidobacterium* проживают в кишечнике животных (насекомых, птиц и млекопитающих, в т.ч. человека). Некоторые представители этого рода, например, *B. dentium* и *B. catenulatum*, крайне устойчивы и могут проживать в любой из многих экологических ниш [61]. Бифидобактерии являются одними из главных обитателей кишечника у детей, однако их число снижается при взрослении человека [62], [63]. Известно несколько случаев бактериемии, вызванной бифидобактериями [64], [65], однако каждый из этих случаев касался пациента с историей предшествующих заболеваний, например недоношенных детей или пожилых пациентов с диабетом. Таким образом, при наличии медицинских осложнений следует использовать бифидобактерии в качестве пробиотиков с осторожностью.

Углеводный метаболизм бифидобактерий

Из-за показанных положительных эффектов на человеческое здоровье, некоторые штаммы *Bifidobacterium animalis subsp. lactis* и *Bifidobacterium longum* продаются в качестве пищевых добавок [55], [66]. Типичный способ культивации бифидобактерий в кишечном микробиоме – это использование специальных пищевых полисахаридов, или пребиотиков, которые могут быть переварены бифидобактериями, но не могут быть переварены многими другими организмами, населяющими кишечник [67], [68], [58], [69], [70]. Бифидобактерии – это сахаролитические организмы; для них, успех в борьбе за ресурсы зависит от их способности переваривать многие полисахариды, происходящие как из клеток организма хозяина, так и из его питания. Распространенность генов, которые могут участвовать в метаболизме углеводов, в геномах бифидобактерий отражает адаптацию к кишечному тракту как к среде обитания [71], [72], [73]. Для поддержания работы такого большого количества катаболических путей необходима сложная регуляторная сеть, координирующая экспрессию генов в ответ на появления субстратов в окружающей среде. Пути утилизации углеводов в бактериях часто контролируются ДНК-связывающими ТФ (см. выше). Несмотря на недавно опубликованные работы, описывающие отдельные углеводные метаболические пути и регулоны в бифидобактериях [74], [75], [76], [60], [77], [78], [79], [80], транскрипционные регулоны большинства катаболических путей и генов до сих пор не изучены. Таким образом, изучение

путей катаболизма углеводов и их регуляции у бифидобактерий может привести к разработке новых диет, улучшающих человеческое здоровье.

Происхождение семейства транскрипционных факторов LacI

Название семейства LacI произошло от лактозного оперона LacI у *E. coli*, который является модельным объектом для изучения регуляции транскрипции у бактерий со времен классической статьи Жакоба и Моно [81]. Семейство было определено посредством анализа сходства белковых последовательностей; проведенный анализ также обнаружил сходства мотивов сайтов связывания ДНК [82]. В то же время, было отмечено, что ДНК-связывающие домены ТФ из семейства похожи на домены типа спираль-поворот-спираль (НТН), распространенные среди других ТФ [83], тогда как эффектор-связывающие домены ТФ из семейства LacI гомологичны периплазматическим белкам ABC-транспортеров [84], [47]. Любопытно, что этот домен также часто встречается вместе с другим ДНК-связывающим доменом, а именно – НТН -доменом из семейства GntR [85]. Если результаты ранних исследований, основанных на небольшом количестве белков, показывали, что на ранних этапах эволюции семейства произошло значительно больше дупликаций, чем на более поздних этапах [83], то более новые исследования, основанные на более полных данных, показали, что дупликации случались на протяжении всей истории семейства [47]. Из-за большого размера семейства и из-за высокого структурного сходства между его представителями, семейство LacI часто используется как модель для разработки новых алгоритмов для нахождения функционально важных аминокислотных остатков (как консервативных, так и тех, которые определяют специфичность связывания) [47], [86], [87], [88], [89], [90]. Детальное изучение большого количества ТФ из семейства LacI и их регулонов позволит одновременно получить более полную картину эволюции семейства и получить высококачественную модель для разработки новых методов изучения транскрипционных факторов.

Роль метаболизма витаминов группы В в формировании структуры микробиома

Кишечный микробиом человека представляет собой сложное сообщество микробных организмов. Вместе, они содержат огромное количество генов, организованных в сложные

метаболические и транскрипционные сети [91], [92]. Кишечный микробиом синтезирует множество необходимых метаболитов, в т.ч короткоцепочечные жирные кислоты и витамины группы В, которые затем всасываются эпителиальными клетками организма-хозяина. В последнее десятилетие было показано много примеров взаимосвязи между здоровьем человека и составом его кишечного микробиома [93], [94], [95], [96]. Среди факторов, влияющих на состав микробиома, самым важным считается питание [97], [98], [99]. Однако, метаболические потребности большинства представителей человеческого генома плохо изучены; их детальное изучение поможет нам лучше предсказывать последствия дисбактериоза и разрабатывать новые методы терапевтической модуляции кишечной микробиоты в медицинских целях.

Несколько витаминов группы В, в т.ч В1 (тиамин), В2 (рибофлавин), В3 (ниацин), В5 (пантотенат), В9 (фолат) and В6 (пиридоксин), являются метаболическими предшественниками необходимых кофакторов ТПФ, НАД, ФМН, НАДФ, КоА, ТГФ и ПФ, соответственно [100]. Витамин В7 служит кофактором для биотинкарбоксилаз и биотиндекарбоксилаз, участвующих в биосинтезе жирных кислот, глюконеогенезе и ферментации [101]. Витамин В12 (кобаламин) – это предшественник семейства коферментов В12, содержащего цианокобаламин, метилкобаламин и аденозилкобаламин; он необходим для всех животных, но не для всех бактерий. Доступность различных витаминов группы В может влиять на состав бактериальных сообществ [102]. С другой стороны, недавние исследования самоподдерживающихся бактериальных сообществ с постоянным видовым составом подчеркивают важность синтрофического метаболизма витаминов и потенциальных путей обмена метаболитами между различными организмами [103]. Однако, способности бактерий, населяющих кишечный микробиом, к синтезу и обмену витаминами и их предшественниками на данный момент плохо изучены. Детальное изучение метаболических возможностей представителей человеческого микробиома позволит лучше понимать и предсказывать структуру и динамику всего сообщества как в нормальных, так и в патогенных условиях.

Глава 2. Изучение транскрипционной регуляции путей утилизации углеводов в бактериях рода *Bifidobacterium*

Методы

Для проведения сравнительного анализа мы выбрали 10 полностью секвенированных геномов бактерий из рода *Bifidobacterium*, исключив штаммы, которые состоят в слишком близком родстве друг с другом (Таблица 1). Геномные последовательности выбранных штаммов были получены из базы данных Genbank [104]. Мы использовали программу поиска белковых доменов, доступную на портале Pfam [105], для поиска начального набора ТФ, которые могли бы контролировать экспрессию генов сахарного метаболизма. Шесть семейств ТФ содержат, среди прочего, регуляторы сахарного метаболизма; мы выбрали все гены, содержащие домены любого из этих семейств, в изучаемых геномах (LacI: PF00356 и PF00532; ROK: PF00480; DeoR: PF00455; RpiR: PF01418 и PF01380; SorC: PF04198; GntR: PF00392; AraC: PF02311). Семейства AraC and GntR содержат множество ТФ, специфичных к разнообразным лигандам, и контролирующим разнообразные метаболические пути. Поэтому среди регуляторов этого семейства мы выбрали только те, которые расположены по соседству с предсказанными генами сахарного метаболизма на хромосоме (Таблица 1). Позже мы также включили в анализ некоторые регуляторы семейства TetR (для которого нет известных примеров регуляции сахарного метаболизма), которые ко-локализовались с генами утилизации глюкозидов и галактозидов (Таблица 1). Мы построили филогенетические деревья для каждого семейства ТФ при помощи PhyML [106]. Далее мы определили ортологичные группы ТФ. Мы использовали три основных критерия для определения ортологичных групп: мы считали группу ТФ ортологичной в том случае, если ТФ из этой группы: (1) формируют моно- или пара- филогенетическую группу на филогенетическом дереве; (2) находятся в консервативном геномном контексте и (3) связывают похожие ДНК мотивы. Мы идентифицировали связывающие сайты ТФ и реконструировали регулоны ТФ при помощи сравнительно-

геномного метода, доступного на веб-сервере RegPredict [107]. Вкратце, метод состоит из: (1) нахождения маленькой группы предполагаемых регулируемых генов, (2) нахождения связывающих сайтов ТФ в регуляторных областях этих генов, (3) создания нуклеотидной позиционной матрицы весов для найденных связывающих сайтов, и (4) поиска дополнительных членов регулона по всему геному при помощи сконструированной матрицы. ТФ зачастую контролируют гены, расположенные недалеко от своих собственных генов (кодирующих регулятор), поэтому мы определяли первичный набор потенциально регулируемых генов для данного ТФ при помощи анализа геномного контекста гена изучаемого ТФ и его ортологов. Для этого мы использовали геномные браузеры MicrobesOnline [108] и IMG [109]. Мы искали возможные сайты связывания в регуляторных областях первичного набора потенциально регулируемых генов при помощи программы Discover Profile, доступной на сервере RegPredict [107]. Мы независимо проверяли каждый найденный мотив при помощи метода филогенетического футпринтинга [110]. Для множественных выравниваний последовательностей ДНК мы использовали программу ClustalW2 [111]. Для каждого найденного ДНК-связывающего мотива, мы составили нуклеотидную позиционную матрицу весов. Далее, мы использовали ее для поиска остальных членов регулона в геномах, содержащих ортологов изучаемого ТФ, сканируя регуляторные области генов (от -350 до +50 нуклеотидов относительно

предполагаемого сайта начала транскрипции) программой Run Profile, доступной на сервере RegPredict. Для поиска сайтов в геноме *B. breve* UCC2003, недоступном на сервере, мы использовали программу Genome Explorer [112], которая использует схожий алгоритм и параметры. Обе программы присваивают каждому найденному сайту определенный балл; мы рассматривали лишь те сайты, балл которых был не менее чем на 10% ниже, чем у сайта с наименьшим баллом из первоначальной выборки. Мы также рассматривали некоторые из более слабых сайтов, но только при условии, что (1) их расположение было таким же, как и у более сильных сайтов (с баллом выше порогового значения) из других геномов, и (2) регуляторная область, в которой они располагались, не содержала других, более сильных, сайтов связывания. Чтобы минимизировать количество ложно-положительных предсказаний, мы рассматривали ген как новый член регулона только в том случае, если регуляторные области ортологов этого гена содержат сайты связывания ТФ по крайней мере в нескольких родственных геномах. В тех случаях, когда новый ген, включаемый в регулон, являлся частью оперона, мы также включали в регулон все остальные гены из этого оперона. Мы считали гены

членами одного оперона, если они (1) транскрибируются в одном и том же направлении, (2) разделены межгенными участками длиной 200 или меньше нуклеотидов, и (3) расположены в одном и том же порядке в нескольких геномах. Мы рисовали лого ДНК-связывающих мотивов при помощи программы WebLogo (Crooks et al., 2004). Вся информация о реконструированных регулонах, включая ТФ, связывающие их сайты ДНК и регулируемые гены (для всех геномов, кроме *B. breve* UCC2003) была загружена в общедоступную базу данных RegPrecise [48].

Чтобы реконструировать регулируемые метаболические пути и функционально аннотировать регулируемые гены, мы совмещали использование сравнительно-геномных методов [113] с поиском известных функциональных описаний из литературы, а также баз данных SEED [114] и UniProt [115]. Кроме того, ортологи экспериментально изученных генов из близкородственных геномов были найдены при помощи программы BLAST [116]. Наличие данных о функции хотя бы одного члена регулона может сделать возможным предсказание специфичности всего регулируемого пути. Анализ белковых доменов (Pfam) позволил нам предположить общий класс выполняемых биохимических реакций для многих ферментов с неизученной специфичностью. Например, одним из важных в метаболизме сахаров классов ферментов являются гликозил гидролазы; они обычно выполняют первый шаг катаболических путей. Для их классификации и аннотации мы использовали базу данных CAZy database [117]. Мы основывали предсказания эффекторов ТФ на доступных экспериментальных данных (в т.ч. на опубликованных транскриптомных данных) и на анализе функционального содержания реконструированных регулонов.

Результаты

Репертуар транскрипционных факторов, участвующих в регуляции сахарного метаболизма, в геномах рода *Bifidobacterium*

Чтобы оценить число и разнообразие транскрипционных регуляторов путей утилизации углеводов в геномах бифидобактерий, мы идентифицировали все гены, кодирующие транскрипционные факторы из семейств LacI, ROK, DeoR, RpiR, SorC, AraC, BglG и GntR, и проанализировали их геномный контекст в 10 бактериях рода *Bifidobacterium* (Таблица 1). Из предыдущих исследований было известно, что транскрипционные факторы, принадлежащие этим семействам, часто вовлечены в регуляцию сахарных метаболических путей [113], [118],

Таблица 1. Распределение ТФ, контролирующих метаболизм углеводов у бифидобактерий

Геном	Семейства транскрипционных факторов								Всего ТФ	ГУМ ² , %
	LacI	ROK	RpiR	DeoR/ SorC	BglG	TetR ¹	AraC ¹	GntR ¹		
<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

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

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

[118], [119], [120]. Согласно базе данных RegPrecise [48], большинство регуляторов из семейств LacI, ROK, BglG, DeoR, SorC и RpiR вовлечены в регуляцию метаболизма углеводов. Однако, транскрипционные факторы из семейств GntR и AraC могут контролировать самые разные метаболические пути, в том числе и пути утилизации сахаров. Из-за этого мы выбрали лишь те транскрипционные факторы из семейств GntR и AraC, гены которых находились рядом с другими генами, предположительно участвующими в сахарном метаболизме. Транскрипционные факторы из семейства TetR регулируют устойчивость к антибиотикам, клеточную сигнализацию и метаболизм различных веществ [121]. Мы обнаружили новую группу транскрипционных факторов из семейства TetR, которые могут контролировать пути утилизации сахаров; поэтому мы также проанализировали распределение генов регуляторов из данного семейства в геномах бифидобактерий и выбрали те регуляторы, чьи гены соседствовали с генами ферментов, участвующих в метаболизме сахаров.

В результате, мы идентифицировали 308 транскрипционных факторов, потенциально регулирующих сахарный метаболизм (Таблица 1). 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 встречаются в геномах бифидобактерий значительно реже. Семейство LacI вмещает 63% идентифицированных регуляторов сахарного метаболизма (от 7 до 34 транскрипционных факторов на геном); 16% и 7% регуляторов принадлежат к семействам ROK и TetR,

соответственно, тогда как оставшиеся регуляторы принадлежат к семействам AraC, BglG, DeoR, GntR, RpiR и SorC.

Анализируя филогенетические деревья и геномный контекст идентифицированных регуляторов для каждого семейства по отдельности, мы выделили ортологичные группы транскрипционных факторов (Рис. 1 и Таблица 2). В результате, мы смогли выделить 34 ортологичные группы, состоящие из двух и более белков, и 22 одиночных регулятора среди 194 изучаемых регуляторов, принадлежащих к семейству LacI. Для семейства ROK, 42 белка

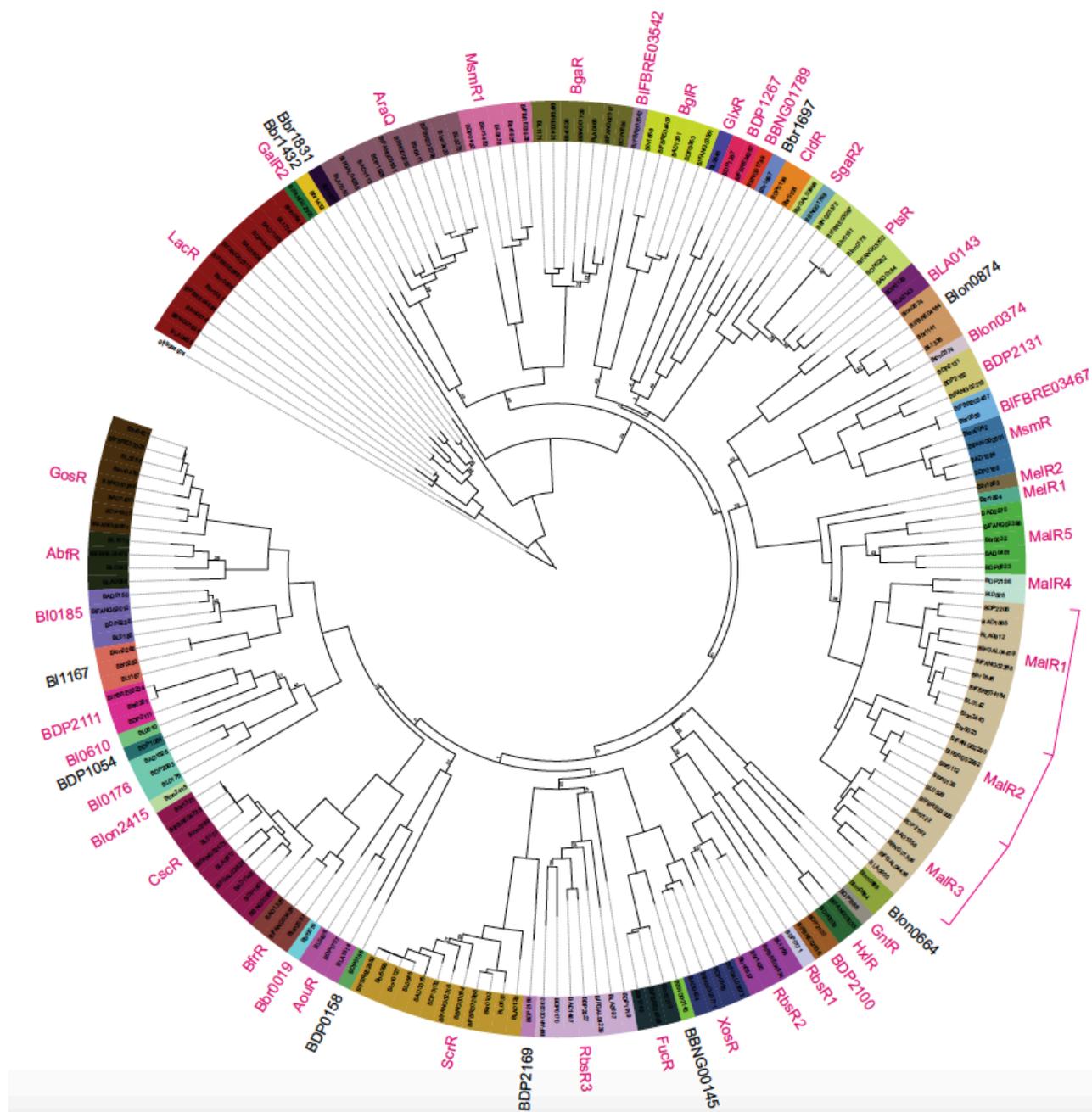


Рис. 1А Филогенетическое дерево ТФ из семейства *LacI*, предположительно участвующих в контроле генов утилизации углеводного метаболизма у бифидобактерий. ТФ из одной и той же ортологичной группы показаны одинаковым фоном. Названия ортологичных групп показаны наружным кругом. Группы ТФ, регулоны которых нам удалось реконструировать, подсвечены красным.

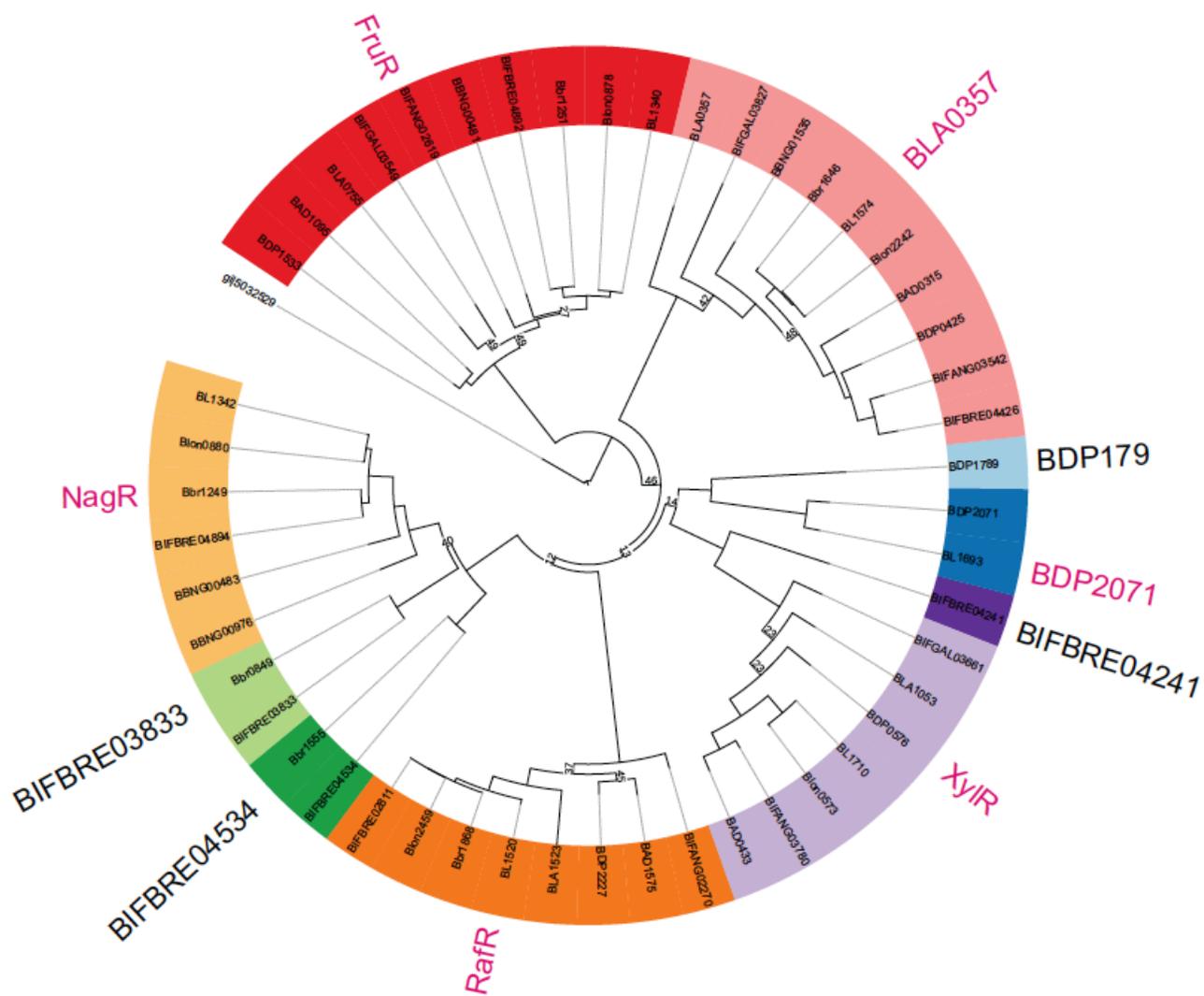


Рис. 1Б Филогенетическое дерево ТФ из семейства *ROK*, предположительно участвующих в контроле генов утилизации углеводного метаболизма у бифидобактерий. ТФ из одной и той же ортологичной группы показаны одинаковым фоном. Названия ортологичных групп показаны наружным кругом. Группы ТФ, регулоны которых нам удалось реконструировать, подсвечены красным

составляют 8 ортологичных групп, тогда как остальные белки являются одиночными регуляторами. 23 регулятора сахарного метаболизма из семейства TetR были поделены на 6 ортологичных групп и 2 одиночных регулятора.

В целом, ортологи идентифицированных регуляторов сахарного метаболизма неравномерно распределены среди 10 проанализированных геномов. К консервативным регуляторам, представленным во всех 10 геномах, относятся регулятор GlcR из семейства DeoR, регуляторы CscR, MalR и AgaQ из семейства LacI, и регуляторы NagR и BLA_0357 из семейства ROK.

Геномная реконструкция регулонов транскрипционных факторов

Для анализа транскрипционных регулонов идентифицированных транскрипционных факторов, мы использовали методы сравнительной геномики (см. Методы выше). Мы смогли найти мотивы сайтов связывания ТФ и реконструировать регулоны для 48 из 55 ортологичных групп и для 16 из 33 одиночных регуляторов. Большинство транскрипционных факторов, чьи регулоны были успешно реконструированы, мозаично распределены среди изучаемых геномов. 6 транскрипционных факторов, включая потенциальные регуляторы путей утилизации сахарозы, мальтозы, фруктозы, галактозы, а также регулятор центрального углеводного метаболизма AgaQ, были идентифицированы во всех изучаемых геномах бифидобактерий. Детализированная информация об ортологичных транскрипционных факторах, их сайтах связывания и регулируемых генах/оперонах для каждого реконструированного регулона доступна в базе данных RegPrecise. Всего, реконструированные регулоны содержат 268 транскрипционных факторов, 755 сайтов связывания и более 600 регулируемых оперонов, содержащих более 1000 генов.

8 регуляторов, чьи регулоны были успешно реконструированы, были ранее экспериментально охарактеризованы в *B. breve UCC2003* (Таблица 2). К ним относятся регулятор утилизации целлодекстрина (CldR), галактана (GalR, был переименован в GosR для того, чтобы избежать путаницы с регулятором метаболизма галактозы из семейства DeoR), мелезитозы и ее производных (MelR1, MelR2), раффинозы (RafR), сиаловой кислоты (NanR), сахарозы (ScrR) и рибозы (RbsR). Информация, полученная из реконструкции регулонов (в т.ч регулируемые гены, их предсказанные функции и сайты связывания транскрипционных факторов) хорошо согласуется с опубликованными экспериментальными данными [74], [76], [60], [77], [78], [80].

Таблица 2. Состав ортологичных групп транскрипционных факторов, регулирующих метаболизм углеводов в 10 геномах бифидобактерий

Регулон ¹	Изучаемые геномы <i>Bifidobacteria</i> ²										Метаболический путь ³
	1	2	3	4	5	6	7	8	9	10	
AbfR			+		+				++		Арабинозные олигосахариды
AouR			+				+		+		Арабинозные олигосахариды
BBNG_01789				+							?
Bbr_0019						+					?
BDP_1267					+		+				?
BDP_2071							+		+		?
BDP_2100					+		+				?
BDP_2111					+	+	+				?
BDP_2131		+						++			?
BfrR	+	+								+	Фруктоолигосахариды
BgaR		+	+	+	+	+	+		+		Бета-галактозиды
BglR	+	+			+	+	+				Бета-глюкозиды
BgrT1	+	+			+		+			+	Бета-глюкозиды
BgrT2	+	+			+	+	+		+		Бета-глюкозиды
BgrT3	+		+				+				Бета-галактозиды
BgrT4	+						++				Бета-глюкозиды
BgrT5	+						+				Бета-глюкозиды
BgrT6	+										Бета-галактозиды
BgrT7							+				Бета-галактозиды
BgrT8			+			+					Бета-глюкозиды
BIFBRE_03467					+	+					?
BIFBRE_03542					+						?
BL0176	+						+		+		?
BL0185	+	+					+		+		?
BL0610									+		?
BLA_0143			+				+				?
BLA_0357	+	+	+	+	+	+	+	+	+	+	?
Blon_0374										+	?
Blon_2415										+	?
CldR*							+	+			<i>Целлодекстрин</i>
CscR	+	+	+	+	+	+	+	+	+	+	Сахароза
FruR	+	+	+	+	+	+	+	+	+	+	Фруктоза
FucR					+	+				+	Фукоза
GalR	+	+	+	+	+	+	+	+	+	+	<i>Галактоза</i>
GalR2		+									Галактоза; лактоза
GlcR	+	+		+	+	+	+		+	+	Глюкоза
GlxR									+		Глицерат
GntR							+				Глюконат
GosR (or GalR)*	+	+		+	+	+	+		+	+	<i>Галактан</i>
HxlR		+					+				Гекеулоза

LacR	++	+	+	+	++	++	+	+	++	Лактоза
MalR/MalR2/MalR3	++	++	++	+	+++	++++	++	++	++	Мальтоза; мальтодекстрин
MalR4							+		+	Мальтоза; мальтодекстрин
MalR5	++	+					+		+	Мальтоза; мальтодекстрин
MelR1*							+			<i>Мелезитоза</i>
MelR2*							+			?
MsmR	+	+							+	?
MsmR1					+	+	+		+	Альфа-галактозиды
NagR				++	+	+			+	<i>Галакто-N-биоза, лакто-N-биоза</i>
NanR*					+	+			+	<i>Сиаловые кислоты</i>
PtsR	+	+		+	+	+	+		+	Компоненты фосфотрансферазных систем
RafR*	+	+	+		+	+	+		+	<i>Раффиноза</i>
RbsR							+			Рибоза
RbsR2 (or RbsR)*					+	+			+	<i>Рибоза; (рибонуклеозиды)</i>
RbsR3	+	+	+				+++	+		(Рибонуклеозиды)
RbtR	+				+					Рибитол; ксилитол
ScrR*	+	+	+	+	++	++	+		++	<i>Сахароза</i>
SgaR				+						L-ксилулоза
SgaR2				+						Аскорбат
XosR	+	+					+	+		Ксилоолигосахариды
XylR	+	+	+				+	+	+	Ксилоза
AraQ	+	+	+	+	+	+	+	+	+	Центральный углеводный метаболизм; арабиноза

¹ Экспериментально изученные ТФ отмечены звездочками; альтернативные названия ТФ, используемые в других исследованиях, показаны в скобках.

² Пронумерованные столбцы соответствуют изучаемым геномам *Bifidobacteria*: 1, *B. adolescentis* ATCC 15703; 2, *B. angulatum* DSM 20098; 3, *B. animalis* subsp. *lactis* AD011; 4, *B. bifidum* NCIMB 41171; 5, *B. breve* DSM 20213; 6, *B. breve* UCC2003; 7, *B. dentium* Bd1; 8, *B. gallicum* DSM 20093; 9, *B. longum* NCC2705; 10, *B. longum* subsp. *infantis* ATCC 15697. Клетки, отмеченные несколькими знаками '+' показывают наличие нескольких паралогичных транскрипционных факторов.

³ Перечислены предсказанные субстраты регулируемых метаболических путей. В отдельных случаях более широкая группа регулируемых метаболических путей, не специфичных к данному субстрату, указана в скобках. Ранее охарактеризованные пути катаболизма углеводов отмечены курсивом.

Анализ функций реконструированных регулонов

Чтобы предсказать биологические функции и молекулы-эффекторы изучаемых транскрипционных факторов, мы анализировали функции регулируемых генов для каждого регулона. Мы провели метаболическую реконструкцию соответствующих биохимических путей и предсказали функции ко-регулируемых генов, анализируя консервативные генетические кластеры из геномных баз данных. Это позволило нам предсказать метаболические пути и биологические функции для 46 из 64 ортологичных групп и одиночных регуляторов (Таблица 2). Для остальных 18 групп транскрипционных факторов нам не удалось обнаружить специфических путей утилизации сахаров. Большинство изученных

транскрипционных факторов контролируют локальные регулоны; такие регулоны содержат 1-5 сайтов связывания, расположенных в промотерных областях оперонов, соседствующих с геном транскрипционного фактора (см. в базе данных RegPrecise). Гены таких локальных регулонов, как правило, кодируют ферменты специфических метаболических путей (Рис. 2).

Несмотря на то, что биохимические реакции, составляющие пути катаболизма различных сахаров крайне разнообразны, большинство из них осуществляется при помощи лишь небольшой группы ферментативных активностей, включая оксидоредуктазы, киназы, альдолазы, гидролазы и изомеразы. 35 реконструированных регулонов содержат в своем составе гликозил гидролазы из 20 или более семейств (согласно базе данных CAZy). На этом основании можно предположить, что эти регулоны принимают участие в контроле катаболизма ди-, олиго- и полисахаридов. Мы предсказываем, что проанализированные гликозил гидролазы из регулонов сахарного метаболизма режут гликозидные связи в некоторых дисахаридах (сахароза, лактоза, целлобиоза), разнообразных олигосахаридах (бета-галактозиды, бета-глюкозиды, фрукто-, арабино-, галакто- и мальто-олигосахаридами), и в некоторых полисахаридах, например мальтодекстрин и галактан (Рис. 2).

Кроме ферментов, 48 транскрипционных регулонов также включают в своем составе от одного до нескольких углеводных транспортеров. Большинство из них относится к суперсемейству АТФ-связывающих кассетных транспортеров (ABC) и к суперсемейству мембранных транспортёров (MFS). Кроме того, фосфотрансферная транспортная система (PTS), специфичная к глюкозе, была найдена в составе регулона GlcR, тогда как гены *ptsI* и *hpr*, кодирующая компоненты PTS, общие для различных PTS-систем, контролируются регулятором PtsR из семейства LacI. Хотя, некоторые регулоны, например GlxR (глицерат) и GalR (галактоза), не содержат в своем составе транспортеров, а содержат только метаболические ферменты. Некоторые регулоны, наоборот, содержат углеводные транспортеры, но не содержат полных катаболических путей (например, SgaR2, MsmR1, MsmR2, RafR); возможно, недостающие ферменты контролируются в составе других регулонов, или не регулируются вообще.

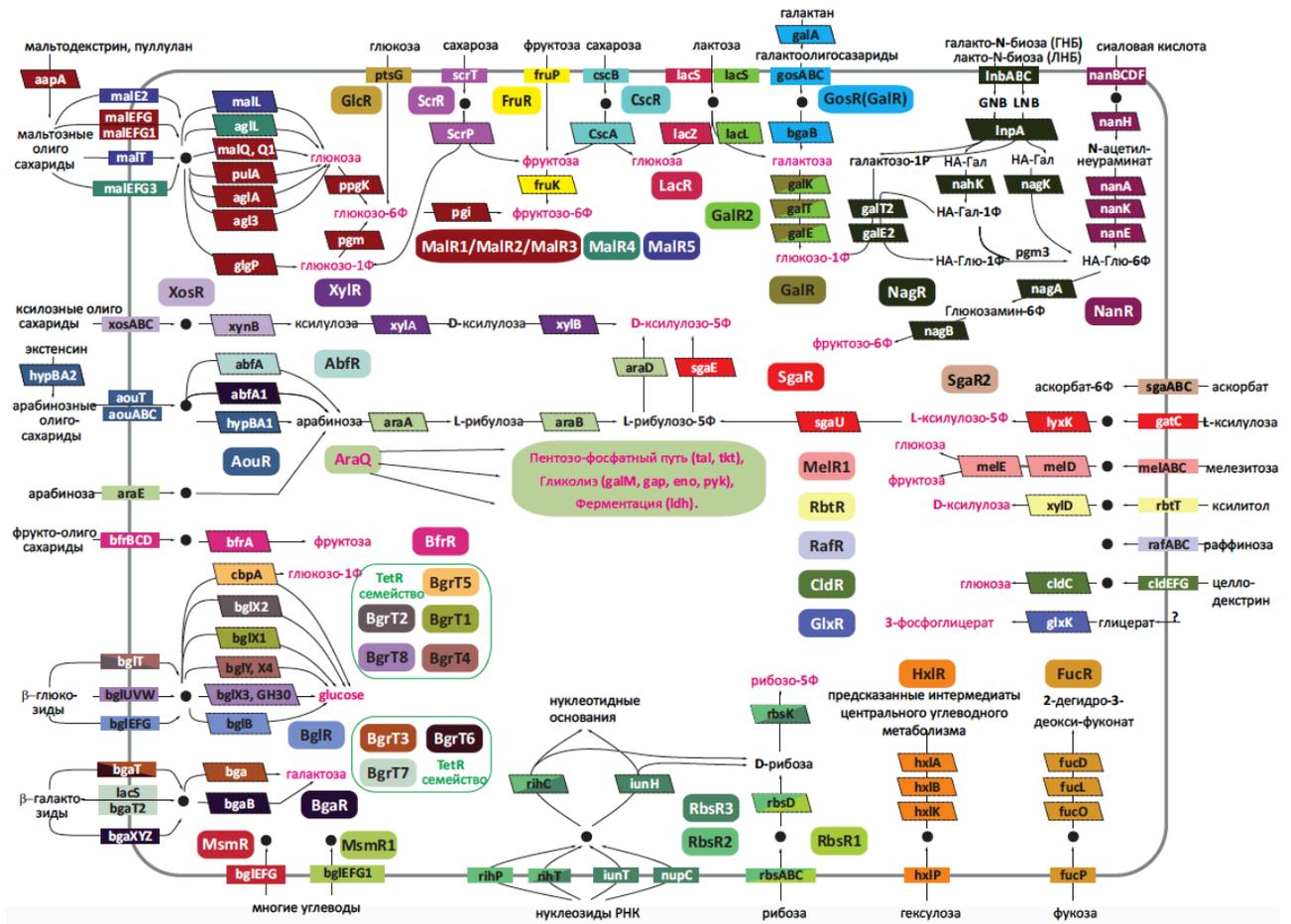


Рис. 2. Пути утилизации углеводов, участвующие в реконструированной сети транскрипционной регуляции в геномах бифидобактерий. Регуляторы обозначены овалами, транспортеры – прямоугольниками, а ферменты – скошенными прямоугольниками. Гены, регулируемые одним и тем же ТФ, отмечены одним и тем же цветом. 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

Реконструированная метаболическая сеть в 10 бактериях рода *Bifidobacterium* покрывает пути утилизации разнообразных сахаров (Рис. 2). Некоторые пути контролируются более чем одним регулятором. Например, мы реконструировали регулоны трех ортологичных групп транскрипционных факторов, контролирующих пути утилизации мальтозы и мальтодекстрина; 8 групп ТФ, контролирующих пути утилизации бета-галактозы и бета-гликозидов и 3 группы ТФ, контролирующих пути утилизации рибозы и/или пиримидиновых нуклеозидов. Мы описываем содержание этих мульти-ТФ регулонов в следующих параграфах. Возможные объяснения избыточной регуляции путей утилизации сахаров

транскрипционными факторами включают: (1) неортологичное замещение транскрипционных факторов для одного и того же метаболического пути в разных организмах, (2) существование альтернативных вариантов одного и того же пути, выполняемых паралогичными ферментами, которые регулируются различными транскрипционными факторами в одном и том же организме. Например, у бифидобактерий существуют два альтернативных пути утилизации сахарозы; утилизация сахарозы может осуществляться либо через гидролазу CscA, или через фосфорилазу ScrP; эти два фермента регулируются транскрипционными факторами CscR и ScrR, соответственно. Ферменты пути катаболизма галактозы ко-регулируются транскрипционным фактором GalR из семейства DeoR в 9 из 10 изученных геномах. Однако, эти гены ко-регулируются транскрипционным фактором GalR2 из семейства LacI вместе с генами лактозного катаболизма в бактерии *B. angulatum*. Другой пример - в двух метаболических путях катаболизма лакто-N-биозы/галакто-N-биозы и сиаловой кислоты последние две ферментативные реакции осуществляются одними и теми же ферментами (NagA и NagB, см. Рис. 2). Однако, регуляторные стратегии для этих двух метаболических путей различаются: общие для двух путей гены *nagA* и *nagB* ко-регулируются вместе с генами регуляции лакто-N-биозы/галакто-N-биозы (*lnbABC-lnpA-nahK-galT2-galE2*) при помощи транскрипционного фактора из ROK-семейства NagR, тогда как остальные гены катаболизма сиаловой кислоты (*nanBCDFA-nagB1-nanH*) контролируются регулятором NanR [80].

Ко-эволюция рибозных и нуклеозидных регулонов

В *Bifidobacterium breve* UCC2003, гены катаболизма рибозы собраны в оперон *rbsACBDK*, который кодирует гены ABC-транспортера, рибозной мутаротазы и рибокиназы, соответственно. Транскрипционный фактор RbsR из семейства LacI, специфичный к D-рибозе, закодирован геном, расположенным рядом с *rbs*-опероном, и контролирует экспрессию *rbs* оперона [76]. 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. *B. dentium* кодирует RbsR1, контролирующей оперон *rbsACBD*. Однако, предсказанные ДНК-связывающие мотивы RbsR1 и RbsR из *B. breve* UCC2003 сильно различаются, поэтому мы переименовали второй регулятор в RbsR2. Мы нашли ортологи RbsR2 в геномах *B. breve* DSM20213 и в двух штаммах *B. longum*, хотя их реконструированные регулоны не содержали генов катаболизма рибозы.

Вместо этого, регулоны RbsR2 в этих трех геномах содержали рибокиназу *rbsK*, неспецифичную рибонуклеозид-гидролазу *rihC*, и предсказанные рибонуклеозидные транспортеры *rihP* и *rihT*. RbsR3-регуляторы в геномах *B.angulatum*, *B.dentium*, *B.adolescentis*, *B.animalis* и *B.gallicum* формируют отдельную ветвь на филогенетическом дереве транскрипционных факторов из семейства LacI (Рис. 1А). Однако, оба регулятора (RbsR2 и RbsR3) характеризуются похожими ДНК-связывающим мотивами с общим консенсусом TGATAAAACGTTTTATCA. Регулон RbsR3 включает неспецифичную рибонуклеозид-гидролазу *rihC*, рибонуклеозидные транспортеры *rihT* и *nupC* и рибокиназу *rbsK*. Кроме того, в геноме *B. dentium* есть второй паралог RbsR3, который, возможно, контролирует экспрессию нуклеозид-гидролазы *iunH*, специфичной к инозину и/или уридину, и предсказанного инозинового/уридинового транспортера *iunT*. Таким образом, мы предполагаем, что несколько рибозо-специфичных регулонов RbsR обнаруженных в различных бифидобактериальных геномах контролируют катаболизм рибозо-содержащих нуклеозидов.

Гены утилизации мальтозы и мальтодекстрина регулируются паралогичными ТФ из семейства LacI

10 изученных геномов содержат 29 ТФ из семейства LacI, которые, согласно нашим предсказаниям, регулируют экспрессию широкого спектра белков, участвующих в катаболизме мальтозы и мальтодекстрина, а также различных полимеров, содержащих α -глюкозидные связи, таких как крахмал, амилопектин, амилозу и пуллулан (Таблица 2). Анализ соответствующего филогенетического дерева показал, что мальтозо-специфичные ТФ содержат пять паралогичных групп регуляторов; здесь и далее мы называем их MalR1 - MalR5 (Рис. 1А). Реконструкция регулонов показала, что три из пяти мальтозных регуляторов обладают похожими мотивами сайтов связывания, и потому могут контролировать один и тот же общий регулон; мы называем его регулон MalR1/MalR2/MalR3 (Рис. 3а). Гены, кодирующие регуляторы *malR1*, *malR2*, и *malR3* мозаично распределены среди изученных геномов. Однако, в тех случаях, когда любые два из них встречаются вместе в каком-либо геноме, они зачастую расположены рядом друг с другом на одной и той же хромосоме. Более того, все эти гены, кодирующие ТФ, 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.

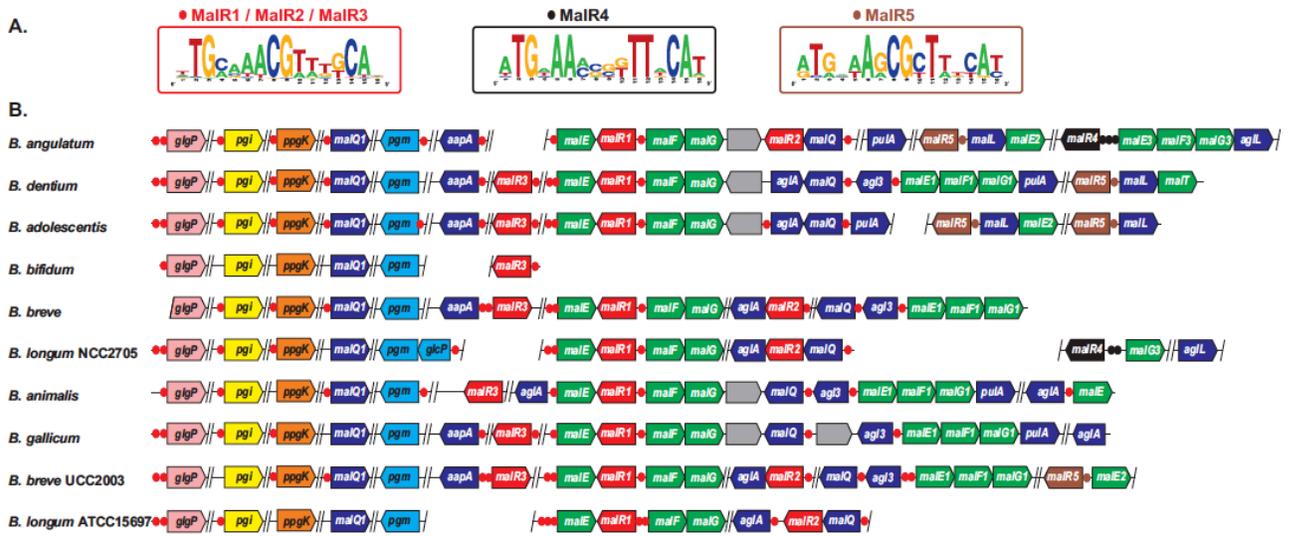


Рисунок 3а. Геномная организация регулонов утилизации мальтозы и мальтодекстрина у бифидобактерий. (А) Лого консенсусных последовательностей предсказанных ДНК-связывающих сайтов для 5 ортологичных групп MalR регуляторов. Мотивы групп MalR1, MalR2 и MalR3 одинаковые. (Б) Геномный контекст MalR регулонов 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/мальтозосахаридов *malEFG/malEFG1/malE2/malEFG3* показаны зеленым. Гликозил гидролазы, участвующие в утилизации мальтодекстрина, показаны темно-синим. Гликоген фосфорилаза *glgP*, глюкозо-6-фосфат изомеразы *pgi*, глюкокиназа *prgK* и фосфоглюкомутаза *prgM* 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.

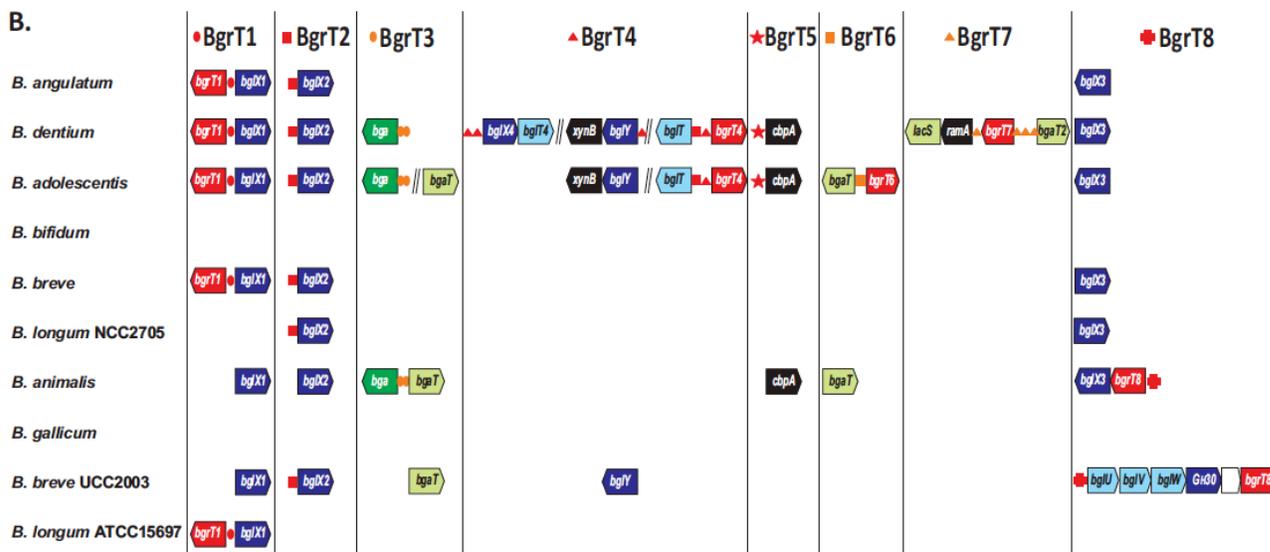


Рис. 3б. Геномная организация регулонов семейства TetR, контролирующих утилизацию бета-глюкозидов и бета-галактозидов в бифидобактериях. (А) Лого консенсусных последовательностей предсказанных ДНК-связывающих сайтов для 8 ортологичных групп BgrT регуляторов. Показана специфичность регуляторов для бета-глюкозидов 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.

Найденные регулоны катаболизма мальтозы/мальтодекстрина содержат широкий спектр генов, включающих две паралогичные системы ABC-транспортеров (*malEFG*, *malEFG1*), секретлируемую α-амилазу (*aapA*, ранее называемую *apuB*), различные цитоплазматические гликозил-гидролазы, такие как α-глюкозидазы (*aglA*, *agl3*), амиломальтазы (*malQ*, *malQ1*), пуллуназы (*pulA*), и другие цитоплазматические ферменты, в том числе гликоген-фосфорилазу (*glgP*), глюкокиназу (*ppgK*), фосфоглюкомутазу (*pgm*) и глюкозо-6-фосфат изомеразу (*pgi*). Три последних фермента участвуют в центральном метаболизме глюкозы и глюкозо-1-фосфата,

которые конвертируются в фруктозо-6-фосфат и таким образом поступают в специфичный для бифидобактерий гликолитический путь («бифидный шунт») (Рис. 4). Таким образом, ферменты и транспортеры, контролируемые регулоном MalR1/MalR2/MalR3, могут составлять полный путь катаболизма мальтозы/мальтодекстрина/пуллулана (Рис. 3а).

Стоит отметить, что идентифицированные опероны транспорта мальтозы/мальтодекстрина у бифидобактерий кодируют только две из трех типичных компонент транспортеров из ABC семейства: субстрат-связывающий белок (*malE*) и две гомологичные пермеазы (*malF*, *malG*), но не АТФ-связывающую компоненту (АТФ-азу). Мы предполагаем, что недостающая АТФ-аза кодируется отдельным консервативным геном, не регулируемым MalR1/MalR2/MalR3. Например, в *B. longum* NCC2705 таким геном может служить BL0673, который на 63% похож на ген *msmK*, служащий общей АТФ-азой для многих сахарных ABC-транспортеров у *Streptococcus pneumoniae* [122].

Реконструированный регулон MalR1/MalR2/MalR3 хорошо консервативен в 9 из 10 изученных геномов, когда как геном *B. bifidum* не содержит ортологов многих генов утилизации мальтозы/мальтодекстрина, но содержит единственный ортологичный регулятор (MalR3). Реконструкция регулона MalR3 у *B. bifidum*, позволяет предположить, что данный ТФ контролирует экспрессию только гликоген фосфорилазы *glgP*, а также самого гена *malR3*. Оставшиеся две группы ортологичных мальтозных регуляторов, а именно 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 специфичных к мальтодекстриновым олигосахаридов (*aglL*, *malL*; Рис. 3а).

Контроль утилизации глюкозидов и галактозидов новыми ТФ из семейства TetR

Бактериальные регуляторы семейства TetR контролируют экспрессию генов, участвующих в различных биологических процессах, включая устойчивость к антибиотикам, межклеточные сигнальные системы и метаболизм различных веществ [121]. В данной работе, мы нашли новую группу регуляторов из TetR семейства, гены которых закодированы в хромосомных кластерах, которые также включают гены, участвующие в катаболизме сахаров. Реконструкция филогенетического дерева найденных бифидобактериальных регуляторов из

семейства TetR и гомологичных белков из других таксономических групп показала, что все

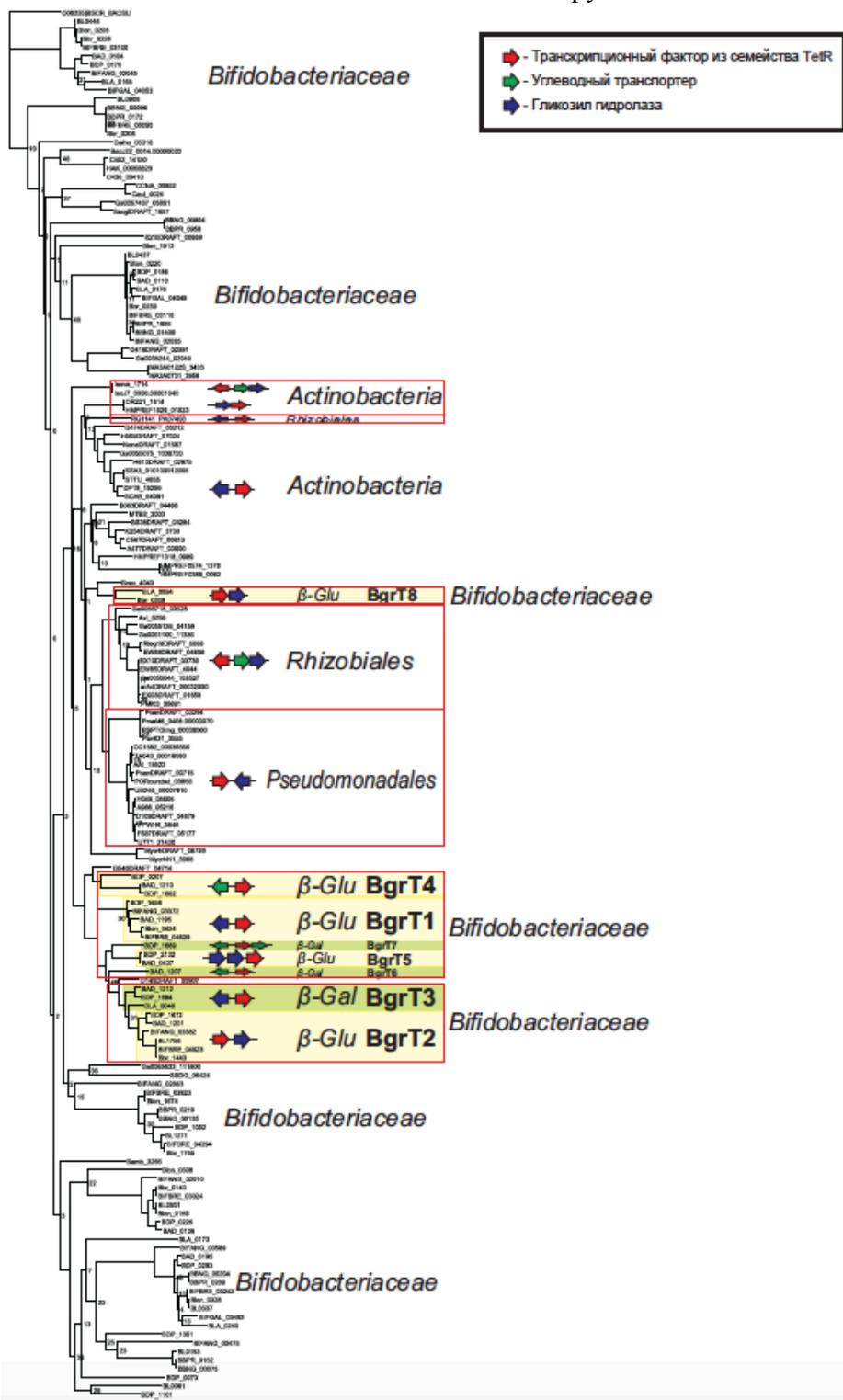


Рис. 5 Филогенетическое дерево ТФ из семейства TetR, предположительно участвующих в контроле генов утилизации углеводного метаболизма у бифидобактерий. ТФ, кодируемые генами сцепленными на хромосоме с генами гликозил гидролаз и/или углеводных транспортеров показаны в красных рамках. Геномный контекст этих регуляторов из семейства TetR показан стрелками, обозначающими гены регуляторов (красный), гликозил гидролазы (синий) и углеводные транспортеры (зеленый). Те 8 ортологических групп

регуляторов, для которых нам удалось произвести геномную реконструкцию в бифидобактериях (*BgrT1-BgrT8*), отмечены цветным фоном, светло зеленым и желтым для регуляторов, контролирующих утилизацию бета-глюкозидов и бета-галактозидов, соответственно.

регуляторы, гены которых образуют кластеры на хромосоме с генами катаболизма сахаров, формируют единую ветвь на дереве, которую мы называем *BgrT* (бета-глюкозидные регуляторы из семейства *TetR*, Рис. 5). Большинство регуляторов *BgrT* было найдено в геномах рода *Bifidobacterium*, а также в нескольких других группах *Actinobacteria* и некоторых протеобактериях (*Rhizobiales*, *Pseudomonadales*). В большинстве случаев, гены *bgrT* ко-локализуются с генами, кодирующими различные β -глюкозид или β -галактозид гидролазы (например, *bglB*, *bgaB*, *bglX*, *bglY*), целлодекстрин фосфорилазу (*cbpA*) и предсказанные β -глюкозид-/ β -галактозид-специфичные транспортеры из семейств MFS (*bglT*, *bgaT*, *lacS*) и ABC (*bglUVW*) (Рис. 3б).

Мы использовали анализ филогенетического дерева и геномного контекста, чтобы классифицировать 23 найденных белка *BgrT* из 10 бифидобактериальных геномов по 8 ортологичным группам, которые мы будем называть здесь и далее *BgrT1-BgrT8*. Для каждой из этих групп, мы смогли предсказать уникальный мотив для сайтов связывания (представленный палиндромной последовательностью длиной 18-22 нуклеотида), и реконструировать локальный регулон (Рис. 3б). 5 из 8 реконструированных регулонов содержат гены утилизации β -глюкозидов, тогда как оставшиеся 3 регулона контролируют гены утилизации β -галактозидов. Большинство реконструированных регулонов представлено в нескольких видах. Наиболее широко распространенные регулоны – это *BgrT1* (представлен в 5 геномах) и *BgrT2* (представлен в 6 геномах); и тот, и другой включают ген β -галактозидазы *bglX*, ко-локализованный с геном регулятора. Наибольшее число регуляторов *bgrT* найдено в геномах *B. dentium* и *B. adolescentis*; оба содержат 3 различных регулона, контролирующих катаболизм β -глюкозидов и 2 регулона, контролирующих катаболизм β -галактозидов. Любопытно, что два упомянутых генома содержат уникальный регуляторный каскад, где ТФ *BgrT2* контролирует экспрессию ТФ *BgrT4*, поскольку регуляторная область гена *bgrT4* содержит сайты связывания обоих регуляторов (*BgrT2* и *BgrT4*). Таким образом, регуляторная сеть утилизации β -глюкозидов в *B. dentium* и *B. adolescentis* имеет более сложную структуру по сравнению с остальными изученными бифидобактериями.

Предсказанный глобальный регулон контролирует гены центрального углеводного метаболизма

Ген, кодирующий новый ТФ из семейства LacI (мы называем его *araQ*), ко-локализован с опероном генов катаболизма арабинозы *araBDA* в 6 из 10 изучаемых бифидобактериальных геномов. В геноме *B. angulatum*, этот оперон также включает арабинозную пермеазу *araE*. Оставшиеся 4 генома (*B. bifidum*, *B. longum* subsp. *infantis* ATCC 15697, и оба штамма *B. breve*) не содержат генов катаболизма арабинозы; однако, мы заметили, что они тем не менее содержат ортологов гена *araQ*. Проанализировав регуляторные области оперонов *araBDA* (далее – *ara*) и генов *araQ*, мы нашли возможный AraQ-связывающий мотив: палиндромную последовательность с консенсусом TGTGAGCGCTCACA, который отдаленно похож на связывающие мотивы других регуляторов из семейства LacI (см. ниже). Поиск дополнительных сайтов связывания по всем изучаемым геномам позволил нам идентифицировать глобальный регулон AraQ, содержащий до 19 генов на геном (см. в базе данных RegPrecise).

Наиболее консервативные члены регулона AraQ кодируют ферменты из центральных гликолитических путей. К числу этих генов относятся, например, альдозо-1-эпимераза (*galM*), гены пентозо-фосфатного пути и бифидного шунта (*tal*, *tkl*), гены гликолиза (*gap*, *eno*, *pyk*) и лактат дегидрогеназа (*ldh*) (Рис. 4).

Любопытно, что в большинстве геномов предсказанный регулон AraQ не пересекается с регулоном MalR, который контролирует некоторые другие гены метаболизма глюкозы и фруктозы (Рис. 4). Однако, во всех изученных штаммах *B. breve* и *B. longum*, регулон AraQ включает в себя несколько дополнительных генов утилизации мальтозы/мальтодекстрина, а именно мальтозный транспортер (*malE*) и амиломальтазу (*malQ1*), а также 1,4-альфа-глюкан ветвящий фермент (*glgB*), а также биотин-лигазу (*birA*) и два предсказанных белка (*Blon_2289-90*). Таким образом, гены *malE* и *malQ1* находятся под двойной регуляцией ТФ AraQ и MalR в некоторых геномах бифидобактерий. Также, в геномах *B. adolescentis* и *B. dentium* реконструированный регулон AraQ включает ацетилтрансферазу мальтозы (*maa*), которая обладает широкой специфичностью и может ацетилировать различные сахара. Физиологическая роль этого фермента неизвестна, но предполагают, что он предотвращает накопление сахаров до чрезмерно высоких концентраций при помощи ацетилирования, и таким образом играет защитную роль в клетке. Другие геном-специфичные члены регулона AraQ включают предсказанный гликозил гидролазный оперон в *B. adolescentis* (PF07944-*aga*) и оперон утилизации L-ксилолата в *B. bifidum*, который также контролируется локальным регулятором SgaR.

В недавнем исследовании ген *araQ* был найден в числе ключевых генов, присутствующих у всех 45 изучаемых геномах бифидобактерий [123], [124], [125]. Арабинозный оперон *araBDA*, напротив, отсутствует у 7 штаммов бифидобактерий, а именно *B. bifidum*, *B. boum*, *B. breve*, *B.*

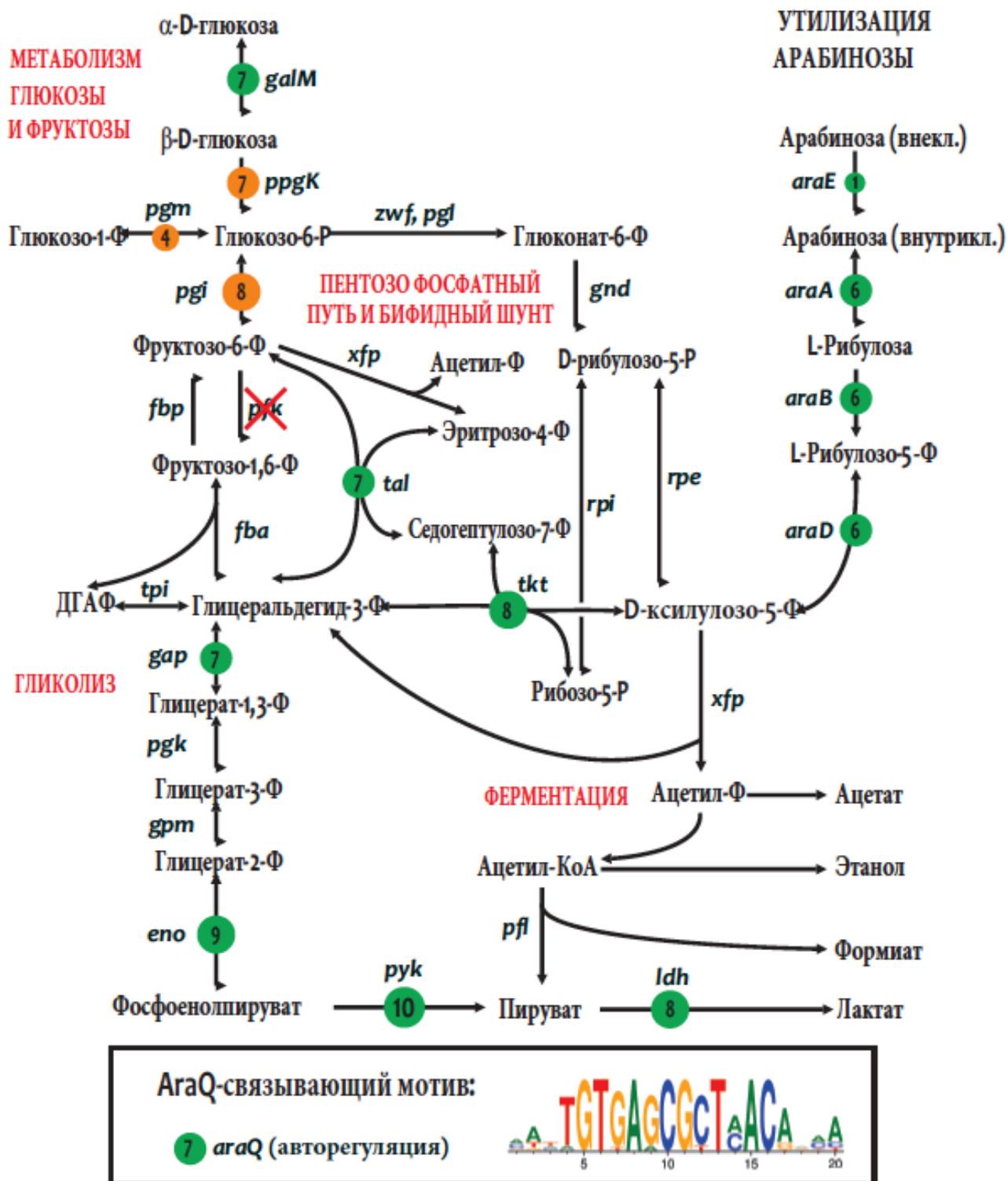


Рис. 4. Метаболический контекст и консервативное ядро реконструированных регулонов

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.

choerinum, *B. longum* subsp. *infantis*, *B. ruminantium* и *B. thermophilum* (данные не показаны). Результаты нашего анализа 10 геномов бифидобактерий позволяют предположить, что глобальная регуляция множества генов центрального углеводного метаболизма является общим свойством реконструированных регулонов *AraQ* у всех бифидобактерий (Рис. 4). Геном еще одного представителя семейства *Bifidobacteriaceae*, *Gardnerella vaginalis*, содержит ортологичный регулятор *AraQ* (аминокислотная последовательность похожа на 51%), который, вероятно, контролирует гены центрального сахарного метаболизма (*eno*, *gap*, *pyk*). Однако, гены утилизации арабинозы *ara* отсутствуют в этом геноме. Поиск дополнительных ортологичных регулонов вне родов *Bifidobacterium* и *Gardnerella* позволил обнаружить далеких гомологов *araQ* в нескольких группах *Actinomycetales* (аминокислотные последовательности похожи на 35-40%). Большинство этих гомологов ко-локализуется на хромосоме с оперонами *ara*, содержащими *AraQ*-связывающие сайты в своих регуляторных областях (данные не показаны); однако, мы не смогли найти дополнительные *AraQ*-связывающие сайты в других удаленных областях этих геномов. Эти наблюдения позволяют предположить возможный сценарий эволюции регулонов *AraQ*. Локальные регуляторы генов утилизации арабинозы *AraQ*, вероятно, были представлены в общем предке *Actinomycetales*. В общем предке бактерий из родов *Bifidobacterium* и *Gardnerella* *AraQ* регулоны расширились, включив в себя гены центрального углеводного метаболизма. Позже, гены пути утилизации арабинозы (вместе с соответствующими *AraQ*-связывающими сайтами) были утрачены во многих геномах семейства *Bifidobacteriaceae*, тогда как глобальные *AraQ* регулоны сохранились. Молекулярный эффектор *AraQ* пока неизвестен. Можно предположить, что *AraQ* диссоциирует от своих сайтов связывания ДНК при взаимодействии с интермедиатом центрального гликолитического пути, например с D-ксилоулозо-5-фосфатом – конечным продуктом арабинозного катаболизма.

Обсуждение

Регуляция транскрипции углеводного метаболизма была изучена ранее у нескольких модельных организмов из рода *Bifidobacterium*; наиболее широкие исследования были проведены для штамма *B. breve* UCC2003. В этой работе, мы использовали сравнительно-геномные методы для более широкого и глубокого изучения регуляции транскрипции углеводного метаболизма у 10 различных представителей бифидобактерий. Изученная нами сеть транскрипционной регуляции углеводного катаболизма управляется широким набором ТФ из различных белковых семейств. Среди них большинство принадлежит к семействам LacI and ROK, остальные являются представителями не более 10 других семейств, также участвующие в описываемой регуляторной сети. Любопытно, что нам удалось обнаружить несколько регуляторов из семейства TetR (для которого не была ранее показана роль в регуляции углеводного метаболизма), регулирующих гены утилизации глюкозидов и галактозидов. Мы обнаружили мотивы ДНК-связывающих сайтов и реконструировали регулоны для 268 из 308 потенциальных углеводо-специфичных ТФ (всего 64 ортологичные группы регуляторов). Большинство анализируемых ТФ контролируют локальные регулоны, где ТФ связывает сайты, расположенные близ гена регулятора на хромосоме, и таким образом контролирует небольшой набор генов, как правило, участвующих лишь в одном метаболическом пути. Всего лишь несколько регулонов представлены во всех 10 изучаемых геномах; такое «мозаичное» распределение регулонов может быть объяснено необходимостью адаптации к специфическим экологическим нишам. Напротив, некоторые регуляторные системы, например регуляторы катаболизма сахарозы, фруктозы, лактозы, глюкозы, галактозы и мальтозы, консервативны и сохраняются в большинстве изучаемых геномов. Вышеперечисленные сахара могут встречаться в местах обитания всех анализируемых бифидобактерий, и потому потеря возможности их метаболизировать может быть сопряжена с потерей эволюционного преимущества. Среди новых ТФ, найденных в процессе данной работы, одним из наиболее интересных является AgaQ - первый предсказанный глобальный регулятор в бифидобактериях, контролирующий в том числе и гены центрального углеводного метаболизма. Неудивительно, что бифидобактерии регулируют экспрессию генов, составляющих их центральные метаболические пути. Функциональные аналоги (но не ортологи) AgaQ были описаны ранее в других группах бактерий, включая Cra/FruR у бактерий из семейства *Enterobacteriaceae* и HexR у бактерий из рода *Shewanella* [126]. Многие

бифидобактерии обитают в желудочно-кишечном тракте у людей и у животных, т.е в среде, богатой сложными углеводами, происходящими как из клеток эпителия организма-хозяина, так и из его питания. Таким образом, способность бактерий – обитателей кишечного тракта модулировать экспрессию генов, важных для утилизации сложных углеводов, напрямую связана с их приспособленностью и способностью соревноваться с другими обитателями той же экологической ниши.

Реконструированная регуляторная сеть показывает высокую корреляцию (коэффициент корреляции 0.95) между числом углеводно-специфичных ТФ и суммарным числом генов, участвующих в углеводном метаболизме среди изучаемых геномов (Таблица 1). Это соответствует интуиции: очевидно, что геномы с меньшим числом генов утилизации углеводов могут содержать меньше углеводно-специфичных регуляторов. Например, согласно нашим предсказаниям, геном оппортунистического патогена *B. dentium* (одного из возбудителей кариеса), содержит 54 углеводно-специфичных ТФ, и 15.5% генов, участвующих в метаболизме углеводов. Напротив, геном *B. gallicum* содержит лишь 12 углеводно-специфичных ТФ, и лишь 9.2% генов в этом геноме участвуют в метаболизме углеводов.

Данное биоинформатическое исследование значительно расширяет наше понимание углеводного метаболизма и его регуляции у бифидобактерий. В дальнейшем, это знание облегчит разработку новых пищевых добавок, укрепляющих здоровье пациентов. Однако, множество предсказанных регуляторных взаимодействий между ТФ и их связывающими сайтами требует дальнейшего экспериментального подтверждения. Систематическое сравнение регулонов, реконструированных исходя из геномных данных, с результатами экспериментальных исследований (например, транскриптомных исследований мутантов с «выключенными» генами ТФ), позволит получить исчерпывающие знания о регуляции углеводного метаболизма в бифидобактериях.

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

Методы

Исследуемые бактериальные геномы были получены из базы данных MicrobesOnline [108]. Гены потенциальных ТФ из семейства LacI были найдены при помощи программы поиска белковых доменов, доступной на сервере Pfam [105]. ТФ из семейства LacI состоят из двух характеристических доменов: N-концевого НТН ДНК-связывающего домена (PF00356) и C-концевого эффектор-связывающего домена, гомологичного периплазматическим связывающим белкам ABC-транспортеров (PF00532, PF13377 или PF13407). Для поиска групп ортологичных генов среди нескольких геномов использовалась программа GenomeExplorer [112] и филогенетические деревья для семейств генов из базы данных MicrobesOnline [108], при этом использовались три основных критерия для определения ортологичных групп (см. выше). Для реконструкции регулонов мы использовали сравнительно-геномный метод, реализованный на веб-сервере RegPredict [107] (см. выше). Используемый метод в целом повторяет процедуры, описанные в предыдущей секции раздела «Методы», за исключением нескольких модификаций. Во-первых, в отличие от регулонов бифидобактерий отдельные регулоны ТФ семейства LacI были уже реконструированы и проанализированы ранее. Мы извлекли известные регулируемые гены и связывающие мотивы изученных регулонов из баз данных RegTransBase [127], RegulonDB [128], DBTBS [129] и CoryneRegNet [130]. Все изученные в данном исследовании ТФ, их предсказанные ДНК сайты связывания, а также консенсусы ДНК мотивов и реконструированные регулоны были загружены в открытую базу данных RegPrecise [48].

Метод сравнительно-геномного анализа, реализованный на Web-сервере RegPredict [107], состоит из трех шагов: (i) выбор таксономических групп родственных друг другу бактерий; (ii) выбор набора геномов, представляющих каждую из этих групп; и (iii) реконструкция регулонов в выбранных геномах. Для анализа регулонов ТФ из семейства LacI, мы выбрали набор из 344 геномов, представляющих 39 таксономических групп из 7 бактериальных типов (см. в базе данных RegPrecise). Мы проанализировали 19 таксономических групп относящихся к типу *Proteobacteria* (183 генома), 9 групп *Firmicutes* (72 генома), 7 групп *Actinobacteria* (57

геномов). Типы бактерий *Bacteroides*, *Chloroflexi*, *Deinococcus-Thermus* и *Thermotogae* были представлены лишь одной таксономической группой каждый (32 генома в сумме).

Результаты

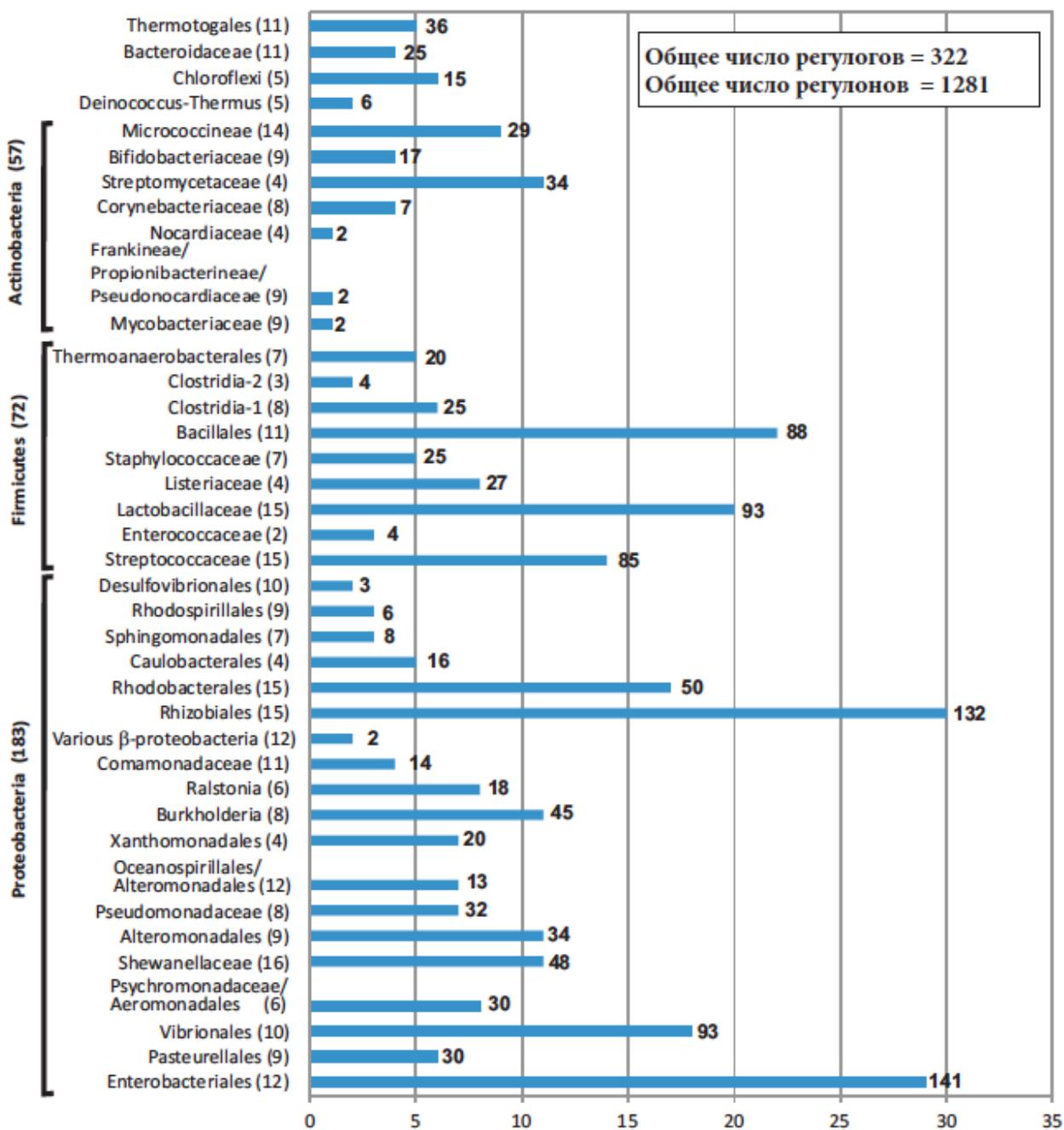
Репертуар генов ТФ из семейства LacI в бактериальных геномах

Чтобы оценить распространенность ТФ семейства LacI в изучаемых геномах, мы использовали поиск по схожести и опубликованные наборы прокариотических ТФ. Найденные ТФ были неравномерно распределены среди изучаемых геномов: 90% геномов (309/344) содержали все 2572 найденных белка, тогда как оставшиеся 10% не содержали ТФ из семейства LacI. (см. в базе данных RegPrecise). Наибольшее среднее число регуляторов из семейства LacI на геном достигается в нескольких группах бактерий относящихся к типу *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). В остальных таксономических группах геномы в среднем содержат менее чем 10 ТФ из семейства LacI (Рис. 6). Стоит отметить, что мы не нашли ТФ из семейства LacI в геномах групп *Methylophilales*, *Neisseriales*, *Nitrosomonadales*, *Oceanospirillales*, *Magnetospirillum/Rhodospirillum* и *Desulfovibrionales*. Отсутствие регуляторов LacI в этих группах протеобактерий может быть связано (1) с малым количеством генов сахарного катаболизма в их геномах (тогда как ТФ из семейства LacI регулируют главным образом сахарный метаболизм, см. ниже), или (2) с повышенным использованием ТФ из других семейств для регуляции метаболических путей.

Статистика реконструированных регулонов и регулогов

Все множество найденных ТФ из семейства LacI было разбито на ортологичные группы, специфичные для отдельных таксономических групп; полученные ортологичные группы были подвергнуты детальному сравнительно-геномному анализу при помощи Web-сервера RegPredict [107]. Несколько ТФ были объединены в ортологичную группу при условии, что они формировали моно- или парафилетическую ветвь на филогенетическом дереве (Рис 18). В большинстве случаев, ортологичные группы содержат не более одного ТФ на геном. Многие

семейства ТФ содержат множество паралогов, образовавшихся в результате частых геномных



Число изученных регулонов для различных таксономических групп

*Общее число регулонов показано рядом с полосками

Рис. 6 Распределение регулонов ТФ из семейства *LacI* среди анализируемых таксономических групп бактерий

дупликаций, делеций и горизонтальных переносов. Это затрудняло точное распознавание ортологичных генов, поэтому в некоторых случаях нам приходилось случайно выбирать одно

из нескольких возможных решений, поддерживаемых геномным контекстом или функциональными свойствами реконструированных регулонов. В результате, изучаемые регуляторы были классифицированы по 190 ортологичным группам, характеризующим консервативными ДНК-мотивами и регулирующими путями. Анализируя ортологичные группы регуляторов в каждой таксономической группе, 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 RegPrecise). Реконструированные регулоны содержат 7465 предсказанных сайтов связывания, 6076 оперонов и 13558 генов.

Две трети всех найденных ортологичных групп (125/190) содержат ТФ лишь из одного регулога, определяемого как набор ортологичных регулонов в группе близко родственных геномов. 37 ортологичных групп содержат 2 регулога, тогда как остальные 28 групп содержат 3 и более регулога (Рис. 7А). На каждую ортологичную группу ТФ приходится от 1 до 59 регулонов (и соответствующих ТФ), в среднем по 6.7 (Рис. 7Б). Чаще всего, ортологичные группы содержат 2 регулона (37 групп); 60% всех групп содержат 5 или меньше регулонов, 20% всех групп содержат между 6 и 10 регулонами. Наиболее широкие группы ортологичных ТФ из семейства LacI были найдены для глобального регулятора катаболизма CsrA у *Firmicutes* (59 регулонов, 6 регулогов), для рибозного репрессора RbsR in *Proteobacteria* (52 регулона, 10 регулогов), для мальтозного репрессора MalR у *Firmicutes* (42 регулона, 6 регулогов), а также у регуляторов сахарного метаболизма FruR (40 регулонов, 6 регулогов), GntR (37 регулонов, 7 регулогов), GalR (27 регулона, 5 регулогов) у *γ-Proteobacteria*.

Глобальные и локальные регулоны

Реконструированные регулоны ТФ семейства LacI существенно различаются по числу предсказанных регулируемых генов и оперонов. Большинство регулонов (1198/1288, 93%) содержат не более 20 генов (Рис. 8А), три четверти этих регулонов содержат от 2 до 7 генов, тогда как 26 регулонов содержат только по одному регулируемому гену. Что касается числа регулируемых оперонов (Рис. 8Б), большинство изученных ТФ из семейства LacI регулирует

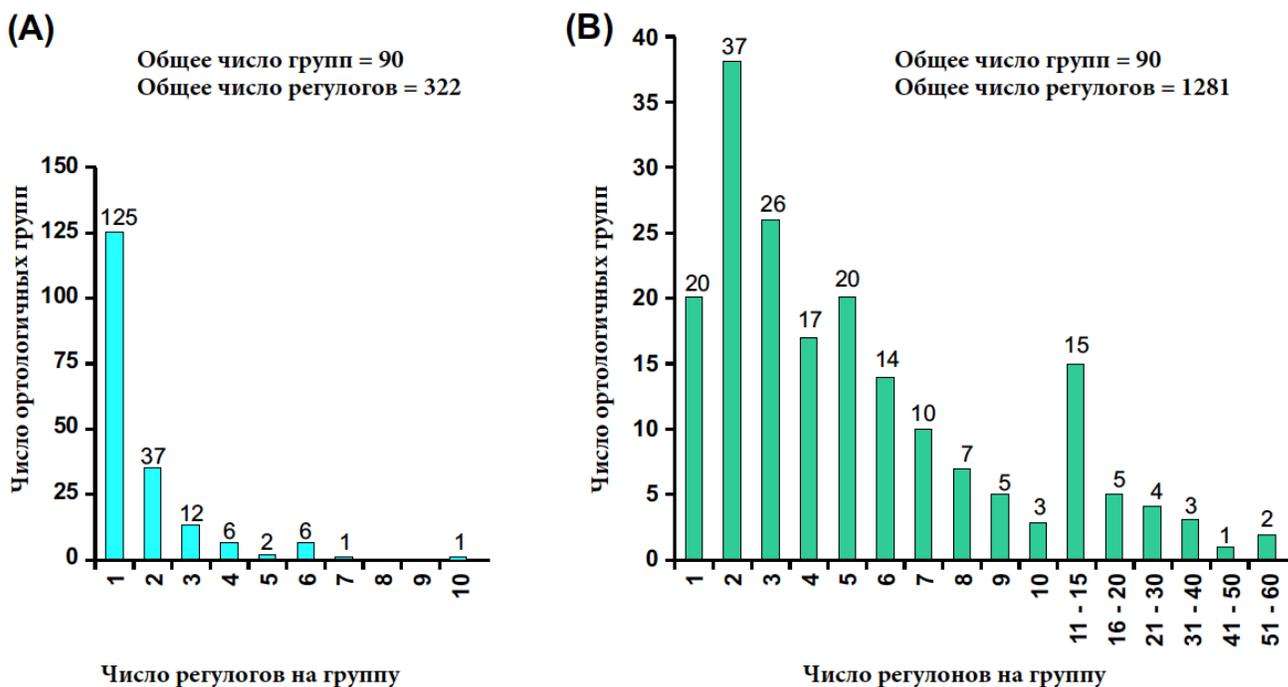


Рис. 7 Содержание регулонов и регулогов изученных ТФ из семейства LacI. (А) Содержание регулонов. (Б) Содержание регулогов

один (31%) или два (38%) оперонов. 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 из типа *Firmicutes*. Регулоны CsrA были детально описаны ранее для бактерий из таксонов *Bacillaceae* [131], [132], [120], *Staphylococcaceae* [133], [119], *Lactobacillales* [134], [135], [136], [137], [97], и *Clostridiaceae* [139]. Другая большая группа глобальных регулонов (31 регулон, 3 регулога) содержит регулоны ортологов пуринового

репрессора PurR из *E. coli* в трех близких таксономических группах γ -*Proteobacteria*:

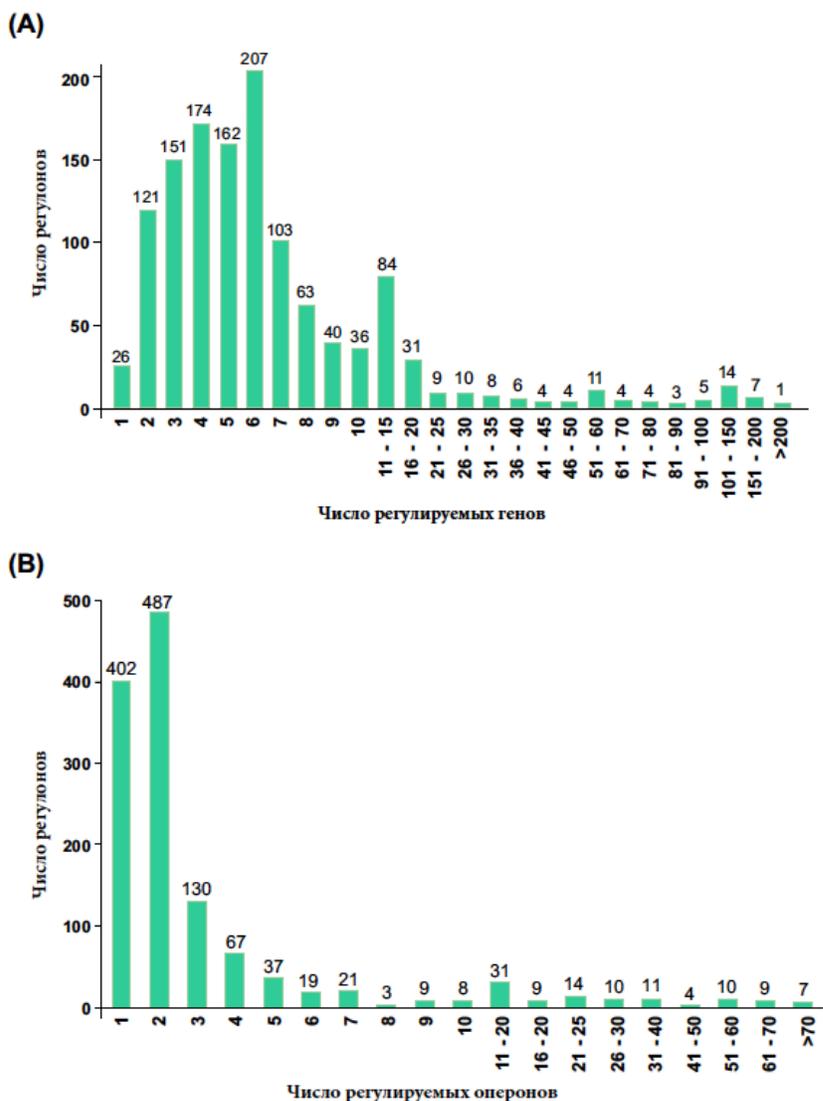


Рис. 8. Распределение реконструированных регулонов ТФ из семейства *LacI*. (А) Распределение по числу регулируемых генов. (Б) Распределение по числу регулируемых оперонов

Enterobacteriales, *Pasteurellales* и *Vibrionales*. Глобальный транскрипционный регулятор PurR контролирует биосинтез пуринов, некоторые реакции в пути синтеза пиримидинов, метаболизм полиаминов и ассимиляцию азота [140], [141]. 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 *Rhodobacteriales* (всего 35 регулонов) – детально описаны ниже.

Для 72% изученных транскрипционных факторов (всего 943 ТФ) была присуща авторегуляция. Ожидаемо, 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).

Связывающие сайты и мотивы

Мотивы связывающих сайтов изучаемых ТФ из семейства LacI - это палиндромы, сформированные высоко консервативными обратными повторами, что соответствует результатам предыдущих исследований [142], [143], [144]. Расстояние между повторами, как правило, постоянна для каждой группы ортологичных ТФ, хотя в отдельных случаях мы наблюдали некоторое разнообразие. Подавляющее большинство мотивов ТФ из семейства LacI (1251/1303 ТФ) – это четные палиндромы (16, 18, 20 или 22 нуклеотидов в длину), тогда как неканонические нечетные палиндромы (17, 19 или 21 нуклеотидов) были найдены только для 4% регуляторов (например, LacR, GalR и EbgR). Характерное свойство четных палиндромов - это наличие динуклеотида CG в центре палиндрома (найденно у 1173 ТФ). В то же время, динуклеотид в центре палиндрома может иметь другой консенсус (49 ТФ) или различаться у разных связывающих сайтов одного и того же регулона (29 ТФ). Более 75% найденных сайтов изучаемых ТФ расположены на участке между 140 и 30 основаниями перед старт кодоном (Рис. 9А). Менее 1% связывающих сайтов расположены внутри открытой рамки считывания, включая экспериментально подтвержденные сайты LacI в гене *lacZ* [145], и PurR в гене *purB* [146] в *E. coli*. Примерно 7% связывающих сайтов расположены более чем 200 нуклеотидов перед старт кодоном. Несмотря на то, что некоторые из них могут регулировать дивергентно транскрибируемые гены, существуют примеры экспериментально подтвержденных сайтов с такой удаленной локализацией, например, сайт PurR перед геном *prsA* [147], опять в *E. coli*.

Около 20% регулируемых оперонов регулируются двумя и более связывающими сайтами. Например, двойные сайты были найдены в 1 +96118 оперонах. Количество сайтов в регуляторной области дивергентно транскрибируемых генов (здесь и далее – дивергонов)

увеличивается с длиной дивергона. Группы их 3 соседних сайтов были найдены в 116

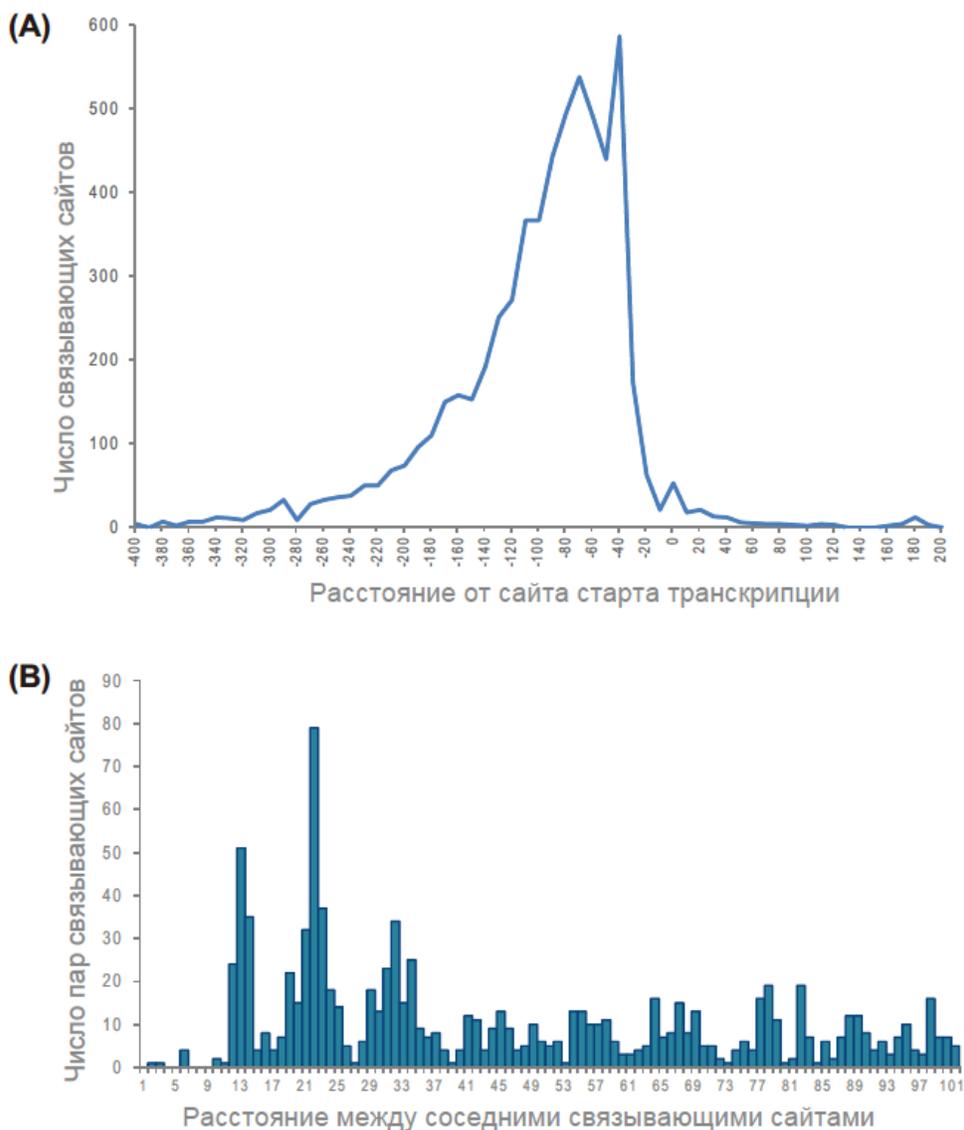


Рис. 9 Распределение расстояний. (А) Расстояния между связывающим сайтом ТФ из семейства *LacI* и сайтом начала транскрипции. (Б) Расстояния между соседними связывающими сайтами

оперонах, группы из 4 сайтов – в 5 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. 9Б). 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 *E. coli* [148]. Тривиальное предположение, что наблюдения, перечисленные выше, лишь артефакты, выглядит неправдоподобным по двум причинам. Во-первых, такие множественные сайты консервативны и даже предпочтительны в случаях некоторых ортологичных групп. Во-вторых, гипотеза артефактов не объясняет предпочитаемую дистанцию в 13 нуклеотидов. В доступных трехмерных структурах комплексов ТФ и ДНК [149], [150], [151], семь пар оснований, ближайших к центру связывающего сайта, образуют контакты между основаниями ДНК и аминокислотными остатками белка. Таким образом, область перекрывания двух сайтов также является зоной перекрывания специфичных взаимодействий между ТФ и ДНК. Предложенная ранее модель связывания тетрамеров не соответствует структурным данным, т.к димеры не могут взаимодействовать при наличии таких перекрывающихся участков. Возможно, существует специальный механизм связывания, при котором цепочки ДНК частично расплетаются.

Регулируемые метаболические пути и эффекторы

Анализируя функциональное содержание реконструированных регулонов, мы предсказали возможные биологические функции и эффекторы для 190 ортологичных групп ТФ из семейства LacI. Так, мы предсказали метаболические пути для 182 групп ТФ. 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 (см. в базе данных RegPrecise). Мы сравнили предсказанные функции регулонов с результатами ранее опубликованных экспериментальных работ, доступными для 24 ТФ из семейства LacI. Функции этих регуляторов, предсказанные в данной работе, соответствуют функциям, установленным ранее экспериментально. Основываясь на реконструкции метаболических путей и на известных метаболитах – интермедиатах метаболических путей, мы предложили возможные молекулярные эффекторы для 108 групп ТФ из семейства LacI (см. в базе данных RegPrecise). Для 21 из 108 групп, эффекторы известны из более ранних работ. Ожидаемо, подавляющее большинство изученных ортологичных групп ТФ из семейства LacI контролирует метаболизм углеводов (176/182, 96% групп, для которых были предсказаны

метаболические пути). Для большинства ортологичных групп мы предсказали пути утилизации конкретных сахаров. Еще пять ортологичных групп локальных регуляторов (в т.ч. AdeR, HpxR и UriR) контролируют пути утилизации нуклеозидов, тогда как локальный регулятор NtdR контролирует биосинтез неотрегалоадиаминов. Большинство глобальных регуляторов из семейства LacI также вовлечены в регуляцию метаболизма углеводов (в т.ч. FruR, CsrA, PckR и GapR), тогда как PurR регулирует несколько ключевых метаболических путей, включая биосинтез пуринов и пиримидинов.

В соответствии с наблюдаемой у ТФ из семейства LacI тенденцией контролировать сахарный метаболизм, результаты нашей работы говорят о том, что сахара представляют наибольший класс эффекторов для регуляторов этого семейства (103/107, 96% групп, для которых были предсказаны метаболические пути). Большинство предсказанных углеводных эффекторов – это моносахариды и их производные (26/45), включая гексозы (например, глюкозу, галактозу и маннозу), пентозы (например, рибозу и ксилозу), фосфаты сахаров (например, фруктозо-1-фосфат, аллозо-6-фосфат), сахарные кислоты (например, глюкуронат и галактуронат), сахарные спирты (например, рибитол), и аминсахара (например, ацетилгликозамин). Вторая по размеру группа сахарных эффекторов включает различные олигосахариды (14/26), включая различные дисахариды, например, целлобиозу, мальтозу, сахарозу, трегалозу, или их фосфорилированные производные (в т.ч. целлобиозо-6-фосфат или сахарозо-6-фосфат).

Наконец, несакхарные эффекторы ТФ из семейства LacI включают нуклеотидные основания (например, гуанин и гипоксантин являются ко-репрессорами PurR в *E. coli* [149]) и нуклеозиды (например, цитидин и аденозин индуцируют CytR в *E. coli* [150]).

Функциональный анализ реконструированных регулонов показал, что многие пути контролируются двумя или более неортологичными регуляторами. К таким путям относятся пути катаболизма не менее чем 10 различных типов сахаров, контролируемые более чем 90 ортологичными группами ТФ из семейства LacI (Таблица 3 и Рис. 14). Большое количество групп регуляторов контролирующих катаболизм глюкозидов и галактозидов соответствует структурному разнообразию глюкозо- и галактозо-содержащих олигосахаридов, которые могут быть утилизированы бактериями в различных естественных средах обитания. С другой стороны, разнообразие регуляторов для нескольких других сахаров, таких, как рибоза и сахароза, позволяет предположить высокую частоту конвергентных событий, при которых специфичность к одному и тому же лиганду развивалась независимо множество раз в процессе эволюции семейства LacI. Например, путь утилизации сахарозы регулируется ТФ из не менее 11 ортологических групп (Рис. 10). При анализе соответствующих регулонов сахарозы мы

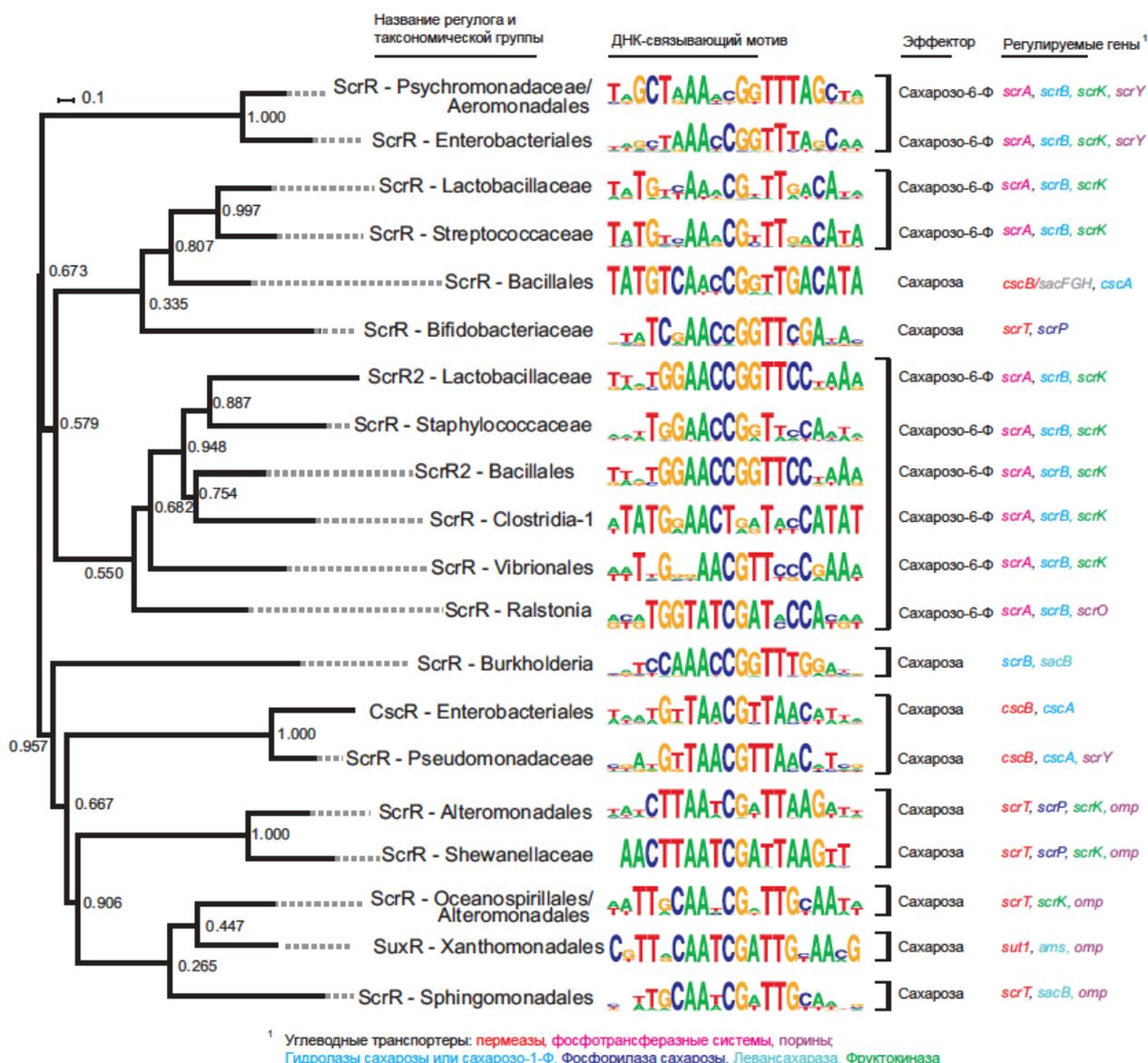


Рисунок 10. Филогенетическое дерево, лого консенсусных последовательностей связывающих сайтов, эффекторы и содержание регулонов утилизации сахарозы. Ортологичные группы ТФ показаны квадратными скобками

обнаружили множество комбинаций сахарозных транспортеров, включая пермеазы (*scrT, cscB, sut1*), фосфотрансферазные системы (*scrA*) и порины (*scrY, scrO, omp*), и ферментов катаболизма сахарозы, включая фосфорилазы (*scrP*), гидролазы (*scrB*) и фруктокиназы (*scrK*). Руководствуясь типом регулируемых сахарозных транспортеров, мы предсказываем, что соответствующие лиганды являются эффекторами регуляторов ScrR (например, сахароза в случае пермеазы и сахарозо-6-фосфат в случае фосфотрансферазной системы). Стоит отметить, что некоторые семейства, такие как *Bacillales* и *Enterobacteriales*, содержат не-

Таблица 3. Пути катаболизма углеводов, регулируемые не-ортологичными ТФ из семейства LacI

Пути катаболизма углеводов	Ортологичные группы ТФ из семейства LacI	
	Число групп	Примеры регуляторов
Глюкоза и глюкозиды	20	BglR, BglZ, CelR, AscG, KojR
Сахароза	11	CscR, ScrR, SuxR
Галактоза и галактозиды	10	BgaR, GalR, GanR, MsmR, EbgR, LacI, LacR
Рибоза	9	RbsR
Мальтоза; мальтодекстрин	8	MalR, MdxR, Mall
Инозитол	7	IolR
Глюконат и идонат	7	GntR, IdnR
Глюкуронат и галактуронат	6	ExuR, KdgR, UxaR, UxuR
Манноза и маннозиды	5	ManR
Фруктоза и фруктоолигосахариды	5	FruR, BfrR
Трегалоза	4	TreR, ThuR

ортологичные регуляторы ScrR с разными эффекторами. Ожидаемо, ДНК-связывающие мотивы регуляторов, контролирующих пути катаболизма сахарозы, консервативны внутри ортологичных групп, но различаются между неортологичными регуляторами (Рис. 10)

Глобальные регуляторы, контролирующие центральный сахарный метаболизм

К семейству LacI относятся несколько глобальных регуляторов центральных путей метаболизма углеводов, например, ранее известные регуляторы катаболитной репрессии CsrA у фирмикут и FruR у энтеробактерий. В данной работе мы (совместно с Д.А.Равчевым) провели сравнительно-геномную реконструкцию регулонов FruR в различных таксономических группах гамма-протеобактерий, тогда как реконструкция регулонов CsrA в различных таксономических группах фирмикут была опубликована ранее [119], [138], [139], [120] и доступна в базе данных RegPrecise. Более того, в данной работе мы идентифицировали три новых неортологичных регулятора (названных PckR, GapR и GluR), реконструкция регулонов которых показала, что они предположительно контролируют центральный

углеводный метаболизм в трех таксономических группах альфа-протеобактерий: *Rhizobiales*, *Rhodobacterales*, и *Caulobacterales*. Ниже мы проводим полный функциональный анализ регулонов четырех вышеуказанных ТФ у протеобактерий.

Фруктозный репрессор FruR (также известный как репрессор и активатор катаболизма Cra) известен как глобальный регулятор центрального метаболизма в *E. coli* [152]. FruR/Cra контролирует поток углерода, репрессируя гликолитические гены ферментов пентозо-фосфатного пути, и путей Эмбдена-Мейерхофа и 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 (Рис. 11). У порядков *Vibrionales* и *Pseudomonadales*, регулоны FruR состоят лишь из оперона утилизации фруктозы *fruBKA* и из гена регулятора *fruR*; таким образом, FruR в данных двух таксономических группах выступает в роли локального регулятора. У порядка *Enterobacteriales*, регулон FruR значительно больше и покрывает гены центральных гликолитических путей, части цикла трикарбоновых кислот и

произошло в группах бактерий относящихся к родам *Escherichia*, *Salmonella*, *Citrobacter*, *Enterobacter* и *Klebsiella*, где FruR также контролирует гены глиоксилатного шунта (*aceBA*). Наконец, в бактериях из группы *Escherichia/Salmonella* регулон распространился на гены пути Энтнера-Дудорова (*edd*, *eda*). Похожие тренды в эволюции глобальных регулонов в близкородственных бактериальных видах были продемонстрированы ранее для регулона PhoP в энтеробактериях [153]. С другой стороны, FruR регулон деградировал в бактериях порядка *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 карбоксикиназы фосфоенолпирувата *pckA* (идентификационный номер EMBL AF004316.1); однако, этот ген не был изучен экспериментально. Ортологи PckR были обнаружены в 10 из 15 проанализированных геномов из порядка *Rhizobiales*, включая виды из семейств *Rhizobiaceae*, *Brucellaceae*, *Phyllobacteriaceae* и *Xanthobacteraceae*. Используя сравнительно-геномный метод, мы идентифицировали возможный ДНК-связывающий мотив PckR и реконструировали регулоны PckR во всех перечисленных геномах (Рис. 12). Более того, основываясь на расположении связывающих сайтов в промотерных областях, мы предсказали, активирует или репрессирует PckR члены своего регулона. По итогам нашего анализа мы предсказываем, что PckR работает как многофункциональный транскрипционный регулятор, репрессируя гликолитические гены из путей Эмбдена-Мейерхофа и Энтнера-Дудорова (*glk*, *fba*, *pykA*, *zwf-pgl-edd*, *eda*) и активируя гены глюконеогенеза и цикла Кребса (*pckA*, *mdh-sucCDAB*, *sdhABCD*). 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). В большинстве изучаемых геномов гены регулятора *pckR* не соседствуют с регулируемым геном на хромосоме; это свойственно многим глобальным регуляторам у бактерий. Однако, в геноме *Xanthobacter autotrophicus* *pckR* расположен дивергентно относительно гена 6-фосфоглюконат дегидрогеназы *gnd*; общая регуляторная область этих двух генов содержит возможный сайт связывания PckR. Мы обнаружили ортологи PckR и их

возможные сайты связывания в бактериях из семейства *Rhizobiales*, но не обнаружили их в

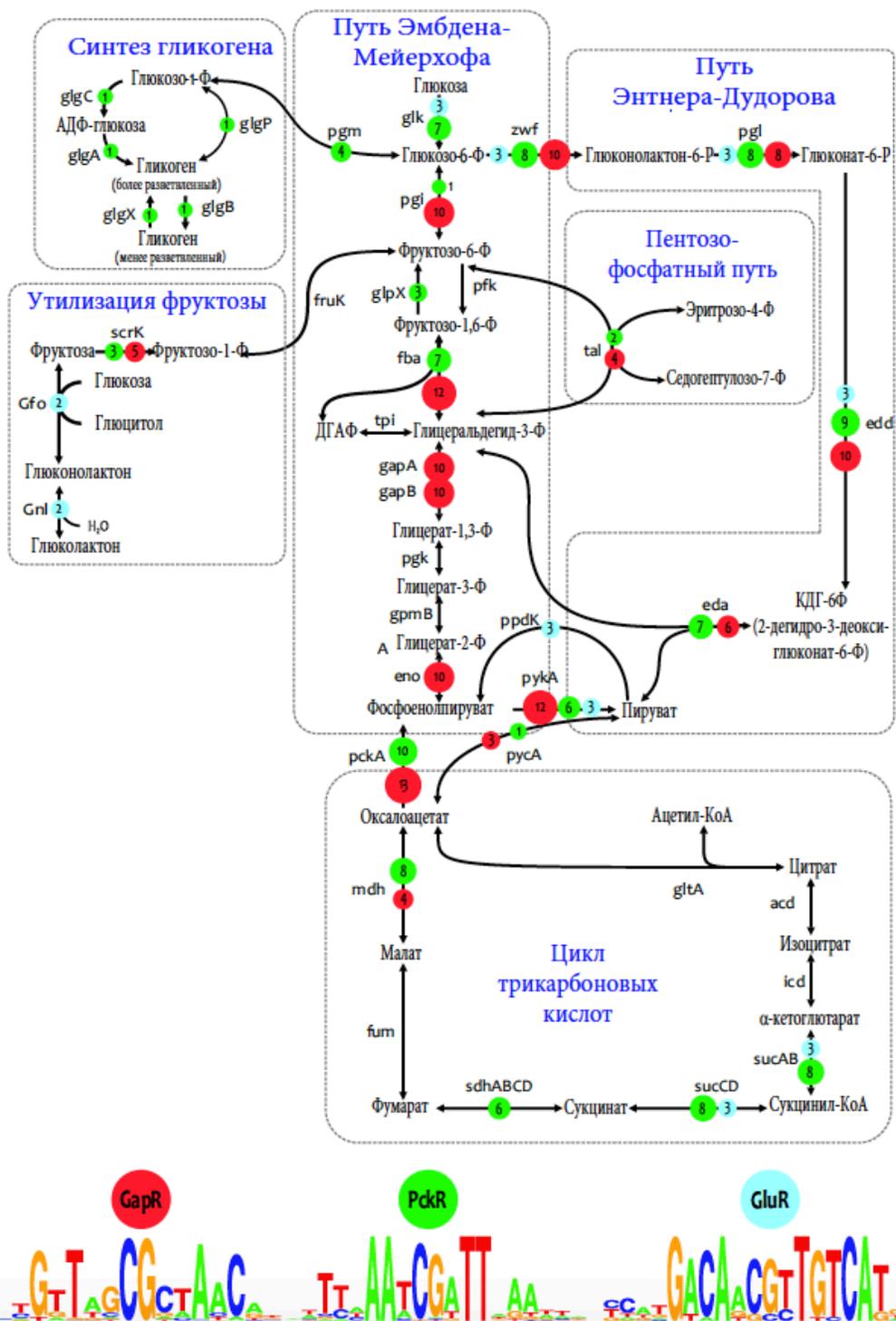


Рис. 12 Предсказанные глобальные регулоны центрального углеводного метаболизма у *Alphaproteobacteria*. (A) Регулон *GapR* у *Rhodobacteraceae*. (B) Регулон *PckR* у *Rhizobiales*. (C) Регулон *GluR* у *Caulobacter spp.* Названия регулируемых генов и оперонов показаны рядом со стрелками. Цифры в кружках показывают число геномов, в которых данный ген/оперон регулируется соответствующим ТФ

бактериях из других семейств, что позволяет предположить, что регулон PckR появился относительно недавно в процессе эволюции альфа-протеобактерий. Регулятор PckR у *Rhizobiales* можно рассматривать как частичную функциональную замену регуляторам Cra/FruR (см. выше) у *Enterobacteria* и HexR у *Shewanella* [126]. Оба регулятора модулируют направление потока углерода через различные пути углеводного метаболизма.

После публикации нашей работы вышла работа, в которой регулятор PckR был изучен экспериментально у *Sinorhizobium meliloti*; в этой работе многие наши предсказания были подтверждены экспериментально [154]. В частности, при помощи анализа дифференциальной экспрессии генов у бактерий - мутантов по PckR и у нормальных бактерий было показано, что PckR контролирует экспрессию генов центрального углеводного метаболизма у *Sinorhizobium meliloti*. В частности, PckR репрессирует гены гликолиза *zwf-pgl-edd*, *eda2* и *mgsA* и активирует гены глюконеогенеза *pckA* и *fba*. Это полностью соответствует нашим предсказанием за единственным исключением гена *fba*: в соответствии с расположением связывающего сайта PckR в промоторе этого гена, мы предсказывали его репрессию PckR, однако экспериментальные данные говорят о его активации. Далее, при помощи замедления ДНК в геле было показано, что PckR связывает предсказанные сайты с мотивом TTTMAATCGATTWAAA *in vitro*. Кроме того, авторы исследования предположили, что эффектором, индуцирующим PckR является фосфоенолпируват, и показали, что связывание PckR с фосфоенолпируватом изменяет специфичность связывания PckR с ДНК (тоже при помощи замедления ДНК в геле). В целом, результаты этой работы подтверждают и уточняют наши предсказания.

Новый глобальный регулятор генов сахарного метаболизма GapR (ген *RSP_1663* в геноме *Rhodobacter capsulatus*) был найден в 13 изученных геномах семейства *Rhodobacteraceae* (Рис. 12). Мы предсказали возможный мотив связывания GapR; его консенсусная последовательность представляет собой палиндром длины 20. Этот мотив отличается от мотива связывания PckR. 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-pgl*, *pgi*, *edd-eda*, *fba*, *gapA*, *gapB*, *eno*, *pykA*), глюконеогенеза (*pckA*, *pusA*), пути утилизации фруктозы (*scrK*), пентозо-фосфатного пути (*tal*) и цикла трикарбоновых кислот (*mdh*). Похожий мотив 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. Геном *Roseobacter* содержит две копии гена *gapR*. Один из этих паралогов расположен дивергентно относительно гена *gapB*; регуляторная область этих двух генов содержит возможный сайт связывания GapR.

Роль регулятора GapR была также позднее подтверждена в независимом исследовании [155]. Авторы данного исследования, вероятно, не знали о нашей работе и идентифицировали регулятор GapR независимо, назвав его CseR. В данном исследовании, авторы подтвердили ряд наших предположений о GapR, используя в качестве модельного организма *Rhodobacter sphaeroides*. Во-первых, штаммы, мутантные по GapR, росли значительно медленнее нормальных организмов на многих субстратах - источниках углерода, что может поддерживать гипотезу о роли GapR как регулятора центрального углеводного метаболизма. В пользу этой гипотезы также говорят и данные по анализу дифференциальной экспрессии генов у бактерий - мутантов по GapR и у нормальных бактерий. Эти данные указывают на то, что GapR контролирует экспрессию многих генов центрального углеводного метаболизма. В частности, 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*). В процессе реконструкции регулонов GapR мы смогли идентифицировать лишь часть из этих генов как регуляторные мишени GapR, а именно - *mdh*, *pgi*, *fba*, *pckA*, *zwf*, *pgl* и *eda*. При помощи замедления ДНК в геле было показано, что GapR связывает предсказанные сайты с мотивом YGTT-N6-AACW *in vitro*. Также, авторы предположили, что в роли эффектора GapR выступает 6-фосфоглюконат, и показали, что связывание с 6-фосфоглюконатом *in vitro* уменьшает специфичность связывания CseR с ДНК. Таким образом, эксперименты, проведенные в данном исследовании, подтверждают наше предположение о функциональной роли GapR и показывают ранее неизвестные особенности механизма его регуляции.

Несмотря на то, что PckR и GapR регулируют пересекающиеся множества генов в двух разных группах альфа-протеобактерий, эти регуляторы не ортологичны друг другу (Рис. 14) и связываются с различными ДНК-мотивами (Рис. 12). В альфа-протеобактериях из семейства *Caulobacteraceae* нами был обнаружен еще один новый глобальный регулятор из семейства LacI, названный GluR (CC2053 в *Caulobacter crescentus*) (Рис. 12). GluR не является ортологом описанным выше регуляторам PckR и GapR и распознает консервативную палиндромную последовательность длиной 20 нуклеотидов, которая отличается от мотивов PckR и GapR. Реконструированный регулон GluR в *Caulobacter* spp. включает в себя опероны гликолитических генов *zwf-pgl-edd-glK*, *pykA* и *gnl-gfo*, а также ген глюконеогенеза *ppdK*,

кодирующий пируват-фосфат дикиназу и оперон *sucABCD*, кодирующий гены цикла трикарбоновых кислот. Ген регулятора *gluR* расположен сразу после регулируемого оперона *zwf-pgl-edd-glk* и содержит в своей регуляторной области связывающий сайт GluR во всех трех проанализированных геномах *Caulobacter*. Мы не нашли ортологов GluR в других альфа-протеобактериях. Молекулярный эффектор GluR не известен. Основываясь на схожести GluR и предсказанных репрессоров BglR у альфа-протеобактерий, 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* [156], что подтверждает нашу гипотезу об эффекторе GluR.

Обсуждение

Мы использовали методы сравнительно-геномной реконструкции бактериальных ТФ из семейства LacI. Этот выбор был обусловлен несколькими соображениями. Во-первых, семейство LacI – это большое, разнообразное и широко распространенное среди бактерий семейство ТФ. Во-вторых, белки из этого семейства обладают постоянной доменной структурой и консервативной структурой сайтов связывания. Результатом данного исследования является детальная реконструкция 1281 регулонов ТФ из семейства LacI в геномах 272 бактерий. Большинство (90%) реконструированных регулонов – локальные регулоны, т.е. каждый из них контролирует лишь небольшое число генов и оперонов, участвующих лишь в одном метаболическом пути. Однако, некоторые из LacI-регулонов – глобальные регулоны, т.е. каждый из них контролирует от десятков до сотен генов, участвующих во множестве различных метаболических путей. Вдобавок к реконструкции ранее известных глобальных регуляторов, таких как FruR/Cra, PurR и CsrA, мы идентифицировали три новых регулятора центрального углеводного метаболизма у альфа-протеобактерий, а именно PckR, GluR и GapR, и реконструировали их регулоны. Мы также реконструировали эволюционную историю глобального регулона FruR, который наиболее вероятно произошел из локального регулона у предковых групп гамма-протеобактерий и постепенно расширился в ходе эволюции энтеробактерий. Филогенетический анализ ТФ из семейства LacI показал множество примеров различных эволюционных процессов, включая дивергентную эволюцию (разделение функций и специфичности связывания ТФ после

дубликации), конвергентную эволюцию (появление одинаковых функций в ходе независимых процессов в разных частях дерева), и образование парафилетических групп (появление новых функций и специфичностей, нетипичных для данной ветки филогенетического дерева). Ниже мы обсуждаем эти эволюционные процессы в деталях и показываем примеры функциональной диверсификации в семействе LacI.

Анализ филогенетического дерева TF из семейства LacI показывает, что некоторые ортологичные группы TF соответствуют таксономии (TF регулируют ортологичные гены и распознают похожие мотивы), однако некоторые из таких веток содержат внутреннюю подгруппу либо с отличающейся структурой мотива сайтов связывания, либо с отличающейся специфичностью к эффектору, либо с отличающимся содержимым регулона, т.е. эти TF формировали так называемые парафилетические группы. Наиболее любопытный пример парафилии – это ветка рибозных репрессоров RbsR у β - и γ -*Proteobacteria*, которая содержит вложенную ветку пуриновых репрессоров PurR.

Филогенетический анализ показал наличие ортологов PurR только в трех порядках бактерий, а именно *Enterobacteriales*, *Pasteurellales* и *Vibrionales*. Ближайшие паралоги PurR в этих группах – это рибозные репрессоры (RbsRs) (Рис. 14). Скорее всего, PurR произошел при дубликации RbsR у предковой гамма-протеобактерии, а ее предшественники имели лишь единственный RbsR репрессор. Ортологи RbsR из трех вышеупомянутых порядков сохраняют специфичность к лиганду, но мотивы их связывающих сайтов слегка изменены по сравнению с RbsR порядков *Pseudomonadales* и β -*Proteobacteria* (Рис. 13). Регуляторы PurR, напротив, сохраняют мотив, но их специфичность к лиганду и основной режим регуляции изменены (PurR связывает ДНК в присутствии лиганда, тогда как большинство регуляторов из семейства LacI связывают ДНК в отсутствие своих лигандов).

Этот пример показывает возможный путь образования парафилетических веток посредством дубликации предкового гена с последующим расхождением образовавшихся копий, приводящим к появлению нового регулятора. В случае RbsR и PurR, дубликация и расхождение случились, скорее всего, очень давно, и мы можем наблюдать лишь результаты их эволюции, но не можем наблюдать промежуточные состояния. Однако, в случае регулона утилизации альфа-глюкозида AglR и регулона утилизации трегалозы ThuR мы можем наблюдать такие промежуточные состояния. Регуляторы AglR в порядках *Rhizobiales* и

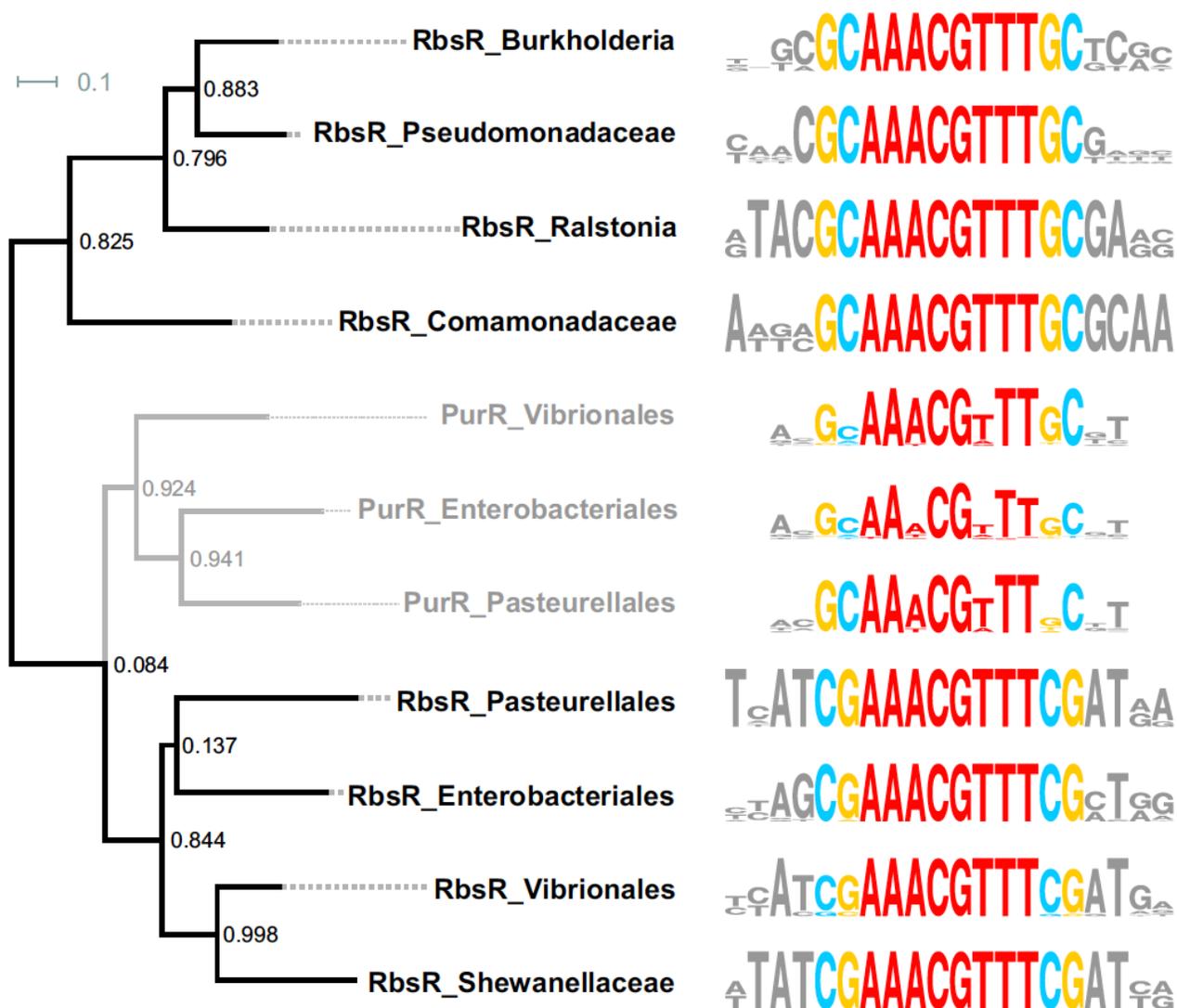


Рис. 13 Мотивы сайтов связывания регулонов *PurR* и *RbsR* у бета- и гамма-протеобактерий. Те позиции, которые консервативны во всех показанных мотивах, отмечены красным. Позиции, специфичные для отдельных веток, отмечены синим (консенсус – C) и желтым (консенсус – G). Неконсервативные позиции отмечены серым.

Rhodobacteriales формируют парафилетическую ветку со вложенной веткой, содержащей регуляторы *ThuR* в *Rhizobiales* (Рис. 13). Оба регулятора (*AglR* и *ThuR*) – локальные, каждый из них контролирует экспрессию гена самого регулятора и еще одного оперона, располагающихся дивергентно друг относительно друга. Регулируемые опероны содержат гомологичные гены, кодирующие киназы и АВС-транспортеры, а также негомологичные гены гидролаз. Мотив сайтов связывания *ThuR* (natchAAAAnCGnTTTngatt) отличается от мотива сайтов связывания *AglR* (nnntcAAAGCGCTTTgannn). Таким образом, в процессе

диверсификации ThuR поменял и специфичность ДНК-связывающего мотива, и специфичность связывания к лиганду. Парафилетическая группа CelR в *γ-Proteobacteria* содержит вложенную группу, содержащую регуляторы AscG в *Enterobacteriales*. В случае регуляторов CelR и AscG, их мотивы сайтов связывания и набор регулируемых генов полностью разделились после дупликации, но специфичность к эффектору (целлобиозо-6-фосфат) и регулируемый метаболический путь (утилизация целлобиозы) остались неизменными.

Ветки филогенетического дерева анализируемых ТФ из семейства LacI (Рис. 13) удалены друг от друга, что усложняет реконструкцию их эволюционной истории. Однако, в некоторых случаях мы можем наблюдать две или более близкие друг к другу ветки, соответствующими двум ортологичным группам. Интуитивно, появление таких двух групп можно объяснить событием дупликации предкового ТФ с последующим разделением копий. Дерево ТФ из семейства LacI содержит множество примеров монофилетических таксон-специфичных веток, состоящих лишь из одной группы бактерий каждая. Одна из таких веток содержит регуляторы UxaR, UxuR и KdgR в *Bacteroides*, контролирующие пути катаболизма галактуроната, глюкуроната и 2-кето-3-деоксиглюконата, соответственно. Две другие монофилетические ветки содержат (1) регулятор катаболизма коджибиозы KojR и регулятор катаболизма неизвестного сахара Caug_3448 из *Chloroflexus* и (2) рибозный регулятор RbsR и уридиновый регулятор UriR из *Corynebacteria*. Ранее, похожий сценарий был описан для нескольких регуляторов из семейства ROK в бактериях из группы *Thermotogales* [157].

Другие примеры дивергентной эволюции специфичности регуляторов могут быть продемонстрированы соседними ветками на филогенетическом дереве. Так, репрессор катаболизма идоната IdnR в *Enterobacteriales* и репрессор катаболизма глюконата GntR в многих группах *γ-Proteobacteria* являются ближайшими (по филогенетическому дереву) паралогами друг друга (Рис. 13). В регулонах IdnR, гены *idnK* и *idnT* являются ближайшими гомологами генов *gntK* и *gntU*, контролируемых регулятором GntR. Таким образом, дупликация повлияла не только на гены регуляторов, но и на регулируемые гены. Тот факт, что GntR может также распознавать связывающие сайты IdnR в регуляторных областях *idnK* и оперона *idnDOTR* в *E. coli* [158], также подтверждает возможность недавней дупликации этих регуляторов. Структурное сходство сахаров-эффекторов GntR и IdnR также указывает на недавнюю дупликацию и последующую специализацию IdnR в *Enterobacteriales*.

Другой возможный сценарий диверсификации регулонов состоит в дупликации предкового эффектора с последующим расширением одного из новообразовавшихся регулонов на новые

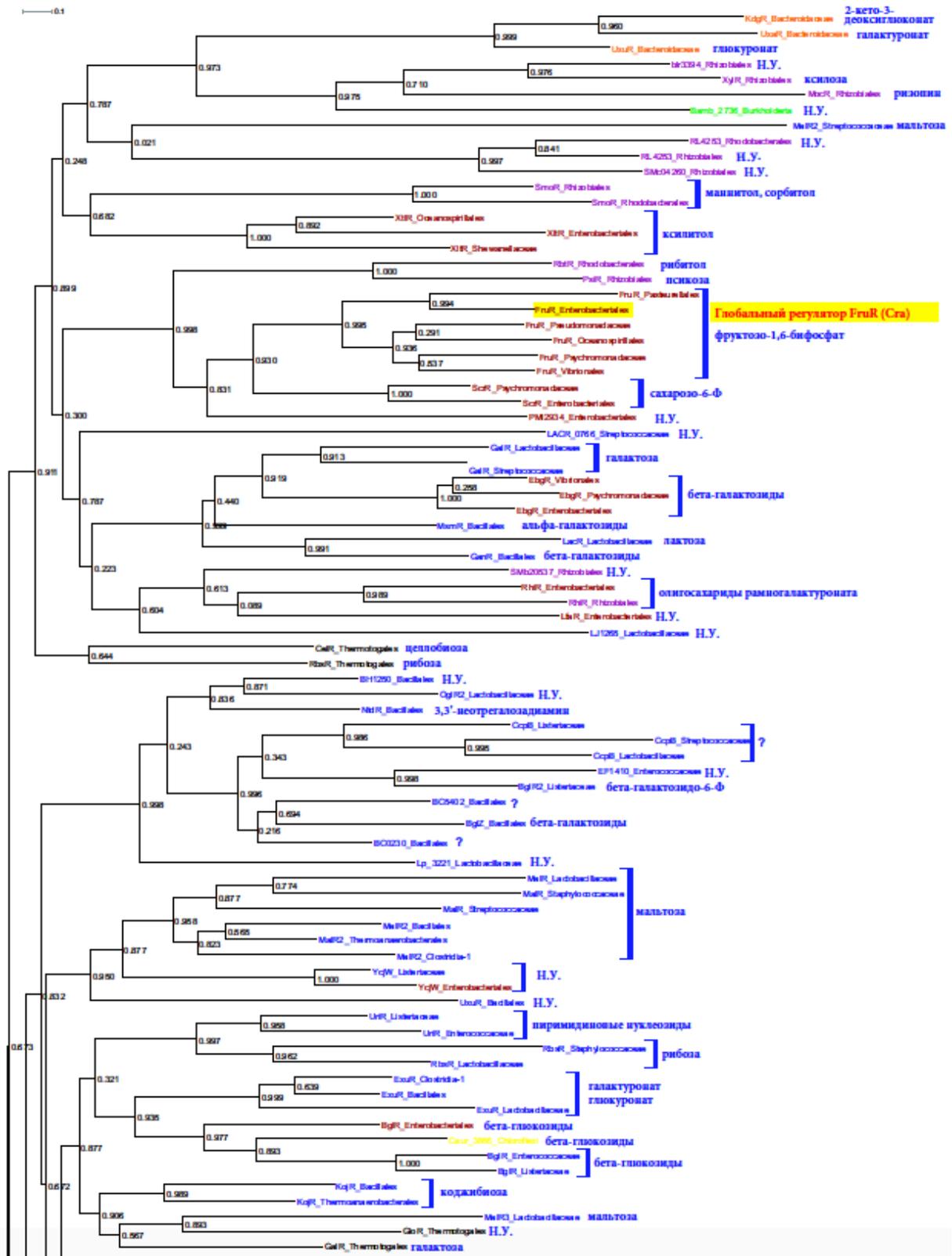
гены и, соответственно, с приобретением специфичности ТФ к новому эффектору. В качестве примера такого сценария можно привести ветку, содержащую регуляторы катаболизма фруктозы (FruR) и сахарозы (ScrR) в *γ-Proteobacteria*. Сахароза – это фруктозосодержащий дисахарид. Вместе с приобретением новой регуляторной функции у одного из этих регуляторов также поменялась специфичность связывания ДНК. При этом, эффекторы этих двух регуляторов (фруктозо-6-фосфат для FruR и сахарозо-6-фосфат для ScrR) похожи друг на друга структурно.

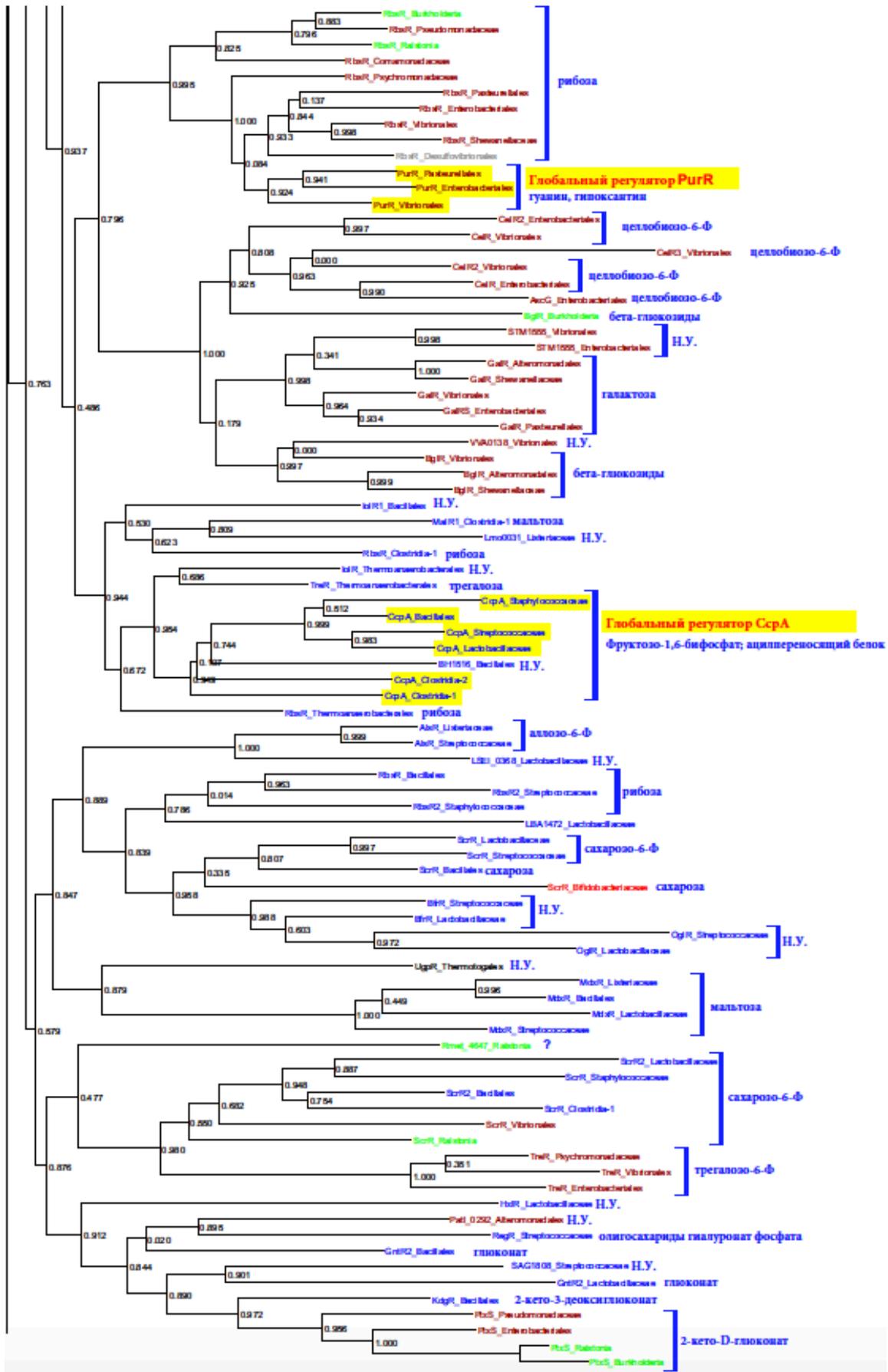
Основываясь на анализе парафилетических веток и соседних монофилетических веток на филогенетическом дереве ТФ из семейства LacI, мы описали три основных способа появления регуляторов с новыми функциями. Первый способ – это одновременная дупликация гена регулятора и оперона, содержащего регулируемые гены, как произошло в случае ThuR в *Rhizobiales*. В процессе диверсификации, и ТФ, и регулируемые гены меняют специфичность, некоторые регулируемые гены могут быть потеряны, и новые гены могут быть включены в составе регулона. Второй способ – это дупликация регулятора с последующим приобретением новых регулируемых генов, как в случаях PurR в *γ-Proteobacteria* и AscG в *Enterobacteriales*. Третий способ более редкий: единственный пример, который мы нашли в ходе нашего исследования – это регулон SCO5692 в *Streptomyetaceae*. В этом случае, новые специфичности ТФ появляются без дупликации, возможно в результате потери части регулируемых генов.

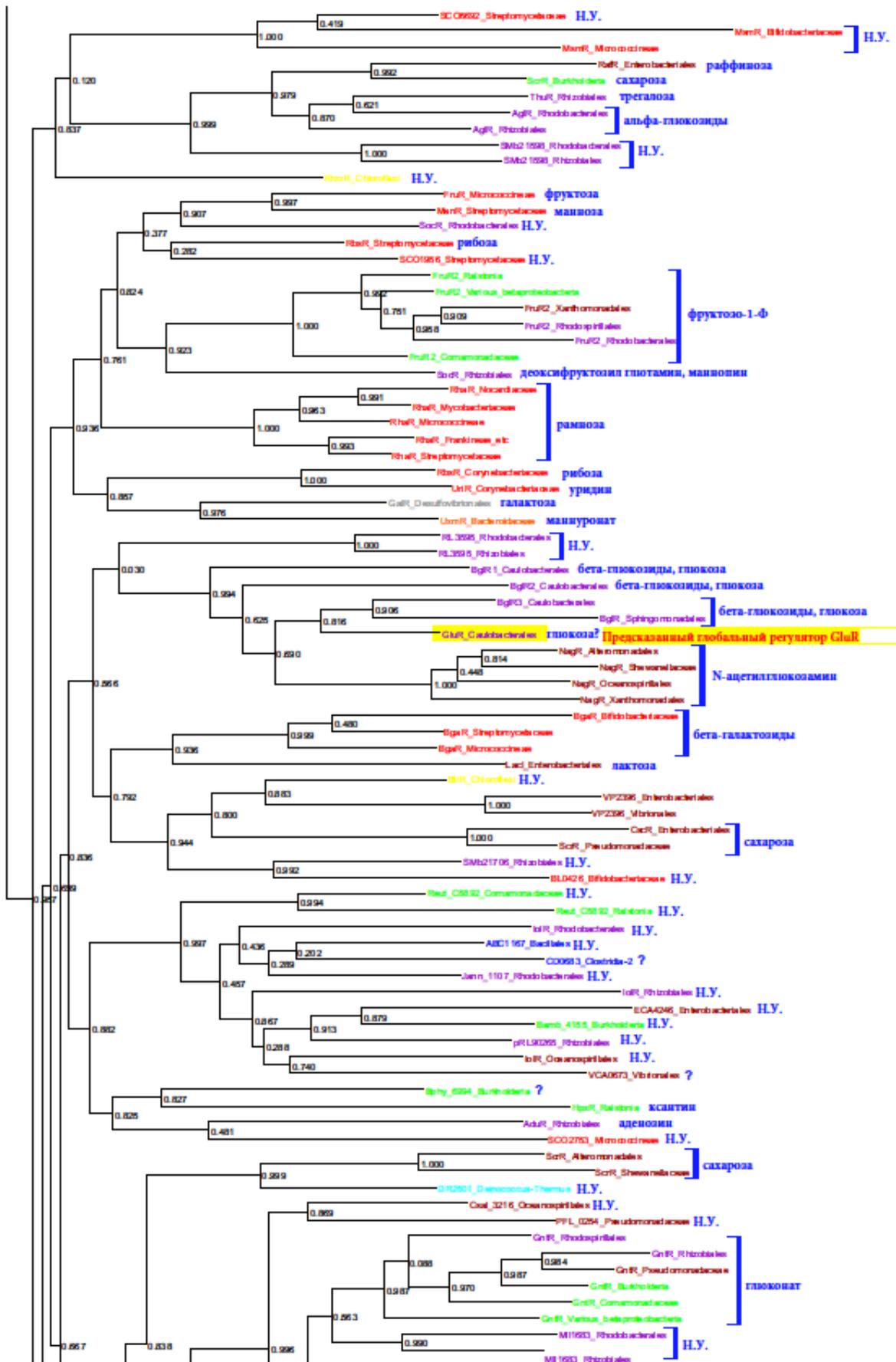
Есть три основные характеристики регулона, которые могут меняться в процессе приобретения новых функций. Это (1) набор регулируемых генов, (2) специфичность к эффектору, (3) структура мотива связывающих сайтов. Для большинства событий появления ТФ с новой функцией, мы наблюдали изменение двух из этих трех характеристик. Изменение всех трех характеристик встречается реже и наблюдается скорее в случае ТФ из давно отделившихся групп бактерий, таких, как *Thermotogales*. Скорее всего, в этих случаях изменение всех трех характеристик – это результат долгой эволюционной истории. Изменение лишь одной из этих характеристик наблюдается лишь для недавно дублировавшихся ТФ, для которых процесс расхождения еще не завершился.

Полученные нами данные по регуляторам семейства LacI могут предоставить множество примеров различных эволюционных процессов для регуляторов и соответствующих им регулонов. Эти данные находятся в открытом доступе в базе данных RegPrecise. В дальнейшем, эти данные могут быть использованы для нахождения ключевых аминокислотных остатков в ДНК- и лиганд-связывающих доменах ТФ и установления связи между этими остатками и предпочитаемыми нуклеотидами в сайтах связывания ДНК, что приведет к разработке более

точных методов реконструкции регулонов ТФ и, возможно, к разработке новых синтетических регуляторов.







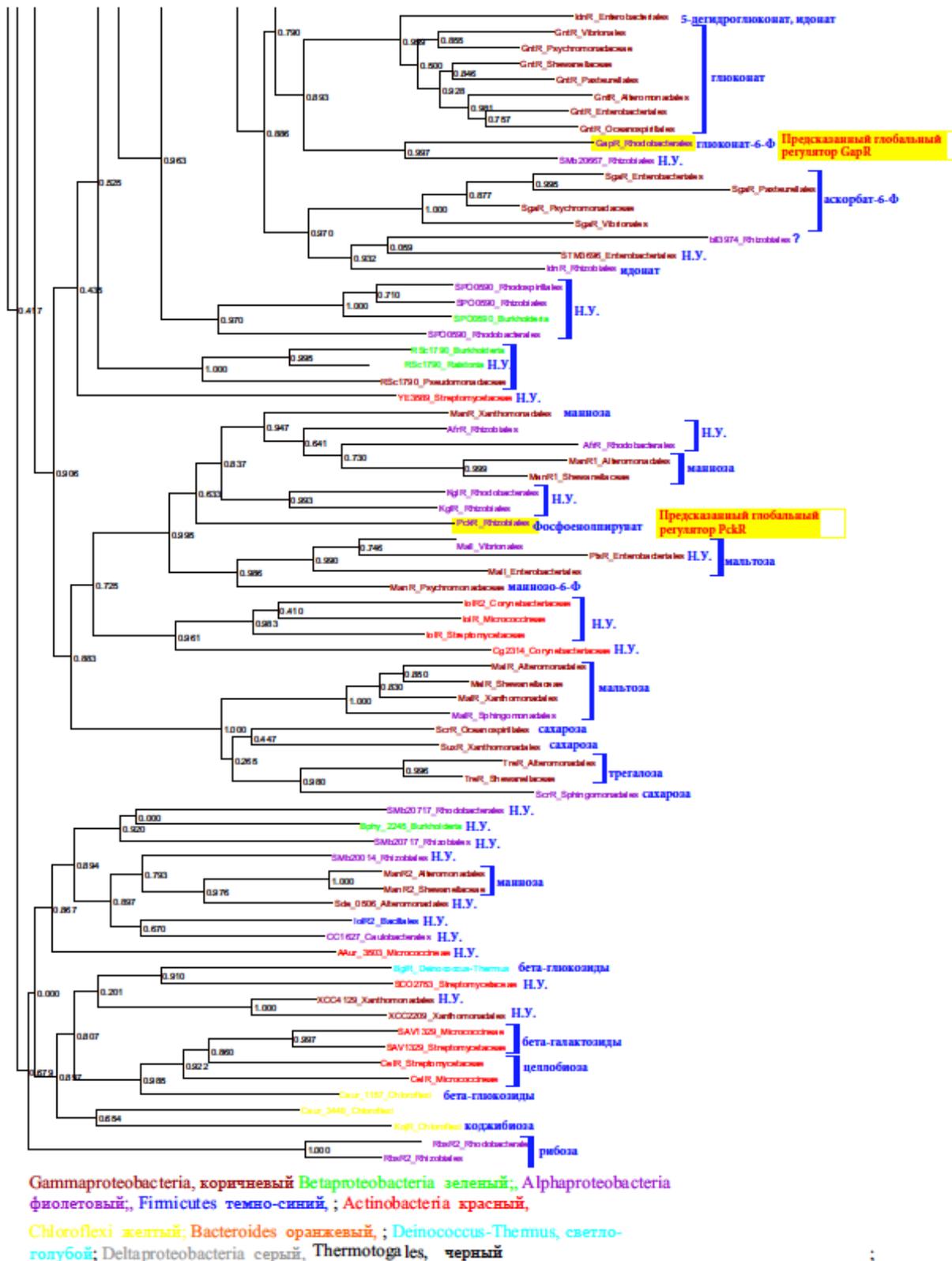


Рис. 14. Филогенетическое дерево максимального правдоподобия для 1324 ТФ из семейства LacI, распределенных среди 321 реконструированного регулога. Каждый регулог (группа ортологических ТФ в одной таксономической группе) представлен одним ТФ на данном дереве. Дерево было построено программой PhyML с 1000 бутстрепами. Номера в

узлах показывают значения бутстрепов. Полный список изученных ТФ, регулогов и геномов доступен в базе данных RegPrecise. Предсказанные эффекторы показаны квадратными скобками. Неизвестные углеводы обозначены как U.S. Глобальные регуляторы подсвечены желтым. Регулоги обозначены различными цветами по таксономическим группам: Gammaproteobacteria - коричневый; Betaproteobacteria - зеленый; Alphaproteobacteria - фиолетовый; Firmicutes – темно-синий, Actinobacteria – красный, Thermotoga - черный; Chloroflexi – желтый, Bacteroides – оранжевый, Deinococcus-Thermus – светло-синий, Deltaproteobacteria – серый.

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

Методы

Составление выборки геномов-представителей микробиома кишечника человека

Мы построили выборку геномов, представляющих различные виды бактерий – обитателей микробиома человеческого кишечника, следующим образом. Во-первых, мы использовали список из 194 геномов таких бактерий, составленный консорциумом MetaHIT в 2010 году [159, p. 20]. Этот список включает в себя 151 геном, отсеквенированный в рамках Human Microbiome Project (HMP), 17 геномов, отсеквенированных консорциумом MetaHIT и 26 геномов, доступных в базе данных GenBank. К этому списку мы добавили 450 геномов бактерий, выделенных из кишечного тракта, отсеквенированных и собранных в базе данных Human Microbiome Reference Genome Database в 2012 году (<https://hmpdacc.org/hmp/HMRGD/>) [91]. Далее, мы проанализировали коллекцию из ~1000 культивированных видов из микробиома человеческого кишечника, опубликованную в 2014 году [160]. Для каждого генома из этой коллекции мы нашли геномы нескольких родственных бактерий в базе данных геномов the PATRIC (<https://www.patricbrc.org>) [161]; всего мы выбрали около 2000 геномов (как полных геномов, так и неполных геномов высокого качества). Совместив вышеуказанные коллекции

геномов, мы получили список из 2228 геномов, который мы использовали в рамках данного исследования в качестве выборки представителей кишечного микробиома. Мы составили выравнивания 11 рибосомальных белков (L5, L6, L9, L10, L15, L20, S2, S4, S5, S6, S8) из 2228 геномов, сшили их вместе и использовали полученное единое выравнивание, чтобы составить филогенетическое дерево анализируемых видов. Для построения множественных выравниваний мы использовали программу Muscle [162]. Для построения и визуализации филогенетических деревьев максимального правдоподобия мы использовали программы RAxML version 8 [163] и iTOL [164], соответственно.

Геномная реконструкция подсистем метаболизма витаминов

Для реконструкции метаболических путей и предсказания фенотипов в геномах из изучаемой выборки мы использовали онлайн-платформу для интеграции и анализа геномов SEED (<http://www.theseed.org>) [49], [114]. Эта платформа построена на использовании метода подсистем. Мы описали метаболические пути биосинтеза, транспорта и сохранения каждого из восьми изучаемых витаминов/кофакторов при помощи соответствующей подсистемы в SEED. При разработке подсистем мы основывались на ранее опубликованных исследованиях геномного метаболизма [165], [166], [167], [168], [169], [170], [171], [172], [173]. Кроме того, мы использовали базу данных метаболических путей KEGG [174] для независимого подтверждения. Для реконструкции метаболических путей в SEED мы проводили функциональную аннотацию генов, используя уже известные аннотации, а также критерии гомологичности и схожести геномного контекста. Применяя эти критерии, мы различали паралогичные гены со схожими, но различными функциями (чаще всего – транспортеры) [175], и заполняли пропуски («отсутствующие гены») в метаболических путях [176], [177]. Три критерия анализа геномного контекста таковы: (1) кластеризация генов на хромосоме (опероны), (2) ко-регуляция генов общим регулятором или рибосвичом (см. ниже), и (3) встречаемость генов в родственных геномах [178], [29], [179].

У витаминов B1, B2, B9 и B12 метаболические и транспортные гены регулируются специфичными рибосвичами [180], тогда как метаболические гены витаминов B3, B6 и B7 регулируются ДНК-связывающими транскрипционными факторами из семейств BirA, BioQ, NrtR, NiaR и PdxR [181], [170], [182], [183]. Для поиска рибосвичей *in silico* мы сканировали геномы при помощи программы Infernal [184] и ковариационных моделей из базы данных Rfam

[185]. Для предсказания сайтов связывания ТФ мы скачивали позиционные матрицы весов из базы данных RegPrecise [48] и искали новые сайты в изучаемых геномах при помощи программ поиска, доступных на сервере RegPredict [107]. Реконструкция регулонов позволила классифицировать неточно проаннотированные транспортеры витаминов из больших семейств (например, транспортеры из семейства ABC), обладающие специфичностью к различным субстратам.

Варианты метаболических путей – это биохимически и физиологически устойчивые функциональные модули, представленные определенными комбинациями ферментов, катализирующих связанные реакции [186]. Мы установили фенотипические правила, описывающие различные варианты метаболических путей для каждого витамина/кофактора, основываясь на диаграммах метаболических путей (Приложение 2) и на распределениях метаболических ферментов и транспортеров в изучаемых геномах (Приложение 1). Варианты метаболических путей были представлены наборами функциональных ролей, необходимых для биосинтеза витамина/кофактора (или для сохранения витамина/кофактора после захвата метаболических предшественников из окружающей среды). Используя фенотипические правила, мы смогли различить прототрофные и ауксотрофные организмы, основываясь на наличии или отсутствии ферментов, необходимых для биосинтеза витаминов/кофакторов. Фенотипические правила для прототрофов по витаминам В1, В3, В6 и В12 включают в себя по два альтернативных пути биосинтеза (см. ниже). Мы классифицировали ауксотрофы по витаминам В1, В3, В5 и В7 более детально, в соответствии с наличием или отсутствием у них ферментов для сохранения витаминов или их метаболических предшественников. Например, мы классифицировали ауксотрофы по тиамину в зависимости от того, кодируют ли их геномы ферменты сохранения витамина В1 и его метаболических предшественников - гидроксиметилпиримидин (ГМП) и гидроксипиридин (ГЭТ). Мы также разработали индивидуальные варианты для описания метаболических путей с одним или двумя отсутствующими ферментами. Кроме того, мы классифицировали изучаемые организмы по вариантам путей сохранения витаминов, основываясь на наличии или отсутствии витамин-специфичных транспортеров и ферментов, используемых для сохранения витаминов, таких как киназа тиазола ThiM или фосфорибозилтрансфераза никотината PncB (Приложение 2). Мы классифицировали организмы, способные сохранять витамины В1, В3 или В5 более детально, в зависимости от наличия в их геномах транспортеров для соответствующих метаболических предшественников, таких как ГМП, ГЭТ, пантоат и никотинамид рибозид (НР). Наконец, мы перевели назначенные различным организмам фенотипы в бинарную форму с

соответствующими значениями для прототрофов («1») и ауксотрофов («0») и собрали их вместе в таблицу бинарных фенотипов для 8 витаминов и 2228 геномов (данные не показаны).

Предсказания уровней ауксотрофии образцов человеческого микробиома.

Мы использовали данные секвенирования генов 16S рРНК из образцов кишечного микробиома, полученных в ходе нескольких исследований (описание каждого исследования см. в секции Результаты). В этих исследованиях из собранных образцов была выделена бактериальная ДНК, проведено обогащение последовательностями 16S рРНК при помощи полимеразной цепной реакции, и последовательности полученной ДНК были расшифрованы при помощи секвенирования нового поколения. Далее, была проведена первичная обработка полученных наборов последовательностей. Были отфильтрованы последовательности, не проходящие по порогу качества, и все оставшиеся последовательности были сгруппированы в операционные таксономические единицы (ОТЕ). В исследованиях [187], [188], [98] для группировки последовательностей по ОТЕ использовали QIIME [189] с порогом схожести последовательностей 97%, в [190] использовали DOTUR [191] с порогом схожести последовательностей 99%. Таким образом, в результате первичной обработки данных для каждого образца были доступны: (1) набор ОТЕ, найденных в данном образце, (2) набор уникальных последовательностей, относящихся к каждому ОТЕ, (3) “встречаемость” различных ОТЕ. Предположительно, каждый ОТЕ соответствует одной группе родственных бактерий, причем размер этой группы зависит от используемого порога сходства последовательностей. Чаще всего используется пороговое значение в 97% сходства последовательностей; при этом размер групп бактерий, определяемых ОТЕ, примерно соответствует делению бактерий на виды [192]. Далее, “встречаемость” ОТЕ вычисляется, как число последовательностей 16S рРНК, относящихся к данному ОТЕ, деленное на общее число последовательностей 16S рРНК в данном образце. Таким образом, “встречаемость” ОТЕ отражает численную долю бактерий из данной группы в изучаемом бактериальном сообществе.

В качестве начальных данных для анализа уровня ауксотрофии образцов микробиома мы использовали первично обработанные данные, т.е. (1) встречаемость ОТЕ для каждого образца и (2) последовательности, относящиеся к каждому ОТЕ. По результатам проведенной реконструкции витаминных фенотипов для бактерий из кишечного микробиома мы

наблюдали, что у большинства бактерий фенотипы консервативны на уровне видов. Поэтому мы предположили, что бактерии, относящиеся к одному ОТЕ, обладают похожими фенотипами. Для того, чтобы поставить в соответствие каждому ОТЕ один фенотип по каждому витамину, мы искали представителей данного ОТЕ среди изучаемой выборки геномов. Мы создали базу данных генов 16S рРНК из исследуемых нами геномов с помощью программы RNAmmer [193]. Далее, для каждого ОТЕ мы искали бактерии со схожими последовательностями генов 16S рРНК в созданной базе данных при помощи программы BLAST [116]. Мы сопоставляли каждому ОТЕ фенотип генома с наибольшим сходством по последовательности, но только в том случае, если сходство 16S рРНК гена из этого генома хоть с одним представителем ОТЕ превышало 97%.

Для вычисления уровня ауксотрофии мы рассматривали каждый образец микробиома как совокупность различных ОТЕ в различных пропорциях. Для большинства ОТЕ мы предсказали фенотипы по витаминам группы В, как описано выше. Мы называем суммарную долю таких ОТЕ с предсказанными фенотипами в каждом образце “покрытием” данного образца; покрытие отражает численную долю организмов в образце кишечного микробиома, для которых мы можем предсказать фенотипы по витаминам группы В. Отметим, что покрытие на зависит от конкретного витамина, поскольку для всех ОТЕ, представители которых были найдены в изучаемой выборке геномов, мы предсказываем фенотипы по всем изученным витаминам группы В. Мы вычисляли уровень ауксотрофии образца по определенному витамину как долю всех ОТЕ с фенотипом “ауксотроф”, нормализованную по встречаемости, среди всех ОТЕ, для которых были предсказаны фенотипы.

Результаты

Реконструкция путей биосинтеза и сохранения витаминов

Мы начали данное исследование с составления выборки бактериальных геномов, представляющих микробиом кишечного тракта (см. Методы). 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). Типы *Fusobacteria* и

Tenericutes представлены 41 и 25 геномами, соответственно. Оставшиеся пять типов (*Verrucomicrobia*, *Synergistetes*, *Spirochaetes*, *Lentisphaerae*, и *Planctomycetes*) были представлены от одного до пяти геномами каждый.

Для реконструкции путей биосинтеза и сохранения витаминов и кофакторов и для поиска транспортеров, импортирующих витамины, мы использовали сравнительно-геномный метод, основанный на использовании подсистем (см. Методы). Анализируемые функциональные роли включают 126 различных ферментов, участвующих в путях биосинтеза/сохранения восьми витаминов, и 83 транспортера (или их компоненты), участвующих в поглощении экзогенных витаминов и их метаболитических предшественников (Рис. 15). Детальные схемы, показывающие вариации метаболитических путей биосинтеза и сохранения витаминов и кофакторов на уровне индивидуальных ферментов (неортологичные замены) и топологий (альтернативные биохимические пути) представлены в Приложении 2. Основываясь на распределении ферментов и транспортеров из путей биосинтеза витаминов, мы (совместно с А. А. Арзамасовым) установили фенотипические правила, чтобы затем классифицировать анализируемые геномы в соответствии с вариантами метаболитических путей, которые они используют.

Фенотипические правила описывают набор генов, чье присутствие или отсутствие в анализируемом геноме позволяет точно определить используемый вариант метаболитического пути и предсказать возможные транспортируемые субстраты (Приложение 1). Для каждого витамина, мы разделили варианты метаболитических путей на две большие категории: прототрофы, которые могут синтезировать витамин сами (мы обозначаем их Р, от английского “prototrophs”), и ауксотрофы, которые не могут синтезировать витамин сами и которым поэтому приходится транспортировать витамины (или их метаболитические предшественники) из окружающей среды (мы обозначаем их А, от английского “auxotrophs”). Варианты метаболитических путей соответствуют альтернативным биосинтетическим путям, используемым прототрофами для синтеза витаминов (Рис. 15). Среди них: два альтернативных пути синтеза ГЭТ (предшественника тиамина), в т.ч. бактериальный путь (ThiFGHS) и эукариотический путь (Thi4); два пути синтеза квинолината (предшественника НАД), в т.ч. распространенный бактериальный путь (NadAB) и эукариотический путь (Tdo-Kmo-Kyn); два пути синтеза пиридоксальфосфата (PdxAJ и PdxST); и два пути синтеза пимелоил-КоА (предшественника биотина), основывающиеся на ферментах BioC и BioW. Наконец, кофермент В12 может синтезироваться либо через анаэробный, либо через аэробный путь биосинтеза корринового кольца, которые различаются тем, на какой стадии (ранней или

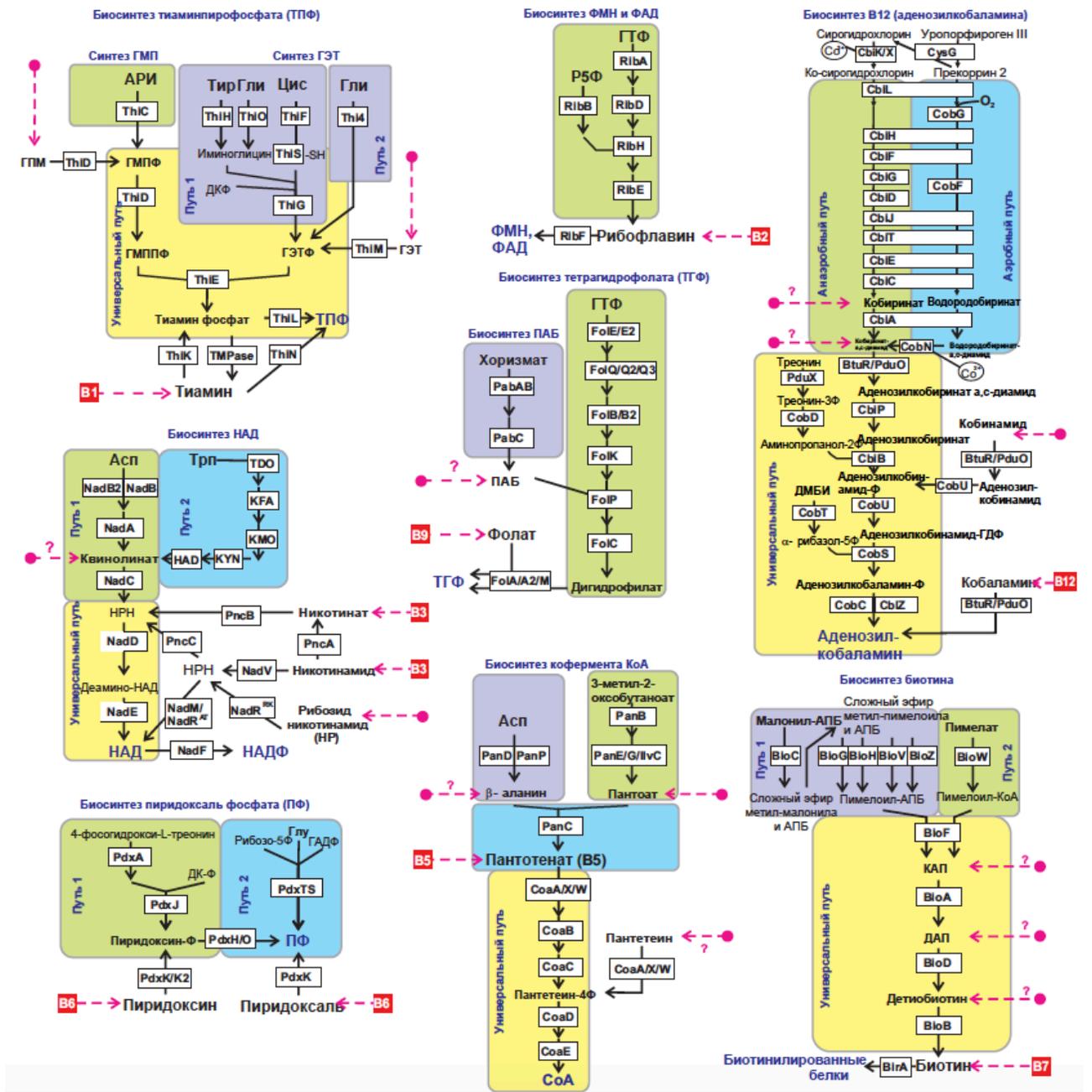


Рис. 15 Схема основных вариантов путей биосинтеза и сохранения витаминов группы В. Метаболические реакции обозначены черными стрелками, транспорт предшественников показан малиновыми стрелками; если соответствующий транспортер неизвестен, стрелка отмечена знаком вопроса. Каждый блок метаболических реакций, отличающийся между различными вариантами путей, отмечен фоновым цветом.

поздней) происходит включение кобальта специфическими хелатазами (CblK/X или CobN) и ферментами (CblDG or CobFG).

В геномах многих бактерий полностью отсутствуют ферменты для самостоятельного биосинтеза витаминов, однако присутствуют гены путей “сохранения” витаминов из окружающей среды и их последующего превращения в биологически активные кофакторы. Такие геномы обычно классифицируют как ауксотрофные. Несмотря на значительную вариабельность путей сохранения предшественников витаминов (см. ниже), в пользу такой вариабельности говорит то, что кофакторы, несомненно, необходимы для жизни бактерий (за исключением В12, см. ниже). Для большинства витаминов (кроме В6, см. ниже) ауксотрофы и прототрофы используют одни и те же конечные шаги путей превращения витаминов в кофакторы. Обычно это касается нескольких ферментов, выполняющих заключительные превращения в метаболическом пути. Единственным исключением является вариант ауксотрофии по коферменту А у паразитических видов *Mollicutes* (таких, как *Mycoplasma/Ureaplasma*), способных транспортировать готовый кофактор из окружающей среды (Приложение 1).

Для сохранения витаминов В1, В3 и В6 используются ферменты, не участвующие в основном пути биосинтеза и/или последующей трансформации витамина в кофактор. Для В1 это ферменты ThiK, ThiM и ThiN, для В3 - PncA, PncB и, реже, NadV, а для В6 - PdxK/K2. В случае других витаминов, те метаболические предшественники, которые могут быть транспортированы в клетку, сами являются интермедиатами биосинтетического пути соответствующего витамина, и потому не требуют дополнительных ферментов. Мы (совместно с А. А. Арзамасовым) провели полный анализ всех известных и предсказанных витаминных транспортеров и систем сохранения витаминов (см. ниже), что позволило лучше изучить метаболические потребности многих видов бактерий. Однако, наличие определенных генов сохранения витаминов не является необходимым признаком ауксотрофных бактерий: во многих ауксотрофных геномах мы не смогли обнаружить гены путей сохранения витаминов. Например, в 116 геномах - ауксотрофах по В6 мы не смогли идентифицировать киназу PdxK/K2, что позволяет предположить наличие альтернативного фермента. Кроме того, полное описание пути сохранения витаминов требует нахождения генов соответствующего транспортера, что представляет собой сложную техническую задачу из-за высокой эволюционной пластичности транспортеров и из-за нехватки экспериментальных сведений о них во многих таксономических группах бактерий.

Важно отметить, что пути сохранения витаминов часто (хотя и не всегда) присутствуют в прототрофных бактериях, что, вероятно, позволяет им вести оппортунистический образ жизни, переключаясь между энергетически затратным самостоятельным биосинтезом витаминов и

более экономичным (но менее универсальным по отношению к условиям окружающей среды) сохранением витаминов из окружающей среды. Мы включили описание таких метаболических свойств в общую схему реконструкции. Кроме того, известно, что ферменты сохранения витаминов часто используются для реутилизации кофакторов внутри клетки. Лучше всего это изучено на примере кофакторов НАД(Ф) [194]. В результате анализа транспортеров и ферментов биосинтеза и сохранения витаминов стало возможно идентифицировать потенциальных доноров витаминов в бактериальных сообществах (см. обсуждение ниже) [103].

Наконец, принимая во внимание то, что классификация ауксотрофных фенотипов, как правило, основана на отсутствии тех или иных генов, мы не можем исключать, по крайней мере теоретически, существования тех или иных неизвестных и неизученных биохимических путей. Однако, результаты проведенного анализа витаминных подсистем в тысячах бактериальных геномов не подтверждают этой гипотезы, поскольку за редкими исключениями мы не наблюдаем бактерий, которых нельзя было бы отнести к одной из нескольких известных категорий в соответствии с их метаболическими возможностями. Некоторые виды несут “неполные” варианты известных метаболических путей с одним или двумя отсутствующими генами (как обсуждается ниже). Стоит подчеркнуть, что разнообразие таких “неполных” вариантов крайне ограничено, что поддерживает предположение, что наше знание биосинтетических путей витаминов и кофакторов близко к насыщению.

Неполные пути биосинтеза кофакторов и сохранение витаминов

Одна из трудностей реконструкции метаболических путей состоит в том, чтобы отличить неполные прототрофные пути с “пропущенными” генами [176], 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 (в т.ч. 62 генома бифидобактерий) 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. 15). 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, в

т.ч. в 17 штаммах *Helicobacter pylori* и в 8 штаммах *Micrococcus luteus*, 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. Основываясь на этих наблюдениях, мы классифицируем некоторые пути как неполные прототрофные пути с “отсутствующими” генами, а не как ауксотрофные пути.

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. Наконец, в пути биосинтеза В9, у 454 штаммов недостает целого пути синтеза пара-аминобензоата (ПАБ, предшественника В9), тогда как все остальные ферменты биосинтеза фолата присутствуют в их геномах. Существование таких неполных путей можно объяснить одним из двух способов: (1) наличие неизученных путей и ферментов биосинтеза ПАБ, или (2) сохранение ПАБ из окружающей среды [195], [196], [197]. Мы решили классифицировать этот вариант как «Р*», то есть как неполный прототрофный путь (несмотря на схожесть данного сценария с неполными путями, классифицируемыми как ауксотрофные, для других витаминов), главным образом из-за того, что ПАБ нельзя рассматривать только лишь как альтернативный витамин - предшественник В9, т.к. ПАБ также является интермедиатом биосинтетических путей других важных метаболитов (например, ароматических аминокислот). 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 неканоничные метаболические предшественники соответствующих кофакторов (Приложение 1).

Наибольшее разнообразие таких вариантов наблюдается в биосинтезе тиамин пирофосфата (ТПФ). Синтез ТПФ происходит при химической реакции между фосфорилированным гидроксиметилпиримидином (ГМП) и гидроксипиримидином (ГЭТ), катализируемой тиаминфосфат синтазой ThiE, с последующим фосфорилированием, катализируемым киназой

ThiL, либо ThiN (Рис. 15) [166]. Примерно половина проанализированных геномов может синтезировать и ГМП и ГЭТ (прототрофы по В1), тогда как остальные ауксотрофные штаммы были поделены на 4 группы в зависимости от присутствия или отсутствия специфических путей сохранения витамина. Ауксотрофы первой группы (452 генома, 40%) не могут синтезировать ни один из этих предшественников, но их геномы содержат гены киназ, при помощи которых они, тем не менее, могут синтезировать тиаминпирофосфат ТПФ из предшественников ГМП и ГЭТ, сохраненных из окружающей среды. Две другие группы ауксотрофов содержат частичные пути биосинтеза ТПФ. Ауксотрофы по ГМП (199 геномов, 18%) могут синтезировать фосфорилированный ГЭТ (но не ГМП), и кодируют киназу, позволяющую сохранять ГМП из окружающей среды, тогда как у ауксотрофов по ГЭТ (114 геномов, 10%) нет путей биосинтеза ГЭТ, но есть пути киназа для сохранения экстрацеллюлярного ГЭТ. Оставшиеся ауксотрофы по В1 (355 геномов, 32%) могут сохранять лишь внешнеклеточный витамин В1, но не его предшественников (вариант Ahz в приложении 1).

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), наиболее распространенной формы, при помощи фермента PncB, и (2) никотиламида, через один из трех альтернативных путей: либо PncA-PncB, либо (реже встречается в бактериях) NadV-PncC, либо NadV-NadM, в обход ферментов NadD-NadE (которые являются обязательными ферментами во всех остальных вариантах путей биосинтеза никотина) (Рис. 15). Среди ауксотрофов по В3 (1012 геномов), мы нашли 86 штаммов из таких таксонов, как *Campylobacter*, *Corynebacterium*, *Lactobacillus* и *Streptococcus*, содержащие неполный вариант пути биосинтеза НАД (вариант Aq в приложении 1), 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 таком случае, эти бактерии могут получать НАД из двух источников: либо синтезируя НАД из квинолината, либо сохраняя ниацин (в форме никотината либо в форме никотиламида). Функциональный путь сохранения квинолината был ранее экспериментально подтвержден в бактерии *Streptococcus pyogenes*, однако транспортер квинолината найден не был [198]. Стоит отметить, что квинолинат является интермедиатом биосинтеза НАД не только в бактериях, но и в некоторых тканях человека, где НАД синтезируется при анаэробной деградации триптофана (кинурениновый путь). Другой редкий путь сохранения НАД - это синтез НАД из никотинамид рибозы (НР)

[194] при помощи фермента NadR. Этот путь является единственным способом биосинтеза НАД в *Haemophilus influenza* и в родственных видах из порядка *Pasteurellales*. Мы рассматриваем эти геномы как ауксотрофы по НР, а не как прототрофы по В3 (Вариант Аг в приложении 1). Однако, в тех немногочисленных геномах бактерий из кишечника человека, в которых есть фермент NadR, этот фермент представлен вдобавок к пути сохранения В3, а не вместо него.

Универсальный путь биосинтеза кофермента А (КоА) из пантотената (витамин В5) присутствует (с некоторыми неортологичными замещениями первой реакции) во всех прототрофах по В5 (1263 генома, 57%), а также в большинстве ауксотрофов по В5 (874 генома, 39%). Вдобавок, в 91 геноме (в основном *Bifidobacterium* и различные *Clostridiales*) был найден неполный путь биосинтеза пантотената, с отсутствующими ферментами CoaBC, которые катализируют превращение фосфо-пантотената в фосфо-пантетеин. Мы предполагаем, что эти бактерии сохраняют пантетеин (продукт деградации КоА) из окружающей среды, который далее вновь превращается в КоА при помощи реакций, последовательно катализируемых киназой CoaX/CoaW и ферментами CoaD и CoaE. Детальный анализ генов биосинтеза КоА среди 88 геномов бифидобактерий показал, что именно гены *coaBC* (но не ген *coaX*) отсутствуют в 63 геномах, что позволяет предположить, что киназа CoaX у бифидобактерий бифункциональна и может фосфорилировать как пантотенат, так и пантетеин. В самом деле, некоторые изученные киназы пантотената, в т.ч. CoaA у *E. coli* и CoaW у *S. aureus* могут фосфорилировать и пантотенат, и пантетеин [199]. Также стоит отметить, что такой набор генов (вариант Арп в приложении 1) не может быть классифицирован как ауксотрофия по В5, т.к. такие виды не могли бы синтезировать КоА из пантотената. Среди ауксотрофов по В5 с неполными биосинтетическими путями есть два варианта, в которых не синтезируется пантонат: (1) вариант Арт, при котором путь состоит из двух ферментов - PanD и PanC (найден в 40 геномах), (2) вариант Арт*, где путь биосинтеза пантотената состоит лишь из одного фермента PanC (найден в 16 геномах). Такие варианты пути позволяют предположить сохранение пантоата (с известным либо неизвестным источником бета-аланина, соответственно) вдобавок к витамину В5.

Биотин (витамин В7) - это необходимый кофактор для биотин-зависимых карбоксилаз, играющих важную роль в метаболизме углеводов, метаболизме аминокислот и в липогенезе. Биотин синтезируется из предшественника пимелоила через универсальный путь, состоящий из четырех реакций, катализируемых ферментами BioF, BioA, BioD и BioB. Интермедиатами пути являются 7-кето-8-аминопеларгоновая кислота (КАП), 7,8-диаминопеларгоновая кислота

(ДАП) и детиобиотин [200]. В большинстве ауксотрофных штаммов (836 геномов, 75%) все четыре биосинтетических фермента отсутствуют. У остальных ауксотрофов могут быть неполные биосинтетические пути; которые можно классифицировать на три основные группы: (A1) есть только фермент BioB; (A2) есть ферменты BioB и BioD; (A3) есть ферменты BioB, BioD и BioA, но BioF отсутствует. Большинство ауксотрофов по B7 с неполными путями оказались в первой группе (вариант A1, 169 геномов), тогда как группы A2 и A3 содержат 38 и 68 геномов, соответственно. Тот факт, что эти неполные пути присутствуют во геномах разнообразных бактерий из типов *Actinobacteria*, *Proteobacteria* и *Firmicutes*, позволяет предположить, что бактерии могут сохранять интермедиаты пути биосинтеза биотина из окружающей среды, а именно - детиобиотин (бактерии с вариантами пути A1, A2 и A3), ДАП (варианты A2 и A3) и КАП (вариант A3).

В отличие от других витаминов, витамин B12 и его производные не являются строго необходимым для всех бактерий. При том, что большинство бактерий используют B12-зависимые метионин-синтазу и рибонуклеотид редуктазу, эти ферменты могут быть необязательны из-за существования альтернативных B12-независимых ферментов и метаболических путей [168]. Это указывает на принципиальное различие между ауксотрофами по B12 (сохранение B12 для них выгодно, но не всегда необходимо) и ауксотрофами по другим витаминам группы В (где соответствующие витаминными абсолютно необходимы для существования бактерии). Тем не менее, было отмечено, что наличие B12 в окружающей среде положительно влияет на жизнеспособность некоторых бактерий кишечного микробиома из группы *Bacteroides* [201].

У большинства ауксотрофов по B12 (1235 геномов, 82%) нет ни ферментов пути синтеза корринового кольца, ни ферментов, выполняющих более поздние шаги синтеза B12, если не считать аденозилтрансферазы *VtuR/PduO*, необходимой для превращения витамина B12 (кобаламин) в активный кофермент B12 (аденозилкобаламин). Руководствуясь теми же принципами, как и в случае других витаминов, мы обнаружили 3 варианта неполных путей биосинтеза B12 (варианты *Aba*, *Acbg* и *Acbi* в приложении 1). Мы предполагаем, что бактерии, несущие эти варианты путей, могут сохранять такие предшественники B12, как кобинамид, диамид кобирината и кобириновую кислоту, соответственно (Рис. 15).

В заключение, проведенный анализ неполных метаболических путей позволил нам объяснить большинство пропусков и несоответствий в *in silico* реконструкции биосинтеза и сохранения витаминов и кофакторов. Это приближает нас к относительно полному пониманию витаминного метаболизма среди всего набора секвенированных геномов микробиома

кишечника человека и подготавливает возможность для дальнейшего изучения витаминных фенотипов среди всего микробиома (см. следующие секции). Что не менее важно, результаты анализа создают много новых вопросов, в т.ч. об отсутствующих генах и неизвестных биохимических путях. Вдобавок, результаты данного анализа позволяют предположить существование нескольких дополнительных и/или альтернативных метаболических предшественников соответствующих кофакторов. Вероятно, эти предшественники, так же как и известные витамины, могут участвовать в метаболическом обмене между бактериями-донорами (некоторыми прототрофами) и бактериями-акцепторами (ауксотрофами), однако такой обмен должен быть обеспечен множеством специализированных систем транспорта.

Геномное распределение транспортеров витаминов

Несмотря на то, что понимание транспортных систем, несомненно, крайне важно для изучения путей поглощения и сохранения витаминов, особенно в ауксотрофных бактериях, мы, как правило, не включали наличие/отсутствие транспортеров в качестве одного из определяющих признаков в наши фенотипические правила. Это связано с тем, что, как упомянуто выше, гены транспортеров обладают высокой эволюционной пластичностью, а также в большинстве своем недостаточно экспериментально изучены, и поэтому определение специфичности транспортеров сугубо биоинформатическими методами представляет собой трудно решаемую задачу. 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. Транспортные системы витаминов и их предшественников, представленные у прототрофных бактерий, могут помогать их оппортунистическому образу жизни и играть роль в меж-организменном обмене метаболитами, как обсуждается ниже. 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.

Среди тех организмов, что являются ауксотрофами и по ГМП, и по ГЭТ, 259 (57%) геномов содержат транспортеры для обоих предшественников, и еще 100 геномов содержат транспортеры либо для одного, либо для другого соединения. Несмотря на неполное знание транспортеров предшественников тиамина, результаты анализа поддерживают гипотезу о том, что сохранение предшественников витамина В1 распространено среди бактерий кишечного микробиома человека и играет важную роль в формировании этого бактериального сообщества. Транспортеры предшественников витамина В1 также представлены во многих прототрофных геномах, что указывает на дополнительные метаболические возможности данных организмов, важные для сохранения и, возможно, выделения предшественников витаминов.

Недостаток информации о транспортных системах и механизмах особенно очевиден в случае хорошо изученной системы метаболизма витамина В3 и кофактора НАД. Так, даже после многочисленных попыток не был найден транспортер В3 у *Escherichia coli* K12, несмотря на то, что способность этой бактерии сохранять никотинат и никотинамид из окружающей среды была показана экспериментально [202]. В самом деле, транспортеры витамина В3, относящиеся к трем известным семействам (NiaP, NiaX и NiaY), были найдены только в 25% ауксотрофов по В3 (по большей части, в различных фирмикутах, но также в актинобактериях и энтеробактериях). Мы не смогли идентифицировать транспортную систему, ассоциированную с путем сохранения квинолината. С другой стороны, транспортер PnuC, который, по нашим предположениям, специфично переносит никотинамид рибозу (НР), был найден лишь в 31 ауксотрофных геномах порядка *Pasteurellales*, полностью зависящих от пути сохранения НР, что поддерживает нашу гипотезу.

Транспорт витамина В5 в бактериях обеспечивается транспортерами PanF и PanT, тогда как PanS был охарактеризован как транспортер пантоата у *Salmonella* [203]. 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%. Мы также идентифицировали ортологи транспортера PanS в 153 геномах, в т.ч. в геномах 35 ауксотрофов из порядков *Clostridiales*, *Tissierellales* и *Veillonellales*. Все данные геномы содержат консервативный оперон *panD-panC-panS*, кодирующий путь сохранения пантоата. Это указывает на то, что бактерии кишечного микробиома, вероятно, могут обмениваться пантоатом.

К известным транспортерам биотина относятся транспортер BioY из семейства ECF (использует энергию гидролиза АТФ), широко распространенный среди многих

бактериальных типов, и вторичный транспортер YigM, распространенный среди энтеробактерий (в т.ч. у *E. coli*) и нескольких других таксонов протеобактерий. В геномах большинства ауксотрофов по B7 есть ген транспортера BioY (78%), тогда как среди геномов с неполными путями синтеза биотина эта цифра достигает 87% (240 из 275 геномов). Вдобавок, транспортер YigM был найден в 13 геномах Бактерий из порядка *Campylobacterales* (*Epsilonproteobacteria*), которые характеризуются неполными путями биосинтеза биотина из групп A1 или A3. Эти наблюдения указывают на то, что оба типа биотиновых транспортеров могут иметь более широкую специфичность к интермедиатам этого биосинтетического пути. В самом деле, было показано, что у *E. coli* детиобиотин репрессирует транспорт биотина [204], что указывает на то, что детиобиотин может конкурентно связывать биотиновый транспортер. Другие исследования показали накопление интермедиатов пути биосинтеза биотина прототрофными штаммами энтеробактерий [205].

Основываясь на наших наблюдениях и результатами предыдущих исследований мы предполагаем, что бактерии кишечного микробиома обмениваются не только биотином, но и его биохимическими предшественниками.

Мы смогли идентифицировать ортологи и паралоги известных транспортеров B12 (транспортер VtuFCD(B) из семейства ABC и транспортер CbrG из семейства ECF) примерно у половины организмов - ауксотрофов по B12, в том числе у 268 организмов с неполными путями биосинтеза B12. Это может указывать на роль этих транспортеров в переносе метаболитических предшественников B12. В самом деле, было показано, представители обоих семейств транспортеров (и VtuFCD/VtuB, и CbrG), могут переносить и B12, и кобинамид [206], [207], [208], [209]. Таким образом, поиск транспортных систем, переносящих витамин B12 и его предшественников, в оставшихся представителях кишечного микробиома человека остается открытой задачей.

Из литературы известно, что метаболический обмен аминокислотами влияет на структуру бактериальных сообществ [210]. Распределение реконструированных метаболитических путей и транспортеров в изученных геномах указывает на то, что обмен витаминами и их предшественниками также может играть важную роль в поддержании стабильности бактериальных сообществ. Возможны два пути секреции витаминов: (1) витамины, наряду с другими метаболитами, выходят в окружающую среду при лизисе клеток, и (2) секреция витаминов из живых клеток при помощи белков-транспортеров, либо специфичных к витаминам, либо с более широкой специфичностью. В среднем 3-16% генов бактерий кодируют компоненты транспортеров, которые можно разделить на первично-активные

транспортеры (например, транспортеры ABC и ECF семейств), вторично-активные транспортеры или активные переносчики (например, пермеазы MFS семейства), и переносчики-каналы [211], [212]. Если транспортеры, переносящие метаболиты внутрь клетки (импортеры) изучены для многих витаминов, то транспортеры, секретирующие метаболиты из клетки во внеклеточную среду (экспортеры) остаются малоизученными. Так, на сегодняшний день, единственный известный пример таких транспортеров - это экспортер рибофлавина, обнаруженный в нескольких штаммах *E. coli* и некоторых *Streptomyces* spp. [213], [214]. Мы предполагаем, что 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). Чтобы лучше оценить возможные механизмы обмена витаминами, мы проанализировали со-встречаемость найденных витаминных транспортеров, их распределение по семействам, а также распределение фенотипов витаминного биосинтеза среди изучаемых геномов (Рис. 16).

Несмотря на то, что в целом транспортеры распределены среди изучаемых геномов мозаично, можно наблюдать отдельные любопытные тренды. 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. Среди остальных трендов стоит отметить следующие: (1) пермеаза биотина YigM чаще встречается у бактерий - прототрофов по B7, тогда как транспортер BioY (из семейства ECF) чаще встречается у ауксотрофов; (2) аналогично, пермеаза пантотената PanF чаще встречается у прототрофов по B5, тогда как транспортер PanT из ECF семейства чаще встречается у ауксотрофов; (3) схожий тренд наблюдается у транспортеров тиамин - пермеазы PnuT и транспортера ThiT из ECF семейства. Несмотря на несколько важных оговорок (таких, как неполные данные и возможные альтернативные гипотезы), наблюдаемые тренды поддерживают предположение о возможном участии некоторых вторичных транспортеров в выделении витаминов и их предшественников в окружающую среду и в обмене метаболитами между бактериями. Если это предположение будет подтверждено экспериментально, вышеперечисленные тенденции могут быть использованы для того, чтобы отличать бактерий - доноров витаминов от остальных прототрофов, ведущих оппортунистический образ жизни, но не выделяющих витамины в окружающую среду.

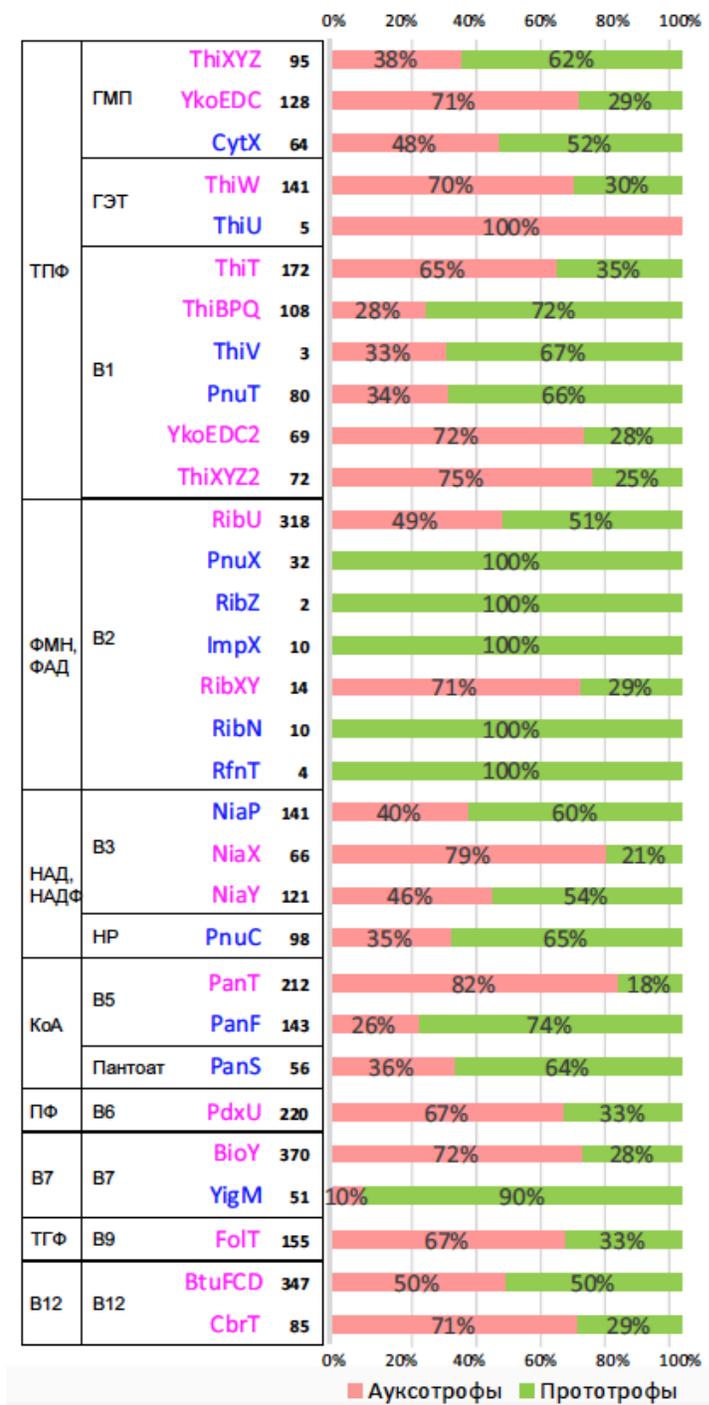


Рис. 16 Распределение транспортеров витаминов и их предшественников в геномах микробиома человека. Транспортеры сгруппированы в соответствии с кофакторами (первая колонка) и витамином/предшественниками (вторая колонка). Активные транспортеры из семейств ABC и ECF (используют энергию гидролиза АТФ) выделены розовым; АТФ-независимые пермеазы и выделены синим. Число геномов, содержащих тот или иной

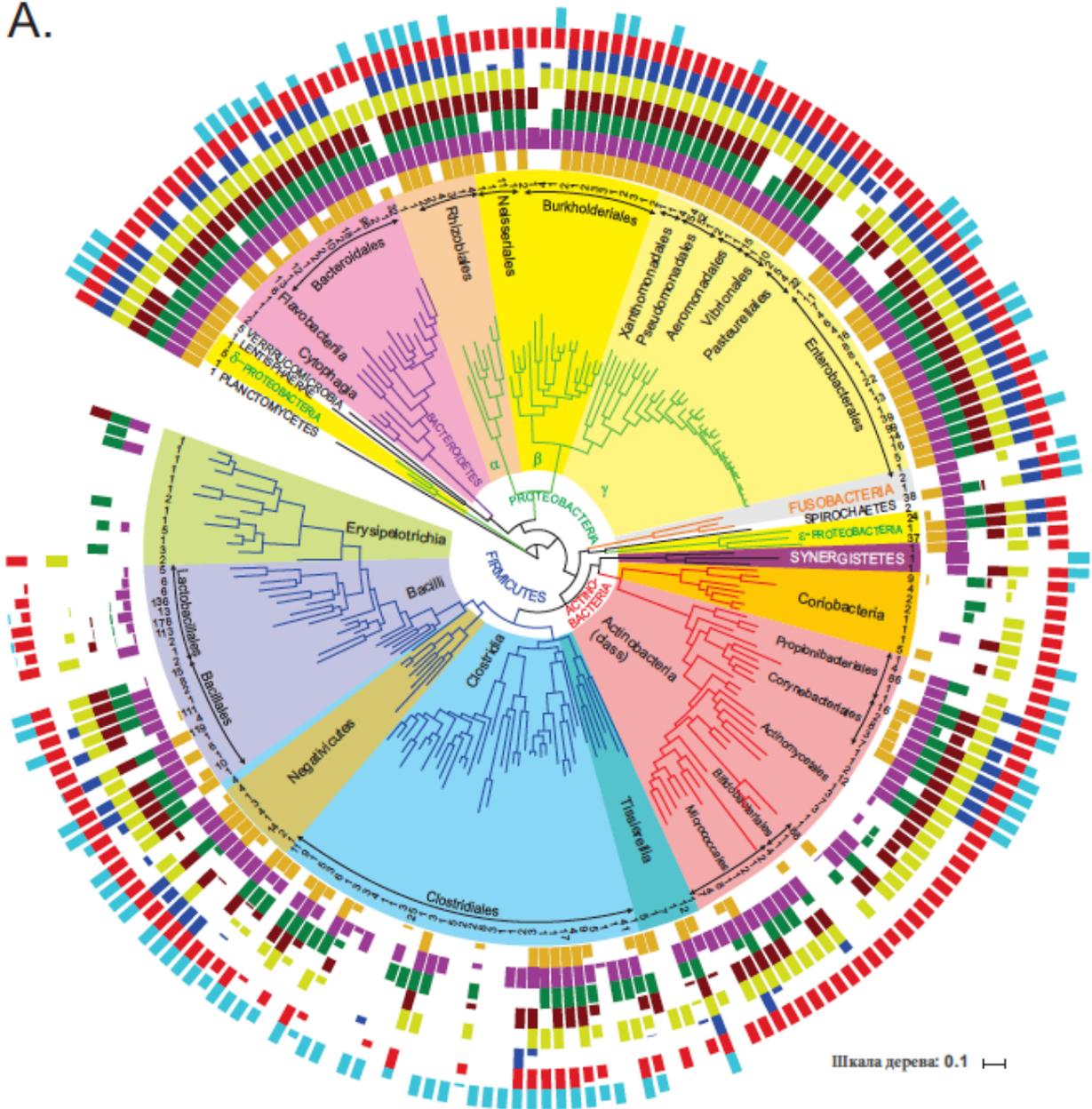
транспортер, выделено жирным; доли ауксотрофных и прототрофных бактерий среди них изображены при помощи розовых и зеленых полос.

Бинарные витаминные фенотипы: филогенетическое распределение и вариабельность

Для того, чтобы иметь возможность оценивать и сравнивать фенотипические свойства образцов микробиомных сообществ на основании полученных метаболических реконструкций, был разработан концепт цифровых “бинарных фенотипов”, где значение витаминного фенотипа определенного организма может равняться либо “1” (прототроф), либо “0” (ауксотроф). Для последующего анализа микробиомных сообществ, мы перевели все детальные коды, описывающие варианты метаболических путей, в упрощенные бинарные формы, присваивая значения “1” и “0” для каждого из восьми исследуемых витаминов во всех выбранных геномах (данные не показаны). Такое упрощенное представление фенотипов позволило нам разрешить несколько важных вопросов.

Во-первых, такое представление упрощает анализ филогенетического распределения метаболических потребностей и возможностей среди широкой выборки генов бактерий человеческого микробиома. Мы использовали бинарные фенотипы, чтобы вычислить усредненные уровни витаминной прототрофии на разных таксономических уровнях, включая: виды, роды, семейства, порядки, классы и типы. Эти усредненные уровни варьируют на шкале от 0 до 1 (что равносильно шкале от 0 до 100%), отражая долю прототрофных организмов в данной таксономической группе. Мы изобразили наиболее часто встречающийся фенотип для каждой таксономической группы на уровне видов (Приложение 3), а также на уровне родов и типов (Рис. 17) на филогенетическом дереве. Типы *Proteobacteria* и *Bacteroidetes* содержат по большей части прототрофные организмы, способные синтезировать все витамины, не считая кобаламин: только лишь 30% протеобактерий и 42% бактероидов могут синтезировать этот витамин самостоятельно. Большинство представителей типа *Fusobacteria* (кроме родов *Cetobacterium* и *Leptotrichia*) могут синтезировать все витамины, кроме пантотената (B5). Тип *Verrucomicrobia* (представлен пятью видами *Akkermansia*), а также тип *Planctomycetes*, представленный единственным штаммом, могут синтезировать все витамины, кроме B12. Тип *Lentisphaerae*, также представленный единственным организмом (*Victivalis vadensis*), может

A.



B.

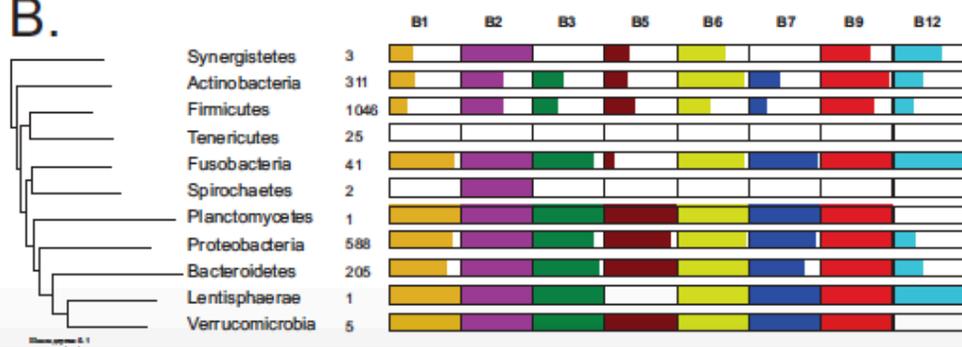


Рис. 17 Распределение прототрофных бактерий среди анализируемых геномов представителей кишечного микробиома. (A) Усредненные фенотипы по витаминам

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. (В) Усредненные витаминные фенотипы для 11 типов. Мы получили филогенетическое дерево типов, схлопнув филогенетическое дерево родов. Число проанализированных представителей показано рядом с названием каждого типа.

синтезировать все витамины, кроме В5. Оба вида-представителя типа *Spirochaetes* ауксотрофны по всем витаминам, кроме рибофлавина (В2). Штаммы из родов *Ureaplasma* и *Mycoplasma* из типа *Tenericutes* ауксотрофны по всем восьми исследуемым витаминам. Наибольшее число организмов ауксотрофов по всем витаминам относится к типам *Firmicutes*, *Actinobacteria* и *Synergistetes*. В исследуемой выборке геномов есть 3 порядка *Actinobacteria* (*Propionibacteriales*, *Corynebacteriales*, *Coriobacteriales*), представители которых могут синтезировать В12, в то время как среди *Firmicutes* большинство бактерий продуцентов витамина В12 относится к порядкам *Clostridiales*, *Selenomonadales* и *Veillonellales* (Рис. 17). В целом, проведенный филогенетический анализ демонстрирует наличие корреляций между филогенией (даже на уровне широких таксономических групп) и витаминными фенотипами, но при этом наблюдается существенная вариабельность витаминных фенотипов внутри таксономических групп.

Оценка межвидовой и внутривидовой вариабельности бинарных витаминных фенотипов необходима для оценки того, насколько точно можно предсказывать витаминные фенотипы неизученных бактерий, основываясь на их филогенетическом родстве с изученными бактериями. Такая оценка полезна для практического применения, например в таких задачах, как оценка фенотипического профиля бактериальных сообществ (см. следующую секцию). Для оценки фенотипической вариабельности на уровне видов и более широких таксономических групп мы использовали такую метрику, как число вариабельных фенотипов (варьирует от 1 до 8 для восьми витаминов). 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. Видом с наибольшей фенотипической вариабельностью оказался *Ruminococcus torques*: фенотипы этого вида по 6 из 8 витаминов отличаются у двух штаммов, представляющих этот вид в нашей выборке. Это указывает на возможную неверную классификацию одного из штаммов; в самом деле, основываясь на фенотипических, биохимических, филогенетических и геномных признаках один из этих двух штаммов был недавно переклассифицирован в *Mediterraneibacter torques* ATCC 27756 [215]. С эволюционной точки зрения, наблюдаемые случаи внутривидовой вариабельности витаминных фенотипов могут быть объяснены либо потерей, либо приобретением генов биосинтеза витаминов, часто ко-локализующихся вместе на хромосоме. Например, сравнительный анализ 62 штаммов *Enterococcus faecalis* показал, что 27 штаммов могли потерять геномный локус *panDBGC*, кодирующий все необходимые ферменты пути биосинтеза пантотената. 4 из 5 штаммов *Faecalibacterium prausnitzii* и 3 из 5 штаммов *Lactobacillus reuteri* содержат оперон биосинтеза рибофлавина *ribADHBE*, тогда как близкородственные штаммы из тех же видов утратили гены биосинтеза рибофлавина. Напротив, один из пяти штаммов *Streptococcus parasanguinis*, скорее всего, приобрел геномный локус, кодирующий 20 ферментов, необходимых для биосинтеза B12.

Мы также проанализировали межвидовую фенотипическую вариативность и вариативность на более высоких таксономических уровнях, используя ту же метрику (число вариабельных фенотипов). 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*, подчеркивая вклад этого класса в вариабельность витаминных фенотипов среди бактерий кишечного микробиома. На уровне порядков, 30 из 36 таксонов, представленных более чем одним геномом (83%) демонстрируют вариабельные фенотипы. Наиболее вариабельными типами бактерий являются *Actinobacteria* and *Firmicutes*, с 7 и 8 вариабельными фенотипами, соответственно.

Таким образом, согласно нашим наблюдениям, фенотипы по витаминам консервативны на внутривидовом уровне, но вариативны на более высоких таксономических уровнях. Поскольку

группировка бактерий по операционным таксономическим единицам (ОТЕ) с порогом сходства последовательностей 97% ближе к группировке бактерий по видам, чем по другим таксономическим единицам, мы пришли к выводу, что можем сопоставлять каждому ОТЕ единый фенотип по каждому витамину без существенной потери точности. Мы использовали такое усреднение для оценки общей доли ауксотрофных организмов среди бактерий в образцах кишечного микробиома.

Уровни ауксотрофии в образцах человеческого кишечного микробиома

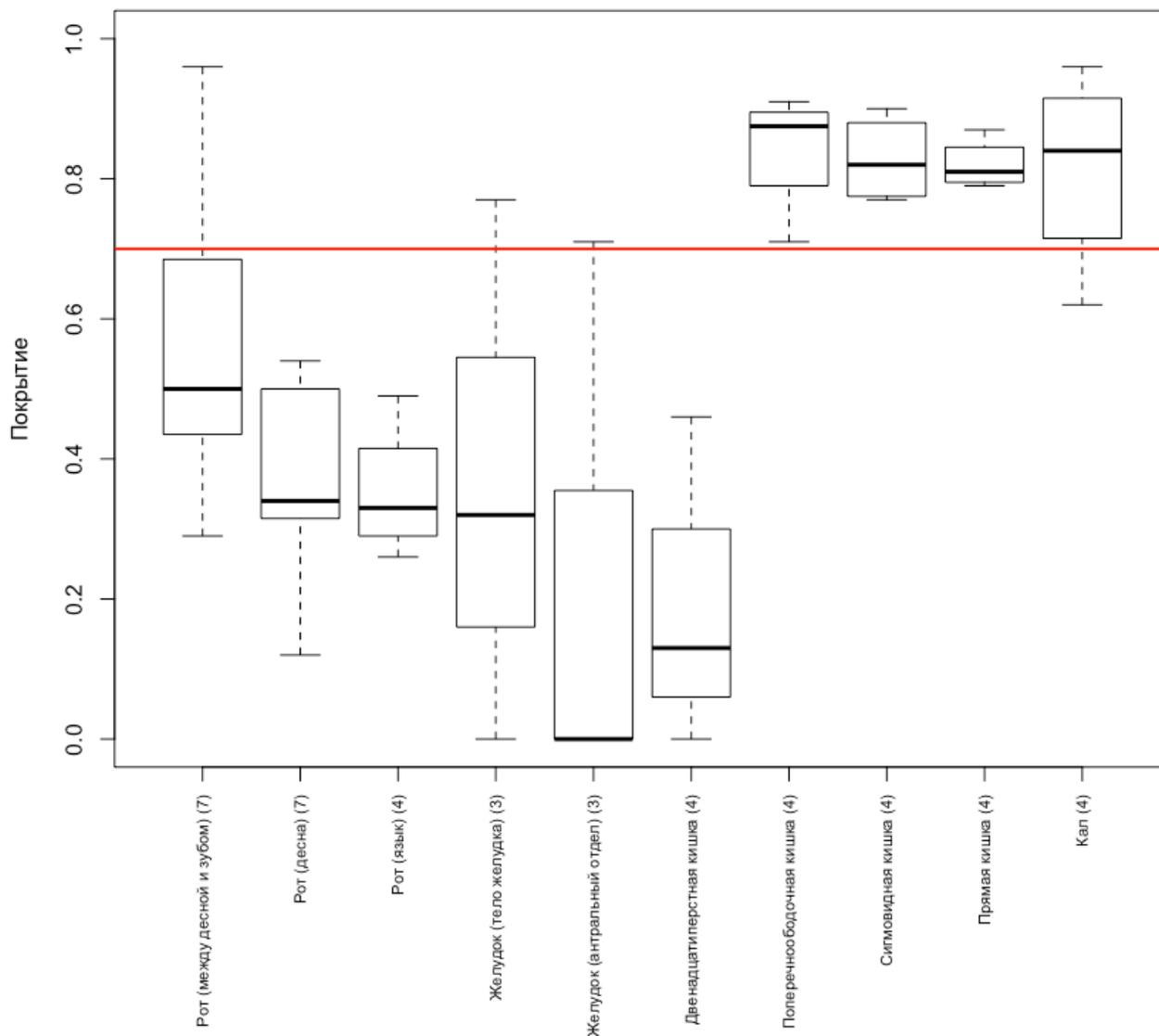
Вышеописанные наблюдения о распределении метаболических путей биосинтеза и сохранения витаминов интересны с точки зрения филогенетического разнообразия бактерий и их эволюции. Однако, в бактериальных сообществах, и, в частности, в кишечном микробиоме, численность различных бактериальных таксонов распределена 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. В последние 10 лет было опубликовано много таких работ, оценивающих филогенетическое разнообразие человеческого микробиома, основывающихся на секвенировании генов 16S рибосомальной РНК (16S рРНК) [97], [92], [216], [217], [190], [218], [219], [220], [221], [222], [223]. Распространенный метод оценки филогенетического разнообразия по итогам секвенирования 16S рРНК - это разбиение всех найденных последовательностей 16S рРНК на кластеры, где все представители одного кластера имеют сходство последовательностей не ниже определенного порогового значения. Такие кластеры называют “операционными таксономическими единицами” (ОТЕ от английского “Operational taxonomic units”) [224]. Степень сходства геномов бактерий внутри одного ОТЕ примерно соответствует степени сходства геномов внутри одного вида [192]. Как обсуждается выше, большинство бактерий демонстрируют консервативные фенотипы по биосинтезу витаминов группы В на уровне видов, поэтому, мы допускаем, что у всех бактерий внутри одного ОТЕ фенотипы по биосинтезу витаминов схожи. Каждому ОТЕ мы сопоставили набор бинарных фенотипов по восьми витаминам группы В. Таким образом, для каждого образца кишечного микробиома, для которого доступны данные секвенирования 16S рРНК, мы смогли оценить численность бактерий, несущих тот или иной

фенотип по каждому изучаемому витамину. Для численной оценки баланса ауксотрофов и прототрофов мы использовали такую метрику, как средний уровень ауксотрофии, т.е. численную долю ауксотрофных бактерий в исследуемом образце.

Мы проанализировали баланс прототрофов и ауксотрофов в образцах, собранных в процессе четырех независимых исследований (см. более подробное обсуждение ниже). 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 [225]. Для того, чтобы оценить, насколько сильно может варьировать уровень ауксотрофии между разными здоровыми индивидуумами, мы проанализировали уровни ауксотрофии в образцах кишечного микробиома, собранных в рамках проекта «Микробиом человека» [187]. Наконец, чтобы наблюдать зависимость среднего уровня ауксотрофии от рациона питания организма хозяина, мы использовали данные из двух исследований, где были собраны образцы одной группы людей, употребляющей разную пищу. В одном исследовании, испытуемые ели только мясную пищу на протяжении пяти дней, в остальное время совмещая мясную и растительную пищу; в другом исследовании, были собраны образцы людей, ведущих кочевой образ жизни, чей рацион питания меняется сезонно.

Кишечник человека населен чрезвычайно разнообразными бактериями, и некоторые группы бактерий могут не быть представлены в анализируемой выборке геномов. Для оценки того, насколько хорошо исследуемая нами выборка геномов представляет разнообразие бактерий кишечного микробиома, мы использовали такую метрику, как «покрытие» образца. Мы вычисляли покрытие как суммарную встречаемость тех ОТЕ, для которых мы смогли предсказать фенотипы по витаминам группы В. Таким образом, покрытие образца отражает долю организмов в данном образце, для которых мы можем предсказать фенотипы. Средний уровень ауксотрофии можно считать адекватной оценкой баланса между ауксотрофами и прототрофами в бактериальном сообществе только в случае высокого покрытия.

Для оценки того, насколько хорошо выборка геномов представляет состав кишечного микробиома, мы проанализировали образцы, взятые из разных участков желудочно-кишечного тракта у четырех индивидуумов [225]. Мы наблюдали высокое покрытие только в образцах кишечного микробиома, но не в образцах из вышележащих участков (рот, желудок) (Рис. 18), что говорит о том, что выборка анализируемых геномов высоко специфична для кишечного микробиома, но покрывает его в достаточной мере, чтобы адекватно оценивать баланс *Рис.*



18. Покрытие разнообразия организмов таксонами с предсказанными витаминными фенотипами у образцов микробиома из разных участков желудочно-кишечного тракта. Для каждого участка желудочно-кишечного тракта показан бокс-плот распределения значений покрытия для всех изученных образцов. Количество изученных образцов указано в скобках. Участки желудочно-кишечного тракта упорядочены слева направо. Красная линия отображает минимальное значение покрытия (0.7), которое мы считали достаточным для дальнейшего изучения образца.

ауксотрофов и прототрофов в сообществе. Весь дальнейший анализ проводился только для образцов с покрытием выше 70%.

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, рибофлавинолу, 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. 19). 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. 19). Особенно высок уровень ауксотрофии для кобаламина (медианное значение 72%). 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, сколько 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 [188] 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

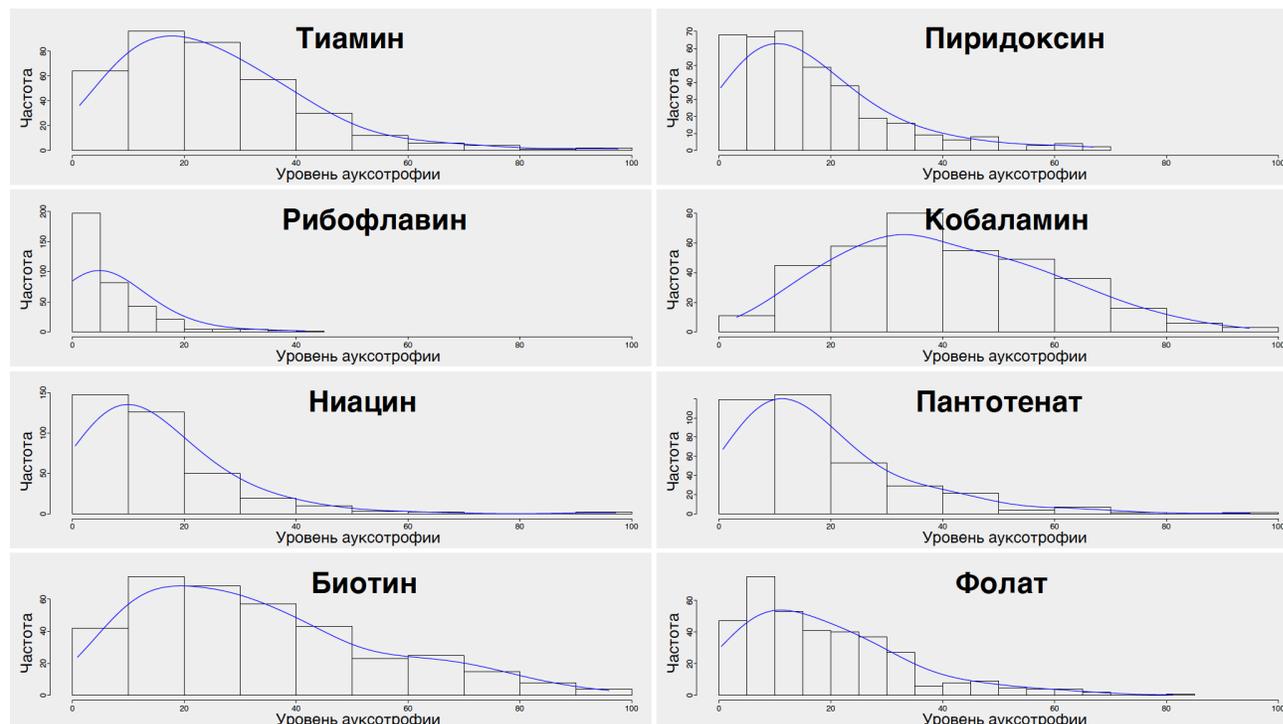


Рис. 19. Распределения средних уровней ауксотрофии для витаминов группы В в образцах кишечного микробиома, собранных из 388 индивидуумов в рамках проекта «Микробиом человека». Средний уровень ауксотрофии вычисляется как численная доля ауксотрофов среди геномов с предсказанными фенотипами.

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. 20).

В другом исследовании [98] 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

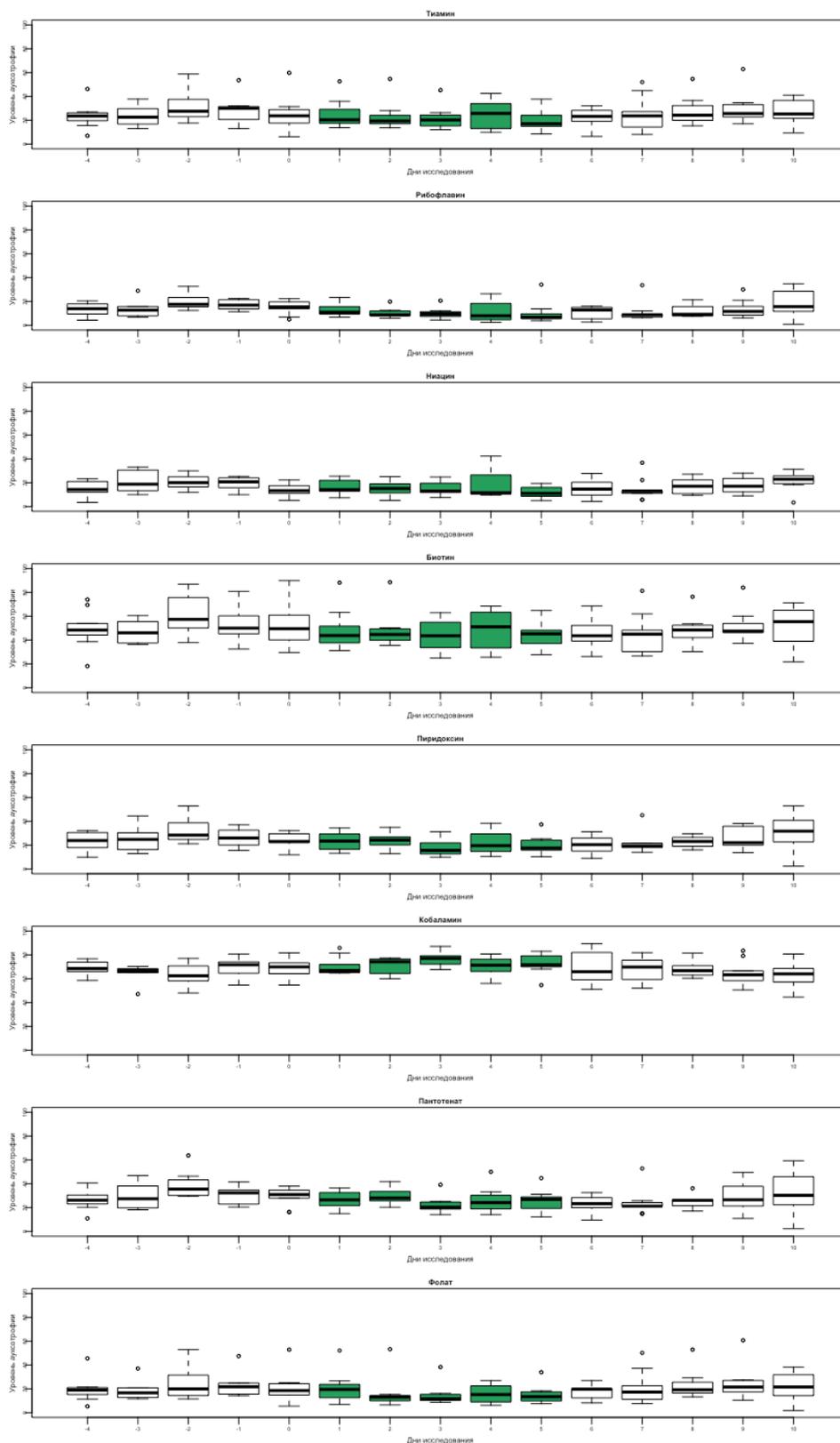


Рис. 20. Распределения средних уровней ауксотрофии для витаминов группы В в образцах кишечного микробиома, собранных у волонтеров в исследовании [188]. Слева направо

отмечены дни проведения исследования. Первые 4 дня волонтеры питаются согласно своему обычному рациону, следующие 5 дней они переходят на мясную диету (отмечено зеленым), далее 6 дней волонтеры переходят обратно к своему обычному рациону питания. Для каждого дня исследования распределение средних уровней ауксотрофии в микробиомах волонтеров в этот день показано бокс-плотом. Каждому витамину соответствует индивидуальная панель.

данном случае, мы тоже не наблюдали значительных сезонных изменений среднего уровня ауксотрофии, соответствующих изменению рациона питания организма-хозяина (Рис. 21).

То, что мы не смогли пронаблюдать значительного изменения среднего уровня ауксотрофии при изменении диеты для обоих исследований, можно объяснить несколькими способами. Во-первых, разработанный нами метод оценки среднего уровня ауксотрофии может быть недостаточно точен, чтобы распознать изменения в бактериальном сообществе. Даже при том, что выборка геномов, которую мы используем, покрывает большую часть распространенных видов бактерий, она безусловно далеко не полна. Далее, оценка разнообразия и структуры микробиома строится на секвенировании генов 16S рНК, и этот метод также обладает рядом недостатков: например, различное число генов 16S рНК в геномах разных бактерий, а также слишком большое сходство генов 16S рНК между родственными, но не идентичными штаммами снижают точность этого метода. Общий уровень шума, происходящий как от экспериментальных методов обработки образца, так и от последующей биоинформатической обработки данных секвенирования, может превышать реальный наблюдаемый эффект. Во-вторых, влияние витаминов, поступающих с пищей организма-хозяина, может быть достаточно ограниченным: если большинство витаминов всасывается клетками верхних отделов желудочно-кишечного тракта, и в нижние отделы кишечника поступают лишь незначительные количества витаминов, то изменения в диете могут менять численный состав представителей микробиома, но не менять баланс ауксотрофов и прототрофов. Наконец, в-третьих, средний уровень ауксотрофии может быть некорректным инструментом оценки баланса ауксотрофов и прототрофов в бактериальном сообществе кишечника. Например, активную роль в метаболическом обмене могут играть не все прототрофы, а лишь те из них, которые способны активно экспортировать витамины. В том случае, если они представляют лишь малую долю всего сообщества по численности, уровень ауксотрофии полученный в данной работе может не отражать реального баланса доноров и акцепторов витаминов в кишечнике человеческого микробиома.

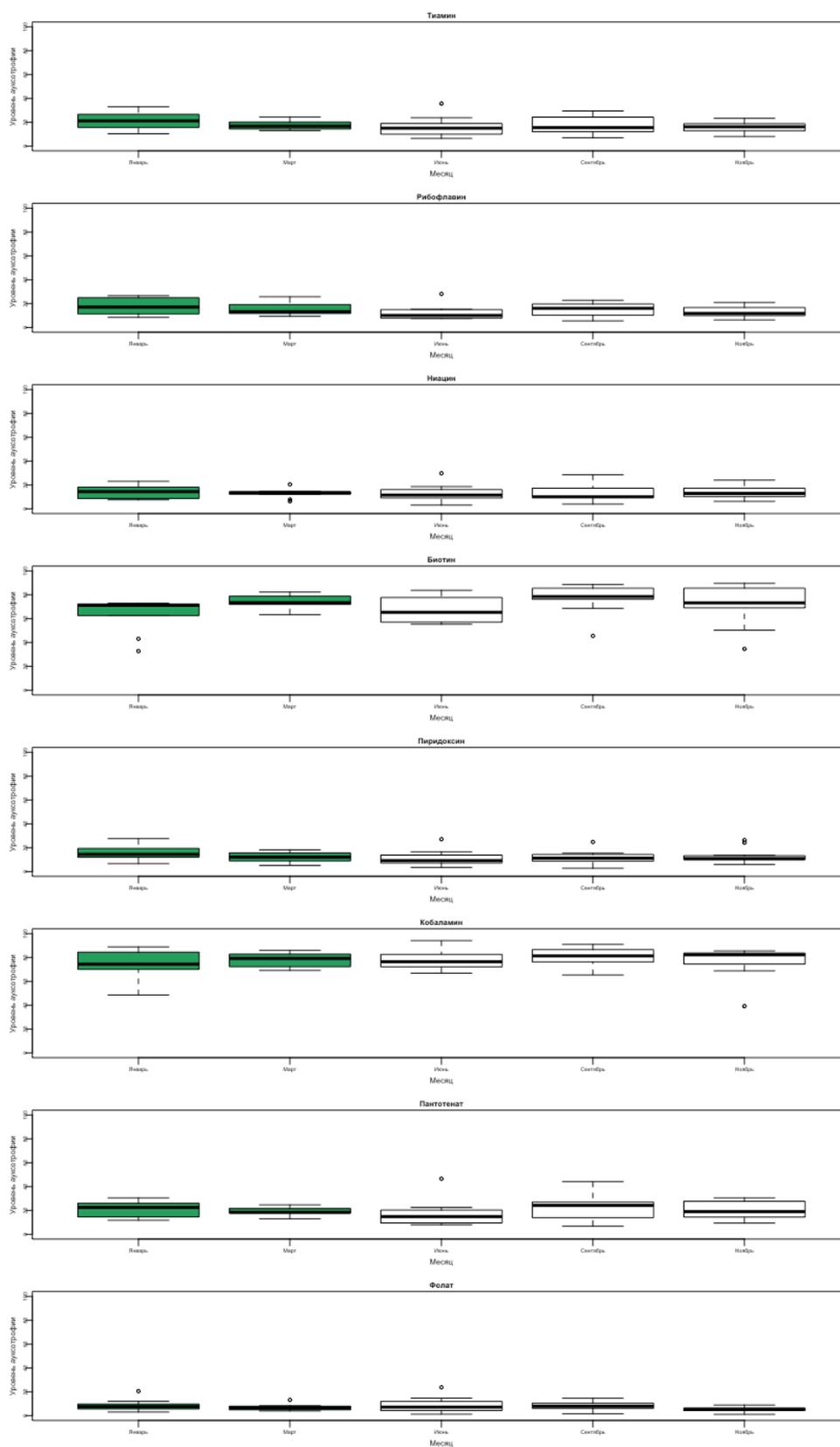


Рис. 21. Распределения средних уровней аукотрофии для витаминов группы В в образцах кишечного микробиома, собранных у жителей провинции Хентий [98]. Слева направо

отмечены месяцы сбора образцов (январь, март, июнь, сентябрь, ноябрь). Зимние месяцы отмечены зеленым (ноябрь и январь). Для каждого месяца распределение средних уровней ауксотрофии в микробиомах исследуемых индивидуумов показано бокс-плотом. Каждому витамину соответствует индивидуальная панель.

Влияние генов витаминного метаболизма на выживаемость бактерий из рода *Bacteroides* в кишечном микробиоме человека

Проведенные реконструкции метаболических путей биосинтеза витаминов были также использованы для анализа генов, важных для выживания бактерий в бактериальных сообществах. Данный проект был осуществлен в рамках сотрудничества с лабораторией Джеффри Гордона в Университете Вашингтона. В лаборатории Джеффри Гордона был разработан экспериментальный метод для определения генетических факторов, важных для выживаемости бактерий в бактериальных сообществах [226]. Для четырех штаммов *Bacteroides* (а именно, *B. cellulosilyticus* WH2, *B. ovatus* ATCC 8483, *B. thetaiotaomicron* 7330 и *B. thetaiotaomicron* VPI-5482) были созданы библиотеки мутантов. Каждый мутантный геном содержал вставку транспозона, и у разных мутантов транспозон располагался в различных позициях в геноме. Для подавляющего большинства белок-кодирующих генов хотя бы один мутантный штамм содержал вставку транспозона в кодирующую последовательность, тем самым нарушая функциональность гена (от 81,5% до 91,8% генов, в зависимости от бактерии). Поскольку было создано на порядок больше мутантов, чем количество генов в организме этих бактерий, можно заключить, что, вероятно, при нарушении функции оставшихся генов бактерии не могут расти и, тем самым, эти гены являются необходимыми для роста даже в питательной среде (*in vitro*). Нуклеотидная последовательность транспозона была модифицирована: на обоих концах последовательности транспозона располагаются сайты рестрикции MmeI. Рестриктаза MmeI разрезает молекулу ДНК в 20-21 нуклеотидах от своего связывающего сайта. Таким образом, наличие сайтов рестрикции MmeI позволяет секвенировать вставки транспозонов вместе с прилегающими фрагментами, и таким образом можно определить расположение каждой транспозонной вставки в геноме. Кроме того, каждая транспозонная вставка содержала таксон-специфичный баркод.

Созданные библиотеки мутантов были введены в кишечник мыши, лишенной кишечного микробиома, в составе искусственных бактериальных сообществ, состоящих из 15 различных штаммов бактерий. Доля штаммов с различными мутациями в данном сообществе была сперва измерена до внесения бактерий в кишечник мыши, и затем измерена последовательно 4, 10 и 16 дней спустя. Изменение относительной численности для любого выбранного мутантного штамма во время проведения эксперимента отражает, насколько сильно влияет данная мутация на выживаемость организма в данном сообществе. Фенотипическое действие гена зависит от окружающей среды. В данном эксперименте одними из главных факторов внешней среды являются (1) состав бактериального сообщества и (2) питание хозяина-организма (мыши). Для того, чтобы первый фактор (состав бактериального сообщества) был единообразен и сравним между проводимыми экспериментами, мыши получали инъекцию заранее искусственно созданного бактериального сообщества, состоящего из 4 библиотек транспозонных мутантов (для штаммов *B. cellulosilyticus* WH2, *B. ovatus* ATCC 8483, *B. thetaiotaomicron* VPI-5482 и *B. thetaiotaomicron* 7330) и из еще 11 штаммов бактерий (*Bacteroides caccae* ATCC 43185, *Bacteroides uniformis* ATCC 8492, *Bacteroides vulgatus* ATCC 8482, *Parabacteroides distasonis* ATCC 8503, *Eubacterium rectale* ATCC 33656, *Ruminococcus obeum* ATCC 29174, *Clostridium scindens* ATCC 35704, *Clostridium symbiosum* ATCC 14940, *Clostridium spiroforme* DSM 1552, *Dorea longicatena* DSM 13814, and *Collinsella aerofaciens* ATCC 25986). Для того, чтобы изучить фенотипические эффекты генов изучаемых бактерий в разных условиях окружающей среды, варьировали питание мышей (второй фактор): в одном эксперименте мыши получали диету с низким содержанием жира и высоким содержанием длинноцепочечных углеводов (LF/HPP), в другом - диету с высоким содержанием насыщенных жирных кислот и высоким содержанием короткоцепочечных углеводов (HF/HS). Кроме того, в качестве дополнительного контроля был проведен эксперимент *in vitro*, в котором библиотеки мутантов растили *in vitro* в минимальной среде с добавкой единственного источника углеводорода.

Проведенные эксперименты позволили наблюдать фенотипические эффекты на двух различных уровнях одновременно: (1) приспособленность каждого индивидуального штамма в условиях данного сообщества и данной диеты организма-хозяина (что в данном случае представляет меньший интерес), и (2) влияние пертурбации генов в изучаемых штаммах на их приспособленность в бактериальном сообществе. Более того, фенотипический эффект пертурбации генов в данном эксперименте можно было вычислить количественно. Все гены в четырех изучаемых штаммах *Bacteroides* были разделены на три основные группы (Рис. 22):

(1) незаменимые гены (не представлены в библиотеке мутантов, выращенной на богатой среде *in vitro*), (2) гены - факторы выживаемости *in vitro* (численность мутантов уменьшается при росте на минимальной среде *in vitro*) (2) гены - факторы выживаемости *in vivo* (численность мутантов уменьшается при росте в контексте бактериального сообщества *in vivo*), (4) остальные гены. Далее, первая и третья группы были разделены на несколько подгрупп: среди группы (3) есть гены, которые являются факторами выживаемости *in vivo* независимо от питания организма-хозяина, и есть гены, пертурбация которых вызывает фенотипический эффект лишь в одном из двух экспериментов, т.е специфично в условиях одного режима питания организма-хозяина. Такие гены были обозначены красными кругами и полукругами соответственно на Рис. 22. Среди группы (1) красными и желтыми квадратами соответственно были показаны гены с нулевой и с ненулевой численностью в библиотеке мутантов. Пороговые значения были выбраны с помощью алгоритма максимизации ожидания.

Информация о том, какие гены являются факторами выживаемости *in vivo*, позволяет узнать нам больше о метаболических потребностях и селективном отборе, который бактерии испытывают, будучи частью бактериального сообщества в кишечном микробиоме мыши. В рамках данного коллаборативного проекта мы проводили реконструкцию путей биосинтеза витаминов группы В у изучаемых бактерий, и затем анализировали данные о генах - факторах выживаемости, чтобы установить, какие из витаминов группы В могут быть ограничивающими ресурсами в микробиоме кишечника мыши. В следующих параграфах мы описываем результаты проведенного анализа.

Большинство ферментов биосинтеза витаминов группы В консервативно среди трех изучаемых видов *Bacteroides* (Рис. 22), с двумя исключениями: (1) у *B. cellulosilyticus*, фермент биосинтеза тиазола Thi4 замещает более распространенный вариант пути (через ThiF, ThiH, ThiS и ThiG), и (2) ферменты биосинтеза В12 представлены только у *B. cellulosilyticus*, но не у двух других видов. У всех трех видов *Bacteroides* отсутствуют ортологи известных витаминных транспортеров, за исключением тиамина (В1) и кобаламина (В12). Предсказанные транспортеры тиамина - транспортер внутренней мембраны PnuT и TonB-зависимый транспортер внешней мембраны Omp-Thi - были идентифицированы ранее в результате сравнительно-геномной реконструкции регулонов тиаминового рибосвича у *Bacteroides* [166]. Функция пермеазы PnuT у *B. thetaiotaomicron* была недавно подтверждена экспериментально: было показано, что она дополняет функцию транспортера ThiBPQ у *E. coli* [227]. Аналогично, транспортеры В12 - транспортер VtuFCD из семейства ABC и TonB-зависимый транспортер внешней мембраны VtuB - были ранее найдены в результате анализа регулонов рибосвича В12

и транспортеры показаны жирными линиями и пунктирными линиями, соответственно. Для ферментов (белые прямоугольники), транспортеров (синие прямоугольники) и транскрипционных факторов (черные прямоугольники) указаны названия генов и номера классификации ферментов (ЕС). Мы используем сокращения VT1, VT2, VO и VC для геномов *B. thetaiotaomicron* VPI-5482, *B. thetaiotaomicron* 7330, *B. ovatus* ATCC 8483 и *Bacteroides cellulosilyticus* WH2, соответственно. Незаменимые гены показаны красными и оранжевыми квадратами, в зависимости от встречаемости мутантов данного гена в исходной библиотеке мутантов. Зеленые квадраты показывают необязательные гены. Если ортолог данного гена отсутствует в данном геноме, это показано белым квадратом. Красные круги и полукруги внутри зеленых квадратов отображают гены - факторы выживаемости *in vivo*. Левый и правый полукруг соответствуют диетам "HF/HS" и "LF/HPP", соответственно. Гены - факторы выживаемости *in vitro* у *B. cellulosilyticus* WH2 показаны красными треугольниками, вписанными в зеленые квадраты.

[168]. Специфичность трех паралогичных транспортеров BtuB у *B. thetaiotaomicron* к различным корриноидам была экспериментально изучена [201].

Незаменимые гены и гены - факторы выживаемости *in vitro*

Гены, кодирующие ферменты из реконструированных путей биосинтеза витаминов/кофакторов, демонстрируют разные паттерны в плане своей значимости для выживания бактерии в бактериальном сообществе (Рис. 22). Гены биосинтеза рибофлавина, кофермента А и тетрагидрофолата незаменимы во всех трех изученных видах *Bacteroides*, что указывает на их неспособность сохранять соответствующие витамины - рибофлавин (B2), фолат (B9) и пантотенат (B5) из окружающей среды. Любопытно, что если неспособность сохранения рибофлавина и фолата распространена среди грам-отрицательных бактерий (например, *E. coli* [228]), то неспособность сохранять пантотенат является скорее исключением из правила, как обсуждается ниже. Ферменты биосинтеза НАД NadA, NadB и NadC являются незаменимыми только у *B. cellulosilyticus*, но не у *B. thetaiotaomicron* или *B. ovatus*. Эти два генома содержат фермент сохранения ниацина (B3) PncB, что позволяет предположить, что они могут сохранять ниацин из окружающей среды. Однако, транспортер ниацина у *Bacteroides* пока не найден. Бактерии рода *Bacteroides* могут как синтезировать тиамин самостоятельно, так и захватывать и сохранять его из окружающей среды. Это согласуется с

наблюдениями о том, что большинство генов биосинтеза и транспорта тиамин не являются незаменимыми для штаммов *Bacteroides in vitro*. Гены биосинтеза парааминобензоата (ПАБ) PabA и PabB также не являются незаменимыми; возможно, бактероиды могут сохранять ПАБ из окружающей среды и использован далее для биосинтеза тетрагидрофолата (ТГФ). Все гены биосинтеза биотина (B7) не были идентифицированы в качестве незаменимых генов, указывая на потенциальную возможность сохранения биотина из окружающей среды. Однако, транспортеры биотина у *Bacteroides* еще не были идентифицированы. Ферменты биосинтеза пиридоксина (B6) демонстрируют смешанный паттерн: PdxA незаменим у *B. thetaiotaomicron* и *B. ovatus*, но не является таковым у *B. cellulosilyticus*, что указывает на межвидовые различия в пути сохранения пиридоксина. Как и в случае биотина, на данный момент не известно ни об одном транспортере, специфичном к пиридоксину, у бактерий рода *Bacteroides*. Все ферменты биосинтетического пути кобаламина (B12) у *B. cellulosilyticus*, так же как и многие паралоги транспортера VtuB, найденные у всех трех видов *Bacteroides*, не являются незаменимыми. Дальнейший анализ *B. cellulosilyticus* WH2 показал, что все ферменты биосинтеза биотина являются факторами выживаемости *in vitro*, т.е. при росте библиотеки мутантов на минимальной среде с единственным источником углерода (не содержащей витаминов) численность мутантов этих генов значительно уменьшается. Этот означает, что биотин может быть лимитирующим фактором при росте бактерий в таких условиях. Также, ферменты биосинтеза пиридоксина PdxA, PdxB и SerC (но не PdxJ) являются факторами выживаемости *in vitro* при росте на минимальной среде.

Гены - факторы выживаемости *in vivo*

Отображение генов - факторов выживаемости *in vivo* на схемы метаболических путей биосинтеза витаминов показал, что большинство генов, кодирующих “заменяемые” ферменты (т.е. ферменты, мутации которых все же позволяют бактерии расти на богатой среде), не являются факторами выживаемости *in vivo*. К исключениям относятся гены биосинтеза двух предшественников КоА, бета-аланина (PanD) и пантоата (PanE), которые являются незаменимыми (т.е. бактерии с мутациями в этих генах не могут расти даже на богатой среде), можно предположить, что изучаемые штаммы *Bacteroides* не могут сохранять пантотенат из окружающей среды. Декарбоксилаза аспартата PanD заменяема и в штаммах *B. thetaiotaomicron*, и в *B. ovatus*, указывая на то, что, возможно, эти два штамма могут сохранять бета-аланин (несмотря на то, что транспортер бета-аланина у *Bacteroides* не известен). Тот

факт, что PanD является фактором выживаемости *in vivo*, позволяет предположить, что бета-аланин - это лимитирующий фактор в среде кишечника.

Далее, были идентифицированы многие штамм-специфичные факторы выживаемости *in vivo*, в том числе среди ферментов путей биосинтеза биотина (BioC, BioG, BioA, BioB у *B. cellulosilyticus*), НАД (NadA и NadC у *B. thetaiotaomicron* 7330), тиамина (ThiD и ThiE у *B. thetaiotaomicron* 7330), пиридоксина (PdxB, SerC, PdxJ и PdxK у *B. thetaiotaomicron* 7330), кобаламина (CbiL, CbiH, CbiT, CbiE, CbiC, CbiA, CobD, CobU, CobS и CobT у *B. cellulosilyticus* WH2) и ПАБ для биосинтеза фолата (PabA и PabB у *B. thetaiotaomicron* 7330). Можно предположить, что системы сохранения витаминов у соответствующих штаммов работают менее эффективно, либо что соответствующие штаммы нуждаются в больших концентрациях витаминов в условиях растущего *in vivo* бактериального сообщества. С другой стороны, витаминные транспортеры внешней мембраны BtuB и OMP-Thi являются факторами выживаемости *in vivo* в некоторых, но не во всех штаммах *Bacteroides*. В целом, проведенный анализ показывает, что у четырех изучаемых штаммов *Bacteroides* различаются метаболические потребности по некоторым витаминам, и что функциональная значимость генов витаминного метаболизма для выживаемости *in vivo* отличается между штаммами.

Обсуждение

Возможность предсказывать метаболические возможности и потребности бактериальных сообществ, населяющих человеческий кишечник, открывает новые возможности для диагностики и терапии болезней и синдромов, связанных с дисбиозом, используя персонализированный подбор терапевтических пробиотиков, микронутриентов и пребиотиков. Чрезвычайно важными микронутриентами для бактерий, населяющих человеческий кишечник, являются витамины группы В, подробно проанализированные в данной работе. Несмотря на то, что химически эти соединения значительно различаются между собой, их объединяет ряд свойств и функций, перечисленных ниже. Во-первых, все витамины группы В являются предшественниками важных метаболических кофакторов (приложение 1), незаменимых как для организма-хозяина, так и для всех видов бактерий (кроме витамина В12, который может быть необязателен для некоторых бактерий, как обсуждается выше). Во-вторых, тогда как некоторые бактерии могут синтезировать все (или

по крайней мере большинство) кофакторов самостоятельно (прототрофы), многие другие бактерии не способны этого делать, и для их роста требуется внешний источник витаминных предшественников (ауксотрофы). В-третьих, в отличие от других метаболитов, выступающих в роли источника энергии и углерода (углеводы) или строительных блоков (аминокислоты), витамины группы В нужны бактериям в лишь в очень малых количествах (отсюда происходит слово “микронутриенты”), что говорит в пользу возможного обмена витаминами между различными организмами. В самом деле, известно, что растворимые витамины группы В впитываются клетками верхнего отдела желудочно-кишечного тракта [229], и поэтому обмен витаминами между ауксотрофами и прототрофами был бы правдоподобным объяснением механизма поддержания такого большого бактериального сообщества в человеческом кишечнике. Подробный анализ метаболических возможностей и потребностей необходим для изучения как метаболического обмена между бактериями, так и возможных эффектов питания организма-хозяина на развитие и структуру его кишечного микробиома.

Используя метод геномной реконструкции при помощи подсистем, мы проанализировали метаболические пути биосинтеза, сохранения и транспорта 8 витаминов группы В у бактерий человеческого кишечного микробиома, представленных выборкой из 2228 геномов. 8 реконструированных метаболических подсистем включают более 200 функциональных ролей, представленных ферментами из различных белковых семейств (Рис. 15 и Приложение 1). Выделяя специфические комбинации найденных компонентов биосинтетических путей и транспортеров витаминов (и их предшественников), мы классифицировали все геномы из изучаемой выборки согласно их метаболическим возможностям и предсказали их прототрофные либо ауксотрофные фенотипы. Полученные результаты говорят в пользу возможного межорганизменного обмена витаминами и указывают на специфические семейства транспортеров, которые могут играть роль в таких метаболических взаимодействиях в кишечном микробиоме человека (Рис. 16). Мы выделяем группы ауксотрофных геномов, кодирующих неполные варианты метаболических путей, которые указывают на возможное участие в метаболическом обмене также и предшественников канонических витаминов, например таких, как тиазол, квинолилат, детиобiotин и пантоат. Мы называем такие соединения “альтернативными витаминами” и предполагаем, что они дополняют набор метаболитов, обмен которыми влияет на структуру бактериальных сообществ. Наши предсказания витаминных фенотипов согласуются с опубликованными экспериментальными данными, что указывает на корректность предсказанных метаболических потребностей и возможностей.

С целью проведения сравнений фенотипических профилей образцов микробиомных сообществ, мы перевели все детальные описания вариантов метаболических путей в бинарные фенотипы, принимающие значения либо “1” (прототроф), либо “0” (ауксотроф). Ауксотрофные фенотипы оказались широко распространены среди бактерий кишечного микробиома, и только лишь небольшое количество бактерий может синтезировать все витамины самостоятельно, что поддерживает гипотезу об обмене витаминами в контексте микробиома кишечника (Рис. 17). Анализируя филогенетическое распределение витаминных фенотипов, мы наблюдали, что они консервативны на уровне видов, но существенно варьируют на более высоких таксономических уровнях. Это наблюдение также помогло нам установить порог на степень филогенетического родства организмов, при котором мы допускаем идентичность их фенотипов. В частности, при установке порога сходства 16S рРНК генов в 97% достигается деление бактерий на группы, более близкое к делению на виды, чем к делению на другие филогенетические таксоны; поскольку наблюдаемые нами фенотипы были консервативны на уровне видов, мы допускали, что каждой такой группе (т.е. ОТЕ) мы ставили в соответствие единственный фенотип. Такой метод позволял нам максимизировать покрытие образцов человеческого микробиома группами бактерий с предсказанными фенотипами, минимизируя при этом ошибку предсказаний. В будущем, ожидаемый рост числа секвенированных и проанализированных геномов, так же как и рост точности определения филогении бактерий (например, при секвенировании более длинных участков гена 16S рРНК или при анализе дополнительных филогеномных признаков) увеличит аккуратность предсказаний витаминных фенотипов по филогенетическим данным.

Отображая предсказания витаминных фенотипов на данные секвенирования генов 16S рРНК в образцах кишечного микробиома человека, отражающие филогенетическое разнообразие образцов, мы оценивали средний уровень ауксотрофии в бактериальных сообществах, населяющих кишечник различных индивидуумов. Таким образом мы косвенно оценивали баланс между ауксотрофами и прототрофами в изучаемых бактериальных сообществах. Проанализировав >300 образцов кишечного микробиома, собранных в рамках проекта “Микробиом Человека”, мы получили оценку разнообразия средних уровней ауксотрофии среди разнообразной популяции индивидуумов. Несмотря на то, что в изучаемой нами выборке геномов преобладали ауксотрофные фенотипы, при взвешенном (по численности) подсчете числа ауксотрофов и прототрофов в образцах микробиома человека мы наблюдали, что для 6 из 8 витаминов большинство бактерий в сообществе прототрофны. Исключениями являются биотин и кобаламин. Кобаламин не является необходимым для некоторых бактерий, поэтому

большое количество бактерий - ауксотрофов по кобаламину в бактериальных сообществах ожидаемо. Также мы оценили средние уровни ауксотрофии для образцов кишечного микробиома, собранных у индивидуумов при переходе с одного рациона питания на другой и обратно. При том, что общий филогенетический состав микробиома этих индивидуумов сильно меняется, мы не наблюдали значительных изменений распределения средних уровней ауксотрофии. Это можно объяснить либо тем, что (1) микробиом кишечника обладает устойчивым балансом ауксотрофов и прототрофов, который не меняется даже в случае значительных изменений филогенетического состава сообщества, либо (2) недостаточной точностью измерения соотношения ауксотрофов и прототрофов в нашем исследовании, либо (3) тем, что (3) соотношение ауксотрофов и прототрофов не отражает реальный баланс доноров и акцепторов витаминов в бактериальном сообществе. Чтобы уточнить, какая из этих гипотез верна, предстоит проводить более детальные исследования метаболических взаимодействий в микробиоме кишечника.

Ожидаемое развитие геномных и метагеномных методов анализа микробиома человека, главным образом (1) секвенирование новых геномов, (2) улучшение методов секвенирования образцов микробиома, и (3) дальнейшее изучение метаболических путей представителей микробиома, приведет к увеличению точности оценки баланса между ауксотрофами и прототрофами в образцах кишечного микробиома. Этот подход можно также применять к другим классам метаболитов, например к аминокислотам, сахарам [230] и короткоцепочечным жирным кислотам, а также к неметаболическим фенотипам (например, устойчивость к антибиотикам). К возможным применениям анализа баланса прототрофов и ауксотрофов в образцах кишечного микробиома можно отнести диагностику и терапию заболеваний, связанных с дисбиозом.

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

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**,

Список используемых сокращений и обозначений

НТН – спираль-поворот-спираль
АМ - S-аденозил-L-метионин
АПБ - ацилпереносающий белок
АРИ - 5-аминоимидазол риботид
АТФ – аденозинтрифосфат
ГМП – 4-амино-5-гидроксиметил-2-метилпиримидин
ГМП – 4-амино-5-гидроксиметил-2-метилпиримидин
ГМПФ – 4-амино-5-гидроксиметил-2-метилпиримидин-пирофосфат
ГМПФ – 4-амино-5-гидроксиметил-2-метилпиримидин-фосфат
ГТФ - гуанинтрифосфат
ГЭТ – 5-(2-гидроксиэтил)-4-метилпиримидин
ГЭТ – 5-(2-гидроксиэтил)-4-метилпиримидин
ГЭТФ – 5-(2-гидроксиэтил)-4-метилпиримидин-фосфат
ДАП – 7,8-диаминопеларгоновая кислота
ДНК – дезоксирибонуклеиновая кислота
КАП – 7-кето-8-аминопеларгоновая кислота
КоА – кофермент А
НАД – никотинамидадениндинуклеотид
НАДФ – никотинамидадениндинуклеотидфосфат
НР – никотинамид рибозид
ОТЕ – операционные таксономические единицы
ПАБ – парааминобензоат
ПМВ – позиционная матрица весов
ПФ – пиридоксальфосфат
Р5Ф - рибулозо-5-фосфат
РНК - рибонуклеиновая кислота
ТГФ – тетрагидрофуран
ТПФ – тиамин пирофосфат
ТФ - транскрипционный фактор
ФАД - флавин аденин динуклеотид
ФЕП - фосфоенолпируват
ФМН – флавиномононуклеотид
тРНК – транспортная рибонуклеиновая кислота

Список использованной литературы

1. Crick F. Central Dogma of Molecular Biology // Nature. 1970. Vol. 227, № 5258. P. 561–563.
2. Thurmond J. et al. FlyBase 2.0: the next generation // Nucleic Acids Research. 2019. Vol. 47, № D1. P. D759–D765.

3. Franks A., Airoidi E., Slavov N. Post-transcriptional regulation across human tissues // PLOS Computational Biology / ed. Vogel C. 2017. Vol. 13, № 5. P. e1005535.
4. Browning D.F., Busby S.J.W. The regulation of bacterial transcription initiation // Nature Reviews Microbiology. 2004. Vol. 2, № 1. P. 57–65.
5. Wigneshweraraj S. et al. Modus operandi of the bacterial RNA polymerase containing the σ^{54} promoter-specificity factor // Molecular Microbiology. 2008. Vol. 68, № 3. P. 538–546.
6. Newell K.V. et al. The RNA polymerase-binding protein RbpA confers basal levels of rifampicin resistance on *Streptomyces coelicolor* // Molecular Microbiology. 2006. Vol. 60, № 3. P. 687–696.
7. Paul B.J., Berkmen M.B., Gourse R.L. DksA potentiates direct activation of amino acid promoters by ppGpp // Proceedings of the National Academy of Sciences. 2005. Vol. 102, № 22. P. 7823–7828.
8. Cavanagh A.T., Wassarman K.M. 6S RNA, a Global Regulator of Transcription in *Escherichia coli*, *Bacillus subtilis*, and Beyond // Annual Review of Microbiology. 2014. Vol. 68, № 1. P. 45–60.
9. Swint-Kruse L., Matthews K.S. Allostery in the LacI/GalR family: variations on a theme // Current Opinion in Microbiology. 2009. Vol. 12, № 2. P. 129–137.
10. Krebs J.E., Goldstein E.S., Kilpatrick S.T. Lewin's genes XII. Burlington, MA: Jones & Bartlett Learning, 2018. 837 p.
11. He X., Duque T.S.P.C., Sinha S. Evolutionary Origins of Transcription Factor Binding Site Clusters // Molecular Biology and Evolution. 2012. Vol. 29, № 3. P. 1059–1070.
12. Guenther U.-P. et al. Hidden specificity in an apparently nonspecific RNA-binding protein // Nature. 2013. Vol. 502, № 7471. P. 385–388.
13. Reid S.J., Abratt V.R. Sucrose utilisation in bacteria: genetic organisation and regulation // Applied Microbiology and Biotechnology. 2005. Vol. 67, № 3. P. 312–321.
14. Bintu L. et al. Transcriptional regulation by the numbers: models // Current Opinion in Genetics & Development. 2005. Vol. 15, № 2. P. 116–124.
15. Kamenšek S. et al. Silencing of DNase Colicin E8 Gene Expression by a Complex Nucleoprotein Assembly Ensures Timely Colicin Induction // PLOS Genetics / ed. Hughes D. 2015. Vol. 11, № 6. P. e1005354.
16. Valentin-Hansen P., Sogaard-Andersen L., Pedersen H. A flexible partnership: the CytR anti-activator and the cAMP-CRP activator protein, comrades in transcription control // Molecular Microbiology. 1996. Vol. 20, № 3. P. 461–466.
17. Lee D.J., Minchin S.D., Busby S.J.W. Activating Transcription in Bacteria // Annual Review of Microbiology. 2012. Vol. 66, № 1. P. 125–152.
18. Philips S.J. et al. Allosteric transcriptional regulation via changes in the overall topology of the core promoter // Science. 2015. Vol. 349, № 6250. P. 877–881.
19. Vigouroux A. et al. Tuning dCas9's ability to block transcription enables robust, noiseless knockdown of bacterial genes // Molecular Systems Biology. 2018. Vol. 14, № 3. P. e7899.
20. Qi X. et al. The applications of CRISPR screen in functional genomics // Briefings in Functional Genomics. 2017. Vol. 16, № 1. P. 34–37.
21. Ness R.O., Sachs K., Vitek O. From Correlation to Causality: Statistical Approaches to Learning Regulatory Relationships in Large-Scale Biomolecular Investigations // Journal of Proteome Research. 2016. Vol. 15, № 3. P. 683–690.
22. Marinov G.K. A decade of ChIP-seq // Briefings in Functional Genomics. 2018. Vol. 17, № 2. P. 77–79.
23. Nakato R., Shirahige K. Recent advances in ChIP-seq analysis: from quality management to whole-genome annotation // Briefings in Bioinformatics. 2016. P. bbw023.

24. Genome List - Genome - NCBI [Electronic resource]. URL: <https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/> (accessed: 25.01.2019).
25. Levy S., Hannehalli S., Workman C. Enrichment of regulatory signals in conserved non-coding genomic sequence // *Bioinformatics*. 2001. Vol. 17, № 10. P. 871–877.
26. Ma H.T. et al. An inducible system for expression and validation of the specificity of short hairpin RNA in mammalian cells // *Nucleic Acids Research*. 2007. Vol. 35, № 4. P. e22–e22.
27. Conant G.C., Wolfe K.H. Turning a hobby into a job: How duplicated genes find new functions // *Nature Reviews Genetics*. 2008. Vol. 9, № 12. P. 938–950.
28. El-Gebali S. et al. The Pfam protein families database in 2019 // *Nucleic Acids Research*. 2019. Vol. 47, № D1. P. D427–D432.
29. Rodionov D.A. Comparative genomic reconstruction of transcriptional regulatory networks in bacteria // *Chem. Rev.* 2007. Vol. 107, № 8. P. 3467–3497.
30. Ureta-Vidal A., Ettwiller L., Birney E. Comparative genomics: genome-wide analysis in metazoan eukaryotes // *Nature Reviews Genetics*. 2003. Vol. 4, № 4. P. 251–262.
31. Гомология (биология) // Википедия. 2018.
32. Blanchette M., Tompa M. FootPrinter: A program designed for phylogenetic footprinting // *Nucleic Acids Res.* 2003. Vol. 31, № 13. P. 3840–3842.
33. Liu B. et al. An integrative and applicable phylogenetic footprinting framework for cis-regulatory motifs identification in prokaryotic genomes // *BMC Genomics*. 2016. Vol. 17, № 1.
34. Bailey T.L. et al. MEME SUITE: tools for motif discovery and searching // *Nucleic Acids Res.* 2009. Vol. 37, № Web Server issue. P. W202–208.
35. Claeys M. et al. MotifSuite: workflow for probabilistic motif detection and assessment // *Bioinformatics*. 2012. Vol. 28, № 14. P. 1931–1932.
36. Schumacher M.A. et al. 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 Research*. 2010. Vol. 39, № 7. P. 2931–2942.
37. Roberson E.D.O. Motif scraper: a cross-platform, open-source tool for identifying degenerate nucleotide motif matches in FASTA files // *Bioinformatics* / ed. Hancock J. 2018. Vol. 34, № 22. P. 3926–3928.
38. Crooks G.E. WebLogo: A Sequence Logo Generator // *Genome Research*. 2004. Vol. 14, № 6. P. 1188–1190.
39. Zhou Q., Liu J.S. Modeling within-motif dependence for transcription factor binding site predictions // *Bioinformatics*. 2004. Vol. 20, № 6. P. 909–916.
40. Zhao Y. et al. Improved Models for Transcription Factor Binding Site Identification Using Nonindependent Interactions // *Genetics*. 2012. Vol. 191, № 3. P. 781–790.
41. Zhao Y., Stormo G.D. Quantitative analysis demonstrates most transcription factors require only simple models of specificity // *Nature Biotechnology*. 2011. Vol. 29, № 6. P. 480–483.
42. Narlikar L., Ovcharenko I. Identifying regulatory elements in eukaryotic genomes // *Briefings in Functional Genomics and Proteomics*. 2009. Vol. 8, № 4. P. 215–230.
43. Rodionov D.A. Comparative genomic reconstruction of transcriptional regulatory networks in bacteria // *Chem. Rev.* 2007. Vol. 107, № 8. P. 3467–3497.
44. Oehler S. et al. The three operators of the lac operon cooperate in repression. // *The EMBO journal*. 1990. Vol. 9, № 4. P. 973.
45. Camas F.M., Poyatos J.F. What Determines the Assembly of Transcriptional Network Motifs in *Escherichia coli*? // *PLoS ONE* / ed. Isalan M. 2008. Vol. 3, № 11. P. e3657.
46. Kranz R.G., Foster-Hartnett D. Transcriptional regulatory cascade of nitrogen-fixation genes in anoxygenic photosynthetic bacteria: oxygen-and nitrogen-responsive factors // *Molecular Microbiology*. 1990. Vol. 4, № 11. P. 1793–1800.

47. Fukami-Kobayashi K. Parallel Evolution of Ligand Specificity Between LacI/GalR Family Repressors and Periplasmic Sugar-Binding Proteins // *Molecular Biology and Evolution*. 2003. Vol. 20, № 2. P. 267–277.
48. Novichkov P.S. et al. RegPrecise 3.0 – A resource for genome-scale exploration of transcriptional regulation in bacteria // *BMC Genomics*. 2013. Vol. 14, № 1. P. 745.
49. Overbeek R. The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes // *Nucleic Acids Research*. 2005. Vol. 33, № 17. P. 5691–5702.
50. Dam P. et al. Operon prediction using both genome-specific and general genomic information // *Nucleic Acids Research*. 2007. Vol. 35, № 1. P. 288–298.
51. Wu H. Prediction of functional modules based on comparative genome analysis and Gene Ontology application // *Nucleic Acids Research*. 2005. Vol. 33, № 9. P. 2822–2837.
52. Hill C. et al. The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic // *Nature Reviews Gastroenterology & Hepatology*. 2014. Vol. 11, № 8. P. 506–514.
53. Metchnikoff É. *The nature of man: Studies in optimistic philosophy*. 1908.
54. Ouwehand A.C., Salminen S., Isolauri E. // *Antonie van Leeuwenhoek*. 2002. Vol. 82, № 1/4. P. 279–289.
55. Stanton C. et al. Fermented functional foods based on probiotics and their biogenic metabolites // *Current Opinion in Biotechnology*. 2005. Vol. 16, № 2. P. 198–203.
56. Bezirtzoglou E., Stavropoulou E. Immunology and probiotic impact of the newborn and young children intestinal microflora // *Anaerobe*. 2011. Vol. 17, № 6. P. 369–374.
57. Goldin B.R. Health benefits of probiotics // *Br. J. Nutr.* 1998. Vol. 80, № 4. P. S203–207.
58. Turrone F., van Sinderen D., Ventura M. Genomics and ecological overview of the genus *Bifidobacterium* // *International Journal of Food Microbiology*. 2011. Vol. 149, № 1. P. 37–44.
59. Ventura M. et al. Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum // *Microbiology and Molecular Biology Reviews*. 2007. Vol. 71, № 3. P. 495–548.
60. Pokusaeva K. et al. Cellodextrin Utilization by *Bifidobacterium breve* UCC2003 // *Applied and Environmental Microbiology*. 2011. Vol. 77, № 5. P. 1681–1690.
61. Lamendella R. et al. *Bifidobacteria* in Feces and Environmental Waters // *Applied and Environmental Microbiology*. 2007. Vol. 74, № 3. P. 575–584.
62. Lee J.-H., O’Sullivan D.J. Genomic Insights into *Bifidobacteria* // *Microbiology and Molecular Biology Reviews*. 2010. Vol. 74, № 3. P. 378–416.
63. Pokusaeva K., Fitzgerald G.F., van Sinderen D. Carbohydrate metabolism in *Bifidobacteria* // *Genes & Nutrition*. 2011. Vol. 6, № 3. P. 285–306.
64. Bertelli C. et al. *Bifidobacterium longum* Bacteremia in Preterm Infants Receiving Probiotics // *Clinical Infectious Diseases*. 2014. Vol. 60, № 6. P. 924–927.
65. Weber E. et al. *Bifidobacterium* Species Bacteremia: Risk Factors in Adults and Infants: Table 1. // *Clinical Infectious Diseases*. 2015. Vol. 61, № 3. P. 482–484.
66. Russell D.A. et al. Metabolic activities and probiotic potential of *bifidobacteria* // *International Journal of Food Microbiology*. 2011. Vol. 149, № 1. P. 88–105.
67. Guarner F., Malagelada J.-R. Gut flora in health and disease // *The Lancet*. 2003. Vol. 361, № 9356. P. 512–519.
68. Vaughan E.E. et al. Diversity, vitality and activities of intestinal lactic acid bacteria and *bifidobacteria* assessed by molecular approaches // *FEMS Microbiology Reviews*. 2005. Vol. 29, № 3. P. 477–490.
69. Bindels L.B. et al. Towards a more comprehensive concept for prebiotics // *Nature Reviews Gastroenterology & Hepatology*. 2015. Vol. 12, № 5. P. 303–310.

70. Rastall R.A., Gibson G.R. Recent developments in prebiotics to selectively impact beneficial microbes and promote intestinal health // *Current Opinion in Biotechnology*. 2015. Vol. 32. P. 42–46.
71. Ventura M. et al. Genomics as a means to understand bacterial phylogeny and ecological adaptation: the case of bifidobacteria // *Antonie van Leeuwenhoek*. 2006. Vol. 91, № 4. P. 351–372.
72. Milani C. et al. Bifidobacteria exhibit social behavior through carbohydrate resource sharing in the gut // *Scientific Reports*. 2015. Vol. 5, № 1.
73. Turrone F. et al. Glycan cross-feeding activities between bifidobacteria under in vitro conditions // *Frontiers in Microbiology*. 2015. Vol. 6.
74. Trindade M.I., Abratt V.R., Reid S.J. Induction of Sucrose Utilization Genes from *Bifidobacterium lactis* by Sucrose and Raffinose // *Applied and Environmental Microbiology*. 2003. Vol. 69, № 1. P. 24–32.
75. Gilad O. et al. Combined Transcriptome and Proteome Analysis of *Bifidobacterium animalis* subsp. *lactis* BB-12 Grown on Xylo-Oligosaccharides and a Model of Their Utilization // *Applied and Environmental Microbiology*. 2010. Vol. 76, № 21. P. 7285–7291.
76. Pokusaeva K. et al. Ribose utilization by the human commensal *Bifidobacterium breve* UCC2003 // *Microbial Biotechnology*. 2010. Vol. 3, № 3. P. 311–323.
77. O’Connell Motherway M., Fitzgerald G.F., van Sinderen D. Metabolism of a plant derived galactose-containing polysaccharide by *Bifidobacterium breve* UCC2003 // *Microbial Biotechnology*. 2010. Vol. 4, № 3. P. 403–416.
78. O’Connell K.J. et al. Transcription of Two Adjacent Carbohydrate Utilization Gene Clusters in *Bifidobacterium breve* UCC2003 Is Controlled by LacI- and Repressor Open Reading Frame Kinase (ROK)-Type Regulators // *Applied and Environmental Microbiology*. 2014. Vol. 80, № 12. P. 3604–3614.
79. De Bruyn F. et al. Unraveling the Leloir Pathway of *Bifidobacterium bifidum*: Significance of the Uridyltransferases // *Applied and Environmental Microbiology*. 2013. Vol. 79, № 22. P. 7028–7035.
80. Egan M., O’Connell Motherway M., van Sinderen D. A GntR-type transcriptional repressor controls sialic acid utilization in *Bifidobacterium breve* UCC2003 // *FEMS Microbiology Letters*. 2015. Vol. 362, № 4. P. 1–9.
81. Jacob F., Monod J. Genetic regulatory mechanisms in the synthesis of proteins // *Journal of Molecular Biology*. 1961. Vol. 3, № 3. P. 318–356.
82. Weickert M.J., Adhya S. A family of bacterial regulators homologous to Gal and Lac repressors // *J. Biol. Chem*. 1992. Vol. 267, № 22. P. 15869–15874.
83. Nguyen C.C., Saier M.H. Phylogenetic, structural and functional analyses of the LacI-GalR family of bacterial transcription factors // *FEBS Letters*. 1995. Vol. 377, № 2. P. 98–102.
84. Mauzy C.A., Hermodson M.A. Structural homology between repressor and ribose binding protein implies functional similarity // *Protein Science*. 1992. Vol. 1, № 7. P. 843–849.
85. Franco I.S. et al. Functional Domains of the *Bacillus subtilis* Transcription Factor AraR and Identification of Amino Acids Important for Nucleoprotein Complex Assembly and Effector Binding // *Journal of Bacteriology*. 2006. Vol. 188, № 8. P. 3024–3036.
86. Mirny L.A., Gelfand M.S. Using Orthologous and Paralogous Proteins to Identify Specificity-determining Residues in Bacterial Transcription Factors // *Journal of Molecular Biology*. 2002. Vol. 321, № 1. P. 7–20.
87. Kalinina O.V. et al. SDPpred: a tool for prediction of amino acid residues that determine differences in functional specificity of homologous proteins // *Nucleic Acids Research*. 2004. Vol. 32, № Web Server. P. W424–W428.

88. Pei J. et al. Prediction of functional specificity determinants from protein sequences using log-likelihood ratios // *Bioinformatics*. 2005. Vol. 22, № 2. P. 164–171.
89. Tungtur S., Egan S.M., Swint-Kruse L. Functional consequences of exchanging domains between LacI and PurR are mediated by the intervening linker sequence // *Proteins: Structure, Function, and Bioinformatics*. 2007. Vol. 68, № 1. P. 375–388.
90. Parente D.J., Swint-Kruse L. Multiple Co-Evolutionary Networks Are Supported by the Common Tertiary Scaffold of the LacI/GalR Proteins // *PLoS ONE*. 2013. Vol. 8, № 12. P. e84398.
91. Structure, function and diversity of the healthy human microbiome // *Nature*. 2012. Vol. 486, № 7402. P. 207–214.
92. Yatsunenko T. et al. Human gut microbiome viewed across age and geography // *Nature*. 2012. Vol. 486, № 7402. P. 222–227.
93. Backhed F. et al. The gut microbiota as an environmental factor that regulates fat storage // *Proceedings of the National Academy of Sciences*. 2004. Vol. 101, № 44. P. 15718–15723.
94. Ley R.E. et al. Obesity alters gut microbial ecology // *Proceedings of the National Academy of Sciences*. 2005. Vol. 102, № 31. P. 11070–11075.
95. MacDonald T.T., Gordon J.N. Bacterial Regulation of Intestinal Immune Responses // *Gastroenterology Clinics of North America*. 2005. Vol. 34, № 3. P. 401–412.
96. Ley R.E. et al. Human gut microbes associated with obesity // *Nature*. 2006. Vol. 444, № 7122. P. 1022–1023.
97. David L.A. et al. Diet rapidly and reproducibly alters the human gut microbiome // *Nature*. 2013. Vol. 505, № 7484. P. 559–563.
98. Zhang J. et al. Mongolians core gut microbiota and its correlation with seasonal dietary changes // *Scientific Reports*. 2014. Vol. 4, № 1.
99. Sheflin A.M. et al. Linking dietary patterns with gut microbial composition and function // *Gut Microbes*. 2016. Vol. 8, № 2. P. 113–129.
100. McCormick D.B. Metabolism of vitamins in microbes and mammals // *Biochemical and Biophysical Research Communications*. 2003. Vol. 312, № 1. P. 97–101.
101. Lin S., Cronan J.E. Closing in on complete pathways of biotin biosynthesis // *Molecular BioSystems*. 2011. Vol. 7, № 6. P. 1811.
102. Sañudo-Wilhelmy S.A. et al. The Role of B Vitamins in Marine Biogeochemistry // *Annual Review of Marine Science*. 2014. Vol. 6, № 1. P. 339–367.
103. Romine M.F. et al. Underlying mechanisms for syntrophic metabolism of essential enzyme cofactors in microbial communities // *The ISME Journal*. 2017. Vol. 11, № 6. P. 1434–1446.
104. Benson D.A. et al. GenBank // *Nucleic Acids Research*. 2014. Vol. 43, № D1. P. D30–D35.
105. Finn R.D. et al. Pfam: the protein families database // *Nucleic Acids Research*. 2013. Vol. 42, № D1. P. D222–D230.
106. Guindon S. et al. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0 // *Systematic Biology*. 2010. Vol. 59, № 3. P. 307–321.
107. Novichkov P.S. et al. RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach // *Nucleic Acids Research*. 2010. Vol. 38, № Web Server. P. W299–W307.
108. Dehal P.S. et al. MicrobesOnline: an integrated portal for comparative and functional genomics // *Nucleic Acids Research*. 2009. Vol. 38, № Database. P. D396–D400.
109. Markowitz V.M. et al. IMG 4 version of the integrated microbial genomes comparative analysis system // *Nucleic Acids Research*. 2013. Vol. 42, № D1. P. D560–D567.

110. Ravcheev D.A. et al. Polysaccharides utilization in human gut bacterium *Bacteroides thetaiotaomicron*: comparative genomics reconstruction of metabolic and regulatory networks // *BMC Genomics*. 2013. Vol. 14, № 1. P. 873.
111. Larkin M.A. et al. Clustal W and Clustal X version 2.0 // *Bioinformatics*. 2007. Vol. 23, № 21. P. 2947–2948.
112. Mironov A.A., Vinokurova N.P., Gelfand M.S. Software for analysis of bacterial genomes // *Molecular Biology*. 2000. Vol. 34, № 2. P. 222–231.
113. Rodionov D.A. et al. Genomic encyclopedia of sugar utilization pathways in the *Shewanella* genus // *BMC Genomics*. 2010. Vol. 11, № 1. P. 494.
114. Overbeek R. et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST) // *Nucleic Acids Research*. 2013. Vol. 42, № D1. P. D206–D214.
115. UniProt: a hub for protein information // *Nucleic Acids Research*. 2014. Vol. 43, № D1. P. D204–D212.
116. Altschul S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs // *Nucleic Acids Research*. 1997. Vol. 25, № 17. P. 3389–3402.
117. Lombard V. et al. The carbohydrate-active enzymes database (CAZy) in 2013 // *Nucleic Acids Research*. 2013. Vol. 42, № D1. P. D490–D495.
118. Rodionov D.A. et al. Transcriptional regulation of the carbohydrate utilization network in *Thermotoga maritima* // *Frontiers in Microbiology*. 2013. Vol. 4.
119. Ravcheev D.A. et al. Inference of the Transcriptional Regulatory Network in *Staphylococcus aureus* by Integration of Experimental and Genomics-Based Evidence // *Journal of Bacteriology*. 2011. Vol. 193, № 13. P. 3228–3240.
120. Leyn S.A. et al. Genomic Reconstruction of the Transcriptional Regulatory Network in *Bacillus subtilis* // *Journal of Bacteriology*. 2013. Vol. 195, № 11. P. 2463–2473.
121. Cuthbertson L., Nodwell J.R. The TetR Family of Regulators // *Microbiology and Molecular Biology Reviews*. 2013. Vol. 77, № 3. P. 440–475.
122. Marion C. et al. Identification of an ATPase, MsmK, Which Energizes Multiple Carbohydrate ABC Transporters in *Streptococcus pneumoniae* // *Infection and Immunity*. 2011. Vol. 79, № 10. P. 4193–4200.
123. Lugli G.A. et al. Investigation of the Evolutionary Development of the Genus *Bifidobacterium* by Comparative Genomics // *Applied and Environmental Microbiology*. 2014. Vol. 80, № 20. P. 6383–6394.
124. Milani C. et al. Genomic Encyclopedia of Type Strains of the Genus *Bifidobacterium* // *Applied and Environmental Microbiology*. 2014. Vol. 80, № 20. P. 6290–6302.
125. Sun Z. et al. Comparative Genomic Analysis of 45 Type Strains of the Genus *Bifidobacterium*: A Snapshot of Its Genetic Diversity and Evolution // *PLOS ONE*. 2015. Vol. 10, № 2. P. e0117912.
126. Leyn S.A. et al. Control of Proteobacterial Central Carbon Metabolism by the HexR Transcriptional Regulator // *Journal of Biological Chemistry*. 2011. Vol. 286, № 41. P. 35782–35794.
127. Cipriano M.J. et al. RegTransBase – a database of regulatory sequences and interactions based on literature: a resource for investigating transcriptional regulation in prokaryotes // *BMC Genomics*. 2013. Vol. 14, № 1. P. 213.
128. Salgado H. et al. RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more // *Nucleic Acids Research*. 2012. Vol. 41, № D1. P. D203–D213.

129. Siervo N. et al. DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information // *Nucleic Acids Research*. 2007. Vol. 36, № suppl_1. P. D93–D96.
130. Pauling J. et al. CoryneRegNet 6.0--Updated database content, new analysis methods and novel features focusing on community demands // *Nucleic Acids Research*. 2011. Vol. 40, № D1. P. D610–D614.
131. Sonenshein A.L. Control of key metabolic intersections in *Bacillus subtilis* // *Nature Reviews Microbiology*. 2007. Vol. 5, № 12. P. 917–927.
132. FUJITA Y. Carbon Catabolite Control of the Metabolic Network in *Bacillus subtilis* // *Bioscience, Biotechnology, and Biochemistry*. 2009. Vol. 73, № 2. P. 245–259.
133. Seidl K. et al. Effect of a glucose impulse on the CcpA regulon in *Staphylococcus aureus* // *BMC Microbiology*. 2009. Vol. 9, № 1. P. 95.
134. Mahr K., Hillen W., Titgemeyer F. Carbon Catabolite Repression in *Lactobacillus pentosus*: Analysis of the ccpA Region // *Applied and Environmental Microbiology*. 2000. Vol. 66, № 1. P. 277–283.
135. Zheng L. et al. CcpA regulates biofilm formation and competence in *Streptococcus gordonii* // *Molecular Oral Microbiology*. 2011. Vol. 27, № 2. P. 83–94.
136. Zheng L. et al. Catabolite Control Protein A Controls Hydrogen Peroxide Production and Cell Death in *Streptococcus sanguinis* // *Journal of Bacteriology*. 2010. Vol. 193, № 2. P. 516–526.
137. Zotta T. et al. Inactivation of ccpA and aeration affect growth, metabolite production and stress tolerance in *Lactobacillus plantarum* WCFS1 // *International Journal of Food Microbiology*. 2012. Vol. 155, № 1–2. P. 51–59.
138. Ravcheev D.A. et al. Genomic reconstruction of transcriptional regulatory networks in lactic acid bacteria // *BMC Genomics*. 2013. Vol. 14, № 1. P. 94.
139. Antunes A. et al. Global transcriptional control by glucose and carbon regulator CcpA in *Clostridium difficile* // *Nucleic Acids Research*. 2012. Vol. 40, № 21. P. 10701–10718.
140. Ravcheev D.A. et al. // *Russian Journal of Genetics*. 2002. Vol. 38, № 9. P. 1015–1025.
141. Cho B.-K. et al. The PurR regulon in *Escherichia coli* K-12 MG1655 // *Nucleic Acids Research*. 2011. Vol. 39, № 15. P. 6456–6464.
142. Francke C. et al. A generic approach to identify Transcription Factor-specific operator motifs; Inferences for LacI-family mediated regulation in *Lactobacillus plantarum* WCFS1 // *BMC Genomics*. 2008. Vol. 9, № 1. P. 145.
143. Camas F.M., Alm E.J., Poyatos J.F. Local Gene Regulation Details a Recognition Code within the LacI Transcriptional Factor Family // *PLoS Computational Biology*. 2010. Vol. 6, № 11. P. e1000989.
144. Milk L., Daber R., Lewis M. Functional rules for lac repressor-operator associations and implications for protein-DNA interactions // *Protein Science*. 2010. Vol. 19, № 6. P. 1162–1172.
145. Lewis M. The lac repressor // *Comptes Rendus Biologies*. 2005. Vol. 328, № 6. P. 521–548.
146. He B., Smith J.M., Zalkin H. *Escherichia coli* purB gene: cloning, nucleotide sequence, and regulation by purR. // *Journal of Bacteriology*. 1992. Vol. 174, № 1. P. 130–136.
147. He B., Choi K.Y., Zalkin H. Regulation of *Escherichia coli* glnB, prsA, and speA by the purine repressor. // *Journal of Bacteriology*. 1993. Vol. 175, № 11. P. 3598–3606.
148. Tsunedomi R. et al. The Activator of GntII Genes for Gluconate Metabolism, GntH, Exerts Negative Control of GntR-Regulated GntI Genes in *Escherichia coli* // *Journal of Bacteriology*. 2003. Vol. 185, № 6. P. 1783–1795.
149. Schumacher M. et al. Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices // *Science*. 1994. Vol. 266, № 5186. P. 763–770.

150. Barbier C.S., Short S.A., Seneor D.F. Allosteric Mechanism of Induction of CytR-regulated Gene Expression // *Journal of Biological Chemistry*. 1997. Vol. 272, № 27. P. 16962–16971.
151. Kalodimos C.G. Plasticity in protein-DNA recognition: lac repressor interacts with its natural operator O1 through alternative conformations of its DNA-binding domain // *The EMBO Journal*. 2002. Vol. 21, № 12. P. 2866–2876.
152. Saier M.H., Ramseier T.M. The catabolite repressor/activator (Cra) protein of enteric bacteria. // *Journal of Bacteriology*. 1996. Vol. 178, № 12. P. 3411–3417.
153. Perez J.C., Groisman E.A. Transcription factor function and promoter architecture govern the evolution of bacterial regulons // *Proceedings of the National Academy of Sciences*. 2009. Vol. 106, № 11. P. 4319–4324.
154. diCenzo G.C. et al. A Key Regulator of the Glycolytic and Gluconeogenic Central Metabolic Pathways in *Sinorhizobium meliloti* // *Genetics*. 2017. Vol. 207, № 3. P. 961–974.
155. Imam S., Noguera D.R., Donohue T.J. CceR and AkgR Regulate Central Carbon and Energy Metabolism in Alphaproteobacteria // *mBio*. 2015. Vol. 6, № 1.
156. Hottes A.K. et al. Transcriptional Profiling of *Caulobacter crescentus* during Growth on Complex and Minimal Media // *Journal of Bacteriology*. 2004. Vol. 186, № 5. P. 1448–1461.
157. Kazanov M.D. et al. Functional diversification of ROK-family transcriptional regulators of sugar catabolism in the Thermotogae phylum // *Nucleic Acids Research*. 2012. Vol. 41, № 2. P. 790–803.
158. Tsunedomi R. et al. Dual Control by Regulators, GntH and GntR, of the GntII Genes for Gluconate Metabolism in *Escherichia coli* // *Journal of Molecular Microbiology and Biotechnology*. 2003. Vol. 6, № 1. P. 41–56.
159. Qin J. et al. A human gut microbial gene catalogue established by metagenomic sequencing // *Nature*. 2010. Vol. 464, № 7285. P. 59–65.
160. Rajilić-Stojanović M., de Vos W.M. The first 1000 cultured species of the human gastrointestinal microbiota // *FEMS Microbiology Reviews*. 2014. Vol. 38, № 5. P. 996–1047.
161. Wattam A.R. et al. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center // *Nucleic Acids Research*. 2016. Vol. 45, № D1. P. D535–D542.
162. Edgar R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput // *Nucleic Acids Research*. 2004. Vol. 32, № 5. P. 1792–1797.
163. Stamatakis A. Using RAxML to Infer Phylogenies // *Current Protocols in Bioinformatics*. John Wiley & Sons, Inc., 2015. P. 6.14.1-6.14.14.
164. Letunic I., Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees // *Nucleic Acids Research*. 2016. Vol. 44, № W1. P. W242–W245.
165. Rodionov D.A. Conservation of the Biotin Regulon and the BirA Regulatory Signal in Eubacteria and Archaea // *Genome Research*. 2002. Vol. 12, № 10. P. 1507–1516.
166. Rodionov D.A. et al. Comparative Genomics of Thiamin Biosynthesis in Prokaryotes // *Journal of Biological Chemistry*. 2002. Vol. 277, № 50. P. 48949–48959.
167. Vitreschak A.G. Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation // *Nucleic Acids Research*. 2002. Vol. 30, № 14. P. 3141–3151.
168. Rodionov D.A. et al. Comparative Genomics of the Vitamin B12 Metabolism and Regulation in Prokaryotes // *Journal of Biological Chemistry*. 2003. Vol. 278, № 42. P. 41148–41159.
169. Tanaka T., Tateno Y., Gojobori T. Evolution of Vitamin B6 (Pyridoxine) Metabolism by Gain and Loss of Genes // *Molecular Biology and Evolution*. 2004. Vol. 22, № 2. P. 243–250.
170. Rodionov D.A. et al. Transcriptional regulation of NAD metabolism in bacteria: NrtR family of Nudix-related regulators // *Nucleic Acids Research*. 2008. Vol. 36, № 6. P. 2047–2059.

171. Zhang Y. et al. Comparative genomic analyses of nickel, cobalt and vitamin B12 utilization // *BMC Genomics*. 2009. Vol. 10, № 1. P. 78.
172. de Crécy-Lagard V. Variations in metabolic pathways create challenges for automated metabolic reconstructions: Examples from the tetrahydrofolate synthesis pathway // *Computational and Structural Biotechnology Journal*. 2014. Vol. 10, № 16. P. 41–50.
173. Rodionova I.A. et al. Genomic distribution of B-vitamin auxotrophy and uptake transporters in environmental bacteria from the Chloroflexiphylum // *Environmental Microbiology Reports*. 2014. Vol. 7, № 2. P. 204–210.
174. Kanehisa M. et al. Data, information, knowledge and principle: back to metabolism in KEGG // *Nucleic Acids Research*. 2013. Vol. 42, № D1. P. D199–D205.
175. Gelfand M.S., Rodionov D.A. Comparative genomics and functional annotation of bacterial transporters // *Physics of Life Reviews*. 2008. Vol. 5, № 1. P. 22–49.
176. Osterman A. Missing genes in metabolic pathways: a comparative genomics approach // *Current Opinion in Chemical Biology*. 2003. Vol. 7, № 2. P. 238–251.
177. Omelchenko M.V. et al. Non-homologous isofunctional enzymes: A systematic analysis of alternative solutions in enzyme evolution // *Biology Direct*. 2010. Vol. 5, № 1. P. 31.
178. Overbeek R. et al. Annotation of Bacterial and Archaeal Genomes: Improving Accuracy and Consistency // *Chemical Reviews*. 2007. Vol. 107, № 8. P. 3431–3447.
179. Haft D.H. Using comparative genomics to drive new discoveries in microbiology // *Current Opinion in Microbiology*. 2015. Vol. 23. P. 189–196.
180. Sun E.I. et al. Comparative genomics of metabolic capacities of regulons controlled by cis-regulatory RNA motifs in bacteria // *BMC Genomics*. 2013. Vol. 14, № 1. P. 597.
181. Suvorova I.A., Rodionov D.A. Comparative genomics of pyridoxal 5'-phosphate-dependent transcription factor regulons in Bacteria // *Microbial Genomics*. 2016. Vol. 2, № 1.
182. Rodionov D.A. et al. Transcriptional regulation of NAD metabolism in bacteria: genomic reconstruction of NiaR (YrxA) regulon // *Nucleic Acids Research*. 2008. Vol. 36, № 6. P. 2032–2046.
183. Brune I. et al. Negative transcriptional control of biotin metabolism genes by the TetR-type regulator BioQ in biotin-auxotrophic *Corynebacterium glutamicum* ATCC 13032 // *Journal of Biotechnology*. 2012. Vol. 159, № 3. P. 225–234.
184. Nawrocki E.P., Kolbe D.L., Eddy S.R. Infernal 1.0: inference of RNA alignments // *Bioinformatics*. 2009. Vol. 25, № 10. P. 1335–1337.
185. Burge S.W. et al. Rfam 11.0: 10 years of RNA families // *Nucleic Acids Research*. 2012. Vol. 41, № D1. P. D226–D232.
186. Ye Y. et al. Automatic detection of subsystem/pathway variants in genome analysis // *Bioinformatics*. 2005. Vol. 21, № Suppl 1. P. i478–i486.
187. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome // *Nature*. 2012. Vol. 486, № 7402. P. 207–214.
188. David L.A. et al. Diet rapidly and reproducibly alters the human gut microbiome // *Nature*. 2014. Vol. 505, № 7484. P. 559–563.
189. Caporaso J.G. et al. QIIME allows analysis of high-throughput community sequencing data // *Nature Methods*. 2010. Vol. 7, № 5. P. 335–336.
190. Eckburg P.B. et al. Diversity of the human intestinal microbial flora // *Science*. 2005. Vol. 308, № 5728. P. 1635–1638.
191. Schloss P.D., Handelsman J. Introducing DOTUR, a Computer Program for Defining Operational Taxonomic Units and Estimating Species Richness // *Applied and Environmental Microbiology*. 2005. Vol. 71, № 3. P. 1501–1506.
192. Gevers D. et al. Opinion: Re-evaluating prokaryotic species // *Nat. Rev. Microbiol.* 2005. Vol. 3, № 9. P. 733–739.

193. Lagesen K. et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes // *Nucleic Acids Res.* 2007. Vol. 35, № 9. P. 3100–3108.
194. Sorci L. et al. Genomics and Enzymology of NAD Biosynthesis // *Comprehensive Natural Products II.* 2010. P. 213–257.
195. Hugenschmidt S. et al. Screening of a natural biodiversity of lactic and propionic acid bacteria for folate and vitamin B12 production in supplemented whey permeate // *International Dairy Journal.* 2010. Vol. 20, № 12. P. 852–857.
196. Santos F. et al. High-Level Folate Production in Fermented Foods by the B12 Producer *Lactobacillus reuteri* JCM1112 // *Applied and Environmental Microbiology.* 2008. Vol. 74, № 10. P. 3291–3294.
197. Wegkamp A. et al. Development of a minimal growth medium for *Lactobacillus plantarum*: Minimal medium for *Lactobacillus plantarum* // *Letters in Applied Microbiology.* 2010. Vol. 50, № 1. P. 57–64.
198. Sorci L. et al. Quinolinate Salvage and Insights for Targeting NAD Biosynthesis in Group A Streptococci // *Journal of Bacteriology.* 2012. Vol. 195, № 4. P. 726–732.
199. Strauss E., de Villiers M., Rootman I. Biocatalytic Production of Coenzyme A Analogues // *ChemCatChem.* 2010. Vol. 2, № 8. P. 929–937.
200. Marquet A. Biosynthesis of Biotin // *Comprehensive Natural Products II.* 2010. P. 161–180.
201. Degan P.H., Taga M.E., Goodman A.L. Vitamin B 12 as a Modulator of Gut Microbial Ecology // *Cell Metabolism.* 2014. Vol. 20, № 5. P. 769–778.
202. Rowe J.J., Lemmon R.D., Tritz G.J. Nicotinic acid transport in *Escherichia coli* // *Microbios.* 1985. Vol. 44, № 179–180. P. 169–184.
203. Ernst D.C., Downs D.M. The STM4195 Gene Product (PanS) Transports Coenzyme A Precursors in *Salmonella enterica* // *Journal of Bacteriology.* 2015. Vol. 197, № 8. P. 1368–1377.
204. Prakash O., Eisenberg M.A. Active transport of biotin in *Escherichia coli* K-12 // *J. Bacteriol.* 1974. Vol. 120, № 2. P. 785–791.
205. OHSUGI M. et al. Biosynthesis of biotin-vitimers by family enterobacteriaceae. // *Journal of Nutritional Science and Vitaminology.* 1990. Vol. 36, № 5. P. 447–456.
206. Kenley J.S., Leighton M., Bradbeer C. Transport of vitamin B12 in *Escherichia coli*. Corrinoid specificity of the outer membrane receptor // *J. Biol. Chem.* 1978. Vol. 253, № 5. P. 1341–1346.
207. Butzin N.C. et al. *Thermotoga lettingae* Can Salvage Cobinamide To Synthesize Vitamin B12 // *Applied and Environmental Microbiology.* 2013. Vol. 79, № 22. P. 7006–7012.
208. Santos J.A. et al. Functional and structural characterization of an ECF-type ABC transporter for vitamin B12 // *eLife.* 2018. Vol. 7.
209. Wexler A.G. et al. Human gut *Bacteroides* capture vitamin B12 via cell surface-exposed lipoproteins // *eLife.* 2018. Vol. 7.
210. Mee M.T. et al. Syntrophic exchange in synthetic microbial communities // *Proceedings of the National Academy of Sciences.* 2014. Vol. 111, № 20. P. E2149–E2156.
211. Shi Y. Common Folds and Transport Mechanisms of Secondary Active Transporters // *Annual Review of Biophysics.* 2013. Vol. 42, № 1. P. 51–72.
212. Saier M.H. et al. The Transporter Classification Database // *Nucleic Acids Research.* 2013. Vol. 42, № D1. P. D251–D258.
213. Hemberger S. et al. RibM from *Streptomyces davawensis* is a riboflavin/roseoflavin transporter and may be useful for the optimization of riboflavin production strains // *BMC Biotechnology.* 2011. Vol. 11, № 1. P. 119.
214. McAnulty M.J., Wood T.K. YeeO from *Escherichia coli* exports flavins // *Bioengineered.* 2014. Vol. 5, № 6. P. 386–392.

215. Togo A.H. et al. Description of *Mediterraneibacter massiliensis*, gen. nov., sp. nov., a new genus isolated from the gut microbiota of an obese patient and reclassification of *Ruminococcus faecis*, *Ruminococcus lactaris*, *Ruminococcus torques*, *Ruminococcus gnavus* and *Clostridium glycyrrhizinilyticum* as *Mediterraneibacter faecis* comb. nov., *Mediterraneibacter lactaris* comb. nov., *Mediterraneibacter torques* comb. nov., *Mediterraneibacter gnavus* comb. nov. and *Mediterraneibacter glycyrrhizinilyticus* comb. nov. // *Antonie van Leeuwenhoek*. 2018. Vol. 111, № 11. P. 2107–2128.
216. Ley R.E. et al. Microbial ecology: human gut microbes associated with obesity // *Nature*. 2006. Vol. 444, № 7122. P. 1022–1023.
217. Turnbaugh P.J. et al. A core gut microbiome in obese and lean twins // *Nature*. 2009. Vol. 457, № 7228. P. 480–484.
218. Frank D.N. et al. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases // *Proc. Natl. Acad. Sci. U.S.A.* 2007. Vol. 104, № 34. P. 13780–13785.
219. Costello E.K. et al. Bacterial community variation in human body habitats across space and time // *Science*. 2009. Vol. 326, № 5960. P. 1694–1697.
220. Turnbaugh P.J., Gordon J.I. The core gut microbiome, energy balance and obesity // *J. Physiol. (Lond.)*. 2009. Vol. 587, № Pt 17. P. 4153–4158.
221. Arumugam M. et al. Enterotypes of the human gut microbiome // *Nature*. 2011. Vol. 473, № 7346. P. 174–180.
222. Palmer C. et al. Development of the human infant intestinal microbiota // *PLoS Biol.* 2007. Vol. 5, № 7. P. e177.
223. Walujkar S.A. et al. Characterization of bacterial community shift in human Ulcerative Colitis patients revealed by Illumina based 16S rRNA gene amplicon sequencing // *Gut Pathog.* 2014. Vol. 6. P. 22.
224. Blaxter M. et al. Defining operational taxonomic units using DNA barcode data // *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 2005. Vol. 360, № 1462. P. 1935–1943.
225. Stearns J.C. et al. Bacterial biogeography of the human digestive tract // *Scientific Reports*. 2011. Vol. 1, № 1.
226. Wu M. et al. Genetic determinants of in vivo fitness and diet responsiveness in multiple human gut *Bacteroides* // *Science*. 2015. Vol. 350, № 6256. P. aac5992–aac5992.
227. Costliow Z.A., Degnan P.H. Thiamine Acquisition Strategies Impact Metabolism and Competition in the Gut Microbe *Bacteroides thetaiotaomicron* // *mSystems* / ed. Gilbert J.A. 2017. Vol. 2, № 5.
228. Gerdes S.Y. et al. Experimental Determination and System Level Analysis of Essential Genes in *Escherichia coli* MG1655 // *Journal of Bacteriology*. 2003. Vol. 185, № 19. P. 5673–5684.
229. Said H.M. Intestinal absorption of water-soluble vitamins in health and disease // *Biochemical Journal*. 2011. Vol. 437, № 3. P. 357–372.
230. Blanton L.V. et al. Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children // *Science*. 2016. Vol. 351, № 6275. P. aad3311–aad3311.

Приложения

Приложение 1. Фенотипические правила для путей биосинтеза и сохранения витаминов и кофакторов

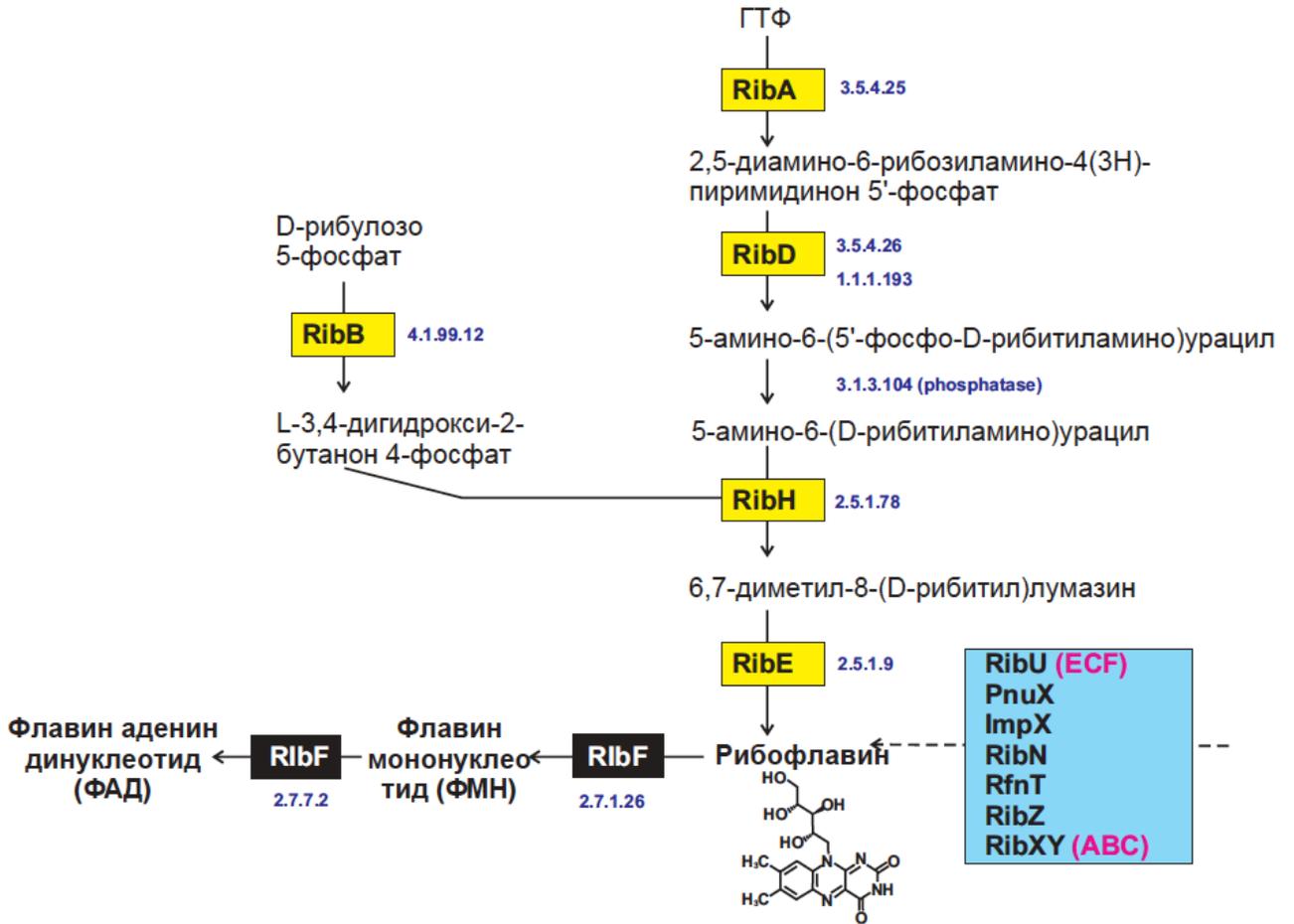
Витамин [Предшественники] 1	Геномный паттерн ³	Вариант пути ⁴	Бинарный фенотип ⁵	Число геномов ⁶	Метаболические потребности ⁷	Транспортеры витаминов и предшественников
B1: Тиамин [HET, HMP]	(ThiF), (ThiS), ThiH/ThiO, ThiG, ThiC, ThiD, ThiE, ThiL/ThiN	P1	1	975	--	<u>B1</u> : ThiT/ThiBPQ/ PnuT/ThiV/ ThiXYZ2/YkoE
	(ThiF), (ThiS), ThiG, ThiC, ThiD, ThiE, ThiL/ThiN	P1*	1	89	(Источник Ig?)	
	Thi4, ThiC, ThiD, ThiE, ThiL/ThiN	P2	1	45	--	
	(ThiF), (ThiS), ThiH/ThiO, ThiG, ThiD, ThiE, ThiL/ThiN	Ah	0	197	B1; HMP	DC2
	ThiC, ThiD, ThiE, ThiM, ThiL/ThiN	Az	0	116	B1; HET	<u>HMP</u> : CytX/
	ThiD, ThiE, ThiM, ThiL/ThiN	Ahz	0	451	B1; (HMP+HET)	ThiXYZ/YkoE
	ThiL/ThiN	A	0	355	B1	DC <u>HET</u> : ThiW/ThiU
B2: Рибофлавин	(RibA), RibB, RibD, RibH, RibE, RibF	P	1	1644	--	<u>B2</u> : RibU/PnuX/ ImpX/RibN/RfnT/ RibZ/RibXY/RfnT
	RibF	A	0	584	B2	

В3: Ниацин (Никотинат или никотинамид) [Qn, Nr]	NadA, NadB/NadB2, NadC, NadD/NadM, NadE	P1	1	1170	--	В3: NiaP/NiaX/NiaY Nr: PnuC
	NadA, NadC, NadD/NadM, NadE	P1*	1	34	(Отсутствует NadB?)	
	Tdo, (Kfa), Kmo, Kyn, (Had), NadC, NadD/NadM, NadE	P2	1	12	--	
	PncB/NadV, (PncA), NadD/NadM, NadE	A	0	895	B3	
	PncB, (PncA), NadC, NadD/NadM, NadE NadR	Aq Ar	0 0	86 31	B3; Qn Nr	
В5: Пантотенат [Pne, Pnt]	PanD/PanP, PanB, (PanE/PanG), PanC, CoaA/CoaX/CoaW, CoaB, CoaC, CoaD, CoaE	P	1	1168	--	В5: PanT/PanF Pnt: PanS
	PanB, (PanE/PanG), PanC, CoaA/CoaX/CoaW, CoaB, CoaC, CoaD, CoaE	P*	1	95	(источник b-ala?)	
	PanD/PanP, PanC, CoaA/CoaX/CoaW, CoaB, CoaC, CoaD, CoaE	Apt	0	40	B5; Pnt	
	PanC, CoaA/CoaX/CoaW, CoaB, CoaC, CoaD, CoaE	Apt*	0	16	B5; Pnt (источник b-ala?)	
	CoaA/CoaX/CoaW, CoaB, CoaC, CoaD, CoaE	A	0	793	B5	
	CoaA/CoaX/CoaW, CoaD, CoaE -	Apn A*	0 0	91 25	Пантетеин Созранение КоА	
В6: Пиридоксин	PdxS, (PdxT)	P1	1	862	--	В6: PdxU/PdxU2
	PdxJ, (PdxA), (PdxH/PdxO)	P2	1	711	--	
	PdxK/PdxK2	A	0	541	B6	
	-	A*	0	114	B6 (Отсутствует PdxK?)	
В7: Биотин [Dtb, КАРА, DAPA]	BioF, BioA, BioB, BioD, BioC, (BioG/BioH/BioZ/BioV), BirA	P1	1	797	--	В7: BioY/YigM
	BioF, BioA, BioB, BioD, BioW, BirA	P2	1	246	--	
	BioF, BioA, BioB, BioD, BirA	P*	1	74	(источник пимелоила?)	
	BioA, BioB, BioD, BirA	A3	0	68	B7; Dtb; DAPA; КАРА	
	BioD, BioB, BirA	A2	0	38	B7; Dtb; DAPA	
	BioB, BirA	A1	0	169	B7; Dtb	
	BirA	A	0	836	B7	

В9: Фолат	FolE1/FolE2, (FolQ/FolQ1/FolQ2), (FolB/FolB2), FolK, FolP, FolC, PabC, (PabAB), FolA/FolA2/FolM	P	1	1471	--	<u>B9:</u> FolT
	FolE1/FolE2, (FolQ/FolQ1/FolQ2), (FolB/FolB2), FolK, FolP, FolC, FolA/FolA2/FolM	P*	1	415	(Источник ПАБ?)	
	FolA/FolA2/FolM	A	0	342	B9	
В12: Кобалам ин [Cbi, Cbr, Ba]	CbiK/CbiX/CbiX2, CbiL, CbiH, CbiF, CbiG, CbiD, CbiJ, CbiT, CbiE, CbiC, CbiA, Co transporter, CbiP, CbiB, CobU, CobS, CobC/CblZ, CobT, CobD, BtuR/PduO	P1	1	628	--	<u>B12:</u> CbrUVT/ BtuBCDF
	ChlID, CobN, CbiL, CobG, CbiH, CbiF, CobF, CbiJ, CbiT, CbiE, CbiC, CbiA, Co transporter, CbiP, CbiB, CobU, CobS, CobC/CblZ, CobT, CobD, BtuR/PduO	P2	1	97	--	
	CbiA, CbiP, CbiB, CobU, CobS, CobC/CblZ, CobT, CobD, BtuR/PduO	Aba	0	32	B12; Cbi; Cbr; Ba	
	CbiP, CbiB, CobU, CobS, CobC/CblZ, CobT, CobD, BtuR/PduO	Acbr	0	43	B12; Cbi; Cbr	
	CobU, CobS, CobC/CblZ, CobT, CobD, BtuR/PduO BtuR/PduO	Acbi	0	193	B12; Cbi	
	A	0	1235	B12		

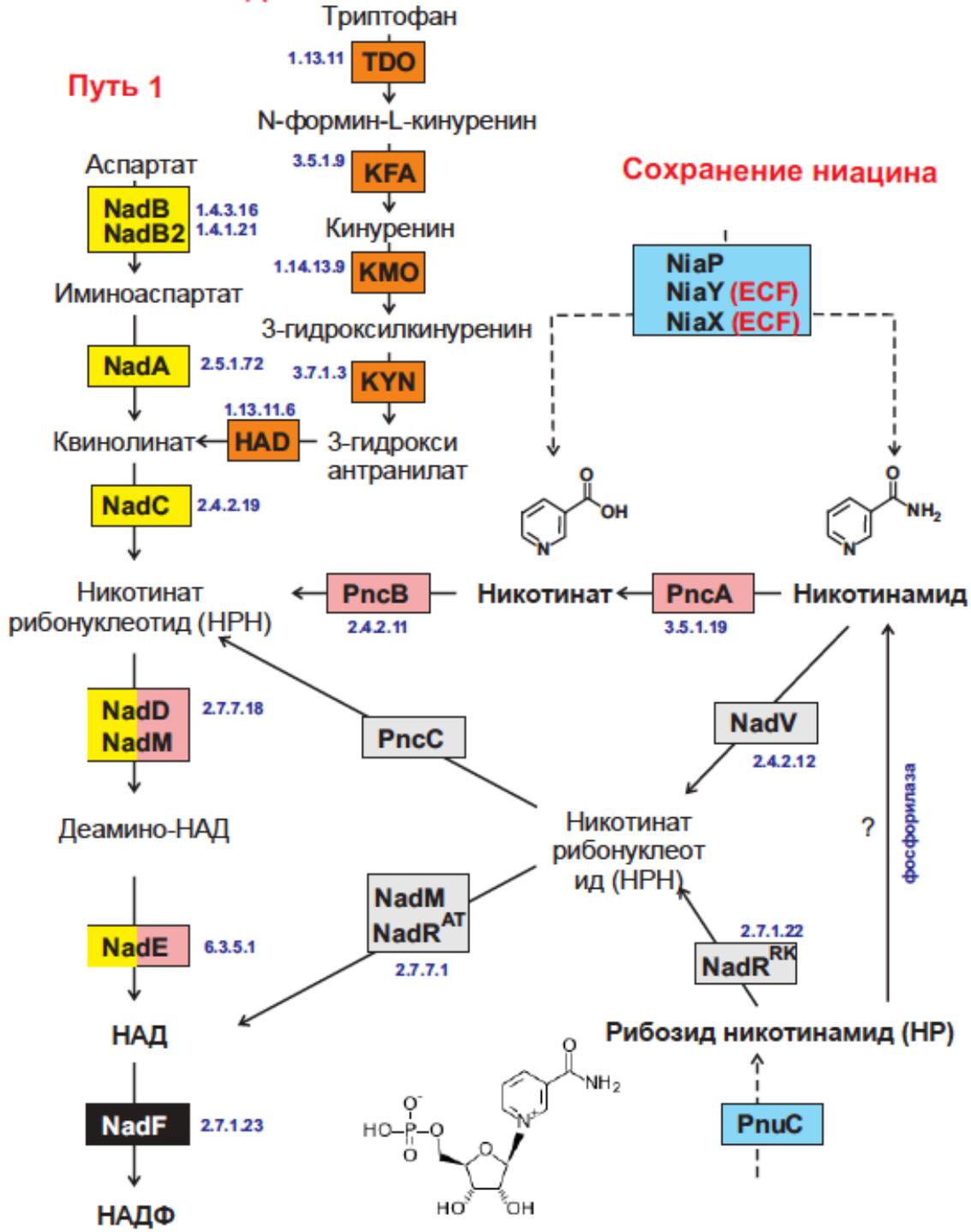
Биосинтез и сохранение рибофлавина (B2)

Самостоятельный биосинтез B2



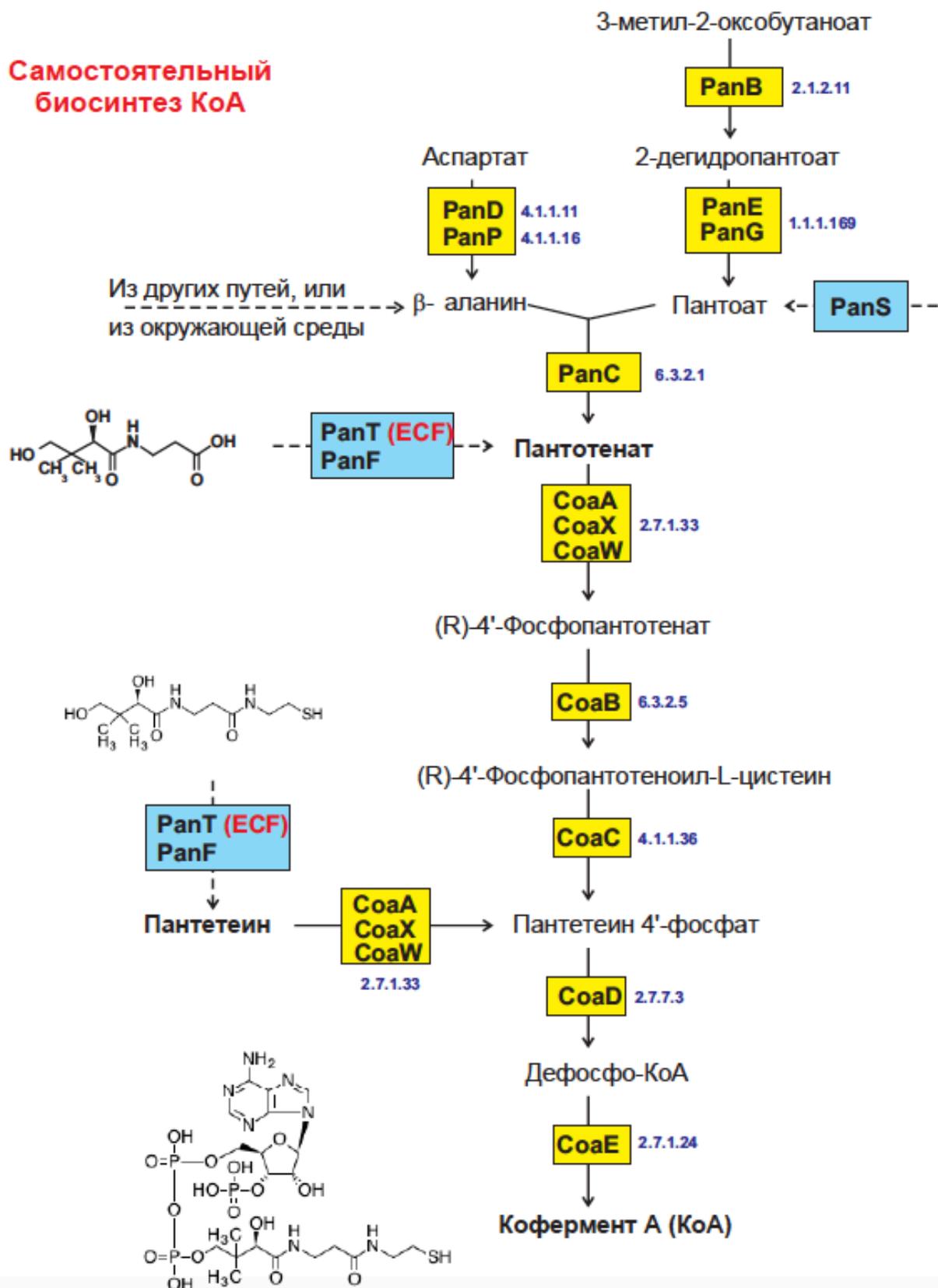
Биосинтез НАД и сохранение ниацина (В3)

Самостоятельный Путь 2 биосинтез НАД



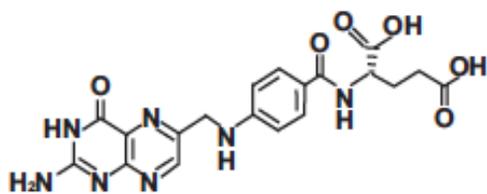
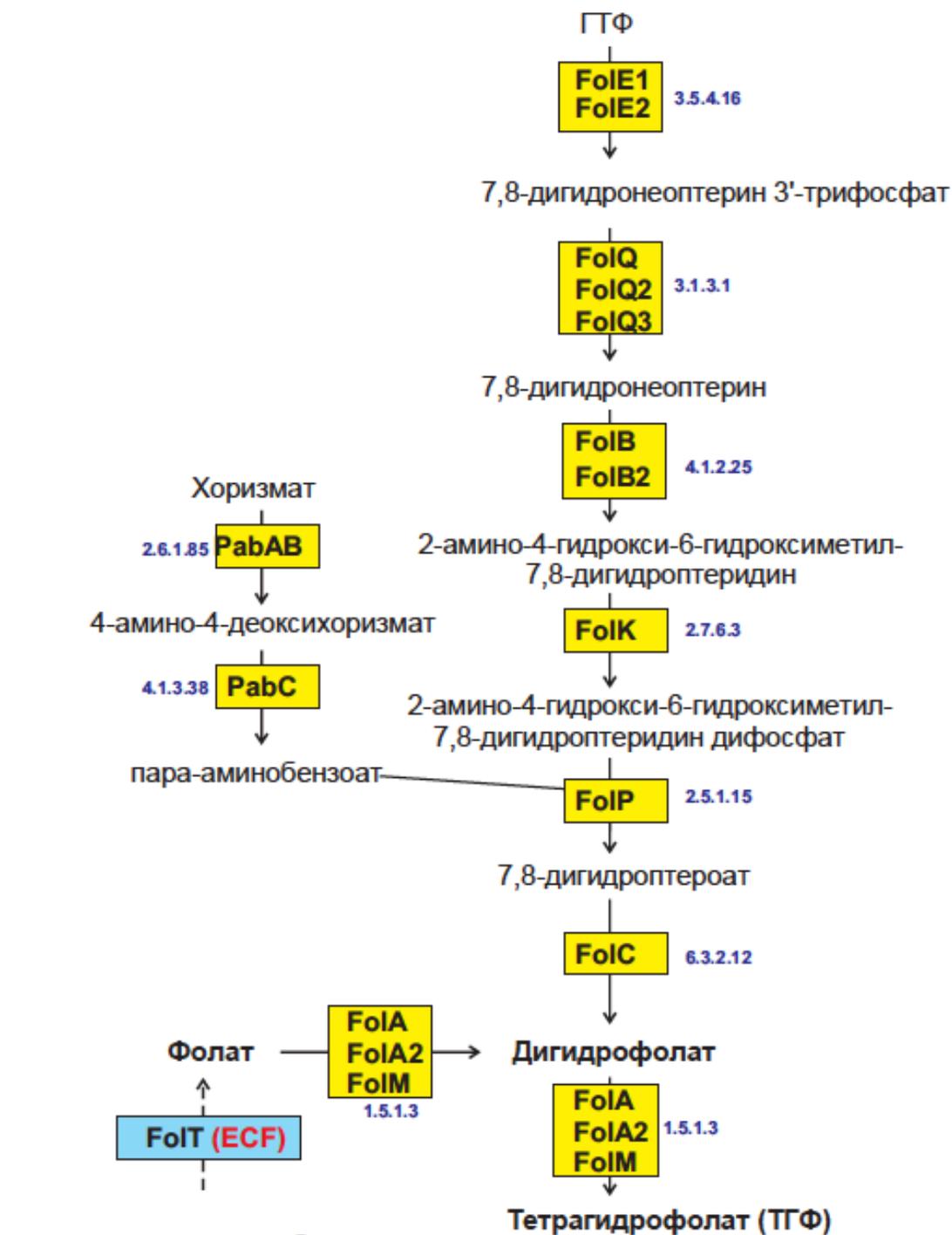
Биосинтез и сохранение пантотената (B5)

Самостоятельный биосинтез КоА



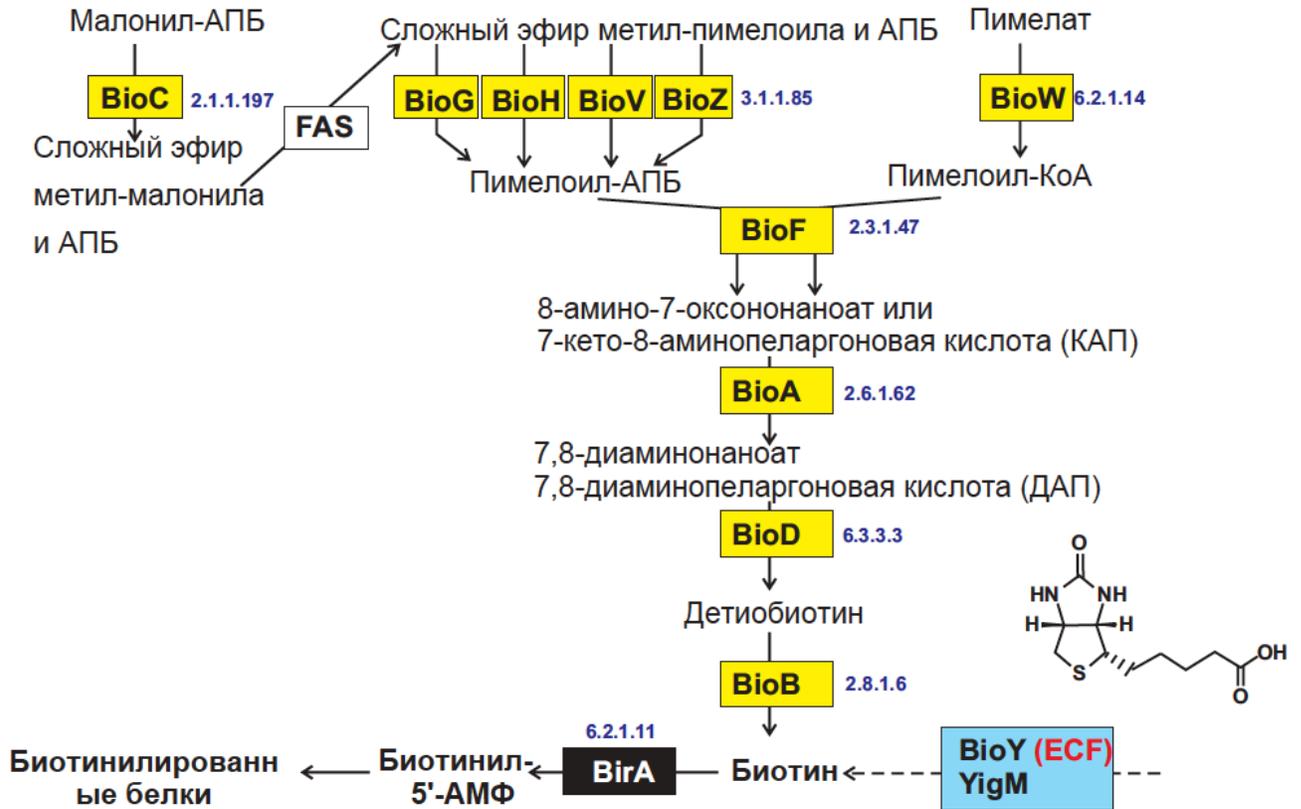
Биосинтез и сохранение фолата (B9)

Самостоятельный биосинтез фолата

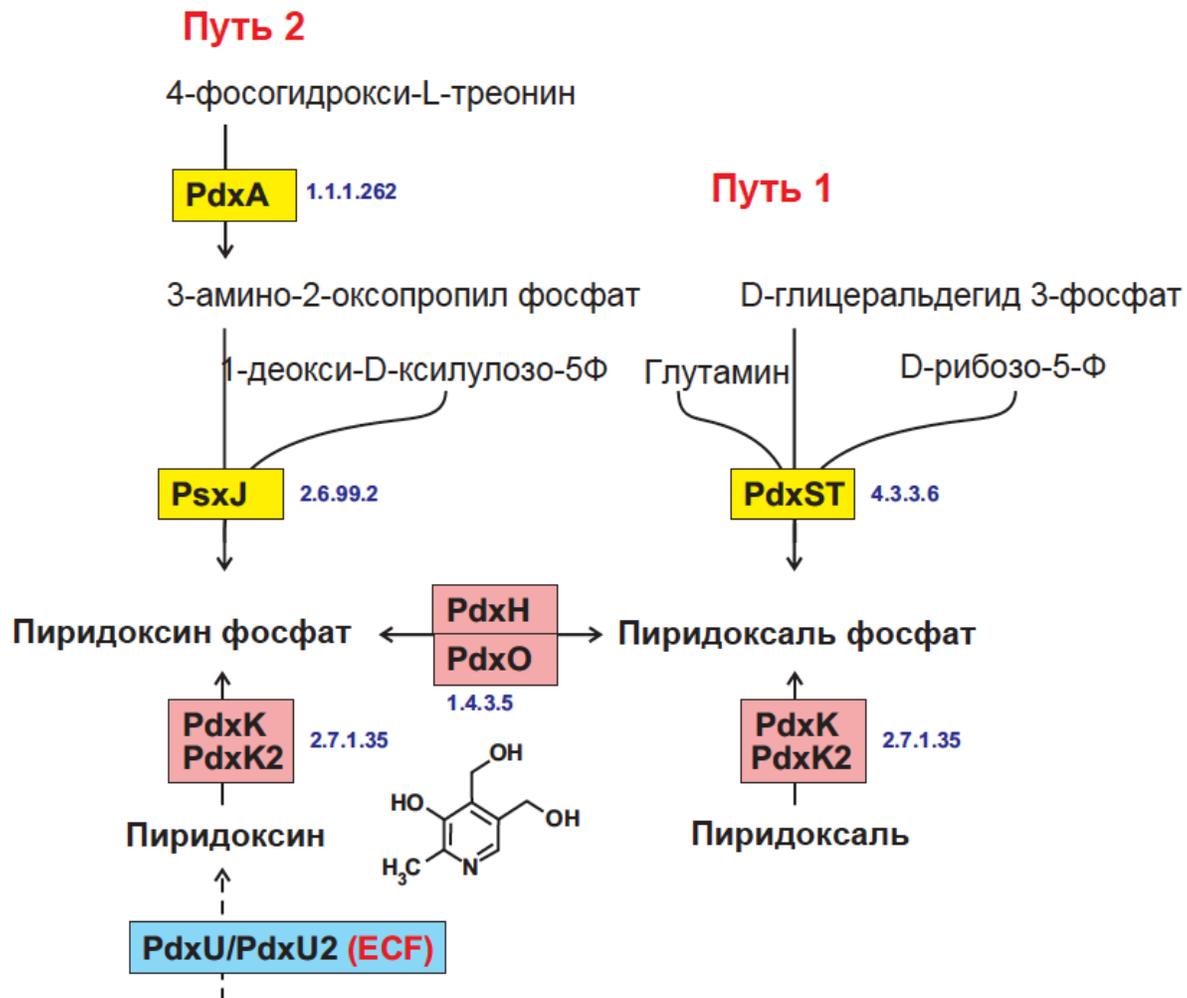


Биосинтез и сохранение биотина (B7)

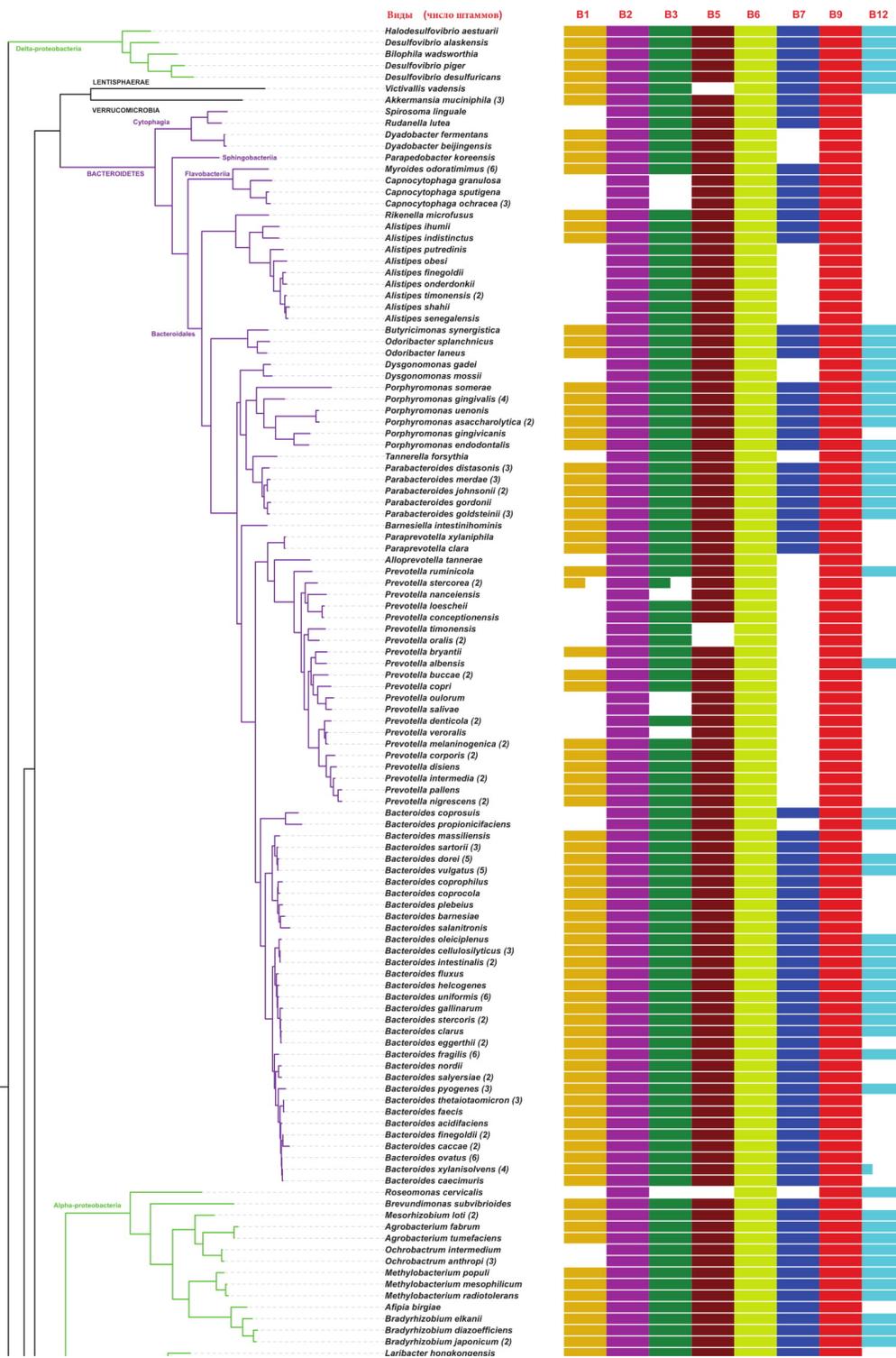
Самостоятельный биосинтез B7

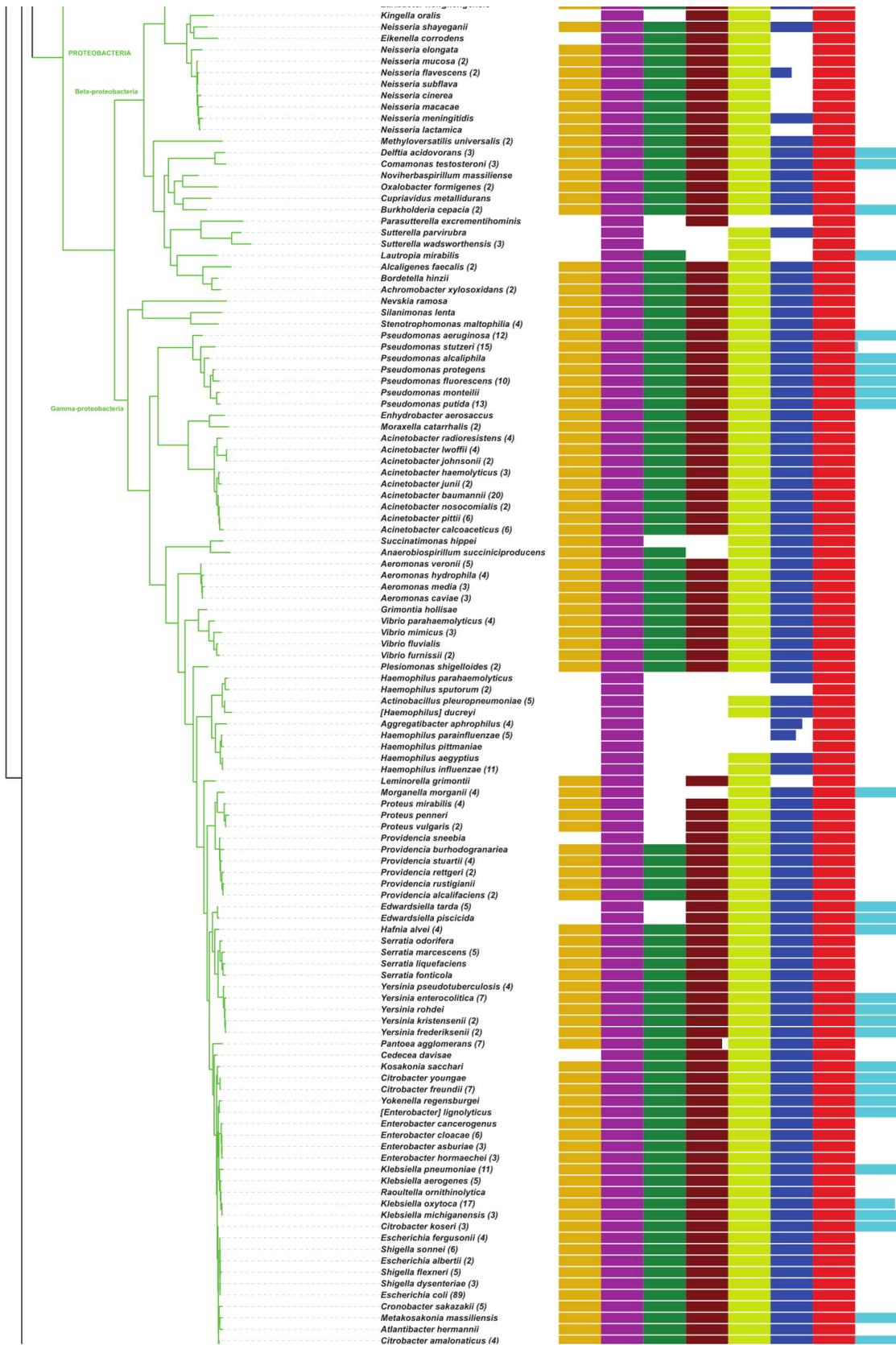


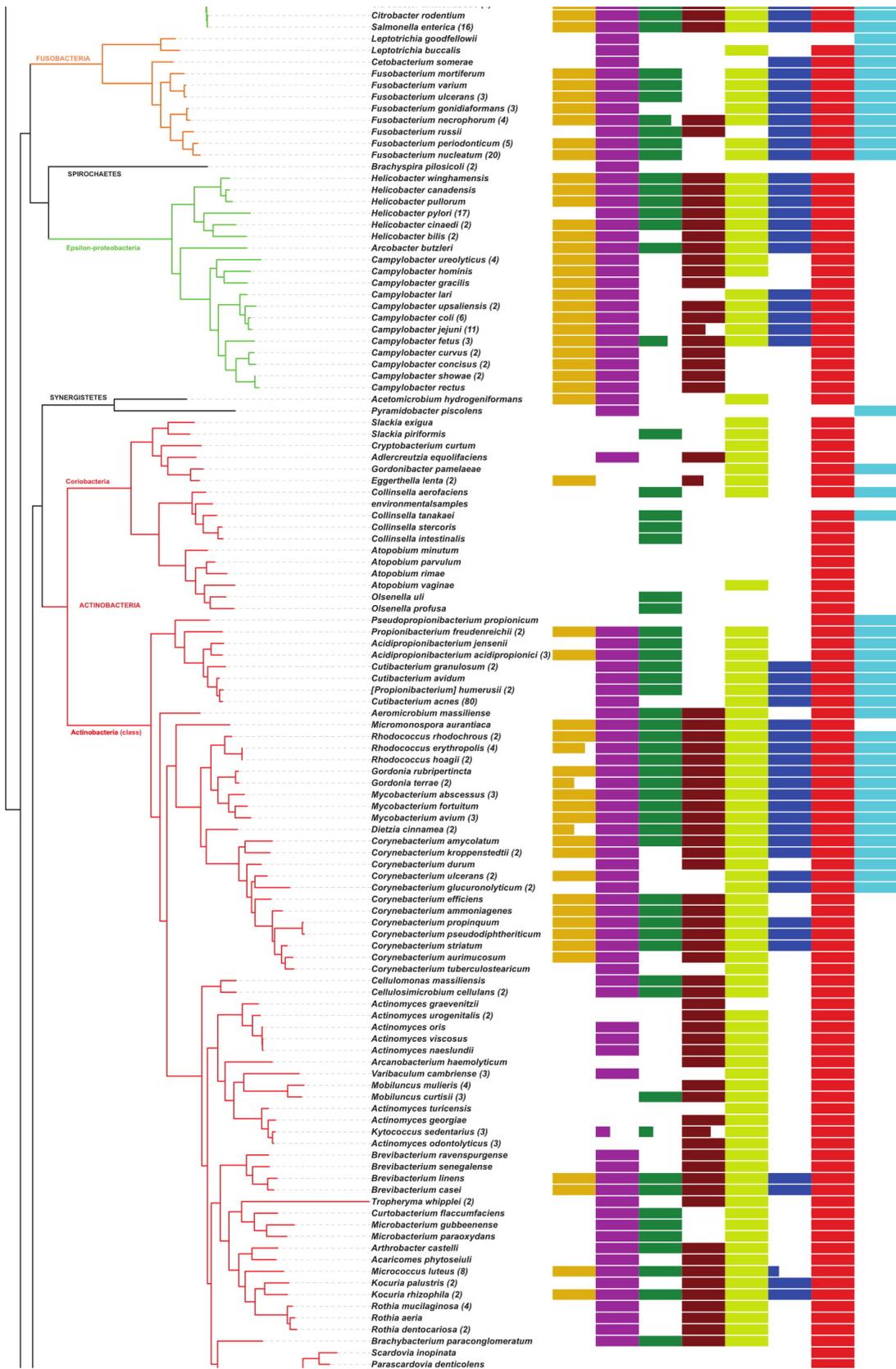
Биосинтез и сохранение пиридоксина (В6)

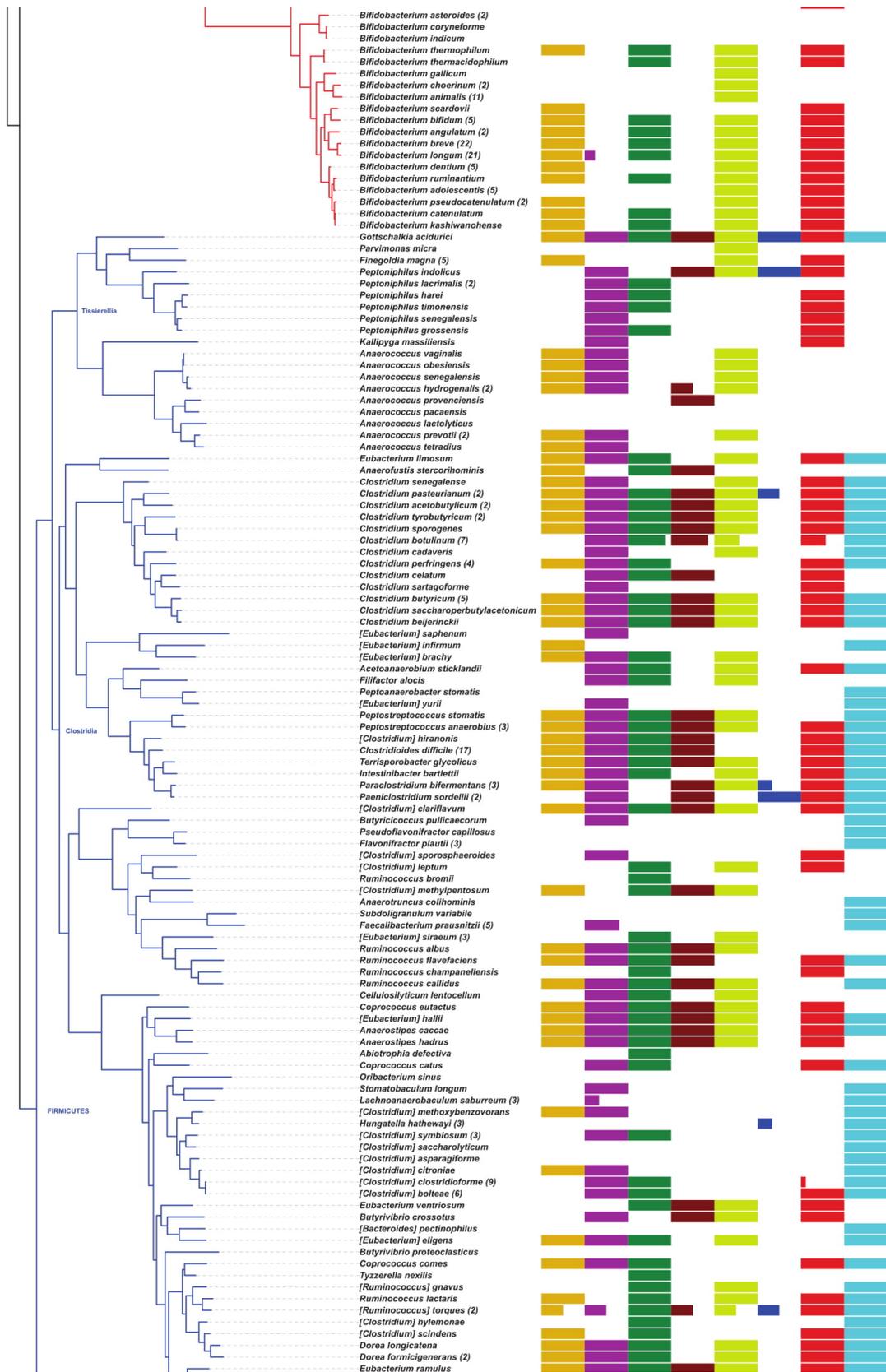


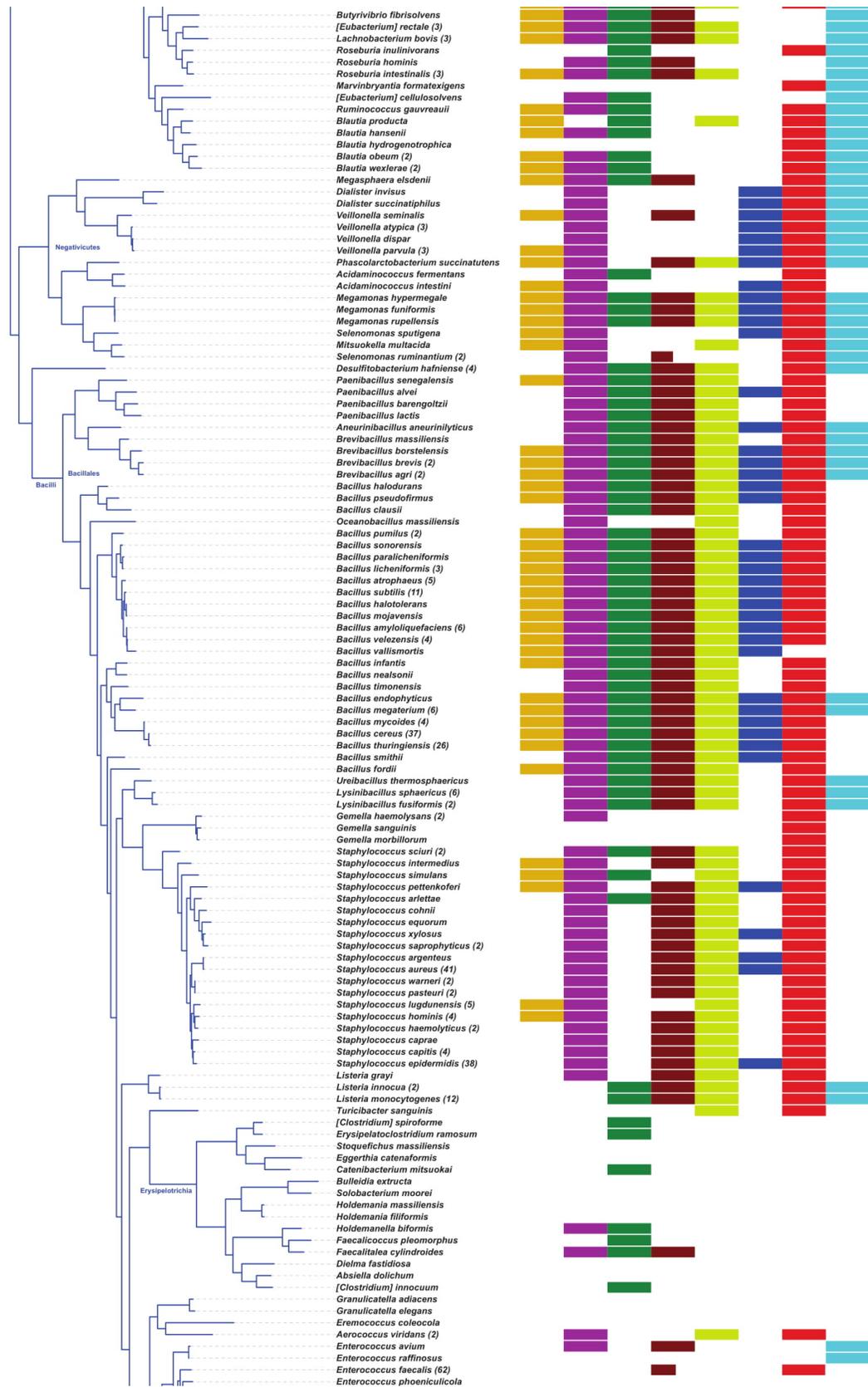
Приложение 3. Распределение фенотипов по витаминам группы В среди видов бактерий – обитателей кишечного микробиома











Tree scale: 0.1

