

CRISPR-CAS СИСТЕМЫ *MYCOBACTERIUM TUBERCULOSIS*: СТРУКТУРА МОДУЛЯ, ИЗМЕНЕНИЕ В ПРОЦЕССЕ ЭВОЛЮЦИИ У РАЗЛИЧНЫХ ЛИНИЙ, ВОЗМОЖНАЯ РОЛЬ В ФОРМИРОВАНИИ ВИРУЛЕНТНОСТИ И ЛЕКАРСТВЕННОЙ УСТОЙЧИВОСТИ

М. В. Зайчикова¹✉, Н. В. Захаревич¹, М. С. Чекалина¹, В. Н. Даниленко^{1,2}

¹ Лаборатория генетики микроорганизмов, Институт общей генетики имени Н. И. Вавилова Российской академии наук, Москва

² Кафедра биоинформатики, факультет биологической и медицинской физики, Московский физико-технический институт (государственный университет), Долгопрудный

CRISPR-Cas системы широко распространены у бактерий и архей. Они обеспечивают адаптивный иммунитет к бактериофагам и плазмидам, а также выполняют другие функции, включая регуляцию экспрессии генов, репарацию ДНК, формирование вирулентности. Нами был проведен анализ CRISPR-Cas систем полностью секвенированных геномов *M. tuberculosis* из семи линий: Beijing, B0/W-148, EAI, Haarlem, Ural, LAM, S. Проанализированные геномы содержат CRISPR-Cas систему типа III-A. Линиям в составе вида *M. tuberculosis* свойственны различия в строении CRISPR-Cas системы, в том числе редукция части системы у линии Beijing. Для *cas*-генов нами был осуществлен поиск возможных функциональных партнеров и компенсаторных механизмов с использованием метода филогенетического профайлинга. В ходе анализа филогенетических профилей (ФП) были обнаружены гены со сходным характером эволюционных событий. Установлено, что потеря части системы CRISPR-Cas у представителей линии Beijing сопровождалась по крайней мере двумя эволюционными событиями потери и одним событием приобретения участков генома. Возможность изучения альтернативных функций CRISPR-Cas систем у *M. tuberculosis* и их предполагаемая связь с другими генными системами представляет значительный интерес.

Ключевые слова: системы CRISPR-Cas, вирулентность, *M. tuberculosis*, линия Beijing, филогенетический профайлинг

✉ **Для корреспонденции:** Зайчикова Марина Викторовна
ул. Губкина, д. 3, г. Москва, 119991; marinaz15@yandex.ru, valerid@vigg.ru

Статья получена: 15.03.2018 **Статья принята к печати:** 20.03.2018

DOI: 10.24075/vrgmu.2018.016

CRISPR-CAS SYSTEMS OF *MYCOBACTERIUM TUBERCULOSIS*: THE STRUCTURE, TRANSFORMATION IN DIFFERENT LINEAGES IN THE PROCESS OF EVOLUTION AND A POSSIBLE ROLE IN THE FORMATION OF VIRULENCE AND DRUG RESISTANCE

Zaychikova MV¹✉, Zakharevich NV¹, Chekalina MS¹, Danilenko VN^{1,2}

¹ Laboratory of Bacterial Genetics, Vavilov Institute of General Genetics, RAS, Moscow

² Department of Bioinformatics, Faculty of Biological and Medical Physics, Moscow Institute of Physics and Technology (State University), Dolgoprudny

CRISPR-Cas systems are widespread in bacteria and archaea. They provide adaptive immunity against bacterial phages and plasmids and exert a few important functions like regulation of gene expression, DNA repair or virulence formation. We have analyzed the CRISPR-Cas systems of 7 *M. tuberculosis* lineages with fully sequenced genomes, namely Beijing, B0/W-148, EAI, Haarlem, Ural, LAM, and S. The CRISPR-Cas systems present in the analyzed genomes belong to type III-A. *M. tuberculosis* lineages differ in their CRISPR-Cas structure; in the Beijing lineage a part of the system is reduced. We have conducted a search for the functionally related partners and compensatory mechanisms of *cas*-genes using a method of phylogenetic profiling. The obtained phylogenetic profiles show that some genes have undergone similar evolutionary events. The reduction of the system's part in the Beijing lineage was accompanied by at least two evolutionary losses and one acquisition of genome regions. Exploration of alternative CRISPR-Cas functions in *M. tuberculosis* and their possible associations with other gene systems remains an exciting challenge.

Keywords: CRISPR-Cas systems, virulence, *M. tuberculosis*, Beijing lineage, phylogenetic profiling

✉ **Correspondence should be addressed:** Marina Zaychikova
Gubkina 3, Moscow, 119991; marinaz15@yandex.ru, valerid@vigg.ru

Received: 15.03.2018 **Accepted:** 20.03.2018

DOI: 10.24075/brsmu.2018.016

CRISPR-CAS СИСТЕМЫ У БАКТЕРИЙ: СТРОЕНИЕ И КЛАССИФИКАЦИЯ

В настоящее время системы CRISPR-Cas идентифицированы в геномах примерно 40% бактерий и большинства архей (90%) [1, 2]. Эти системы состоят из двух обязательных компонентов: CRISPR¹-кассет и Cas²-белков. Повторяющиеся последовательности равной длины, перемежающиеся уникальными участками (спейсерами), были описаны еще в 1987 г. в геноме *E. coli* [3], но об их роли тогда ничего не было известно. В 2000-х годах было описано участие CRISPR-Cas систем в реализации бактериального иммунитета [4, 5], а на настоящий момент уже показано их участие в целом ряде различных клеточных процессов, в том числе в процессах репарации ДНК, в регуляции экспрессии генов, в процессах вирулентности и т. д. [6]. Стоит отметить, что в геноме *M. tuberculosis* прямые повторы (direct repeats, DR) были обнаружены ещё в 1990-х годах, их полиморфизм подробно изучен, и еще до описания иммунной функции CRISPR-Cas систем DR применялись и применяются в настоящее время при генотипировании (сполиготикировании) микобактерий [7, 8].

CRISPR-Cas системы очень разнообразны, при этом каждая функциональная кассета обязательно содержит три следующих элемента: повторы, спейсеры и лидерную последовательность. Рядом с кассетой располагается локус из *cas*-генов, кодирующих белки с различными функциональными доменами, взаимодействующими с нуклеиновыми кислотами [9]. Разные наборы *cas*-генов определяют особенности молекулярного механизма CRISPR-Cas систем, но все же в них можно выделить общие черты. Так, два белка, Cas1 и Cas2, обнаружены в большинстве известных на сегодняшний день активных CRISPR-Cas системах. Эти белки образуют комплекс, который участвует в процессе интеграции новых спейсеров в кассету. Новые спейсеры встраиваются в кассету рядом с лидерной последовательностью. В течение жизни кассеты часть спейсеров может быть утеряна в результате рекомбинации между повторами кассеты [10]. Часть кассеты или даже вся кассета целиком может быть перемещена посредством горизонтального переноса генов (ГПГ) [11].

CRISPR-Cas системы классифицируются по составу *cas*-локусов; так, согласно современной классификации, они делятся на 2 больших класса, 5 типов и множество подтипов [12]. К 1 классу (типы I, III и IV) относят системы CRISPR-Cas, обладающие мультисубъединичными эффекторными комплексами, в то время как в системах класса 2 (типы II и V) все функции эффекторного комплекса выполняются одним белком, таким как Cas9 [12]. Учитывая исключительную важность CRISPR-Cas систем типа II в биотехнологии, в том числе для геномного редактирования, стоит отметить, что они довольно редко встречаются и были обнаружены исключительно в геномах бактерий [12]. Большинство CRISPR-Cas систем можно однозначно отнести к одному из пяти основных типов. Тем не менее, существует ряд организмов, *cas*-локусы которых не вписываются в текущую классификацию.

Описывая CRISPR-Cas системы, стоит сделать акцент на генах *cas1* и *cas2*, задействованных в процессе интеграции новых спейсеров. Несмотря на различные доказательства участия обоих генов в процессе интеграции, все ферментативные активности, необходимые для вставки нового спейсера, присущи Cas1, тогда как каталитическая активность Cas2 не

требуется ни для образования комплекса Cas1–Cas2, ни для вставки нового спейсера. На сегодняшний день известно, что белок Cas2 представляет собой интерфазу мПНК, которая специфически расщепляет связанные с рибосомой мПНК — такая активность кажется, на первый взгляд, неуместной при интеграции новых спейсеров. Однако в ряде работ упоминается о возможном происхождении Cas2 от древних мобильных элементов, в частности от систем токсин–антитоксин (TA) [13, 14]. В связи с этим, можно предположить, что Cas2 сохраняет предковую токсиноподобную функцию эндонуклеазы в системах CRISPR–Cas, но такая ее активность обратимо контролируется посредством ингибирования при взаимодействии с Cas1 и образовании комплекса Cas1–Cas2. Согласно этой гипотезе, когда системы CRISPR-Cas не могут сдерживать рост вируса, может активироваться Cas2 (возможно, путем деградации Cas1) и останавливать транскрипцию, что, вероятно, приводит клетку к самоубийству либо переводит в состояние покоя. Участие же в интеграции спейсеров для Cas2 может быть связано с регулированием или стабилизацией Cas1 посредством образования комплекса Cas1–Cas2, что одновременно обратимо инактивирует Cas2 [15]. Потенциальное участие Cas2 в процессах перевода бактериальной клетки в персистирующее состояние представляет собой перспективное направление исследований для различных патогенов, в том числе для *M. tuberculosis*.

ФУНКЦИИ СИСТЕМ CRISPR-CAS У БАКТЕРИЙ

Учитывая широкую распространенность и разнообразие систем CRISPR-Cas, нетрудно понять, почему с каждым днем публикуется все больше доказательств участия этих систем в различных клеточных процессах [6]. Помимо участия CRISPR-Cas систем в адаптивном иммунитете, наиболее известной и описанной функцией данных систем является регуляция экспрессии генов. Так, у присутствующей в почве бактерии *Mycosoccus xanthus* жизненный цикл включает в себя стадии формирования плодового тела и споруляции. Процессы формирования плодового тела и последующая дифференциация его клеток в микроспоры жестко регулируются различными межклеточными сигналами и внутриклеточными сигнальными каскадами, в которых CRISPR-Cas системы типа I-C *M. xanthus* входят в состав петли положительной обратной связи и принимают участие в споруляции бактерии [16].

На сегодняшний день существуют также доказательства того, что системы CRISPR-Cas могут принимать участие в процессах репарации ДНК. Обнаружено, что очищенный белок Cas1 (YgbT) из *Escherichia coli* способен физически и генетически взаимодействовать с ключевыми компонентами систем репарации ДНК, включая такие гены, как *recB*, *recC* и *ruvB* [17]. Авторы статьи показали, что штамм, у которого произошла делеция *ugbT*, демонстрирует повышенную чувствительность к повреждению ДНК. Аналогичные фенотипы авторы наблюдали и у штаммов с удаленным CRISPR-кластером; это свидетельствует по меньшей мере о том, что некоторые компоненты CRISPR-Cas систем вовлечены в процессы репарации ДНК.

¹ От англ. clustered regularly interspaced short palindromic repeats — короткие полидромные повторы, регулярно расположенные группами.

² От англ. CRISPR-associated proteins.

Была описана еще одна альтернативная функция систем CRISPR-Cas — участие в формировании биопленок [18]. Авторы исследовали CRISPR-Cas систему 1-F оппортунистического патогена *Pseudomonas aeruginosa* и показали, что данная система ингибирует образование биопленки. Описанная авторами CRISPR-зависимая способность образования биопленки зависела от взаимодействия конкретного спейсера с его прототипом — протоспейсером, расположенным в геноме бактериофага. Результат данного взаимодействия в итоге приводил к индукции генов, связанных с фагом, индуцирующих в свою очередь гибель поверхностных клеток. Эти данные свидетельствуют о наличии еще одного механизма, присущего системам CRISPR-Cas и не связанного с адаптивным иммунитетом.

У бактерий контроль экспрессии генов обычно происходит на посттранскрипционном уровне различными малыми некодирующими РНК. И хотя описываемые РНК контролируют многие физиологические процессы в клетках, мало кто из них участвует в распознавании интрузивных нуклеиновых кислот — такую роль берут на себя системы CRISPR-Cas. В отличие от различных систем эукариот, бактериальные системы CRISPR-Cas расщепляют ДНК, а это означает, что если они будут участвовать в регуляции эндогенных генов, то неизбежно разрушат бактериальную хромосому. Однако в 2013 г. в *Nature* была опубликована статья, в которой авторы сообщили о механизме посттранскрипционного контроля у *Francisella novicida*, при котором ген вирулентности регулируется белком Cas (а именно Cas9) и CRISPR-ассоциированными малыми РНК [19]. Предполагается, что в этом случае Cas9 действует на эндогенную мРНК, а не на ДНК. В настоящий момент взаимосвязь между CRISPR-Cas системами и способностью бактериальных штаммов проявлять повышенную вирулентность или даже резистентность к лекарственным средствам представлена в ряде исследований [20].

Говоря о «новых» функциях CRISPR-Cas систем, некоторые авторы считают, что такие функции, как формирование биопленок у *Pseudomonas aeruginosa*, являются «побочными продуктами» классической иммунной функции CRISPR-Cas, тогда как другие, например вирулентность у *Francisella novicida* и регуляция развития у *Mycococcus xanthus*, являются отдельными, самостоятельными функциями [6]. История последовательного открытия различных функций, присущих системам CRISPR-Cas, начиная с иммунной, очень сходна с историей исследования РНК-интерференции эукариот, для которой тоже первоначально было описано

участие в иммунитете, и только потом обнаружено, что РНК-интерференция влияет на многие клеточные процессы, в том числе на регуляцию генов и образование гетерохроматина [21]. Ряд авторов проводят параллель между CRISPR-Cas системами и РНК-интерференцией [22, 23].

CRISPR-CAS СИСТЕМЫ У МИКОБАКТЕРИЙ: ОБЩИЙ ПЛАН СТРОЕНИЯ И ОСОБЕННОСТИ CAS-ОПЕРОНА НА ПРИМЕРЕ *M. TUBERCULOSIS* H37RV

Род *Mycobacterium* представлен широким спектром микроорганизмов, включая патогенные для человека, из которых наиболее значимы микобактерии туберкулеза комплекса (МТК). Среди прочих этот комплекс включает в себя *Mycobacterium tuberculosis* — основного возбудителя туберкулеза. *M. tuberculosis* генетически гетерогенен и делится на несколько групп, так называемых линий, для каждой из которых характерен определенный набор мутаций, постепенно накапливавшихся в ходе эволюции [24–26]. Изоляты из разных линий различаются фенотипически, в том числе по тенденции к развитию лекарственной устойчивости (ЛУ), уровню вирулентности и патогенности, что влияет на тяжесть течения заболевания [27, 28]. К числу наиболее распространенных и клинически значимых в мире линий *M. tuberculosis* принадлежат Beijing, Haarlem, LAM, S. Линия Beijing (в особенности недавно сформировавшаяся в ее составе сублиния B0/W-148) является наиболее эпидемиологически значимой в связи с широким распространением и склонностью к формированию ЛУ-форм [29, 30]. Для линии Haarlem также характерен повышенный уровень вирулентности [28]. Помимо этого, определенный интерес вызывают представители линий EAI и Ural, обладающие, наоборот, сниженной вирулентностью, и в связи с этим менее распространенные [28]. EAI — это древняя линия, имеющая ограниченное распространение на сегодняшний день (преимущественно Юго-Восточная Азия) [31]. Линия Ural, родственная Haarlem, как и EAI не особенно широко распространена и, по-видимому, обладает сниженной трансмиссивностью [32] (рис. 1).

Если учитывать потенциальную роль CRISPR-Cas в вирулентности [19, 20], представляется интересным изучение этой системы у различных линий *M. tuberculosis*.

На сегодняшний день системы CRISPR-Cas обнаружены у 14 видов микобактерий [34]. Все обнаруженные системы CRISPR-Cas располагаются на хромосоме. CRISPR-кассеты с числом повторов более 5 идентифицированы только у трех микобактериальных видов: *M. tuberculosis* и

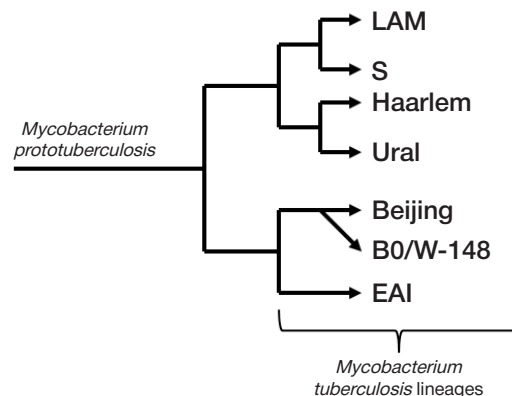


Рис. 1. Схематическое представление филогении рассматриваемых линий *M. tuberculosis* [33]

M. bovis, относящихся к МТК, и у патогена *M. avium*. При этом у *M. avium* рядом с CRISPR-кассетой отсутствуют *cas*-гены; а структуры CRISPR-локусов *M. tuberculosis* и *M. bovis* очень похожи — это отражает тесную эволюционную взаимосвязь между ними и согласуется с их филогенией [34, 35]. Для систем CRISPR-Cas *M. tuberculosis* характерно строение, типичное для систем типа III-A [34].

Были проанализированы CRISPR-Cas системы из 41 полногеномной последовательности штаммов из различных линий *M. tuberculosis*, доступных в базе данных RefSeq NCBI: 13 геномов линии Beijing; 3 генома линии B0/W-148; 2 генома линии EAI; 10 геномов линии Haarlem; 1 геном линии Ural; 2 генома линии S; 10 геномов линии LAM. Дополнительно были проанализированы 7 геномов сублинии B0/W-148, 4 генома линии URAL, 3 генома линии EAI и 3 генома линии S в статусе «draft», что было связано с недостаточным количеством «complete»-геномов для данных линий. Генотипирование осуществляли по маркерным полиморфизмам [36–38]. Для ряда геномов генотип изолята был заранее известен из литературных источников. Поиск и анализ CRISPR-Cas систем проводили с помощью двух алгоритмов: CRISPRFinder и CRISPR Recognition Tool [39, 40]. На рис. 2 представлено типичное строение системы CRISPR-Cas у *M. tuberculosis* на примере штамма H37Rv, являющегося общепринятым референсным геномом.

Большинство проанализированных штаммов *M. tuberculosis* содержат две длинные CRISPR-кассеты (см. рис. 2) [8]. Исключением является штамм *M. tuberculosis* 7199–99, относящийся к линии Haarlem, у которого произошла редукция кассеты CRISPR2 с 12-го спейсера и разделяющего кассеты участка, что привело к формированию единой кассеты, содержащей 33 спейсера. Максимальное количество спейсеров в одном геноме — 57 [8], наименьшее количество — 10 (часть штаммов сублинии B0/W-148). Рядом с кассетой CRISPR1 расположено 9 *cas*-генов: *cas2*, *cas1*, *cas6*, *cas5*, *cas4*, *cas3*, *cas2*, *cas10* (*cas1*) и *cas6* (см. рис. 2). Отличительной особенностью *cas*-генов у *M. tuberculosis* является их высокая консервативность. Среди проанализированных геномов не были обнаружены мутации в генах *cas1*, *cas2*, *cas4*, *cas2* и *cas6*. Мутации в других генах единичны и

носили случайный характер (табл. 1). Кассета CRISPR2 отделена от первой кассеты участком длиной около ~ 1300 п.н. (см. рис. 2). В данном промежутке аннотированы две транспозазы принадлежащие к семейству IS6110 [34]. Стоит также отметить, что для систем CRISPR-Cas *M. tuberculosis* характерна короткая лидерная последовательность из 48 п.н. [34].

ОСОБЕННОСТИ CRISPR-CAS СИСТЕМ У РАЗЛИЧНЫХ ЛИНИЙ В СОСТАВЕ *M. TUBERCULOSIS*

Линия Beijing

Для изолятов линии Beijing характерна делеция региона, содержащего гены *cas1*, *cas2*, *cas5*, *cas6* (см. табл. 1) и кассету CRISPR1 [8, 34]. Оставшаяся кассета CRISPR2 содержит не 18, а 14 спейсеров, 10 из которых (Sp1-Sp10) являются общими для всех линий *M. tuberculosis*, а 4 спейсера (SpB11-SpB14, где B обозначает Beijing) специфичны для рассматриваемой линии (рис. 3). Эти спейсеры не встречаются у представителей других линий *M. tuberculosis*.

Следует отметить, что в промежутке между геном *cas4* и кассетой CRISPR2 у большинства штаммов аннотировано две транспозазы. А сам ген *cas4* заметно короче своего ортолога из других линий — длина кодируемого им белка составляет либо 76 а. о., либо 116–118 а. о., тогда как в штаммах из других линий *M. tuberculosis* — 302 а. о. При длине около 100 а. о. в данном белке не сохраняются консервативные домены, необходимые для его взаимодействия с геном *cas3* (внутри комплекса *cas1-cas4-cas3*) [41]. В связи с этим в линии Beijing, возможно, нарушена стадия интерференции.

Линия Beijing начала формироваться в Северном Китае, Корее и Японии около 7000 лет назад [37] (рис. 4). По всей вероятности, после отделения данной линии формирование кассеты CRISPR2 продолжалось с участием специфичных для неё спейсеров (SpB11-SpB14), что могло быть вызвано отличиями в факторах окружающей среды, с которыми встречался патоген. Далее в линии Beijing произошла потеря нескольких *cas*-генов, в том числе *cas1* и *cas2*, участвующих в интеграции новых спейсеров, и формирование кассеты прекратилось. Таким образом,

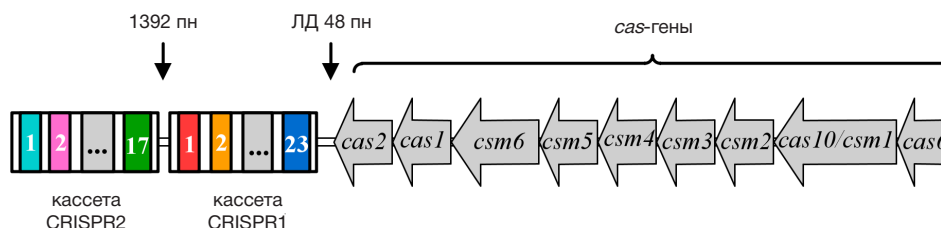


Рис. 2. Общая схема строения системы CRISPR-Cas *M. tuberculosis* на примере штамма H37Rv. ЛД — лидерная последовательность

Таблица 1. Сравнительный анализ *cas*-генов из шести линий (и одной сублинии) *M. tuberculosis*

Линия \ Ген	<i>cas2</i>	<i>cas1</i>	<i>cas6</i>	<i>cas5</i>	<i>cas4</i>	<i>cas3</i>	<i>cas2</i>	<i>cas10</i>	<i>cas6</i>
Beijing и B0/W-148	гены отсутствуют				100%	99%	100%	99%	100%
EAI	100%*	100%	100%	100%		100%			
Haarlem			100%	100%		100%			
Ural			99%	99%		99%			
S			99%	99%		99%			
LAM			99%	99%	99%				

Примечание: * — представлены проценты идентичности, рассчитанные по программе BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

для всей линии характерно наличие единственной кассеты CRISPR2, включающей в себя 14 спейсеров. Однако у ряда изолятов в составе молодой в эволюционном отношении сублинии B0/W-148, входящей в состав линии Beijing, обнаружена утрата нескольких спейсеров. У одной части этих изолятов отсутствуют спейсеры SpB13 и SpB14, а у другой части нет всех 4 спейсеров (SpB11-SpB14), специфичных для линии Beijing. Стоит отметить, что спейсеры SpB11-SpB14 были также обнаружены нами в составе кассеты у некоторых штаммов *M. bovis*.

Высокая частота возникновения мутаций и снижение репарации ДНК, описанные в литературе у линии Beijing [42], возможно связанные с редукцией CRISPR-Cas системы, потенциально могут служить причиной изменчивости линии и приводить к возникновению ЛУ. Предположение, что редуцированные или отсутствующие CRISPR-Cas системы связаны с ЛУ, соответствует результатам, полученным в недавно проведенном исследовании патогенной бактерии *Campylobacter jejuni*, в котором штаммы, вызывающие наиболее тяжелые гастроэнтериты и постинфекционные осложнения, также имели укороченные CRISPR-кассеты или полностью были лишены CRISPR-Cas систем [20, 43].

Линии Ural и Haarlem

Для линий Ural и Haarlem характерны инсерции спейсеров. Они происходят в кассете CRISPR1 после Sp3. Наличие вставки характерно только для части проанализированных изолятов Haarlem и всех изолятов линии Ural. Стоит отметить, что данные спейсеры встречаются среди некоторых изолятов *M. bovis*, а также у двух изолятов линии EAI, в связи с чем нельзя исключать наличие событий рекомбинации и горизонтального переноса генов (ГПГ).

Нами были также зафиксированы единичные случаи потери и приобретения спейсеров в кассете CRISPR2 у линий Ural и Haarlem. Так, у 3 изолятов линии Ural наблюдается потеря спейсеров Sp4-Sp6 в кассете CRISPR2, а у 2 изолятов линии Haarlem — потеря спейсера Sp6 из второй кассеты.

Линия EAI

Для линии EAI характерны наиболее протяженные CRISPR-кассеты среди всех линий *M. tuberculosis*. Возможно, это отчасти связано с тем, что линия EAI одна из наиболее

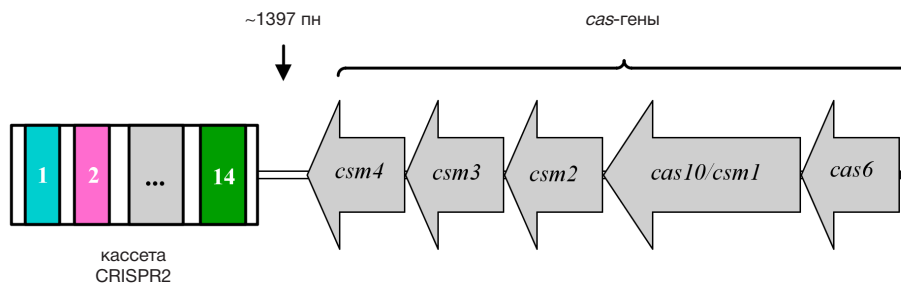


Рис. 3. Строение системы CRISPR-Cas линии Beijing *M. tuberculosis*. Десять высококонсервативных спейсеров, общих для всех линий *M. tuberculosis*, располагаются на дистальном по отношению к лидерной последовательности конце кассеты CRISPR2 и являются предковыми спейсерами, соответствующими более древнему состоянию CRISPR-иммунитета [10]. «Новые» спейсеры, приобретенные недавно, как уже было сказано выше, располагаются рядом с лидерной последовательностью

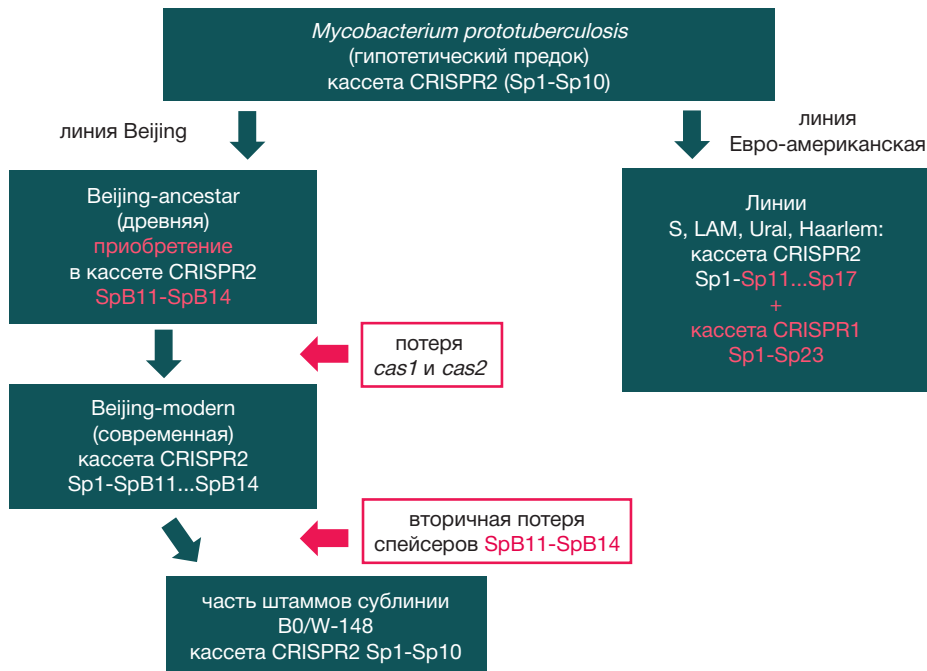


Рис. 4. Эволюция CRISPR-кассеты у линии Beijing *M. tuberculosis*. После формирования первых 10 спейсеров кассеты CRISPR2 (наиболее древних) произошло расхождение Евро-американской линии к линии Beijing. Спейсеры SpB11-SpB14 линии Beijing не тождественны Sp11-Sp14 из других линий *M. tuberculosis*. В связи с утратой генов *cas1* и *cas2*, участвующих в интеграции новых спейсеров, рост кассеты CRISPR2 в линии Beijing, по-видимому, остановился. Строение системы CRISPR-Cas у линии Beijing оставалось неизменным на протяжении достаточно долгого периода, но у наиболее молодой сублинии (B0/W-148) мы наблюдаем потерю четырех спейсеров (SpB11-SpB14), приобретенных самыми последними

древних. В ряде случаев длина кассеты CRISPR2 превышает 24 спейсера, а длина кассеты CRISPR1 — 30. Наибольшее количество спейсеров нами было обнаружено у изолята HN-024: 25 спейсеров для кассеты CRISPR1 и 34 спейсера для кассеты CRISPR2, среди которых присутствует ряд уникальных для этой линии спейсеров.

Линии S и LAM

Для линий S и LAM в общем характерно каноническое для *M. tuberculosis* строение CRISPR-Cas системы (см. рис. 2). В этих линиях может наблюдаться некоторый полиморфизм, например, у двух изолятов линии LAM в одном случае встречается потеря спейсеров Sp4-Sp6 кассеты CRISPR1, а в другом случае — потеря спейсера Sp20 из этой же кассеты.

В общем, описывая особенности кассет систем CRISPR-Cas *M. tuberculosis*, можно сказать, что кассета CRISPR1 отличается значительной вариабельностью, что и позволяет использовать ее для генотипирования [8]. Несмотря на то, что делеция спейсеров довольно распространена, среди десяти дистальных спейсеров Sp1-Sp10 кассеты CRISPR2 (по отношению к лидерной последовательности), являющихся предковыми и отличающихся высокой консервативностью, она практически не наблюдается, так же, как и мутации. Протоспейсеры для них до сих пор не установлены. Кроме того, хотя древние спейсеры и оцениваются часто как малозначимые по причине высокой изменчивости и быстрой эволюции профагов, защиту от которых они обеспечивали, они сохраняются неизменными у всех рассмотренных нами линий *M. tuberculosis*, не подвергаясь делеции. Поэтому возможно и другое объяснение: данные спейсеры могут иметь большое значение для жизнедеятельности бактерии, и их роль еще только предстоит узнать.

ПОИСК ФУНКЦИОНАЛЬНЫХ ПАРТНЕРОВ И КОМПЕНСАТОРНЫХ МЕХАНИЗМОВ ПРИ РЕДУКЦИИ ЧАСТИ СИСТЕМЫ CRISPR-CAS У ЛИНИИ BEIJING

Поиск возможных функциональных партнеров и механизмов компенсации функций генов *cas1*, *cas2*, *csm5*, *csm6* у представителей линии Beijing осуществляли с помощью метода филогенетического профайлинга генов в последовательностях геномов штаммов различных линий *M. tuberculosis* (анализировали 130

полногеномных последовательностей, доступных в NCBI). Филогенетический профиль (ФП) представляет собой бинарный вектор, определяющий наличие кодирующей последовательности изучаемого белка в геномах группы организмов [44]. Предполагается, что эволюция генов, принадлежащих к одному функциональному пути, происходит совместно, поэтому гены, характеризующиеся сходными или инвертированными ФП, могут выступать в качестве кандидатов на роль функциональных партнеров или компенсаторных механизмов соответственно.

Филогенетический профайлинг включал в себя определение групп ортологов в геномах штаммов различных линий *M. tuberculosis*, построение бинарных векторов, построение матрицы попарных расстояний между векторами и кластеризацию ФП. Построение и визуализацию ФП генов проводили с использованием программ OrthoFinder v.2.0.0 [45] и Count [46]. Матрицу попарных расстояний между ФП получали с использованием значений взаимной информации (MI): $D_{MI} = 1 - MI$. Кластерный анализ проводили с помощью метода невзвешенного попарного среднего (UPGMA) [47].

Оценка результатов кластерного анализа ФП позволила установить гены, характеризующиеся сходным характером эволюционных событий по отношению к анализируемому гену *cas1*, *cas2*, *csm5*, *csm6* (рис. 5, А). Потеря части системы CRISPR-Cas у ряда изолятов *M. tuberculosis* (принадлежащих линии Beijing) сопровождалась, по видимому, по крайней мере двумя эволюционными событиями потери и событием приобретения участков генома (в разных участках хромосомы) (рис. 5, Б, В).

В первом случае, в результате анализа ФП в геномах линии Beijing *M. tuberculosis* были обнаружены специфичные для линии протяженные делеции. В результате данных событий ортологи генов Rv0071, Rv0072, Rv0073 и Rv1761c, Rv1760, Rv1758 (идентификаторы соответствуют генам в геноме *M. tuberculosis* H37Rv) (табл. 2) приобрели схожие ФП (профили партнеров, см. рис. 5, Б). Стоит отметить, что хромосомный регион, содержащий гены Rv1761c, Rv1760 и Rv1758, фланкирован инвертированными повторами IS-элементов *IS6110*, относящихся к семейству IS3. Во втором случае, в результате анализа ФП в геномах линии Beijing *M. tuberculosis* была обнаружена специфичная для линии протяженная инсерция, и в результате этих событий ортологи генов CFBS_RS10335, CFBS_RS10345, CFBS_RS10350, CFBS_RS10355, CFBS_RS10360, CFBS_RS10365, CFBS_RS21395 (идентификаторы соответствуют

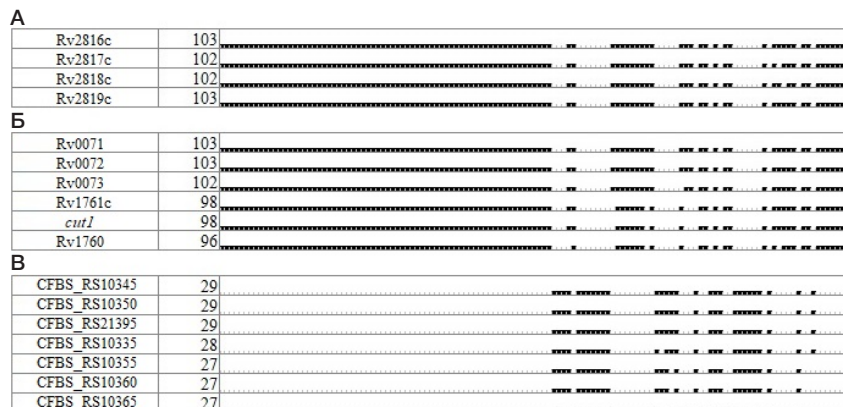


Рис. 5. Филогенетические профили генов в последовательностях геномов штаммов различных линий *M. tuberculosis*. В каждой строчке последовательно указаны идентификатор гена, количество геномов, содержащих ортолог, а также ФП гена, где пробелы соответствуют отсутствию гена в геноме. А. ФП *cas*-генов *M. tuberculosis* H37Rv: *cas2* (Rv2816c), *cas1* (Rv2817c), *csm6* (Rv2818c), *csm5* (Rv2819c). Б. ФП предполагаемых функциональных партнеров *cas*-генов *M. tuberculosis* в геноме штамма H37Rv. В. ФП генов, предположительно участвующих в формировании компенсаторных механизмов в геноме штамма CDC5079 (линия Beijing) в результате утраты части системы CRISPR-Cas

Таблица 2. Характеристика генов — возможных функциональных партнеров и генов, участвующих в компенсаторных механизмах для систем CRISPR-Cas*

Эволюционное событие	Идентификатор гена	Продукт (белок)	Функция белка
Инсерция	CFBS_RS10360 (<i>tuf</i>)	Железорегулируемый фактор элонгации трансляции Tu Tuf	Принимает участие в трансляции белка
	CFBS_RS10365	MDR-дегидрогеназа	Для группы MDR-дегидрогеназ характерны различные виды активности, включая основную алкогольдегидрогеназную, хинонредуктазную, сорбитолдегидрогеназную, формальдегиддегидрогеназную, кеторедуктазную и другие
Делеция	Rv1758 (<i>cut1</i>)	Кутиназа	Фермент, катализирующий гидролиз кутина [48]
	Rv1760	Диацилглицеролацетилтрансфераза	Катализирует заключительные стадии биосинтеза триацилглицерина и других компонентов клеточной стенки микобактерий [49]
	Rv0072	Глутамин ABC-транспортер пермеаза	ABC-транспортер, участие в транспорте глутамин (пермеаза); вероятно, отвечает за транслокацию субстрата через мембрану [50]
	Rv0073	Глутамин ABC-транспортер АТФ-связывающий белок	ABC-транспортер, участие в транспорте глутамин (связывающий протеин); вероятно, отвечает за энергетическую связь с транспортной системой [50]

Примечание: * — в таблице описаны только аннотированные гены; оставшиеся гены кодируют гипотетические белки с неизвестными функциями (Rv1761с и CFBS_RS10335, CFBS_RS10345, CFBS_RS10350, CFBS_RS10355, CFBS_RS21395) либо, в случае Rv0071, представляют собой мобильный самосплайсирующийся ретроэлемент, так называемый интрон группы II.

генам в геноме *M. tuberculosis* CCDC5079) (см. табл. 2) приобрели близкие к инвертированным ФП (профили компенсаторов, см. рис. 5, В).

Таким образом, в ходе анализа ФП были обнаружены гены со сходным характером эволюционных событий: было установлено, что потеря генов *cas1*, *cas2*, *csm5*, *csmb* у представителей линии Beijing *M. tuberculosis* сопровождалась событиями потери и приобретения ряда других генов (см. табл. 2). Данные гены-кандидаты потенциально способны участвовать в механизмах компенсации функций *cas*-генов или являться функциональными партнерами *cas*-генов у *M. tuberculosis*, что служит предметом дальнейших экспериментальных исследований.

ЗАКЛЮЧЕНИЕ

У *M. tuberculosis* наблюдается значительная вариативность систем CRISPR-Cas, от протяженных кассет, характерных для линии EAI, до частично редуцированных систем CRISPR-Cas у линии Beijing. Таким образом, наличие активной системы CRISPR-Cas типа III-A не является обязательным условием для эволюционного успеха (в отношении патогенности, вирулентности, трансмиссивности и адаптируемости) линии.

Утрата части кассеты и ряда *cas*-генов представителями линии Beijing *M. tuberculosis* привела, по-видимому, к полной или частичной потере способности их систем CRISPR-Cas разрушать чужеродную ДНК. Нарушение функционирования системы CRISPR-Cas у одной из наиболее успешных линий *M. tuberculosis* могло сопровождаться механизмами компенсации утерянных генов (например, в результате проведенного анализа была обнаружена специфичная для линии Beijing протяженная инсерция); а также утратой функциональных партнеров *cas*-генов, так как в общем случае предполагается, что ген, утративший своего функционального партнера, не будет сохраняться в геноме в результате отбора и подлжет элиминации, как подтверждение этому — обнаруженные специфичные для линии Beijing протяженные делеции. Следует отметить, что обнаруженные в ходе анализа закономерности в характере эволюционных событий потери и приобретения генов могут носить случайный

характер и подлежат экспериментальной верификации. Филогенетический профайлинг позволяет выдвинуть гипотезы и предоставляет необходимый для дальнейшего изучения материал.

Несмотря на то, что система CRISPR-Cas у представителей линии Beijing *M. tuberculosis* может быть неактивна, предполагается, что в случае линий, у которых присутствуют полный набор *cas*-генов и повторы, система сохраняет активность и способна участвовать в защите от чужеродной ДНК [34]. Учитывая короткую лидерную последовательность, характерную для всех систем CRISPR-Cas *M. tuberculosis*, возможно, при изучении этих систем у данного организма стоит отдать предпочтение исследованию их альтернативных функций, таких как регуляция экспрессии генов, репарация ДНК, формирование вирулентности и других.

Строение CRISPR-Cas системы *M. tuberculosis* изучено и описано достаточно подробно [8], а роль этих систем остается неизвестной. Для большинства микобактериальных спейсеров до настоящего времени не удалось найти протоспейсеры среди известных микобактериофагов. Это может быть связано с тем, что большая часть бактериофагов *M. tuberculosis* была выделена с использованием в качестве хозяина *M. smegmatis* с последующей проверкой способности поражать *M. tuberculosis* [34]. Разница в наборе спейсеров, наблюдаемая у различных линий, может привести к потенциальному различию в реализации бактериального иммунитета и, как следствие, к различной устойчивости к фагам, а также разнице в регуляции экспрессии. Помимо спейсеров интерес вызывает изучение роли белка Cas2 вне комплекса Cas1-Cas2, в связи с выдвинутой ранее гипотезой [15], что при «самостоятельной» активации данный белок способен прекращать трансляцию, и, вероятно, тем самым переводить клетку в состояние покоя, либо провоцировать апоптоз. Также представляет интерес поиск ингибиторов для Cas2. Особое внимание стоит обратить на возможную функциональную связь между системами CRISPR-Cas, в частности *cas*-генами, и TA-системами [38]. Возможное участие систем CRISPR-Cas в вирулентности и устойчивости к лекарственным препаратам позволит разработать новые подходы в борьбе с лекарственно-устойчивыми штаммами *M. tuberculosis*.

Литература

- Marraffini LA, Sontheimer EJ. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet.* 2010; 11 (3): 181–190.
- Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics.* 2007; 8: 172. doi:10.1186/1471-2105-8-172.
- Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol.* 1987; 169 (12): 5429–33.
- Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology.* 2005; 151 (8): 2551–61.
- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science.* 2007; 315: 1709–1712.
- Westra ER, Buckling A, Fineran PC. CRISPR-Cas systems: beyond adaptive immunity. *Nat Rev Microbiol.* 2014; 12 (5): 317–326. DOI: 10.1038/nrmicro3241.
- Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *Journal of Clinical Microbiology.* 1997; 35 (4): 907–14.
- van Embden JD, van Gorkom T, Kremer K, Jansen R, van Der Zeijst BA, Schouls LM. Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. *J Bacteriol.* 2000 May; 182 (9): 2393–401.
- Haft DH, Selengut J, Mongodin EF, Nelson KE. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol.* 2005; 1 (6): e60.
- Gogleva AA, Gelfand MS, Artamonova II. Comparative analysis of CRISPR cassettes from the human gut metagenomic contigs. *BMC Genomics.* 2014; 15 (1): 202. DOI:10.1186/1471-2164-15-202.
- Takeuchi N, Wolf YI, Makarova KS, Koonin EV. Nature and Intensity of Selection Pressure on CRISPR-Associated Genes. *J Bacteriol.* 2012; 194 (5): 1216–25.
- Makarova KS, Wolf YI, Alkhnbashi OS, et al. An updated evolutionary classification of CRISPR–Cas systems. *Nature reviews Microbiology.* 2015; 13 (11): 722–36. DOI:10.1038/nrmicro3569.
- Makarova KS, Zhang F, Koonin EV. SnapShot: Class 1 CRISPR-Cas Systems. *Cell.* 2017; 168 (5): 946–46.
- Goeders N, van Melderen L. Toxin-Antitoxin Systems as Multilevel Interaction Systems. *Toxins.* 2014; 6 (1): 304–24.
- Makarova KS, Anantharaman V, Aravind L, Koonin EV. Live virus-free or die: coupling of antiviral immunity and programmed suicide or dormancy in prokaryotes. *Biol Direct.* 2012; 7: 40.
- Viswanathan P, Murphy K, Julien B, Garza AG, Kroos L. Regulation of dev, an Operon That Includes Genes Essential for *Myxococcus xanthus* Development and CRISPR-Associated Genes and Repeats. *J Bacteriol.* 2007; 189 (10): 3738–50.
- Babu M, Beloglazova N, Flick R, Graham C, Skarina T, Nocek B, et al. A dual function of the CRISPR-Cas system in bacterial antiviral immunity and DNA repair. *Mol Microbiol.* 2011; 79 (2): 484–502.
- Heussler GE, Cady KC, Koeppen K, Bhuju S, Stanton BA, O'Toole GA. Clustered Regularly Interspaced Short Palindromic Repeat-Dependent, Biofilm-Specific Death of *Pseudomonas aeruginosa* Mediated by Increased Expression of Phage-Related Genes. *mBio.* 2015; 6 (3): e00129-15.
- Sampson TR, Saroj SD, Llewellyn AC, Tzeng Y-L, Weiss DS. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature.* 2013; 497: 254–7.
- Louwen R, Horst-Krefit D, de Boer AG, van der Graaf L, de Knecht G, Hamersma M, et al. A novel link between *Campylobacter jejuni* bacteriophage defence, virulence and Guillain-Barré syndrome. *Eur J Clin Microbiol Infect Dis.* 2013; 32 (2): 207–26.
- Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. *Nature.* 2012; 482, 331–38.
- Bhaya D, Davison M, Barrangou R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet.* 2011; 45: 273–97.
- Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct.* 2006; 1: 7.
- Gagneux S, Small PM. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect Dis.* 2007; 7 (5): 328–37.
- Прозоров А. А., Даниленко В. Н. Микобактерии туберкулезного комплекса: геномика, молекулярная эпидемиология, пути эволюции. *Успехи современной биологии.* 2011; 131 (3): 227–43.
- Jagielski T, van Ingen J, Rastogi N, Dziadek J, Mazur PK, Bielecki J. Current methods in the molecular typing of *Mycobacterium tuberculosis* and other *Mycobacteria*. *Biomed Res Int.* 2014; 2014:645802.
- Прозоров А. А., Федорова И. А., Беккер О. Б., Даниленко В. Н. Факторы вирулентности *Mycobacterium tuberculosis*: генетический контроль, новые концепции. *Генетика.* 2014; 50 (8): 885–908.
- Reiling N, Homolka S, Walter K, Brandenburg J, Niwinski L, Ernst M et al. Clade specific virulence patterns of *Mycobacterium tuberculosis* complex strains in human primary macrophages and aerogenically infected mice. *MBio.* 2013; 4 (4): pii: e00250-13.
- Lasunskaja E, Ribeiro SC, Manicheva O, Gomes LL, Suffys PN, Mokrousov I, et al. Emerging multidrug resistant *Mycobacterium tuberculosis* strains of the Beijing genotype circulating in Russia express a pattern of biological properties associated with enhanced virulence. *Microbes Infect.* 2010; 12 (6): 467–75.
- Hanekom M, Gey van Pittius NC, McEvoy C, Victor TC, Van Helden PD, Warren RM. *Mycobacterium tuberculosis* Beijing genotype: a template for success. *Tuberculosis (Edinb).* 2011; 91 (6): 510–23.
- Wirth T, Hildebrand F, Allix-Béguec C, Wölbeling F, Kubica T, Kremer K, et al. Origin, spread and demography of the *Mycobacterium tuberculosis* complex. *PLoS Pathog.* 2008; 4 (9): e1000160.
- Mokrousov I. The quiet and controversial: Ural family of *Mycobacterium tuberculosis*. *Infect Genet Evol.* 2012; 12 (4): 619–29.
- Mikhecheva NE, Zaychikova MV, Melerzanov AV, Danilenko VN. A nonsynonymous SNP catalog of *Mycobacterium tuberculosis* virulence genes and its use for detecting new potentially virulent sublineages. *Genome Biol Evol.* 2017; 9 (4): 887–99.
- He L, Fan X, Xie J. Comparative genomic structures of *Mycobacterium* CRISPR-Cas. *J Cell Biochem.* 2012; 113 (7): 2464–73.
- Brosch R, Pym AS, Gordon SV, Cole ST. The evolution of mycobacterial pathogenicity: clues from comparative genomics. *Trends Microbiol.* 2001; 9 (9): 452–58.
- Homolka S, Projahn M, Feuerriegel S, Ubben T, Diel R, Nübel U et al. High resolution discrimination of clinical *Mycobacterium tuberculosis* complex strains based on single nucleotide polymorphisms. *PLoS One.* 2012; 7 (7): e39855.
- Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, et al. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat Genet.* 2015; 47 (3): 242–49.
- Zaychikova MV, Zakharevich NV, Sagaidak MO, Bogolubova NA, Smirnova TG, Andreevskaya SN, et al. *Mycobacterium tuberculosis* Type II Toxin-Antitoxin Systems: Genetic Polymorphisms and Functional Properties and the Possibility of Their Use for Genotyping. *PLoS One.* 2015; 10: e0143682.
- Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpidis NC,

- et al. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics*. 2007; 8: 209.
40. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res*. 2007; 35: 52–7.
 41. Numata T, Inanaga H, Sato C, Osawa T. Crystal structure of the Csm3-Csm4 subcomplex in the type III-A CRISPR-Cas interference complex. *J Mol Biol*. 2015; 427 (2): 259–73.
 42. Mestre O, Luo T, Dos Vultos T, Kremer K, Murray A, Namouchi A, et al. Phylogeny of *Mycobacterium tuberculosis* Beijing strains constructed from polymorphisms in genes involved in DNA replication, recombination and repair. *PLoS One*. 2011; 6 (1): e16020.
 43. Freidlin PJ, Nissan I, Luria A, Goldblatt D, Schaffer L, Kaidar-Shwartz H, et al. Structure and variation of CRISPR and CRISPR-flanking regions in deleted-direct repeat region *Mycobacterium tuberculosis* complex strains. *BMC Genomics*. 2017; 18 (1): 168.
 44. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci USA*. 1999; 96 (8): 4285–8.
 45. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*. 2015; 16: 157.
 46. Count CM.: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics*. 2010; 26 (15): 1910–2.
 47. Sokal R, Michener C. A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*. 1958; 38: 1409–1438.
 48. Monu, Meena LS. Roles of Triolein and Lipolytic Protein in the Pathogenesis and Survival of *Mycobacterium tuberculosis*: a Novel Therapeutic Approach. *Appl Biochem Biotechnol*. 2016; 178 (7): 1377–89.
 49. Deb C, Lee CM, Dubey VS, Daniel J, Abomoelak B, Sirakova TD, et al. A Novel In Vitro Multiple-Stress Dormancy Model for *Mycobacterium tuberculosis* Generates a Lipid-Loaded, Drug-Tolerant, Dormant Pathogen. *PLoS ONE*. 2009; 4 (6): e6077.
 50. Braibant M, Gilot P, Content J. The ATP binding cassette (ABC) transport systems of *Mycobacterium tuberculosis*. *FEMS Microbiol Rev*. 2000; 24 (4): 449–67.

References

1. Marraffini LA, Sontheimer EJ. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet*. 2010; 11 (3): 181–190.
2. Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics*. 2007; 8: 172. DOI: 10.1186/1471-2105-8-172.
3. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol*. 1987; 169 (12): 5429–33.
4. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*. 2005; 151 (8): 2551–61.
5. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. 2007; 315: 1709–1712.
6. Westra ER, Buckling A, Fineran PC. CRISPR-Cas systems: beyond adaptive immunity. *Nat Rev Microbiol*. 2014; 12 (5): 317–326. DOI: 10.1038/nrmicro3241.
7. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol*. 1997; 35 (4): 907–14.
8. van Embden JD, van Gorkom T, Kremer K, Jansen R, van Der Zeijst BA, Schouls LM. Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. *J Bacteriol*. 2000 May; 182 (9): 2393–401.
9. Haft DH, Selengut J, Mongodin EF, Nelson KE. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol*. 2005; 1 (6): e60.
10. Gogleva AA, Gelfand MS, Artamonova II. Comparative analysis of CRISPR cassettes from the human gut metagenomic contigs. *BMC Genomics*. 2014; 15 (1): 202. DOI: 10.1186/1471-2164-15-202.
11. Takeuchi N, Wolf YI, Makarova KS, Koonin EV. Nature and Intensity of Selection Pressure on CRISPR-Associated Genes. *J Bacteriol*. 2012; 194 (5): 1216–25.
12. Makarova KS, Wolf YI, Alkhnbashi OS, et al. An updated evolutionary classification of CRISPR–Cas systems. *Nature reviews Microbiology*. 2015; 13 (11): 722–36. DOI: 10.1038/nrmicro3569.
13. Makarova KS, Zhang F, Koonin EV. SnapShot: Class 1 CRISPR-Cas Systems. *Cell*. 2017; 168 (5): 946–46.
14. Goeders N, van Melderen L. Toxin-Antitoxin Systems as Multilevel Interaction Systems. *Toxins*. 2014; 6 (1): 304–24.
15. Makarova KS, Anantharaman V, Aravind L, Koonin EV. Live virus-free or die: coupling of antiviral immunity and programmed suicide or dormancy in prokaryotes. *Biol Direct*. 2012; 7: 40.
16. Viswanathan P, Murphy K, Julien B, Garza AG, Kroos L. Regulation of dev, an Operon That Includes Genes Essential for *Myxococcus xanthus* Development and CRISPR-Associated Genes and Repeats. *Journal of Bacteriology*. 2007; 189 (10): 3738–50.
17. Babu M, Beloglazova N, Flick R, Graham C, Skarina T, Nocek B, et al. A dual function of the CRISPR-Cas system in bacterial antiviral immunity and DNA repair. *Mol Microbiol*. 2011; 79 (2): 484–502.
18. Heussler GE, Cady KC, Koeppen K, Bhuju S, Stanton BA, O'Toole GA. Clustered Regularly Interspaced Short Palindromic Repeat-Dependent, Biofilm-Specific Death of *Pseudomonas aeruginosa* Mediated by Increased Expression of Phage-Related Genes. *mBio*. 2015; 6 (3): e00129-15.
19. Sampson TR, Saroj SD, Llewellyn AC, Tzeng Y-L, Weiss DS. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature*. 2013; 497: 254–257.
20. Louwen R, Horst-Kreft D, de Boer AG, van der Graaf L, de Knecht G, Hamersma M, et al. A novel link between *Campylobacter jejuni* bacteriophage defence, virulence and Guillain-Barré syndrome. *Eur J Clin Microbiol Infect Dis*. 2013; 32 (2): 207–26.
21. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. *Nature*. 2012; 482, 331–38.
22. Bhaya D, Davison M, Barrangou R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet*. 2011; 45: 273–97.
23. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct*. 2006; 1: 7.
24. Gagneux S, Small PM. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect Dis*. 2007; 7 (5): 328–37.
25. Prozorov AA, Danilenko VN. *Mycobacterii tuberculoznogo kompleksa: genomika, molekularnaja epidemiologija, puti evoluzii*. *Usphehi sovremennoy biologii*. 2011; 1 (6): 483–95.
26. Jagielski T, van Ingen J, Rastogi N, Dziadek J, Mazur PK, Bielecki J. Current methods in the molecular typing of *Mycobacterium tuberculosis* and other *Mycobacteria*. *Biomed*

- Res Int. 2014; 2014:645802.
27. Prozorov AA, Fedorova IA, Bekker OB, Danilenko VN. Faktory virulentnosti Mycobacterium tuberculosis: geneticheskiy kontrol, novie konceptzii. *Genetika*. 2014; 50 (8): 775–97.
 28. Reiling N, Homolka S, Walter K, Brandenburg J, Niwinski L, Ernst M et al. Clade specific virulence patterns of Mycobacterium tuberculosis complex strains in human primary macrophages and aerogenically infected mice. *MBio*. 2013; 4 (4): pii: e00250-13.
 29. Lasunskaja E, Ribeiro SC, Manicheva O, Gomes LL, Suffys PN, Mokrousov I, et al. Emerging multidrug resistant Mycobacterium tuberculosis strains of the Beijing genotype circulating in Russia express a pattern of biological properties associated with enhanced virulence. *Microbes Infect*. 2010; 12 (6): 467–75.
 30. Hanekom M, Gey van Pittius NC, McEvoy C, Victor TC, Van Helden PD, Warren RM. Mycobacterium tuberculosis Beijing genotype: a template for success. *Tuberculosis (Edinb)*. 2011; 91 (6): 510–23.
 31. Wirth T, Hildebrand F, Allix-Béguec C, Wölbeling F, Kubica T, Kremer K, et al. Origin, spread and demography of the Mycobacterium tuberculosis complex. *PLoS Pathog*. 2008; 4 (9): e1000160.
 32. Mokrousov I. The quiet and controversial: Ural family of Mycobacterium tuberculosis. *Infect Genet Evol*. 2012; 12 (4): 619–29.
 33. Mikheeva NE, Zaychikova MV, Melerzanov AV, Danilenko VN. A nonsynonymous SNP catalog of Mycobacterium tuberculosis virulence genes and its use for detecting new potentially virulent sublineages. *Genome Biol Evol*. 2017; 9 (4): 887–99.
 34. He L, Fan X, Xie J. Comparative genomic structures of Mycobacterium CRISPR-Cas. *J Cell Biochem*. 2012; 113 (7): 2464–73.
 35. Brosch R, Pym AS, Gordon SV, Cole ST. The evolution of mycobacterial pathogenicity: clues from comparative genomics. *Trends Microbiol*. 2001; 9 (9): 452–58.
 36. Homolka S, Projahn M, Feuerriegel S, Ubben T, Diel R, Nübel U et al. High resolution discrimination of clinical Mycobacterium tuberculosis complex strains based on single nucleotide polymorphisms. *PLoS One*. 2012; 7 (7): e39855.
 37. Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, et al. Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. *Nat Genet*. 2015; 47 (3): 242–49.
 38. Zaychikova MV, Zakharevich NV, Sagaidak MO, Bogolubova NA, Smirnova TG, Andreevskaya SN, et al. Mycobacterium tuberculosis Type II Toxin-Antitoxin Systems: Genetic Polymorphisms and Functional Properties and the Possibility of Their Use for Genotyping. *PLoS One*. 2015; 10: e0143682
 39. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC et al. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics*. 2007; 8: 209.
 40. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res*. 2007; 35: 52–57.
 41. Numata T, Inanaga H, Sato C, Osawa T. Crystal structure of the Csm3-Csm4 subcomplex in the type III-A CRISPR-Cas interference complex. *J Mol Biol*. 2015; 427 (2): 259–73.
 42. Mestre O, Luo T, Dos Vultos T, Kremer K, Murray A, Namouchi A, et al. Phylogeny of Mycobacterium tuberculosis Beijing strains constructed from polymorphisms in genes involved in DNA replication, recombination and repair. *PLoS One*. 2011; 6 (1): e16020.
 43. Freidlin PJ, Nissan I, Luria A, Goldblatt D, Schaffer L, Kaidar-Shwartz H, et al. Structure and variation of CRISPR and CRISPR-flanking regions in deleted-direct repeat region Mycobacterium tuberculosis complex strains. *BMC Genomics*. 2017; 18 (1): 168.
 44. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. USA*. 1999; 96 (8): 4285–8.
 45. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*. 2015; 16: 157.
 46. Count CM.: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics*. 2010; 26 (15): 1910–2.
 47. Sokal R, Michener C. A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*. 1958; 38: 1409–1438.
 48. Monu, Meena LS. Roles of Triolein and Lipolytic Protein in the Pathogenesis and Survival of Mycobacterium tuberculosis: a Novel Therapeutic Approach. *Appl Biochem Biotechnol*. 2016; 178 (7): 1377–89.
 49. Deb C, Lee CM, Dubey VS, Daniel J, Abomoelak B, Sirakova TD, et al. A Novel In Vitro Multiple-Stress Dormancy Model for Mycobacterium tuberculosis Generates a Lipid-Loaded, Drug-Tolerant, Dormant Pathogen. *PLoS ONE*. 2009; 4 (6): e6077.
 50. Braibant M, Gilot P, Content J. The ATP binding cassette (ABC) transport systems of Mycobacterium tuberculosis. *FEMS Microbiol Rev*. 2000; 24 (4): 449–67.