

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}

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

²Department 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 *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}

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

²Кафедра 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 IN BACTERIA: STRUCTURE AND CLASSIFICATION

To date, CRISPR-Cas systems have been identified in approximately 40% of bacterial and 90% of archaeal genomes [1, 2]. These systems consist of two essential components: CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) arrays and Cas (CRISPR-associated) proteins. Repetitive sequences of equal length alternating with unique regions (spacers) were described as early as 1987 in the *E. coli* genome [3], but at that time their function was unclear. In the 2000s, CRISPR-Cas systems were shown to have a role in bacterial immunity [4, 5]. By now, they have been proved to participate in a number of various cellular processes, including DNA repair, regulation of gene expression, virulence formation, etc. [6]. Interestingly, direct repeats (DR) discovered in the *M. tuberculosis* genome as early as 1990s have been used in the genotyping (spoligotyping) of mycobacteria even before the immune function of CRISPR-Cas systems was described, and their polymorphism is well-studied [7, 8].

CRISPR-Cas systems are very diverse. Each functional array contains three essential elements: repeats, spacers and a leader sequence. Adjacent to the array is a set of *cas* genes coding for proteins with various functional domains interacting with nucleic acids [9]. Although the sets of *cas* genes ensuring the performance of different components of CRISPR-Cas molecular mechanisms are different, they do have common features. For example, the majority of known active CRISPR-Cas systems contain two proteins called Cas1 and Cas2. These proteins form a complex that integrates new spacers into the array. New spacers are inserted into the array next to the leader sequence. Throughout the array's lifetime some spacers can be lost as a result of recombination between repeats [10]. Partially or fully, the array can be acquired through horizontal gene transfer (HGT) [11].

CRISPR-Cas systems are classified based on the composition of their *cas*-loci. According to the currently used classification, they are subdivided into 2 big classes, 5 types and multiple subtypes [12]. Class 1 (types I, III and IV) comprises CRISPR-Cas systems with multisubunit effector complexes; in class 2 systems (types II and V) all functions of the effector complex are exerted by one protein, such as Cas9 [12]. Type II CRISPR-Cas systems are of paramount importance for biotechnology and specifically for genome editing, but they are quite rare and have been detected only in bacterial genomes [12]. The majority of CRISPR-Cas systems can be unambiguously assigned to one of its 5 main types. However, there are organisms whose *cas*-loci do not fit into the current classification.

Cas1 and *cas2* genes, the CRISPR-Cas components involved in the integration of new spacers into the array, deserve particular attention. Although there is evidence that both of them have their role in spacer integration, all enzymatic activities necessary for this process can be found in Cas1, whereas the catalytic activity of Cas2 is not required to form a Cas1-Cas2 complex or insert a new spacer. So far, we know that Cas2 is an mRNA interferase that specifically cleaves ribosome-bound mRNA. On the face of it, such activity seems to be "inappropriate" when it comes to the integration of new spacers. However, some researchers suggest that Cas2 may have originated from ancient mobile elements, such as toxin-antitoxin (TA) systems [13, 14]. In view of this, it may be assumed that Cas2 retains its ancestral toxin-like endoribonuclease activity in the CRISPR-Cas system, but the latter is reversely controlled through inhibition during interaction with Cas1 and formation of the Cas1-Cas2 complex. According to this hypothesis, if the CRISPR-Cas system fails to inhibit viral growth, Cas2 is activated (possibly through Cas1 degradation) and stops translation, driving the cell to suicide or into the dormant state. Cas2 participation in spacer integration may be connected to Cas1 regulation or stabilization following the formation of Cas1-Cas2 complex, which at the same time reversely inactivates Cas2 [15]. The possible participation of Cas2 in getting the cell into a persistent state is a promising area of pathogen research (*M. tuberculosis* research, in particular).

FUNCTIONS OF CRISPR-CAS SYSTEMS IN BACTERIA

Because CRISPR-Cas systems are widely spread and very diverse, it is no wonder why more evidence of their involvement in different cellular processes appears in the literature [6]. Apart from the role in the adaptive immunity, the most well-known and well-described of CRISPR-Cas functions is regulation of gene expression. For example, the life cycle of the soil bacteria *Myxococcus xanthus* includes stages of fruit body formation and sporulation. Formation of the fruit body and further differentiation of its cells into microspores is rigorously regulated by intercellular signals and intracellular signaling cascades in which type I-C CRISPR-Cas systems of *M. xanthus* act as a component of the positive feedback loop and participate in sporulation [16].

Today, there is evidence that CRISPR-Cas systems can engage in DNA repair. It has been established that purified Cas1 (YgbT) obtained from *Escherichia coli* is capable of interaction, both at the physical and genetic levels, with key components of DNA repair systems, such as genes *recB*, *recC* and *ruvB* [17]. The researchers have demonstrated that the *ygbT* deletion strain has increased sensitivity to DNA damage.

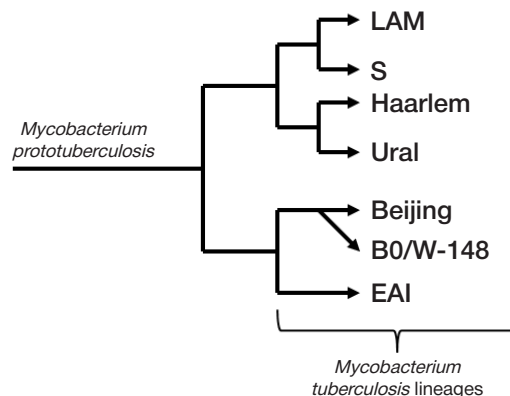


Fig. 1. The schematic representation of the phylogeny of the described *M. tuberculosis* lineages [33]

Similar phenotypes have been observed in the strains with a deleted CRISPR cluster; this indicates, at least, that some of CRISPR-Cas components are involved in DNA repair.

Another alternative function of CRISPR-Cas systems pertains to their participation in biofilm formation [18]. The study of the type 1-F CRISPR-Cas systems of the opportunistic pathogen *Pseudomonas aeruginosa* has revealed that this system inhibits biofilm formation. Such CRISPR-dependent ability relies on the interaction between a certain spacer and its prototype, the protospacer located in the bacteriophage genome. This interaction eventually leads to the induction of phage-related genes that, in turn, trigger death of surface cells. These findings suggest that CRISPR-Cas systems possess another mechanism unrelated to the adaptive immunity.

Bacteria usually regulate their gene expression post-transcriptionally by various small non-coding RNA. Although these RNA molecules control a great deal of cell physiology, only a few of them participate in the recognition of intrusive nucleic acids, ceding this role to CRISPR-Cas systems. Unlike eukaryotic systems, bacterial CRISPR-Cas systems cleave DNA, which means that if they should engage in the regulation of endogenous genes, the bacterial chromosome will be inevitably destroyed. Surprisingly, though, in 2013 an article was published in *Nature* reporting a mechanism of post-transcriptional regulation in *Francisella novicida*, in which the virulence gene is regulated by the Cas9 protein and CRISPR-associated small RNA [19]. Hypothetically, Cas9 directs its activity against endogenous mRNA (but not DNA). So far, the association between CRISPR-Cas systems and the ability of bacterial strains to exhibit increased virulence or even drug resistance has been shown in a number of research works [20].

Speaking of alternative functions of CRISPR-Cas systems, some authors hypothesize that biofilm formation in *Pseudomonas aeruginosa* is a by-product of a “classical” CRISPR-Cas immune function, whereas virulence in *Francisella novicida* or development regulation in *Myxococcus xanthus* have come about independently [6]. The history of gradual discovery of different CRISPR-Cas functions, starting with immune, resembles the exploration of RNA interference in eukaryotes. At first, RNA interference was shown to have a role in the immune defense, and it was not until later that its effects on various cellular processes were discovered, including gene

regulation and heterochromatin formation [21]. Some authors draw a parallel between CRISPR-Cas systems and RNA interference [22, 23].

CRISPR-CAS SYSTEMS IN MYCOBACTERIA: GENERAL STRUCTURE AND PECULIARITIES OF CAS-OPERONE IN *M. TUBERCULOSIS* H37RV

The *Mycobacterium* genus is represented by a wide range of organisms, including human pathogens among which members of the *Mycobacterium tuberculosis* complex (MTBC) are the most important. This complex includes *Mycobacterium tuberculosis*, the major causative agent of tuberculosis. *M. tuberculosis* is genetically heterogeneous and can be divided into several groups, or the so-called lineages. Each lineage is characterized by a certain set of mutations that have accumulated in the course of evolution [24–26]. Isolates of different lineages can be distinguished by their phenotype, specifically by the ability to develop drug resistance (DR), virulence and pathogenicity, all of which determine the severity of the disease [27, 28]. The most widespread and clinically significant lineages of *M. tuberculosis* are Beijing, Haarlem, LAM, and S. The Beijing lineage (in particular, the B0/W-148 sublineage that has emerged recently) is the most epidemiologically important one due to its high prevalence and propensity to develop DR [29, 30]. The Haarlem lineage is characterized by increased virulence [28]. Of certain interest are the lineages EAI and Ural, with their reduced virulence that makes them less prevalent [28]. EAI is an ancient lineage territorially limited to South East Asia [31]. Related to Haarlem, the Ural line is not very widespread, just like EAI, and appears to have reduced transmissibility [32] (Fig. 1).

Given its possible role in virulence formation [19, 20], CRISPR-Cas systems could become an interesting research object, especially in different *M. tuberculosis* lineages.

To date, CRISPR-Cas systems have been identified in 14 mycobacterial species [34]. All such systems are located on a chromosome. CRISPR arrays with more than 5 repeats have been identified in only 3 mycobacterial species: *M. tuberculosis* and *M. bovis*, which belong to the MTBC, and in the pathogenic *M. avium*. *M. avium* misses cas-genes that should be adjacent to the CRISPR array, and CRISPR loci in *M. tuberculosis* and

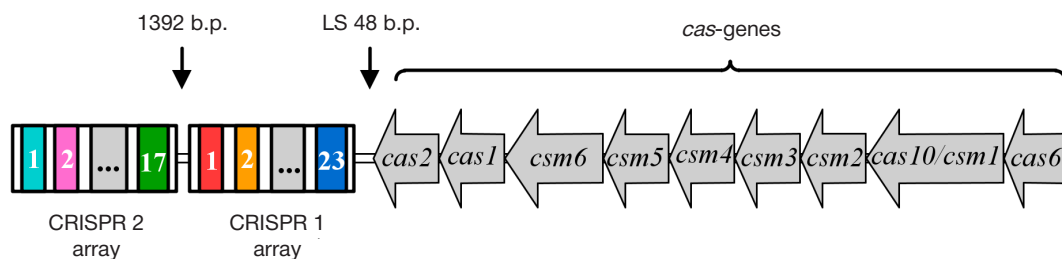


Fig. 2. Schematic of CRISPR-Cas structure in *M. tuberculosis* exemplified by the strain H37Rv. LS is leader sequence

Table 1. The comparative analysis of cas-genes found in 6 different lineages and one sublineage of *M. tuberculosis*

Lineage	Gene	cas2	cas1	csm6	csm5	csm4	csm3	csm2	cas10	cas6
Beijing и B0/W-148		missing				100%	99%	100%	99%	100%
EAI		100%*	100%	100%	100%		100%			
Haarlem				100%	100%		100%			
Ural				99%	99%		99%			
S										
LAM										

Note: * — represents percent identity computed in BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

M. bovis are very similar in terms of their structure. This reflects a close evolutionary relationship between them and is consistent with their phylogeny [34, 35]. *M. tuberculosis* CRISPR-Cas systems have a structure typically found in type III-A systems [34].

We have analyzed the CRISPR-Cas systems in 41 complete genome sequences of different *M. tuberculosis* lineages available in the NCBI RefSeq database, including 13 Beijing genomes, 3 B0/W-148 genomes, 2 EAI genomes, 10 Haarlem genomes, 1 Ural genome, 2 S genomes, and 10 LAM genomes. Additionally, we have analyzed a few draft genomes, including 7 B0/W-148 genomes, 4 URAL genomes, 3 EAI genomes, and 3 S genomes, the reason being the low number of complete genomes available. Genotyping was based on marker polymorphisms [36–38]. For some genomes the genotype of the isolate was already known from the literature. The search and analysis of CRISPR-Cas systems was conducted using two algorithms: CRISPRFinder and CRISPR Recognition Tool [39, 40]. Fig. 2 shows a typical structure of *M. tuberculosis* CRISPR-Cas systems exemplified by the H37Rv strain, the standard reference genome.

The majority of the analyzed *M. tuberculosis* strains had two long CRISPR-arrays (Fig. 2) [8]. The only exception was the strain 7199-99, which belongs to the Haarlem lineage; its

CRISPR2 array had been reduced starting from spacer 12 and including the region between the arrays, leading to the formation of a single array of 33 spacers. The largest number of spacers in an *M. tuberculosis* genome is 57 [8], the smallest is 10, as was the case with some of B0/W-148 strains. Adjacent to the CRISPR1 array were 9 *cas*-genes, namely *cas2*, *cas1*, *csml*, *csm2*, *csm3*, *csm4*, *csm5*, *cas10* (*cas1*), and *cas6* (Fig.2). The *cas*-genes of *M. tuberculosis* are highly conserved. In our study no mutations were detected in *cas1*, *cas2*, *csml*, *csm2* and *cas6*. Other analyzed genes had single random mutations (Table 1). The CRISPR2 array was separated from the CRISPR1 array by a sequence of ~ 1300 b.p. (Fig. 2) containing two annotated transposases that belong to the IS6110 family [34]. Of note, the CRISPR-Cas systems of *M. tuberculosis* typically have a short leader sequence of 48 b.p. [34].

DISTINCTIVE CHARACTERISTICS OF CRISPR-CAS SYSTEMS IN DIFFERENT *M. TUBERCULOSIS* LINEAGES

Beijing lineage

The region containing *cas1*, *cas2*, *csml*, *csm6* (Table 1) and the CRISPR1 array were missing in the analyzed Beijing isolates [8,

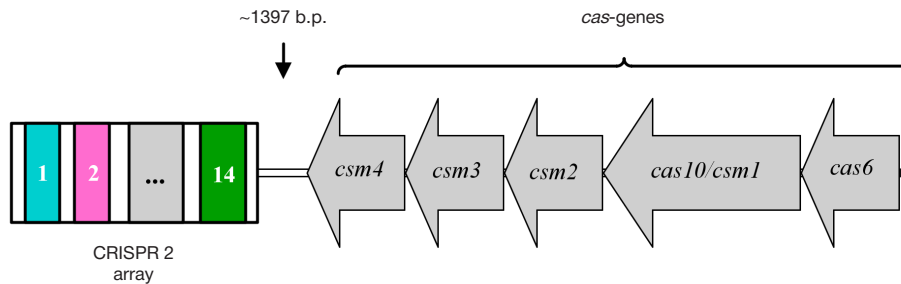


Fig. 3. The structure of the CRISPR-Cas system of the *M. tuberculosis* Beijing lineage. Ten highly conserved spacers shared by all *M. tuberculosis* lineages are located at the end of the CRISPR2 array distal to the leader sequence and are ancestral spacers mirroring the ancient state of CRISPR immunity [10]. More recent spacers are located next to the leader sequence

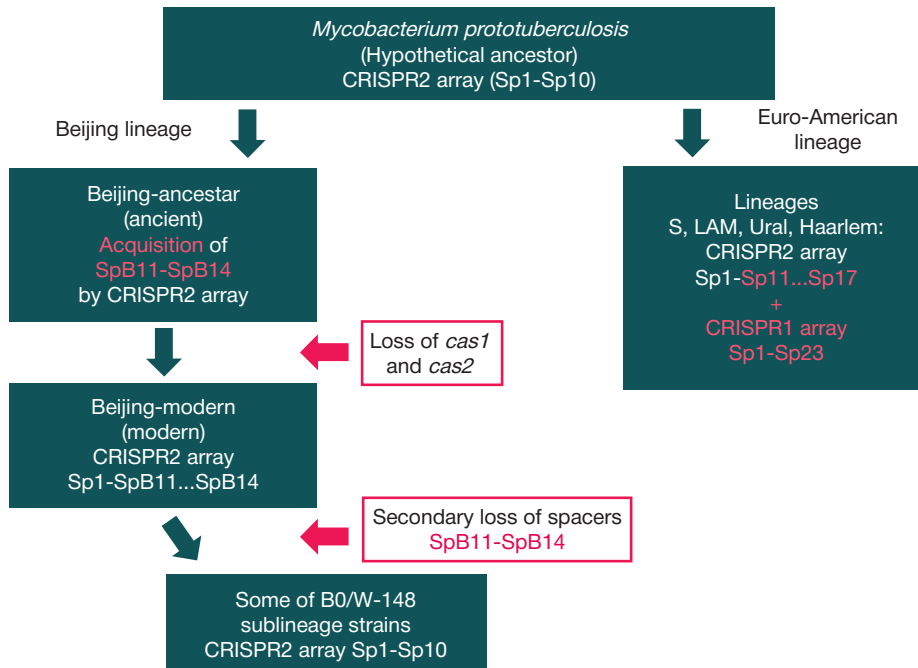


Fig. 4. Evolution of the CRISPR array in the *M. tuberculosis* Beijing lineage. After the first 10 spacers of the CRISPR2 array had been integrated (they are the most ancient ones), the Euro-American and Beijing lines separated. The spacers SpB11-SpB14 of the Beijing lineage are not identical to the spacers Sp11-Sp14 found in other *M. tuberculosis* lineages. Due to the loss of *cas1* and *cas2* genes involved in the integration of new spacers, formation of the CRISPR2 array probably stopped. The CRISPR-Cas structure of the Beijing lineage has remained intact for a long time, but its youngest sublineage B0/W-148 demonstrates a loss of 4 spacers SpB11-SpB14 that are the array's most recent acquisitions

34]. The remaining CRISPR2 array had only 14 spacers instead of 18, of which 10 (Sp1-Sp10) are shared by all *M. tuberculosis* lineages and 4 (SpB11-SpB14, B stands for Beijing) are specific to the Beijing lineage (Fig. 3). These spacers are absent in other *M. tuberculosis* lineages.

Curiously, the majority of the analyzed strains had two annotated transposases in the region between the gene *csm4* and the CRISPR2 array. In the Beijing lineage *csm4* is significantly shorter than its ortholog from other lines: the length of the protein it codes for is either 76 a.a.r. or 116 to 118 a.a.r., whereas in other *M. tuberculosis* lineages the protein length is 302 a.a.r. If the encoded protein is about 100 a.a.r. long, it cannot retain its conserved domains necessary for the interaction with *csm3* inside the *csm1-csm4-csm3* complex [41]. This implies that the interference stage may be disrupted in the Beijing lineage.

The Beijing lineage originated in North China, Korea and Japan about 7,000 years ago [37] (Fig. 4). It appears that after this lineage separated from others, its CRISPR2 array continued to incorporate CRISPR2-specific SpB11-SpB14 spacers. This could be due to the differences in the environmental factors the pathogen had to face. Then the lineage lost a few *cas*-genes, including *cas1* and *cas2* involved in the integration of new spacers, and the array growth stopped. As a result, representatives of the Beijing line normally have only one array (CRISPR2) with 14 spacers in it. However, some isolates of the evolutionary young Beijing sublineage B0/W-148 appear to have lost a few of them. A number of these isolates lack SpB13 and SpB14, while others have lost all 4 SpB11-SpB14 spacers specific to the Beijing lineage. Interestingly, we have found these 4 spacers in the CRISPR arrays of some *M. bovis* strains.

High frequency of mutations and decreased DNA repair in the Beijing lineage described in the literature [42] may result from the reduction of the CRISPR-Cas system and can be a potential cause of variability and drug resistance observed in the lineage. A hypothesized association between reduced or missing CRISPR-Cas systems and DR is consistent with the findings of the recent study of the *Campylobacter jejuni* pathogen, which demonstrated that the strains causing the most severe gastroenteritis and post-infectious complications have shortened CRISPR-arrays or totally lack the CRISPR-Cas system [20, 43].

Ural and Haarlem lineages

A typical feature of the Ural and Haarlem lineages is spacer insertions. They occur in the CRISPR array at the locus following the Sp3 spacer. Insertions are found in only some of the analyzed Haarlem isolates and all Ural isolates. Importantly, we observed those spacers in some *M. bovis* and two EAI isolates; therefore, past recombination events and horizontal gene transfer cannot be ruled out.

We also observed a few cases of spacer loss or acquisition by the CRISPR2 array in the Ural and Haarlem lineages. For example, 3 Ural isolates lacked the Sp4-Sp6 spacers in the CRISPR2 array, and 2 Haarlem isolates were missing the Sp6 spacer in it.

EAI lineage

Of all *M. tuberculosis* lineages, EAI has the longest CRISPR-arrays. EAI is one of the most ancient lineages, so this could be the reason. In some isolates, the CRISPR2 array is more than 24 spacers long, and the CRISPR1 array contains over 30 spacers. The largest number of spacers was found in the

isolate HN-024: 25 spacers in CRISPR1 and 34 spacers in CRISPR2; some of them were unique.

S and LAM lineages

On the whole, the S and LAM lineages have a canonical *M. tuberculosis* CRISPR-Cas structure (Fig. 2). A certain polymorphism can be observed. For example, one LAM isolate was missing the Sp4-Sp6 spacers in its CRISPR1 array, and another LAM isolate had lost the Sp20 spacer from the same array.

To sum up, the CRISPR1 array of *M. tuberculosis* is highly variable and therefore can be conveniently used in genotyping [8]. Although spacer deletions are common, they almost never occur in 10 highly conserved ancestral CRISPR2 spacers Sp1-Sp10 distal to the leader sequence. The same is true for mutations. Protospacers of Sp1-Sp10 remain unidentified. Although ancient spacers are regarded as barely significant because of their high variability and a rapid evolution of prophages which they protected the bacteria against, they look intact in all analyzed *M. tuberculosis* lineages and do not undergo deletions. This brings in another possible explanation: the Sp1-Sp10 spacers are vital for bacteria, and their role is yet to be elucidated.

THE SEARCH FOR FUNCTIONALLY RELATED PARTNERS AND COMPENSATORY MECHANISMS IN THE BEIJING LINEAGE WITH REDUCED CRISPR-CAS SYSTEMS

The search for functionally related partners and mechanisms compensating for the functions of *cas1*, *cas2*, *csm5*, and *csm6* in the Beijing lineage was conducted using a method of phylogenetic profiling and the genomic sequences of different *M. tuberculosis* lineages (in total, 130 complete genome sequences available in NCBI were analyzed). The phylogenetic profile (PP) is a binary vector determining the presence of a sequence coding for a protein of interest in the genomes of a group of organisms [44]. Hypothetically, the evolution of genes belonging to the same functional pathway happens simultaneously, therefore the genes with similar or inverted PP can be used as functionally related candidate partners or candidate compensatory mechanisms, respectively.

Using phylogenetic profiling we identified orthologous gene groups in different *M. tuberculosis* lineages, constructed binary vectors and a pairwise distance matrix for the vectors, and performed PP clusterization. Construction and visualization of PP were done in OrthoFinder v.2.0.0 [45] and Count [46]. The pairwise distance matrix was constructed based on the mutual information values (MI): $DMI=1-MI$. Cluster analysis was performed using the unweighted pair group method with arithmetic mean (UPGMA) [47].

The cluster analysis of PP allowed us to identify genes that had undergone evolutionary events similar to those undergone by *cas1*, *cas2*, *csm5*, and *csm6* (Fig. 5, A). The loss of some CRISPR-Cas components in a number of Beijing isolates of *M. tuberculosis* may have been accompanied by at least two evolutionary losses and one acquisition of a genome region (in different regions of a chromosome) (Fig. 5 B and C).

The analysis of PP of Beijing *M. tuberculosis* genomes revealed long deletions specific to this lineage. Because of those deletions, the orthologs of Rv0071, Rv0072, Rv0073 and Rv1761c, Rv1760, Rv1758 (identifiers correspond to the genes in the *M. tuberculosis* H37Rv genome; see Table 2) now have similar phylogenetic profiles (partner profiles, Fig. 5B). It should be noted that the chromosomal region harboring

genes Rv1761c, Rv1760 and Rv1758 is flanked with the inverted repeats of IS6110 IS-elements belonging to the IS3 family. In the second round of the analysis, a long insertion was revealed specific to the Beijing lineage; because of that insertion the orthologs of CFBS_RS10335, CFBS_RS10345, CFBS_RS10350, CFBS_RS10355, CFBS_RS10360, CFBS_RS10365, and CFBS_RS21395 (identifiers correspond to the genes in the *M. tuberculosis* CDC5079 genome) (Table 2) now have phylogenetic profiles very much resembling inverted profiles (compensator profiles, Fig. 5 C).

To sum up, our PP analysis has identified a number of genes that have undergone similar evolutionary events. The loss of *cas1*, *cas2*, *csm5*, and *csm6* in the Beijing lineage of *M. tuberculosis* was accompanied by the loss and acquisition of other genes (Table 2). Those candidate genes have a potential to participate in the mechanisms of compensation for *cas*-gene functions or be their functionally related partners in *M. tuberculosis*, creating a subject for further research.

CONCLUSION

The CRISPR-Cas systems of *M. tuberculosis* vary considerably between the lineages: some (EAI) have long arrays, others (Beijing) are partially reduced. Therefore, the presence of an active type III-A CRISPR-Cas system is not an essential prerequisite for the evolutionary success in terms of pathogenicity, virulence, transmissibility and adaptability of the lineage.

The partial loss of the array and a few *cas*-genes in the Beijing lineage of *M. tuberculosis* seems to have resulted in a fully or partially lost ability of their CRISPR-Cas to destroy a foreign DNA. Disturbances in the functioning of the CRISPR-Cas system in one of the most successful *M. tuberculosis* lineages may have been accompanied by the activation of mechanisms compensating for the lost genes (for example, our analysis revealed a long insertion in the Beijing lineage) and by the loss of the functional partners of *cas*-genes; because it is assumed that the gene that has lost its functional partner will not be retained through selection in the genome and will be eliminated, which can be illustrated by the detected long deletions specific to the Beijing lineage. It should be noted that the observed regularities in the pattern of evolutionary losses and acquisitions could be random and require further experimental verification. Phylogenetic profiling provides a basis for generating a hypothesis and material for further research.

Although the CRISPR-Cas system of *M. tuberculosis* Beijing strains can be inactive, it is assumed that in the lineages that have a full set of *cas*-genes and repeats, the CRISPR-Cas system retains its activity and is capable of contributing to the defense against foreign DNA [34]. Considering a short leader sequence typical for all CRISPR-Cas systems of *M. tuberculosis*, it may be more productive to focus on the exploration of their alternative functions, such as regulation of gene expression, DNA repair, virulence formation, etc.

The structure of *M. tuberculosis* CRISPR-Cas systems has been studied and described in great detail [8], but the

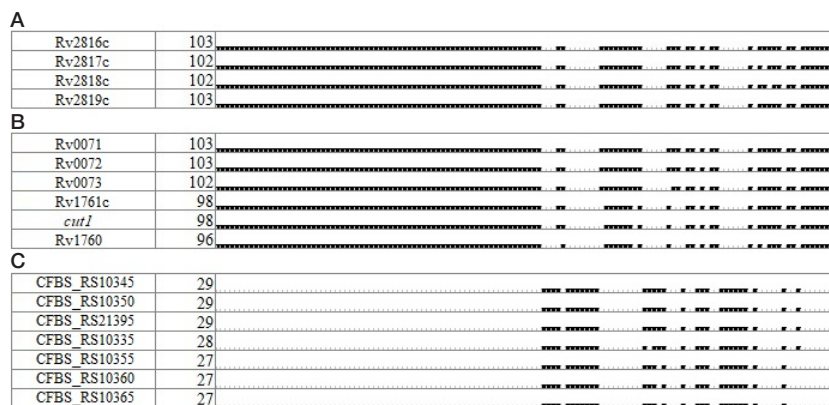


Fig. 5. Phylogenetic profiles of genes found in the genomes of different *M. tuberculosis* lineages and strains. Each line contains gene ID, the number of genomes with an orthologous gene and the PP of the gene; spaces represent missing genes. **A.** PP of *cas*-genes of *M. tuberculosis* H37Rv: *cas2* (Rv2816c), *cas1* (Rv2817c), *csm6* (Rv2818c), and *csm5* (Rv2819c). **B.** PP of candidate functionally related partners of *M. tuberculosis cas*-genes in the genome of the H37Rv strain. **C.** PP of genes putatively involved in the formation of compensatory mechanisms in the genome of the CDC5079 strain (the Beijing lineage) following the loss of some CRISPR-Cas parts

Table 2. Characteristics of functional partner candidates and genes involved in the compensatory mechanisms for CRISPR-Cas systems*

Evolutionary event	Gene ID	Product (protein)	Protein function
Insertion	CFBS_RS10360 (<i>tuſ</i>)	Iron-regulated elongation factor Tu Tuf	Participates in protein translation
	CFBS_RS10365	MDR-dehydrogenase	MDR-dehydrogenases exert different activities, such as alcohol dehydrogenase activity, quinone reductase activity, sorbitol dehydrogenase activity, formaldehyde dehydrogenase activity, keto reductase activity, and others
Deletion	Rv1758 (<i>cut1</i>)	Cutinase	Enzyme that catalyzes cutin hydrolysis [48]
	Rv1760	Diacylglycerol acyltransferases	Catalyzes final stages of biosynthesis of triacylglycerol and other components of mycobacterial cell wall [49]
	Rv0072	Glutamine ABC transporter permease	ABC transporter, participates in glutamine transport (permease); may be responsible for substrate translocation across the membrane [50]
	Rv0073	Glutamine ABC transporter ATP-binding protein	ABC transporter, participates in glutamine transport (the binding protein); may be responsible for energy coupling to the transport system [50]

Note: * — only annotated genes are presented in this table; other genes code for hypothetical proteins with unknown functions (Rv1761c and CFBS_RS10335, CFBS_RS10345, CFBS_RS10350, CFBS_RS10355, CFBS_RS21395) or, as with Rv0071, are a mobile self-splicing retro-element, the so-called group II intron.

role of these systems is still unclear. For the majority of mycobacterial spacers, no protospacers among the known mycobacteriophage have been identified so far. This is probably because the majority of *M. tuberculosis* bacteriophages have been isolated from *M. smegmatis* and then tested for their ability to invade *M. tuberculosis*. [34]. Different lineages have different sets of spacers and therefore – possibly – different immunity, which means varying degrees of resistance to phages and different regulation of gene expression. Apart from spacers, of interest is the role of the Cas2 protein outside

the Cas1-Cas2 complex, in light of the previously proposed hypothesis [15] suggesting that with independent activation this protein can stop translation and drive the cell into the dormant state or promote apoptosis. A search for Cas2 inhibitors also presents a certain interest. Particular attention should be paid to the possible functional link between CRISPR-Cas systems, specifically the *cas*-genes, and TA-systems [38]. Possible participation of CRISPR-Cas systems in virulence formation and drug resistance may allow to develop novel approaches to combating drug-resistant strains of *M. tuberculosis*.

References

- 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. *J Clin Microbiol.* 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. *Journal of Bacteriology.* 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: 254–257.
- 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.
- 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.
- Prozorov AA, Danilenko VN. *Mycobacterii tuberculoznogo kompleksa: genomika, molekularnaja epidemiologija, puti evoluzii. Uspehi sovremennoy biologii.* 2011; 1 (6): 483–95.
- 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.
- Prozorov AA, Fedorova IA, Bekker OB, Danilenko VN. Faktory virulentnosti *Mycobacterium tuberculosis*: geneticheskiy kontrol, novie konceptzii. *Genetika.* 2014; 50 (8): 775–97.
- 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.
32. Mokrousov I. The quiet and controversial: Ural family of *Mycobacterium tuberculosis*. *Infect Genet Evol.* 2012; 12 (4): 619–29.
 33. 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.
 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, Kypides 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.

Литература

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. *Journal of Clinical Microbiology.* 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 *Mycococcus xanthus* Development and CRISPR-Associated Genes and Repeats. *J Bacteriol.* 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: 254–7.
 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. Прозоров А. А., Даниленко В. Н. Микобактерии туберкулезного комплекса: геномика, молекулярная эпидемиология, пути эволюции. *Успехи современной биологии*. 2011; 131 (3): 227–43.
 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. 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.
 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-Béguet 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: 52–7.
 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.