We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter 11

Mosaic Structure as the Main Feature of *Mycobacterium bovis* BCG Genomes

Voronina Olga Lvovna,
Aksenova Ekaterina Ivanovna,
Kunda Marina Sergeevna,
Ryzhova Natalia Nikolaevna,
Semenov Andrey Nikolaevich,
Sharapova Natalia Eugenievna and
Gintsburg Alexandr Leonidovich

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.75005

Abstract

Background: The genome stability of attenuated live BCG vaccine preventing the acute forms of childhood tuberculosis is an important aspect of vaccine production. The purpose of our study was a whole genome comparative analysis of BCG sub-strains and identification of potential triggers of sub-strains' transition.

Results: Genomes of three BCG Russia seed lots (1963, 1982, 2006 years) have been sequenced, and the stability of vaccine sub-strain genomes has been confirmed. A comparative genome analysis of nine *Mycobacterium bovis* BCG and three *M. bovis* strains revealed their specific genome features associated with prophage profiles. A number of prophage-coded homologs to Caudovirales ORFs were common to all BCG genomes. Prophage profiles of BCG Tice and BCG Montreal genomes were unique and coded homologs to herpes viruses ORFs. The data of phylogenetic analysis of BCG sub-strain groups based on whole genome sequences and genome restriction maps were in congruence with prophage profiles. The only fragmentary similarity of specific prophage sequences of BCG Tice, BCG Montreal, and BCG Russia 368 in pair-wise alignments was observed, suggesting the impact of prophages on mosaic structure of genomes.

Conclusions: The whole genome sequencing approach is essential for genomes with mosaic structure, harboring numerous prophage sequences. Tools for prophage search are effective instruments in this analysis.

Keywords: BCG sub-strains, *Mycobacterium bovis*, genome stability, genome rearrangements, prophages

IntechOpen

© 2018 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

1. Introduction

Tuberculosis (TB) is one of the top causes of death in the world. Currently, the only authorized vaccine for primary vaccination of children from TB remains BCG, first applied in 1921. It is broadly used in different countries as part of the national childhood immunization program. Despite the attempts of TB control through widespread introduction of vaccination it was estimated in 2014 worldwide that 9.6 million people have fallen ill with TB. Nevertheless, vaccination against TB reduced TB prevalence by 42% in 2015 compared to that in 1990 [1].

World Health Organization (WHO) controls BCG vaccine, and the WHO Expert Committee on Biological Standardization (ECBS) has developed the international requirements for the manufacture and control of BCG vaccine. In 2009, for BCG vaccines of three different sub-strains (Danish1331, Tokyo 172-1 and Russian BCG-I), WHO Reference Reagents were established by WHO ECBS. In addition, quality control requirements comprising molecular genetic characterization of final lots and working seeds of BCG vaccines were suggested [2]. Russian research laboratories performed whole genome sequencing (WGS) of BCG Russia sub-strain genome as WHO and good manufacturing practice (GMP) recommended [3–5]. Currently, ten whole genome sequences of BCG sub-strains including BCG Russia are available in GenBank. It should be noted that since the 1920s, cultivation of the original strain BCG resulted in the emergence of numerous sub-strains that have evolved from it. So, now we could investigate the evolution of BCG sub-strains and the endpoints of this evolution could be assessed likewise in the study of Darwinian biological species evolution [6]. The reason for BCG sub-strains' transition remains unclear because the progenitor of BCG strains was lost. The comparative analyses of genome features of different BCG sub-strains can help in solving this problem.

The attention was focused on mobile elements of BCG sub-strain genomes especially on prophage sequences because of their contribution to the bacterial genome patterning. Following Brüssow et al. [7], 12 years later, we can reaffirm that there is a renaissance of phage research because now we have a lot of information about bacterial and phage genomes in the international databases. It was noticed that reintroduction of the fitness factor by phages usually influences the pathogenic factors of bacteria cells [8]. Thus, phages are of great importance for bacterial short-term adaptation and our goal was to estimate a potential contribution of prophage sequences on the mosaic structure of vaccine BCG sub-strain formation.

2. BCG genome sequencing

M. bovis AF2122/97 (Accession Number NC_002945) was the first *M. bovis* strain, where complete genomic sequence was determined [9]. The first BCG genomic sequence was performed for BCG Pasteur 1173P2 (NC_008769.1) [10]. The sequences of these strains such as Tokyo 172 (NC_012207.1) and Moreau RDJ (NZ_AM412059.1) were generated from the small-*insert libraries (*1–4 kb) by using BigDye terminator chemistry on ABI377- or ABI3700-automated

DNA sequencers [9–12]. The large-*insert library (40 bp) preparation method was used for* Mexico BCG sub-strain (NC_016804.1) sequencing [13]. Orduña et al. [13] were first who used for BCG genome analysis – the next generation sequencing (NGS) based on 454 technology and Sanger method. Late shotgun DNA libraries for strain sequencing were performed by commercial NGS kits. The strains were sequenced by the combination of 454 and Illumina platforms and the Sanger method, for example, BCG 3281 (NZ_CP008744.1) isolated from the human patient with TB and Korea 1168P (NC_020245.2) [14]. Single-molecule real-time sequencing (SMRT) based on PacBio Systems was allowed to directly sequence DNA and achieve long sequencing reads (>10,000 bp) with uniform coverage [15]. SMRT in combination with Illumina was used for *M. bovis* 1595 (NZ_CP012095.1) complete genomic sequence determination [16]. The further development of SMRT technology was allowed to use this platform alone for complete genomic sequencing. Sequences of BCG sub-strain 26/ATCC 35735/Montreal (CP010331.1) and *M. bovis* 30 (CP010332.1) could be considered as the example of this approach [17]. Whole genome sequence of BCG Russia sub-strain was performed using 454 and Sanger technology for short-gun and paired-end libraries. Whole genome map (WGM) creation was useful for the control of the repeat regions.

3. Comparative genome analyses as proof of BCG Russia genome stability

In the vaccine manufacture, one of the important features of BCG sub-strain is the genome stability. So, BCG vaccines' quality control and production now include characterization of BCG sub-strain genome. The importance of molecular genetic characterization is confirmed by the WHO requirements. According to these requirements, WGS of the last seed lot of BCG Russia (BCG Russia 368, 2006 year) was performed. Besides, two BCG Russia sub-strains from seed lots of 1963 and 1982 years (BCG Russia 311 and BCG Russia 977) were analyzed on the basis of WGS.

Comparative analyses of three BCG Russia sub-strains from different seed lots revealed only two differences. The first difference was the *single-nucleotide polymorphism* (SNP) in the position 3,175,301 (numeration according to reference strain BCG Tokyo) in the sub-strain BCG Russia 368. This SNP leads to the synonymous mutation in the uridylyltransferase gene. In the generation of 1963 and 1982, this mutation was not registered.

The second change in genome-affected glycerol-3-phosphate acyltransferase gene is shown in **Figure 1**. The mutation that occurred in this gene in the position 2,744,580 (an insertion of TGT bases instead of C base) truncated the protein. Nevertheless, the mutation was not concerned with the conservative domain of glycerol-3-phosphate acyltransferase and the protein could be functional. It should be noted that not all reads had the insertion of TGT. The changes were registered only for 14% reads of BCG Russia 311 and 54% reads of BCG Russia 977. In the last BCG Russia 368 generation, this mutation wasn't found. So, the genome structures of three different BCG Russia seed lots remain stable. The last BCG Russia 368 generation that is discussed in the text later was deposited in GenBank with the Accession Number NZ_CP009243.1.



Figure 1. Comparison of different glycerol-3-phosphate acyltransferase variants in the three generations of BCG Russia. (A) Alignment of whole and truncated variants of glycerol-3-phosphate acyltransferase. (B) Fragment of glycerol-3-phosphate acyltransferase with mutation. Hash-amino acids residues in conservative domain of glycerol-3-phosphate acyltransferase important for the enzymatic activity.

4. *In silico* genotyping of BCG-Russia sub-strain

Genomic feature of BCG Russia, as well as all BCG sub-strains, is a large deletion of the 10-kb genomic region of difference 1 (RD1) [18].

Spoligotyping profile is the second known characteristic of BCG sub-strain. This method is based on detection of Direct Repeats (DR) on the right and left sides of IS6110. DR loci are members of a universal family of sequences, designated as clustered regularly interspaced short palindromic repeats (CRISPR) sequence family.

Spoligotype profile of BCG Russia was typical for *M. bovis* with the absence of spacers 3, 9, 16, and 39–43, the *in silico* pattern corresponded to spoligo-international-type number (SIT) 482 according to the SPOLDB4 Database [19, 20].

The Mycobacterial interspersed repetitive unit (MIRU) profile of BCG Russia sub-strain based on *12 MIRU loci was* 232,324,253,222 according to *in silico* genome analysis.

The whole data of BCG Russia MIRU-variable-number tandem repeats (VNTR) loci are summarized in **Table 1**. The repeat unit size (bp) and repeat number are indicated in the brackets,

Locus	Copy number of locus, bp*	Locus	Copy number of locus, bp
MIRU_2	(53 × 2) + 8	MIRU_23	(53 × 4) + 5
Mtub04	(51 × 0) + 30	MIRU_24	(54 × 2) + 30
ETR_C	(58 × 4) + 37	MIRU_26	(51 × 5) + 13
MIRU_4/ETR_D	(77 × 3) + 4	MIRU_27/QUB-5	(53 × 3) + 25
MIRU_40	(54 × 2) + 19	Mtub34	(54 × 2) + 51
MIRU_10	(53 × 1) + 51	MIRU_31/ETR_E	(53 × 2) + 3
MIRU_16	(53 × 3) + 18	Mtub39	(58 × 2)
Mtub21	(57 × 0) + 34	QUB-26	(111 × 4) + 24
MIRU_20	(77 × 2) + 11	QUB-4156	(32 + (59 × 0) + 19)
QUB-11b	(69 × 3) + 10	MIRU_39	(53 × 1) + 29
ETR_A	(75 × 5) + 20	QUB-3232	(56 × 5) + 48
Mtub29	(13 + (57 × 1) + 35)	VNTR-3820	(59 × 5) + 47
Mtub30	(58 × 1) + 53	VNTR-4120	(57 × 2) + 23
ETR_B	(57 × 5) + 8		

*(Repeat unit size (bp) × Repeat number) + Partial repeat size (bp).

Table 1. The copy number of MIRU-VNTR loci in BCG Russia genome.

partial repeat size (bp) is outside. One discrepancy in polymerase chain reaction (PCR) and *in silico* results was revealed. The number of repeat units in MIRU_4/ETR_D was three in accordance with MIRU-VNTR analysis of BCG Russia genome data obtained in three different laboratories (NZ_CP009243.1, CP011455.1, CP013741.1). The number of repeat units in MIRU_4/ETR_D of BCG Tokyo was identical. However, PCR analyzes of MIRU-VNTR performed by Supply et al. [21] for BCG Russia sub-strain and by Mokrousov et al. [22] for the strains isolated from BCGitis patients revealed only two repeat units in MIRU_4/ETR_D. The discrepancy in PCR and *in silico* analyses of MIRU-VNTR could be explained by difficulties in the amplification of high GC genomes. The similar discrepancy in PCR and *in silico* results was described by Iwamoto et al. [23] for *M. tuberculosis* H37Rv. The repeat number in Mtub39 locus was five in PCR analysis but two copies were revealed in *in silico* genome investigation.

5. Is the original BCG Russia sub-strain *recA* a mutant?

High degree of genomic stability of BCG Russia sub-strain is seen as an inexplicable fact by some scientists. One of the explanations of this fact proposed by Keller et al. is in the highly cited paper [24]. They postulated *recA gene inactivation* in BCG Russia sub-strain. RecA is a multifunctional and ubiquitous recombinase protein involved both in general recombination and in DNA repair. RecA-dependent recombination mediates genetic rearrangements

resulting in increased genetic instability, while RecA-mediated DNA repair mechanisms have been shown to be essential for intracellular survival and persistence [25].

Among the mechanisms of bacterial evolution, the leading role belongs to recombination events. The large-scale rearrangements, deletions and duplications were revealed during comparative genomics analyses in *M. leprae* [26], *M. tuberculosis* [27] and *M. bovis* BCG [10]. Gene duplication has led to the origin of the half tubercle bacillus proteins [28]. Tandem duplications and homologous recombination also make a significant contribution to the diversity of mycobacteria. As an example, recombination between adjacent repeats of IS6110 elements resulted in deletions of several genome regions in *M. tuberculosis* H37Rv [27], https//www.ncbi.nlm.nih.gov/pmc/articles/PMC2483709/ - B33 [29].

Keller et al. detected the single-nucleotide insertion of "C" at the 5′ end of the *recA* gene of BCG Russia sub-strain received from TD Allergen [24]. As a result, the stop codon was formed and recombinase A synthesis was absent. These data have not been confirmed by the genome analysis of the original BCG Russia sub-strain. Whole genome sequence (NZ_CP009243), obtained in our laboratory, and other sequences (CP011455.1, CP013741.1), did not have the single nucleotide insertion in the *recA* gene. So the complete reading frame for recombinase A was annotated. Thus, the original BCG Russia sub-strain is not a *recA* mutant and the stability of the BCG Russia genome cannot be associated with *recA inactivation.*

Keller et al. findings may indicate that the sub-strain used by authors was not original or has been changed during cultivation. The last one is possible. According to our data whole and truncated variants of glycerol-3-phosphate acyltransferase gene was identified in one of the BCG Russia generations (**Figure 1**).

6. The "early" sub-strain genomes comparison

Genome sequences were compared using BCG sub-strain Tokyo 172 genome, a member of the "early" sub-strains group as reference. First, among this group, BCG Tokyo sub-strain is closest to BCG Russia sub-strain as regards the time of its provision by the Pasteur Institute to Tokyo (in 1925). Second, it was lyophilized in the 1940s and used later as a freeze-dried vaccine, as BCG Russia sub-strain. Then, in 1960, the 172nd transfer on bile-potato medium was freeze dried and adopted as a primary seed lot [30]. Finally, one of the first BCG genomes that were accurately sequenced, assembled, and submitted to GenBank was the genome of this seed lot [11].

We observed no significant diversity in the sequences of the BCG Russia 368 and BCG Tokyo 172 genomes. The revealed genomic differences were summarized in **Table 2** and could be subdivided into three groups: region of differences (RDs), ins/del and SNP. Only two RDs were detected between the "early" sub-strains. First, a 22 bp insertion was found in the TetR family transcriptional regulator gene of BCG Russia 368 genome. One variant of Japan BCG vaccine (Type I), submitted in GenBank, included this deletion (RD16). The RD16 band identical to those of other BCG sub-strains was found in the Type II strain [31]. A 1602 bp deletion in BCG Russia

Type of differences	Number of differences	
	BCG Russia368/BCG Tokyo 172	**BCG Tokyo 172/BCGPasteur1173P2 [11]**
Region of Differences (more than 20 bp)	2	20
Insertions/deletions <20 bp (1–9 bp)	10	20
SNP in total	52	68
intergenic SNP	11	
synonymous SNP	8	
nonsynonymous SNP (without nonsense)	31	
Nonsense SNP as variant of nonsynonymous	2	

Table 2. Genomic differences of BCG Russia 368, Tokyo 172, and Pasteur 1173P2 sub-strains.

368 genome was the second RD, corresponding to the region from 4,110,452 to 4,112,053 bp in BCG Tokyo 172, beginning in JTY_RS19265 (ribonuclease gene), including JTY_RS19270 (antitoxin VapB48 gene) and finishing inside JTY_RS19275 (glutamate-cysteine ligase gene).

The sub-strains used for vaccine production in Bulgaria (BCG Sofia) and India were obtained from BCG Russia. Nowadays, UNICEF uses four variants of BCG vaccine on behalf of the Global Alliance for Vaccines and immunization. The Statens Serum Institute in Denmark produces BCG-Denmark; Bulbio (BBNCIPD) in Bulgaria; and the Serum Institute in India produces BCG-Russia (genetically identical to BCG-Bulgaria) and the Japan BCG laboratory produces BCG-Japan [32].

We could trace the genome characteristics of BCG Russia daughter sub-strains using published data. Stefanova et al. analyzed the BCG sub-strain used for production in Bulgaria (named Sofia SL222) with *M. tuberculosis* microarrays. They detected a 1.6-kb deletion that affects Rv3697c and Rv3698 homologs. The deletion of this region was also noted in BCG Russia but not in any other strains [33]. The authors concluded that RD 1602 bp is an old deletion, because BCG Pasteur was replaced with BCG Russia in Bulgaria BCG laboratory in the 1950s.

According to Seki M. et al. differences between the "early" sub-strain Tokyo and the "late" sub-strain Pasteur were more significant and the number of RD increased tenfold [11].

Less ins/del differences were found between BCG Russia and BCG Tokyo genomes, then between BCG Tokyo and BCG Pasteur genomes. The size of ins/del differences was small: only 1–9 bp.

However, the number of SNPs was nearly the same in the two pairs of the genomes. Nonsynonymous SNP in BCG Russia 368 amounted to 60%, but most of them were associated with conservative substitutions in the proteins. Only seven proteins had radical substitutions, though three of them were from the PE-PGRS/PPE family. This finding has emphasized the significance of these proteins for BCG sub-strain adaptation.

7. A whole genome restriction map analysis

The large array of published literature accentuated the important role of RD in BCG sub-strains differentiation. For checking these statements the methods by OpGen Incorporated Company was used. So, first of all, the assembly of DU2 region and the number of tandem duplications in this region in BCG Russia 368 genome were performed by the Argus™ Optical Mapping System. WGM of the sub-strain BCG Russia 368 was created by the laboratory of OpGen Incorporated Company (Maryland, the USA), according to the Argus™ Optical Mapping System user manual [34]. DNA was digested with *NheI*. Map Solver software version 3.2 was employed for creating the final circular WGM; the whole genome map of BCG Russia 368 is represented in **Figure 2**. The separate comparison of DU2 regions (**Figure 3**) has shown that genomes BCG Russia 368 and BCG Tokyo 172 are identical in this region, unlike from the BCG Pasteur optical map, which can be confirmed by presence of three copies (triple tandem duplications) in the DU2 region of BCG Russia 368.

The cluster construction based on map similarity of the six references of BCG sub-strains is shown in **Figure 4**. As you can see, the cluster was split into two groups: BCG Tice (ATCC 35743) was attributed to the group of the "early", while BCG Mexico to the "late" group of sub-strains in accordance with the *NheI* restriction fragments.

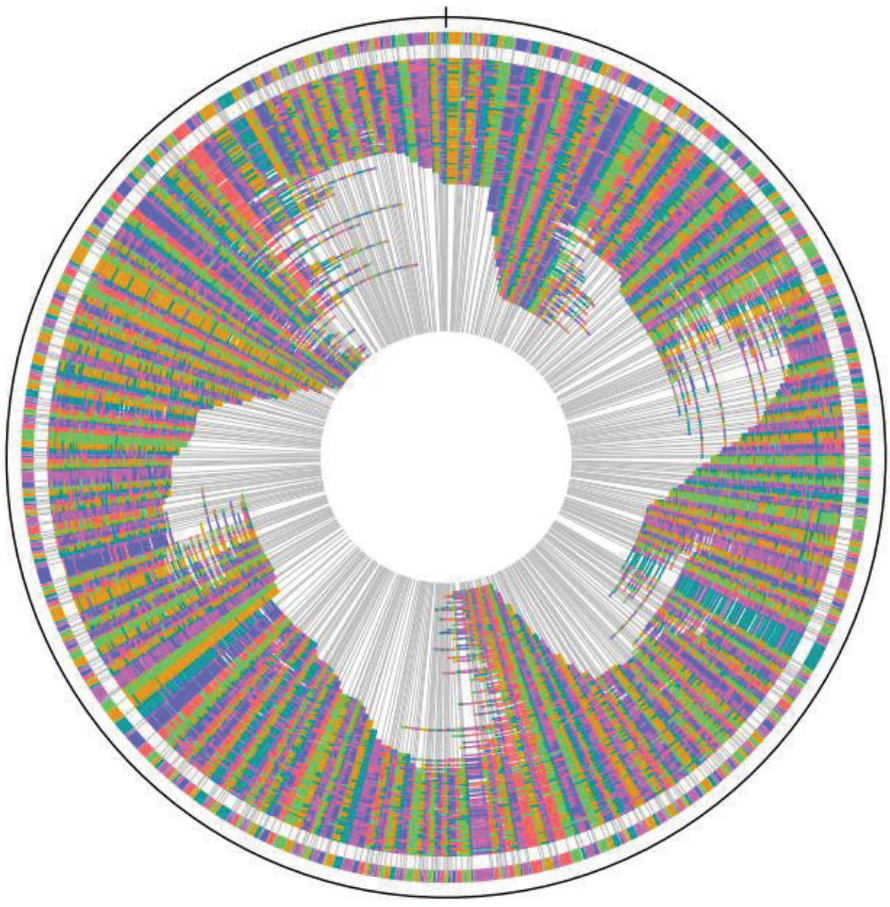


Figure 2. The circular restriction map of BCG Russia 368 whole genome. The restriction map was obtained by DNA digestion with *NheI*.

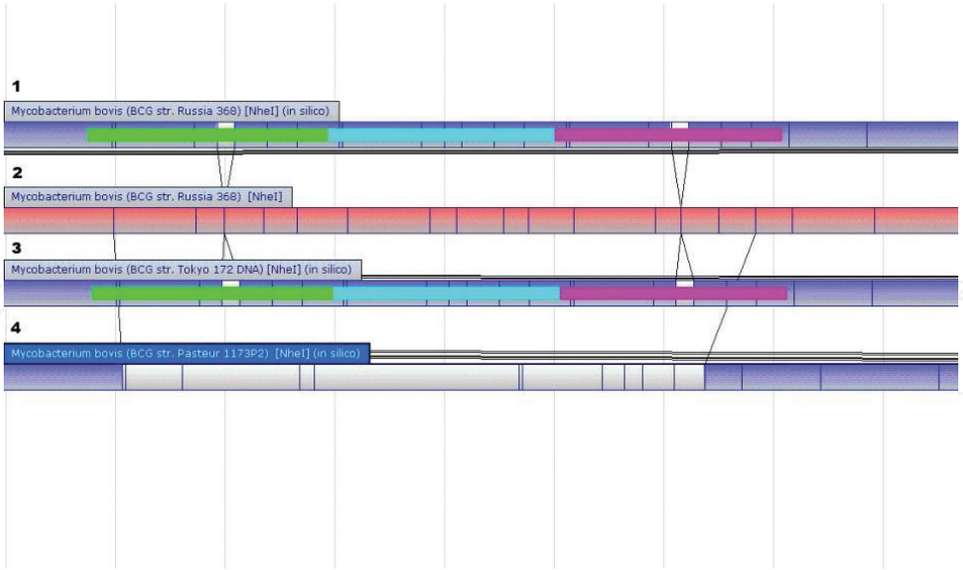


Figure 3. Aligned OpGen maps for BCG Russia 368 and reference BCG sub-strains created for DU2 region. (1) OpGen map created *in silico* for BCG Russia 368 genome fragment. (2) OpGen map of BCG Russia 368 whole genome digestion with *NheI in vitro*. (3 and 4) OpGen map created *in silico* for BCG Tokyo 172 and BCG Pasteur 1173P2 genome fragments. All OpGen maps were created by DNA digestion with *NheI*. Vertical lines are pointed out at restriction sites. Tree copies of the DU2 region in BCG Russia 368 and BCG Tokyo 172 genomes are marked as green, blue and purple bars. The genome region from the *astB* to the *sdhD* genes (DU2 region) is represented as the green bar. The *astB* gene in the second and the third copies of DU2 region (blue and purple bars) was truncated.

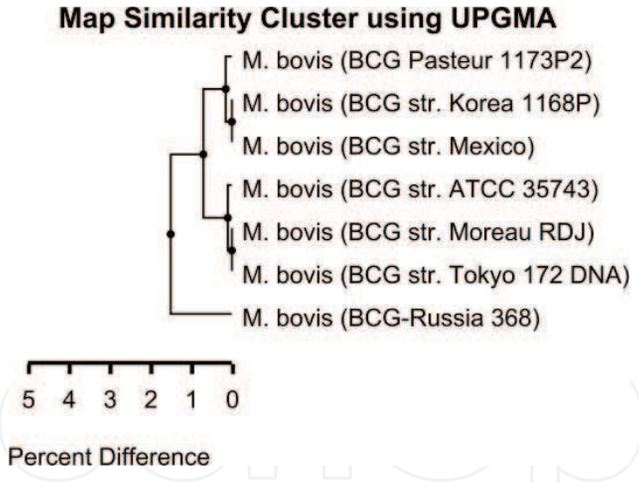


Figure 4. Map similarity cluster reconstruction for seven BCG sub-strain. The optical restriction maps of BCG Russia 368 and six reference BCG sub-strains obtained *in silico* were used for map similarity cluster reconstruction. The cluster construction was carried out using UPGMA (unweighted pair-group method using arithmetic averages) algorithm in OpGen MapSolver v.3.2.0. Program.

8. Genome map construction and the analyses of repetitive elements of BCG Russia 368

The whole genome gap-less BCG Russia 368 chromosome after the verification of the number of DU2 repeats was visualized in GeneWiz [35] (**Figure 5**). Genome atlas option of GeneWiz primarily GC Skew was selected as an appropriate instrument for verifying the accuracy of

genome assemblies and OriC detection. The place of the change in GC Skew agreed with the OriC and the first nucleotide position in the BCG Russia 368 genome. Other DNA properties, intrinsic curvature, stacking energy, position preference, global direct repeats, global inverted repeats, and AT-content, were essential for genome structure description.

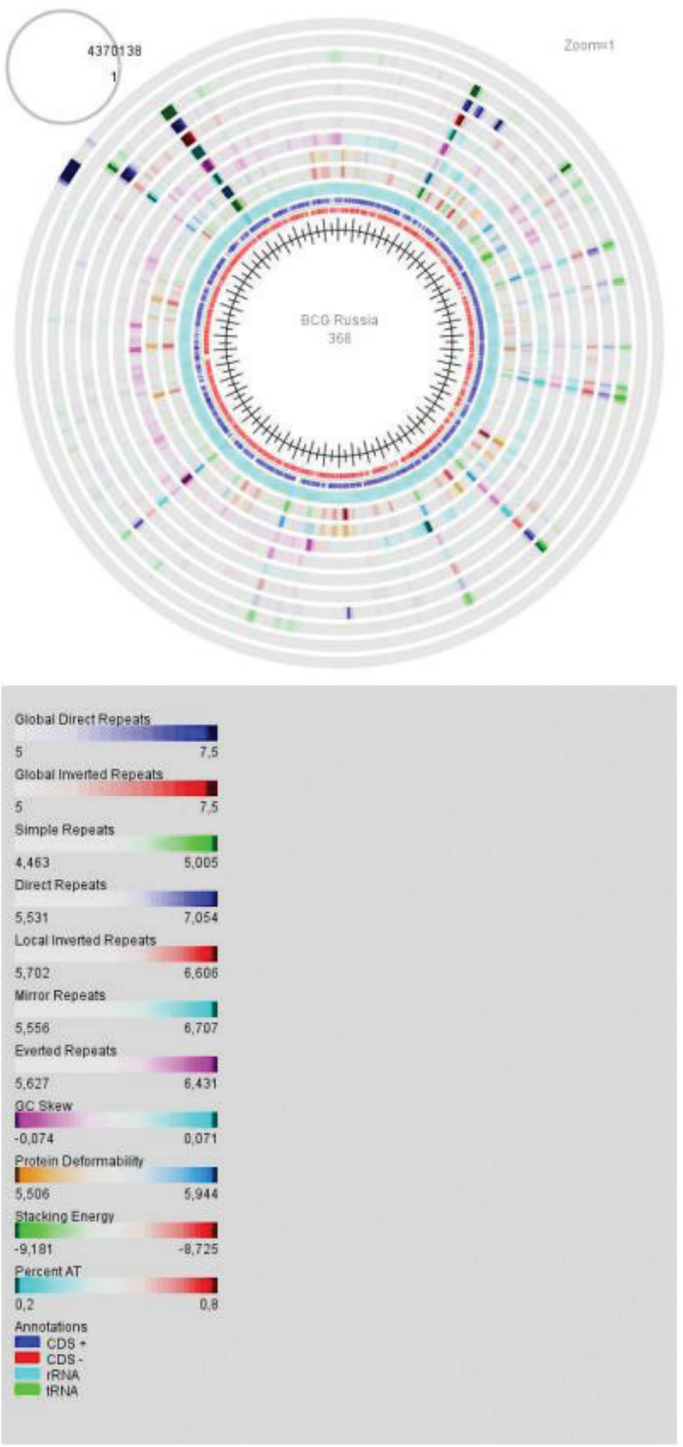


Figure 5. *M. bovis* BCG Russia 368 genome map.

Different types of repeats visualized by GeneWiz and shown in **Figure 5** correlated with the specific genome elements identified with the specific resources (**Table 3**).

Name	Number
REP	6
CRISPR	5
Prophage	2
PPE protein gene	66
PE protein gene	33
PE_PGRS protein gene	69
IS elements	41

Table 3. Specific genome elements in BCG Russia genome.

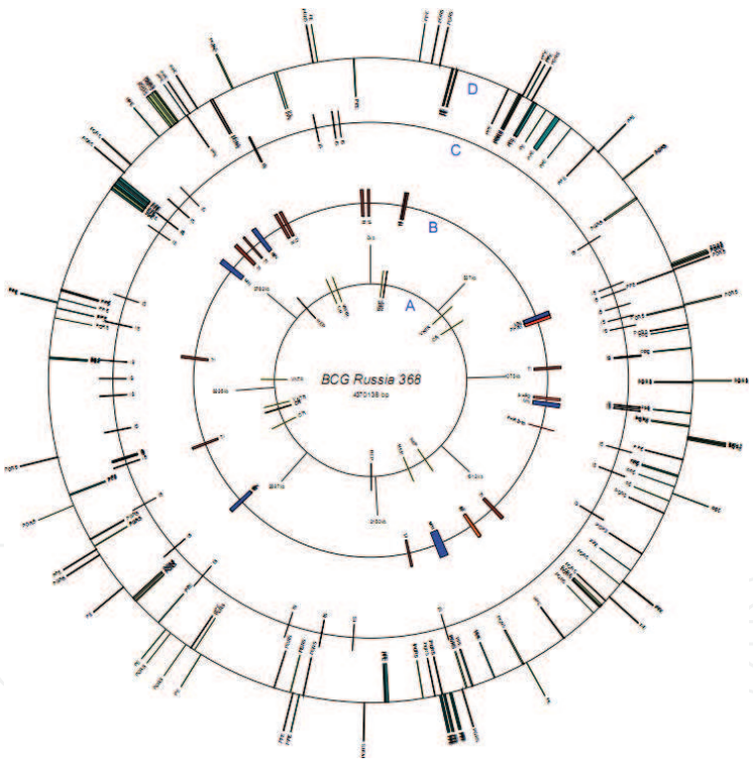


Figure 6. Localization of mobile elements (IS, repeats, prophage sequences), PE, PPE, and PE_PGRS genes in BCG Russia 368 genome. The circular map of BCG Russia 368 genome was visualized by the GenomeVx program. All of the prophage sequences were predicted by PHAST. Description of scheme: Repeats (REP, VNTR, and CRISPR elements) – circle A; phage sequences (according to PHAST) – B; IS elements – C; genes for PE, PPE, and PE_PGRS proteins – D. Accepted abbreviations: REP – Repetitive extragenic palindrome element; CR – CRISPR or possible CRISPR sequences predicted by CRISPRfinder; VNTR – Variable number tandem repeat; IS – Insertion sequence elements. Phages sequences: TI – BCG tice (CP003494.1); MN – BCG Montreal (CP010331.1); AF – *M. bovis* AF2122/97 (BX248333.1); PHR-2-rep – (922 bp repeat of 7.5 kb), PHR-1 – (11 kb), and PHR-2 – (7.5 kb) of BCG Russia. In the color code-ciphered phage sequences discovered in different *M. bovis* genomes: BCG tice (CP003494.1) – purple; BCG Montreal (CP010331.1) – blue; *M. bovis* AF2122/97 (BX248333.1) – orange; BCG Russia 368 (CP009243) – red.

The **insertion sequences (IS) elements and** affiliated resolvases, transposases, and integrases genes **were predicted by** ISfinder. Classification of IS elements and determination of inverted repeats flanking IS were made by the use of ISfinder database [36, 37]. Clustered regularly interspaced short palindromic repeats (CRISPR) and RISPR-associated Cas and Csm family proteins were predicted by CRISPRfinder [38, 39].

Locations of IS elements, repeats, prophage sequences, and PE, PPE, and PE_PGRS genes in BCG Russia 368 genome are visualized in **Figure 6**. Most of the repetitive elements of BCG Russia 368 genome, including some prophage sequences, are coinciding, overlapping, or interconnecting. So, it is hard to annotate some fragments of BCG Russia 368 genome. Special difficulties have arisen in the differentiation of bacterial and phage genes during PE/PGRS genes characterization.

9. The phages predicted in *M. bovis* genomes

Along with other mobile elements, the variability of predicted prophages may be the best indexes for characterization of mosaic structure genome. All of the prophages described in *M. bovis* genome were computed us by PHAge Search Tool (PHAST) [40, 41]. GenVision Plug-In of the DNASTAR Lasergene program package was selected for the visualization of prophage sequences. According to PHAST data, all of the predicted phages could be shared into three groups (see **Figure 7**). So, the first group composed of the common ones for *M. bovis* and *M. bovis* BCG prophages. A 7.5-kb prophage was revealed in most of the BCG genomes and in three *M. bovis*; exceptions include– BCG sub-strains Tice and Montreal. A 20.3-kb prophage in "early" sub-strain genomes (BCG Tokyo, Moreau, Russia) was replaced by an 11.2-kb one but was lost in "late" sub-strain genomes. The second group represented six BCG Montreal prophages, the third represented 15 BCG Tice sub-strain prophages. The prophages in the second and third groups were unique and did not coincide with the prophages of other sub-strains (see circle B in **Figure 6**). Like most of phage ORFs in the common ones, *M. bovis* prophages were annotated as genes belonging to the order Caudovirales (Myoviridae, Siphoviridae, and Podoviridae family), while most of phage ORF in BCG Tice or BCG Montreal prophages were similar to the genes of various Herpesviruses (Human, Bovine, Macaci, Alcela, Anguil).

The mosaic BCG genome structure has been verified by comparative prophage analyses. A partial similarity of BCG Tice/BCG Montreal prophage fragments has been identified after pair-wise alignment of BCG Montreal and BCG Tice phage sequences with BCG Russia 368 whole genome. The regions of similarity defined as the purple (BCG Tice) and blue (BCG Montreal) blocks on circle B are phage sequences discovered in different *M. bovis* genomes (**Figure 6**). In BCG Russia 368 genome, Tice-specific prophages (13.4 and 13.9 kb) were split into five and three parts, respectively. Fragments which are homologous to BCG Montreal-specific prophages represented in the genome of BCG Russia 368 as a sequence with multiple gaps ranging from 14 to 128 bp.

In turn, the 7.5-kb BCG Russia 368 prophage was split on 0.9- and 6.6-kb fragments located in different regions of BCG Tice/Montreal genomes. Moreover, these fragments lacked the transposase gene, which was specific to 7.5-kb BCG Russia 368 prophage.

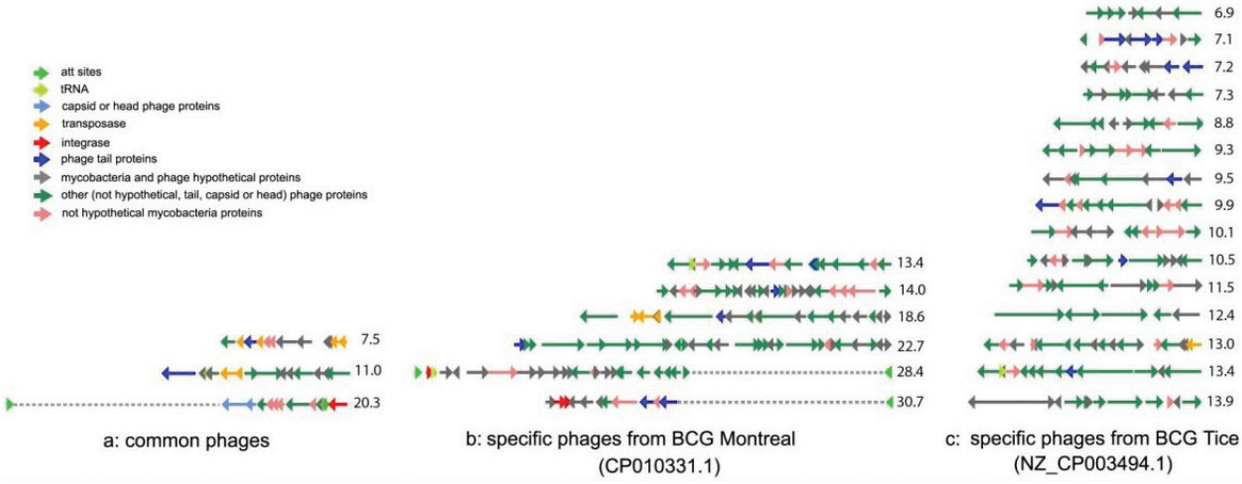


Figure 7. Comparative analyses of phage types from BCG Russia 368 and references *M. bovis* genomes. Color arrows indicate the location of phage genes. (a) Phages common for all analyzed genomes: 7.5, 11.0, and 20.3 kb – In all the analyzed strains (accession NZ_CP009243.1, NC_012207.1, NC_008769.1, NZ_CP003494.1, CP010331.1, NC_016804.1, NC_002945.3, NZ_AM412059.1, NZ_CP008744.1, NZ_CP012095.1, CP010332.1, NC_020245.2); (b) phages specific for BCG tice: 13.4, 14.0 18.6, 22.7, 28.4, and 30.7 kb; (c) phages specific for BCG Montreal: 6.9, 7.1, 7.2, 7.3, 8.8, 9.3, 9.5, 9.9, 10.1 10.5, 11.5, 12.4, 13.0, 13.4, and 13.9 kb.

Also, the 7.5-kb BCG Russia 368 prophages had 922-bp repeats in BCG Russia 368 genome (red bar in **Figure 6**, circle B) and was located near ISMt1 insertion element. Interestingly, besides prophage fragments, two intact prophages associated with the insertion elements have been predicted by PHAST in BCG Russia 368 genome. Thus, 11-kb and 7.5-kb prophage sequences were linked with IS6110 and IS1560 elements, respectively. So, the connection between prophage sequences and the IS elements has a considerable impact on the BCG genome evolution.

10. Phylogeny reconstruction

Phylogeny reconstruction was made using the genome sequences of analyzed *M. bovis* strains and BCG sub-strains. The full-genome comparison and phylogeny reconstruction were based on BLAST alignment and neighbor-joining algorithm [42] used in NCBI BLAST. The trees were represented by MEGA 6.0 [43]. Taking into account the prophage profile data, the congruence of obtained phylogenetic tree and vaccine sub-strains genealogy based on the DU2 region [10, 44] has been evaluated. *M. bovis* strains and BCG sub-strains formed different clusters on the tree (see **Figure 8**). As was expected, the "early" (Russia, Tokyo) and the "late" (Pasteur 1173P2, Korea 1168P, Mexico) BCG sub-strains formed separate but closely related groups. The basal branch in the BCG cluster was represented by BCG Moreau sub-strain. The unexpected position of the BCG Tice and the BCG Montreal has been revealed. The BCG Montreal showed some relationship with the "early" sub-strains. The BCG Tice has been placed in the most divergent basal position on the tree. In turn, BCG 3281, isolated from a pulmonary TB, patient had the relationship

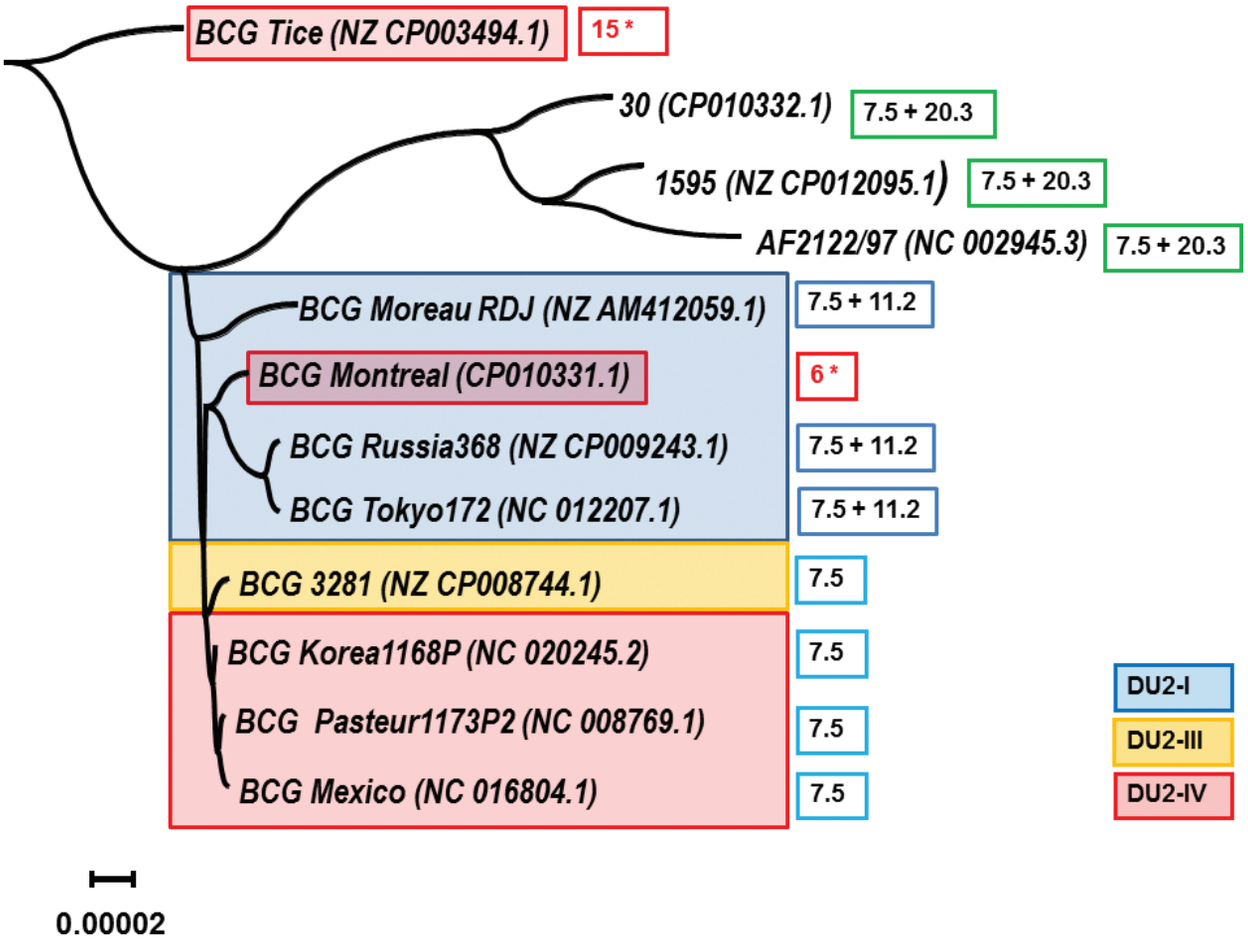


Figure 8. Whole genome phylogeny of *M. bovis* strains. Colored blocks describe the DU2 region type in BCG sub-strains: blue – DU2-I, yellow – DU2-III, red – DU2-IV. Unpainted blocks with colored borders and numbers inside define types of prophage profiles and prophage size in kb. Red borders with asterisks indicate the number of unique prophages identified in BCG tice and BCG Montreal genomes.

with the "late" sub-strains. Remarkably, all phylogenetic groups of the tree were characterized by specific sets of prophage sequences. The DU2 region genealogy only partially correlated with the whole genome phylogeny obtained in this study. If the DU2-I "early" sub-strain group showed common origin but the DU2-IV group split apart (see **Figure 8**). The BCG 3281 that represented DU2-III group [45] took an intermediate position. One of the possible explanations for this discrepancy may be numerous prophage-associated genome rearrangements.

11. Discussion

Numerous comparative genomics investigations of BCG sub-strains confirmed significant genomic polymorphism of BCG sub-strains which arose from one progenitor. RDs, indels and SNPs are real evidences still going on in the *in vitro* evolution of BCG sub-strains.

Here we supposed that genomic evolution and the BCG sub-strains diversity is a direct consequence of prophage-associated genome rearrangements. It is well known that 10–20% of bacterial genomes represented prophage sequences. Most of the prophages are damaged

and mutated. Nevertheless, recombinant events between homologous prophage sequences are possible. Moreover, some genes of defective prophages can be still working [46]. So, prophage genome content is an important biological driver/trigger of genomic rearrangements and evolution. Extensive contribution of prophages to bacterial fitness was supposed as a result of unexpected evolutionary prophage patterns [47]. Suggesting our assumption, outstanding differences between prophage profiles have been revealed in our comparative genome analysis of nine BCG sub-strains and three *M. bovis* strains. Big differences in the number and composition of prophages in the genomes of the late strains Tice and Montreal were discovered. According to the Brosch et al., both BCG Tice and BCG Montreal or Frappier were taken from the Pasteur Institute after 1934. They had close phylogenetic relations because they fall in one phylogenetic group, "DU2 IV, Δint" [10]. Dr. Rosenthal, who received the first Tice sub-strain from the Pasteur Institute, demonstrated heterogeneity of the "late" BCG sub-strain. It was a progenitor of at least six different daughter BCG sub-strains: H, K, E, L, LH, and BL. The sub-strain BL was strongly attenuated in laboratory studies. In 1952, BL was mixed with a new routine 'P' strain, received from the Pasteur Institute in 1951, in the ratio 3:1. This new sub-strain was called BLP. Since 1953 only freeze-dried BCG vaccine from this mixed strain has been produced [30]. The history of BCG Montreal sub-strain is also well known. Three times these BCG sub-strains were sent to Canada from the Pasteur Institute [30]. Significant changes of BCG genomes have been reflected in the appearance of new prophage profiles in BCG Tice and BCG Montreal sub-strains. It could also impact on vaccine properties of the sub-strains. According to Zhang et al. [48], BCG Tice, BCG Montreal/Frappier along with BCG Prague, and BCG Phipps sub-strains have lost the largest number of T-cell epitopes, defining its vaccine properties. In contrast, BCG Russia and BCG Tokyo sub-strains still have the largest number of T-cell epitopes among other BCGs. So an extended genomic sequencing is very important to identify prophages as potential markers of genomic rearrangement. Prophage studies could enhance our understanding of the genetic features of various BCG sub-strains and may also be useful for checking the genetic stability of the seed-lot sub-strain.

12. Conclusions

People migration from regions with a high incidence of TB and the growth of the number of HIV-infected individuals last decade resulted in the necessity of TB vaccination not only among children but also among adolescents and adults. In 2015, 15 vaccine candidates were considered in clinical trials. BCG vaccine replacement or vaccine boosting for the protection of adolescents and adults were considered. Recombinant BCGs, recombinant viral-vectored platforms, protein/adjuvant combinations, attenuated *M. tuberculosis* strains and mycobacterial extracts were included in the list [1]. A subunit vaccine developed in N.F. Gamaleya Research Center was based on the fusion of mycobacterial proteins with cellulose-binding domain [49]. On the other hand, new areas of BCG vaccine application have been proposed. As most humans are born in bacteriological environments characterized by a low microbial diversity, the effects of BCG vaccine administrated immediately after birth, as a modulator of Th-1/Th-2 responses, is very important and should be analyzed [50]. In this situation, the control of BCG genome stability is the important task, which will continue to be relevant.

Author details

Voronina Olga Lvovna*, Aksenova Ekaterina Ivanovna, Kunda Marina Sergeevna, Ryzhova Natalia Nikolaevna, Semenov Andrey Nikolaevich, Sharapova Natalia Eugenievna and Gintsburg Alexandr Leonidovich

*Address all correspondence to: olv550@gmail.com

N.F. Gamaleya National Research Center for Epidemiology and Microbiology, Ministry of Health, Russia

References

[1] Global Tuberculosis Report 2015. WHO. 20th ed. Available from: http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf

[2] Ho MM, Southern J, Kang H-N, Knezevic I. Meeting Report. WHO Informal Consultation on Standardization and Evaluation of BCG Vaccines. Geneva, Switzerland. 22-23 September, 2009; Available from: http://www.who.int/biologicals/publications/meetings/areas/vaccines/bcg/BCG_meeting_report_2009v7_FOR_WEB_10JUNE.pdf

[3] Kunda MS, Voronina OL, Aksenova EI, Semenov AN, Ruzhova NN, Lunin VG, Gintsburg AL. Analyzing of the BCG substrains diversity formed by the human influence (2014; p. 34). In: Troitsky A, Rusin L, Petrov N, editors. Molecular Phylogenetics: Contributions to the 4th Moscow International conference "Molecular Phylogenetics" (MolPy-4). Moscow: Torus Press; 2014. 90 p. ISSN: 978-5-94588-153-2

[4] Ludannyy R, Alvarez Figueroa M, Levi D, Markelov M, Dedkov V, Aleksandrova N, Shipulin G. Whole-Genome Sequence of *Mycobacterium bovis* BCG-1 (Russia). Genome Announcements. Nov 12, 2015;**3**(6). pii: e01320-15. DOI: 10.1128/genomeA.01320-15

[5] Sotnikova EA, Shitikov EA, Malakhova MV, Kostryukova ES, Ilina EN, Atrasheuskaya AV, Ignatyev GM, Vinokurova NV, Gorbachyov VY. Complete Genome Sequence of *Mycobacterium bovis* Strain BCG-1 (Russia). Genome Announcements. Mar 31, 2016;**4**(2). pii: e00182-16. DOI: 10.1128/genomeA.00182-16

[6] Arnoldt H, Strogatz SH, Timme M. Toward the Darwinian transition: Switching between distributed and speciated states in a simple model of earlylife. Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics. 2015;**92**(5):052909. DOI: 10.1103/PhysRevE.92.052909

[7] Brüssow H, Canchaya C, Hardt WD. Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. Microbiology and Molecular Biology Reviews. 2004;**68**(3):560-602. DOI: 10.1128/MMBR.68.3.560-602.2004

[8] Fortier LC, Sekulovic O. Importance of prophages to evolution and virulence of bacterial pathogens. Virulence. 2013;**4**(5):354-365. DOI: 10.4161/viru.24498

[9] Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, Duthoy S, Grondin S, Lacroix C, Monsempe C, Simon S, Harris B, Atkin R, Doggett J, Mayes R, Keating L, Wheeler PR, Parkhill J, Barrell BG, Cole ST, Gordon SV, Hewinson RG. The complete genome sequence of *Mycobacterium bovis*. Proceedings of the National Academy of Sciences of the United States of America. Jun 24, 2003;**100**(13):7877-7882

[10] Brosch R, Gordon SV, Garnier T, Eiglmeier K, Frigui W, Valenti P, Dos Santos S, Duthoy S, Lacroix C, Garcia-Pelayo C, Inwald JK, Golby P, Garcia JN, Hewinson RG, Behr MA, Quail MA, Churcher C, Barrell BG, Parkhill J, Cole ST. Genome plasticity of BCG and impact on vaccine efficacy. Proceedings of the National Academy of Sciences of the United States of America. Mar 27, 2007;**104**(13):5596-5601

[11] Seki M, Honda I, Fujita I, Yano I, Yamamoto S, Koyama A. Whole genome sequence analysis of *Mycobacterium bovis* bacillus Calmette-Guérin (BCG) Tokyo 172: A comparative study of BCG vaccine substrains. Vaccine. 2009;**27**(11):1710-1716. DOI: 10.1016/j.vaccine.2009.01.034

[12] Gomes LH, Otto TD, Vasconcellos EA, Ferrao PM, Maia RM, Moreira AS, Ferreira MA, Castello-Branco LR, Degrave WM, Mendonça-Lima L. Genome sequence of Mycobacterium bovis BCG Moreau. The Brazilian vaccine strain against tuberculosis. Journal of Bacteriology. Oct 2011;**193**(19):5600-5601. DOI: 10.1128/JB.05827-11

[13] Orduña P, Cevallos MA, de León SP, Arvizu A, Hernández-González IL, Mendoza-Hernández G, López-Vidal Y. Genomic and proteomic analyses of *Mycobacterium bovis* BCG Mexico 1931 reveal a diverseimmunogenic repertoire against tuberculosis infection. BMC Genomics. Oct 8, 2011;**12**:493. DOI: 10.1186/1471-2164-12-493

[14] Joung SM, Jeon SJ, Lim YJ, Lim JS, Choi BS, Choi IY, Yu JH, Na KI, Cho EH, Shin SS, Park YK, Kim CK, Kim HJ, Ryoo SW. Complete genome sequence of Mycobacterium bovis BCG Korea, the Korean vaccine strain for substantial production. Genome Announcements. 2013 Mar 14;**1**(2):e0006913. DOI: 10.1128/genomeA.00069-13

[15] PacBio Systems. Available from: URL:http://www.pacb.com/products-and-services/pacbio-systems/

[16] Kim N, Jang Y, Kim JK, Ryoo S, Kwon KH, Kang SS, Byeon HS, Lee HS, Lim YH, Kim JM. Complete genome sequence of Mycobacterium bovis clinical strain 1595, isolated from the laryngopharyngeal lymph node of South Korean cattle. Genome Announcements. Oct 1, 2015;**3**(5). pii: e01124-15. DOI: 10.1128/genomeA.01124-15

[17] Zhu L, Zhong J, Jia X, Liu G, Kang Y, Dong M, Zhang X, Li Q, Yue L, Li C, Fu J, Xiao J, Yan J, Zhang B, Lei M, Chen S, Lv L, Zhu B, Huang H, Chen F. Precision methylome characterization of mycobacterium *tuberculosis* complex (MTBC) using PacBio single-molecule real-time (SMRT) technology. Nucleic Acids Research. Jan 29, 2016;**44**(2):730-743. DOI: 10.1093/nar/gkv1498

[18] Behr MA, Small PM. A historical and molecular phylogeny of BCG strains.Vaccine. 1999;**17**(7-8):915-922

[19] SPOLDB4 Database. Available from: http://www.pasteur-guadeloupe.fr/tb/bd_myco.html

[20] Brudey K, Driscoll JR, Rigouts L, et al. Mycobacterium tuberculosis complex genetic diversity: Mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiology. 2006;**6**:23

[21] Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, Locht C. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. Journal of Clinical Microbiology. 2001;**39**(10):3563-3571

[22] Mokrousov I, Vyazovaya A, Potapova Y, Vishnevsky B, Otten T, Narvskaya O. Mycobacterium bovis BCG-Russia clinical isolate with noncanonical spoligotyping profile. Journal of Clinical Microbiology. 2010;**48**(12):4686-4687. DOI: 10.1128/JCM.01368-10

[23] Iwamoto T, Yoshida S, Suzuki K, Tomita M, Fujiyama R, Tanaka N, Kawakami Y, Ito M. Hypervariable loci that enhance the discriminatory ability of newly proposed15-loci-and24-locivariable-number tandem repeat typing method on *Mycobacterium tuberculosis* strains predominated by the Beijing family. FEMS Microbiology Letters. 2007;**270**(1):67-74

[24] Keller PM, Böttger EC, Sander P. Tuberculosis vaccine strain *Mycobacterium bovis* BCG Russia is a natural recA mutant. BMC Microbiology. 2008 Jul 17;**8**:120. DOI: 10.1186/1471-2180-8-120

[25] Sander P, Papavinasasundaram KG, Dick T, Stavropoulos E, Ellrott K, Springer B, Colston MJ, Böttger EC. Mycobacterium bovis BCG recA deletion mutant shows increased susceptibility to DNA-damaging agents but wild-type survival in a mouse infection model. Infection and Immunity. Jun 2001;**69**(6):3562-3568

[26] Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honoré N, Garnier T, Churcher C, Harris D, Mungall K, Basham D, Brown D, Chillingworth T, Connor R, Davies RM, Devlin K, Duthoy S, Feltwell T, Fraser A, Hamlin N, Holroyd S, Hornsby T, Jagels K, Lacroix C, Maclean J, Moule S, Murphy L, Oliver K, Quail MA, Rajandream MA, Rutherford KM, Rutter S, Seeger K, Simon S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Taylor K, Whitehead S, Woodward JR, Barrell BG. Massive gene decay in the leprosy bacillus. Nature. Feb 22, 2001;**409**(6823):1007-1011

[27] Dubos RJ, Pierce CH, Schaefer WB. Differential characteristics in vitro and in vivo of several substrains of BCG. III. Multiplication and survival in vivo. American Review of Tuberculosis. 1956;**74**(5):683-698

[28] Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. Jun 11, 1998;**393**(6685):537-544

[29] Gordon SV, Eiglmeier K, Garnier T, Brosch R, Parkhill J, Barrell B, Cole ST, Hewinson RG. Genomics of *Mycobacterium bovis*. Tuberculosis (Edinburgh, Scotland). 2001;**81**(1-2): 157-163

[30] Oettinger T, Jørgensen M, Ladefoged A, Hasløv K, Andersen P. Development of the *Mycobacterium bovis* BCG vaccine: Review of the historical and biochemical evidence for a genealogical tree. Tubercle and Lung Disease. 1999;**79**(4):243-250

[31] WHO Consultation on the Characterization of BCG Vaccine. Geneva, Switzerland: WHO; December 8-9, 2004. Available from: http://www.who.int/biologicals/publications/meetings/areas/vaccines/bcg/

[32] Luca S, Mihaescu T. History of BCG vaccine. MAEDICA – A Journal of Clinical Medicine. 2013;**8**(1):53-58

[33] Stefanova T, Chouchkova M, Hinds J, Butcher PD, Inwald J, Dale J, Palmer S, Hewinson RG, Gordon SV. Genetic composition of *Mycobacterium bovis* BCG substrain Sofia. Journal of Clinical Microbiology. 2003;**41**(11):53-49

[34] Argus™ Optical Mapping System User Manual. MAN-11207-001.02 OpGen, Inc. ©2010 All Rights Reserved

[35] The Main Site of the Center for Biological Sequence Analysis at the Technical University of Denmark, Kemitorvet, Denmark. Available from: http://www.cbs.dtu.dk/services/gwBrowser

[36] Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: The reference centre for bacterial insertion sequences. Nucleic Acids Research. 2006;**34**(Database issue):D32-D36. DOI: 10.1093/nar/gkj014

[37] The Main Site of the Laboratory of Microbiology and Molecular Genetics, National Center for Scientific Research, Toulouse Cedex, France. Available from: http://www-is.biotoul.fr

[38] CRISPRs web server of the Institute of Genetic and Microbiology at the Paris-Sud University, France. Available from: http://crispr.u-psud.fr/Server/

[39] Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: A webtool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Research. 2007;**35**(Web Server issue):W52-W57. DOI: 10.1093/nar/gkm360

[40] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: A fastphagesearch tool. Nucleic Acids Research. 2011;**39**(Web Server issue):W347-W352. DOI: 10.1093/nar/gkr485

[41] The PHAST website is maintained by Dept. of Biological Sciences, University of Alberta, Edmonton, AB, Canada. Available from: http://phast.wishartlab.com/contact.html

[42] Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 1987;**4**(4):406-425

[43] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution. 2013;**30**(12):2725-2729

[44] Joung SM, Ryoo S. BCG vaccine in Korea. Clinical and Experimental Vaccine Research. 2013;**2**(2):83-91. DOI: 10.7774/cevr.2013.2.2.83

[45] Li X, Chen L, Zhu Y, Yu X, Cao J, Wang R, Lv X, He J, Guo A, Huang H, Zheng H, Liu S. Genomic analysis of a *Mycobacterium bovis* bacillus [corrected] Calmette-Guérin strain isolated from an adult patient with pulmonary tuberculosis. PLoS One. 2015; **10**(4):e0122403. DOI: 10.1371/journal.pone.0122403. eCollection 2015

[46] Casjens S. Prophages and bacterial genomics: What have we learned so far? Molecular Microbiology. 2003;**49**(2):277-300. DOI: 10.1046/j.1365-2958.2003.03580.x

[47] Bobay LM, Touchon M, Rocha EP. Pervasive domestication of defective prophages by bacteria. Proceedings of the National Academy of Sciences of United States of America. 2014;**111**(33):12127-12132. DOI: 10.1073/pnas.1405336111

[48] Zhang W, Zhang Y, Zheng H, Pan Y, Liu H, Du P, Wan L, Liu J, Zhu B, Zhao G, Chen C, Wan K. Genome sequencing and analysis of BCG vaccine strains. PLoS One. 2013;**8**(8):e71243. DOI: 10.1371/journal.pone.0071243

[49] Sergienko OV, Liashchuk AM, Aksenova EI, Galushkina ZM, Poletaeva NN, Sharapova NE, Semikhin AS, Kotnova AR, Veselov AM, Bashkirov VN, Kulikova NL, Khlebnikov VS, Kondrat'eva TK, Kariagina-Zhulina AS, Apt AS, Lunin VG, Gintsburg AL. Production of mycobacterial antigenes merged with cellulose binding protein domain in order to produce subunit vaccines against tuberculosis. Molecular Genetics, Microbiology and Virology. 2012;(1):16-20

[50] Odent MR. The future of neonatal BCG. Medical Hypotheses. 2016;**91**:34-36. DOI: 10.1016/j.mehy.2016.04.010

IntechOpen