


RESEARCH ARTICLE

Open Access



First insight into the whole-genome sequence variations in *Mycobacterium bovis* BCG-1 (Russia) vaccine seed lots and their progeny clinical isolates from children with BCG-induced adverse events

Olga Narvskaya^{1,2*} , Daria Starkova¹, Diana Levi³, Natalia Alexandrova³, Vladimir Molchanov^{1,4*}, Ekaterina Chernyaeva⁵, Anna Vyazovaya¹, Alexander Mushkin², Viacheslav Zhuravlev², Natalia Solovieva², Boris Vishnevskiy² and Igor Mokrousov¹

Abstract

Background: The only licensed live Bacille Calmette-Guérin (BCG) vaccine used to prevent severe childhood tuberculosis comprises genetically divergent strains with variable protective efficacy and rates of BCG-induced adverse events. The whole-genome sequencing (WGS) allowed evaluating the genome stability of BCG strains and the impact of spontaneous heterogeneity in seed and commercial lots on the efficacy of BCG-vaccines in different countries. Our study aimed to assess sequence variations and their putative effects on genes and protein functions in the BCG-1 (Russia) seed lots compared to their progeny isolates available from immunocompetent children with BCG-induced disease (mainly, osteitis).

Results: Based on the WGS data, we analyzed the links between seed lots 361, 367, and 368 used for vaccine manufacture in Russia in different periods, and their nine progeny isolates recovered from immunocompetent children with BCG-induced disease. The complete catalog of variants in genes relative to the reference genome (GenBank: CP013741) included 4 synonymous and 8 nonsynonymous single nucleotide polymorphisms, and 3 frameshift deletions. Seed lot 361 shared variants with 2 of 6 descendant isolates that had higher proportions of such polymorphisms in several genes, including *ppsC*, *eccD5*, and *eccA5* involved in metabolism and cell wall processes and reportedly associated with virulence in mycobacteria. One isolate preserved variants of its parent seed lot 361 without gain of further changes in the sequence profile within 14 years.

(Continued on next page)

* Correspondence: onarvskaya@gmail.com; Vladimir.Molchanov@vai.edu

¹Laboratory of Molecular Epidemiology and Evolutionary Genetics, St. Petersburg Pasteur Institute, St. Petersburg 197101, Russia

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

Conclusions: The background genomic information allowed us for the first time to follow the BCG diversity starting from the freeze-dried seed lots to descendant clinical isolates. Sequence variations in several genes of seed lot 361 did not alter the genomic stability and viability of the vaccine and appeared accumulated in isolates during the survival in the human organism. The impact of the observed variations in the context of association with the development of BCG-induced disease should be evaluated in parallel with the immune status and host genetics. Comparative genomic studies of BCG seed lots and their descendant clinical isolates represent a beneficial approach to better understand the molecular bases of efficacy and adverse events during the long-term survival of BCG in the host organism.

Keywords: BCG, *M. bovis* BCG-1 (Russia), Seed lot, Clinical isolate, Osteitis, Whole-genome sequencing

Background

Treatment of tuberculosis (TB) and vaccination remain inevitable health-care interventions to prevent new *Mycobacterium tuberculosis* infections and their progression to disease. The live Bacille Calmette-Guérin (BCG) vaccine is the only licensed vaccine that is successfully used for nearly a century to prevent childhood TB, although it shows a modest protective effect in adults [1, 2]. The history of BCG development and spreading was thoroughly reviewed [3, 4]. Briefly, the ancestral form of BCG was derived from wild-type *Mycobacterium bovis* progenitor strain, while attenuation by numerous in vitro passages resulted in world distribution of parental BCG strain in 1924. During the next decades, maintenance of the BCG strain progenies by further passages using different culture media and transfer schedules in different countries contributed to the in vitro evolution of BCG. The production of seed stocks without lyophilization until the 1960s resulted in the generation of BCG daughter strains (substrains) with variable morphological, culture, antigenic characteristics, and residual virulence due to adaptation in vitro [3–7].

The up-to-date genealogy of widely used BCG vaccine strains is based on genomic variations that occurred due to large deletions, tandem duplications, insertion sequences, and the subsequent series of insertions/deletions and single nucleotide polymorphisms (SNPs). Accordingly, BCG strains were classified into “early” strains represented by BCG Japan, Moreau, Russia (and the descendant BCG Sofia) strains that show fewer chromosomal deletions than “late” strains - Danish, Pasteur, Glaxo, and others [8–11]. Altogether, these variations underpin the molecular basis of attenuation, different levels of gene expression of immunodominant proteins leading to variable protective efficacy, and different rates of adverse events of currently used genetically divergent BCG strains otherwise considered different BCG vaccines [3, 8].

Historically, the original *M. bovis* BCG strain was transferred to Russia in 1924, and since 1925 it was referred to as *M. bovis* BCG-1 (also spelled as BCG-I). In

the 1940s, *M. bovis* BCG-1 was lyophilized, and in 1954 the seed lot system was adopted in Russia. The first seed lot used in 1963 for the manufacturing of freeze-dried BCG vaccines in the former Soviet Union was seed lot 352 (“ch”). Since 2006, the 368 (“shch”) generation of *M. bovis* BCG-1 (Russia) strain no. 700001 is in use for the production of the BCG vaccine in Russia [12, 13]. The same BCG generation 368 is also adopted by two manufacturers in India for vaccine production and distribution worldwide under the name “BCG Russia 368” [14].

The attenuated BCG vaccines are believed safe and effective mainly to prevent the most severe meningitis and miliary forms of childhood TB associated with high mortality in infants and young children, especially in high TB-burden countries [1, 2]. However, adverse events following immunization with BCG in 1–10% of vaccinees are likely to be substantially underreported since assays to confirm isolates as BCG are still relatively rare [15–20] despite the advent of molecular techniques and commercially available kits for BCG identification [21]. In Russia, the estimated frequency of complications per 100,000 BCG-immunized decreased from 35.3 in 2005 to 11.2 in 2013 [22].

Vaccine-related complications, i.e., BCG-induced disease (BCG-ID), may occur later than 12 months after vaccination as local site reactions or distal to the site of inoculation, and they appear to differ from cutaneous lesions, lymphadenopathy/lymphadenitis, and hypersensitivity reactions to osteitis and disseminated BCG infections (almost exclusively in individuals with compromised cellular immunity) [1, 16, 19, 23–25].

The BCG-induced adverse reactions are believed to be due to the use of genetically distant vaccine strains with variable residual virulence, the culture technique, the route of administration and the dose of viable cells in the BCG vaccine delivered, along with the age, immune status, and host genetics of the vaccinees [4, 8, 9, 16, 17, 26, 27].

To standardize the vaccine production and to prevent adverse reactions related to BCG vaccination, the WHO

has established the international requirements for the manufacture and quality control including genetic characterization of seeds and final commercial lots of the reference strains, i.e., BCG Danish 1331, Tokyo 172–1, and Russia BCG-1 used for vaccine manufacture and distribution worldwide [11, 28, 29].

Given recommendations, the identity and genome stability of seven seed lots of *M. bovis* BCG-1 (Russia) used for immunization against TB in Russia since 1948 till present, and 32 commercial lots of the vaccine produced by Russian manufacturers were proved by multiplex polymerase chain reaction [12].

During the last decades, the whole-genome sequencing (WGS) was applied to assess the genome stability of BCG vaccine strains from different culture collections and to evaluate the impact of minor sequence variations occurring in seed and commercial lots on the efficacy of BCG-vaccines used in different countries [30–34]. Recently, sequence variations were identified in clinical isolates from two infants with BCG-induced disease compared with the sub-cultured *M. bovis* BCG Moscow strain (Serum Institute of India, India) in South Africa [15].

The complete genome sequences of the current *M. bovis* BCG-1 (Russia) 368 (“shch”) freeze-dried seed lot and its working seed lot used for vaccine production by one of the Russian manufacturers are available in GenBank (NCBI) under accession numbers CP011455 [13], CP009243 [35], and CP013741 [36]. The genome constancy of *M. bovis* BCG-1 (Russia) 368 within the entire vaccine production flow (from working seed lot, the last production passage to final vaccine product) was confirmed by spoligotyping, 24 loci VNTR-typing, and WGS [36, 37]. However, so far, there is a lack of data on the contribution of BCG genomic diversity to the viability, fitness, and residual virulence of vaccine populations in vivo in the context of the development of BCG-induced adverse events. Therefore, our study aimed to assess sequence variations and their putative effects on genes and protein functions in the BCG-1 (Russia) seed lots compared to their progeny clinical isolates available from immunocompetent children with BCG-induced osteitis and lymphadenitis.

Results

Sequence variations identified in BCG-1 (Russia) parent seed lots and their progeny clinical isolates

Raw reads of seed lots 361, 367, 368, and descendant clinical isolates were mapped to the reference genome of BCG-1 (Russia) (GenBank: CP013741). Variant calling allowed to produce the complete catalog of 15 differences, including 12 SNPs (four synonymous and eight nonsynonymous) and three disruptive deletions (frame-shift variants) in CDS of the seed lots and clinical

isolates relative to the reference genome. Overview of sequence variation patterns identified in BCG seed lots and their progeny clinical isolates are presented in Table 1, Additional Files 1 and 2.

His57Asp mutation in *pncA* (Rv2043c) gene conferring high intrinsic resistance to pyrazinamide in *M. bovis* was detected in all BCG-1 (Russia) progeny isolates. A single-nucleotide insertion in the *recA* gene was not detected in any of three seed lots, nor their descendant clinical isolates.

Comparative analysis of sequence variants in BCG-1 (Russia) seed lots and their progeny clinical isolates

We examined a set of nine BCG-1 (Russia) clinical isolates recovered from duly BCG-vaccinated immunocompetent children (except one, immunized with delay for 4 months after birth) with a culture-confirmed diagnosis of BCG-ID. For each case, analysis of medical records allowed to follow the history of primary BCG vaccination, including information on commercial vaccine products used for intradermal inoculation, its manufacture, and finally, the parent seed lot (Tables 2 and 3).

BCG clinical isolates shared the majority of genomic features with their parent seed lots. However, the proportions of reads with alternative alleles varied in BCG seed lots, and their progeny isolates. These findings were summarized in Table 1 and Additional file 1. Based on the comparative analyses of the data obtained, we built a schematic view of the relationships between BCG seed lots and their descendant clinical isolates (Fig. 1).

The population of BCG-1 (Russia) seed lot 361 was polymorphic with variable proportions of SNPs and deletions in reads spanning seven CDS (Table 1, Fig. 1, Additional file 1).

Synonymous SNP in the ornithine aminotransferase gene (*rocD*) was detected in almost 1/3 reads of seed lot 361, but it was absent in the descendant isolates. Alternatively, seed lot 361 shared nonsynonymous (missense) SNPs in type VII secretion system ESX-5 (*eccD5* and *eccA5*), dolichol-phosphate mannosyltransferase (*ppm1*) and hypothetical protease genes with its progeny clinical isolates 2925 and 5448 that had higher proportions (up to 100%) of alternative variants. However, sequences of these isolates differed in amino acid transporter (*cycA*) and polyketide synthase (*ppsC*) genes. A nonsynonymous SNP in the PE-PGRS family protein PE_PGRS7 gene was found in 38% reads of isolate 5340, the progeny of seed lot 361.

Isolates 1986, 4159, and 5075, progenies of the seed lot 361, and seed lot 367 did not show sequence polymorphisms compared to the reference sequence, i.e., BCG-1 (Russia) current seed lot 368 (Table 1, Additional file 1).

Table 1 Sequence variants in CDS of BCG-1 (Russia) parent seed lots 361, 367 and 368 and their progeny clinical isolates compared to the reference sequence BCG-1 (Russia) (GenBank: CP013741)

Parent gene BVR	Chain	Seed lot 361	1986	2925	4159	5075	5340	5448	Seed lot 367	3363	Seed lot 368	577	1032	Parent gene BVR amino acid change	Product name	Effect (putative impact)
01860	-	CG	CG	CG	CG	CG	CG	CG	CG	C	CG	CG	CG	Pro170fs	thioesterase	high
03020	-	C	C	C	T	C	C	C	C	C	C	C	C	Ala41Thr	PE-PGRS family protein PE_PGRS7	moderate
03865	+	CA	CA	CA	CA	CA	CA	CA	CA	CA	C	C	CA	Thr229fs	methionine aminopeptidase	high
06950	+	T	T	T	T	T	T	T	T	T	T	C	T	Leu468Ser	ATP synthase subunit beta	moderate
08645	+	G	A	G	A	A	A	A	A	A	A	A	A	Val668Val	polyketide synthase	low
08855	-	G	A	A	A	A	G	A	A	A	A	A	A	Leu372Pro	amino acid transporter	moderate
09330	+	A	G	A	G	G	A	A	G	G	G	G	G	Ala328Thr	type VII secretion integral membrane protein EccD	moderate
09345	+	C	T	T	T	T	T	T	T	T	T	T	T	Ile43Thr	type VII secretion AAA-ATPase EccA	moderate
10,325	+	GC	GC	GC	GC	GC	GC	GC	GC	GC	GC	GC	GC	Leu212fs	universal stress protein UspA	high
10,615	-	T	C	C	C	T	C	T	C	C	C	C	C	Gly196Asp	dolichol-phosphate mannosyltransferase	moderate
12,050	-	A	C	C	C	C	C	C	C	C	C	C	C	Leu26Leu	ornithine aminotransferase	low
13,820-	+	T	C	C	C	C	T	C	C	C	C	C	C	Ala252Val	protease	moderate
15,090	-	G	G	G	G	G	G	G	G	G	G	G	G	Ala713Ala	protein-PII uridylyltransferase	low
16,925	-	C	C	C	C	C	C	C	C	C	C	C	T	Leu282Leu	N-acetylglucosaminyl-diphospho-decaprenol L-rhamnosyltransferase	low
17,690	+	A	A	A	A	A	A	A	A	G	A	A	A	Gln278Arg	amidase	moderate

Sequence variants are presented as the proportion of reads with alternative allele/all reads; *del*/deletion, *fs* frameshift

Table 2 Characteristics of patients and *M. bovis* BCG-1 (Russia) clinical isolates

Patient ID/sex ^a	Date of birth ^b	Region of birth	Vaccination	Vaccine lot number used for inoculation	Vaccine manufacturer	Vaccine seed lot	Diagnosis ^c	Age of surgical intervention ^d	Date of surgical intervention ^h	Clinical isolate	NCBI SRA accession No
1/f	05/2017	St. Petersburg	delayed ^e	411	Medgamal ^f	368 "shch"	A18.2	7 m	01/2018	577	SRR9593177
2/m	06/2016	Dagestan	at birth	537	Microgen ^g	368 "shch"	A18.0	1y8m	02/2018	1032	SRR9593172
3/m	07/2006	Moscow	at birth	621	Medgamal	361 "sh"	A18.0	9 m	04/2007	1986	SRR9593171
4/f	03/2002	St. Petersburg	at birth	281	Medgamal	361 "sh"	A18.0	14y	04/2016	2925	SRR9593170
5/f	07/2004	Ekaterinburg	at birth	25	Microgen	367 "sh"	A18.0	2y10m	05/2007	3363	SRR9593169
6/f	12/2004	St. Petersburg	at birth	437	Medgamal	361 "sh"	A18.0	2y6m	06/2007	4159	SRR9593168
7/m	04/2006	St. Petersburg	at birth	615	Medgamal	361 "sh"	A18.0	1y5m	09/2007	5075	SRR9593175
8/m	12/2004	St. Petersburg	at birth	725	Medgamal	361 "sh"	A18.0	4y9 m	09/2009	5340	SRR9593174
9/m	06/2005	Kabardino-Balkaria	at birth	567	Medgamal	361 "sh"	A18.0	1y3m	09/2006	5448	SRR9593173

^af female, m male; ^bmonth/year; ^cdiagnosis according to International Statistical Classification of Diseases and Related Health Problems (ICD-10): Tuberculosis of bones and joints (A18.0), A18.2 (Tuberculosis peripheral lymphadenopathy) [International statistical classification of diseases and related health problems. - 10th revision, edition 2010]; ^dy year, m month; ^edelayed for 4 months due to respiratory infection; ^fBranch "Medgamal," Federal State Budgetary Institution "National Research Center for Epidemiology and Microbiology named after Honorary Academician N.F. Gamalei", Ministry of Health of the Russian Federation, Russia; ^gScientific and Production Association for Immunobiological Preparations "Microgen," Russia; ^hmonth/year

Isolate 3363, a single progeny of the seed lot 367, carried a missense variant in amidase (*amiD*) gene and a frameshift deletion in the thioesterase (*pks13*) gene (Table 1, Fig. 1, Additional file 1).

The sequenced seed lot 368 shared a single base deletion in methionine aminopeptidase (*mapA*) with descendant isolate 577, also demonstrating a polymorphism (43% reads) in the ATP synthase subunit beta gene (*atpD*). Polymorphism in the N-acetylglucosaminyl-diphospho-decaprenol L-rhamnosyltransferase gene (*wbbL1*) and a frameshift deletion in the universal stress protein UspA gene were identified in clinical isolate 1032 (18 and 28% reads, respectively), the progeny of seed lot 368 (Table 1, Fig. 1; Additional file 1 and 3).

Discussion

All nine BCG clinical isolates recovered from tissue samples taken during the surgery were susceptible to first-line antibiotics: streptomycin, isoniazid, rifampin, ethambutol. However, they were pyrazinamide-resistant due to the typical His57Asp mutation in *pncA* (Rv2043c)

gene conferring high intrinsic resistance in *M. bovis*, and consequently, all BCG vaccine strains [38].

Previously, spontaneous heterogeneity of BCG seed lots and commercial vaccines during vaccine production determined by variant calling was analyzed in the Tokyo-172 vaccine strain [32]. In line with the published data, we observed a synonymous A2139G single nucleotide base substitution (position 3,175,529, "-" chain) in BOVR_15,090, i.e., the *glnD* (ortholog Rv2918c in *M. tuberculosis* H37Rv) gene, without Ala replacement in position 713 of 808 amino acids in the conservative domain of the PII uridylyltransferase protein in all genomes studied. The G variant was present in 100% reads of sequenced seed lots 361 and 367, and eight of nine clinical isolates. However, it was detected only in 25% reads of current seed lot 368 (sequenced in this study), and 46% reads of its progeny isolate 577.

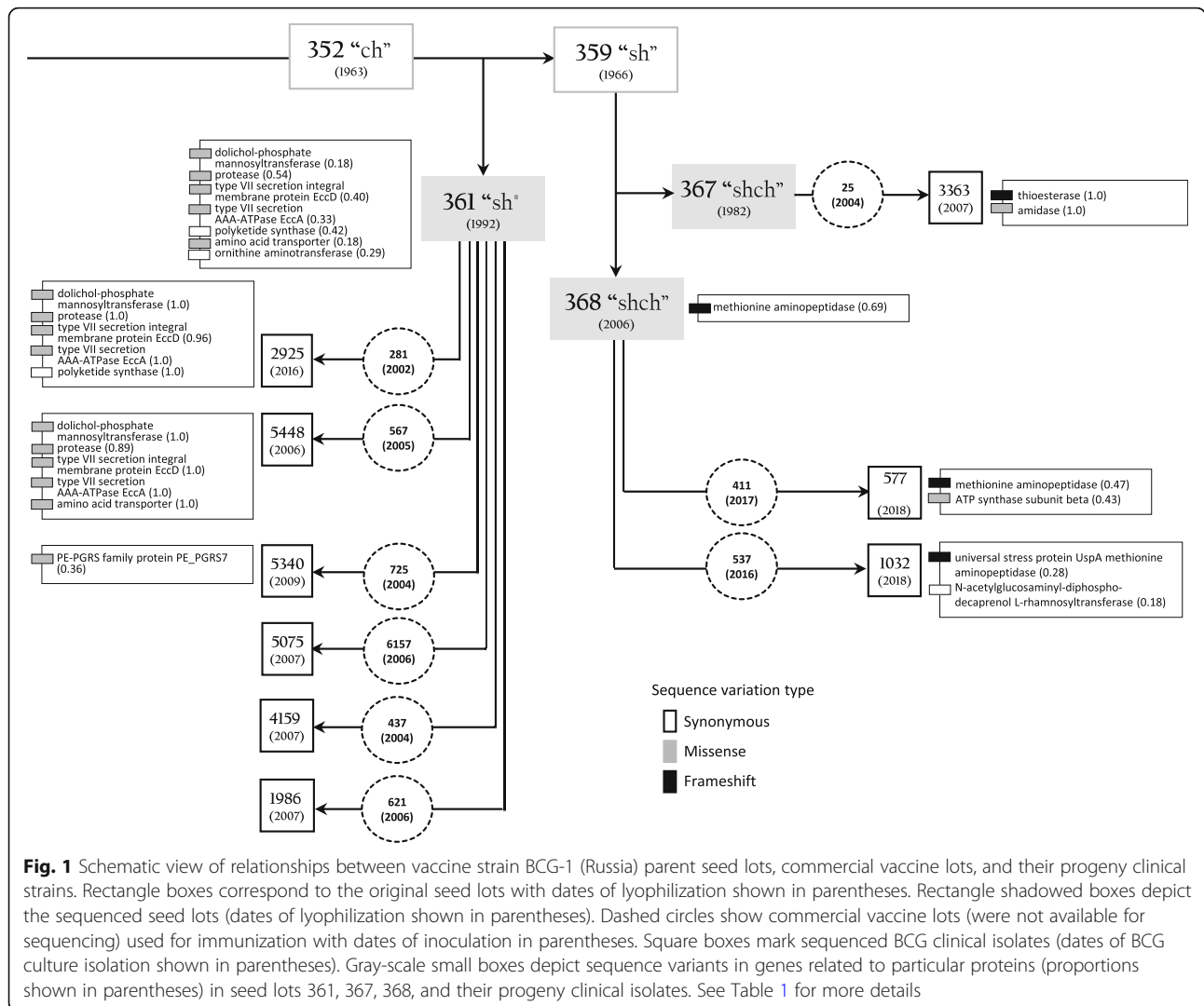
These findings inspired further examination of A/G polymorphisms in the *glnD* in other sequenced BCG genomes.

The G variant was previously reported in the *glnD* in BCG-1 (Russia) current seed lot 368 and some earlier

Table 3 Seed lots of *M. bovis* BCG-1 (Russia) substrain No. 700001

Freeze-dried seed lot ^a	Date of lyophilization ^b	Origin	Used for vaccine manufacturing	Manufacture	NCBI SRA accession number
361 "sh"	09/1992	Freeze-dried Seed lot 352 "ch" (lyophilization 1963)	2002–2007	Medgamal	SRR9593176
367 "shch"	11/1982	Freeze-dried Seed lot 359 "sh" (lyophilization 1966)	2001–2008	Microgen	SRR9593167
368 "shch"	08/2006	Freeze-dried Seed lot 359 "sh" (lyophilization 1966)	2008 till present	Medgamal, Microgen	SRR9593166

^a the lyophilized seed lots (generations 361 "sh," 367 "shch," and 368 "shch") of *M. bovis* BCG-1 (Russia) substrain No. 700001 in ampoules were obtained from the National State Collection of Pathogenic Microorganisms (NSCPM) of Scientific Center for Expertise of Medical Application Products of the Ministry of Health, Russia along with the characteristics (the origin and manufacture) of corresponding BCG vaccine commercial lots used for immunization; ^b month/year



vaccine products [35], though alternative variant A was detected in two other BCG-1 (Russia) 368 complete genomes [13, 36]. The stability of variant A in the *glnD* gene of the BCG-1 (Russia) 368 working seed lot (GenBank: CP013741) and consistent vaccine batches within 2 years was confirmed by WGS [36, 37]. Noteworthy, all three complete genomes of BCG-1 (Russia) were sequenced independently from the same generation 368 (“shch”), which is still in use since 2008 for the BCG vaccine production by two Russian manufactures: Medgamal (Moscow) and Microgen (Stavropol).

Interestingly, allele G was recorded in the genomes of BCG str. Tokyo 172 which represents a well-defined Type I subpopulation of BCG Tokyo-172, the closest to BCG Russia (GenBank: NZ_CUWO01000001) and Russia ATCC 35740, members of presumably monophyletic “early” substrain group of BCG [10, 30, 32]. The G variant was also common in the reference sequences of some other BCG strains, e.g., Pasteur 1173P2, Danish

1331, ATCC 35743 (Tice), Moreau RDJ, used worldwide for vaccine production and research purposes.

Altogether, these data suggest that intrinsic polymorphism in the *glnD* gene of BCG-1 (Russia) strain represented by different seed lots and descendant clinical isolates occurs due to the selection of variants during laboratory and vaccine production manipulations, rather than mutations in patients’ organisms. Otherwise, the observed heterogeneity can be partly explained as a result of processing nucleotide variants among the mapped reads using different bioinformatics tools and quality criteria.

Another interesting finding: a notorious single-nucleotide insertion in the *recA* gene leading to non-functional truncated recombinase A reported previously in a vaccine production lot of *M. bovis* BCG Russia (corresponding to ATCC 35740) [39] was not detected in seed lots studied, nor their descendant clinical isolates. De facto, there were no alterations in *recA* in the

previously sequenced complete genomes of BCG-1 (Russia) [13, 35, 36] along with other hyperconserved BCG strains [33].

Based on the comparative study of WGS data, we attempted to analyze the links between BCG-1 (Russia) seed lots used for vaccine manufacture and their progeny clinical isolates associated with specific adverse events, e.g., BCG-osteitis and an abscess at the point of injection.

The population of BCG-1 (Russia) seed lot 361 appeared heterogeneous with variable proportions of SNPs and deletions in reads spanning seven CDS. Seed lot 361 shared SNPs in *eccD5*, *eccA5*, *ppm1* genes, and hypothetical protease gene with its progeny clinical isolates 2925 and 5448. The latter typically had higher proportions (reaching up to 100%) of alternative alleles than the parent seed lot 361, thus suggesting the likely accumulation of such changes under selective pressure in the host organism. The similar SNPs we observed in corresponding genomic positions in *M. tuberculosis* H37Rv, BCG strains Tokyo-172 type I and Pasteur 1173P2. These polymorphisms affected type VII specialized bacterial secretion ESX-5 (integral membrane protein EccD and secretion AAA-ATPase), reportedly associated with virulence in mycobacteria, and dolichol-phosphate mannosyltransferase and protease genes involved in the vital cell processes (protein modification, lipoprotein biosynthesis), intermediary metabolism, and respiration [8, 40, 41]. The data obtained confirmed the close genetic relatedness between seed lot 361 and both progeny isolates 2925 and 5448 that otherwise differed from each other in amino acid transporter (*cycA*) and polyketide synthase (*ppsC*) sequences.

A unique missense variant in the PE-PGRS family protein PE_PGRS7 gene with an unclear function in antigenic variation and interactions of bacteria with the host immune system was found in 38% reads of isolate 5340, the progeny of seed lot 361.

Interestingly, no sequence polymorphisms (except *glnD*) were detected in isolates 1986, 4159, and 5075, progenies of the seed lot 361 (used for BCG manufacture in 2002–2007) compared to the reference sequence, i.e., BCG-1 (Russia) current seed lot 368.

The sequences of seed lot 367 (used for vaccine production in 2001–2008) also did not show variations in 14 of 15 CDS compared to the reference genome and thus appeared the most conservative in our study. However, its single progeny, isolate 3363, carried a missense variant in amidase (*amiD*) gene essential for protein translation and required for mycobacterial survival during infection (virulence factor) and a frameshift deletion in the thioesterase region of polyketide synthase 13 (*pks13*) gene involved in the final assembly of mycolic acids, major and specific lipid components of the envelope essential for the survival of a mycobacterial cell [40, 42].

The population of the sequenced seed lot 368 was almost homogenous containing a single base deletion (69% reads) in methionine aminopeptidase (*mapA*) gene compared to the previously reported complete genomes of BCG-1 (Russia) current-generation 368 [13, 35, 36] and BCG Tokyo 172 type 1 [32]. This frameshift leading to protein truncation was inherited in the descendant isolate 577 (47% reads), also bearing an isolate-specific allele (43% reads) in the ATP synthase subunit beta gene (*atpD*). Both proteins encoded by *mapA* and *atpD* genes are involved in the processes of intermediary metabolism and respiration [8].

A synonymous SNP in the N-acetylglucosaminyl-diphospho-decaprenol L-rhamnosyltransferase gene (*wbbL1*) not altering the enzyme structure and function in the cell wall and cell processes was identified in clinical isolate 1032 (18% reads). Another finding in isolate 1032 (28% reads) was a frameshift due to a single-base cytosine deletion in codon 630 of the universal stress protein UspA gene (orthologue Rv1996 in *M. tuberculosis* H37Rv), one of eight Usp paralogues essential for the intracellular survival of pathogenic mycobacteria in hypoxia, low levels of nitric oxide and carbon monoxide environment. This deletion completely altering 27 amino acids from the position 212 onwards resulted in a gained opal stop codon (TGA). The latter was leading to a premature protein truncation in the residue 239 yielding to the loss of 3 ligand-binding loci in the putative ATP binding motif of the UspA domain 2 [43].

It appears that the accumulation of shared SNPs and unique variations occurred in clinical isolates 577 and 1032 (both progenies of the sequenced seed lot 368), 3363 (progeny of seed lot 367), and 5340 (progeny of seed lot 361) did not affect the vaccine survival in the organisms of vaccinees. Moreover, none of the frameshifts, which presumably could alter essential cell functions, were crucial for the microbial population in vitro, nor in vivo, thus emphasizing the long-term viability of BCG-1 (Russia) strain in the human organism.

Adverse events following immunization with BCG are considered underreported since assays to confirm isolates as BCG are still relatively rare. Moreover, based on the routine laboratory diagnostics and even molecular techniques, it is difficult to associate the clinical BCG isolates with the particular vaccine lot administered [15–20]. Therefore, the contribution of BCG genomic heterogeneity to the viability and residual virulence of vaccine populations in vivo remains unclear in terms of the development of BCG-related disease that otherwise can be associated with immune responses and human genetics.

Unfortunately, the administered commercial vaccines were unavailable for WGS, which limited the comparative analysis of sequence variations through the whole vaccine production chain. Nevertheless, bioinformatics

tools provided a possibility to identify sequence variants and their putative effects on genes and protein functions in BCG-1(Russia) seed lots 361 and 368 used for vaccine manufacture in Russia in different periods relative to their progeny clinical isolates from patients with confirmed BCG-related disease and the genomes of internationally recognized BCG reference strains.

BCG osteitis is a rare but severe adverse event that typically presents in children before the age of 5 years, and there are few reports concerning the late onset of the disease in teenagers [17, 19]. In our setting, we analyzed confirmed cases of BCG–osteitis typically developed at the age under 5 years (median 22 months) among immunocompetent children immunized with BCG-1 (Russia) vaccine after, or soon after birth. The latest onset of the osteitis affecting the left eighth rib was diagnosed in the patient aged 14 years who once received as a newborn the BCG vaccine derived from the seed lot 361. The isolate 2925 recovered from tissue samples taken during surgery accumulated four nonsynonymous and one synonymous SNPs assigned to its parent seed lot 361 but did not gain any further sequence variants in the patient’s organism. Noteworthy, such a prolonged period of survival in a human organism of the live BCG vaccine without significant changes in its genome appears remarkable.

Conclusions

To our knowledge, we reported for the first time the results of a WGS-based comparative analysis of the BCG-1 (Russia) seed lots used for the vaccine production in different periods and their progeny clinical isolates recovered from immunocompetent children with BCG-induced adverse reactions (mainly, BCG osteitis). Our study provided the background genomic information, which allowed us to follow the BCG diversity starting from the freeze-dried seed lots to descendant clinical isolates presenting the results of the long-term survival of vaccine populations in the human organism.

Two of three BCG-1 (Russia) seed lots were heterogeneous with various proportions of sequence variants in several genes, including *ppsC*, *eccD5*, and *eccA5* involved in metabolism and cell wall processes and reportedly associated with virulence in mycobacteria. However, such polymorphisms did not alter the genomic stability and viability of BCG-1 (Russia)-derived vaccines in vivo. Sequence variants occurred in polymorphic CDS of seed lot 361 used for vaccine manufacture before 2006 appeared accumulated in two of six progeny isolates, hypothetically, under selective pressure in the human organism. Although polymorphisms were identified in two unrelated clinical isolates 2925 and 5448 from children immunized in different years with different vaccine lots produced from the BCG-1 (Russia) seed lot 361 by

the same manufacturer, the impact of such changes in the context of association with the development of BCG-induced disease remains questionable due to the lack of detailed information on the immune status and host genetics of presumably immunocompetent patients with BCG-osteitis.

Despite promising results in the development of a new generation of BCG vaccines, there is no reliable alternative to those which are still in use in the nearest future.

The contributions of genomic diversity and variations in gene expression profiles during the BCG vaccine manufacture flow and adaptations in vivo are yet to be uncovered in parallel with the molecular mechanisms of innate and adaptive immune responses in the human organism. Thereby, the comparative studies of genomic variations in production strains, their seeds, administered vaccine lots, and the descendant clinical isolates (if available) represent a beneficial approach to better understand the bases of vaccine efficacy and adverse reactions of present and future BCG-based vaccines at genomic, transcriptomic, and proteomic levels.

Methods

Study design

We retrospectively reviewed 65 medical records at the clinics of St. Petersburg Research Institute of Phthisiopulmonology selected as codes A18.0 (Tuberculosis of bones and joints, acute or chronic osteomyelitis) and A18.2 (Tuberculosis peripheral lymphadenopathy) according to the International Statistical Classification of Diseases and Related Health Problems (ICD-10). The probable BCG-ID was suspected in nine immunocompetent children (eight patients aged under 5 years and the 14-year old patient) with developed osteitis (eight patients), or cutaneous lesion near the BCG inoculation site (one patient), well-defined BCG inoculation records, and without a history of TB contacts admitted for treatment (including surgical intervention) in 2006–2018 (Table 2).

Inclusion criteria

The definition for the diagnosis of BCG-ID (“BCG-itis”) included 1) the histology showing granulation tissue with caseous necrosis in biopsy samples taken while surgery and 2) mycobacteria cultures recovered from tissue samples on the Löwenstein-Jensen (L-J) medium if they were identified as *M. bovis* BCG using GenoType MTBC (Hain Lifescience, Germany). The *M. bovis* BCG cultures were stored at – 80 °C for further examination.

Drug susceptibility testing

Drug susceptibility to first-line antibiotics (streptomycin, isoniazid, rifampin, ethambutol) and pyrazinamide of the *M. bovis* BCG cultures was assessed by using conventional

absolute concentration method and/or BACTEC MGIT 960 System (Becton-Dickinson, Sparks, MD, USA).

BCG-1 (Russia) seed lots and clinical isolates used for WGS

For genomic analysis, we used genomic DNA extracted as described [44] from *M. bovis* BCG-1 (Russia) freeze-dried seed lots and nine BCG clinical isolates cultured on L-J medium, each recovered from a child patient with confirmed BCG-ID.

The lyophilized seed lots of three generations 361 (“sh”), 367 (“shch”), and 368 (“shch”) of *M. bovis* BCG-1 (Russia) substrain No. 700001 in ampoules were obtained from the National State Collection of Pathogenic Microorganisms (NSCPM) of Scientific Center for Expertise of Medical Application Products of the Ministry of Health, Russia along with the characteristics (the origin and manufacture) of corresponding BCG vaccine commercial lots used for immunization (Table 3).

Clinical isolates 1986, 2925, 4159, 5340, 5075, and 5448 - progenies of BCG-1 (Russia) seed lot 361 (“sh”), isolate 3363 – of seed lot 367 (“shch”) and isolate 1032 – of seed lot 368 (“shch”) were obtained from biopsies of patients with BCG-osteitis affecting femur, tibia, calcaneus or ribs; isolate 577 – the progeny of seed lot 368 (“shch”) was obtained from a patient with an abscess at the point of injection (Tables 2 and 3).

Library construction and DNA sequencing bioinformatics analysis

DNA samples were purified by RNase A (Thermo Scientific™ #EN0531) and sonicated using Bioruptor® UCD-200 system (Diagenode, Denville, NJ, USA) to prepare paired-end (P.E.) libraries by following NEB Library Preparation Protocol, as seen in the product manual for the NEBNext Ultra II DNA Library Preparation Kit for Illumina (NEB #E7645). P.E. genomic libraries were subjected to WGS performed using the MiSeq platform (Illumina, San Diego, CA, USA). The average coverage of 28x was achieved. Raw PE reads were qualitatively analyzed using FastQC software (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) to retrieve proper parameters for trimming and filtering with Trimmomatic software version 0.32 [45]. Trimmed reads were aligned to the initially indexed reference genome of *M. bovis* BCG-1 (Russia), available at NCBI GenBank under the accession number [CP013741](https://www.ncbi.nlm.nih.gov/nuccore/CP013741) (RefSeq: NZ_CP013741.1) using BWA-MEM algorithm [46]. Sequence alignment maps were subsequently sorted by coordinates with conversion to binary alignment maps, deduplicated and indexed using Picard software version 2.17.8 (<http://broadinstitute.github.io/picard/>) SortSam, MarkDuplicates, and BuildBamIndex tools, respectively. Next, deduplicated reads were subjected to the

realignment based on the realignment target list performed with the Genome Analysis Toolkit (GATK) software version 4.1.1.0 IndelRealigner tool [47]. Realigned reads were used for variant identification using the GATK version 4.1.1.0 HaplotypeCaller tool. Identified variants were selected to generate separate SNPs- and indels-containing variant call format (VCF) files using the GATK version 4.1.1.0 SelectVariants tool. SNPs and indels then were processed with the GATK version 4.1.1.0 VariantFiltration tool according to the following parameters such that SNPs with the quality per depth (Q.D.) less than 2.0, Fisher strand bias (F.S.) more than 60.0, strand odds ratio (SOR) more than 4.0, root mean square mapping quality (M.Q.) less than 40.0, mapping quality rank-sum test (MQRankSum) less than -12.5, read position rank-sum test (ReadPosRankSum) less than -8.0 and indels with Q.D. < 2.0, F.S. > 200.0, ReadPosRankSum < -20.0, SOR > 10.0 were filtered out. Next, realigned reads were subjected to the base quality score recalibration (BQSR) stage using the GATK version 4.1.1.0 BaseRecalibrator and PrintReads tools to model systematic sequencing technical errors, adjust base quality scores accordingly basing on filtered VCF files and generate recalibrated reads, which improves the accuracy of variant calls. Recalibrated reads were used for adjusted variant identification using the GATK version 4.1.1.0 HaplotypeCaller tool. Identified variants were selected to generate separate SNPs- and indels-containing variant call format (VCF) files using the GATK version 4.1.1.0 SelectVariants tool. SNPs and indels were then processed with the GATK version 4.1.1.0 VariantFiltration tool, as described above. Effects on genes and putative impact (modifier, low, moderate, high) of identified variants were predicted using SnpEff software version 4.3 T [48] basing on annotations retrieved from general feature format version 3 (GFF3) file of *M. bovis* BCG-1 (Russia) current seed lot 368 (“shch”) available at NCBI under the accession number CP013741. Particular genome features were retrieved from the PATRIC platform 3.6.2 (<https://www.patricbrc.org/view/Taxonomy/1763>) [49], UniProt (<https://www.uniprot.org/>) [50], and MycoBrowser portal (<https://mycobrowser.epfl.ch>) [51].

The *pncA* gene was searched for mutations conferring resistance to pyrazinamide in BCG clinical isolates [38].

Comparative genomic analysis

Reference genomic sequences of *M. tuberculosis* H37Rv (NC_000962.3) and *M. bovis* BCG strains: BCG-1 (Russia) (NZ_CP013741; NZ_CP011455; NZ_CP009243), Tokyo-172 type I (NC_012207), Pasteur 1173P2 (NC_008769), BCG Russia (NZ_CUWO01000001), and Danish 1331 (NZ_CP039850) were used to analyze variations in coding DNA sequences (CDS) by the NCBI basic local alignment

and search tool (BLAST) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-020-06973-5>.

Additional file 1: Table S1. Sequence variants in BCG-1 (Russia) parent seed lots 361, 367, and 368 and their progeny clinical isolates (complete information). Sequence variants are presented as reads with alternative allele/all reads, ratio, %; del, deletion; fs, frameshift.

Additional file 2: Figure S1. Comparative circular genome diagram of *M. bovis* BCG 1 (Russia) vaccine seeds and clinical isolates (circo plot was created with Circa (<http://omgenomics.com/circa>)). Tracks from inside: scale in megabases; G.C. content histogram (yellow fill with black stroke) calculated from the ratio $(G + C)/(A + T + G + C)$ using a 1 kilobase (kb) non-overlapping sliding window; G.C. bias plots calculated from the ratio $(G/C)/(G + C)$ using a 1 kb non-overlapping window (blue plot) and a 10 kb non-overlapping window (translucent black plot); strain number labeled tracks (alternating grey and white) each depicting strain-specific insertions/deletions (blue/red transverse marks) and SNPs (dots*) identified relative to the reference genome of *M. bovis* BCG-1 (Russia) (GenBank accession number CP013741); tracks depicting ORFs (blue transverse marks) alongside affected CDSs (black, orange and red transverse marks[†]) on forward and reverse strands; names of affected genes colored according to an impact type of the most severe variant affecting a particular gene. Notes. *Sizes and colors determine a particular SNP type: small black – synonymous SNP, large orange – missense SNP; [†]color determines an impact type of a particular variant on a particular CDS assigned according to the following concept: low impact (black mark) – synonymous SNPs, moderate impact (orange mark) – missense SNPs and conservative insertions/deletions, high impact (red mark) – nonsense SNPs and frameshift insertions/deletions. (PPTX 311 kb)

Additional file 3: Figure S2. UspA domain 2: sequence alteration and truncation. (A) Sequences of affected *UspA* gene region. The top two rows depict reference region aligned against respective loci of strain 1032, indicating cytosine deletion in the codon 210. The third row represents the consensus sequence of the region highlighting premature opal (TGA) stop codon. (B) UspA amino acid sequence alteration and protein truncation. Domain 2 altered region and corresponding changed residues indicated in red. Ligand-binding sites depicted as green regions. (PPTX 59 kb)

Abbreviations

BCG: Bacille Calmette-Guérin; TB: Tuberculosis; SNP: Single nucleotide polymorphisms; BCG-ID: BCG-induced disease; WGS: Whole-genome sequencing; CDS: Coding DNA sequences

Acknowledgments

We thank Drs. Ruslan Ludannyy, Egor Shitikov, and Olga Voronina for helpful conversations and thoughtful comments. Scientific research was performed using equipment of the Resource Center “Biobank”, Research park of St. Petersburg State University.

Authors' contributions

ON, DL, AM, and VZ conceptualized the study and designed the experiments. NA, AV, EC, VZ, and NS performed data curation and laboratory experiments. VM, ON, and DS performed the computational analysis. ON, DS, VM, BV, and IM wrote and edited the manuscript. All authors read and approved the manuscript.

Funding

This research received funding from the Federal Government of the Russian Federation under Rospotrebnadzor Scientific Research Program AAAA-A16-116061410036-3.P.1.3.3.1. The funding body played no role in the design of the study and collection, analysis, and interpretation of data and in writing of the manuscript.

Availability of data and materials

Raw sequencing data of BCG-1 (Russia) seed lots 361, 367, and 368 (accession numbers: SRR9593176, SRR9593167, SRR9593166), and nine BCG clinical isolates (accession numbers: SRR9593177, SRR9593172, SRR9593171, SRR9593170, SRR9593169, SRR9593168, SRR9593175, SRR9593174, SRR9593173) are available in the NCBI Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) related to the BioProject “Whole-genome sequence variation of *Mycobacterium bovis* BCG-1 (Russia) vaccine substrain” (Accession: PRJNA523499). Reference genomic sequences of *M. tuberculosis* H37Rv (NC_000962.3) and *M. bovis* BCG strains BCG-1 (Russia) (CP013741, NZ_CP013741; NZ_CP011455; NZ_CP009243), Tokyo-172 type I (NC_012207), Pasteur 1173P2 (NC_008769), BCG Russia (NZ_CUW001000001), and Danish strain 1331 (NZ_CP039850) are available at GenBank NCBI (<https://www.ncbi.nlm.nih.gov/genbank/>). All data generated or analyzed during this study are included in this published article, in particular, in the Methods section (sub-section Comparative genomic analysis) and Tables 2 and 3.

Ethics approval and consent to participate

Our retrospective, descriptive study was approved by the Independent Ethics Committee of St. Petersburg Research Institute of Phthisiopulmonology of Ministry of Health, Russia (protocol №17/12–2017, 18.12.2017). All participants were under age of 16 and informed written consent to participate was obtained from their parents or legal guardians. Administrative permissions were not required to access the samples and data used in this research.

Consent for publication

Not applicable.

Competing interests

Igor Mokrousov is a Section Editor in one of the BMC Series journals (BMC Microbiology). The authors declare that no other competing interests exist.

Author details

¹Laboratory of Molecular Epidemiology and Evolutionary Genetics, St. Petersburg Pasteur Institute, St. Petersburg 197101, Russia. ²St. Petersburg Research Institute of Phthisiopulmonology, St. Petersburg 191036, Russia. ³Scientific Center for Expert Evaluation of Medical Products, Moscow 127051, Russia. ⁴Present address: Van Andel Institute, Grand Rapids, MI 49503-2518, USA. ⁵Saint Petersburg State University, St. Petersburg 199034, Russia.

Received: 3 June 2020 Accepted: 5 August 2020

Published online: 18 August 2020

References

- World Health Organization. Weekly Epidemiological Record. 2018;93(08):73–96. <https://www.who.int/wer/2018/wer9308/en/>. Accessed 13 Aug 2020.
- Trunz BB, Fine P, Dye C. Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness. *Lancet*. 2006;367:1173–80. [https://doi.org/10.1016/S0140-6736\(06\)68507-3](https://doi.org/10.1016/S0140-6736(06)68507-3).
- Behr MA. BCG - Different strains, different vaccines? *Lancet Infect Dis*. 2002; 2:86–92. [https://doi.org/10.1016/S1473-3099\(02\)00182-2](https://doi.org/10.1016/S1473-3099(02)00182-2).
- Leung AS, Tran V, Wu Z, Yu X, Alexander DC, Gao GF, et al. Novel genome polymorphisms in BCG vaccine strains and impact on efficacy. *BMC Genomics*. 2008;9:413. <https://doi.org/10.1186/1471-2164-9-413>.
- Mostowy S, Tsolaki AG, Small PM, Behr MA. The *in vitro* evolution of BCG vaccines. *Vaccine*. 2003;21:4270–4. [https://doi.org/10.1016/S0264-410X\(03\)00484-5](https://doi.org/10.1016/S0264-410X(03)00484-5).
- Seki M, Udagawa T, Sugawara I, Iwama K, Honda I, Fujita I, et al. The effect of passages during Japanese BCG vaccine production on genetic stability and protective efficacy. *Vaccine*. 2012;30:1460–4. <https://doi.org/10.1016/j.vaccine.2011.12.101>.
- Zhang L, Ru HW, Chen FZ, Jin CY, Sun RF, Fan XY, et al. Variable virulence and efficacy of BCG vaccine strains in mice and correlation with genome polymorphisms. *Mol Ther*. 2016;24:398–405. <https://doi.org/10.1038/mt.2015.216>.
- Brosch R, Gordon SV, Garnier T, Eiglmeier K, Frigui W, Valenti P, et al. Genome plasticity of BCG and impact on vaccine efficacy. *Proc Natl Acad Sci U S A*. 2007;104(13):5596–601. <https://doi.org/10.1073/pnas.0700869104>.

9. Ritz N, Curtis N. Mapping the global use of different BCG vaccine strains. *Tuberculosis*. 2009;89:248–51. <https://doi.org/10.1016/j.tube.2009.03.002>.
10. Abdallah AM, Hill-Cawthorne GA, Otto TD, Coll F, Guerra-Assunção JA, Gao G, et al. Genomic expression catalogue of a global collection of BCG vaccine strains show evidence for highly diverged metabolic and cell-wall adaptations. *Sci Rep*. 2015;5:15443. <https://doi.org/10.1038/srep15443>.
11. Stefanova T. Quality control and safety assessment of BCG vaccines in the post-genomic era. *Biotechnol Biotechnol Equip*. 2014;28:387–91. <https://doi.org/10.1080/13102818.2014.927200>.
12. Levi DT, Obukhov YI, Aleksandrova NV, Volkova RA, Elbert EV, Alvarez Figueroa MV, et al. Identity and stability assessment of BCG vaccine by multiplex PCR. *BIOpreparations. Prev Diagn Treat*. 2016;16(1):49–54.
13. Ludannyy R, Alvarez Figueroa M, Levi D, Markelov M, Dedkov V, Aleksandrova N, et al. Whole-genome sequence of *Mycobacterium bovis* BCG-1 (Russia). *Genome Announc*. 2015;3(6):e01320–15. <https://doi.org/10.1128/genomeA.01320-15>.
14. Cernuschi T, Malvolti S, Nickels E, Friede M. Bacillus Calmette-Guérin (BCG) vaccine: a global assessment of demand and supply balance. *Vaccine*. 2018;36:498–506. <https://doi.org/10.1016/j.vaccine.2017.12.010>.
15. Modipane L, Reva O, Magazi BT, Antiabong JF, Osei Sekyere J, Mbelle NM. Phylogenomic and epidemiological insights into two clinical *Mycobacterium bovis* BCG strains circulating in South Africa. *Int J Infect Dis*. 2019;87:32–8. <https://doi.org/10.1016/j.ijid.2019.08.010>.
16. Venkataraman A, Yusuff M, Liebeschuetz S, Riddell A, Prendergast AJ. Management and outcome of Bacille Calmette-Guérin vaccine adverse reactions. *Vaccine*. 2015;33:5470–4. <https://doi.org/10.1016/j.vaccine.2015.07.103>.
17. Rermruay R, Rungmaitree S, Chatpornvorarux S, Brukesawan C, Wittawatmongkol O, Lapphra K, et al. Clinical features and outcomes of Bacille Calmette-Guérin (BCG)-induced diseases following neonatal BCG Tokyo-172 strain immunization. *Vaccine*. 2018;36(28):4046–53. <https://doi.org/10.1016/j.vaccine.2018.05.098>.
18. Jou R, Huang WL, Su WJ. Tokyo-172 BCG vaccination complications, Taiwan. *Emerg Infect Dis*. 2009;15(9):1525–6. <https://doi.org/10.3201/eid1509.081336>.
19. Lin WL, Chiu NC, Lee PH, Huang ASE, Huang FY, Chi H, et al. Management of Bacillus Calmette-Guérin osteomyelitis/osteitis in immunocompetent children—a systematic review. *Vaccine*. 2015;33:4391–7. <https://doi.org/10.1016/j.vaccine.2015.07.039>.
20. Wansaula Z, Wortham JM, Mindra G, Haddad MB, Salinas JL, Ashkin D, et al. Bacillus Calmette-Guérin cases reported to the national tuberculosis surveillance system, United States, 2004–2015. *Emerg Infect Dis*. 2019;25:451–6. <https://doi.org/10.3201/EID2503.180686>.
21. Richter E, Weizenegger M, Rüsck-Gerdes S, Niemann S. Evaluation of genotype MTBC assay for differentiation of clinical *Mycobacterium tuberculosis* complex isolates. *J Clin Microbiol*. 2003;41:2672–5. <https://doi.org/10.1128/JCM.41.6.2672-2675.2003>.
22. Sevost'yanova TA, Aksenova VA, Sterlikov SA. Epidemiology and monitoring of complications after immunization with BCG/BCG-M in the Russian Federation. *Medicinskij Almanach*. 2019;60(3–4):56–60. <https://doi.org/10.21145/2499-9954-2019-3-56-60>.
23. Choi YY, Han MS, Lee HJ, Yun KW, Shin CH, Yoo WJ, et al. *Mycobacterium bovis* osteitis following immunization with Bacille Calmette-Guérin (BCG) in Korea. *J Korean Med Sci*. 2019;34(1):e3. <https://doi.org/10.3346/jkms.2019.34.e3>.
24. Saldaña NG, Ranero ARDC, Trujillo DMG, Garza EAD I, Tortoriello AIQ, Ruiz BV, et al. Osteitis secondary to BCG vaccine in immunocompetent patients: Three case reports. *Medicine (Baltimore)*. 2019;98:e13871. <https://doi.org/10.1097/MD.00000000000013871>.
25. Aksenova VA, Mushkin AI, Kovalenko KN, Kaz'mina EA, Bakin MN, Isaeva NI, et al. BCG ostitis in children: the epidemiological parameters of some regions of the Russian Federation. *Probl Tuberk Bolezn Legk*. 2007;1:9–12.
26. Korppi M, Teräsjärvi J, Liehu-Martiskainen M, Lauhkonen E, Vuononvirta J, Nuolivirta K, et al. Haplotype of the interleukin 17A gene is associated with osteitis after Bacillus Calmette-Guérin vaccination. *Sci Rep*. 2017;7(1):11691. <https://doi.org/10.1038/s41598-017-12113-z>.
27. Fortin A, Abel L, Casanova JL, Gros P. Host genetics of mycobacterial diseases in mice and men: forward genetic studies of BCG-osis and tuberculosis. *Annu Rev Genomics Hum Genet*. 2007;8:163–92. <https://doi.org/10.1146/annurev.genom.8.080706.092315>.
28. Dagg B, Hockley J, Rigby P, Ho MM. The establishment of sub-strain specific WHO reference reagents for BCG vaccine. *Vaccine*. 2014;32:6390–5. <https://doi.org/10.1016/j.vaccine.2014.09.065>.
29. Markey K, Ho MM, Choudhury B, Seki M, Ju L, Castello-Branco LRR, et al. Report of an international collaborative study to evaluate the suitability of multiplex PCR as an identity assay for different sub-strains of BCG vaccine. *Vaccine*. 2010;28:6964–9. <https://doi.org/10.1016/j.vaccine.2010.08.045>.
30. 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;27:1710–6. <https://doi.org/10.1016/j.vaccine.2009.01.034>.
31. Pan Y, Yang X, Duan J, Lu N, Leung AS, Tran V, et al. Whole-genome sequences of four *Mycobacterium bovis* BCG vaccine strains. *J Bacteriol*. 2011;193:3152–3. <https://doi.org/10.1128/JB.00405-11>.
32. Wada T, Maruyama F, Iwamoto T, Maeda S, Yamamoto T, Nakagawa I, et al. Deep sequencing analysis of the heterogeneity of seed and commercial lots of the Bacillus Calmette-Guérin (BCG) tuberculosis vaccine substrain Tokyo-172. *Sci Rep*. 2015;5:17827. <https://doi.org/10.1038/srep17827>.
33. Zhang W, Zhang Y, Zheng H, Pan Y, Liu H, Du P, et al. Genome sequencing and analysis of BCG vaccine strains. *PLoS One*. 2013;8:e71243. <https://doi.org/10.1371/journal.pone.0071243>.
34. Borgers K, Ou JY, Zheng PX, Tiels P, Van Hecke A, Plets E, et al. Reference genome and comparative genome analysis for the WHO reference strain for *Mycobacterium bovis* BCG Danish, the present tuberculosis vaccine. *BMC Genomics*. 2019;20(1):561. <https://doi.org/10.1186/s12864-019-5909-5>.
35. Voronina OL, Kunda MS, Aksenova EI, Semenov AN, Ryzhova NN, Lunin VG, et al. Mosaic structure of *Mycobacterium bovis* BCG genomes as a representation of phage sequences' mobility. *BMC Genomics*. 2016;17:1009. <https://doi.org/10.1186/s12864-016-3355-1>.
36. Sotnikova EA, Shitikov EA, Malakhova MV, Kostryukova ES, Ilna EN, Otrasheskaya AV, et al. Complete genome sequence of *Mycobacterium bovis* strain BCG-1 (Russia). *Genome Announc*. 2016;4(2):e00182–16. <https://doi.org/10.1128/genomeA.00182-16>.
37. Otrasheskaya E, Vinokurova V, Shitikov E, Sotnikova E, Perevyshina T, Kolchenko S, et al. *M. bovis* BCG-1 (Russia) sub-strain genome stability investigation within the entire production process. *Zh Mikrobiol (Moscow)*. 2018;2:58–67.
38. World Health Organization. The use of next-generation sequencing technologies for the detection of mutations associated with drug resistance in *Mycobacterium tuberculosis* complex: technical guide: World Health Organization; 2018. <https://apps.who.int/iris/handle/10665/274443>. Accessed 13 Aug 2020.
39. Keller PM, Böttger EC, Sander P. Tuberculosis vaccine strain *Mycobacterium bovis* BCG Russia is a natural recA mutant. *BMC Microbiol*. 2008;8:120. <https://doi.org/10.1186/1471-2180-8-120>.
40. Bottai D, Di Luca M, Majlessi L, Frigui W, Simeone R, Sayes F, et al. Disruption of the ESX-5 system of *Mycobacterium tuberculosis* causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation. *Mol Microbiol*. 2012;83(6):1195–209. <https://doi.org/10.1111/j.1365-2958.2012.08001.x>.
41. Saelens JW, Viswanathan G, Tobin DM. Mycobacterial evolution intersects with host tolerance. *Front Immunol*. 2019;10:528. <https://doi.org/10.3389/fimmu.2019.00528>.
42. Sassetti CM, Rubin EJ. Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A*. 2003;100:12989–94. <https://doi.org/10.1073/pnas.2134250100>.
43. Zvi A, Ariel N, Fulkerson J, Sadoff JC, Shafferman A. Whole genome identification of *Mycobacterium tuberculosis* vaccine candidates by comprehensive data mining and bioinformatic analyses. *BMC Med Genet*. 2008;1:18. <https://doi.org/10.1186/1755-8794-1-18>.
44. Van Embden JDA, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol*. 1993;31:406–9. <https://doi.org/10.1097/MD.00000000000013871>.
45. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
46. Li H. Exploring single-sample snp and indel calling with whole-genome de novo assembly. *Bioinformatics*. 2012;28:1838–44. <https://doi.org/10.1093/bioinformatics/bts280>.
47. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–303. <https://doi.org/10.1101/gr.107524.110>.

48. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6:80–92. <https://doi.org/10.4161/fly.19695>.
49. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, et al. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Res*. 2017;45:D535–42. <https://doi.org/10.1093/nar/gkw1017>.
50. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019;47:D506–15. <https://doi.org/10.1093/nar/gky1049>.
51. Kapopoulou A, Lew JM, Cole ST. The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis*. 2011;91:8–13. <https://doi.org/10.1016/j.tube.2010.09.006>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

