International Journal of Infectious Diseases 87 (2019) 32–38

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

Phylogenomic and epidemiological insights into two clinical *Mycobacterium bovis* BCG strains circulating in South Africa

Lesedi Modipane[a], Oleg Reva[b], Beki Themba Magazi[a], John Francis Antiabong[c], John Osei Sekyere[a,*], Nontombi Marylucy Mbelle[a,d]

[a] Department of Medical Microbiology, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa
[b] Department of Biochemistry, Genetics and Microbiology, Centre for Bioinformatics and Computational Biology, University of Pretoria, Pretoria, South Africa
[c] Centre for Infectious Diseases Research and Experimental Therapeutics, Baylor Research Institute, TX, USA
[d] National Health Laboratory Services, Department of Medical Microbiology, Pretoria, South Africa

ARTICLE INFO

Article history:
Received 12 March 2019
Received in revised form 8 August 2019
Accepted 9 August 2019
Corresponding Editor: Eskild Petersen, Aarhus, Denmark

Keywords:
Mycobacterium bovis BCG
Tuberculosis
TB adenitis
BCG disease
Phylogenomics
South Africa

ABSTRACT

Background: Mycobacterium bovis BCG is a live, attenuated tuberculosis vaccine. While the vaccine protects infants from tuberculosis, complications including disseminated infections have been reported following vaccination. Genetically diverse BCG sub-strains now exist following continuous passaging of the original Pasteur strain for vaccine manufacture. This genetic diversity reportedly influences the severity of disseminated BCG infections and the efficacy of BCG immunization.
Methods: M. bovis BCG was isolated from infants suspected of being infected with tuberculosis. The whole genome of the clinical isolates and BCG Moscow were sequenced using Illumina Miseq and the sequences were analysed using CLC Genomics Workbench 7.0, PhyResSE v1.0, and Parsnp.
Results and conclusions: Genetic variations between the clinical strains and the reference BCG Copenhagen were identified. The clinical strains shared only one mutation in a secretion protein. Mutations were identified in various antibiotic resistance genes in the BCG isolates, which suggests their potential as multidrug-resistant (MDR) phenotypes. Phylogenetic analysis showed that the two isolates were distantly related, and the M1_S48 clinical isolate was closely related to *M. bovis* BCG Moscow. The phylogenomics results imply that two different BCG strains may be circulating in South Africa. However, it is difficult to associate the BCG vaccine strain administered and the BCG strain supplied with specific adverse events, as BCGiosis is under-reported. This study presents background genomic information for future surveillance and tracking of the distribution of BCGiosis-associated mycobacteria. It is also the first to report on the genomes of clinical BCG strains in Africa.

Introduction

Mycobacterium bovis bacillus Calmette–Guerin (BCG) is an attenuated live vaccine conferring protection against disseminated forms of tuberculosis (TB) in infants. Besides various complications following vaccination with the BCG vaccine (Movahedi et al., 2011; Ying et al., 2014), the protective efficacy of the vaccine against pulmonary TB is inconsistent, ranging from 0 to 80% (Copina et al., 2014; Pan et al., 2011). The inconsistency in protective efficacy may be associated with the host's immune state, routes of vaccine administration, and possible genetic heterogeneity of circulating BCG strains, resulting from continuous passaging of the original BCG Pasteur strain (Dockrell and Smith, 2017).

The BCG vaccine is derived from virulent *M. bovis* through serial passaging on potato bile media. The BCG vaccine shares biological and genetic properties with virulent *Mycobacterium tuberculosis* and *M. bovis*, creating a risk of reversion to the invasive phenotype (Jia et al., 2017). Pathogenic *M. bovis* and *M. tuberculosis* share genes that are absent in BCG strains (Mahairas et al., 1996; Nieuwenhuizen and Kaufmann, 2018). Additionally, different BCG sub-strains may either have or lack the region of difference two (RD2), and only BCG Pasteur has the tandem duplication 1, DU1 (Li et al., 2015). These studies show that BCG strains are different from pathogenic *Mycobacterium* species and that different BCG strains are phylogenetically related.

M. tuberculosis complex (MTBC) is routinely identified using the MPT64 antigen test and the line probe assay GenoType MTBDR*plus* in our setting, at the National Health Laboratory Services, Tshwane

* Corresponding author.
E-mail addresses: leeprianca92@gmail.com (L. Modipane), oleg.reva@up.ac.za (O. Reva), beki106@gmail.com (B.T. Magazi), negribj@hotmail.com (J.F. Antiabong), u18392467@tuks.co.za (J. Osei Sekyere), nontombi.mbelle@up.ac.za (N.M. Mbelle).

https://doi.org/10.1016/j.ijid.2019.08.010

Academic Division (NHLS/TAD). The GenoType MTBDR*plus* assay simultaneously identifies MTBC, as well as rifampicin (RIF) and isoniazid (INH) resistance, but does not identify MTBC to the species level.

MTBC includes the following *Mycobacterium* species: *M. tuberculosis*, *M. bovis* ssp. *bovis*, *M. bovis* ssp. *caprae*, *Mycobacterium africanum*, *Mycobacterium microti*, *Mycobacterium canetti*, and the vaccine strain, *M. bovis* BCG. The different species can be identified using the GenoType MTBC line probe assay (Neonakis et al., 2007). This line probe assay identifies and differentiates MTBC species based on *gyrB* polymorphisms and the RD1 deletion; the test strips used are coated with specific oligonucleotides targeting *gyrB* polymorphisms and the RD1 deletion (Richter et al., 2003). However, this test is expensive and BCG infections are only diagnosed upon request. In most cases, BCG infections would be missed and reported as TB.

In this study, two clinical BCG vaccine strains were isolated from infants suspected of being infected with TB and BCGiosis, respectively. The isolates were previously confirmed as BCG strains using the Hain MTBC line probe assay (Hain Lifescience, Nehren, Germany) as part of the routine diagnosis at the NHLS/TAD; the results were obtained from the laboratory information system (LIS). The invasiveness of the isolates was suspected to have been caused by mutations in the clinical isolates. This study therefore aimed to identify genetic variations within the clinical isolates and to analyse the phylogenetic relationship between the clinical isolates and other *M. bovis* species. To our knowledge, this is the first study in South Africa and Africa to compare clinical *M. bovis* BCG genomes with reference *M. bovis* BCG genomes.

Methods

Study design and sample description

This was a retrospective, descriptive study of two clinical *M. bovis* BCG isolates. The study was approved by the Research Ethics Committee, University of Pretoria (reference number 59/2017).

The two clinical isolates (designated M1_S48 and M2_S49) were previously characterized as susceptible to RIF and INH, using the GenoType MTBDR*plus* assay (Hain Lifescience, Nehren, Germany). The lyophilized BCG Moscow (designated BCG_S50; Serum Institute of India, India) was used as a live BCG reference. The live BCG Danish (Copenhagen) strain was not available due to a shortage and change of the BCG vaccine strain used in South Africa; hence the BCG Danish genome stored at GenBank was used as a reference.

Phenotypic methods

The two clinical isolates (M1_S48 and M2_S49) and the *M. bovis* BCG Moscow strain (Serum Institute of India, India) were sub-cultured in Lowenstein–Jensen (LJ) (NICD, South Africa) and the Middlebrook 7H9 MGIT (Mycobacterial Growth Indicator Tube) media (Becton, Dickson and Company, Sparks, MD, USA). Isolates that grew on the Middlebrook 7H9 MGIT medium and flagged positive on the Bactec MGIT 960 instrument within 5 days were tested for pyrazinamide (PZA) drug susceptibility using the BD Bactec MGIT 960 system (Becton, Dickson and Company, Sparks, MD, USA). The manufacturers' instructions were followed for all of the tests.

Genomic DNA extraction and sequencing

Genomic DNA (gDNA) was extracted using the Zymo Research Bacterial/Fungal DNA extraction kit (The Epigenetic Company, USA) according to the manufacturer's instructions. The gDNA was extracted from isolates grown on LJ medium and the gDNA libraries were prepared using the Nextera DNA Sample Prep Kit (Illumina Inc., San Diego, CA, USA). Paired-end whole-genome sequencing was done using the Illumina Miseq sequencing platform (Illumina Inc., San Diego, CA, USA) with 100× coverage.

Genomic analysis

The sequence data were analysed using CLC Genomics Workbench 7.0, Mauve 2.4.0 (http://darlinglab.org/mauve/download.html), and the PhyResS v1.0 server (www.phyresse.org/). The genomes of different *Mycobacterium* strains were downloaded from GenBank and PATRIC (https://www.patricbrc.org/) for genomic comparison and phylogenetic analysis.

The genomes of the M1_S48, M2_S49, and BCG_S50 isolates were aligned against BCG Copenhagen using CLC Genomics Workbench 7.0 (Qiagen) and against the *M. tuberculosis* H37Rv reference strain using progressive Mauve, as described by Darling et al. (2004). A probabilistic variant detection algorithm was used to identify genetic variations between genomes. The following parameters were set: the read filter was set to ignore non-specific variation and broken read pairs. The significance threshold was set to a minimum coverage of 10 and variant probability of 95.0, to require the presence of both forward and reverse sequences, ignore variants in non-specific region(s), and require a variant count of 2; the maximum expected variant was set to 2. PhyResSE was used to search for antibiotic resistance mutations, and Parsnp was used to analyse phylogenetic relatedness between the clinical BCG genomes (M1_S48 and M2_S49) and the downloaded reference genomes. The Parsnp command line with the "– C 1000 –c" flag was run to force alignment over collinear regions. The trees generated were viewed with Gingr (https://harvest.readthedocs.io/en/latest/content/gingr.html) and edited with Figtree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

Table 1
Phenotypic pyrazinamide drug susceptibility test results obtained from the BD Bactec MGIT 960 instrument.

Isolate	Drug name	Concentration (μg/ml)	Growth unit	Status
M. tuberculosis H37Rv	Growth control	0	400	
	Pyrazinamide	100	2	Susceptible
M1_S48	Growth control	0	400	
	Pyrazinamide	100	400	Resistant
M2_S49	Growth control	0	400	
	Pyrazinamide	100	400	Resistant
BCG_S50	Growth control	0	400	
	Pyrazinamide	100	400	Resistant

The growth control tubes only contained the isolates being tested. Each drug was added into a separate tube with the tested isolate and incubated along with the growth control. The BD Bactec MGIT 960 instrument monitors and compares growth of the isolates in the drug-containing tubes with the isolates in the growth control tubes and automatically interprets the results.

Table 2
Genetic variation between the M1_S48 clinical isolate from South Africa and the reference BCG Copenhagen.

Region	Type	Reference	Allele	Zygosity	Annotation
40845	SNV	G/A	A/P	Codon 223	Membrane-anchored mycosin
310910	SNV	T/C	G/L	Codon 618	Multidrug transporter
439148	SNV	G/A	G/D	Codon 224	Hypothetical protein
578486	SNV	A/G	G/P	Codon 371	13E12 repeat family protein
578489	Deletion	T/-	G/D	Codon 372	13E12 repeat family protein
578491^578492	Insertion	-/C	G/P	Codon 372	13E12 repeat family protein
578498	SNV	C/G	C/P	Codon 375	13E12 repeat family protein
750575	SNV	C/T		Non-coding	
750610	SNV	C/G		Non-coding	
910739	SNV	G/C	T/G	Codon 142	Hypothetical protein
910745..910746	MNV	GT/CC	G/P	Codon 144	Hypothetical protein
1039504	SNV	T/G	T/D	Codon 265	Hypothetical protein
1044681	SNV	A/G	G/L	Codon 112	Hypothetical protein
1383793	SNV	T/G	C/T	Codon 89	2,4-dienoyl-CoA reductase
1702069..1702071	MNV	CCG/TTA	AC/NT	Codon 901	Hypothetical protein
1702074..1702076	MNV	AGC/GTT	CA/NG	Codon 899	Hypothetical protein
1706189	SNV	C/T		Non-coding	
1717367	SNV	T/C	C/A	Codon 265	DNA-binding protein
2420970	SNV	A/C	T/K	Codon 55	Hypothetical protein
2655975	SNV	A/G	G/L	Codon 22	Secretion protein
2818510	SNV	T/C		Non-coding	
2837412	SNV	C/T	T/V	Codon 48	Hypothetical protein
2959825	SNV	A/C		Non-coding	
2959834	SNV	C/G		Non-coding	
2960142^2960143	Insertion	-/GG		Non-coding	
2960144^2960145	Insertion	-/ATCGC		Non-coding	
3123378^3123379	Insertion	-/A		Non-coding	
3123379^3123380	Insertion	-/T		Non-coding	
3123381	SNV	G/T		Non-coding	
3123383..3123384	Replacement	TG/AACC		Non-coding	
3123386	SNV	G/C		Non-coding	
3123388	SNV	G/C		Non-coding	
3123391	SNV	G/C		Non-coding	
3123394	SNV	C/T		Non-coding	
3123397..3123402	Replacement	GGTGGG/AAAAA		Non-coding	
3123404..3123405	MNV	CG/GT		Non-coding	
3123409..3123410	Replacement	CC/A		Non-coding	
3123412	SNV	C/T		Non-coding	
3123414	SNV	C/A		Non-coding	
3140696..3140697	MNV	CG/GC	T/C	Codon 127	Hypothetical protein
3382235	SNV	G/A	T/A	Codon 308	Ribulose-phosphate 3-epimerase
3684593	SNV	G/A	G/V	Codon 1390	DEAD/DEAH box helicase
3737307	SNV	G/C	G/L	Codon 816	Hypothetical protein
3737309	SNV	G/A	G/F	Codon 816	Hypothetical protein
3737313	SNV	A/G	G/F	Codon 814	Hypothetical protein
3737315..3737316	MNV	AG/CA	G/V	Codon 814	Hypothetical protein
3737322..3737323	MNV	CC/AA	T/F	Codon 811	Hypothetical protein
3738624	SNV	G/C	C/T	Codon 377	Hypothetical protein
3739001	SNV	G/A	T/L	Codon 252	Hypothetical protein
3741178	SNV	G/C		Non-coding	
3741188	SNV	T/G		Non-coding	
3746466	SNV	T/C		Non-coding	
3754823	SNV	G/C		Non-coding	
3756268	SNV	G/T		Non-coding	
3756270	SNV	G/A		Non-coding	
3756276	SNV	G/C		Non-coding	
3756278..3756279	MNV	CC/TG		Non-coding	
3756283	SNV	G/C		Non-coding	
3756285	SNV	C/G		Non-coding	
3888768..3888769	MNV	CC/GG	C/W	Codon 57	Hypothetical protein
3920728^3920729	Insertion	-/G	G/A	Codon 234	PE-PGRS family protein
3920731	SNV	A/G	C/A	Codon 235	PE-PGRS family protein
3920737..3920738	MNV	TC/GG	C/A	Codon 237	PE-PGRS family protein
3920741	SNV	A/G	T/A	Codon 238	PE-PGRS family protein
3920746..3920747	MNV	AC/GG		Non-sense mutation	PE-PGRS family protein
3920749	SNV	A/G	C/A	Codon 241	PE-PGRS family protein
3920752..3920753	MNV	GC/AA	G/T	Codon 242	PE-PGRS family protein
3920756..3920757	Replacement	AT/G	G/A	Codon 243	PE-PGRS family protein
3932779	SNV	G/C		Non-coding	
4071249	SNV	A/C	C/A	Codon 170	Hypothetical protein
4132102	SNV	C/G	C/G	Codon 509	2-isopropylmalate synthase

A (allele), alanine; A (reference), adenine; C (allele), cysteine; C (reference), cytosine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G (allele), glycine; G (reference), guanine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; MNV, multiple nucleotide variant; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; SNV, single nucleotide variant; T, threonine; W, tryptophan; Y, tyrosine; V, valine.

Results and discussion

M. bovis and its derivative *M. bovis* BCG vaccine are intrinsically resistant to PZA (Ritz et al., 2009). As expected, the M1_S48, M2_49, and BCG Moscow in this study were also resistant to PZA (Table 1).

The draft genome sequences of the two isolates, M1_S48 and M2_S49, were 2 767 203 bp and 1 432 828 bp, respectively. These genome sizes are significantly smaller than the previously sequenced BCG genomes (Borgers et al., 2019); the reasons for this are not known, although it may suggest that some genes are missing in these isolates, making their genomes smaller. Annotation with the NCBI Prokaryotic Genome Annotation Pipeline identified 3999 coding genes, 3999 coding CDS, 51 RNA (3 rRNA, 45 tRNA, 3 ncRNA), 187 pseudogenes, and three CRISPR arrays in the genome of M1_48. In the genome of M2_49, 4027 coding genes, 51 RNA genes (3 rRNA, 45 tRNA, 3 ncRNA), 205 pseudogenes, and two CRISPR arrays were identified.

Genomic analysis using CLC Genomics Workbench 7.0 identified variations in several protein coding genes in the clinical isolates and the reference BCG Copenhagen genome. The M1_S48 genome had mutations in 23 non-coding sequences and 48 coding sequences including hypothetical proteins (23.9%), proline glutamate polymorphic GC-rich repetitive sequences (PE-PGRs, 11.3%), membrane anchored mycosin (11.4%), and several other functional genes (Table 2). The M2_S49 genome had two mutations (Table 3). Both isolates shared a mutation causing K → G amino acid substitution in the secretion protein, believed to be the ESX-5 protein (CLC Genomics Workbench 7.0). *M. tuberculosis* complex consists of five type IIV or ESX (ESX-1 to ESX-5) secretion systems. ESX-1, ESX-3, and ESX-5 are associated with protein secretion and it was suspected that one of them could be the secretion protein identified with mutation in this study (Ates et al., 2015). ESX-1 is encoded by the RD1 region and this region is deleted in all *M. bovis* BCG sub-strains and was therefore ruled out. ESX-3 plays a role in the acquisition of metal ions and survival of *M. tuberculosis* (Knudsen et al., 2014). There are, however, no data on ESX-3 associated with *M. bovis* BCG. On the other hand, ESX-5 plays a role in outer membrane permeability and has been shown to play an essential role in *M. bovis* BCG growth (Ates et al., 2015).

As expected, when compared to BCG Copenhagen, the BCG Moscow genome had more genetic variability (Supplementary material, Table S1). Different BCG sub-strains are genetically diverse, and it has been shown that BCG Moscow (also known as BCG Russia) is distantly related to BCG Copenhagen (Abdallah et al., 2015; Copina et al., 2014). This explains the genetic variability observed between these two strains.

The membrane-anchored mycosin genes encode a family of subtilisin-like proteases, designated MycP1–MycP5 (Brown et al., 2000; Fang et al., 2016). These genes comprise a part of the ESX system that is involved in the virulence of MTBC and they are also present in BCG strains. Only the mycosins MycP2, MycP3, and MycP5 have previously been identified in BCG strains and they are believed to be essential for the growth of the bacteria. Additionally, MycP5 reportedly facilitates the secretion of the PE/PPE (proline–glutamate/proline–proline–glutamate) proteins. Mutations in these genes may therefore interfere with mycobacterial growth and secretions of the PE/PPE proteins (Chen, 2016; Fang et al., 2016). The PE-PGRs are a subfamily of the PE proteins. These proteins contain conserved sequences at the N-terminus and polymorphic repeat units at the C-terminus. Reports suggest that the PE-PGR proteins elicit an immune response and are essential for mycobacterial pathogenesis (Tiwari et al., 2012). The role of the PE-PGRs in BCG virulence is not well known.

The RD1 region was found to be missing in M1_S48, M2_S49, and BCG_S50, when aligned against *M. tuberculosis* H37Rv (Figure 1). This result therefore confirms the clinical isolates to be *M. bovis* BCG.

Analysis with PhyResSE identified various resistance mutations in the clinical isolates, with both isolates having PZA resistance mutations in the *pncA* gene. The BCG M2_S49 isolate had PZA resistance mutations in *pncA*: His57Asp (cac/gac). Previous studies have also reported a His57Asp (cac/gac) mutation in *M. bovis* and *M. bovis* BCG, which is associated with intrinsic PZA resistance (Feuerriegel et al., 2014; Jureen et al., 2008). BCG M1_S48 had the following antibiotic resistance-conferring mutations: Ser450Leu (tcg/ttg) in *rpoB* (RIF resistance); Ile194Thr (atc/acc) in *inhA* and Arg268His (cgc/cac) in *ndh* (INH and ethionamide resistance); Gly97Asp (ggt/gat) in *pncA* (PZA resistance); and Met306Val (atg/gtg) in *embB* (ethambutol resistance). This suggests that the strain could be multidrug-resistant.

For M1_S48, the MTBDR*plus* assay and genomic analysis showed discordant drug susceptibility results. As mentioned above, both isolates were susceptible to INH and RIF according to the GenoType MTBDR*plus* assay, despite identifying mutations in the *rpoB* and *inhA* genes, which confer resistance to RIF and INH, respectively. Routine drug susceptibility testing at the NHLS, Pretoria is done using the GenoType MTBDR*plus* line probe assay. The assay uses strips that are coated with probes specific for some regions of the *inhA* and *rpoB* genes, and mutations occurring outside of the included regions cannot be detected. Failing to detect these mutations would result in false-negative results and negatively impact patients, as incorrect drug regimens would be prescribed. Incorrect drug regimens have also been associated with the spread of drug-resistant TB strains (Lange et al., 2018). To improve the prescription of correct drug regimens, the results of the MTBDR*plus* assay should be validated with the phenotypic drug-susceptibility testing (DST) method, and more resistance-mediating regions should be added to test strips to improve the diagnosis of drug-resistant strains. Moreover, whole genome-based diagnostics for TB should be adopted in the long-term in South Africa, as has been done recently in the UK (Cabibbe et al., 2018), to improve the efficient diagnosis of MTBC and drug resistance.

Confirmatory phenotypic drug susceptibility testing for RIF, INH, and ethambutol could not be performed to confirm the resistance genotype in this study. The isolates failed to grow when sub-cultured from storage (at −70 °C). This is therefore a limitation of the study.

The phylogenetic analysis showed that M1_S48 was more closely related to the BCG Moscow strain (Serum Institute of India, India); these strains clustered together on the phylogenetic tree (Figure 2). These two strains also branched from the same ancestor as the East African Indian strain (*M. tuberculosis* EAI5/NITR2060),

Table 3
Genetic variation between the M2_S49 clinical isolate from South Africa and the reference BCG Copenhagen.

Region	Type	Reference	Allele	Zygosity	Annotation
528445	SNV	G/A	G/H	Codon 24	MerR family transcriptional regulator
2655975	SNV	A/G	G/L	Codon 22	Secretion protein

A, adenine; G (allele), glycine; G (reference), guanine; H, histidine; L, leucine; SNV, single nucleotide variant.

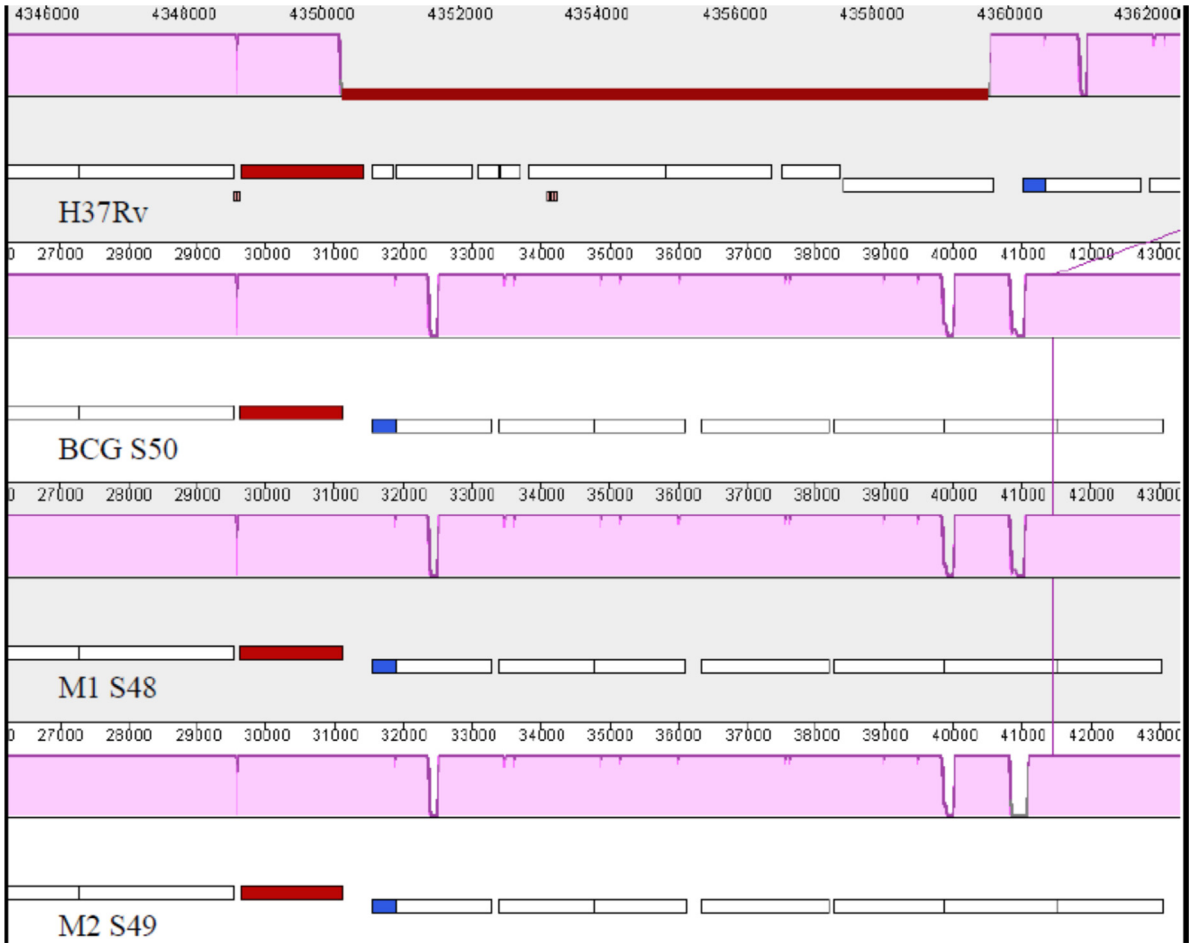

Figure 1. RD1 deletion in sequenced *Mycobacterium bovis* BCG strains. Genome sequence alignment was performed using Mauve 2.4.0. Reference sequence *Mycobacterium tuberculosis* H37Rv in the top line is followed by the aligned sequences of BCG_S50, M1_S48, and M2_S49. Pink colour histograms show the level of sequence similarity in the sequenced genomes to the reference genome H37Rv. Gene locations are depicted by boxes. The genes Rv3871 and Rv3880c flanking the RD1 deletion are shown in red and blue colours, respectively. RD1 sequences in H37Rv are highlighted with a thick brown line. The RD1 sequence spans the C-terminal part of Rv3871 and eight other genes of H37Rv, which are absent in *M. bovis* BCG strains used for vaccination (Kurenuma et al., 2009).

attesting to the homogeneity of the MTBC species. Surprisingly, the M1_S49 and BCG Moscow clade was more closely related to MTB and *M. bovis* strains than to the other BCG strains isolated from different countries. It was expected that the BCG Moscow and the M1_S48 isolate would branch from the same root as the other BCG strains, and not from the East African Indian strain. In contrast, M2_S49 was distantly related to both M1_S48 and the *M. bovis* BCG Moscow. The M2_S49 clustered with the other BCG strains, including BCG Glaxo, and branched with BCG Russia, Pasteur, Tokyo, Prague, and several other BCG strains from different countries. Previous studies have shown that BCG Danish (BCG Copenhagen) clusters around BCG Glaxo, which is closely related to a cluster of BCG Prague (Abdallah et al., 2015; Bottai and Brosch, 2016). This suggests that M2_S49 is closely related to the BCG Danish strain. As expected *M. bovis* isolated from different animals in different countries branched out from the same root in different clades.

Few cases of BCGiosis have been documented, making it difficult to build an adequate sample size. Similarly, most cases of BCGiosis are not reported, making it difficult to follow up on the clinical outcomes associated with genetic variations of BCG strains. The mutated genes identified in this study have not been reported previously as virulence factors, and this study did not include immunological assays to identify the impact of the mutation on the patient. In addition, it was not possible to follow up on the clinical outcomes of the infected patients, limiting the clinical application of this study. It is also worrying that two genetically diverse BCG strains are circulating in South Africa simultaneously, because adverse events cannot be attributed to a specific strain. This study thus presents valuable unprecedented information on the genomic characteristics of BCGiosis in South Africa and lays a foundation for further exploration in this field.

In conclusion, the clinical BCG isolates were genetically distinct when compared to BCG Copenhagen. Various mutations were identified in coding sequences in the form of deletions, insertions, and single nucleotide variations. Both isolates shared a mutation causing a substitution of glycine to lysine amino acid in the ESX-5 secretion protein. Isolate, M1_S49 contained multiple mutations in antibiotic resistance genes, which would potentially complicate treatment.

The two isolates were phylogenetically distinct, clustering far apart when compared to global BCG isolates. The study results imply that two different BCG strains may be circulating in South Africa, but as BCGiosis is under-reported, it is difficult to associate the BCG vaccine strain administered and the BCG strain supplied with specific adverse events. Whole-genome sequencing was used successfully to identify two of the BCG vaccine strains circulating in

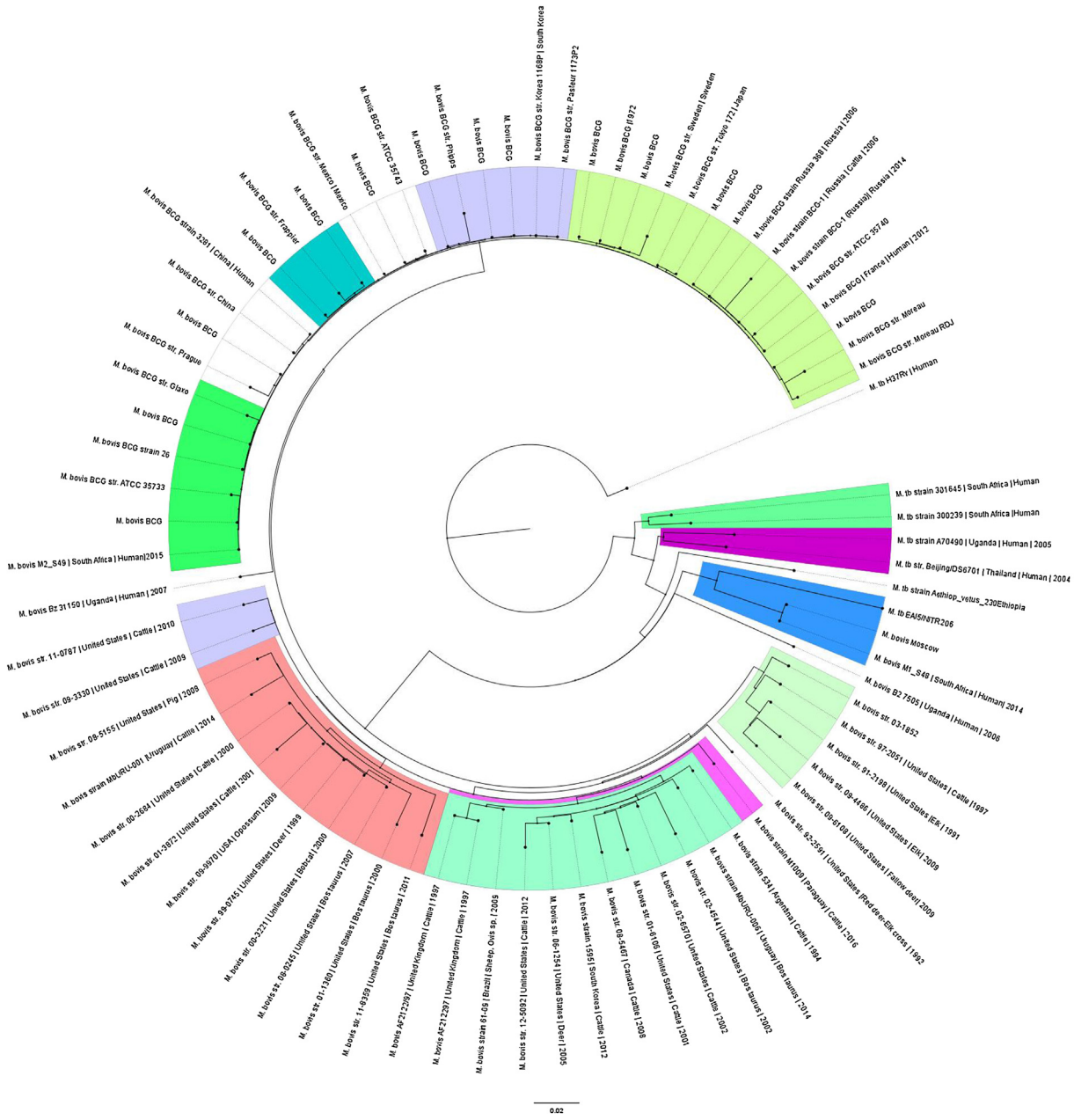

Figure 2. Phylogenetic relationship between the clinical *Mycobacterium bovis* BCG vaccine strains and the reference *Mycobacterium* strains. *M. bovis* Moscow represents the *M. bovis* BCG Moscow (BCG_S50; Serum Institute of India, India). *Mycobacterium tuberculosis* H37Rv was used as a reference genome for all of the *M. tuberculosis* complex strains. The clinical strains M1_S48 and M2_S49 are distantly related; M1_S48 clustered with *M. bovis* BCG Moscow and the MTB EAM strains. M2_S49 clustered with *M. bovis* BCG Glaxo and other *M. bovis* BCG strains from different countries.

South Africa and this tool is recommended in routine diagnosis to improve MTBC and resistance detection. The data gleaned from this study could be useful for molecular tracking of the distribution of these BGC strains in South Africa.

Funding

This study was funded by an NHLS grant (grant number 94498).

Conflict of interest

The authors declare no conflict of interest.

Accession number(s)

This whole-genome shotgun project has been deposited in GenBank under accession numbers **CP033310** (SRA: SRS3974045) and **CP033311** (SRA: SRS3974046) for BCG_S48 and BCG_S49, respectively, under BioProject ID PRJNA498011. The versions described in this paper are the first versions **CP033310.1** and **CP033311.1**.

Acknowledgements

We would like to acknowledge the Department of Microbiology, National Health Laboratory Services/Tshwane Academic

Division for providing the BCG isolates and funding the study, the Department of Medical Microbiology, University of Pretoria for providing facilities to conduct the study, and the National Institute of Communicable Diseases for sequencing the BCG isolates. We would also like to offer our gratitude to Mrs O.O. Onwegbuna for her assistance with sample collection and her laboratory expertise.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ijid.2019.08.010.

References

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.

Ates LS, Ummels R, Commandeur S, van der Weerd R, Sparrius M, Weerdenburg E, et al. Essential role of the ESX-5 secretion system in outer membrane permeability of pathogenic mycobacteria. PLoS Genet 2015;11(5)e1005190, doi:http://dx.doi.org/10.1371/journal.pgen.1005190.

Borgers K, Ou JY, Zheng PX, 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;20:561, doi:http://dx.doi.org/10.1186/s12864-019-5909-5.

Bottai D, Brosch R. The BCG strain pool: diversity matters. Mol Ther 2016;24:201–3.

Brown GD, Dave JA, van Pittius NCG, Stevens L, Ehlers MRW, Beyers AD. The mycosins of *Mycobacterium tuberculosis* H37Rv: a family of subtilin-like proteases. Gene 2000;254:147–55.

Cabibbe Andrea Maurizio, Trovato Alberto, De Filippo Maria Rosaria, Ghodousi Arash, Rindi Laura, Garzelli Carlo, et al. Countrywide implementation of whole genome sequencing: an opportunity to improve tuberculosis management, surveillance and contact tracing in low incidence countries. Eur Respir J 2018;(January)1800387, doi:http://dx.doi.org/10.1183/13993003.00387-2018.

Chen JM. Mycosins of mycobacterial type VII ESX secretion system: the glue that holds the party together. mBio 2016;7: e02062-16.

Copina R, Coscollá M, Efstathiadis E, Gagneux S, Ernsta J. Impact of in vitro evolution on antigenic diversity of *Mycobacterium bovis* bacillus Calmette-Guerin (BCG). Vaccine 2014;32:5998–6004, doi:http://dx.doi.org/10.1016/j.vaccine.2014.07.11.

Darling Aaron CE, Mau Bob, Blattner Frederick R, Perna Nicole T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004;14(7):1394–403.

Dockrell HM, Smith SG. What have we learnt about BCG vaccination in the last 20 years?. Front Immunol 2017;8:1134, doi:http://dx.doi.org/10.3389/fimmu.2017.01134.

Fang Z, Schubert W, van Pittius NCG. Expression and production of soluble *Mycobacterium tuberculosis* H37Rv mycosin-3. Biochem Biophys Rep 2016;5:448–52, doi:http://dx.doi.org/10.1016/j.bbrep.2016.02.005.

Feuerriegel S, Köser CU, Niemann S. Phylogenetic polymorphisms in antibiotic resistance genes of the *Mycobacterium tuberculosis* complex. J Antimicrob Chemother 2014;69:1205–10.

Jia X, Yang L, Dong M, Chen S, Lv L, Cao D, et al. Comparative genomics of *Mycobacterium tuberculosis* complex (MTBC) provides insight into dissimilarities between intraspecific groups differing in host association, virulence, and epitope diversity. Front Cell Infect Microbiol 2017;7:88.

Jureen P, Werngren J, Toro J, Hoffner S. Pyrazinamide resistance and *pncA* gene mutations in *Mycobacterium tuberculosis*. J Antimicrob Chemother 2008;1852–4, doi:http://dx.doi.org/10.1128/AAC.00110-08.

Knudsen NP, Nørskov-Lauritsen S, Dolganov GM, Schoolnik GK, Lindenstrøm T, Andersen P, et al. Tuberculosis vaccine with high predicted population coverage and compatibility with modern diagnostics. PNAS 2014;111(3):1096–101.

Kurenuma Takeshi, Kawamura Ikuo, Hara Hideki, Uchiyama Ryosuke, Daim Sylvia, Dewamitta Sita Ramyamali, et al. The RD1 locus in the *Mycobacterium tuberculosis* Genome contributes to activation of caspase-1 via induction of potassium ion efflux in infected macrophages. Infect Immun 2009;77(August (9)):3992–4001, doi:http://dx.doi.org/10.1128/IAI.00015-09.

Li X, Chen L, Zhu Y, Yu X, Cao J, Wang R, et al. Genomic analysis of a *Mycobacterium bovis* bacillus Calmette-Guérin strain isolated from an adult patient with pulmonary tuberculosis. PLoS One 2015;10(4)e0122403.

Lange C, Chesov D, Heyckendorf J, Leung CC, Udwadia Z, Dheda K. Drug-resistant tuberculosis: an update on disease burden, diagnosis and treatment. Respirology 2018;23(7):656–73.

Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. J Bacteriol 1996;178:1274–82.

Movahedi Z, Norouzi S, Mamishi S, Rezaei N. BCGiosis as a presenting feature of a child with chronic granulomatous disease. Braz J infect Dis 2011;15:83–6, doi:http://dx.doi.org/10.1590/S1413-86702011000100016.

Neonakis IK, Gitti Z, Petinaki E, Maraki S, Spandidos DA. Evaluation of the GenoType MTBC assay for differentiating 120 clinical *Mycobacterium tuberculosis* complex isolates. Eur J Clin Microbiol Infect Dis 2007;26:151–2, doi:http://dx.doi.org/10.1007/s10096-007-0255-y.

Nieuwenhuizen NE, Kaufmann SH. Next-generation vaccines based on bacille Calmette–Guérin. Front Immunol 2018;9:121.

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, doi:http://dx.doi.org/10.1128/JB.00405-11.

Richter E, Weizenegger M, Rüsch-Gerdes S, Niemann Stefan. Evaluation of genotype MTBC assay for differentiation of clinical *Mycobacterium tuberculosis* complex isolates. J Clin Microbiol 2003;41:2672–5, doi:http://dx.doi.org/10.1128/JCM.41.6.2672–2675.

Ritz N, Tebruegge M, Connell TM, Sievers A, Robins-Browne R, Curtis N. Susceptibility of *Mycobacterium bovis* BCG vaccine strains to antituberculous antibiotics. 2009.

Tiwari BM, Kannan N, Vemu L, Raghunand TR. The *Mycobacterium tuberculosis* PE proteins Rv0285 and Rv1386 modulate innate immunity and mediate bacillary survival in macrophages. PLoS One 2012;7:e51686, doi:http://dx.doi.org/10.1371/journal.pone.0051686.

Ying W, Sun J, Liu D, Hui X, Yu Y, Wang J, et al. Clincal characteristics and immunogentics of BCGosis/BCGitis in Chines children: a year follow-up study. PLoS One 2014;9:e94485, doi:http://dx.doi.org/10.1371/journal.pone.0094485.