JOURNAL OF BACTERIOLOGY, June 2011, p. 3152–3153 Vol. 193, No. 12
0021-9193/11/$12.00 doi:10.1128/JB.00405-11


# Whole-Genome Sequences of Four *Mycobacterium bovis* BCG Vaccine Strains▽

Yuanlong Pan,[1,2] Xi Yang,[1,2] Jia Duan,[1,2] Na Lu,[1] Andrea S. Leung,[3] Vanessa Tran,[3] Yongfei Hu,[1] Na Wu,[1,2] Di Liu,[1] Zhiming Wang,[1,2] Xuping Yu,[4] Chen Chen,[5] Yuanyuan Zhang,[5] Kanglin Wan,[5] Jun Liu,[3]* and Baoli Zhu[1]*

*Microbial Genome Research Center, CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, People's Republic of China[1]; Graduate University of the Chinese Academy of Sciences, Beijing 100049, People's Republic of China[2]; Department of Molecular Genetics, University of Toronto, 1 King's College Circle, Toronto, Ontario M5S 1A8, Canada[3]; Department of Veterinary Medicine, College of Animal Sciences, Zhejiang University, Hangzhou 310029, People's Republic of China[4]; and State Key Laboratory for Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, People's Republic of China[5]*

Received 24 March 2011/Accepted 31 March 2011

***Mycobacterium bovis* Bacille Calmette-Guérin (BCG) is the only vaccine available against tuberculosis (TB). A number of BCG strains are in use, and they exhibit biochemical and genetic differences. We report the genome sequences of four BCG strains representing different lineages, which will help to design more effective TB vaccines.**

*Mycobacterium bovis* Bacille Calmette-Guérin (BCG) was derived from *M. bovis* by continuous *in vitro* passaging from 1908 to 1921 (12). Distribution and widespread use of BCG started around 1924 and were accompanied by further *in vitro* passaging until 1960s. The *in vitro* evolution of BCG has resulted in a number of BCG substrains that are heterogenic (12, 17). The protective efficacy of BCG against pulmonary TB varies from 0 to 80%, and the heterogeneity of BCG strains is thought to be one of the contributing factors (1).

A molecular genealogy of BCG based on genomic deletions and duplications has been established (2, 4, 10). More than a dozen BCG strains were placed into four major groups (4). Thus far, the complete genome sequence has been determined for only two BCG strains, BCG-Pasteur (4) and BCG-Tokyo (15). Here we present draft genome sequences of four BCG strains obtained by using a whole-genome shotgun strategy.

BCG-China, BCG-Denmark 1331 (ATCC 35733), BCG-Russia (ATCC 35740), and BCG-Tice (ATCC 35743) were described previously (2, 10). Genomic DNA was sequenced with an Illumina genome analyzer. The genome coverages were 63.2-, 65.6-, 59.1-, and 70.7-fold for BCG-China, -Danish, -Russia, and -Tice, respectively. The pair-end reads were assembled by SOAPdenovo (11). Nearly 500 gaps were filled by multiplex PCR and primer walking, which yielded 29, 32, 36, and 28 contigs for the four strains, respectively.

Annotation was done using MetaGeneAnnotator (14), tRNAscan-SE 1.21 (13), RNAmmer 1.2 (9), and Tandem Repeats Finder 4.04 (3). In addition, the contigs were searched against the KEGG (8), Pfam (6), COGs (16), and NCBI NR protein databases.

The draft genome sequences of four BCG strains have similar sizes, approximately 4.27 Mb, and all have about 4,030 predicted genes, with a G+C content of 65%. Each genome has a single copy of predicted 5S, 16S, and 23S rRNA genes and 45 copies of predicted tRNAs genes. The repeat regions of each genome vary from 638 (BCG-Russia) to 679 (BCG-Tice). Genes annotated by the COGs database can be classified into 21 COG (clusters of orthologous groups) categories.

Comparative genomic analyses were performed using the genome sequence of *M. bovis* (7) and *M. tuberculosis* H37Rv (5) as references. More than 350 single-nucleotide polymorphisms (SNPs) between BCG and *M. bovis* genomes were identified. About 170 SNPs were specific to BCG strains (not present in *M. bovis* or *M. tuberculosis*), which likely contributes to the attenuation of BCG. Each BCG strain contains dozens of SNPs not shared by other BCG strains. Consistent with previous studies (4, 10), we confirmed the presence of tandem duplications: DU-I in BCG-Russia, DU-III in BCG-China and BCG-Denmark, and DU-IV in BCG-Tice. Future studies of these genomic polymorphisms will help to identify molecular factors that influence BCG clinical properties, including safety, immunogenicity, and protective efficacy, and provide novel insights for the rational design of the next generation of TB vaccines.

**Nucleotide sequence accession numbers.** The whole-genome shotgun sequences have been deposited at DDBJ/EMBL/GenBank under the accession codes AEZE00000000, AEZF00000000,

* Corresponding author. Mailing address for Jun Liu: Department of Molecular Genetics, University of Toronto, 1 King's College Circle, Toronto, Ontario M5S 1A8, Canada. Phone: (416) 946-5067. Fax: (416) 978-6885. E-mail: jun.liu@utoronto.ca. Mailing address for Baoli Zhu: Microbial Genome Research Center, CAS Key Lab of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, NO.1 West Beichen Road, Chaoyang District, Beijing 100101, People's Republic of China. Phone: 86-10-64807362. Fax: 86-10-64807438. E-mail: zhubaoli@im.ac.cn.

▽ Published ahead of print on 8 April 2011.

AEZG00000000, and AEZH00000000 for BCG-China, -Danish, -Russia, and -Tice, respectively.

This work was supported by funding from the National Basic Research Program of China (973 Program: 2009CB522605 and 2007CB513002 to B.Z.), the Scientific Program for Key Infectious Diseases, The Ministry of Science and Technology (2009ZX10601 to B.Z.), and funding from Canadian Institutes of Health Research (CIHR) (MOP-489636 and MOP-482204 to J.L.).

## REFERENCES

1. **Behr, M. A.** 2002. BCG—different strains, different vaccines? Lancet Infect. Dis. **2:**86–92.
2. **Behr, M. A., et al.** 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science **284:**1520–1523.
3. **Benson, G.** 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic. Acids Res. **27:**573–580.
4. **Brosch, R., et al.** 2007. Genome plasticity of BCG and impact on vaccine efficacy. Proc. Natl. Acad. Sci. U. S. A. **104:**5596–5601.
5. **Cole, S. T., et al.** 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature **393:**537–544.
6. **Finn, R. D., et al.** 2009. The Pfam protein families database. Nucleic. Acids Res. **38:**D211–D222.
7. **Garnier, T., et al.** 2003. The complete genome sequence of Mycobacterium bovis. Proc. Natl. Acad. Sci. U. S. A. **100:**7877–7882.
8. **Kanehisa, M., S. Goto, S. Kawashima, Y. Okuno, and M. Hattori.** 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res. **32:**D277–D280.
9. **Lagesen, K., et al.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35:**3100–3108.
10. **Leung, A. S., et al.** 2008. Novel genome polymorphisms in BCG vaccine strains and impact on efficacy. BMC Genomics **9:**413.
11. **Li, R., et al.** 2010. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. **20:**265–272.
12. **Liu, J., V. Tran, A. S. Leung, D. C. Alexander, and B. L. Zhu.** 2009. BCG vaccines Their mechanisms of attenuation and impact on safety and protective efficacy. Hum. Vaccin. **5:**70–78.
13. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic. Acids Res. **25:**955–964.
14. **Noguchi, H., T. Taniguchi, and T. Itoh.** 2008. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA Res. **15:**387–396.
15. **Seki, M., et al.** 2009. Whole genome sequence analysis of Mycobacterium bovis bacillus Calmette-Guérin (BCG) Tokyo 172: A comparative study of BCG vaccine substrains. Vaccine **27:**1710–1716.
16. **Tatusov, R. L., M. Y. Galperin, D. A. Natale, and E. V. Koonin.** 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. **28:**33–36.
17. **Walker, K. B., et al.** 2010. The second Geneva Consensus: Recommendations for novel live TB vaccines. Vaccine **28:**2259–2270.