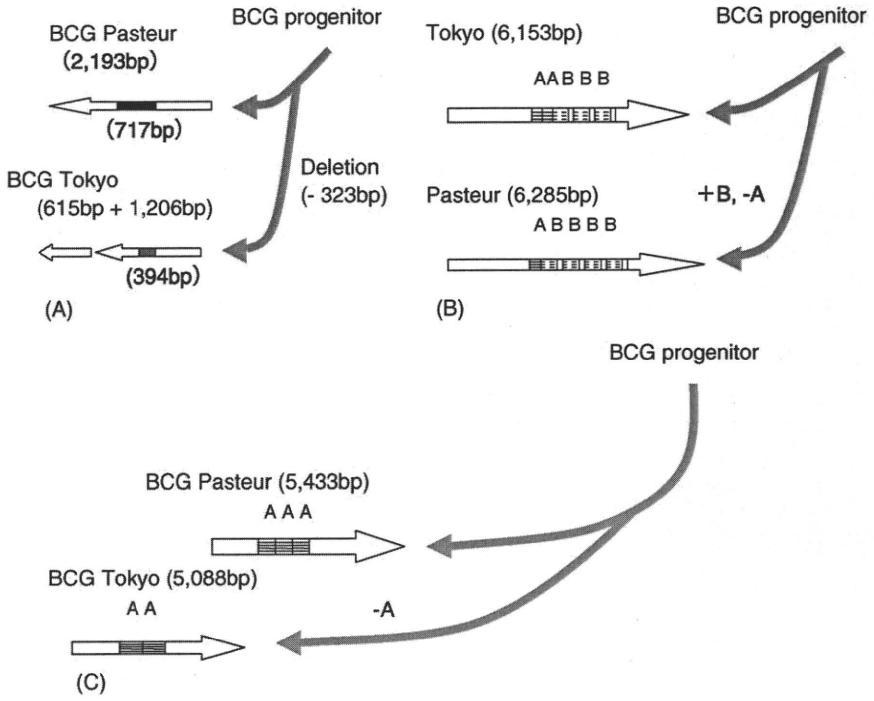

Fig. 2. Possible process of PE_PGRS mutation. (A) PE_PGRS20. (B) PE_PGRS54, repeat of A (465-bp), repeat of B (552-bp). (C) PE_PGRS55, repeat of A (345-bp).

As 345 bp is the length of a repeat in the gene, the difference reflected the copy number of both BCG strains (Fig. 2C). The PE_PGRS55 gene also shows polymorphism among *M. tuberculosis* complex.

3.5. Tandem duplications (DU1 and DU2)

There was no tandem duplication of DU1 in BCG Tokyo like that present in BCG Pasteur. However, DU2 (20,704 bp) existed in BCG Tokyo with a different length and position from that of BCG Pasteur, as reported previously [14]. BCG Tokyo had three copies of the region from the *astB* gene (*Rv3299c*) to the *sdhC* (*Rv3316*) and *sdhD* genes (*Rv3317*), including 20 genes. In the copied region, the *astB* gene was truncated.

3.6. RD2

RD2 (including 11 genes) existed in BCG Tokyo, as it was previously reported to be present in early substrains and absent in later substrains [18]. The sequence homology of the 11 genes (*Rv1978–Rv1988*) were 100% with those of *M. bovis* or *M. tuberculosis*. The 11 genes included those encoding for the transcriptional regulator (*Rv1985c*) and immunogenic protein MPT64 (*Rv1980c*). The sequence of *Rv1978* in BCG Tokyo was same as that of *M. tuberculosis*, but the sequences of BCG Tokyo and *M. tuberculosis* had a SNP compared with that of *M. bovis*. The other sequences of 10 genes were same as those of *M. bovis*.

3.7. RD14

Although RD14 (including 10 genes) existed in BCG Tokyo, this region was deleted in BCG Pasteur. The sequences of 10 genes (*Rv1765c–Rv1773c*) in BCG Tokyo showed 100% homology with those of *M. bovis* or *M. tuberculosis*. These 10 genes included one coding for the transcriptional regulator (*Rv1773c*). The sequence homology of PE_PGRS31 (*Rv1768*) was 100% with that of *M. bovis*, but there was an 18-bp insertion and a SNP in *M. tuberculosis* H37Rv. As a result of the 18-bp insertion, 6 aa were inserted into the protein and one amino acid (N207H) was replaced by the SNP. The sequence homology of BCG Tokyo (*Rv1773c*) was 100% with that of *M. tuberculosis*, but showed a SNP compared with *M. bovis*. The other sequences of 8 genes showed 100% homology with those of *M. bovis*.

3.8. N-RD18

N-RD18 contained three intact genes (*sigI*, *Rv1190*, and *Rv1191*) in BCG Tokyo. *Rv1190* was deleted and *sigI* and *Rv1191* were fused in BCG Pasteur (N-RD20), so the function of *sigI* seemed to be lost [20]. The *sigI* gene was reported to have a role in adaptation to cold shock [31]. The sequence of BCG Tokyo (*Rv1190*) showed 100% homology with that of *M. tuberculosis*, but showed a SNP compared with *M. bovis*. The other 2 sequences of BCG Tokyo (*sigI*, *Rv1191*) showed 100% homology with those of *M. tuberculosis* and *M. bovis*.

3.9. Rv3405c

BCG Tokyo had a 22-bp deletion in *Rv3405c* (possible transcriptional regulatory protein) of RD16. There were two different types of colonies (S: smooth and R: rough) when BCG Tokyo was cultured on Middlebrook 7H10 medium. PCR of each type of colony showed a strong relationship between colony morphology and genotype, since 98.7% of S-colonies had the 22-bp deletion (type I) and 95.9% of R-colonies did not (type II) [25]. In every Tokyo 172 preparation studied, S-colonies (type I) exceeded 90% of the total. Accordingly, it is probable that Tokyo 172 predominantly has the type I genotype [25].

3.10. Comparison of BCGs Tokyo and Pasteur (differences of less than 20 bp)

There were 19 genes and one intergenic region that differed between BCG Tokyo and Pasteur due to insertion or deletion (ins/del) of less than 20 bp. These 20 genes were compared between *M. tuberculosis* and BCG Tokyo or Pasteur. As a result, 15 genes had ins/del mutations between *M. tuberculosis* and BCG Tokyo, and 12 genes had ins/del mutations between *M. tuberculosis* and BCG Pasteur (Fig. 3, SD Table 2).

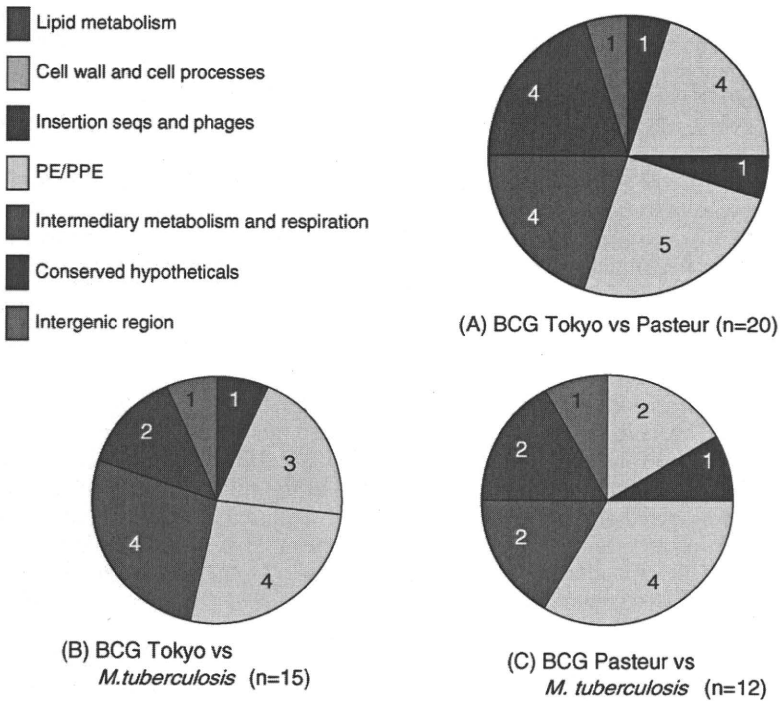

Fig. 3. Functional classification of 20 ins/del (<20 bp) mutations between BCG Tokyo and Pasteur (A), as well as between BCG Tokyo and *M. tuberculosis* (B), and between BCG Pasteur and *M. tuberculosis* (C).

Compared with other genomes, ins/del mutations of 8 genes [*acs* (−1-bp), *ftsW* (−1-bp), *rpfE* (−1-bp), PE_PGRS24 (−9-bp), *sdhA* (+1-bp), *Rv3814c* (+3-bp), *Mb3263c* and *Mb3359c* (*RvD5*)] were found in BCG Tokyo only, and those of 5 genes [*Rv3835* (+1-bp), *Rv1313* (IS1557), *PE_PGRS7* (−9-bp), *Rv1486c* (+1-bp), and *Rv3433c* (−3-bp)] were found in BCG Pasteur only.

When the sequences near ins/del mutations of BCG Tokyo and Pasteur were analyzed, 13 microsatellite polymorphisms (also known as simple sequence repeats comprising tandem repeat motifs of 1–6 bp in length) were found between the two BCG substrains (SD Table 3), as previously reported between *M. tuberculosis* and *M. bovis* [32].

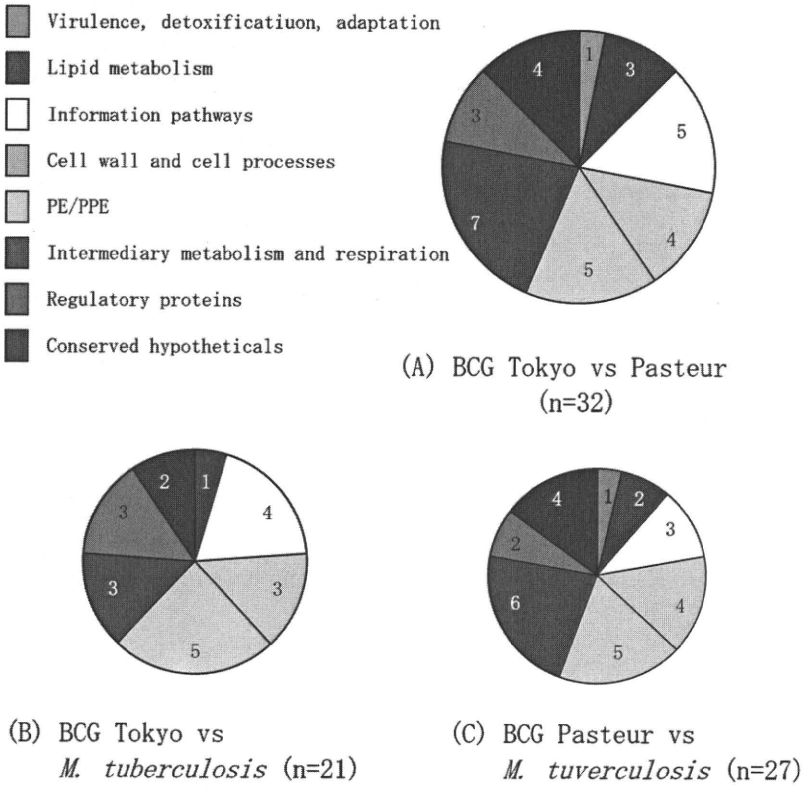

Fig. 4. Functional classification of 32 SNPs mutations between BCG Tokyo and Pasteur (A), as well as SNPs or ins/del mutations between BCG Tokyo and *M. tuberculosis* (B), and SNPs or ins/del mutations between BCG Pasteur and *M. tuberculosis* (C).

When the function of genes with ins/del was assessed, the PE or PPE families were over-represented in each BCG substrain relative to the frequency in the individual genome (BCG Tokyo: 4/15 versus PE/PPE in BCG Tokyo genome of 170/3996, odds ratio = 8.1, $P < 0.01$; Fisher's exact, BCG Pasteur: 4/12 versus PE/PPE in BCG Pasteur of 168/4002, odds ratio = 11.4, $P < 0.01$; Fisher's exact). These results show that the PE_PGRS genes are unstable and ins/del mutation occurs more easily than in other genes of these BCG substrains.

There were 68 SNPs between BCG Tokyo and Pasteur. A total of 56 SNPs were found in 43 genes, in which 37 genes (86%) had nonsynonymous mutations (leading to amino acid substitution) and 6 genes had synonymous mutations, so the majority of these mutations related to SNPs were nonsynonymous. As 32 of the 37 genes with nonsynonymous mutations had no ins/del mutations, these genes were compared. Of the 32 genes with nonsynonymous mutations, 21 of BCG Tokyo and 27 of BCG Pasteur had SNPs or ins/del mutations as compared to the genome with SNPs of *M. tuberculosis* (Fig. 4, SD Table 4).

When the genes with SNPs were assessed on the basis of function, the PE or PPE families were over-represented in each BCG substrain relative to the frequency in the individual genome (BCG Tokyo: 5/21 versus the PE/PPE in BCG Tokyo genome of 170/3996, odds ratio = 6.9, $P < 0.01$; Fisher's exact, BCG Pasteur: 5/27 versus PE/PPE in the BCG Pasteur genome of 168/4002, odds ratio = 5.2, $P < 0.01$; Fisher's exact).

Compared with other genomes, SNP mutations of 5 genes [*rplE*, *pcaA*, *typA*, *Rv3401* and *Rv3583*] were found in BCG Tokyo only, and SNP mutations of 11 genes [*hrcA*, *echA3*, *mmaA3*, *sigK*, *Rv2571*, *narJ*, *pepN*, *ilvN*, *Icd1*, *Rv0552* and *Rv3258*] were found in BCG Pasteur only.

4. Discussion

The BCG substrains are considered to have evolved mainly by gene deletion and gene amplification (DU1 and DU2). The results of this study demonstrated several new VNTRs and PE_PGRS polymorphisms. As polymorphisms are known to exist among other *M. tuberculosis* complex, not only BCG Tokyo and Pasteur but also other BCG substrains might have such polymorphisms frequently.

The PE_PGRS gene family also shows variation among clinical isolates of *M. tuberculosis*, and it might play a role in the variability of antigenicity and persistency of infection [33,34]. Certain proteins are localized on the surface of BCG and may influence interactions between BCG cells and host macrophages [35]. When the genes showing ins/del mutations or SNPs in BCG Tokyo and Pasteur were compared with those of their individual genomes, the PE or PPE families were over-represented in both BCG substrains relative to their genomic frequency. Differences of PE_PGRS gene expression between BCG Tokyo and Pasteur have also been reported [20]. Therefore, mutations of these genes and differences in the level of expression might play a role in the different protective effects of the BCG substrains [36,37].

To examine the genetic characteristics of BCG Sofia from seed lots and commercial batches, VNTR typing was done with six alleles, and a profile identical to that of other BCG substrains was obtained [38]. The 5 alleles that were newly shown to vary between BCG Tokyo and Pasteur in this study may be useful to identify BCG substrains and to examine genetic stability during vaccine production.

The entire genomic sequence of BCG Pasteur was determined previously, but the sequences of RD2 and RD14 were unknown because these regions were absent in BCG Pasteur. In the present study, it was revealed that the genes of RD2 and RD14 in BCG Tokyo were almost 100% identical with those of *M. bovis* AF2122/97. The two regions contained a total of 21 genes, including two genes encoding regulatory proteins. Five genes were classified into functional category 3 (cell wall and cell processes) and an immunogenic protein MPB64 was included in RD2. Therefore, the existence or deletion of these regions might affect the phenotype of BCG, but further study will be necessary to clarify this issue.

Only a few nonsynonymous SNPs are known to affect the phenotype of BCG substrains, such as SNPs of the *mma3*, *sigK* [39], and *crp* [40,41] genes. In the present study, there were no nonsynonymous SNPs in the start or stop codons among the new SNPs, but 32 nonsynonymous SNPs were found between BCGs Tokyo and Pasteur, including those of the *mma3*, *sigK*, and *crp* genes. The effect of other SNPs on the phenotype is unknown.

There have been no reports about ins/del mutations (<20 bp) of BCG substrains, except for a 10-bp deletion of the PhoR gene in DU-2-III substrains [20]. However, 19 genes with ins/del mutations were detected between BCG Tokyo and Pasteur, with 13 being located in microsatellite regions. Eight genes (*acs*, *ftsW*, *rpfE*, PE_PGRS24, *sdhA*, *Rv3814c*, *Mb3263c* and *Rv3583*) showed ins/del in BCG Tokyo. The *acs* gene, which encodes acetyl-CoA synthetase, had a 1 bp deletion in codon 209 and this mutation causes frame shift in the region of the putative AMP-binding domain signature (from codon 260 to codon 271). Acetyl-CoA synthetase is involved in the pyruvate pathway and pyruvate is an initial substrate of the TCA cycle. The *sdhA* gene encodes a subunit of the succinate dehydrogenase complex and is related to the TCA cycle which *sdhB*, *sdhC* and *sdhD*. The *sdhB* gene of BCG Tokyo had 2 SNPs compared with that of *M. tuberculosis*, while *sdhC* and *sdhD* were present in triplicate at the DU2 region. Such these mutations might affect the TCA cycle, but further study is necessary to clarify this issue.

BCG Tokyo had a specific 22-bp deletion within *Rv3405c*, and a strong relationship has been demonstrated between colony morphology and genotype [25]. Therefore, *Rv3405c* might have an important influence on colony morphology. Chen et al. recently showed that BCG Tokyo does not produce phthiocerol dimycocerosates (PDIMs) and phenolic glycolipids (PGLs), two cell wall lipids that are known to be important for the virulence of *Mycobacterium tuberculosis*. They reported that *Rv3405c* is involved in PGL biosynthesis and that deletion of this gene is partially responsible for defective PGL synthesis in BCG Tokyo [42]. Although they did not describe colony morphology, there might be a relationship between PGL biosynthesis and colony morphology. The morphology of BCG colonies has been suggested to be related to cell wall components such as mycoside B [43]. An investigation is now in progress to determine the relationship between colony morphology and cell wall glycolipids (PDIMs/PGLs) in the two genotypes of BCG Tokyo.

The genomic and phenotypic differences demonstrated between early and later BCG substrains by our comparative study possibly also exist between other early and later BCG substrains. These findings may be useful with the respect to the standardization of BCG vaccine substrains and for more the precise understanding of the genotypic and phenotypic differences between early and later substrains.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.vaccine.2009.01.034.

References

[1] Petroff SA, Branch A, Steenken Jr W. A study of Bacillus Calmette–Guérin (BCG). Am Rev Tuberc 1929;19:9–46.
[2] Behner DM. The stability of the colony morphology and pathogenicity of BCG. Am Rev Tuberc 1935;31:174–202.
[3] Osborn TW. Changes in BCG strains. Tuberculosis 1983;64:1–13.
[4] Harboe M, Nagai S, Patarroyo ME, Torres ML, Ramirez C, Cruz N. Properties of proteins MPB64, MPB70, and MPB80 of *Mycobacterium bovis* BCG. Infect Immun 1986;52(1):293–302.

[5] Wiker HG, Nagai S, Hewinson RG, Russell WP, Harboe M. Heterogeneous expression of the related MPB70 and MPB83 proteins distinguish various substrains of *Mycobacterium bovis* BCG and *Mycobacterium tuberculosis* H37Rv. Scand J Immunol 1996;43(4):374–80.
[6] Behr MA, Schroeder BG, Brinkman JN, Slayden RA, Barry III CE. A point mutation in the mma3 gene is responsible for impaired methoxymycolic acid production in *Mycobacterium bovis* BCG strains obtained after 1927. J Bacteriol 2000;182(12):3394–9.
[7] Hesselberg I. Drug resistance in the Swedish/Norwegian BCG strain. Bull World Health Org 1972;46:503–7.
[8] Buriánková K, Doucet-Populare F, Dorson O, Gondran A, Ghnassia JC, Weiser J, et al. Molecular basis of intrinsic marolide resistance in the *Mycobacterium tuberculosis* complex. J Antimicrob Chemother 2004;48(1):143–50.
[9] Grange JM, Gilson JA. Strain to strain variation in the immunogenicity of BCG. Dev Biol Stand 1986;58:37–41.
[10] Lagranderie MR, Balanzuc AM, Deriaud E, Leclerc CD, Gheorghiu M. Comparison of immune responses of mice immunized with five different *Mycobacterium bovis* BCG vaccine strains. Infect Immun 1996;64(1):1–9.
[11] Davids V, Hanekom WA, Mansoor N, Gamieldien H, Gelderbloem SJ, Hawkridge A, et al. The effect of Bacille Calmette–Guérin vaccine strain and route of administration on induced immune responses in vaccinated infants. JID 2006;193:531–6.
[12] Wu B, Huang C, Garcia L, Ponce de Leon A, Osornio JS, Bobadilla-del-Valle M, et al. Unique gene expression profiles in infants vaccinated with different strains of *Mycobacterium bovis* Bacille Calmette–Guérin 2007;75(7):3658–64.
[13] Takeya K, Nomoto K, Muraoka S, Shimotori S, Taniguchi T, Miyake T. Growth of two strains of *Mycobacterium bovis* (BCG) in athymic mice. J Gen Microbiol 1977;100:403–5.
[14] Fomukong NG, Dale JW, Osborn TW, Grange JM. Use of gene probes based on the insertion sequence IS986 to differentiate between BCG vaccine strains. J Appl Microbiol 1992;72:126–33.
[15] Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 1998;393:537–44.
[16] Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, et al. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. J Bacteriol 2002;184(19):5479–90.
[17] Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, et al. The complete genome sequence of *Mycobacterium bovis*. PNAS 2003;100(13):7877–82.
[18] Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, et al. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science 1999;284:1520–3.
[19] Mostowy S, Tsolaki AG, Small PM, Behr MA. The in vitro evolution of BCG vaccines. Vaccine 2003;21(27–30):4270–4.
[20] Brosch R, Gordon SV, Garnier T, Eiglmeier K, Frigui W, Valenti P, et al. Genome plasticity of BCG and impact on vaccine efficacy. PNAS 2007;104(13):5596–601.
[21] Frothingham R, Meeker-O'Connell A. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology 1998;144:1189–96.
[22] Smittipat N, Billamas P, Palittapongarnpim M, Thong-On A, Temu MM, Thanakijcharoen P, et al. Polymorphism of variable-number tandem repeats at multiple loci in *Mycobacterium tuberculosis*. J Clin Microbiol 2005;43(10):5034–43.
[23] Brosch R, Gordon SV, Buchrieser C, Pym AS, Gamier T, Cole ST. Comparative genomics uncovers large tandem chromosomal duplications in *Mycobacterium bovis* BCG Pasteur. Yeast 2000;17:111–23.
[24] Bedwell J, Kairo SK, Behr MA, Bygraves JA. Identification of substrains of BCG vaccine using multiplex PCR. Vaccine 2001;19(15–16):2146–51.
[25] Honda I, Seki M, Ikeda N, Yamamoto S, Yano I, Koyama A, et al. Identification of two subpopulations of bacillus Calmette–Guérin (BCG) Tokyo 172 substrain with different RD16 regions. Vaccine 2006;24(23):4969–74.
[26] Sassetti CM, Boyd DH, Rubin EJ. Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol 2003;48:77–84.
[27] Magdalena J, Supply P, Locht C. Specific differentiation between *Mycobacterium bovis* BCG and virulent strains of the *Mycobacterium tuberculosis* complex. J Clin Microbiol 1998;36(9):2471–6.
[28] Parish T, Smith DA, Roberts G, Betts J, Stoker NG. The *senX3-regX3* two-component regulatory system of *Mycobacterium tuberculosis* is required for virulence. Microbiology 2003;149:1423–35.
[29] Manabe YC, Hatem CL, Kesavan AK, Durack J, Murphy R. Both *Corynebacterium diphtheriae* DtxR (E175K) and *Mycobacterium tuberculosis* IdeR (D177K) are dominant positive repressors of IdeR-regulated genes in *M. tuberculosis*. Infect Immun 2005;73(9):5988–94.
[30] Chanchaem W, Palittapongarnpim PA. A variable number of tandem repeats result in polymorphic alpha-isopropylmalate synthase in *Mycobacterium tuberculosis*. Tuberculosis 2002;82:1–6.
[31] Manganelli R, Dubnau E, Tyagi S, Kramer FR, Smith I. Differential expression of 10 sigma factor genes in *Mycobacterium tuberculosis*. Mol Microbiol 1999;31(2):715–24.
[32] Sreenu VB, Kumar P, Nagaraju J, Nagarajaram HA. Microsatellite polymorphism across the *M. tuberculosis* and *M. bovis* genomes: implication on genome evolution and plasticity. BMC Genomics 2006;7:78.
[33] Banu S, Honoré N, Saint-Joanis B, Philpott D, Prévost M-C, Cole ST. Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? Mol Microbiol 2002;44(1):9–19.
[34] Dheenadhayalan V, Delogu G, Sanguinetti M, Fadda G, Brennan MJ. Variable expression patterns of *Mycobacterium tuberculosis* PE_PGRS genes: evidence that PE_PGRS16 and PE_PGRS26 are inversely regulated in vivo. J Bacteriol 2006;188(10):3721–5.
[35] Brennan M, Delogu G, Chen Y, Bardarov S, Kriakov J, Alavi M, et al. Evidence that Mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. Infect Immun 2001;69(12):7326–33.
[36] Brewer TF, Colditz GA. Relationship beween Bacille Calmette–Guérin (BCG) strains and the efficacy of BCG vaccine in the prevention of tuberculosis. CID 1995;20:126–35.
[37] Behr MA. Correlation between BCG genomics and protective efficacy. Lancet Infect 2002;2:86–92.
[38] Stefanova T, Chouchkova M, Hinds J, Butcher PD, Inwald J, Dale J, et al. Genetic composition of *Mycobacterium bovis* BCG substrain Sofia. J Clin Microbiol 2003;41(11):5349.
[39] Charlet D, Mostowy S, Alexander D, Sit L, Wiker HG, Behr MA. Reduced expression of antigenic proteins MPB70 and MPB83 in *Mycobacterium bovis* BCG strains due to a start codon mutation in *sigK*. Mol Microbiol 2005;56(5):1302–13.
[40] Spreadbury CL, Pallen MJ, Overton T, Behr MA, Mostowy S, Spiro S, et al. Point mutations in DNA- and cNMP-binding domains of the homologue of the cAMP receptor protein (CRP) in *Mycobacterium bovis* BCG: implications for the inactivation of a global regulator and strain attenuation. Microbiology 2005;151:547–56.
[41] Bai G, Gazdik MA, Schaak DD, McDonough KA. The *Mycobacterium bovis* BCG cyclic AMP receptor-like protein is a functional DNA binding protein in vitro and in vivo, but its activity differs from that of its *M. tuberculosis* ortholog, Rv3676. Infect Immun 2007;75(11):5509–17.
[42] Chen JM, Islam ST, Ren H, Liu J. Differential production of lipid virulence factors among BCG vaccine strains and implications on BCG safety. Vaccine 2007;25(48):8114–22.
[43] Abou-Zeid C, Rook GAW, Minnikin DE, Parlett JH, Osborn TW, Grange JM. Effect of the method of preparation of Bacille Calmette–Guérin (BCG) vaccine on the properties of four daughter strains. J Appl Bacteriol 1987;63:449–53.