Comparative Genomics of BCG Vaccines by Whole-Genome DNA Microarray

M. A. Behr,[1]*† M. A. Wilson,[2]† W. P. Gill,[2] H. Salamon,[2]
G. K. Schoolnik,[2] S. Rane,[2] P. M. Small[2]

Bacille Calmette-Guérin (BCG) vaccines are live attenuated strains of *Mycobacterium bovis* administered to prevent tuberculosis. To better understand the differences between *M. tuberculosis*, *M. bovis*, and the various BCG daughter strains, their genomic compositions were studied by performing comparative hybridization experiments on a DNA microarray. Regions deleted from BCG vaccines relative to the virulent *M. tuberculosis* H37Rv reference strain were confirmed by sequencing across the missing segment of the H37Rv genome. Eleven regions (encompassing 91 open reading frames) of H37Rv were found that were absent from one or more virulent strains of *M. bovis*. Five additional regions representing 38 open reading frames were present in *M. bovis* but absent from some or all BCG strains; this is evidence for the ongoing evolution of BCG strains since their original derivation. A precise understanding of the genetic differences between closely related Mycobacteria suggests rational approaches to the design of improved diagnostics and vaccines.

Tuberculosis is an ancient scourge of humankind, caused by the bacterium *M. tuberculosis* and infrequently by other subspecies of the *M. tuberculosis* complex, such as *M. bovis*. The anticipation of 80 million cases of tuberculosis in the coming decade, an increasing proportion of which are likely to be drug-resistant, has revived efforts to develop a new vaccine (*1*). The current vaccine was originally developed by Calmette and Guérin, who passaged a strain of *M. bovis* 230 times in vitro between 1908 and 1921. The resulting vaccine was thought to have struck a balance between reduced virulence and preserved immunogenicity (*2*). However, because of the inability to preserve viable bacteria (such as by freezing), this live vaccine required continued passage, eventually resulting in a profusion of phenotypically different daughter strains that are collectively known as BCG.

By the time lyophilized seed lots of BCG vaccines were created in the 1960s, these vaccines had been separately propagated through about 1000 additional passages (depending on the daughter strain), usually under the very conditions that effected the original attenuation. We previously hypothesized that pressures to minimize adverse effects and maintain tuberculin reactivity during this time resulted in impotent vaccines that consistently induce tuberculin sensitivity with immunization (*3*). If true, such evolution could in part explain why BCG efficacy has varied considerably in human trials. Moreover, such BCG strains would contain antigens that cross-react with the commonly applied diagnostic test for latent tuberculous infection.

To understand the genetic basis of this progression, one would ideally compare contemporary BCG vaccines to their progenitor strain. Because this strain was lost during World War I, the origin of current BCG vaccines can only be inferred through an evolutionary approach (*4*). We therefore curated a collection of BCG daughter strains representing this global dissemination for the purpose of performing genomic comparisons (*5*).

The recent determination of the genomic nucleotide sequence of *M. tuberculosis* H37Rv provides a framework for the genomic analysis of BCG strains (*6*), in that differences between H37Rv and BCG comprise (i) differences between *M. tuberculosis* H37Rv and virulent *M. bovis* and (ii) differences between virulent *M. bovis* and BCG strains. We have used the genomic sequence of H37Rv to assemble a DNA microarray (*7*) representing nearly all open reading frames (ORFs) of H37Rv (*8*) and have used this microarray to perform parallel comparative hybridizations between *M. tuberculosis* H37Rv and *M. bovis* BCG strains (*9*).

Because both the nucleotide sequence of individual genes and the order of genes within the whole genome of *M. tuberculosis* and

M. bovis BCG-Pasteur are extremely similar, one can expect equivalent hybridization by *M. tuberculosis* and BCG strains for loci that are equally present in both organisms (*10, 11*). On the other hand, a mismatched hybridization signal is likely to reflect nonequivalent representation in the genome because of either a repeated element or a relative deletion in the BCG strain.

Our DNA microarray had 4896 spots, representing 3902 of the 3924 ORFs of *M. tuberculosis* H37Rv (99.4%); comparisons of H37Rv and BCG strains provided interpretable data for an average of 3756 ORFs (96%) (*12*). The results can be graphically represented as pseudocolorized spots, or mapped to their respective genomic locations, to highlight ORFs with unequal hybridization ratios (Fig. 1). In this example, BCG-Danish 1331 is shown to have significantly weaker hybridization than H37Rv for the large domain encompassing ORFs Rv1963c to Rv1988. We have used microarray-based comparisons to screen for ORFs deleted from BCG strains and then performed confirmatory analysis with polymerase chain reaction (PCR)–based sequencing across deleted regions (*13*). Using this strategy, we have documented 16 regions deleted in BCG strains varying in length from 1903 to 12,733 base pairs (bp) (Table 1). Four of these regions have been previously described (*14, 15*) and we have extended the nomenclature of the former description to name our deletion regions RD1 to RD16. Of the 16 deletion regions, nine are missing from BCG and all virulent *M. bovis* strains tested, two (representing prophages) are missing from BCG and some of the strains of *M. bovis*, one is missing from all BCG strains, and four are missing only from certain BCG strains.

The identification of nine regions (61 ORFs) that are present in *M. tuberculosis* and consistently absent from *M. bovis* strains (including BCG) may provide some insights into the phenotypic differences between *M. bovis* and *M. tuberculosis*. Pulmonary disease caused by *M. tuberculosis* and *M. bovis* are clinically, radiographically, and pathologically indistinguishable. However, *M. bovis* appears to have a diminished propensity to reactivate and spread from person to person (*16*). Of particular note is the absence from *M. bovis* of a cluster of three phospholipase C genes (*plcA, -B,* and *-C*), containing the only two genes (of the 61 described here) for which empirical evidence exists for both antigenicity and enzymatic function (*17, 18*). Given that phospholipase C activity contributes to the virulence of the opportunistic lung pathogen *Pseudomonas aeruginosa*, it is attractive to speculate that the *M. tuberculosis plcA-plcC* gene cluster manifests a similar, clinically important phenotype. However, before attributing species-specific roles to these

[1]Division of Infectious Diseases, Department of Medicine, McGill University Health Centre, Montreal H3G 1A4, Canada. [2]Division of Infectious Diseases and Geographic Medicine, Department of Medicine, Stanford University, Stanford, CA 94305, USA.

*To whom correspondence should be addressed. E-mail: mbgq@musica.mcgill.ca
†These authors contributed equally to this report.

deletions, a better understanding of the phenotypic relevance of genes missing from *M. bovis* will require concentrated biochemical and genetic analysis of each of these 61 deleted ORFs. It must be stressed that the use of H37Rv sequence as a reference excludes studies of *M. bovis* regions missing from *M. tuberculosis*, including possible *M. bovis*–specific virulence determinants. Therefore, the analysis presented here is unidirectional and represents only a subset of the differences between these organisms.

This collection of ORFs absent from *M. bovis* isolates may prove to be of considerable practical utility. Currently available diagnostics, such as the tuberculin skin test, are unable to distinguish between individuals who have been infected with *M. tuberculosis* and those who have been vaccinated with BCG. This distinction may be possible with an assay based on antigens derived from these *M. tuberculosis*–specific genes. The availability of a diagnostic specific for wild-type infection could greatly facilitate the introduction of vaccination in countries that have a low prevalence of disease and may permit more discriminating treatment of latent tuberculous infections in countries using BCG.

To address the differences specific to BCG, we assumed that regions of H37Rv present in *M. bovis* strains and absent only from BCG were deleted during the derivation and maintenance of BCG vaccines in various vaccine facilities around the world. If this is true, then regions of *M. bovis* missing from BCG strains would indicate unidirectional genetic events from which a phylogeny of BCG strains can be inferred. Because some BCG daughter strains were obtained directly from the Institut Pasteur while other strains were derived from another vaccine facility (for example, BCG Connaught came from BCG Frappier in 1948), it is possible to reconstruct the genealogy of BCG strains and determine when and where BCG-specific deletions occurred (Fig. 2) (*4*). The deletions described in this report can be superimposed on the historical record, demonstrating a unique use of genomic analysis to describe a half-century of in vitro bacterial evolution (*19*). In comparison with *M. bovis*, all BCG vaccines lack one region (RD1) that presumably was lost during the 1908–1921 attenuation (*14*). Another deletion (RD2) occurred at the Institut Pasteur between 1927 and 1931. A further deletion (RD14) specific to BCG-Pasteur indicates an event after Aronson's receipt of BCG-Pasteur 575 in 1938 (*20*) and before the lyophilization of BCG-Pasteur 1173 in 1961 (*21*). The losses of RD8 in Montréal (between 1937 and 1948) and RD16 in Uruguay or Brazil (after 1925) indicate that ongoing evolution of BCG strains was not confined to the Institut Pasteur.

A historical review of the BCG literature reveals reports of decreasing virulence in the Institut Pasteur at various times, consistent with (but not necessarily directly caused by) the documented ongoing evolution. During the original derivation between 1908 and 1921, RD1 was lost and concurrently Calmette described an attenuation of virulence for animals (*22*). Later, between 1927 and 1931, RD2 was lost and various investigators reported decreased vaccine lesions in humans (*23*) and reduced virulence in animals (*24*). As a basis for speculation on the impact of RD14 in humans, one can compare descriptions from the vaccine trials using BCG-Pasteur before and after this deletion. In the American Indian trial,

which used BCG-Pasteur 575, 75% of participants developed a draining abscess at the vaccine site (*20*). In Madras, India, where BCG-Pasteur 1173 was used, abscesses at the vaccination sites were rare (*25*). However, caution must be taken in inferring causality between these documented genetic events and the evolving phenotype of BCG strains. For example, similar results were obtained in the Madras trial with BCG-Danish 1331, a strain that had not suffered any detectable deletion after being obtained from the Institut Pasteur in 1931.

To explore the potential association between BCG deletions and progressive attenuation of virulence, we analyzed the predicted function of the 38 ORFs deleted specifi-

Fig. 1. (A) Scanning fluorimetric representation of a whole-genome DNA microarray comparison of genomic DNA from *M. tuberculosis* H37Rv (red) and *M. bovis* BCG-Danish 1331 (green). Yellow fluorescence indicates equivalent hybridization by both strains. Red fluorescence indicates unopposed hybridization by *M. tuberculosis*, indicating deletion of relevant ORFs in BCG strain. **(B)** Microarray hybridization results comparing H37Rv to BCG-Denmark as displayed by the ProbeBrowser software, which maps microarray results to genomic location. For each spot on the array, the logarithm of the hybridization ratio of H37Rv (Cy3) to BCG-Denmark (Cy5) is displayed in blue marks on the y axis (such that ORFs not present in BCG are displayed above the mean value). The location of ORFs within the entire 4.4-megabase genome maps on the x axis. Regions where more than two contiguous ORFs have a log ratio exceeding 2 standard deviations above the mean are indicated by red bars, and these putative deletions were examined with PCR as described (*13*). **(C)** The scalable software permits an expanded view of one genomic region with each array spot similarly represented in blue on x and y coordinates, but now with the length of the region probed reflected by the length of the blue lines. PCR-based investigation confirmed two deletions, RD15 and RD2, separated by 335 bp. The probe browser software is freely available at http://molepi.stanford.edu/free_software.html.
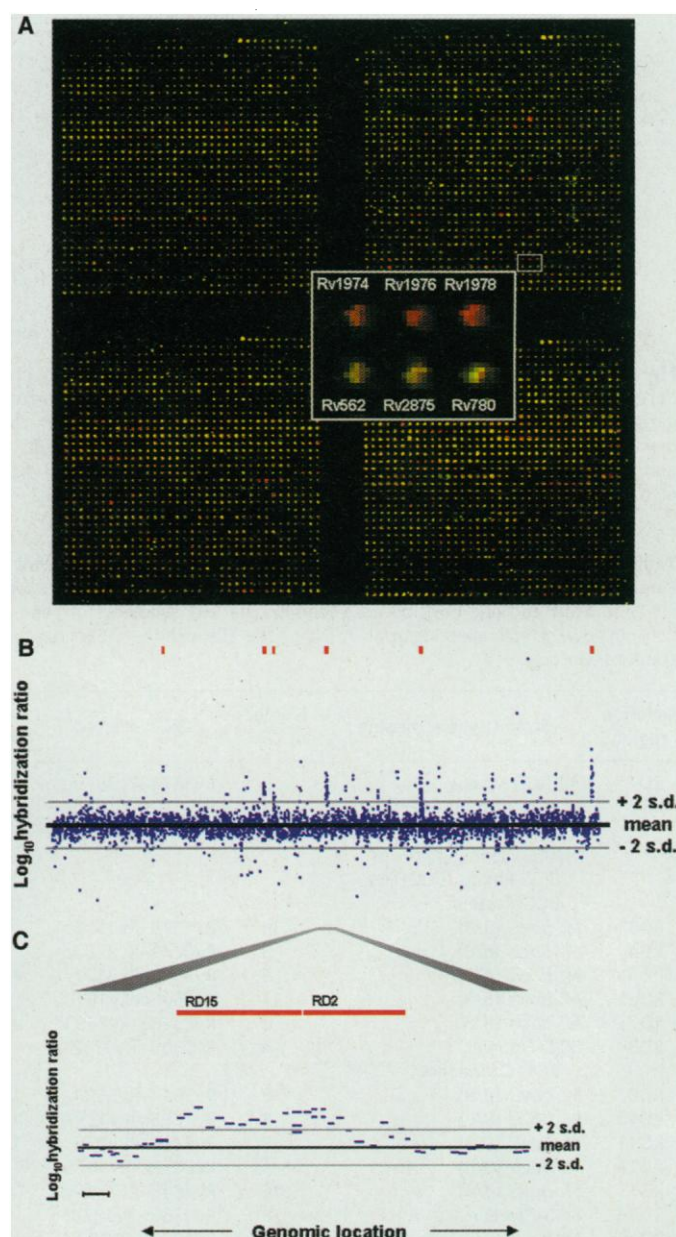
cally from BCG strains by means of the hierarchy of gene function proposed by Cole and colleagues (26). None of the ORFs present in *M. bovis* but deleted from BCG strains were classified as a virulence element, and the gene function predicted by homology search spanned a variety of functions. ORFs classified as transcriptional regulators ("repressors/activators") were overrepresented in BCG deletions relative to their frequency in the H37Rv genome (3/38 versus 37/3924, odds ratio = 9.0, P = 0.006). Transcriptional regulators were deleted in BCG strains ob-

tained after 1927 (Rv1985c in RD2), BCG-Pasteur (Rv1773c in RD14), and BCG-Moreau (Rv3405c in RD16). Although the role of this gene family remains to be determined in *M. tuberculosis*, it is tempting to speculate that regulatory elements may serve a role in adapting to environmental change, such as might be experienced during in vivo infection. In contrast, the loss of such genes in laboratory conditions may have little consequence under the relatively constrained conditions of in vitro growth.

The association between ongoing deletion

of genetic material (including transcriptional regulators) and progressive attenuation of virulence now suggests a testable hypothesis for the loss of protective efficacy over the same era. It has long been known that administration of killed BCG organisms results in a weak and transient immune response, indicating that protective immunity by BCG requires survival and replication in the vaccinated host (27). We propose that the deletions detected reflect a progressive adaptation of BCG strains to laboratory conditions that has compromised their capacity to survive within the host, impairing their ability to stimulate a durable immune response.

The basic approach we have used to determine genetic differences within the *M. tuberculosis* complex can be applied to study genetic variability within any species for which the genomic sequence is known. Our current mycobacterial DNA microarray can detect deletions as small as 2 kb, or 1/2000 of the genome. This array cannot detect smaller deletions, point mutations, deletions restricted to intergenic regions, genetic rearrangements, *M. bovis* ORFs that are not part of the H37Rv genome, and deletions of homologous repetitive elements (such as *PGRS* genes). However, with microarrays of greater resolution it will soon be possible to study these other sources of genetic variability and determine their relative frequency and importance. As a result of the numerous genome projects, complete sequence information will soon be available for dozens of microbial species, many derived from reference strains and laboratory isolates. The use of microarray-based comparative genomics should prove a powerful tool for understanding phenotypic variability among clinical and environmental isolates of similar genetic composition.
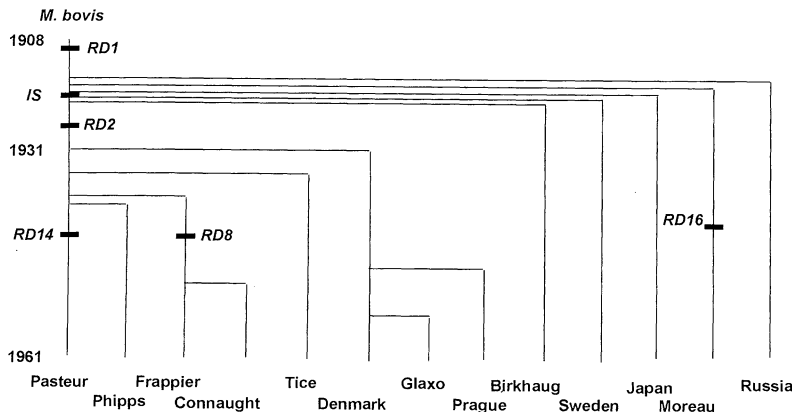

Fig. 2. BCG historical genealogy incorporating genetic differences previously noted and newly detected genetic deletions. The vertical axis represents time. The horizontal axis denotes different geographic locations of BCG propagation. Under this reconstruction, the *M. bovis* strain that was used to develop BCG would be missing RD3, RD4, RD5, RD6, RD7, RD9, RD10, RD11, RD12, RD13, and RD15. During serial propagation of this strain, RD1, RD2, RD8, RD14, RD16, and an IS6110 element (IS) were deleted.

Table 1. Genetic information about distribution of deletions in virulent *M. bovis* and BCG strains. Parentheses after *M. bovis* indicate how many strains of virulent *M. bovis* were missing the genetic element. Start and end position are based on the 162 segments of genome available at http://www.ncbi.nih.gov/Entrez/nucleotide.html. A list of the 129 deleted ORFs may be accessed at http://molepi.stanford.edu/bcg.

Deletion region	Strains where missing	Number of ORFs	ORFs deleted	Start (segment: bp)	End (segment: bp)
RD1*	All BCG strains	9	Rv3871–Rv3879c	160: 7,534	160: 16,989
RD2*	BCG-Danish, BCG-Prague, BCG-Glaxo, BCG-Frappier, BCG-Connaught, BCG-Phipps, BCG-Tice, BCG-Pasteur	11	Rv 1978–Rv1988	88: 14,211	89: 8,598
RD3*	*M. bovis* (3/8)	14	Rv1573–Rv1586c	70: 7,677	70: 16,923
RD4	*M. bovis* (8/8)	3	Rv0221–Rv0223c	12: 17,432	12: 19,335
RD5	*M. bovis* (8/8)	5	Rv3117–Rv3121	135: 27,437	135: 30,212
RD6*	*M. bovis* (8/8)	11	Rv1506c–Rv1516c	65: 23,614	65: 36,437
RD7	*M. bovis* (8/8)	8	Rv2346c–Rv2353c	103: 17,622	103: 26,584
RD8	BCG-Frappier, BCG-Connaught	4	Rv0309–Rv0312	16: 17,018	16: 20,446
RD9	*M. bovis* (8/8)	7	Rv3617–Rv3623	153: 21,131	154: 2,832
RD10	*M. bovis* (8/8)	3	Rv1255c–Rv1257c	55: 3,689	55: 6,696
RD11	*M. bovis* (8/8)	5	Rv3425–Rv3429	145: 30,303	146: 1,475
RD12	*M. bovis* (8/8)	4	Rv2072c–Rv2075c	93: 9,301	93: 11,331
RD13	*M. bovis* (4/8)	16	Rv2645–Rv2660c	118: 12,475	118: 23,455
RD14	BCG-Pasteur	8	Rv1766–Rv1773c	79: 30,573	79: 39,642
RD15	*M. bovis* (8/8)	15	Rv1963c–Rv1977	88: 1,153	88: 13,873
RD16	BCG-Moreau	6	Rv3400–Rv3405c	145: 5,012	145: 12,621

*RD1, RD2, and RD3 were detected by Mahairas *et al.* (*14*) and RD6 was described by Brosch *et al.* (*15*). RD3 and RD13 are prophages phiRv1 and phiRv2, respectively.

References and Notes
1. Centers for Disease Control and Prevention, *Morb. Mortal. Wkly. Rep.* **47** (RR-13), 1 (1998).
2. A. Calmette and C. Guérin, *Ann. Inst. Pasteur* **34**, 553 (1920).
3. M. A. Behr and P. M. Small, *Nature* **389**, 133 (1997).
4. _____, *Vaccine* **17**, 915 (1999).
5. Strains used were *M. tuberculosis* H37Rv, BCG-Russia (ATCC 35740), BCG-Moreau, BCG-Japan, BCG-Sweden, BCG-Prague, BCG-Phipps (ATCC 35744), BCG-Denmark 1331 (ATCC 35733), BCG-Tice (ATCC 35743), BCG-Frappier (ATCC 35735), BCG-Connaught, BCG-Birkhaug (ATCC 35731), BCG-Glaxo (ATCC 35741), and BCG-Pasteur 1173.
6. S. T. Cole *et al.*, *Nature* **393**, 537 (1998).
7. Second Workshop on Methods and Applications of DNA Microarray Technology, *Nature Genet.* **18**, 195 (1998).
8. Using the annotated sequence of *M. tuberculosis* H37Rv made available by the Sanger Centre (ftp://ftp.sanger.ac.uk/pub/tb/sequences), we designed PCR primer pairs to amplify 250- to 1000-bp internal fragments of predicted ORFs (details about primers used are available at http://molepi.stanford.edu/bcg). Primer design was automated using Primer 3 [release 0.6, S. Rozen and H. J. Skaletsky (1996, 1997), code available at www-genome.wi.mit.edu/genome_software/other/primer3.html]. PCR reactions were prepared in 65-μl reactions in 96-well plates; amplicons were evaluated to ensure they were the expected size and then robotically applied to poly-L-lysine–coated microscope slides as described (proto-

cols available at http://cmgm.stanford.edu/pbrown/mguide/index.html). After printing, free amino groups of the poly-L-lysine–coated slides were acetylated with 5.5 g of succinic anhydride in 350 ml of borate-buffered 1-methyl-2-pyrrolidinone (pH 8).

9. Bacteria listed above were grown in 7H9 broth with ADC enrichment, and DNA was extracted after 14 to 21 days' growth as described [D. van Soolingen, P. W. Hermans, P. E. de Haas, D. R. Soll, J. D. van Embden, J. Clin. Microbiol. 29, 2578 (1991)]. Whole genomic DNA from either H37Rv or BCG strain (2 μg) was used as templates for direct incorporation of fluorescent nucleotide analogs (Cy3 and Cy5 dUTP, Amersham) by a randomly primed polymerization reaction. In brief, 50-μl labeling reactions contained 2 μg of template DNA, 5 μl of 10× buffer, 1.5 μl of fluorescent dUTP, 0.5 μl each of 5 mM dATP, dCTP, and dGTP, 1 μl of random hexamers and decamers, and 2 μl of Klenow (Escherichia coli DNA polymerase 3′ to 5′exo–, New England Biolabs). Labeled DNAs from the strains being compared were mixed together to 14.75 μl total volume in a solution with final concentration of 4× SSC buffer, tRNA (0.7 μg/μl), and 0.3% SDS. The hybridization mixture was denatured at 98°C for 2 min, cooled to 65°C, applied to the microarray, and covered with a 22-mm² cover slip. The slide was then placed in a waterproof hybridization chamber for hybridization in a 65°C water bath for 3 hours. After hybridization, slides were washed 2 min in 1× SSC buffer with 0.06% SDS followed by 2 min in 0.06× SSC buffer.

10. S. Sreevatsan et al., Proc. Natl. Acad. Sci. U.S.A. 94, 9869 (1997).

11. W. J. Philipp et al., Microbiology 142, 3135 (1996).

12. Slides were scanned with a custom-built scanning laser microscope as described [J. L. DeRisi, V. R. Iyer, P. O. Brown, Science 278, 680 (1997)]. Fluorescent spot intensities were quantified using custom software as described by DeRisi and available at http://rana.Stanford.EDU/software. For each spot, background fluorescence was subtracted from the average spot fluorescence to produce a channel-specific value. Spots were excluded from analysis when background intensity was greater than the spot intensity. The ratio of Cy3/Cy5 hybridization was calculated and log-transformed for representation according to genomic location implementing ProbeBrowser software (designed and written in our laboratory and available at http://molepi.stanford.edu/free_software.html). This software integrates microarray experimental results with the genomic positions of the hybridization targets and displays corresponding ORF annotations (see Fig. 1, B and C).

13. For any single experimental spot, a mismatched hybridization could represent (i) a repeated element more numerous in H37Rv, such as IS6110, (ii) an element present in H37Rv and absent from BCG, or (iii) a false positive result that may occasionally occur during thousands of simultaneous hybridizations. We observed that spots representing IS6110 reliably resulted in high ratios of H37Rv:BCG, corresponding to their known 16:1 genomic presence. For assigning deletions, we restricted our investigation to regions where probes from at least two adjacent ORFs exhibited mismatched hybridization. This strategy was invoked to maximize the likelihood of detecting true deletions at the expense of decreased resolution. For putative deletions identified by the array experiments, we performed Southern (DNA) hybridization using probes independent of those on the array to confirm their absence in BCG strains. Then, to characterize confirmed deletions, we designed PCR primers to approach the predicted edges of each deletion so that the amplified region would span the missing genomic locus for those strains in which it is deleted. These amplicons were sequenced using the PCR primers at the Protein and Nucleic Acid (PAN) Core Facility, Stanford University, by ABI dye terminator chemistry. Sequences obtained were aligned to the known sequence of H37Rv by BLAST search [S. F. Altschul et al., Nucleic Acids Res. 25, 3389 (1997)] to determine the precise site of each deletion (details available at http://molepi.stanford.edu/bcg). Based on the site and the predicted ORFs from the H37Rv, we documented which reading frames were partially or completely missing from BCG strains. To classify deleted regions as BCG-specific or as differences between H37Rv and virulent M. bovis, we repeated PCR with the primers spanning the missing genetic region using DNA from virulent M. bovis (four human and four bovine isolates).

14. G. G. Mahairas, P. J. Sabo, M. J. Hickey, D. C. Singh, C. K. Stover, J. Bacteriol. 178, 1274 (1996).

15. R. Brosch et al., Infect. Immum. 66, 2221 (1998).

16. L. M. O'Reilly and C. J. Daborn, Tuberc. Lung. Dis. 76 (suppl. 1), 1 (1995).

17. K. A. Johansen, R. E. Gill, M. L. Vasin, Infect. Immun. 64, 3259 (1996).

18. C. A. Parra et al., ibid. 59, 3411 (1991).

19. In the previous work, there were two discrepancies between genotype data and the written record; in both we have subsequently obtained new information indicating that our historical information was incorrect. During the American Indian trial, Aronson obtained BCG-Pasteur (passage) 575 in 1938 [J. D. Aronson, E. I. Parr, R. M. Saylor, Am. Rev. Tuberc. 42, 651 (1940)]; this likely became the stock from which BCG-Phipps was derived. According to Birkhaug himself, BCG-Birkhaug appears to have been obtained from the Institut Pasteur in 1927 [K. Birkhaug, Psychiatr. Q. 21, 453 (1948)].

20. J. D. Aronson, E. I. Parr, R. M. Saylor, Am. Rev. Tuberc. 42, 651 (1940).

21. M. Gheorghiu, J. Augier, P. H. Lagrange, Bull. Inst. Pasteur 81, 281 (1983).

22. A. Calmette and C. Guérin, Ann. Inst. Pasteur 25, 625 (1911).

23. K. A. Jensen, Acta Tuberc. Scand. 20, 1 (1946).

24. G. Dreyer and R. L. Vollum, Lancet i, 9 (1931).

25. Indian J. Med. Res. 72S, 1 (1980).

26. The genomic hierarchy proposed by Cole et al. is available at www.sanger.ac.uk/Projects/M_tuberculosis/Gene_list/.

27. R. J. Dubos and C. H. Pierce, Am. Rev. Tuberc. Pulmon. Dis. 74, 655 (1956).

28. For strains and DNA of M. tuberculosis, M. bovis, and M. bovis BCG, we thank M. Gheorghiu and S. Cole, Institut Pasteur, Paris (BCG-Pasteur 1173); A. Ladefoged, Statens Serum Institut, Copenhagen (BCG-Sweden and BCG-Prague); M. Takahashi, Japanese Anti-Tuberculosis Association (BCG-Japan); P. Rousseau, Institut Armand Frappier, Canada (BCG-Moreau); U. Kosecka, Connaught Laboratories, Canada (BCG-Connaught); M. D. Cave, University of Arkansas; and G. Hewinson, Central Veterinary Laboratories, Surrey, UK. For technical assistance we thank J. DeRisi and members of P. Brown's laboratory. For assistance with the gene annotation database, we thank S. Cole and his laboratory at the Institut Pasteur.

27 January 1999; accepted 8 April 1999

Broadly Protective Vaccine for *Staphylococcus aureus* Based on an in Vivo–Expressed Antigen

David McKenney,[1] Kimberly L. Pouliot,[1] Ying Wang,[1] Vivek Murthy,[1] Martina Ulrich,[3] Gerd Döring,[3] Jean C. Lee,[1] Donald A. Goldmann,[2] Gerald B. Pier[1]*

Vaccines based on preferential expression of bacterial antigens during human infection have not been described. Staphylococcus aureus synthesized poly-N-succinyl β-1-6 glucosamine (PNSG) as a surface polysaccharide during human and animal infection, but few strains expressed PNSG in vitro. All S. aureus strains examined carried genes for PNSG synthesis. Immunization protected mice against kidney infections and death from strains that produced little PNSG in vitro. Nonimmune infected animals made antibody to PNSG, but serial in vitro cultures of kidney isolates yielded mostly cells that did not produce PNSG. PNSG is a candidate for use in a vaccine to protect against S. aureus infection.

Staphylococcus aureus is the most frequently isolated bacterial pathogen in hospital-acquired infections (1) and is a common cause of community-acquired infections, including endocarditis, osteomyelitis, septic arthritis, pneumonia, and abscesses (1). *Staphylococcus aureus* is also a significant pathogen in economically important animals (2). Staphylococcal resistance to first-line drugs such as synthetic penicillins has resulted in major problems in treating methicillin-resistant *S. aureus* (MRSA) strains, which are increas-

ingly common, especially in hospitalized patients. Of greater concern is the recent emergence in several countries of MRSA strains with reduced susceptibility to vancomycin, the antibiotic of last resort (3). The appearance of these vancomycin-intermediate *S. aureus* (VISA) strains raises the specter of untreatable staphylococcal infections, necessitating a search for alternative therapies (1, 3).

One potential therapeutic target for bacterial infections is products of bacterial genes activated during in vivo infections (4). Presumably these genes encode factors critical for infection and disease progression. Strategies designed to discover in vivo–expressed genes have, of necessity, used animal models for gene identification, and in vivo–expressed bacterial factors have not yet been detected in human infections. In addition, the usefulness in vaccines of any factors encoded

[1]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, [2]Department of Medicine, Children's Hospital, Harvard Medical School, Boston, MA 02115, USA. [3]Department of General and Environmental Hygiene, Hygiene-Institute, Universitätsklinikum, D-72074 Tübingen, Germany.

*To whom correspondence should be addressed. E-mail: gpier@channing.harvard.edu