

Use of Genome Level-Informed PCR as a New Investigational Approach for Analysis of Outbreak-Associated *Mycobacterium tuberculosis* Isolates

Kumar Rajakumar,^{1,2†} Jamila Shafi,^{1†} Rebecca J. Smith,^{1†} Richard A. Stabler,^{3†}
Peter W. Andrew,¹ Deborah Modha,³ Gerry Bryant,⁴ Philip Monk,⁴
Jason Hinds,³ Philip D. Butcher,³ and Michael R. Barer^{1,2*}

Department of Infection, Immunity and Inflammation, Leicester Medical School, University of Leicester, Leicester LE1 9HN,¹
Department of Clinical Microbiology, University Hospitals of Leicester NHS Trust, Leicester LE1 5WW,² and
Leicestershire Health Authority,⁴ Leicester, and Department of Cellular and Molecular Medicine (Medical
Microbiology), St. George's Hospital Medical School, London SW17 0RE,³ United Kingdom

Received 15 October 2003/Returned for modification 29 November 2003/Accepted 7 January 2004

***Mycobacterium tuberculosis* strain CH, the index isolate linked to a major tuberculosis outbreak associated with high levels of transmissibility and virulence, was characterized by microarray analysis by use of a PCR product array representative of the genome of *M. tuberculosis* strain H37Rv. Seven potential genomic deletions were identified in CH, five of which were confirmed by PCR analysis across the predicted deletion points. The panel of five PCRs required to individually interrogate these loci was collectively referred to as the genome level-informed PCR (GLIP) assay. GLIP analysis was performed with CH, 12 other epidemiologically linked isolates, and 43 recent, non-outbreak-associated isolates derived from patients within the local area. All 13 outbreak-linked isolates showed a profile corresponding to the presence of all five deletions. These 13 isolates were also found to share common variable-number tandem repeat and mycobacterial interspersed repetitive unit profiles. None of the 43 non-outbreak-associated isolates exhibited the five-deletion profile. Although three individual deletions were present in upwards of 44% of the non-outbreak-associated isolates, no single-deletion isolates were detected. Interestingly, none of these deletions had been previously recognized, and sequence analysis of the immediate flanking regions in CH failed to identify a likely mechanism of deletion for four of the five loci. The GLIP assay also proved valuable in ongoing surveillance of the outbreak, rapidly identifying a further two outbreak-associated cases months after the initial cluster and, importantly, dismissing a further 12 epidemiologically suspect cases, which allowed the optimum deployment of public health resources.**

Mycobacterium tuberculosis remains a major killer, accounting for more than 2 million deaths annually. Its success as a pathogen is highlighted by its remarkable ability to spread among human populations. It is estimated that more than 2 billion people, a third of the world's population, are infected with this bacterium (29). The major burden is borne by underdeveloped nations. However, even in settings with excellent public health and clinical facilities, sporadic outbreaks of tuberculosis remain a continuing threat (7, 23).

In 2001 the largest recognized outbreak of tuberculosis in a United Kingdom school was detected in Leicester. The index patient was a 14-year-old student who had been complaining of a chronic cough for 9 months prior to being diagnosed with sputum smear-positive cavitary pulmonary tuberculosis (9). Subsequent screening and investigation by the health authority of the entire school population led to the diagnosis of a further 77 cases of active disease and 254 cases of latent tuberculosis among students, staff, and family contacts of the index patient and secondary cases. This outbreak occurred in Leicester, a city with rates of tuberculosis exceeding four times the national

average (24; P. Monk, unpublished data). Given the large scale of the outbreak and the elevated rates of tuberculosis in Leicester, there were substantial concerns that the outbreak strain would emerge in the community at large. From a public health perspective, the potential dissemination of the outbreak strain posed a significant hazard. This was particularly so given the high rates of transmission of 20 to 90%, based on measures of proximity and duration of exposure to the index patient, among student contacts and the markedly raised rate of active disease associated with infection (9). Indeed, the potential for the community-wide spread of outbreak strains is well recognized (16, 23). To monitor the situation closely, rapid molecular epidemiological tools were essential. In this report we describe a new investigational approach premised on the hypothesis that deletions detected by a single round of genomic microarray analysis would provide useful strain-specific markers. The microarray-derived data allowed the establishment of a rapid, easily interpretable, PCR-based typing assay that was of value in ongoing outbreak surveillance. Importantly, the technique described was capable of excluding with certainty isolates that were not the result of dissemination of the original clone, allowing a more focused public health effort.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *M. tuberculosis* H37Rv and *M. tuberculosis* CH (the Leicester outbreak index isolate) were grown at 37°C with shaking at 150 rpm in Middlebrook 7H9 broth (Difco) containing 10% (vol/vol)

* Corresponding author. Mailing address: Department of Infection, Immunity and Inflammation, Leicester Medical School, University of Leicester, Maurice Shock Building, University Rd., P.O. Box 138, Leicester LE1 9HN, United Kingdom. Phone: 44 (0) 116 2522933. Fax: 44 (0) 116 2525030. E-mail: mrb19@leicester.ac.uk.

† K.R., J.S., R.J.S., and R.A.S. contributed equally to this work.

TABLE 1. Sequences of oligonucleotide primers used for PCR-based verification of putative CH-borne genomic deletions

Primer no.	H37Rv ORF(s) interrogated	Primer sequence	H37Rv genomic coordinates
1	<i>Rv1519</i>	5'-GGTGGAGAGCGACGACATCAAG-3'	1710043-1710064
2	<i>Rv1519</i>	5'-TGTAGCGACAGCATCGTCATCC-3'	1711863-1711884
3	<i>Rv0180c</i>	5'-GTCCTTCGTTCGCATAGGCA-3'	209943-209964
4	<i>Rv0180c</i>	5'-CGACCACTACGATCCCGACAAC-3'	212901-212922
5	<i>echA19-Rv3517</i>	5'-TCTCCTGTGCGAAAGTCGGTTC-3'	3951857-3951878
6	<i>echA19-Rv3517</i>	5'-GATGATCGACATGGACGATCCC-3'	3955279-3955300
7	<i>PPE66-PPE67</i>	5'-TCGACTCGTTCAGCGGATAACC-3'	4188076-4188097
8	<i>PPE66-PPE67</i>	5'-GACGCAGCACACCCGCCATCAAG-3'	4191655-4191676
9	<i>Rv1995-Rv1996</i>	5'-TCAGCTACCTGTCATCTCGACC-3'	2237628-2237949
10	<i>Rv1995-Rv1996</i>	5'-TCTGATTCTCGGTTCAGCGTCC-3'	2241164-2241185
11	<i>esxR-esxS</i>	5'-CCCACCCTAATCGTTCCGAGTC-3'	3376560-3376582
12	<i>esxR-esxS</i>	5'-TTCCAAAGAGGACGCTGCAGA-3'	3383722-3383742
13	<i>qor</i>	5'-GTTGCACGAATTACTGGATGTG-3'	1638764-1638785
14	<i>qor</i>	5'-CAAGATTGCGGATGTCTGGGTC-3'	1641442-1641463

bovine serum albumin, glucose, and catalase enrichment. Additional local *M. tuberculosis* isolates analyzed in this study were obtained from the Clinical Microbiology Laboratory at the University Hospitals of Leicester or from the Mycobacterial Reference Laboratory at Birmingham.

DNA extraction. For microarray analysis, DNA was extracted from cultures of strains H37Rv and CH as follows. Genomic DNA was extracted from 100-ml stationary-phase cultures by the method of Belisle and Sonnenberg (2). Briefly, bacterial cells were pelleted and washed once with Tris-EDTA (TE; pH 8.0) before being placed at -20°C for 4 h. The cells were delipidated by extraction once with CHCl_3 -methanol (2:1), resuspended in 100 mM Tris-HCl (pH 9.0) containing 200 μg of lysozyme ml^{-1} and 100 μg of RNase A ml^{-1} , and incubated at 37°C for 12 h with gentle shaking. A further 3 h of incubation at 55°C followed after the addition of proteinase K and sodium dodecyl sulfate (SDS) to final concentrations of 200 μg ml^{-1} and 1% (wt/vol), respectively. The cellular matter was extracted twice with CHCl_3 -isoamyl alcohol (24:1), and the genomic DNA was precipitated with 3 M sodium acetate (pH 5.2) and isopropanol. The genomic DNA was further purified by CsCl ultracentrifugation. DNA was resuspended in 805 μl of TE (pH 8.0), to which an equal volume of 3% (wt/vol) Sarkosyl and 1.76 g CsCl was added, and the mixture was ultracentrifuged at $350,000 \times g$ for 24 h at 20°C . Fractions containing genomic DNA were isolated, and the genomic DNA was precipitated with isopropanol.

For genome level-informed PCR (GLIP) analysis, the bacterial biomass collected from BACTEC *Mycobacterium* growth indicator tubes or Lowenstein-Jensen slopes was resuspended in 100 μl of sterile H_2O and boiled (100°C , 10 min) to release the DNA. Lysed cells were extracted once with phenol- CHCl_3 -isoamyl alcohol (25:24:1) and once with CHCl_3 -isoamyl alcohol (24:1), and the DNA was precipitated with isopropanol. The pellet was resuspended in 20 μl of 10 mM Tris-HCl (pH 8.0). A total of 1 to 2 μl was used for GLIP analysis.

DNA labeling and hybridization by use of microarrays. Microarray analyses were performed at the Bacterial Microarray Group, St. George's Hospital Medical School, London, United Kingdom. Whole-genome *M. tuberculosis* microarrays were used. The microarrays comprised PCR amplicons designed to have minimal cross-hybridization and represented the 3,924 predicted open reading frames (ORFs) of the sequenced strain *M. tuberculosis* H37Rv (5). Construction of this microarray is described by Stewart et al. (32). DNA was labeled by incorporation of cyanine 3 (Cy3)- or Cy5-labeled dCTP (Amersham) during random priming of genomic DNA. DNA (2 to 10 μg) was mixed with 3 μg of random primers (Invitrogen Life Technologies) in 41.5 μl of water, heated to 95°C , and snap cooled. The following were then added to this mixture: 5 μl of $10\times$ REact 2 buffer (Invitrogen Life Technologies); 100 μM (each) dATP, dGTP, and dTTP; 40 μM dCTP; 0.75 nM Cy3- or Cy5-labeled dCTP; and 5 U of the Klenow fragment (Invitrogen Life Technologies). The reaction mixture was incubated in the dark at 37°C for 90 min. Cy3-labeled H37Rv DNA (reference strain) and Cy5-labeled CH DNA (Leicester outbreak index isolate) were mixed and purified with a Qiagen MinElute PCR purification kit, with the labeled DNA eluted into 10.5 μl of water.

The microarray was incubated in prehybridization buffer ($3.5\times$ SSC [$1\times$ SSC is 0.15 M NaCl plus 0.015 M sodium citrate], 0.1% [wt/vol] SDS, 10 mg of bovine serum albumin ml^{-1}) at 65°C for 20 min. The slide was rinsed in (each) water and propan-2-ol and dried by centrifugation. Purified Cy3: Cy5-labeled DNA was mixed with hybridization solution ($4\times$ SSC, 0.3% [wt/vol] SDS), incubated at

95°C for 2 min, briefly centrifuged, and applied to the array; and a coverslip was placed over the array. The slide was sealed in a humid hybridization cassette and incubated in the dark at 65°C for 16 to 20 h. The slide was washed at 65°C in $1\times$ SSC-0.05% (wt/vol) SDS for 2 min, followed by two 2-min washes in $0.6\times$ SSC at room temperature, and was dried by centrifugation. After cohybridization of the Cy3- and Cy5-labeled DNA, the microarray was scanned with an Affymetrix 428 scanner (MWG Biotech) and fluorescent spot intensities were quantified with Imagen software (version 4.2; Biodiscovery Inc.). The data were further analyzed with Genespring software (version 5.1; Silicon Genetics), as described by Dorrell et al. (6).

PCR protocol. The oligonucleotide primers used in this study are listed in Table 1. For all reactions, amplifications were performed in a total volume of 10 μl containing 1 μl of DNA, 1 μM each primer, 200 μM each deoxynucleoside triphosphate, $1\times$ PCR buffer, $1\times$ Q solution, and 1 U of Hot start *Taq* polymerase (Qiagen). All reactions were subjected to a hot start for 15 min to activate the *Taq* polymerase, followed by 40 cycles of 30 s at 94°C , 30 s at 62°C , and 3 min at 72°C . PCRs were performed in a Dyad DNA engine (MJ Research); and the products were separated by electrophoresis on 1% (wt/vol) agarose gels, stained with ethidium bromide, and visualized by UV transillumination.

RESULTS

Microarray-based comparative genomics. Genomic DNA from the 2001 Leicester outbreak index *M. tuberculosis* isolate, designated CH, and an equivalent amount of *M. tuberculosis* strain H37Rv DNA were labeled with the fluorescent dyes Cy5 and Cy3, respectively. The labeled DNA was simultaneously hybridized to an in-house-synthesized glass slide-based DNA microarray comprising 0.1- to 0.8-kb PCR amplicons representing the 3,924 protein-encoding ORFs of H37Rv (5, 32). The hybridizations were performed in triplicate to obtain robust, statistically valid data. By using the previously employed cutoff of a twofold change in hybridization intensity and by confining the analysis to spots with a minimum signal fluorescence of >100 units in the Cy3 channel after subtraction of the background fluorescence (6), we identified 11 ORFs that were potentially deleted from the CH sequence relative to the H37Rv sequence. These comprised three single and four adjacent pairs of ORFs dispersed over seven distinct loci on the chromosome. Individual datum points from each of the triplicate arrays corresponding to the 11 ORFs were analyzed by the Student *t* test to calculate a *P* value for the likelihood of deletion (Table 2).

PCR-based verification of deleted loci. The seven loci were further interrogated by PCR analysis of the putative deleted

TABLE 2. Putative deletions identified in *M. tuberculosis* strain CH by ORF-based microarray analysis

H37Rv ORF designation	Normalized Cy5/Cy3 ratio	P value ^a
<i>qor</i> (Rv1454c)	0.498	0.032
<i>echA19</i> (Rv3516)	0.431	0.002
<i>esxR</i> (Rv3019c)	0.258	0.002
<i>Rv1519</i>	0.241	0.001
<i>PPE67</i> (Rv3739c)	0.223	0.002
<i>esxS</i> (Rv3020c)	0.185	0.001
<i>PPE66</i> (Rv3738c)	0.072	0.000
<i>Rv0180c</i>	0.061	0.000
<i>Rv1996</i>	0.052	0.000
<i>Rv3517</i>	0.028	0.000
<i>Rv1995</i>	0.023	0.000

^a By the Student *t* test.

regions identified (Fig. 1). The oligonucleotide primers used (Table 1) were selected from the large set synthesized previously for fabrication of the *M. tuberculosis* H37Rv microarray (32). Individual PCR analyses with either purified H37Rv or CH DNA confirmed the presence of deletions ranging in size from 0.8 to 2.0 kb at five of the seven loci implicated by comparative microarray analysis. Analysis of the sixth locus harboring the single gene *qor* yielded identically sized fragments of 2.4 kb with both H37Rv and CH template DNAs, suggesting that this locus was not deleted in CH. This was consistent with the statistically least secure prediction of the *qor* deletion in CH (Table 2). PCR interrogation of the seventh

locus bearing the tandem genes *esxR-esxS* resulted in fragments of ~7 kb, with both H37Rv and CH template DNAs also failing to confirm this deletion. BLAST analysis of the array elements for both these loci (*qor* and *esxR-esxS*) showed potential partial cross-hybridization with the PE-PGRS and the PE genes (5), respectively, which may differ either by sequence divergence or by GC-rich repeat copy number between H37Rv and the CH strain. This would produce ratios (Table 2) similar to those produced by deletion. These two loci were thus excluded from subsequent characterization. Conversely, *echA19* showed a ratio of 0.431 (Table 2), close to the cutoff used to define deletion events. However, PCR revealed (Fig. 1) partial deletion of *echA19*, which resulted in only a partial loss of the hybridization signal and hence a lower ratio. Verification of deletions by PCR was therefore an essential part of deletion discovery by microarray analysis.

Sequence-level definition of the five loci deleted in CH. Sequence data across the junctions of the five PCR-confirmed loci deleted in CH were generated by limited sequencing of the CH-derived truncated PCR amplicons. Comparison of these data with those for the H37Rv genome demonstrated sequences nearly identical to the H37Rv sequences flanking each of the loci deleted in CH. The predicted shift in PCR amplicon size for each deletion accorded well with the sizes measured by gel electrophoresis. No IS6110-related sequences were identified immediately adjacent to the identified deletion points. Furthermore, besides the *echA19-Rv3517* deletion locus, examination of the immediate flanking sequences failed to reveal

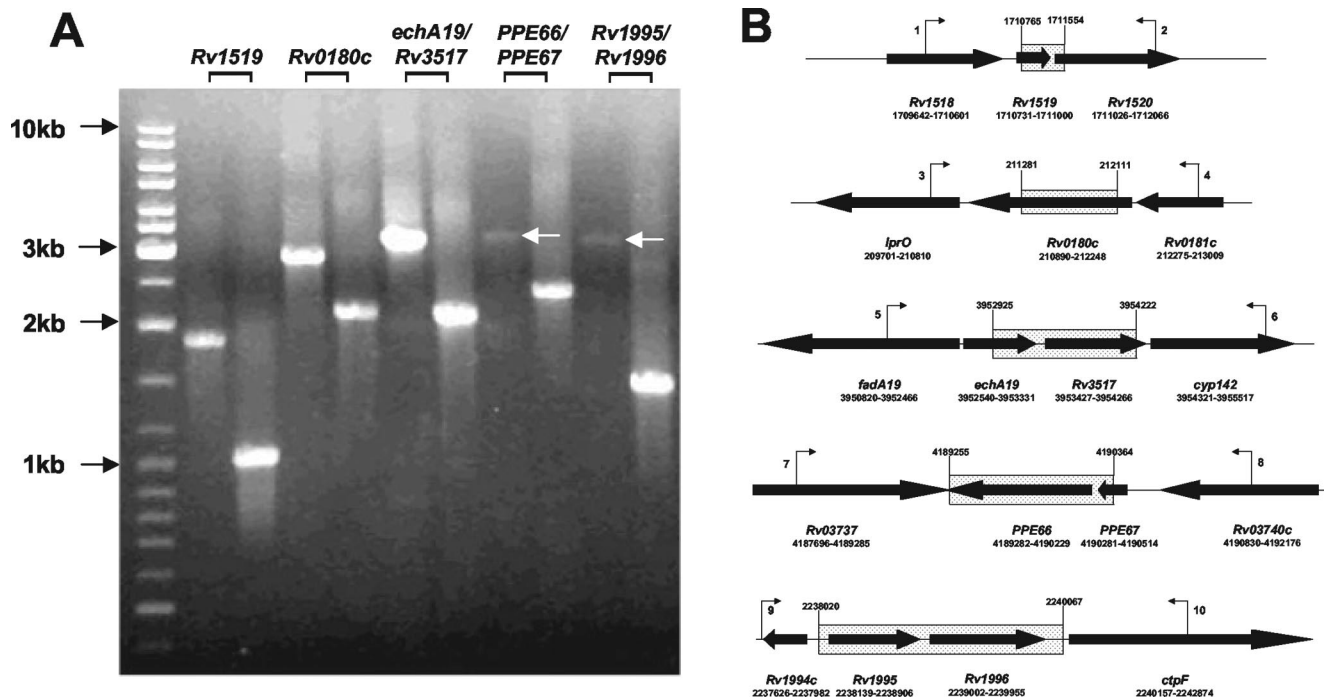


FIG. 1. (A) GLIP analysis of *M. tuberculosis* strain H37Rv and the Leicester outbreak index strain CH indicating the shift in the PCR amplicon size due to the individual CH-specific genomic deletions. Standard DNA size markers are shown on the left. For each locus interrogated, the products from H37Rv and CH are shown in the left- and right-hand lanes, respectively. The white arrows highlight the positions of fainter amplicons. (B) Genetic organization of the five *M. tuberculosis* strain H37Rv loci bearing defined deletions in CH. The coordinates of the ORFs are shown below their designations, while the shaded boxes and flanking coordinates indicate the regions deleted in CH. The bent arrows represent the approximate positions of the GLIP primer binding sites.

TABLE 3. GLIP analysis of *M. tuberculosis* isolates derived from patients presenting in the Leicester area

DT	Result for ORF interrogated ^a					No. of initial outbreak-associated and retrospective isolates	No. of isolates tested during prospective surveillance
	<i>Rv1519</i>	<i>Rv0180c</i>	<i>echA19-Rv3517</i>	<i>PPE66-PPE67</i>	<i>Rv1995-Rv1996</i>		
DT1	+	+	+	+	+	23	4
DT2	-	+	-	-	+	15	7
DT3	-	-	-	-	+	4	1
DT4	-	-	-	-	-	13	2
DT5	-	+	-	+	+	1	0
Outbreak ^b	100	100	100	100	100		
Nonoutbreak ^c	46.5	9.3	46.5	44.2	0		

^a The plus and minus symbols indicate the presence and absence of the interrogated locus in the tested isolate, respectively.

^b Percentage of the 13 outbreak-associated isolates exhibiting each individual deletion.

^c Percentage of the 43 non-outbreak-associated isolates bearing each individual deletion.

potential evidence of short direct repeat-mediated deletion events. The presence of the tetramer repeat GGTTG at one flank and at the distal end of the deleted *echA19-Rv3517* locus suggested the possibility of a recombination-mediated deletion event in this instance. A schematic representation of the five loci deleted in CH is shown in Fig. 1. Two of the complete or partially deleted ORFs (*PPE66* and *PPE67*) encode members of the glycine-rich PPE family of proteins (28). This is consistent with previous evidence that these genes are overrepresented among the *M. tuberculosis* genomic deletions detected (11, 25). The *echA19* gene encodes a possible enoyl coenzyme A hydratase (8). *Rv1996* codes for a protein with a high degree of similarity to bacterial universal stress protein UspA (30). The annotation of the four remaining ORFs provides no clues as to their likely function.

GLIP analysis of *M. tuberculosis* isolates. The five PCR assays that confirmed the deletions in CH were used for subsequent analysis of the *M. tuberculosis* isolates. The maximum amplicon size generated in these reactions with either H37Rv or CH DNA was 3.6 kb. The five-locus PCR genotyping assay, termed the GLIP assay, was applied in an investigator-blinded fashion to CH, 12 other outbreak-associated isolates, and 43 isolates derived from local cases of tuberculosis over the preceding 2 years with no known epidemiological links to the outbreak. The last group of isolates was selected on a random basis, with the exception that representatives of each of the prevalent *IS6110* restriction fragment length polymorphism (RFLP) types were included. PCR products were analyzed by gel electrophoresis, and the sizes of the fragments were assessed by comparison with those derived from H37Rv and CH and with standard DNA size markers (Fig. 1A). The analyses were considered interpretable only if a distinct PCR product was obtained in all five reactions, a nearly routine assay outcome during the course of these studies.

Twelve of the 13 isolates derived from the index patient and the 12 patients with strong epidemiological links to the index patient were assigned to deletion type (DT) 4 (DT4), the DT characterized by the presence of all five deletions. However, one epidemiologically linked isolate (isolate J24), derived from the father of the index patient, was identified as a DT3 strain, possessing only four of the five CH-specific deletions. By contrast, none of the 43 non-epidemiologically linked isolates exhibited a DT4 genotype.

We were surprised by the assignment of isolate J24 to a

profile different from those for the other outbreak isolates. Consequently, as part of a quality control exercise, we retested all four- and five-deletion isolates identified and the single isolate (isolate J23) exhibiting an apparently unique three-deletion profile, along with a random selection of three isolates each of the DT1 and DT2 GLIP groups. This analysis confirmed the findings of the initial investigator-blinded study for all but two isolates (isolates J23 and J24). J24, like all the other outbreak-associated strains, was shown to possess a deletion at the *Rv1995-Rv1996* locus, while J23, contrary to the earlier finding, lacked this deletion, suggesting the likelihood of a gel electrophoresis lane transposition error with adjacent specimens during the initial analysis of the PCR amplicons targeting this locus.

The results of the final GLIP analysis are shown in Table 3. Five distinct repertoires of deletion patterns were identified among the 56 isolates studied. These were designated DT1 to DT5.

DT1 represented the most prevalent DT, accounting for 23 of 56 strains investigated. Isolates exhibiting this genotype lacked all five deletions, resulting in an H37Rv-like GLIP profile. Fifteen isolates bore three of the five deletions (DT2), four isolates possessed four of the five deletions (DT3), and one isolate harbored two of the five deletions. Intriguingly, none of the 56 isolates investigated harbored only one of the five deletions sought.

Established molecular typing of the original outbreak-associated isolates. Strain CH and 9 of the 12 secondary isolates have been reported to possess identical profiles by variable-number tandem repeat (VNTR), mycobacterial interspersed repetitive unit (MIRU), and spoligotyping analyses (9). Furthermore, the isolates were indistinguishable by conventional *IS6110*-based RFLP analysis, with 8 of the 10 isolates possessing an identical 15-band pattern and 2 secondary isolates each differing by only one and three bands, respectively (9). The remaining three initially recognized outbreak-associated isolates reported in this study were also indistinguishable from CH by VNTR and MIRU profiling, supporting the clonal relationship of these 13 isolates. Equally, when data on the profiles obtained by *IS6110*-based RFLP, MIRU, and/or spoligotyping analysis were available for the remaining 43 non-outbreak-associated isolates through routine public health epidemiology, the profiles for the isolates were distinct from those for CH.

Application of GLIP analysis for ongoing, real-time, outbreak-associated surveillance. Thirteen additional *M. tuberculosis* isolates derived from patients presenting after the acute phase of the outbreak were investigated by GLIP analysis. These prospective isolates were selected on the basis of the perceived urgent public health need to exclude any association with the earlier outbreak. Two further isolates derived from patients with epidemiological links to the original outbreak exhibited a DT4 genotype. The GLIP profile data were obtained several months postisolation of CH and strongly suggested a delayed evolution of disease in these individuals following infection with the CH clone. These GLIP typing data were subsequently supported by both VNTR and MIRU analyses.

The practical utility of GLIP analysis was further demonstrated by its rapid and unambiguous ability to exclude the other 12 potentially linked isolates. These isolates possessed at least one of the five loci deleted in CH and, given the absence of evidence of natural horizontal gene dissemination in *M. tuberculosis*, could not represent progeny of CH. The latter group included an isolate cultured from a student in the same tutor group as the index patient; the contacts in the index patient's tutor group were found to be at the greatest risk of contracting the infection (9). The student in the tutor group was found on initial screening to have a positive Mantoux test and was prescribed a 3-month course of chemoprophylaxis comprising rifampin and isoniazid. During the course of prophylaxis he was diagnosed with sputum culture-positive pulmonary tuberculosis, raising concerns about the emergence of resistance in the original outbreak strain. However, prompt application of GLIP analysis dismissed this possibility, identifying the isolate as having a DT2 genotype instead.

Furthermore, for five sputum smear-positive patients we were able to fast-track the process by subjecting primary sputum samples to GLIP analysis directly. On each occasion the results unequivocally excluded these patients as having outbreak-related cases of tuberculosis, as one or more of the CH-deleted loci were detectable within the primary specimen itself.

DISCUSSION

The prompt investigation of outbreaks, supported by sufficiently discriminatory typing methods, remains central to the control of tuberculosis in the developed world (7, 29). Present typing methods based on IS6110-based RFLP (33), spoligotyping (17) and VNTR-MIRU (19, 29) analyses have proved invaluable in these investigations, often defining with a high degree of certainty the true clonality of epidemiologically linked *M. tuberculosis* isolates.

IS6110-based RFLP profiles are known to evolve rapidly, occasionally even during the course of a brief outbreak (33), potentially confounding interpretation, especially in the hands of less experienced personnel. Indeed, two Leicester outbreak-associated isolates exhibited additional IS6110 bands. This variation stems from the inherent mobility of IS6110 elements and their noted ability to mediate genomic deletions (10). Spoligotyping, a technique that interrogates the direct-repeat locus for the presence or absence of individual unique spacer sequences, has proved more stable, although this analysis is frequently insufficiently discriminatory to be solely relied upon

in outbreak investigations (17, 29). The VNTR-MIRU analysis approach assesses the number of repeat units at 5 to 12 distinct loci around the chromosome. To date the technique has proved remarkably robust, exhibiting impressive stability over time and an apparent resolution level comparable to that of IS6110-based RFLP analysis (19, 26, 29). However, the observed stability of VNTR-MIRU profiles could result in the identification of a limited range of types in geographic localities characterized by minimal strain importation. This would affect the utility of this typing tool in the investigation of sporadic outbreaks. There is already evidence of the emergence of a dominant VNTR type, which possesses the five-locus profile 42235, among isolates derived from patients in the cities of Leeds and Bradford in the United Kingdom. Isolates with this profile accounted for 23% of the 210 isolates investigated and was exclusively associated with patients with South Asian surnames (13). Remarkably, a further 27% of isolates differed from the isolates with the VNTR profile associated with patients with a South Asian surname at only one of five loci (13). Indeed, isolate CH, the 12 acute-phase secondary isolates, and the 2 outbreak-associated isolates identified during prospective surveillance were all found to share this same VNTR profile, profile 42235 (Monk, unpublished).

The high degree of DNA sequence conservation among *M. tuberculosis* isolates (31), the lack of evidence of natural horizontal gene transfer, and the growing list of identified deletions found in members of the *M. tuberculosis* complex (1, 3, 4, 11, 14, 18, 25) led us to hypothesize that PCR-based deletion analysis could potentially be used to investigate outbreak-associated isolates. Characterization of the index isolate, isolate CH, by use of an ORF-based genomic microarray led to the identification of five deletions relative to the sequence of the strain H37Rv genome. None of the five deletions were precisely coincident with those identified in the vaccine strain *M. bovis* BCG Pasteur (1, 25), the major U.S. *M. tuberculosis* outbreak strain CDC1551 (11), or the 13 *M. tuberculosis* clones from San Francisco, Calif., investigated to date (18). Only one of the deletions, that encompassing *PPE66* and the 3' terminus of *PPE67*, was found to partially overlap previously identified distinct deletions in BCG Pasteur and the recently sequenced strain of *M. bovis* (12, 25). Furthermore, unlike a significant proportion of previously identified genomic deletions (3, 4, 10, 14), IS6110 elements were not involved in any of the identified gene-loss events that gave rise to CH. Surprisingly, despite the absence of a clearly demonstrable mechanism of deletion, four of the five individual deletions interrogated by the GLIP assay were found among the non-outbreak-associated Leicester strains investigated at frequencies ranging from 9.3 to 46.5%. The fifth deletion was unique to the 15 outbreak-associated isolates studied. We are investigating the basis of these findings among *M. tuberculosis* isolates prevalent in the Leicester area. In particular, it will be interesting to know whether the loci exclusively affected in CH contribute to virulence and/or transmissibility.

Little is known about the factors contributing to the emergence of new deletions in circulating *M. tuberculosis* strains. On the basis of limited comparisons with other phylogenetic data, it has been proposed that since deletions probably represent unidirectional genetic events, individual deletion profiles may permit valid phylogenetic inference (18, 20). This appears to be

the case for the *M. tuberculosis* complex, for which Brosch et al. (3) and Mostowy et al. (20) have put forward elegant scenarios for the evolution of this group of closely related organisms. Indeed, deletion analysis may yet prove to be an extremely robust and simple approach to classifying members of the *M. tuberculosis* complex (15, 21). Fleischmann et al. (11), using *in silico* comparative genomics of the two strains of *M. tuberculosis* that have been sequenced, strains H37Rv and CDC1551, have identified 37 insertions larger than 10 bp in H37Rv and a further 49 insertions in CDC1551. The distributions of 17 of these CDC1551-H37Rv large-sequence polymorphisms among 169 random isolates cultured at a New York City hospital were investigated by hybridization. All isolates exhibited the absence of at least one to seven loci, with an average of 3.7 large-sequence polymorphisms missing per isolate. Further analysis of a set of epidemiologically well-characterized isolates demonstrated an identical deletion pattern in all epidemiologically linked isolates within a common IS6110-based RFLP-defined cluster (11).

Our discovery of five novel *M. tuberculosis* deletions following microarray analysis of a single isolate highlights the potential of gene loss as a major driver of genome diversity in this species. The potential impact of this form of genomic remodeling on virulence and transmissibility remains unknown. On the basis of an analysis of 16 clones, Kato-Maeda et al. (18) have suggested that clinical isolates that have lost a greater proportion of their genomes are less likely to be associated with cavitory pulmonary tuberculosis and, hence, presumably with secondary case transmission. Interestingly, by these criteria both CH and CDC1551, which have lost only approximately 6 to 10 kb of DNA in total, fall into the category of strains exhibiting minimal "reductive evolution" and, hence, potentially pose the greatest public health threat.

In this study, the data from a single round of microarray analysis identified five deletion loci that were immediately amenable to routine PCR-based deletion analysis. The low signal-to-signal ratios on the ORF amplicon array used here reflect a loss of a sequence complementary to individual amplicons in the test DNA relative to the sequence of the strain H37Rv control. In most cases, a specific sequence complementary to an amplicon is confined to a single locus. In a few cases, particularly when PPE or PE-PGRS elements are involved, multiple loci may contribute to a particular amplicon-related signal. Two of the low signal ratios detected here did not identify deletions suitable for routine analysis. In the case of the *gor* locus, this may reflect partial cross-hybridization involving PE-PGRS elements, to which this locus shows a small amount of cross homology. We were also unable to confirm a deletion involving the *esxR* and *esxS* loci. Homology of *esxR* to the *ESAT6-CFP10* gene family may have made a significant contribution in this case. Similarly, homology of *esxS* (PE28) with other PE genes across the genome that may have undergone sequence divergence or GC-repeat copy number changes would have affected the hybridization signal. Hence, not all microarray elements that show a reduced ratio (Table 2) automatically indicate deletion events, as cross-hybridization and copy number events can also produce reduced ratios. The importance of PCR deletion verification is therefore clearly demonstrated. It should be noted that the ORF amplicon array would not have detected genomic rearrangements when a sig-

nificant loss of the target sequence does not occur. Conversely, sequence variation within an amplicon target may lead to lower signal ratios without any change in amplicon size when the region concerned is interrogated by PCR. These points raise the general issue of the sensitivity and specificity of our array procedure for the detection of genomic deletions. Our procedure cannot detect deletions affecting loci outside those targeted by the amplicon set used, and this means that exclusively intergenic deletions would not be detected. However, when CDC1551 DNA was tested against this array, all the deletions defined by sequencing that involved ORFs were detected (R. A. Stabler, J. Hinds, and P. D. Butcher, unpublished data). Finally, we emphasize that our microarray procedure does not provide a comprehensive genomic analysis and that many additional analyses short of complete sequencing could be applied to this end. Rather, the procedure that we describe should be seen as a screen that proved useful in this instance.

We have demonstrated the potential utility of deletion analysis for local tuberculosis outbreak investigation. The five-locus GLIP assay developed was simple to perform, the results were simple to interpret, and the assay had the power to unambiguously exclude isolates as outbreak associated. In the context of this study, its positive predictive value has also proved impressive, with all 15 isolates bearing the five deletions having been cultured from samples from patients with strong direct epidemiological links with the acute phase of the outbreak. Like the multiplex PCR-based assay that has been developed for ongoing surveillance of CDC1551 in Tennessee and Kentucky (22), the GLIP assay, directed at monitoring of the dissemination of the CH clone, will aid with the early detection of community spread, potentially preventing the stable establishment of a highly transmissible and virulent strain within the local area. As discussed previously (27), we suggest that application of an iterative process of genotyping by the GLIP assay and identification of additional discriminatory deletion loci by selective further microarray analyses of isolates within one locality may provide a rapid and useful view of locally circulating strains and their transmission. We also propose that a similar general strategy of microarray profiling of an index isolate followed by selection of strain-specific loci to be interrogated by PCR could prove useful in tracking future tuberculosis outbreaks.

ACKNOWLEDGMENTS

J.S. performed the initial array analyses and established the PCR typing method.

We thank MediSearch (Leicester) for funding J.S. and R.J.S. and the Medical Research Council of the United Kingdom for additional support. The whole-genome *M. tuberculosis* microarray was constructed and analyzed at St. George's Hospital Medical School as part of the multicollaborative microbial pathogen microarray facility, for which funding from The Wellcome Trust's Functional Genomics Resources Initiative is acknowledged.

Strains were kindly provided by Peter Gale and Hemu Patel (Leicester) and Grace Smith and Geoff Brooks (Birmingham). The work of Malcolm Yates and Francis Drobniowski (Central PHLS Mycobacterium Reference Unit, Dulwich, United Kingdom) in providing and interpreting IS6110-based RFLP results is gratefully acknowledged.

REFERENCES

1. Behr, M. A., M. A. Wilson, W. P. Gill, H. Salamon, G. K. Schoolnik, S. Rane, and P. M. Small. 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284**:1520-1523.

2. Belisle, J. T., and M. G. Sonnenberg. Isolation of genomic DNA from mycobacteria. *Methods Mol. Med.* **101**:31–44.
3. Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole. 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci. USA* **99**:3684–3689.
4. Brosch, R., W. J. Philipp, E. Stavropoulos, M. J. Colston, S. T. Cole, and S. V. Gordon. 1999. Genomic analysis reveals variation between *Mycobacterium tuberculosis* H37Rv and the attenuated *M. tuberculosis* H37Ra strain. *Infect. Immun.* **67**:5768–5774.
5. Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry III, F. Tekaiia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, B. G. Barrell, et al. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**:537–544.
6. Dorrell, N., J. A. Mangan, K. G. Laing, J. Hinds, D. Linton, H. Al-Ghusein, B. G. Barrell, J. Parkhill, N. G. Stoker, A. V. Karlyshev, P. D. Butcher, and B. W. Wren. 2001. Whole genome comparison of *Campylobacter jejuni* human isolates using a low-cost microarray reveals extensive genetic diversity. *Genome Res.* **11**:1706–1715.
7. Drobniowski, F. A., A. Gibson, M. Ruddy, and M. D. Yates. 2003. Evaluation and utilization as a public health tool of a national molecular epidemiological tuberculosis outbreak database within the United Kingdom from 1997 to 2001. *J. Clin. Microbiol.* **41**:1861–1868.
8. Dubnau, E., P. Fontan, R. Manganelli, S. Soares-Appel, and I. Smith. 2002. *Mycobacterium tuberculosis* genes induced during infection of human macrophages. *Infect. Immun.* **70**:2787–2795.
9. Ewer, K., J. Deeks, L. Alvarez, G. Bryant, S. Waller, P. Andersen, P. Monk, and A. Lalvani. 2003. Comparison of T-cell-based assay with tuberculin skin test for diagnosis of *Mycobacterium tuberculosis* infection in a school tuberculosis outbreak. *Lancet* **361**:1168–1173.
10. Fang, Z., C. Doig, D. T. Kenna, N. Smittipat, P. Palittapongarnpim, B. Watt, and K. J. Forbes. 1999. IS6110-mediated deletions of wild-type chromosomes of *Mycobacterium tuberculosis*. *J. Bacteriol.* **181**:1014–1020.
11. Fleischmann, R. D., D. Alland, J. A. Eisen, L. Carpenter, O. White, J. Peterson, R. DeBoy, R. Dodson, M. Gwinn, D. Haft, E. Hickey, J. F. Kolonay, W. C. Nelson, L. A. Umayam, M. Ermolaeva, S. L. Salzberg, A. Delcher, T. Utterback, J. Weidman, H. Khouri, J. Gill, A. Mikula, W. Bishai, W. R. Jacobs, Jr., Jr., J. C. Venter, and C. M. Fraser. 2002. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J. Bacteriol.* **184**:5479–5490.
12. Garnier, T., K. Eiglmeier, J. C. Camus, N. Medina, H. Mansoor, M. Pryor, S. Duthoy, S. Grondin, C. Lacroix, C. Monsempe, S. Simon, B. Harris, R. Atkin, J. Doggett, R. Mayes, L. Keating, P. R. Wheeler, J. Parkhill, B. G. Barrell, S. T. Cole, S. V. Gordon, and R. G. Hewinson. 2003. The complete genome sequence of *Mycobacterium bovis*. *Proc. Natl. Acad. Sci. USA* **100**:7877–7882.
13. Gascoyne-Binzi, D. M., R. E. Barlow, A. Essex, R. Gelletlie, M. A. Khan, S. Hafiz, T. A. Collins, R. Frizzell, and P. M. Hawkey. 2002. Predominant VNTR family of strains of *Mycobacterium tuberculosis* isolated from South Asian patients. *Int. J. Tuberc. Lung Dis.* **6**:492–496.
14. Ho, T. B., B. D. Robertson, G. M. Taylor, R. J. Shaw, and D. B. Young. 2000. Comparison of *Mycobacterium tuberculosis* genomes reveals frequent deletions in a 20 kb variable region in clinical isolates. *Yeast* **17**:272–282.
15. Huard, R. C., L. C. de Oliveira Lazzarini, W. R. Butler, D. van Soolingen, and J. L. Ho. 2003. PCR-based method to differentiate the subspecies of the *Mycobacterium tuberculosis* complex on the basis of genomic deletions. *J. Clin. Microbiol.* **41**:1637–1650.
16. Jones, T. F., C. L. Woodley, F. F. Fountain, and W. Schaffner. 2003. Increased incidence of the outbreak strain of *Mycobacterium tuberculosis* in the surrounding community after an outbreak in a jail. *South. Med. J.* **96**:155–157.
17. Kamerbeek, J., L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper, A. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal, and J. van Embden. 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol.* **35**:907–914.
18. Kato-Maeda, M., J. T. Rhee, T. R. Gingeras, H. Salamon, J. Drenkow, N. Smittipat, and P. M. Small. 2001. Comparing genomes within the species *Mycobacterium tuberculosis*. *Genome Res.* **11**:547–554.
19. Mazars, E., S. Lesjean, A. L. Banuls, M. Gilbert, V. Vincent, B. Gicquel, M. Tibayrenc, C. Locht, and P. Supply. 2001. High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proc. Natl. Acad. Sci. USA* **98**:1901–1906.
20. Mostowy, S., D. Cousins, J. Brinkman, A. Aranaz, and M. A. Behr. 2002. Genomic deletions suggest a phylogeny for the *Mycobacterium tuberculosis* complex. *J. Infect. Dis.* **186**:74–80.
21. Parsons, L. M., R. Brosch, S. T. Cole, A. Somoskovi, A. Loder, G. Bretzel, D. Van Soolingen, Y. M. Hale, and M. Salfinger. 2002. Rapid and simple approach for identification of *Mycobacterium tuberculosis* complex isolates by PCR-based genomic deletion analysis. *J. Clin. Microbiol.* **40**:2339–2345.
22. Plikaytis, B. B., N. Kurepina, C. L. Woodley, R. Fleischmann, B. Kreiswirth, and T. M. Shinnick. 1999. Multiplex PCR assay to aid in the identification of the highly transmissible *Mycobacterium tuberculosis* strain CDC1551. *Tuber. Lung Dis.* **79**:273–278.
23. Plikaytis, B. B., J. L. Marden, J. T. Crawford, C. L. Woodley, W. R. Butler, and T. M. Shinnick. 1994. Multiplex PCR assay specific for the multidrug-resistant strain W of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **32**:1542–1546.
24. Rose, A. M., J. M. Watson, C. Graham, A. J. Nunn, F. Drobniowski, L. P. Ormerod, J. H. Darbyshire, and J. Leese. 2001. Tuberculosis at the end of the 20th century in England and Wales: results of a national survey in 1998. *Thorax* **56**:173–179.
25. Salamon, H., M. Kato-Maeda, P. M. Small, J. Drenkow, and T. R. Gingeras. 2000. Detection of deleted genomic DNA using a semiautomated computational analysis of GeneChip data. *Genome Res.* **10**:2044–2054.
26. Savine, E., R. M. Warren, G. D. van der Spuy, N. Beyers, P. D. van Helden, C. Locht, and P. Supply. 2002. Stability of variable-number tandem repeats of mycobacterial interspersed repetitive units from 12 loci in serial isolates of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **40**:4561–4566.
27. Shafi, J., P. W. Andrew, and M. R. Barer. 2002. Microarrays for public health: genomic epidemiology of tuberculosis. *Comp. Funct. Genomics* **3**:362–365.
28. Skjot, R. L., I. Brock, S. M. Arend, M. E. Munk, M. Theisen, T. H. Ottenhoff, and P. Andersen. 2002. Epitope mapping of the immunodominant antigen TB10.4 and the two homologous proteins TB10.3 and TB12.9, which constitute a subfamily of the *esat-6* gene family. *Infect. Immun.* **70**:5446–5453.
29. Sola, C., I. Filliol, E. Legrand, S. Lesjean, C. Locht, P. Supply, and N. Rastogi. 2003. Genotyping of the *Mycobacterium tuberculosis* complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. *Infect. Genet. Evol.* **3**:125–133.
30. Sousa, M. C., and D. B. McKay. 2001. Structure of the universal stress protein of *Haemophilus influenzae*. *Structure (Cambridge)* **9**:1135–1141.
31. Sreevatsan, S., X. Pan, K. E. Stockbauer, N. D. Connell, B. N. Kreiswirth, T. S. Whittam, and J. M. Musser. 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc. Natl. Acad. Sci. USA* **94**:9869–9874.
32. Stewart, G. R., L. Wernisch, R. Stabler, J. A. Mangan, J. Hinds, K. G. Laing, D. B. Young, and P. D. Butcher. 2002. Dissection of the heat-shock response in *Mycobacterium tuberculosis* using mutants and microarrays. *Microbiology* **148**:3129–3138.
33. Warren, R. M., G. D. van der Spuy, M. Richardson, N. Beyers, C. Booyens, M. A. Behr, and P. D. van Helden. 2002. Evolution of the IS6110-based restriction fragment length polymorphism pattern during the transmission of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **40**:1277–1282.