High-Throughput Method for Detecting Genomic-Deletion Polymorphisms

Yves-Olivier Luc Goguet de la Salmonière,[1]* C. C. Kim,[2] A. G. Tsolaki,[1] A. S. Pym,[1]
M. S. Siegrist,[1] and Peter M. Small[1]

*Division of Infectious Diseases and Geographic Medicine, Department of Medicine,[1] and Department of
Microbiology and Immunology,[2] Stanford University Medical Center, Stanford, California*

DNA microarrays have been successfully used with different microorganisms, including *Mycobacterium tuberculosis*, to detect genomic deletions relative to a reference strain. However, the cost and complexity of the microarray system are obstacles to its widespread use in large-scale studies. In order to evaluate the extent and role of large sequence polymorphisms (LSPs) or insertion-deletion events in bacterial populations, we developed a technique, termed deligotyping, which hybridizes multiplex-PCR products to membrane-bound, highly specific oligonucleotide probes. The approach has the benefits of being low cost and capable of simultaneously interrogating more than 40 bacterial strains for the presence of 43 genomic regions. The deletions represented on the membrane were selected from previous comparative genomic studies and ongoing microarray experiments. Highly specific probes for these deletions were designed and attached to a membrane for hybridization with strain-derived targets. The targets were generated by multiplex PCR, allowing simultaneous amplifications of 43 different genomic loci in a single reaction. To validate our approach, 100 strains that had been analyzed with a high-density microarray were analyzed. The membrane accurately detected the deletions identified by the microarray approach, with a sensitivity of 99.9% and a specificity of 98.0%. The deligotyping technique allows the rapid and reliable screening of large numbers of *M. tuberculosis* isolates for LSPs. This technique can be used to provide insights into the epidemiology, genomic evolution, and population structure of *M. tuberculosis* and can be adapted for the study of other organisms.

Bacterial species exist as populations of genotypically similar but distinct organisms. Measuring the genetic diversity within a population is of particular importance for bacterial pathogens, as it can result in differences in virulence, antibiotic susceptibility, and other phenotypes important for the treatment and control of infectious diseases. Knowledge of the population structure of a bacterial species can also shed light on the epidemiology, evolution, and emergence of pathogenic organisms (33).

Although traditional techniques used for studying population genetics of bacteria, such as multilocus enzyme electrophoresis (30) and multilocus sequence typing (25), are useful for determining population structures, they target only a few genomic loci and provide no or limited insight into the functional consequences of genetic diversity. The elucidation of whole genome sequences and the development of DNA microarray technology have led to a new approach in which loci can be analyzed on the whole-genome scale. Genomic DNA from an isolate of interest can be hybridized to a DNA microarray constructed from a sequenced strain, revealing the presence or absence of genomic sequence at every interrogated locus (19).

Comparative genomic experiments using DNA microarrays have been carried out on a rapidly increasing number of bacteria (19). These studies have shown that gene content can vary between members of the same species as well as closely related species. For example, in species such as *Helicobacter pylori* (28) and *Staphylococcus aureus* (11), over 20% of genes are variably distributed among clinical isolates, and comparative genomics of *Listeria* species also revealed such polymorphism (13). Definition of variability in gene content has been useful in understanding short-term in vitro evolution (1), identifying putative virulence factors (28), reconstructing whole-genome-based phylogenies (32), identifying epidemic clones (31), and demonstrating the dissemination of antibiotic resistance genes by horizontal transfer (11).

Although DNA microarrays have been instrumental for the development of comparative bacterial genomics, they have limitations. High-density microarrays, such as the Affymetrix system (8), can represent intergenic sequences as well as coding sequences (CDS) with multiple oligonucleotides and their corresponding control oligonucleotides. This results in a precision tool for detecting genomic deletions but makes them expensive and difficult to adapt or customize. Spotted microarrays, which often represent each gene or CDS with a short PCR product fixed to a glass slide, are more amenable to adaptation but are still costly and technically demanding to develop. Macroarrays target a limited number of loci and are usually constructed by attaching PCR products to a membrane (12). Although more flexible than microarrays, they are not optimized for high-throughput analysis, as each strain normally requires an individual membrane and hybridization experiment. A high throughput capacity is particularly important for population genetics and for making robust associations between specific gene content and clinical phenotypes. More flexible and read-

---

* Corresponding author. Present address: Unité d'Hygiène et Prévention de l'Infection, Hôpital Henri Mondor, Assistance Publique-Hôpitaux de Paris, Université Paris-12, 51 Avenue Maréchal de Lattre de Tassigny, 94010 Créteil, France. Phone: (33) 1 49 81 45 95. Fax: (33) 1 49 81 45 98. E-mail: ygoguet@pasteur.fr.

ily available tools are therefore required to extend the genetic findings of microarray studies to large population-based or global collections of bacterial isolates.

In this study we have developed a rapid membrane-based technique, deligotyping, for simultaneously interrogating 43 genomic loci in 40 test strains of *Mycobacterium tuberculosis* in a single hybridization experiment, taking advantage of a PCR-based system. Deligotyping provides a new system for probing an important source of genetic diversity in this pathogen, with the potential to enhance our understanding of its population genomics and molecular epidemiology. Furthermore, our approach could be adapted for detecting sequences in other prokaryotes or eukaryotic species.

MATERIALS AND METHODS

Deletion detection. Some specific genomic loci targeted had already been well characterized during previous comparative genomic studies using a variety of techniques (1, 4, 15, 21, 24). Others were recently identified by using an Affymetrix-GeneChip (8) system described elsewhere (29). Briefly, it examines the genome of a test strain by using 236,360 oligonucleotides, designed by using the *M. tuberculosis* H37Rv genome (9) and linked to the GeneChip support. These 25-bp oligonucleotides are arranged in near-identical pairs. One corresponds to the exact H37Rv sequence, and the other contains a single-nucleotide mismatch, which acts as a control signal during analysis.

The data for successive probe pairs, obtained from hybridization of genomic DNA to the GeneChip, were analyzed by using Affymetrix software. A second-step analysis was then performed with a semiautomated computational program (29) to identify putative genomic deletions. Putative deletions larger than 1,000 bp, which have a high predictive value, were selected for further analysis. The genomic position of each deletion was confirmed by using PCR and sequencing.

Probe design. The probe selection software was written in Perl and executed under RedHat Linux 7.0. A FASTA-formatted sequence file is required as input, and the script uses command-line access to WU-BLAST 2.0 and PRIMER3. The algorithm splits each input sequence into all possible oligonucleotides of a specified range of lengths and selects optimal oligonucleotides based on melting temperature, uniqueness of the oligonucleotide within the genome sequence (assessed by BLAST), and self-annealing properties (assessed by PRIMER3). Details of the algorithm as well as the source code are available at http://falkow.stanford.edu/whatwedo/supplementarydata/ (probe sequences are available from the author upon request).

Primer design. Primers were designed by using DNAstar with the following parameters: primer length between 20 and 24 bp, melting temperature between 68 and 72°C, stability between 57.3 and 30.0 kcal/M, 3′ pentamer stability at 8.5 −kc/M, limitation of 2 bases for dimer and hairpin duplexing, and ignoring duplexing 8 bp from the 3′ end. (Primer sequences are available from the author upon request.)

M. tuberculosis strains and DNA extraction. One hundred *M. tuberculosis* strains were derived from a collection isolated between 1991 and 1999 in the bacteriological laboratories of San Francisco, Calif., hospitals. These strains were collected as part of an ongoing population-based molecular epidemiology project (18). They had previously been genotyped by IS*6110* restriction fragment length polymorphism analysis (35) and found to have different IS*6110* profiles, indicating that they were distinct strains. Four reference strains were used: H37Rv (9), H37Ra (6), CDC1551 (34), and BCG Pasteur. Genomic DNA was extracted from 10-ml cultures grown in 7H9 medium supplemented with oleic acid-albumin-dextrose-catalase (Difco) from stocks preserved at −80°C. The DNA was extracted by using a standard phenol-chloroform extraction procedure (36).

Multiplex PCR. Approximately 200 ng of genomic DNA was suspended in a final volume of 50 μl containing 1.5 mM MgCl$_2$; 0.250 mM (each) dATP, dGTP, dCTP, and dTTP (Invitrogen); 20 mM Tris-HCl (pH 8.4); 50 mM KCl; 7.5 U of recombinant *Taq* DNA polymerase (Invitrogen); and 10 pmol of each individual primer. For convenience, the primers were maintained as a mixture of the 43 pairs. One primer in each pair was biotinylated at the 5′ end (operon product synthesized by using the 10-1955 BiotinTEG Phosphoramidite from Glen Research). The PCR conditions were 4 min at 95°C, 50 s at 95°C, 50 s at 64°C, and 1 min at 72°C for 35 cycles, followed by one 7-min extension step at 72°C.

Membrane preparation and hybridization. To test which targets (genomic loci) had been amplified for each strain, the product of the multiplex PCR was hybridized simultaneously with the 43 probes covalently bound to a nylon membrane (Biotrans; ICN), in parallel lines (22). Hybridization was performed in a 45-lane blotter (Miniblotter 45; Immunetics, Cambridge, Mass.) by using 20 μl of the PCR products in 150 μl of 2× SSPE (1× SSPE is 0.18 M NaCl, 10 mM NaH$_2$PO$_4$, and 1 mM EDTA [pH 7.7]) at 60°C for 1 h. The membrane was washed twice in 250 ml of 2× SSPE–0.5% sodium dodecyl sulfate for 13 min at 62°C. Bound fragments were revealed by chemiluminescence after incubation with horseradish peroxidase-labeled streptavidin (Boehringer Mannheim) as for spoligotyping (14). The membrane was incubated for 2 min in 30 ml of ECL detection liquid and then exposed to ECL hyperfilm (Amersham).

RESULTS

The aim of this study was to develop a low-cost, widely applicable technique to screen large numbers of *M. tuberculosis* isolates for previously identified large sequence polymorphisms (LSPs). The deligotyping process is a membrane-based technique adapted for screening 40 strains simultaneously for the presence or absence of 43 specific sequences. The probes are oligonucleotides covalently linked to a membrane and specific for a sequence located within each LSP. The targets, used to hybridize with the membrane, are short multiplex-PCR products, spanning each of the oligonucleotide probes, and are generated by using DNA from each test strain. A PCR approach was selected to increase specificity and utility, given the low growth rate of *M. tuberculosis*.

Design of probes. The probes were designed to lie within LSPs. Ten of the probes were designed from LSPs chosen from previously published studies that used different techniques, including full-genome-sequence comparisons of *M. tuberculosis* isolates (2, 12, 29), microarray (1) or bacterial artificial chromosome array (15) analysis of the *M. tuberculosis* complex, and microarray studies of *M. tuberculosis* clinical isolates (21). The remaining 29 LSPs were identified in a microarray study of 100 *M. tuberculosis* clinical isolates from San Francisco, Calif., analyzed by using the Affymetrix-GeneChip system (8, 21, 29). Four of the 43 probes were designed to monitor loci with special properties. Two sequences were internal to the two prophages ϕRV1 and ϕRV2 identified in the H37Rv genome and found to be variable among clinical isolates of *M. tuberculosis* (15). Two other sequences were directed against genomic loci that show homology to regions implicated in horizontal transfer between nonpathogenic mycobacteria (23). The 39 remaining LSPs had a mean length of 3,999 bases (from bp 326 to 12734) and were predicted to delete 110 CDS and disrupt 56 CDS. Two of these LSPs were sequences present in *M. tuberculosis* CDC1551 but not in H37Rv. Eleven LSPs were sequences usually absent in *Mycobacterium bovis* strains (1, 15). One LSP is the overlapping RD1 deletions of BCG and *Mycobacterium microti* (3).

Most of the probes were directed against coding sequences and against genes with known or probable functions. Sequencing of the junction sequences for each deletion revealed that the ends of some LSPs were overlapping. In these cases the probes were preferentially designed against the nonoverlapping portions of each LSP in order to maintain specificity.

The GC-rich *M. tuberculosis* genome and the existence of large gene families with repetitive motifs (9) made manual selection of high-specificity probes very difficult. We therefore developed a script which analyzes every possible oligonucleotide of a given sequence for specificity (by using calls to BLAST) and optimal physical characteristics (by using calls to

PRIMER3). In addition to being specific for the genomic regions they represent, the probes were designed to have physical characteristics similar to one another to allow simultaneous hybridization.

For the oligonucleotide size, an optimum of 28 bp was chosen to avoid poor hybridization signals due to nucleotide polymorphisms. Probes containing different combinations of mismatches and/or nucleotide deletions were constructed and tested for hybridization to two target sequences. The results indicated that with a 28-mer oligonucleotide, single-nucleotide polymorphism insertions or deletions would not give a null hybridization signal suggestive of a deletion (data not shown). To facilitate our choice of possible probes, we allowed flexibility in oligonucleotide size (± 2 nucleotides) in order to provide sufficient leeway for primers of the required specificity to be selected with a GC base content close to the overall 65.6% genomic content estimated for *M. tuberculosis* (9). Therefore, the smallest probe is 26 bases long and the largest one 30 bases, with a mean length of 29 bases and a mean GC content of 60%.

All of the probe oligonucleotides have a 5′ amine group to facilitate covalent linkage to a nylon membrane. We anticipated that the covalent attachment would allow membranes to be used repeatedly, and we confirmed that they could be used for at least five hybridizations (data not shown).

Multiplex-PCR design. In order to maximize the speed and convenience of the technique, the hybridization targets were generated by using multiplex PCR. More than 50 unique amplifications for different targets were successfully performed simultaneously (data not shown), although the technique described here utilizes 43 simultaneous PCRs. To facilitate the design of the PCR primers, some constraints were needed to allow this simultaneous complex reaction (detailed in Materials and Methods). The sizes of the products of each reaction were restricted as much as possible and produced a final set of targets that ranged from 113 to 282 bp, with a mean size of 156 bp. In general, the PCR primers were designed to be within a single CDS or intergenic region located within the selected LSP, although in two cases one primer was located outside the LSP boundary.

To minimize cross-hybridization to other regions of the genome, a BLAST search of each primer was conducted against the appropriate complete genome sequence. One primer of each pair is biotinylated to allow the detection of the hybridization with a streptavidin peroxidase reagent. The mean size of the primers is 24 bp, and their GC content is 65%. Because the 43 different PCRs did not always result in similar yields, the quantity of probe attached to the membrane needed to be adjusted in some cases. Altering the concentration of the primers in the PCR mix did not improve homogeneity.

Deligotyping process. Because one of the potential applications of deligotyping is as a new genotyping technique for molecular epidemiologic studies, we adapted our protocol to be compatible with spoligotyping (10, 14, 37), a technique that is already in common use and that takes advantage of a single locus with variable sequences between conserved direct repeat sequences (16), so that the two techniques could be performed simultaneously with the same reverse line blotting technique, apparatus, buffer, and protocol. The only change to the spoligotyping protocol was to increase the washing temperature by

2°C to enhance the specificity. An example of a hybridized membrane is shown in Fig. 1A.

Evaluation of deligotyping. Various approaches were used to evaluate deligotyping. Two strains of *M. tuberculosis*, H37Rv and the clinical isolate CDC1551, have been completely sequenced and compared (2, 12). Two sequences present in CDC1551 but absent in H37RV are represented on the deligotyping membrane. In lanes 1 and 3 of Fig. 1B are H37Rv and CDC1551, respectively, with rows 3 and 4 corresponding to these deletions. Four other LSPs were present in H37Rv but absent in CDC1551, corresponding to rows 24, 29, 34, and 36, and the hybridization signals were also appropriate for these probes. Thirteen other probe targets were evaluated with *M. bovis*, *M. microti* and BCG (Fig. 1B, lanes 4, 5, and 6). The lack of hybridization for these 13 probes is also in agreement with the sequence data.

Many of the probes on the membrane are targeted at LSPs detected by an Affymetrix-GeneChip-based analysis of 100 strains of *M. tuberculosis*. An opportunity therefore existed to evaluate the sensitivity and specificity of deligotyping by using the Affymetrix microarray-based data as the "gold standard." Genomic DNAs from these 100 isolates were subjected to deligotyping in a blinded fashion. The results in Table 1 show a high degree of reproducibility between the relatively labor-intensive and expensive technique used to discover the LSPs and the macroarray presented here. In total, 201 LSPs were absent in the 100 test isolates represented by 35 probes on the membrane. Deligotyping identified 197 of these deletions accurately. Conversely, of the 3,771 regions that were present in the test isolates, deligotyping gave a false-negative result in only 2 instances. The few discrepancies are rare enough to conclude that this technique is suitable as an efficient screening tool for polymorphisms identified by a microarray-based approach.

DISCUSSION

Bacterial genetics has traditionally concentrated on single-nucleotide polymorphisms as the basis for calibrating and understanding bacterial evolution. However, comparative bacterial genomics following the advent of whole-genome sequencing has suggested that alterations in genome content are also an important source of genetic diversity in some species of bacteria. Previous comparative genomic studies using microarrays and other techniques have identified a group of LSPs in clinical isolates of *M. tuberculosis* (15, 21) as well as among the subspecies of the *M. tuberculosis* complex (1). These polymorphisms can lead to the loss or disruption of multiple CDS and therefore may have functional consequences for the organism. Competing mechanisms of gene acquisition, by horizontal transfer or gene duplication, and gene loss following genomic deletions result in differences in genome content, and these differences may have important phenotypic consequences, such as resistance to antibiotics (27) and enhancement or loss of virulence (17, 26). Understanding more fully how the genome content varies within populations is an important goal of bacterial genetics. In addition, LSPs are thought to occur as a result of a genomic deletion event rather than an insertion of DNA following horizontal transfer (7). These deletions are irreversible and often unique events, and they therefore rep-
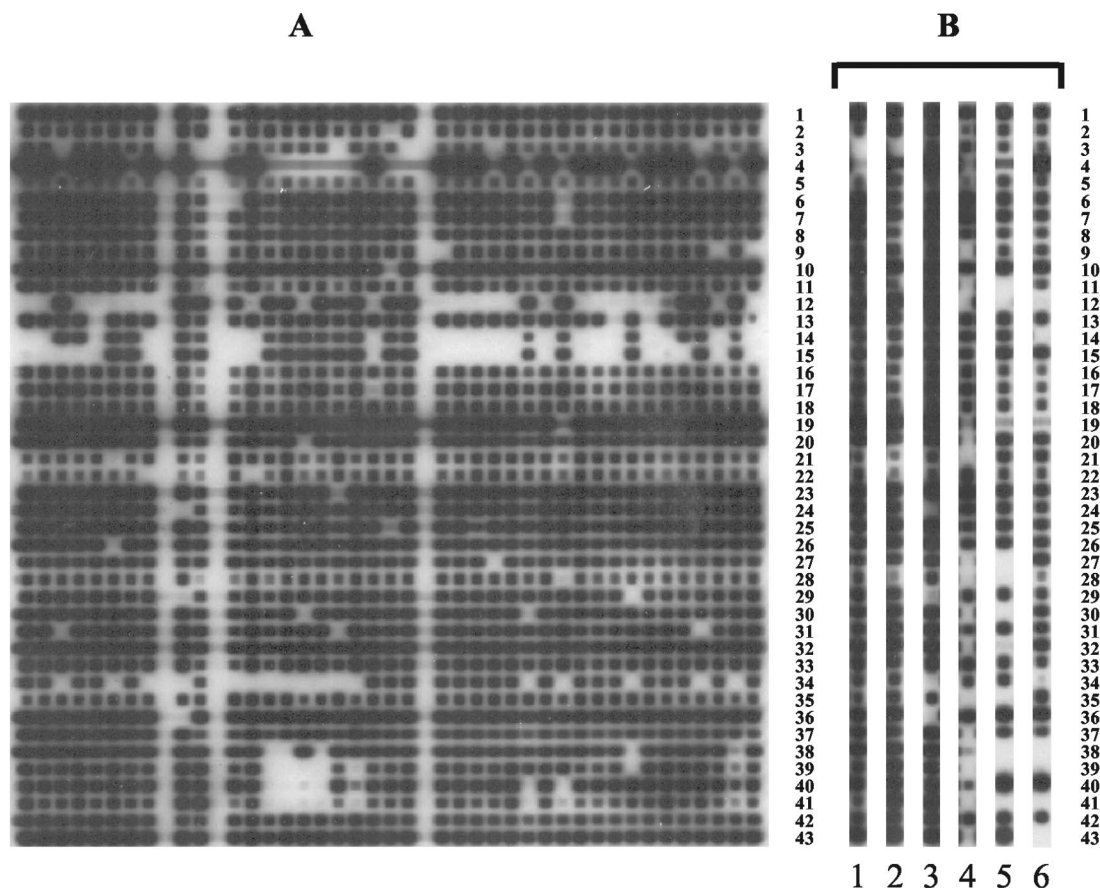
FIG. 1. Example of a hybridized membrane and results for specific strains. (A) A representative hybridized membrane. (B) Hybridization pattern obtained with the three reference strains H37Rv (lane 1), H37Ra (lane 2), and CDC1551 (lane 3); a BCG strain (lane 4); an *M. bovis* strain (lane 5); and an *M. microti* strain (lane 6). The lanes represent different strains; the rows represent different deletions. Black spots indicate that the DNA is present.

resent attractive polymorphisms for molecular typing as well as for constructing robust phylogenies.

We present here a technique, deligotyping, to examine rapidly the genome content at 43 genomic loci of interest in *M. tuberculosis*. Most of these loci are LSPs that are present in some strains but absent from others, and the majority of these LSPs were discovered by using microarrays. The microarray-based approach is unparalleled in detecting a priori whether any of a very large number of loci are missing from an isolate. However, once polymorphic loci are identified, it is a fundamentally different task to determine whether a specific region is missing from a very large number of isolates. We therefore designed deligotyping to be a high-throughput, high-resolution, and simple technique that could be used in laboratories without microarray equipment. Although we developed deligotyping for *M. tuberculosis*, it could readily be adapted for the study of any prokaryotic or eukaryotic species for which LSP information is available.

A key feature of any successful simultaneous hybridization technique is specificity, particularly when the targeted sequences include repetitive sequence or members of homologous gene families. In addition it was important to ensure that none of the targets gave ambiguous "grey" signals that would prevent or hinder an easy (and ultimately automated) reading of the membranes. We therefore focused on the specificity of

the membrane-bound probes, and our software allowed us to predict in silico which sequences would perform well as probes. Our evaluation of deligotyping showed that it performed at a level comparable to that of the Affymetrix-PCR sequencing

TABLE 1. Comparison of results obtained by deligoyping and Affymetrix-GeneChip analysis for 100 isolates[a]

Locus detected as follows by Affymetrix-GeneChip	No. (%) of loci detected as follows by deligotyping:		Total
	Present	Absent	
Present	3,769 (94.9)	2 (0.1)	3,771
Absent	4 (0.1)	197 (5.0)	201
Total	3,773	199	3,972

[a] All strain and locus information from the Affymetrix-GeneChip analysis is included here. Using the GeneChip results as the standard, the deligotyping specificity is 0.980, the sensitivity is 0.999, the positive predictive value is 0.999, and the negative predictive value is 0.990. The sensitivity is calculated by dividing the number of LSPs detected as present by both deligotyping and microarray by the number of LSPs detected as present by microarrray, the specificity is calculated by dividing the number of LSPs detected as absent by both deligotyping and microarray by the number of LSPs detected as absent by microarrray, the positive predictive value is calculated by dividing the number of LSPs detected as present by both deligotyping and microarray by the number of LSPs detected as present by deligotyping, and the negative predictive value is calculated by dividing the number of LSPs detected as absent by both deligotyping and microarray by the number of LSPs detected as absent by deligotyping.

strategy; the results demonstrate 99.8% agreement between the two techniques (Table 1). We are uncertain as to why we did not observe 100% agreement, but the rare discrepancies could be due to a difference in sensitivity to divergence in sequence identity or differences in the probe characteristics between the microarray-based and deligotyping approaches. Nevertheless, the agreement between the approaches in conjunction with analysis of the sequenced *Mycobacterium* strains confirms that deligotyping can detect LSPs with high specificity.

Specificity in deligotyping is also enhanced by using a PCR product as a target, effectively enriching the DNA composition for those sequences of interest. We found that a multiplex PCR involving as many as 50 individual reactions was feasible, and our observations suggest that the number of probes and targets can be further increased.

The majority of the sequences targeted on the deligotyping membrane were within LSPs. Although LSPs could represent the horizontal transfer of DNA from one strain to another, they are thought to be the result of genomic deletions, as their boundaries usually disrupt CDS and their GC content is conserved (5, 9). Since a deletion event is irreversible, it represents a fixed mutation ideal for molecular phylogeny and epidemiology. We therefore ensured that the experimental conditions and apparatus for deligotyping were compatible with spoligotyping, which is widely used for the molecular epidemiology of tuberculosis (20, 37), allowing the two procedures to be carried out in parallel. Spoligotyping targets a single genomic locus that is characterized by short repeats separated by unique sequences (16). Although this polymorphic region is useful for distinguishing strains, its evolutionary history is complex, blurring its phylogenetic signal (38). In contrast, LSPs in *M. tuberculosis* represent irreversible and usually unique events and are therefore attractive mutations for phylogenetics. In summary, deligotyping represents a highly reproducible technique for detecting LSPs among populations of bacterial strains. Genetic data derived from deligotyping studies will be useful for understanding the functional consequences of gene deletion and may provide insights into evolution, phylogenetics, and molecular epidemiology.

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. **Betts, J. C., P. Dodson, S. Quan, A. P. Lewis, P. J. Thomas, K. Duncan, and R. A. McAdam.** 2000. Comparison of the proteome of *Mycobacterium tuberculosis* strain H37Rv with clinical isolate CDC 1551. Microbiology **146 Pt. 12:**3205–3216.
3. **Brodin, P., K. Eiglmeier, M. Marmiesse, A. Billault, T. Garnier, S. Niemann, S. T. Cole, and R. Brosch.** 2002. Bacterial artificial chromosome-based comparative genomic analysis identifies *Mycobacterium microti* as a natural ESAT-6 deletion mutant. Infect. Immun. **70:**5568–5578.
4. **Brosch, R., S. V. Gordon, A. Billault, T. Garnier, K. Eiglmeier, C. Soravito, B. G. Barrell, and S. T. Cole.** 1998. Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. Infect. Immun. **66:**2221–2229.
5. **Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Par-** sons, 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.
6. **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.
7. **Brosch, R., A. S. Pym, S. V. Gordon, and S. T. Cole.** 2001. The evolution of mycobacterial pathogenicity: clues from comparative genomics. Trends Microbiol. **9:**452–458.
8. **Chee, M., R. Yang, E. Hubbell, A. Berno, X. C. Huang, D. Stern, J. Winkler, D. J. Lockhart, M. S. Morris, and S. P. Fodor.** 1996. Accessing genetic information with high-density DNA arrays. Science **274:**610–614.
9. **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. Tekaia, 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, A. Krogh, A. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M.-A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell.** 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature **393:**537–544.
10. **Dale, J. W., D. Brittain, A. A. Cataldi, D. Cousins, J. T. Crawford, J. Driscoll, H. Heersma, T. Lillebaek, T. Quitugua, N. Rastogi, R. A. Skuce, C. Sola, D. Van Soolingen, and V. Vincent.** 2001. Spacer oligonucleotide typing of bacteria of the *Mycobacterium tuberculosis* complex: recommendations for standardised nomenclature. Int. J. Tuberc. Lung Dis. **5:**216–219.
11. **Fitzgerald, J. R., D. E. Sturdevant, S. M. Mackie, S. R. Gill, and J. M. Musser.** 2001. Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. Proc. Natl. Acad. Sci. USA **98:**8821–8826.
12. **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., J. C. Venter, and C. M. Fraser.** 2002. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. J. Bacteriol. **184:**5479–5490.
13. **Glaser, P., L. Frangeul, C. Buchrieser, C. Rusniok, A. Amend, F. Baquero, P. Berche, H. Bloecker, P. Brandt, T. Chakraborty, A. Charbit, F. Chetouani, E. Couve, A. de Daruvar, P. Dehoux, E. Domann, G. Dominguez-Bernal, E. Duchaud, L. Durant, O. Dussurget, K. D. Entian, H. Fishi, F. Garcia-del Portillo, P. Garrido, L. Gautier, W. Goebel, N. Gomez-Lopez, T. Hain, J. Hauf, D. Jackson, L. M. Jones, U. Kaerst, J. Kreft, M. Kuhn, F. Kunst, G. Kurapkat, E. Madueno, A. Maitournam, J. M. Vicente, E. Ng, H. Nedjari, G. Nordsiek, S. Novella, B. de Pablos, J. C. Perez-Diaz, R. Purcell, B. Remmel, M. Rose, T. Schlueter, N. Simoes, A. Tierrez, J. A. Vazquez-Boland, H. Voss, J. Wehland, and P. Cossart.** 2001. Comparative genomics of Listeria species. Science **294:**849–852.
14. **Goguet de la Salmoniere, Y. O., H. M. Li, G. Torrea, A. Bunschoten, J. van Embden, and B. Gicquel.** 1997. Evaluation of spoligotyping in a study of the transmission of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **35:**2210–2214.
15. **Gordon, S. V., R. Brosch, A. Billault, T. Garnier, K. Eiglmeier, and S. T. Cole.** 1999. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. Mol. Microbiol. **32:**643–655.
16. **Groenen, P. M., A. E. Bunschoten, D. van Soolingen, and J. D. van Embden.** 1993. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. Mol. Microbiol. **10:**1057–1065.
17. **Hacker, J., and E. Carniel.** 2001. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. EMBO Rep. **2:**376–381.
18. **Jasmer, R. M., J. A. Hahn, P. M. Small, C. L. Daley, M. A. Behr, A. R. Moss, J. M. Creasman, G. F. Schecter, E. A. Paz, and P. C. Hopewell.** 1999. A molecular epidemiologic analysis of tuberculosis trends in San Francisco, 1991–1997. Ann. Intern. Med. **130:**971–978.
19. **Joyce, E. A., K. Chan, N. R. Salama, and S. Falkow.** 2002. Redefining bacterial populations: a post-genomic reformation. Nat. Rev. Genet. **3:**462–473.
20. **Kato-Maeda, M., P. J. Bifani, B. N. Kreiswirth, and P. M. Small.** 2001. The nature and consequence of genetic variability within *Mycobacterium tuberculosis*. J. Clin. Investig. **107:**533–537.
21. **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.
22. **Kaufhold, A., A. Podbielski, G. Baumgarten, M. Blokpoel, J. Top, and L. Schouls.** 1994. Rapid typing of group A streptococci by the use of DNA amplification and non radioactive allele-specific oligonucleotide probes. FEMS Microbiol. Lett. **119:**19–26.
23. **Le Dantec, C., N. Winter, B. Gicquel, V. Vincent, and M. Picardeau.** 2001. Genomic sequence and transcriptional analysis of a 23-kilobase mycobacte-

rial linear plasmid: evidence for horizontal transfer and identification of plasmid maintenance systems. J. Bacteriol. **183:**2157–2164.

24. **Mahairas, G. G., P. J. Sabo, M. J. Hickey, D. C. Singh, and C. K. Stover.** 1996. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis.* J. Bacteriol. **178:**1274–1282.

25. **Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. USA **95:**3140–3145.

26. **Maurelli, A. T.** 1989. Temperature regulation of virulence genes in pathogenic bacteria: a general strategy for human pathogens? Microb. Pathog. **7:**1–10.

27. **Rowe-Magnus, D. A., A. M. Guerout, P. Ploncard, B. Dychinco, J. Davies, and D. Mazel.** 2001. The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons. Proc. Natl. Acad. Sci. USA **98:**652–657.

28. **Salama, N., K. Guillemin, T. K. McDaniel, G. Sherlock, L. Tompkins, and S. Falkow.** 2000. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. Proc. Natl. Acad. Sci. USA **97:**14668–14673.

29. **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.

30. **Smith, J. M., N. H. Smith, M. O'Rourke, and B. G. Spratt.** 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. USA **90:**4384–4388.

31. **Smoot, J. C., K. D. Barbian, J. J. Van Gompel, L. M. Smoot, M. S. Chaussee, G. L. Sylva, D. E. Sturdevant, S. M. Ricklefs, S. F. Porcella, L. D. Parkins, S. B. Beres, D. S. Campbell, T. M. Smith, Q. Zhang, V. Kapur, J. A. Daly, L. G. Veasy, and J. M. Musser.** 2002. Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks. Proc. Natl. Acad. Sci. USA **99:**4668–4673.

32. **Snel, B., P. Bork, and M. A. Huynen.** 1999. Genome phylogeny based on gene content. Nat. Genet. **21:**108–110.

33. **Spratt, B. G., and M. C. Maiden.** 1999. Bacterial population genetics, evolution and epidemiology. Philos. Trans. R. Soc. London B **354:**701–710.

34. **Valway, S. E., M. P. Sanchez, T. F. Shinnick, I. Orme, T. Agerton, D. Hoy, J. S. Jones, H. Westmoreland, and I. M. Onorato.** 1998. An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis.* N. Engl. J. Med. **338:**633–639.

35. **van Embden, J. D., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach, B. Gicquel, P. Hermans, C. Martin, R. McAdam, T. M. Shinnick, and P. M. Small.** 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J. Clin. Microbiol. **31:**406–409.

36. **van Soolingen, D., P. W. Hermans, P. E. de Haas, D. R. Soll, and J. D. van Embden.** 1991. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. J. Clin. Microbiol. **29:**2578–2586.

37. **van Soolingen, D., L. Qian, P. E. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Z. Qing, D. Enkhsaikan, P. Nymadawa, and J. D. van Embden.** 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J. Clin. Microbiol. **33:**3234–3238.

38. **Warren, R. M., E. M. Streicher, S. L. Sampson, G. D. Van Der Spuy, M. Richardson, D. Nguyen, M. A. Behr, T. C. Victor, and P. D. Van Helden.** 2002. Microevolution of the direct repeat region of *Mycobacterium tuberculosis*: implications for interpretation of spoligotyping data. J. Clin. Microbiol. **40:**4457–4465.