Comprehensive identification of conditionally essential genes in mycobacteria

Christopher M. Sassetti*, Dana H. Boyd†, and Eric J. Rubin*‡

*Department of Immunology and Infectious Diseases, Harvard School of Public Health, 667 Huntington Avenue, Boston, MA 02115; and †Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115

Edited by R. John Collier, Harvard Medical School, Boston, MA, and approved September 10, 2001 (received for review June 1, 2001)

An increasing number of microbial genomes have been completely sequenced, and the identified genes are categorized based on their homology to genes of known function. However, the function of a large number of genes cannot be determined on this basis alone. Here, we describe a technique, transposon site hybridization (TraSH), which allows rapid functional characterization by identifying the complete set of genes required for growth under different conditions. TraSH combines high-density insertional mutagenesis with microarray mapping of pools of mutants. We have made large pools of independent transposon mutants in mycobacteria by using a *mariner*-based transposon and efficient phage transduction. By using TraSH, we have defined the set of genes required for growth of *Mycobacterium bovis* bacillus Calmette–Guérin on minimal but not rich medium. Genes of both known and unknown functions were identified. Of the genes with known functions, nearly all were involved in amino acid biosynthesis. TraSH is a powerful method for categorizing gene function that should be applicable to a variety of microorganisms.

Since the first genome sequence of a free-living organism was determined in 1995, the genomes of more than fifty organisms have been sequenced. Genome sequences representing the three major domains of life—bacteria, archaea, and eukaryota—have been determined. The availability of the sequences of entire genomes promises to allow us to understand the roles of the thousands of genes that make each organism unique. Currently, however, functions can be assigned only to about 60% of the genes of any organism, and the individual analysis of the remaining genes represents a daunting task. Thus, methods that can be used to study the functions of thousands of genes in parallel are needed to make efficient use of genomic information.

A genome sequence provides a catalogue of all of the genes of an organism. This knowledge has allowed the construction of DNA microarrays, which have been used to measure changes in global gene expression in response to varying environmental conditions (1). Genes that are regulated by a particular stimulus are assumed to be important for adapting to that condition. In addition, a putative function can be assigned to genes that are found to be coordinately regulated with genes of a known pathway. Although this kind of analysis is extremely valuable, it has significant limitations. For example, expression analysis is not informative for genes that are constitutively expressed or genes whose expression or function is modified by posttranscriptional mechanisms.

Most of the limitations of expression analysis are circumvented by techniques that allow functional assignments to be made based on whether genes are essential for survival under a particular condition. Transposon mutagenesis has been invaluable for this type of analysis, because very complex mutant libraries can be obtained readily and each mutation is marked by an insertion (2).

Previously, two general approaches have been used to monitor the survival of individual mutants in a pool. Signature tagged mutagenesis (STM) utilizes a series of transposons (generally 50–100) that are each tagged with a unique nucleotide sequence (3). This procedure allows simultaneously monitoring of the presence of each mutant in a pool. Although STM has been particularly useful for the study of pathogenesis (4), it is limited by the relatively small size of each pool and the labor required to perform a comprehensive screen. Alternatively, genetic footprinting utilizes PCR to map transposon insertions relative to a fixed site on the chromosome (5–7). Products that are amplified from a pool of thousands of mutants are analyzed by gel electrophoresis. Loci that are unable to sustain insertions are identified and assumed to be essential for survival under that particular condition. This technique allows the analysis of much larger mutant pools than STM, but only a relatively small number of genes can be screened in each PCR reaction. Although much of the analysis can be automated, a global analysis of the genome is still a very laborious process.

Here, we describe a new technique, transposon site hybridization (TraSH), which combines transposon mutagenesis and microarray hybridization. TraSH comprehensively identifies conditionally essential genes that are required for growth under one condition but not another. We have demonstrated the utility of TraSH by identifying mutations that result in auxotrophy in *Mycobacterium bovis* bacillus Calmette–Guérin (BCG).

Mycobacteria of the tuberculosis complex, which includes both *M. tuberculosis* and *M. bovis,* are the causative agents of tuberculosis (TB) and exert an enormous disease burden worldwide. It is estimated that one-third of the world's population is latently infected with *M. tuberculosis*, and 10% of these individuals will eventually develop disease (8). Because suppression of cell-mediated immunity greatly increases the risk of developing disease, the current HIV pandemic is expected to result in continually rising TB rates. Efforts to control TB would be greatly aided by the availability of a more effective vaccine. The current vaccine (*M. bovis* BCG) is an attenuated strain that was isolated after serial culture *in vitro*. BCG has demonstrated only weak and variable efficacy in adults (9). Many auxotrophic mutants of *M. tuberculosis* and *M. bovis* are unable to cause disease and represent possible attenuated vaccine strains (10–12). By defining mutations in genes of both known and unknown function that result in auxotrophy, this study provides several targets for the rational attenuation of *M. tuberculosis.*

TraSH has proven to be particularly useful for identifying conditionally essential genes in mycobacteria, which have been difficult to study by conventional methods because of their slow growth and the lack of sophisticated genetic tools available. We also expect that TraSH will prove to be generally applicable to any haploid organism for which a significant portion of the genome sequence is known.

Materials and Methods

Bacterial Strains and Growth Conditions. All mycobacterial strains were grown in Middlebrook 7H9 broth (Difco) supplemented with 0.05% Tween 80 and AD (0.5% BSA, 0.2% dextrose, 0.085% NaCl)

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: STM, signature tagged mutagenesis; TraSH, transposon site hybridization; BCG, bacillus Calmette–Guérin; INH, isoniazid.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AF411123).

‡To whom reprint requests should be addressed. E-mail: erubin@hsph.harvard.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

when grown in liquid culture. *Mycobacterium smegmatis mc2–155* (13) and *Mycobacterium fortuitum MBH1* (a clinical isolate provided by Gloria Brown, Massachusetts General Hospital, Boston, MA) were cultured on LB agar. *Mycobacterium bovis* BCG Pasteur and *Mycobacterium tuberculosis* H37Rv were cultured on Middlebrook 7H10 agar supplemented with AD or Middlebrook OADC enrichment (Becton Dickinson), respectively. For selection of mutants, solid media were supplemented with kanamycin at 20 (*M. bovis* BCG and *M. tuberculosis*), 50 (*M. smegmatis*), or 250 μg/ml (*M. fortuitum*). Selection of auxotrophs was performed on either minimal medium (7H10 agar supplemented with AD) or rich medium (7H11 agar supplemented with AD/0.32 mM thymine/5 mM adenine/0.3 mM guanine/50 μM thiamine/0.1 mM diaminopimelic acid).

Construction of an *M. tuberculosis* Microarray. PCR was used to amplify a fragment corresponding to each predicted ORF of *M. tuberculosis* H37Rv. To ensure that each primer pair would amplify only one product, each ORF in the genome was compared with all others by using BLAST (14). Sequences with an E value of <1e-5 were used to generate "mispriming libraries" for the next step. A set of 10 oligonucleotide primer pairs was generated for each gene with PRIMER3 (S. Rozen and H. J. Skaletsky (1997) at http.//www-genome.wi.mit.edu/genome_software/other/primer3.html.), excluding primers that might prime on the mispriming library. Then, BLAST was used to eliminate primer pairs that generate products that could crosshybridize to other genes (>77% identity). Primer pairs that met these criteria were found for 3,855 of the 3,924 predicted ORFs. Many of the remaining ORFs represent transposons or phage-related genes. Thus, each predicted ORF of *M. tuberculosis H37Rv* was represented by a fragment that was predicted to have minimal cross-hybridization with other *M. tuberculosis* sequences. Fragments varied in length from 70 to 499 bp, with a median length of ≈350. Forward and reverse primers contained 5′ extensions: GGCATCTAGAG and CCGCACTAGTCCTC, respectively. PCRs with these gene-specific primers was performed by using 1.25 units of *Taq* and 0.15 units of Pfu polymerase (Stratagene) in 50 μl reactions containing 2.5 mM $MgCl_2$, 10% (vol/vol) DMSO and H37Rv genomic DNA as template. PCR conditions were 94°C for 2 min; 30 cycles of 94°C for 30 sec, 60°C for 30 sec, 72°C for 1 min; and 72°C for 5 min. These products were diluted 1:100, and 2.5 μl was used as a template for PCR with the following universal primers (both contain 5′ amino modifications including a three-carbon linker; Genosys, The Woodlands, TX): GAACCGATAGGCATCTAGAG and GAAATCCACCGCACTAGTCCTC. PCR was performed as follows: 94°C for 2 min; 3 cycles of 94°C for 30 sec, 40°C for 30 sec, 72°C for 1 min; 20 cycles of 94°C for 30 sec, 60°C for 30 sec, 72°C for 1 min; and 72°C for 5 min. PCR of all but 11 ORFs produced single products when analyzed by gel electrophoresis. Those reactions that did not produce products of appropriate size or produced multiple products were not included in further analysis. PCR products were purified by using multiscreen PCR plates (Millipore) and arrayed onto 3D-Link slides (Surmodics, Eden Prairie, MN) in duplicate, as recommended by the manufacturer.

Construction of Transposon Mutant Libraries. The transposon donor phagemid, ϕMycoMarT7, was constructed in several steps starting with pMycoMar (15). ϕMycoMarT7 includes a transposon that encodes a kanamycin resistance gene and T7 promoters oriented so as to promote transcription into adjacent chromosomal DNA. The transposon is flanked by 29-bp inverted repeats. The entire sequence of the MycoMarT7 transposon has been deposited in GenBank (GenBank accession no. AF411123). The phagemid also contains the highly active C9 *Himar1* transposase gene (16). A plasmid containing these elements and a bacteriophage λ cos site was cloned into *Pac*I-digested ϕAE87, and phage stocks were generated as described (17). For transduction, mycobacterial cultures were grown to OD_{600} of 0.8–1.0 in liquid culture. Cells were washed twice with MP buffer (50 mM Tris, pH 7.5/150 mM NaCl/10 mM $MgSO_4$/2 mM $CaCl_2$) and resuspended in 1/10 of the original culture volume in MP. Cells were infected with 10^{10} phage per ml of original culture for 3 h at 37°C. Transduced cells were plated directly onto solid media and cultured at 37°C. Libraries were collected by scraping colonies off plates. Aliquots were used either for genomic DNA isolation (see below) or frozen at −80°C.

Generation of Labeled Target for TraSH Experiments. Genomic DNA was isolated from mutant pools as described (18). A quantity (0.5 μg) of DNA/reaction was partially digested separately with *Hin*PI and *Msp*I (New England Biolabs). Fragments (500–2,000 bp) were purified from a 1% agarose gel with a Qiaquick gel purification kit (Qiagen, Chatsworth, CA). The resulting DNA was ligated to a 1,000-fold molar excess of the following adapter: CGACCACGACCA (includes 3′ C6-TFA-amino modification; Genosys) and AGTCTCGCAGATGATAAGGTGGTCGTGGT, by using T4 DNA ligase at 16°C. Ligated DNA (1 μl) was used as a template in two separate PCR reactions containing the adapter primer (GTCCAGTCTCGCAGATGATAAGG) and one of the transposon primers (primer 1-CCCGAAAAGTGCCACCTAAATTGTAAGCG or primer 2-CGCTTCCTCGTGCTTTACGGTATCG). PCR conditions were as follows: 94°C for 2 min; 5 cycles of 94°C for 30 sec, 72°C or 69°C for 30 sec, and 72°C for 1.5 min; 5 cycles of 94°C for 30 sec, 70°C or 67°C for 30 sec, and 70°C for 1.5 min; 15–25 cycles of 94°C for 30 sec, 68°C or 65°C for 30 sec, and 68°C for 1.5 min + 5 sec per cycle; and 72°C for 5 min. The low and high annealing temperatures were used for transposon primer 1 and 2, respectively. PCR products (250–500 bp) were purified from a 2% agarose gel and used as template for *in vitro* transcription reactions with T7 polymerase (Ambion, Austin, TX). After transcription, the template DNA was removed by using RQ1 DnaseI (Ambion), followed by extraction with phenol:chloroform and gel-filtration by using Centrisep-20 columns (Princeton Separations, Adelphia, NJ). RNA (1 μg) was labeled with either biotin or fluorescein by using ULS reagents (NEN, Boston, MA), purified with a Centrisep column, and concentrated by evaporation.

Microarray Hybridization. Microarrays were prehybridized in 5× SSC (1× SSC = 0.15 M sodium chloride/0.015 M sodium citrate, pH 7), 0.1% SDS, 100 μg/ml tRNA, and 1% BSA for 1 h, rinsed with water, and dried. Labeled RNAs were mixed in 50% (vol/vol) formamide, 5× SSC, 0.1% SDS, and 100 μg/ml tRNA, denatured at 65°C for 3 min, and hybridized to the microarray. Hybridizations were performed in a 40-μl volume under a coverslip at 42°C overnight. The arrays were washed in 5× SSC, 0.1% SDS at 42°C, then washed with 0.2× SSC, 0.1% SDS at room temperature, and finally washed in 0.2× SSC. Hybridized RNA was detected with the Micromax TSA system (NEN) using the manufacturer's protocol. Images were acquired with a Scanarray microarray scanner (GSI Lumonics, Watertown, MA) and analyzed with GENEPIX software (Axon Instruments, Foster City, CA).

Results

High-Density Transposon Mutagenesis of Mycobacteria. We have used a *Himar1*-based mariner transposon to generate diverse libraries of transposon mutants in mycobacteria. To maximize the efficiency of mutagenesis, we used a mutant gene encoding a highly active *Himar1* transposase (16) and transduction using a temperature-sensitive mycobacteriophage (17). As shown in Fig. 1, this transposon construct can be used to produce large numbers of mutants in a variety of mycobacterial species. Surprisingly, even species that are not lysed by the transducing phage when grown at the permissive temperature were transduced with the transposon. For example, this phage does not form plaques on *Mycobacterium fortuitum* lawns at 30°C. However, the transducing phage intro-

MICROBIOLOGY

A

Species	Approximate Number of kanamycin-resistant colonies/ml transduced culture
M. smegmatis	10^3
M. fortuitum	10^3
M. bovis BCG	10^4
M. tuberculosis	10^4

B

Antibiotic	Resistant colonies/10^4 plated: Library	Wild Type	Fold increase
Isoniazid (0.5 mg/ml)	12.5	1	12.5
Ethionamide (10 mg/ml)	5	10^{-3}	5000
Streptomycin (1 mg/ml)	0	0	0

C

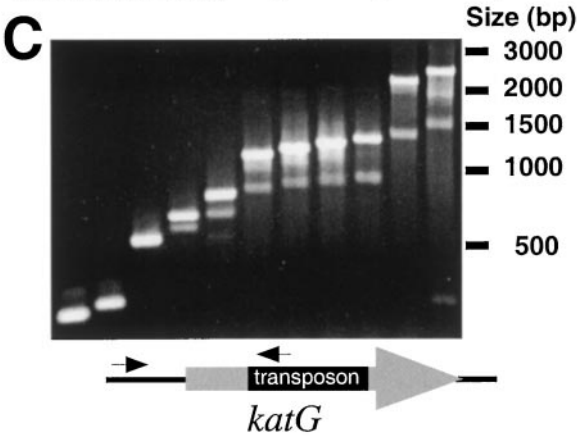

Fig. 1. Efficient transduction of a mariner transposon to generate diverse mutant libraries. (*A*) Cultures of the indicated mycobacterial species were infected with the transducing phage and the resulting kanamycin-resistant colonies were counted. (*B*) *M. bovis* BCG transposon mutants (10^5) or wild-type BCG (10^6 cfu) were plated on the indicated antibiotics and resistant colonies were counted. ''Fold increase'' represents the frequency of antibiotic-resistant strains in the library divided by the frequency of spontaneous resistance. (*C*) Insertions throughout the *katG* gene are found in INH-resistant mutants. PCR was performed on 11 isoniazid-resistant clones by using one primer that hybridizes to a region flanking the *katG* gene and one that hybridizes to the inverted repeat region of the transposon. PCR products between 168 and 2,390 bp represent insertions in the *katG* orf.

duced transposons into this species with the same efficiency obtained in *Mycobacterium smegmatis*, a species in which the phage is lytic. Mutagenesis of the slowly growing pathogenic species, *M. tuberculosis* and *M. bovis* BCG, was more efficient than rapidly growing strains.

By using the transducing phage, we have produced libraries of greater than 10^5 transposon mutants in both *M. bovis* BCG and *M. tuberculosis.* Thus far, the insertion sites from more than 40 mutants have been identified by DNA sequencing, and no evidence of multiple insertions in a single strain has been found. *Himar1* transposons insert only into the dinucleotide TA, which is present >70,000 times in the genome of *M. tuberculosis*. A substantial number of these sites occur in essential regions that cannot sustain mutations, so the number of insertions that can be recovered is somewhat smaller. Therefore, in a library of 10^5 mutants, an insertion will be present at nearly every TA. Assuming these sites are randomly distributed throughout the genome, an insertion will be present approximately every 60 bases. Therefore, a library of this size will contain an average of ≈17 insertions in a gene of 1 kb length.

Transposons Insert into a Large Variety of Sites. To be a useful tool, transposon mutagenesis should produce a wide variety of different mutant phenotypes. To verify the complexity of our mutant libraries, we sought to select strains from our library that were resistant to antimycobacterial antibiotics.

Several drugs used to treat mycobacterial infections require activation by bacterial enzymes. In the absence of such an activity these agents are not converted to their active forms and exert no toxicity. Both isoniazid (INH) and ethionamide require such activation (19, 20), and mutations that alter the function of the activating enzymes are responsible for antibiotic resistance in many clinical *M. tuberculosis* isolates (20, 21). Because transposons often disrupt genes, we expected to find strains resistant to these antibiotics among the transposon mutants. Resistance to other antibiotics, however, does not result from gene inactivation. For example, *M. tuberculosis* strains that are resistant to streptomycin almost uniformly have point mutations in the *rpsL* gene, which alter affinity for the antibiotic (22). Because *rpsL* is essential for survival, we did not expect to find streptomycin resistance in the transposon mutant pool.

We selected an *M. bovis* BCG mutant library containing 10^5 independent mutants on streptomycin, INH, and ethionamide. As shown in Fig. 1, we did not find any streptomycin-resistant strains. However, resistance to INH and ethionamide was present at much higher than spontaneous rates. By using PCR, we mapped the sites of transposon insertions in the INH-resistant mutants. All twenty INH-resistant strains had insertions in the *katG* gene that encodes the enzyme required for INH activation. As seen in Fig. 1, independent insertions were found throughout the INH gene. Similarly, 6 of 12 ethionamide-resistant mutants contained insertions in the gene encoding the known activator, *etaA.* In addition, multiple insertions in the *ndh* gene were found (data not shown). *ndh* mutants also are known to be resistant to ethionamide (23). The rates of INH and ethionamide resistance were consistent with the rates expected if transposition had occurred randomly throughout the chromosome.

Development of Transposon-Site Hybridization. TraSH uses a DNA microarray to map simultaneously all transposon insertions in a pool of mutants. A microarray was constructed that contains a fragment of DNA derived from each predicted ORF in the genome. To identify the location of insertions in a mutant pool, we produced labeled RNA that hybridized to the chromosomal sequences immediately adjacent to each transposon. As shown in Fig. 2*B*, genomic DNA was isolated from a pool of mutants and fragmented by using restriction enzymes with sites that appear frequently in the *M. tuberculosis* genome. Adapters were ligated to the resulting DNA fragments, and PCR was performed by using adapter- and transposon-specific primers to amplify the regions adjacent to the transposon-insertion sites. PCR products were size-selected by agarose gel electrophoresis to produce fragments ≈250–500 bp in size (containing ≈100–350 bp of chromosomal sequence) to minimize overlap into adjacent ORFs. *In vitro* transcription then was performed by using T7 RNA polymerase that recognizes outward-facing T7 promoters located in the transposon. This RNA was labeled and hybridized to the DNA microarray. RNA derived from mutant pools, which have been grown under different selective conditions, can be labeled with fluorophores that emit at different wavelengths. By mixing these labeled RNA's before hybridization, differential gene requirements are identified (see Fig. 2*A*).

TraSH only identifies mutants that lie close to the region represented by the ORF fragment immobilized on the DNA microarray. By convention, these fragments are referred to as probes (24). An average probe is 330 bp in length, whereas an average TraSH target (the RNA derived from *in vitro* transcription) hybridizes to ≈225 bp of chromosomal sequence. Assuming that 50 bp of overlap between the DNA probe and the RNA target are required for hybridization, an average probe can be identified by targets generated from ≈500 bp of chromosome. In a random

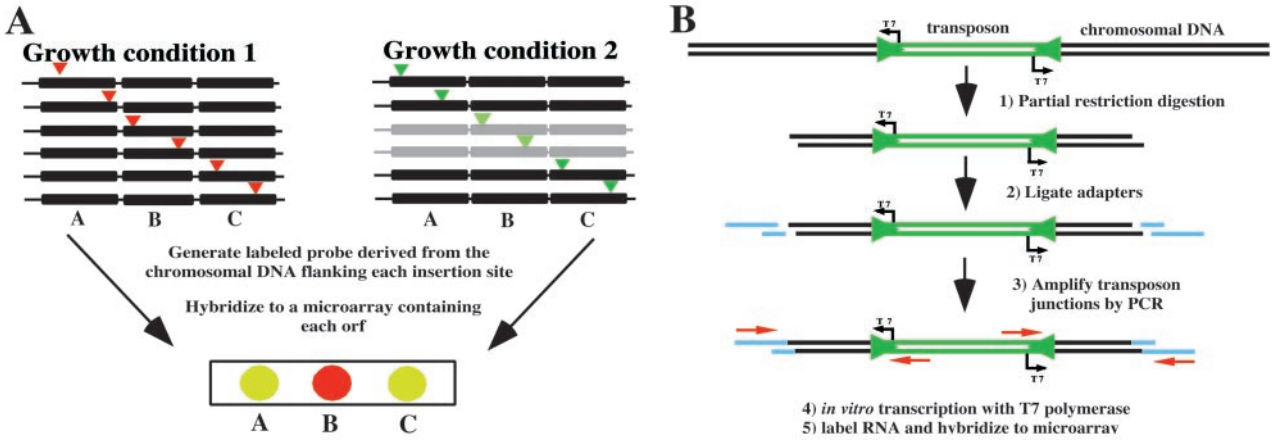

Fig. 2. Schematic representation of TraSH procedure. (*A*) Chromosomal region encompassing genes A–C from six different mutant strains (rectangles) is shown. Each mutant carries a single transposon insertion (triangles) that disrupts the function of a gene. Pools of mutants are grown under two different selective conditions. Genes A and C are nonessential for growth. Gene B is essential only under growth condition 2, and mutants harboring insertions in this gene are lost from this pool (represented by light shading). TraSH target that is complementary to the chromosomal DNA flanking each transposon insertion is generated from the two pools, labeled with different fluorophores, and hybridized to a microarray. The DNA probes representing genes A and C on the microarray will hybridize to the target generated from both pools. However, the target representing gene B will only be present in the pool from growth condition 1. By measuring the ratio of the two fluorophores for each probe, differential gene requirements are detected. (*B*) Method for generating TraSH target. The chromosomal DNA containing a transposon insertion (green) is digested with frequent cutting restriction enzymes, and adapters (blue) are ligated to the ends of the resulting DNA. PCR is then performed with primers (red) that hybridize to transposon and adapter sequences. To avoid the amplification of fragments that do not contain transposon sequence, the adapter contains the adapter–primer sequence but not the complementary strand, which serves as a binding site for this primer. Thus, the binding site must be generated by extension from the transposon-specific primer. To allow for the linear extension of these products, the transposon primer was designed with a higher melting temperature than the adapter primer, and higher annealing temperatures were used in the initial cycles of PCR (see *Materials and Methods*). Extension of the adapter itself, which would generate a binding site for the adapter primer, is prevented by a 3′-amino modification. The resulting PCR product is used as a template for the *in vitro* transcription of RNA target. This RNA is labeled and used for microarray hybridization.

transposon library of 10^5 mutants, an average probe would identify ≈10 different mutants.

TraSH Identifies Mutants That Fail to Grow on Minimal Medium. We compared the growth of libraries containing ≈10^5 *M. bovis* BCG mutants on two different defined media. After mutagenesis, bacteria were grown on rich medium (Middlebrook 7H11 agar, which contains all amino acids with added supplements, as described in *Materials and Methods*) or minimal medium (Middlebrook 7H10 agar). Genomic DNA was prepared from the surviving colonies under both growth conditions. TraSH targets were synthesized from each pool and labeled with different dyes. The targets then were mixed and hybridized to the DNA microarray.

For each DNA probe on the microarray, we determined the ratio of intensity produced by each dye. Probes that produced only a low-intensity signal (less than two times the background intensity) from the dye representing the population grown on rich medium were omitted from the analysis (≈30% of genes). Genes that produced ratios of rich/minimal medium in the top 5% in two independent experiments are reported in Table 1. Fig. 3 depicts the data obtained from the chromosomal regions surrounding several representative auxotrophic mutations.

We found 21 genes that met the above criteria. Of these 21 genes, 12 have annotated functions based on homology, and 10 are known to be involved in biosynthesis. The other two genes, *gmhA* and *fadB,* are similar to genes involved in heptose and fatty-acid metabolism, respectively. It is unclear why these two genes should be required for growth on minimal medium. The remaining identified genes do not have clear homology to genes of known function.

The 10 genes that are predicted to be involved in biosynthesis can be grouped into four different metabolic pathways. First, *leuB*, *leuD*, and *ilvA* are all involved in leucine synthesis (25). Second, mutations in *subI*, a sulfate transporter, result in methionine auxotrophy (12). Similarly, *cysH* and *cysD* are involved in the reduction of sulfate, which is necessary for the synthesis of both cysteine and methionine (26). Third, *glnE*, *gltB*, and *amiB2* are involved in the metabolism of glutamine, glutamate, and nitrogen. Finally, *thiC* is necessary for thiamin synthesis (25). The identification of multiple genes in a pathway provides significant validation for this method of functional analysis. In addition, our results suggest that only a limited number of pathways can be exploited for the generation of auxotrophic mutants in pathogenic mycobacteria (see *Discussion*).

Discussion

Transposon mutagenesis with phage-mediated delivery of a *Himar1*-based transposon is very efficient, and we have used it to construct large and diverse mutant libraries. The mutagenesis seems to occur randomly, as several independent mutants in selected genes can be obtained at frequencies expected for random integration. The ability to create mutant libraries in a variety of mycobacterial species should have many applications.

TraSH has proven to be a useful method for identifying genes required for survival under different growth conditions. By using this technique, we have defined several *M. bovis* BCG mutants that grow well on rich but not on minimal media. Among these are both of the *M. bovis* BCG auxotrophic mutants (*leuD* and *subI*) previously found by transposon mutagenesis (26). We found other mutations in known biosynthetic genes and a number of mutants with insertions in genes not known to be important for survival on minimal medium. This last class contains mostly genes that are not homologous to any gene of known function. In the future, the functions of these genes could be more precisely determined by repeating this experiment using several pools of mutants grown on minimal media that are individually supplemented with each nutrient contained in rich medium.

The limited number of auxotrophic mutants identified by TraSH is consistent with several previous observations. For example, lysine auxotrophs survive only on medium supplemented with concentrations of lysine much higher than that found in 7H11 agar (27), which was used in this study. Similarly, *hisD*, *proD*, and *trpD* mutants in *M. tuberculosis* could not be isolated on 7H11 medium (28). These observations are probably a result of the unusual biology of

Table 1. Auxotrophic mutations identified by TraSH. Intensities are the average of duplicate features from two independent microarray TraSH hybridizations.

Gene name	Intensity		Ratio (Rich/Minimal)	Gene function or metabolic pathway
	Rich medium	Minimal medium		
leuB	1299	200	6.47	Valine, leucine, isoleucine biosynthesis
leuD	1251	317	3.94	Valine, leucine, isoleucine biosynthesis
ilvA	1313	161	8.12	Valine, leucine, isoleucine biosynthesis
cysD	2842	763	3.73	Sulfur metabolism/cysteine and methionine biosynthesis
cysH	472	25	18.7	Sulfur metabolism/cysteine and methionine biosynthesis
subI	2690	463	5.8	Sulfate transport/cysteine and methionine biosynthesis
gltB	10840	217	49.8	Glutamine, glutamate, and nitrogen metabolism
glnE	4193	720	5.82	Glutamine, glutamate, and nitrogen metabolism
amiB2	3076	624	4.93	Multiple pathways including glutamate metabolism
thiC	6770	43	156	Thiamine biosynthesis
gmhA	3558	366	9.7	Phosphoheptose isomerase
fadB	4295	739	5.81	Lipid metabolism
Rv0235c	7832	697	11.2	Unknown
Rv0875c	2649	830	3.19	Unknown
Rv1682	3108	1068	2.91	Unknown
Rv1718	1875	207	9.04	Unknown
Rv2604c	981	147	6.64	Unknown
Rv2757c	1823	309	5.9	Unknown
Rv3129	2152	284	7.57	Unknown
Rv3168	4866	844	5.77	Unknown
Rv3727	26577	6577	4.04	Unknown

slow-growing mycobacteria. The thick, hydrophobic mycobacterial cell wall presents a considerable barrier to the diffusion of many compounds (29). Mycobacteria must, therefore, use transporters to obtain most nutrients. Whereas *M. smegmatis*, an environmental organism, seems capable of obtaining most small molecules, *M. bovis* and *M. tuberculosis*, both of which lack an environmental niche, might be deficient in these transport pathways. In fact, previous work has suggested that *M. tuberculosis* cannot transport lysine efficiently (27). Our data also support this conclusion, as the majority of amino acid biosynthetic genes that were not identified by TraSH are associated with low or absent signal intensity for both pools. This finding suggests that these mutants are unable to grow

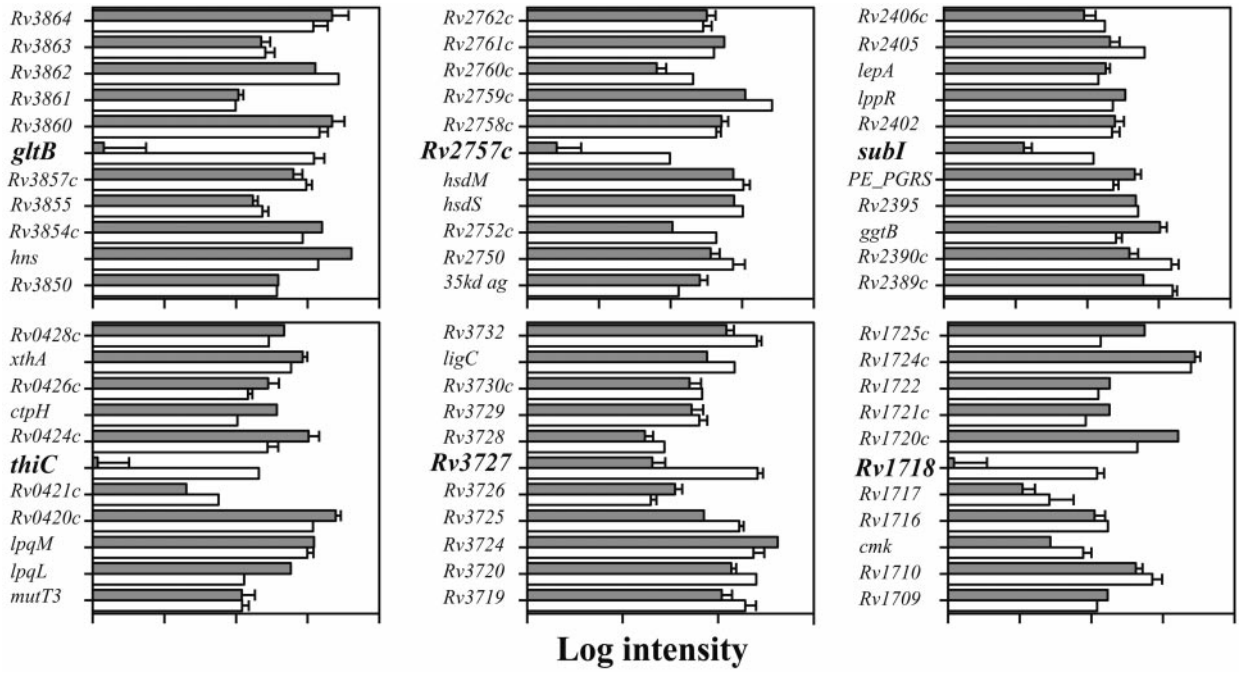

Fig. 3. TraSH identifies mutations that result in auxotrophy. The bar graphs display the fluorescence intensity observed for each gene across the genomic region surrounding auxotrophic mutations (in bold). Gray bars and white bars represent fluorescence observed from pools grown on minimal or rich medium, respectively. Data are the average of duplicate measurements from a single microarray and are represented on a log scale. Error bars representing standard deviations are shown for all features. The absence of error bars indicates that the value was too small to be plotted. Genes for which the intensity from the rich medium pool did not exceed two-times the background were omitted.

on either rich or minimal medium. Similarly, we find little signal associated with almost all genes whose function is known to be essential for growth, such as those encoding ribosomal proteins or proteins involved in cell-wall synthesis (data not shown). The implication of these observations is that pathogenic mycobacteria may not scavenge many nutrients from the host, or that many nutrients are available at high concentrations.

TraSH might fail to identify important mutants for several additional reasons. Some genes lack the dinucleotide TA, which is required for transposition, in proximity to the region of the ORF represented on the microarray. Insertions in such genes would fail to produce detectable signal under any growth condition. Because the genome of *M. tuberculosis* contains >70,000 TA dinucleotides, this limitation applies to only a small number of genes. Likewise, genes could be missed if permissive sites are not saturated with insertions. Finally, some mutants may be able to obtain sufficient nutrients from adjacent protrophic colonies when plated on minimal medium. As a result of such crossfeeding, these mutants might not appear attenuated, although they would not survive on minimal medium if grown in isolation.

Unlike other investigators (11), we did not find mutants in purine biosynthesis. This result may be due to the different growth medium and supplementation used in this study. We did, however, detect mutations in *glnE*, despite convincing evidence that it is essential in *M. tuberculosis* (30). This finding may be due to species differences. More likely, however, this observation is due to differences in technique. TraSH may be able to detect mutants with severe growth defects, even those that cannot form a colony, such as *glnE* mutants. Alternatively, the *glnE* mutants that displayed differential growth in our study could result from insertions in regions of the gene that decrease, but do not eliminate, the activity of the gene. These mutants with attenuated *glnE* activity may have a preferential growth defect on minimal medium.

TraSH produces data that is distinctly different from expression profiling. Whereas expression analysis reveals conditions under which genes are transcriptionally regulated, TraSH defines the conditions under which genes are required for survival. In practice, these methods should produce complementary information. Expression analysis is valuable for identifying pathways and for studying genes that are essential for survival under all conditions. TraSH, on the other hand, can be used to identify unambiguously genes that are conditionally essential for growth under one growth condition but not another.

TraSH has several advantages over other methods used to study transposon mutants. Large-scale sequencing has been used to identify genes that could not sustain insertions and allow growth *in vitro* (31), but this method is probably not a practical way to compare survival under different conditions. Insertion mapping methods, such as GAMBIT (5), can be used to compare pools of mutants but are difficult to perform on a genomic scale. STM is a powerful method that has been extensively used to study virulence mechanisms. STM offers the advantage of ready access to interesting mutants. Unlike TraSH, however, STM can be used to study only relatively small pools that must be assembled in advance. Because of this limitation, it is difficult to study the genome comprehensively. In addition, STM relies on the analysis of single mutants. Such mutants may behave differently from other strains with mutations in the same gene caused by unlinked mutations associated with the tested phenotype.

The transposition of *Himar1* mariner elements occurs in the absence of host cofactors and requires only the presence of transposase enzyme (32). As a result, these elements are active in a wide variety of hosts including Gram-positive and Gram-negative bacteria, archaebacteria, and eukaryotes (33–35). Thus, although we have demonstrated the ability of TraSH for defining gene function in mycobacteria, we expect that this technique will be easily adapted for use in many other microorganisms.

We thank Bill Jacobs and Stoyan Bardarov for ϕAE87, Michelle Dziejman for sharing her microarray expertise, Marc Lipsitch for statistical advice, Simon Simms for facilitating the production of the microarray, Steve Lory and John Mekalanos for their advice and support, and members of the Rubin and Bloom labs. This work was supported in part by Fogarty Center Grant TW0004, National Institute of Allergy and Infectious Diseases Grant AI48704, and National Institute of General Medical Sciences Grants GM54160 and GM38922. C.M.S. was supported by the Cancer Research Fund of the Damon Runyon–Walter Winchell Foundation Fellowship, DRG-1647.

1. Brown, P. O. & Botstein, D. (1999) *Nat. Genet.* **21,** 33–37.
2. Hamer, L., DeZwaan, T. M., Montenegro-Chamorro, M. V., Frank, S. A. & Hamer, J. E. (2001) *Curr. Opin. Chem. Biol.* **5,** 67–73.
3. Mei, J.-M., Nourbakhsh, F., Ford, C. W. & Holden, D. W. (1997) *Mol. Microbiol.* **26,** 399–407.
4. Chiang, S. L., Mekalanos, J. J. & Holden, D. W. (1999) *Annu. Rev. Microbiol.* **53,** 129–154.
5. Akerley, B. J., Rubin, E. J., Camilli, A., Lampe, D. J., Robertson, H. M. & Mekalanos, J. J. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 8927–8932.
6. Smith, V., Botstein, D. & Brown, P. O. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 6479–6483.
7. Reich, K. A., Chovan, L. & Hessler, P. (1999) *J. Bacteriol.* **181,** 4961–4968.
8. Dye, C., Scheele, S., Dolin, P., Pathania, V. & Raviglione, M. C. (1999) *J. Am. Med. Assoc.* **282,** 677–686.
9. Bloom, B. R. & Fine, P. E. M. (1994) in *Tuberculosis*, ed. Bloom, B. R. (Am. Soc. Microbiol., Washington, DC), pp. 531–558.
10. Hondalus, M. K., Bardarov, S., Russell, R., Chan, J., Jacobs, W. R., Jr., & Bloom, B. R. (2000) *Infect. Immun.* **68,** 2888–2898.
11. Jackson, M., Phalen, S. W., Lagranderie, M., Ensergueix, D., Chavarot, P., Marchal, G., McMurray, D. N., Gicquel, B. & Guilhot, C. (1999) *Infect. Immun.* **67,** 2867–2873.
12. McAdam, R. A., Weisbrod, T. R., Martin, J., Scuderi, J. D., Brown, A. M., Cirillo, J. D., Bloom, B. R. & Jacobs, W. R., Jr. (1995) *Infect. Immun.* **63,** 1004–1012.
13. Snapper, S. B., Melton, R. E., Mustafa, S., Kieser, T. & Jacobs, W. R. J. (1990) *Mol. Microbiol.* **4,** 1911–1919.
14. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
15. Rubin, E. J., Akerley, B. J., Novik, V. N., Lampe, D. J., Husson, R. N. & Mekalanos, J. J. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 1645–1650.
16. Lampe, D. J., Akerley, B. J., Rubin, E. J., Mekalanos, J. J. & Robertson, H. M. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 11428–11433.
17. Bardarov, S., Kriakov, J., Carriere, C., Yu, S., Vaamonde, C., McAdam, R. A., Bloom, B. R., Hatfull, G. F. & Jacobs, W. R., Jr. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 10961–10966.
18. Belisle, J. T. & Sonnenburg, M. G. (1998) in *Methods in Molecular Biology: Mycobacteria Protocols*, eds. Parish, T. & Stoker, N. G. (Humana, Totowa, NJ), Vol. 101, pp. 31–44.
19. Zhang, Y., Heym, B., Allen, B., Young, D. & Cole, S. (1992) *Nature (London)* **358,** 591–593.
20. DeBarber, A. E., Mdluli, K., Bosman, M., Bekker, L.-G. & Barry, C. E., 3rd (2000) *Proc. Natl. Acad. Sci. USA* **97,** 9677–9682.
21. Heym, B., Alzari, P. M., Honore, N. & Cole, S. T. (1995) *Mol. Microbiol.* **15,** 235–245.
22. Nair, J., Rouse, D. A., Bai, G. H. & Morris, S. L. (1993) *Mol. Microbiol.* **10,** 521–527.
23. Miesel, L., Weisbrod, T. R., Marcinkeviciene, J. A., Bittman, R. & Jacobs, W. R., Jr. (1998) *J. Bacteriol.* **180,** 2459–2467.
24. Phimister, B. (1999) *Nat. Genet.* **21,** 1.
25. Kanehisa, M. & Goto, S. (2000) *Nucleic Acids Res.* **28,** 27–30.
26. Kredich, N. M. (1996) in *Escherichia coli and Salmonella: Cellular and Molecular Biololgy*, ed. Neidhardt, F. C. (Am. Soc. Microbiol., Washington, DC), Vol. 1.
27. Pavelka, M. S., Jr., & Jacobs, W. R., Jr. (1999) *J. Bacteriol.* **181,** 4780–4789.
28. Parish, T., Gordhan, B. G., McAdam, R. A., Duncan, K., Mizrahi, V. & Stoker, N. G. (1999) *Microbiology* **145,** 3497–3503.
29. Brennan, P. J. & Nikaido, H. (1995) *Annu. Rev. Biochem.* **64,** 29–63.
30. Parish, T. & Stoker, N. G. (2000) *J. Bacteriol.* **182,** 5715–5720.
31. Hutchison, C. A., Peterson, S. N., Gill, S. R., Cline, R. T., White, O., Fraser, C. M., Smith, H. O. & Venter, J. C. (1999) *Science* **286,** 2165–2169.
32. Lampe, D. J., Churchill, M. E. & Robertson, H. M. (1996) *EMBO J.* **15,** 5470–5479.
33. Pelicic, V., Morelle, S., Lampe, D. & Nassif, X. (2000) *J. Bacteriol.* **182,** 5391–5398.
34. Zhang, J. K., Pritchett, M. A., Lampe, D. J., Robertson, H. M. & Metcalf, W. W. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 9665–9670. (First Published August 1, 2000; 10.1073/pnas.160272597)
35. Zhang, L., Sankar, U., Lampe, D. J., Robertson, H. M. & Graham, F. L. (1998) *Nucleic Acids Res.* **26,** 3687–3693.