Impact of Bacterial Genetics on the Transmission of Isoniazid-Resistant *Mycobacterium tuberculosis*

Sebastien Gagneux[1,2*], Marcos V. Burgos[3,4], Kathryn DeRiemer[1,5], Antonio Enciso[1,6], Samira Muñoz[1], Phillip C. Hopewell[7], Peter M. Small[2,8], Alexander S. Pym[1,9*]

1 Division of Infectious Diseases and Geographic Medicine, Stanford University Medical Center, Stanford, California, United States of America, **2** Institute for Systems Biology, Seattle, Washington, United States of America, **3** Department of Internal Medicine, Division of Infectious Diseases, University of New Mexico, Albuquerque, New Mexico, United States of America, **4** Veterans Affairs Medical Center, Albuquerque, New Mexico, United States of America, **5** School of Medicine, University of California Davis, Davis, California, United States of America, **6** Unidad de Investigacion Medica de Zacatecas, Zacatecas, Mexico, **7** Division of Pulmonary and Critical Care Medicine, University of California San Francisco, San Francisco, California, United States of America, **8** Bill and Melinda Gates Foundation, Seattle, Washington, United States of America, **9** Unit for Clinical and Biomedical TB Research, South African Medical Research Council, Durban, South Africa

Understanding the ecology of drug-resistant pathogens is essential for devising rational programs to preserve the effective lifespan of antimicrobial agents and to abrogate epidemics of drug-resistant organisms. Mathematical models predict that strain fitness is an important determinant of multidrug-resistant *Mycobacterium tuberculosis* transmission, but the effects of strain diversity have been largely overlooked. Here we compared the impact of resistance mutations on the transmission of isoniazid-resistant *M. tuberculosis* in San Francisco during a 9-y period. Strains with a KatG S315T or *inhA* promoter mutation were more likely to spread than strains with other mutations. The impact of these mutations on the transmission of isoniazid-resistant strains was comparable to the effect of other clinical determinants of transmission. Associations were apparent between specific drug resistance mutations and the main *M. tuberculosis* lineages. Our results show that in addition to host and environmental factors, strain genetic diversity can influence the transmission dynamics of drug-resistant bacteria.

Citation: Gagneux S, Burgos MV, DeRiemer K, Enciso A, Muñoz S, et al. (2006) Impact of bacterial genetics on the transmission of isoniazid-resistant *Mycobacterium tuberculosis*. PLoS Pathog 2(6): e61. DOI: 10.1371/journal.ppat.0020061

Introduction

Drug resistance in *Mycobacterium tuberculosis* has been encountered worldwide [1]. It has been postulated and mathematical models suggest that drug resistance will seriously impair global tuberculosis control efforts [2,3]. In at least some areas, resistant strains are replacing susceptible ones, leading to "hotspots" of drug resistance [4]. In other areas the introduction of standard control programs has interrupted transmission of drug-resistant strains [5]. Thus, it is important to increase our understanding of the biology of drug resistance to predict and prevent the future spread of resistant strains.

The factors that determine the epidemiology of drug-resistant organisms that can be influenced by human behavior, such as patterns of antimicrobial usage, have been studied extensively. There has been less analysis of the role of intrinsic microbiologic factors, such as the relative fitness of drug-resistant strains. Ecological theory and mathematical models suggest that the relative fitness of drug-resistant strains can profoundly affect the emergence and propagation of these organisms [2,3,6,7].

In vitro models using non-pathogenic organisms have shown that drug resistance–conferring mutations are usually associated with a fitness cost [8]. The degree to which the mutation affects the fitness of the organism varies with the specific drug resistance mutation, the environment, and genetic background of the strain [8–12]. Compensatory evolution can ameliorate the fitness effects of drug resistance mutations [13–16]. However, it is unclear if such experimental models can be extrapolated to explain the behavior of pathogenic drug-resistant bacteria in human populations.

To date, epidemiological studies and surveillance of *M. tuberculosis* have not fully resolved whether the transmission dynamics of drug-resistant strains are different from drug-sensitive strains. Most molecular epidemiological work has focused on multidrug-resistant strains (resistant to at least isoniazid and rifampin) of *M. tuberculosis*. In such studies, multidrug-resistant strains have been associated with both more and less ongoing transmission than sensitive strains [17–20]. These discrepancies suggest that drug-resistant strains may be heterogeneous and that drug resistance–conferring mutations might have a variable impact on strain fitness. In the case of isoniazid resistance, there is some laboratory and epidemiological evidence in support of this [21–24]. Alternatively, the genetic background of a strain might influence how it adapts to antibiotics.

In this study, we have used comprehensive molecular

Editor: Lalita Ramakrishnan, University of Washington, United States of America

Received February 16, 2006; **Accepted** May 5, 2006; **Published** June 16, 2006

DOI: 10.1371/journal.ppat.0020061

Abbreviation: RFLP, restriction fragment length polymorphism

* To whom correspondence should be addressed. E-mail: sgagneux@systemsbiology.org (SG), alexander.pym@mrc.ac.za (ASP)

Synopsis

Resistance to antibiotics is an increasing global public health problem. Multidrug-resistant tuberculosis is threatening the control of this deadly infectious disease in many parts of the world. Studies to identify the factors that lead to the emergence and subsequent spread of drug-resistant organisms have clearly shown that patient adherence to a 6 mo course of treatment is very important to prevent the development of drug-resistant tuberculosis. However, little is known about how bacterial genetics might influence the spread of drug-resistant strains. This study showed that among the strains in San Francisco with resistance to isoniazid, one of the main anti-tuberculosis drugs, only certain strains were transmitted effectively. The transmission of specific isoniazid-resistant strains depended on the type of genetic mutation leading to resistance. Also, isoniazid-resistant strains coming from different parts of the world were associated with different resistance mutations. This study's results show that the spread of drug resistance is due to the strain of tuberculosis as well as the environment and characteristics of the patient, and may partially explain why some areas of the world have worse drug resistance problems than others.

epidemiologic data from San Francisco to examine the role of specific resistance-conferring mutations and strain genetic background on the transmission dynamics of isoniazid-resistant *M. tuberculosis* in a human population.

Results

Bacterial Strains and Resistance Profiles

This study was based on an ongoing population-based molecular epidemiological study of tuberculosis transmission in San Francisco [25]. For the study period, 1991–1999, 2,104 cases of tuberculosis were culture positive for *M. tuberculosis*. Full drug susceptibility testing was performed on 2,081 (98.9%) isolates, which were all recovered from distinct tuberculosis patients. Of these isolates, 298 (14.3%) were initially resistant to at least one first-line anti-tuberculosis drug (Table 1). Resistance to isoniazid was the most common form of drug resistance. Genetic characterization was available for 152 (85.9%) of the 177 isoniazid-resistant isolates. Among these 152 isolates with resistance to isoniazid, with or without resistance to other drugs, 136 had unique restriction fragment length polymorphism (RFLP) patterns. 16 isolates (10.5%) shared an RFLP pattern, drug resistance phenotype, drug resistance–conferring mutation, and lineage specific marker (see below) with at least one other isolate, indicating they were linked in seven genetically distinct isolate clusters. Therefore, there were a total of 143 genetically distinct isoniazid-resistant strains (clones), providing a large, genetically well- characterized and representative group of strains from a defined human population for studying the biology of resistance.

Drug-Resistant Alleles

Resistance to isoniazed has been associated with various chromosomal mutations, but those in *katG,* and in the promoter regions of *inhA,* occur most frequently in clinical isolates of *M. tuberculosis* [26,27]. *katG* encodes a catalase-peroxidase that activates the prodrug isoniazid to its active form and also acts as a virulence factor, protecting against oxidative stress [28]. InhA, encoded by *inhA,* is a component of FAS-II (fatty acid elongation system) that is required for mycolic acid synthesis and is a target for isoniazid [29–31]. *inhA* promoter mutations result in isoniazid resistance through up-regulation of *inhA*. Mutations in the *ahpC* promoter have also been described in isoniazid-resistant isolates, but are thought to be compensatory, helping strains adjust to the additional oxidative stress imposed by mutations in *katG* [32–34].

Sequencing of these three principal isoniazid resistance loci of the 143 isoniazid-resistant strains revealed that 128 (89.5%) had at least one mutation in either the *katG* gene or in the *inhA* promoter region (Table 2). Only 16 (11.2%) of these strains had no resistance mutations in the regions sequenced. After excluding the KatG L463R substitution, a polymorphism not associated with isoniazid resistance [35], 92 strains (64.3%) were found to have *katG* gene mutations. 58 (40.6%) strains had the S315T mutation and a further 34 (23.8%) had other mutations affecting *katG,* including three strains with complete gene deletions. *inhA* promoter mutations were detected in 46 (32.2%) strains, of which 11 (7.7%) also had a *katG* mutation. Characterization of the *ahpC* promoter regions showed that mutations in this region were uncommon, occurring in only 11 (7.7%) strains (Table 2). Pertinently, none of these arose in strains without other isoniazid resistance–conferring mutations, and ten of the 11 strains with mutations in the *ahpC* promoter region had mutations in *katG,* including two of the strains with complete gene deletions. These findings support the notion that *ahpC* promoter mutations are in fact compensatory, rather than causing resistance directly.

Transmission Dynamics of Drug Resistance–Conferring Mutations

Genetic clustering has been used extensively in molecular epidemiological studies as a measure for ongoing tuberculosis transmission [36,37]. The underlying assumption is that in a strict population-based study such as the one reported here,

Table 1. Primary Drug-Resistant Clinical Isolates of *M. tuberculosis* Isolated in San Francisco from 1991–1999

Resistance Profile	Number of Isolates	Genetically Characterized Isolates
	n (Percentage of All)	*n* (Percentage of All with This Profile)
EMB	3 (1.0)	—
PZA	7 (2.3)	—
RIF	27 (9.7)	—
STR	84[a] (28.2)	—
INH	106[b] (35.6)	90 (84.9)
INH + RIF	13[c] (4.7)	12 (92.3)
INH + STR	47[d] (15.8)	39 (83.0)
INH + RIF + STR	11[e] (4.0)	11 (100)
Total	298 (100)	152 (85.9)

[a]Includes one isolate resistant to EMB.
[b]Includes four isolates that were also resistant to EMB, one isolate resistant to PZA, and one isolate resistant to both EMB and PZA.
[c]Includes three isolates resistant to both EMB and PZA and one isolate resistant to PZA only.
[d]Includes four isolates resistant to EMB and two isolates resistant to PZA.
[e]Includes three isolates resistant to EMB and two isolates resistant to both EMB and PZA.
INH, isoniazid; RIF, rifampicin; STR, streptomycin; EMB, ethambutol; PZA, pyrazinamide.
DOI: 10.1371/journal.ppat.0020061.t001

Table 2. Isoniazid Resistance–Conferring Mutations and RFLP Clusters Identified in 152 Isoniazid-Resistant Isolates of *M. tuberculosis* from San Francisco, 1991–1999

Group	Mutations	RFLP Clusters	Clustered Isolates	Unique Isolates	Total Isolates	Total Clones
		n (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)
		7 (100)	**16 (100)**	**136 (100)**	**152 (100)**	**143 (100)**
Group (1) KatG mutations other than S315T[a]	KatG S315N	0	0	2 (1.5)	2 (1.3)	2 (1.4)
	KatG V1L, R104Q, A106V, W107R, A109V, N138D, K143T, Y229F, A264T, K301aag_atag, S315I, Y337C, K345T, T394A, D573N, W728C, D735N, or Deletion	0	0	18 (13.2)	18 (11.8)	18 (12.6)
	KatG D94G + *ahpC* prom-34;g-a	0	0	1 (0.7)	1 (0.7)	1 (0.7)
	KatG L384R + *ahpC* prom-39;c-t	0	0	1 (0.7)	1 (0.7)	1 (0.7)
	KatG S315N + *ahpC* prom-10;c-a	0	0	1 (0.7)	1 (0.7)	1 (0.7)
	KatG A636E + *ahpC* prom-10;c-t	0	0	1 (0.7)	1 (0.7)	1 (0.7)
	KatG L653P + *ahpC* prom-6;g-a	0	0	1 (0.7)	1 (0.7)	1 (0.7)
	KatG DEL[c] + *ahpC* prom-6;g-a	0	0	1 (0.7)	1 (0.7)	1 (0.7)
	KatG DEL + *ahpC* prom-15;c-t	0	0	1 (0.7)	1 (0.7)	1 (0.7)
	KatG T85P, G121C, Y155C, A162T, T251M, P280H, or E607K + *inhA* prom-15;c-t	0	0	7 (5.1)	7 (4.6)	7 (4.9)
	Subtotal	0	0	34 (25.0)	34 (22.4)	34 (23.8)
Group (2) KatG S315T mutation	KatG S315T	3 (42.9)	7 (43.8)	48 (35.3)	55 (36.2)	51 (35.7)
	KatG S315T + *inhA* prom-15;c-t	0	0	3 (2.2)	3 (2.0)	3 (0.7)
	KatG S315T + *inhA* prom-17;g-t	0	0	1 (0.7)	1 (0.7)	1 (0.7)
	KatG S315T + *ahpC* prom 0;t-c	0	0	1 (0.7)	1 (0.7)	1 (0.7)
	KatG S315T + *ahpC* prom-6;g-a	0	0	1 (0.7)	1 (0.7)	1 (0.7)
	KatG S315T + *ahpC* prom-54;t-tatgt	0	0	1 (0.7)	1 (0.7)	1 (0.7)
	Subtotal	3 (42.9)	7 (43.8)	55 (40.4)	62 (40.8)	58 (40.6)
Group (3) *inhA* prom-15;c-t mutation	*inhA* prom-15;c-t	3 (42.9)	7 (43.8)	31 (22.8)	38 (25.0)	34 (23.8)
	inhA prom-15;c-t + *ahpC* prom-30 c-t	0	0	1 (0.7)	1 (0.7)	1 (0.7)
	Subtotal	3 (42.9)	7 (43.8)	32 (23.5)	39 (25.7)	35 (24.5)
Other mutations or no mutations	*inhA* prom-147 c-t	0	0	1 (0.7)	1 (0.7)	1 (0.7)
	No mutations	1 (14.3)	2 (12.5)	14 (10.3)	16 (10.5)	15 (10.5)
	Subtotal	1 (14.3)	2 (12.5)	15 (11.0)	17 (11.2)	16 (11.2)

[a]Excluding the KatG L463R polymorphism.
prom, promoter.
DOI: 10.1371/journal.ppat.0020061.t002

tuberculosis patients harboring isolates with unique RFLP patterns represent cases of reactivation of a latent infection, whereas patients with genetically clustered isolates are epidemiologically linked and represent chains of tuberculosis transmission. This is based on more than a decade of molecular epidemiological studies carried out in many communities from different countries. Outbreaks of tuberculosis in health-care facilities, prisons, and other contained situations have been shown to be caused by genetically identical strains [38,39]. In addition, epidemiological links have been demonstrated between genetically clustered strains identified at the community level [40]. For example, a study from Holland could establish an epidemiological link for 86% of genetically clustered strains [41].

To determine the relationship between drug resistance mutations and tuberculosis transmission dynamics, we compared the percentage of genetic clustering in isolates associated with the different drug resistance mutations (Table 2). Genetic clustering is usually based only on specific RFLP fingerprinting patterns. However, by including the phenotypic drug resistance profile, the specific drug resistance mutation, as well as the markers for the major *M. tuberculosis* lineage (see below), we ensured a more stringent definition of genetic clustering. The isoniazid-resistant isolates with resistance mutations could be divided into three groups (Table 2): those isolates with a mutation in KatG other than the S315T (Group 1), those isolates with a KatG S315T mutation (Group 2), and those isolates with a *inhA*-15;c-t promoter mutation only (Group 3). Whereas isolates with KatG mutations other than S315T were never genetically clustered (Table 2), 11.3% of the KatG S315T mutants (Fisher's 2-tailed exact test: $p < 0.05$) and 17.8% of the *inhA*-15;c-t promoter mutants (Fisher's 2-tailed exact test: $p < 0.002$) showed genetic clustering (Table 2). 11.7% of strains without a defined resistance mutation were genetically clustered. No strains with *ahpC* promoter mutations were genetically clustered.

These data demonstrate that only two isoniazid resistance mutations were effectively propagated in San Francisco during the study period. This is consistent with the role of catalase-peroxidase (KatG) in the pathogenicity of *M. tuberculosis*, which is a known virulence factor [23,42]. KatG S315T is an unusual substitution resulting in a functional enzyme that retains catalase-peroxidase activity but with a diminished

capacity to activate isoniazid [23,43,44]. Similarly, the *inhA*-15;c-t promoter mutation does not affect KatG activity. This is in contrast to the other substitutions in, or deletions of, *katG,* which have a more profound effect on enzyme activity [23,42], although all of the mutations identified in this study have not been empirically evaluated. Thus, KatG activity could explain the different transmission dynamics of the various mutations.

Because resistance to additional drugs could influence the fitness of isoniazid-resistant strains in an unpredictable manner, we repeated our analysis on the 90 isolates resistant to isoniazid alone (Table 1). We found that 7/28 (25.0%) of the isolates with the KatG S315T mutation and 4/23 (17.4%) with the *inhA*-15;c-t were genetically clustered. The comparisons to the group of isolates with other *katG* mutations remained statistically significant (Fisher's 2-tailed exact test: $p < 0.02$ and $p < 0.05$, respectively).

To assess the importance of specific mutations for the propagation of isoniazid-resistant strains, compared to other factors generally considered in conventional molecular epidemiologic studies, we performed univariate and multivariate analyses. The univariate analysis showed that isoniazid resistance due to mutations from Group 2 or 3, which are likely to maintain KatG activity, was the third strongest risk factor for genetic clustering (Table 3). In the final multivariate logistic regression model, these mutations perfectly predicted genetic clustering because all genetically clustered isolates had Group 2 or 3 mutations. Only homelessness remained an independent risk factor for genetic clustering (adjusted odds ratio = 9.6, 95% confidence interval: 1.7–53.3, $p = 0.01$). Taken together, these results indicate that specific drug resistance mutations have an impact on the successful propagation of isoniazid-resistant tuberculosis, as great or greater than other well-recognized risk factors, such as the presence of acid-fast bacilli in sputum smears.

Table 3. Results of the Univariate Analysis Showing Host and Bacterial Factors Associated with Clustering in 152 Isoniazid-Resistant Isolates of *M. tuberculosis* from San Francisco

Risk Factor	Number of Patients (%)	Number of Patients in Clusters (%)	Odds Ratio (95% CI)	*P*
Group (2) and (3) mutations	118 (77.6)	16 (13.6)	5.8 (0.84–248)[a]	0.023[b]
Male sex	97 (63.8)	12 (12.4)	1.8 (0.51–8.04)	0.416[b]
Age ≥ 43 y	78 (51.3)	4 (5.1)	0.3 (0.06–0.99)	0.034[b]
Birth in the United States	29 (19.1)	9 (31.0)	7.5 (2.2–26.1)	<0.001
Ethnicity[c]				
Asian	104 (68.9)	5 (4.8)	0.2 (0.04–1.1)	0.037[b]
Black	8 (5.3)	2 (25.0)	1.3 (0.1–12.4)	>0.99[b]
Hispanic	19 (12.6)	4 (21.1)	1.1 (0.17–6.9)	>0.99[b]
White	20 (13.2)	4 (20.0)	Reference	Reference
HIV-positive	19 (12.5)	4 (21.1)	2.7 (0.56–10.4)	0.119[b]
Homelessness	10/120[d] (8.3)	3 (30.0)	7.4 (0.97–44.0)	0.027[b]
Substance abuse	17 (11.2)	4 (23.5)	3.2 (0.64–12.4)	0.084[b]
Pulmonary tuberculosis	135 (88.8)	14 (10.4)	0.87 (0.17–8.6)	0.695[b]
Sputum smear AFB-positive	70 (46.1)	7 (10.0)	0.85 (0.25–2.7)	0.764
Cavitary disease	23 (15.1)	1 (4.4)	0.35 (0.01–2.5)	0.468[b]
Streptomycin resistance	50 (32.9)	3 (6.0)	0.44 (0.08–1.7)	0.267[b]
Rifampin resistance	23 (15.1)	0 (0.0)	0.28 (0.006–2.0)[a]	0.132[b]

Comparisons were performed using the χ^2 test of proportions unless otherwise noted.
[a]All of the isolates with Group 2 and 3 mutations were in genetic clusters. Because of the resulting zero value in one of the cells of the corresponding 2 by 2-table, the value 1 was therefore added to each cell in order to be able to calculate an odds ratio.
[b]Two-tailed Fisher's exact test.
[c]Excludes one American Indian. Odds ratio and *P* values are from the comparison with Whites.
[d]Denominator is smaller due to missing values.
AFB, acid-fast bacilli; CI, confidence interval.
DOI: 10.1371/journal.ppat.0020061.t003

Strain Lineage and Drug Resistance–Conferring Mutations

The global population structure of *M. tuberculosis* consists of six main phylogenetic lineages [45]. These lineages can be defined by single nucleotide polymorphisms and large sequence polymorphisms, which in *M. tuberculosis* behave as unique event polymorphisms, and therefore allow for robust phylogenetic inference [46,47]. To evaluate the influence of strain genetic background on the genetics of resistance, associations between strain lineage and isoniazid resistance–conferring mutations were sought. All of the 152 isoniazid-resistant isolates were screened for lineage-specific genetic markers and classified as belonging to the East-Asian lineage, the Indo-Oceanic lineage, or the Euro-American lineage [45]. Of the strains examined in this way, 37 (24.3%) belonged to the East-Asian lineage, 51 (33.6%) to the Indo-Oceanic lineage, and 64 (42.1%) to the Euro-American lineage. After exclusion of ten isolates harboring a combination of *katG* and *inhA* promoter mutations (Table 2), we found statistically significant associations between the three lineages and the three groups of isoniazid-resistance mutations (Table 4). The East-Asian lineage was associated with KatG mutations other than S315T, the *inhA* promoter mutation occurred significantly more frequently among the Indo-Oceanic lineage, and the Euro-American lineage was associated with the KatG S315T mutation. These associations suggest that genetic differences between the strain lineages could influence their propensity for different mutations.

We previously showed that within San Francisco, strains belonging to the Euro-American lineage were more likely to result in secondary cases than strains from other lineages [45]. In order to control for this potential confounding factor, we repeated our analysis of transmission and included only the 64 isolates belonging to the Euro-American lineage. We found that 14 out of 52 isolates (26.9%) with isoniazid-resistance mutations from Group 2 and 3 were genetically clustered, whereas none of the 12 isolates with other *katG* mutations were genetically clustered (Fisher's 2-tailed exact test: $p = 0.0539$).

Discussion

Isoniazid resistance is caused by several distinct mechanisms, and data from animal models indicate there are variations in the pathogenicity of isoniazid-resistant strains [22,23]. As reported here, differences are also apparent in human tuberculosis, as among mutations in KatG only the S315T mutation was associated with secondary cases. In contrast, strains with other mutations in *katG* did not produce a single secondary case in the population during the 9-y study period. Catalase-peroxidase encoded by *katG* transforms the prodrug isoniazid into its bioactive form, as well as protecting

Table 4. Association between Three *M. tuberculosis* Strain Lineages and Three Classes of Isoniazid Resistance–Conferring Mutations

Mutation	East-Asian Lineage ($n = 36$) *n* (%)	Indo-Oceanic Lineage ($n = 45$) *n* (%)	Euro-American Lineage ($n = 61$) *n* (%)	Odds Ratio (95% CI)[a]	*P*
KatG mutations other than S315T	15 (41.7)	2 (4.4)	10 (16.4)	5.6 (2.1–15.1)	< 0.001
KatG S315T	13 (36.1)	15 (33.3)	31 (50.8)	2.0 (0.94–4.1)	0.0517
inhA promotor-15;c-t	5 (13.9)	21 (46.7)	13 (21.3)	3.8 (1.6–9.0)	< 0.001
No/other mutation	3 (8.3)	7 (15.6)	7 (11.5)	—	—

[a]For each lineage, the χ^2 test was used to compare the proportion of the most frequent mutation to its proportion in the two other lineages combined.
DOI: 10.1371/journal.ppat.0020061.t004

bacteria against oxidative stress encountered during host infection [28]. Loss of KatG activity is therefore usually associated with reduced virulence [23]. Our data are consistent with the findings of in vitro and in vivo experiments showing that isogenic mutants with the KatG S315T substitution retain KatG activity and associated virulence in mice [23,43,44]. Recently, the crystal structure of KatG has shown how the S315T mutation could block binding of isoniazid without interfering with catalysis [48]. Only one other mutation, the S275T, has been evaluated in mice and has been found to be of low virulence [23]. However, because all of the KatG mutations identified in this study have not been functionally characterized, and some do not abolish enzymatic activity completely, it therefore remains possible that other characteristics linked to both the KatG S315T and *inhA* promoter mutations could account for the differences in transmission between mutants. Nevertheless, our results suggest that only the KatG S315T mutant retains sufficient levels of catalase-peroxidase activity to be effectively propagated within the study setting.

Mutations in the *inhA* gene promoter were also associated with secondary cases. There is no obvious explanation of how *inhA* promoter mutations might affect pathogenicity, but they have not been formally evaluated in animal models. Mutations intrinsic to the *inhA* gene have been formally tested and do not affect pathogenicity [49]. Mutations in the promoter of *ahpC* are thought to compensate for the loss of KatG activity in isoniazid-resistant strains through up-regulation of *ahpC* that encodes a putative alkyl hydroperoxide reductase [32–34]. This notion is supported by our study, as ten out of 11 isolates with *ahpC* promoter mutations were found to have mutations or deletions in *katG*. None of these double mutants were genetically clustered, indicating that *ahpC* promoter mutations are not essential or important for the transmission of isoniazid-resistant strains. Other genetic mechanisms of resistance to isoniazid exist that were not examined here, but these occur infrequently in clinical isolates and are therefore unlikely to be the basis for our results [26].

Mutational events are generally random and individual tuberculosis patients can harbor subpopulations of the same mycobacterial clone with different drug resistance–conferring mutations [50]. However, studies have shown that there seems to be a strong selection for low-cost drug resistance mutations in vivo [10,12,51]. The fact that the KatG S315T and the *inhA*-15;c-t mutations were associated with secondary cases in San Francisco suggests that selection for low-cost mutations in humans may also occur through successful transmission. This is consistent with other studies reporting these two mutations as the clinically most common isoniazid resistance mutations [27,52,53].

An obvious question posed by these results is how the KatG S315T and *inhA* promoter mutants compare to fully susceptible strains. A number of potential confounders make comparisons between resistant and drug-sensitive strains problematic [51,54]. For example, the use of additional genetic and phenotypic markers to epidemiologically link the drug-resistant strains will reduce cluster size in this group. Nevertheless, despite the effective tuberculosis control program in San Francisco [21], the KatG S315T and *inhA* promoter mutants were still transmitting and causing secondary cases of drug-resistant *M. tuberculosis*. Therefore, in less successful programs, or in areas where resistance testing and individualized treatment are not available, a longer period of infectiousness of patients will result in dissemination of these strains.

In addition to showing that drug resistance mutations are affecting pathogenicity, our results demonstrate a second means by which bacterial genetics are influencing the emergence of drug-resistant tuberculosis. We found that different strain lineages were associated with distinct isoniazid resistance mutations. These lineages are dominant in specific geographical areas where the patients are likely to have acquired their tuberculosis infection [45,46]. Differing tuberculosis control and patient adherence practices in each region could favor the emergence of specific drug resistance mutations. However, this seems unlikely as the East-Asian strains were from patients originally from a geographically wide area, including Vietnam, China, and Hong Kong, throughout which treatment practices will have varied. An alternative explanation is that there are genetic differences between the strain lineages that influence their adaptation to isoniazid, as has been observed in other bacteria [11]. For example, the association of the East-Asian lineage, with KatG mutations predicted to be poorly functional, suggests this lineage may be better able to adapt to increases in oxidative stress. Interestingly, a recent study has shown that members of the East-Asian strain lineage produce a phenolic glycolipid that inhibits the innate immune response [55], which might affect the oxidative burst generated by host immunity [28].

Other studies have failed to demonstrate similar associations between specific drug resistance mutations and strain genotypes [56]. This discrepancy is most likely due to the different genetic markers used to define the different mycobacterial lineages. Traditional mycobacterial genotyping techniques used in epidemiological studies of tuberculosis transmission are based on mobile or repetitive genetic

elements that change relatively rapidly [57], accounting for the high discriminatory power of these techniques. However, because of this fast molecular clock, similar DNA patterns can emerge in phylogenetically unrelated strains (homoplasy), making it difficult to infer deep phylogenetic groups unambiguously [46,58–61]. To our knowledge, only one other study has used unique event polymorphisms to define strain lineages in a large collection of drug-resistant strains. In their study, Baker and colleagues defined major strain lineages using synonymous single nucleotide polymorphisms [47]. Consistent with our findings, Baker and co-workers reported the same association between the Indo-Oceanic lineage (corresponding to Lineage IV in the study by Baker et al.) and the *inhA* promoter-15c-t mutation. No association with the East-Asian lineage was found in the study by Baker et al., probably because of the limited number of East-Asian strains that were included [47].

The study presented here is based on the assumption that genetic clustering is an accurate measure of ongoing tuberculosis transmission. Although there is a wealth of published evidence in support of this [25,36–41], alternative hypotheses can be envisaged. For example, certain combinations of strain genetic background and resistance mutations could result in hyper-virulence leading to rapid progression and a greater probability of diagnosis and inclusion in the study. This cannot be ruled out, but is unlikely, as the transmission dynamics of the three main genetic lineages in San Francisco have been studied and considered here [45]. Furthermore, 85% of all isoniazid-resistant isolates collected during a 9-y period by population-based sampling were included in this study, so systematic exclusion of specific genotypes is also unlikely.

In conclusion, our study shows that in addition to well-described clinical determinants of tuberculosis transmission, bacterial genetics also have an important impact on the propagation of drug-resistant *M. tuberculosis*. Further work is required to establish whether these findings can be extended to other classes of anti-tuberculosis drugs and to other settings such as those with high rates of transmission where host susceptibility is altered by HIV infection. Ultimately, a complete understanding of drug resistance will require the integration of host factors, patterns of antibacterial usage, as well as bacterial genetics.

Materials and Methods

Study population. The epidemiological methods and study population have been described previously [21,25]. All patients from San Francisco who met the Centers for Disease Control and Prevention criteria as an incident case of tuberculosis during the years 1991–1999 were included in the study population. There were 2,498 tuberculosis cases reported, 2,142 (85.7%) were culture positive, and 2,104 (98.2%) were identified as *M. tuberculosis*.

Drug resistance testing. All *M. tuberculosis* isolates from incident cases of tuberculosis were tested for drug susceptibility using standard BACTEC methods. The following concentrations were used to define resistance: isoniazid 0.1 mg/l, rifampicin 2 mg/l, streptomycin 2 mg/l, and ethambutol 2.5 mg/l. When detected, drug resistance was confirmed with culture on solid media. Patients infected with strains that were initially fully susceptible, but who subsequently developed resistant strains, were not included in this study.

Molecular epidemiology. Standard epidemiologic data were collected as part of a prospective population-based molecular epidemiologic study of tuberculosis in San Francisco [25]. For each isolate, genomic DNA was extracted from 10-ml liquid cultures and subcultured from the initial Lowenstein Jensen slope used for the primary isolation. RFLP typing was performed following standardized procedures [57], with patterns entered and analyzed using a BioImage Whole Band Analyzer software (version 9 3.3, BioImage Corporation, Ann Arbor, Michigan, United States). Isolates with fewer than six copies of IS*6110* were further genotyped using polymorphic GC-rich sequence (PGRS) RFLP [57]. For the majority of isolates an aliquot of this original DNA was preserved at −20 °C and was used as a template for amplifying resistance genes. In a few cases isolates were re-cultured from −80 °C stocks preserved in glycerol. The protocols and the procedures for the protection of human participants were approved by Stanford University and the University of California, San Francisco.

Isolates with the same RFLP genotype (± one IS*6110* band for strains with > 5 bands; ± one IS*6110* band and identical PGRS patterns in isolates with < 6 IS*6110* copies), the same drug resistance profile, the same drug resistance–conferring mutation, and belonging to the same *M. tuberculosis* lineage were considered genetically clustered and part of a chain of transmission. Isolates that differed in any of these criteria were considered unique and not linked epidemiologically. Two isolates were classified as clustered based on the RFLP pattern and drug resistance profile.

Sequencing of drug resistance genes. To determine the mutations responsible for isoniazid resistance in each isolate, known resistance genes were sequenced. Oligonucleotide primers were designed for the amplification and sequencing of the entire *furA-katG* locus (H37Rv: 2153626–2156657, 3,031 bp), the promoter region of *inhA* (H37Rv: 1673196–1673653, 457 bp), and the *oxyR-ahpC* intergenic region (H37Rv: 2726390–2725888, 502 bp). DNA was amplified in a 25-μl reaction using a GeneAmp system 9700 thermocycler (Applied Biosystems, Foster City, California, United States). Unincorporated nucleotides and primers were removed by filtration with Multiscreen-PCR plates (Millipore, Billerica, Massachusetts, United States). Sequencing reactions were performed with the ET Terminator sequencing kit (Amersham Biosciences, Little Chalfont, United Kingdom) and purified in Multiscreen plates (Millipore) with Sephadex (Amersham Biosciences). Sequence data were generated with an ABI 377 automated sequencer (Applied Biosystems) at the PAN facility, Stanford University (http://cmgm.stanford.edu/pan) and analyzed with SeqMan (DNASTAR, Inc.).

Determination of strain lineage by multiplex real-time PCR. The population structure of *M. tuberculosis* in San Francisco consists of three main phylogenetic lineages [45]. To determine if the study isolates belonged to either the East-Asian lineage, the Indo-Oceanic lineage, or the Euro-American lineage, the corresponding lineage specific markers were targeted by multiplex real-time PCR as reported previously [45,62,63]. The three lineages were defined based on the absence of the region of difference (RD) 105, absence of RD239, and the ctg to cgg substitution at KatG463, respectively.

Statistical analysis. To explore the relationship between RFLP clustering and clinical, demographic, and bacterial variables, univariate analyses using the χ^2 test of proportions or the 2-tailed Fisher's exact test and multivariate logistic regression modeling were performed. All variables with a p-value ≤ 0.2 in the univariate analysis and biological plausibility were considered for the multivariate logistic regression model. We performed forward stepwise model construction and compared the log likelihood ratios of successive models until the final, most parsimonious model was identified [64]. We also tested the significance of potential interaction terms. Statistical analyses were performed with Stata (version 7E; Stata Corporation, College Station, Texas, United States).

Acknowledgments

We would like to thank the staff of the Tuberculosis Section, San Francisco Department of Public Health Tuberculosis, and the Molecular Epidemiology Laboratory at Stanford University for their contributions to the diligent collection and curation of the samples and epidemiological data that made this study possible.

Author contributions. SG, MVB, KD, PMS, and ASP conceived and designed the experiments. SB, MVB, AE, SM, and ASP performed the experiments. SG, KD, and ASP analyzed the data. SG, MVB, KD, PCH, PMS, and ASP wrote the paper.

Funding. This work was supported by the Swiss National Science Foundation and the Novartis Foundation (SG), the National Institutes of Health (grants K01 TW000001 to KD, and AI34238 to PCH), and the Wellcome Trust (PMS). ASP was supported by a Wellcome Trust training fellowship.

Competing interests. The authors have declared that no competing interests exist.

References

1. World Health Organization (2004) Anti-tuberculosis drug resistance in the world: Third global report. Geneva: World Health Organization. Report Number HO/HTM/TB/2004.343.
2. Blower SM, Chou T (2004) Modeling the emergence of the "hot zones": Tuberculosis and the amplification dynamics of drug resistance. Nat Med 10: 1111–1116.
3. Cohen T, Murray M (2004) Modeling epidemics of multidrug-resistant *M. tuberculosis* of heterogeneous fitness. Nat Med 10: 1117–1121.
4. Dye C, Espinal MA, Watt CJ, Mbiaga C, Williams BG (2002) Worldwide incidence of multidrug-resistant tuberculosis. J Infect Dis 185: 1197–1202.
5. DeRiemer K, Garcia-Garcia L, Bobadilla-del-Valle M, Palacios-Martinez M, Martinez-Gamboa A, et al. (2005) Does DOTS work in populations with drug-resistant tuberculosis? Lancet 365: 1239–1245.
6. Levin BR, Lipsitch M, Perrot V, Schrag S, Antia R, et al. (1997) The population genetics of antibiotic resistance. Clin Infect Dis 24: S9–S16.
7. Austin DJ, Anderson RM (1999) Studies of antibiotic resistance within the patient, hospitals, and the community using simple mathematical models. Philos Trans R Soc Lond B Biol Sci 354: 721–738.
8. Andersson DI, Levin BR (1999) The biological cost of antibiotic resistance. Curr Opin Microbiol 2: 489–493.
9. Bjorkman J, Nagaev I, Berg OG, Hughes D, Andersson DI (2000) Effects of environment on compensatory mutations to ameliorate costs of antibiotic resistance. Science 287: 1479–1482.
10. Sander P, Springer B, Prammananan T, Sturmfels A, Kappler M, et al. (2002) Fitness cost of chromosomal drug resistance-conferring mutations. Antimicrob Agents Chemother 46: 1204–1211.
11. Cohan FM, King EC, Zawadzki P (1994) Amelioration of the deleterious pleiotropic effects of an adaptive mutation in *Bacillus subtilis*. Evolution 48: 81–95.
12. Bottger EC, Springer B, Pletschette M, Sander P (1998) Fitness of antibiotic-resistant microorganisms and compensatory mutations. Nat Med 4: 1343–1344.
13. Bjorkman J, Hughes D, Andersson DI (1998) Virulence of antibiotic-resistant *Salmonella typhimurium*. Proc Natl Acad Sci U S A 95: 3949–3953.
14. Maisnier-Patin S, Andersson DI (2004) Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. Res Microbiol 155: 360–369.
15. Reynolds MG (2000) Compensatory evolution in rifampin-resistant *Escherichia coli*. Genetics 156: 1471–1481.
16. Schrag SJ, Perrot V (1996) Reducing antibiotic resistance. Nature 381: 120–121.
17. Garcia-Garcia ML, Ponce de Leon A, Jimenez-Corona ME, Jimenez-Corona A, Palacios-Martinez M, et al. (2000) Clinical consequences and transmissibility of drug-resistant tuberculosis in southern Mexico. Arch Intern Med 160: 630–636.
18. Godfrey-Faussett P, Sonnenberg P, Shearer SC, Bruce MC, Mee C, et al. (2000) Tuberculosis control and molecular epidemiology in a South African gold-mining community. Lancet 356: 1066–1071.
19. van Soolingen D, Borgdorff MW, de Haas PE, Sebek MM, Veen J, et al. (1999) Molecular epidemiology of tuberculosis in the Netherlands: A nationwide study from 1993 through 1997. J Infect Dis 180: 726–736.
20. Cohen T, Sommers B, Murray M (2003) The effect of drug resistance on the fitness of *Mycobacterium tuberculosis*. Lancet Infect Dis 3: 13–21.
21. Burgos M, DeRiemer K, Small PM, Hopewell PC, Daley CL (2003) Effect of drug resistance on the generation of secondary cases of tuberculosis. J Infect Dis 188: 1878–1884.
22. Ordway DJ, Sonnenberg MG, Donahue SA, Belisle JT, Orme IM (1995) Drug-resistant strains of *Mycobacterium tuberculosis* exhibit a range of virulence for mice. Infect Immun 63: 741–743.
23. Pym AS, Saint-Joanis B, Cole ST (2002) Effect of *katG* mutations on the virulence of *Mycobacterium tuberculosis* and the implication for transmission in humans. Infect Immun 70: 4955–4960.
24. van Soolingen D, de Haas PE, van Doorn HR, Kuijper E, Rinder H, et al. (2000) Mutations at amino acid position 315 of the *katG* gene are associated with high-level resistance to isoniazid, other drug resistance, and successful transmission of *Mycobacterium tuberculosis* in the Netherlands. J Infect Dis 182: 1788–1790.
25. Jasmer RM, Hahn JA, Small PM, Daley CL, Behr MA, et al. (1999) A molecular epidemiologic analysis of tuberculosis trends in San Francisco, 1991–1997. Ann Intern Med 130: 971–978.
26. Ramaswamy SV, Reich R, Dou SJ, Jasperse L, Pan X, et al. (2003) Single nucleotide polymorphisms in genes associated with isoniazid resistance in *Mycobacterium tuberculosis*. Antimicrob Agents Chemother 47: 1241–1250.
27. Ramaswamy S, Musser JM (1998) Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. Tuber Lung Dis 79: 3–29.
28. Ng VH, Cox JS, Sousa AO, MacMicking JD, McKinney JD (2004) Role of KatG catalase-peroxidase in mycobacterial pathogenesis: Countering the phagocyte oxidative burst. Mol Microbiol 52: 1291–1302.
29. Banerjee A, Dubnau E, Quemard A, Balasubramanian V, Um KS, et al. (1994) *inhA*, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*. Science 263: 227–230.
30. Dessen A, Quemard A, Blanchard JS, Jacobs WR Jr, Sacchettini JC (1995) Crystal structure and function of the isoniazid target of *Mycobacterium tuberculosis*. Science 267: 1638–1641.
31. Rozwarski DA, Grant GA, Barton DH, Jacobs WR Jr, Sacchettini JC (1998) Modification of the NADH of the isoniazid target (InhA) from *Mycobacterium tuberculosis*. Science 279: 98–102.
32. Wilson T, de Lisle GW, Marcinkeviciene JA, Blanchard JS, Collins DM (1998) Antisense RNA to *ahpC*, an oxidative stress defence gene involved in isoniazid resistance, indicates that AhpC of *Mycobacterium bovis* has virulence properties. Microbiology 144: 2687–2695.
33. Sherman DR, Mdluli K, Hickey MJ, Arain TM, Morris SL, et al. (1996) Compensatory *ahpC* gene expression in isoniazid-resistant *Mycobacterium tuberculosis*. Science 272: 1641–1643.
34. Heym B, Stavropoulos E, Honore N, Domenech P, Saint-Joanis B, et al. (1997) Effects of overexpression of the alkyl hydroperoxide reductase AhpC on the virulence and isoniazid resistance of *Mycobacterium tuberculosis*. Infect Immun 65: 1395–1401.
35. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, et al. (1997) Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci U S A 94: 9869–9874.
36. van Soolingen D (2001) Molecular epidemiology of tuberculosis and other mycobacterial infections: Main methodologies and achievements. J Intern Med 249: 1–26.
37. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, et al. (1993) Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: Recommendations for a standardized methodology. J Clin Microbiol 31: 406–409.
38. Daley CL, Small PM, Schecter GF, Schoolnik GK, McAdam RA, et al. (1992) An outbreak of tuberculosis with accelerated progression among persons infected with the human immunodeficiency virus. An analysis using restriction-fragment-length polymorphisms. N Engl J Med 326: 231–235.
39. Kenyon TA, Valway SE, Ihle WW, Onorato IM, Castro KG (1996) Transmission of multidrug-resistant *Mycobacterium tuberculosis* during a long airplane flight. N Engl J Med 334: 933–938.
40. Kline SE, Hedemark LL, Davies SF (1995) Outbreak of tuberculosis among regular patrons of a neighborhood bar. N Engl J Med 333: 222–227.
41. van Deutekom H, Hoijng SP, de Haas PE, Langendam MW, Horsman A, et al. (2004) Clustered tuberculosis cases: Do they represent recent transmission and can they be detected earlier? Am J Respir Crit Care Med 169: 806–810.
42. Heym B, Alzari PM, Honore N, Cole ST (1995) Missense mutations in the catalase-peroxidase gene, *katG*, are associated with isoniazid resistance in *Mycobacterium tuberculosis*. Mol Microbiol 15: 235–245.
43. Saint-Joanis B, Souchon H, Wilming M, Johnsson K, Alzari PM, et al. (1999) Use of site-directed mutagenesis to probe the structure, function, and isoniazid activation of the catalase/peroxidase, KatG, from *Mycobacterium tuberculosis*. Biochem J 338: 753–760.
44. Wengenack NL, Uhl JR, St Amand AL, Tomlinson AJ, Benson LM, et al. (1997) Recombinant *Mycobacterium tuberculosis* KatG(S315T) is a competent catalase-peroxidase with reduced activity toward isoniazid. J Infect Dis 176: 722–727.
45. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, et al. (2006) Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 103: 2869–2873.
46. Hirsh AE, Tsolaki AG, DeRiemer K, Feldman MW, Small PM (2004) Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. Proc Natl Acad Sci U S A 101: 4871–4876.
47. Baker L, Brown T, Maiden MC, Drobniewski F (2004) Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. Emerg Infect Dis 10: 1568–1577.
48. Bertrand T, Eady NA, Jones JN, Jesmin, Nagy JM, et al. (2004) Crystal structure of *Mycobacterium tuberculosis* catalase-peroxidase. J Biol Chem 279: 38991–38999.
49. Wilson TM, de Lisle GW, Collins DM (1995) Effect of *inhA* and *katG* on isoniazid resistance and virulence of *Mycobacterium bovis*. Mol Microbiol 15: 1009–1015.
50. Post FA, Willcox PA, Mathema B, Steyn LM, Shean K, et al. (2004) Genetic polymorphism in *Mycobacterium tuberculosis* isolates from patients with chronic multidrug-resistant tuberculosis. J Infect Dis 190: 99–106.
51. Bottger EC, Pletschette M, Andersson D (2005) Drug resistance and fitness in *Mycobacterium tuberculosis* infection. J Infect Dis 191: 823–824.
52. Cohen T, Becerra MC, Murray MB (2004) Isoniazid resistance and the future of drug-resistant tuberculosis. Microb Drug Resist 10: 280–285.
53. Mokrousov I, Narvskaya O, Otten T, Limeschenko E, Steklova L, et al. (2002) High prevalence of KatG Ser315Thr substitution among isoniazid-resistant *Mycobacterium tuberculosis* clinical isolates from northwestern Russia, 1996 to 2001. Antimicrob Agents Chemother 46: 1417–1424.
54. Borgdorff MW (2004) Transmission of isoniazid-resistant tuberculosis. J Infect Dis 190: 650.
55. Reed MB, Domenech P, Manca C, Su H, Barczak AK, et al. (2004) A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. Nature 431: 84–87.
56. Lavender C, Globan M, Sievers A, Billman-Jacobe H, Fyfe J (2005) Molecular characterization of isoniazid-resistant *Mycobacterium tuberculosis* isolates collected in Australia. Antimicrob Agents Chemother 49: 4068–4074.

57. Kremer K, van Soolingen D, Frothingham R, Haas WH, Hermans PW, et al. (1999) Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: Interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol 37: 2607–2618.
58. Monot M, Honore N, Garnier T, Araoz R, Coppee JY, et al. (2005) On the origin of leprosy. Science 308: 1040–1042.
59. Keim P, Van Ert MN, Pearson T, Vogler AJ, Huynh LY, et al. (2004) Anthrax molecular epidemiology and forensics: Using the appropriate marker for different evolutionary scales. Infect Genet Evol 4: 205–213.
60. Gutacker MM, Mathema B, Soini H, Shashkina E, Kreiswirth BN, et al. (2006) Single nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from four geographic sites. J Infect Dis 193: 121–128.
61. Filliol I, Motiwala AS, Cavatore M, Qi W, Hernando Hazbon M, et al. (2006) Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: Insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. J Bacteriol 188: 759–772.
62. Tsolaki AG, Gagneux S, Pym AS, Goguet de la Salmoniere YO, Kreiswirth BN, et al. (2005) Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. J Clin Microbiol 43: 3185–3191.
63. Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, et al. (2004) Functional and evolutionary genomics of *Mycobacterium tuberculosis*: Insights from genomic deletions in 100 strains. Proc Natl Acad Sci U S A 101: 4865–4870.
64. Selvin S (1998) Modern applied biostatistical methods using S-Plus. In: Monographs in Epidemiology and Biostatistics. New York: Oxford University Press. pp. 385–391.