MAJOR ARTICLE

Dynamic Population Changes in *Mycobacterium tuberculosis* During Acquisition and Fixation of Drug Resistance in Patients

Gang Sun,[1,a] Tao Luo,[1,a] Chongguang Yang,[1] Xinran Dong,[2] Jing Li,[3] Yongqiang Zhu,[4] Huajun Zheng,[4] Weidong Tian,[2] Shengyue Wang,[4] Clifton E. BarryIII,[5] Jian Mei,[3] and Qian Gao[1]

[1]Key Laboratory of Medical Molecular Virology, Institutes of Biomedical Sciences and Institute of Medical Microbiology, Fudan University, [2]Institute of Biostatistics, School of Life Sciences, Fudan University, [3]Department of TB Control, Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China; [4]Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai, China; and [5]Tuberculosis Research Section, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland

(See the Editorial Commentary by Fortune, on pages 1642–4.)

Background. Drug-resistant tuberculosis poses a growing challenge to global public health. However, the diversity and dynamics of the bacterial population during acquisition of drug resistance have yet to be carefully examined.

Methods. Whole-genome sequencing was performed on 7 serial *Mycobacterium tuberculosis* (*M. tuberculosis*) populations from 3 patients during different stages in the development of drug resistance. The population diversity was assessed by the number and frequencies of unfixed mutations in each sample.

Results. For each bacterial population, 8–41 unfixed mutations were monitored by the fraction of single-nucleotide polymorphisms at specific loci. Among them, as many as 4 to 5 resistance-conferring mutations were transiently detected in the same single sputum, but ultimately only a single type of mutant was fixed. In addition, we identified 14 potential compensatory mutations that occurred during or after the emergence of resistance-conferring mutations.

Conclusions. *M. tuberculosis* population within patients exhibited considerable genetic diversity, which underwent selections for most fit resistant mutant. These findings have important implications and emphasize the need for early diagnosis of tuberculosis to decrease the chance of evolving highly fit drug-resistant strains.

Drug-resistant tuberculosis, especially multidrug resistant (MDR) and extensively drug-resistant tuberculosis (XDR-Tuberculosis) is a growing challenge to global tuberculosis control. *Mycobacterium tuberculosis* (*M. tuberculosis*) acquires antibiotic resistance mainly through chromosomal mutations that either directly affect the binding site of the drug or mutate bacterial enzymes involved in activating a pro-drug [1]. Such mutations often confer fitness defect on the resulting drug-resistant progeny. However, although hundreds of resistance-associated mutations have been identified in vitro and in vivo [2], drug-resistant alleles in clinical isolates are characterized by the prevalence of low- or no-cost mutations [3–5], suggesting that selection for drug-resistant alleles with the least fitness cost may occur in vivo.

Mutations giving rise to drug resistance have also been studied on serial samples from treated tuberculosis patients [6, 7]. These studies have documented that mixed populations of resistant and sensitive *M. tuberculosis* exist in a single clinical sample, and that more than 1 mutation can emerge during treatment. This mixed state of sensitive and resistant alleles, called heteroresistance [8], was detected in about 20% of sputum samples containing resistant organisms [9, 10]. Although these studies consider heteroresistance to be defined by the presence of both resistant and sensitive alleles in the same sputum, heteroresistance may also describe a mixed infection containing different drug resistance–conferring

Received 23 March 2012; accepted 29 June 2012; electronically published 14 September 2012.

[a]G. S. and T. L. contributed equally to this work.

Correspondence: Qian Gao, MD, PhD, Shanghai Medical College, Fudan University, 138 Yi Xue Yuan Rd, Shanghai, 200032, China (qgao99@yahoo.com).

The Journal of Infectious Diseases 2012;206:1724–33

DOI: 10.1093/infdis/jis601

alleles, which then may be selected for those with the lowest fitness cost in vivo.

Drug-resistant clinical isolates are also often found to have enhanced fitness compared to their apparently isogenic in vitro selected counterparts (at least as measured by in vitro growth rate), which suggests the evolution of unidentified compensatory mutations in vivo [4]. Since most drug resistance–associated alleles impose a fitness cost [4, 11, 12], compensatory evolution may play an important role in the stable fixation of such mutants within the bacterial population. However, due to the complexity of compensatory mechanisms, very few of them have been elucidated in *M. tuberculosis* [13–15]. Previous studies on resistant *M. tuberculosis* have mainly focused on resistance-conferring mutations, and the potential compensatory changes have usually been neglected [6, 7].

Recent developments in genome sequencing technologies provide an unprecedented opportunity to study the diversity of populations of microorganisms [16]. Effective statistical approaches to differentiate technical errors from real polymorphisms and high-sequencing coverage enable accurate detection of low-frequency variants in microorganism populations [17, 18]. Deep sequencing of genes from human immunodeficiency virus and hepatitis C virus has begun to reveal patterns of within-host diversity and evolution during infections [18, 19]. A study on a series of mixed-population *Escherichia coli* samples has uncovered changes in the genetic structure of the population during long-term in vitro evolution [17]. Despite these advances, previous whole-genome sequencing studies of *M. tuberculosis* have mainly focused on polymorphisms between transmitted MDR and XDR strains after isolating single-colony isolates [20–22]. A recent study performed whole-genome sequencing on serial sputum samples from an MDR-Tuberculosis patient; however, due to the limitations of analytical methods, very few resistance-conferring mutations were characterized and the evolution of bacterial populations remained unclear [23].

In the present study, we sequenced the bacterial genomes of serially collected samples from treated tuberculosis patients. By identifying the genetic variations during antibiotic treatment, we aim to further understand the in vivo evolution of resistance associated and potential compensatory mutations.

METHODS

Selection of Patients and Samples

M. tuberculosis sputum samples were collected through routine clinical work and cultured on slopes by local hospitals. After about 4 weeks, culture-positive slopes were sent to Shanghai Municipal Center of Disease Control and Prevention (Shanghai CDC), where the bacteria were harvested to 1 mL 15% glycerol and refrigerated in the specimen bank. Approved by the Ethical Review Committee of Shanghai CDC, we selected the tuberculosis patients with serial samples during the treatment, and strains from the same patient shared identical 16 hyperpolymorphic variable-number tandem repeat (VNTR) patterns [24].

The drug susceptibility tests (DSTs) were performed by the concentration method. The concentration for each antibiotic was as follows: RIF, 50 μg/mL; INH, 50 μg/mL; SM, 10 μg/mL; EMB, 5 μg/mL.

Genome Sequencing

In the present study, 0.5 mL of each preserved sample (about 50% of the original culture) was spread onto a fresh Lowenstein–Jensen plate. After 4 weeks' culture, all the bacteria colonies were scratched from the plate and the genomic DNA was extracted following the protocol of cetyltrimethylammonium bromide–lysozyme method [25]. A 300-base-pair (bp) paired-end library was constructed for each purified DNA sample following the standard Illumina paired-end protocol with a low-cycle polymerase chain reaction during the fragment enrichment, and sequencing was performed on the Illumina Genome Analyzer platform with 95 or 115 cycles. The sequencing data were submitted to the National Center for Biotechnology Information Sequence Read Archive under the Accession No. SRA050212.1.

Reads Processing

The pipeline that we applied for analyzing bacteria populations was a combination of previously published protocols (see Supplementary Figure 1). The FASTQ data were transformed to the Sanger format, followed by the trimming of adaptor and low-quality bases at the 3′ ends by Scythe (https://github.com/ucdavis-bioinformatics/scythe) and Sickle (https://github.com/ucdavis-bioinformatics/sickle), respectively. Reads alignment was performed by Burrows–Wheeler Aligner [26]. To improve the accuracy of mutation calling, we employed 2 reference genomes for this and subsequent processes. One is the standard reference strain, *M. tuberculosis* H37Rv (GenBank: AL123456) [27] and the other is CCDC5079 (GenBank: CP001641) [28], a clinical isolate close to our strains. After duplicates marking and local realignment by Genome Analysis ToolKit [29], the software Samtools [30] was used to convert aligned reads into pileup-formatted files.

To ensure the reliability of subsequent analysis, we first discarded low-quality bases with Phred quality scores lower than 20. Then, we removed 2% of sites with the lowest coverage and loci with coverage higher than twice the average level. Indels, mobile genetic elements, and PE/PPE and PE-PGRS gene families that might cause incorrect read alignment were also excluded.

Distinguishing Sequencing From Real Mutations

For sites with only single mapping result, fixed mutations with frequency of 100% were first identified. For the other loci with

heterozygous mapping results, the protocol and parameters for mutation calling, including base error model, likelihood ratio test, and bias filtering were referenced from a published methodology [17].

Identifying Significant Genomic Changes

The detected frequency for each single-nucleotide polymorphism (SNP) was calculated by a maximum likelihood approach [17]. Considering the minimal sequencing depth included for our analysis, SNPs with frequencies higher than 95% were defined as fixed mutations, while those with frequencies from 5% to 95% were defined as unfixed mutations. For fixed SNPs that were repeatedly detected in all serial samples from the same patient, they were grouped as common fixed mutations that emerged before the sampling intervals. For unfixed SNPs that were only detected by one reference but not by the other, they were excluded as false-positive results caused by reference-specific improper read mapping. The other SNPs, including a few fixed mutations that were only observed in part of the samples and unfixed mutations with varying frequencies, were grouped as characteristic mutations, representing genetic variations during treatment.

To identify significant genomic changes at a given locus between serial samples, the numbers of aligned reads harboring wild and mutated genotype were submitted to the Fisher exact test. The p value was then multiplied by the total number of included bases to gain an E value. A significant genomic change was defined by E value lower than 1.

RESULTS

Patients and Samples

Seven sputum samples from 3 patients (denoted A–C) were collected for analysis (Figure 1). Patient A was sampled over a 23-month period during which this subject developed isoniazid monoresistance. This subject had a record indicating poor compliance from the 3rd to the 13th month. Patient B was sampled over an 18-month period during which this subject acquired rifampin resistance in addition to their initial isoniazid and streptomycin resistance. Patient C was a retreatment case and was sampled over an 11-month treatment for MDR-Tuberculosis and showed no change in their underlying resistance pattern. In each subject, we verified that the initial and subsequent samples shared identical 16-VNTR profiles, indicating they originated from the same clonal population. Different VNTR profiles were observed among different patients, suggesting these 3 cases are epidemiologically unrelated. All these cases were infected by the Beijing family strains [31], the most dominant *M. tuberculosis* genotype in Shanghai.

Genome Sequencing

DNA samples were prepared immediately after growing the original sputum sample on Lowenstein–Jensen media to avoid as much as possible enriching for siblings with different growth rates. For each sample we obtained 3.7 to 20.2 million 115 bp or 95 bp single/paired-end reads corresponding to an average sequencing depth ranging from 55- to 269-fold, and these were aligned to the reference genome of *M. tuberculosis* H37Rv and CCDC5079, respectively. The sequencing coverage across the reference genomes was approximately 99%. To ensure the reliability of our results, we discarded 2% of loci with the lowest sequencing depth and loci with depth higher than twice the average level. Indels, mobile elements, and the PE/PPE-PGRS gene families, which may result in inaccurate identification of mutations, were also excluded. In total, bases covering about 90% of the reference genomes were included for subsequent analysis (see Supplementary Table 1).

Intrapatient Population Diversity

Standard tuberculosis treatment for drug-susceptible disease usually lasts for 6 months, but for MDR-Tuberculosis treatment, durations are much longer, typically 18–24 months. We

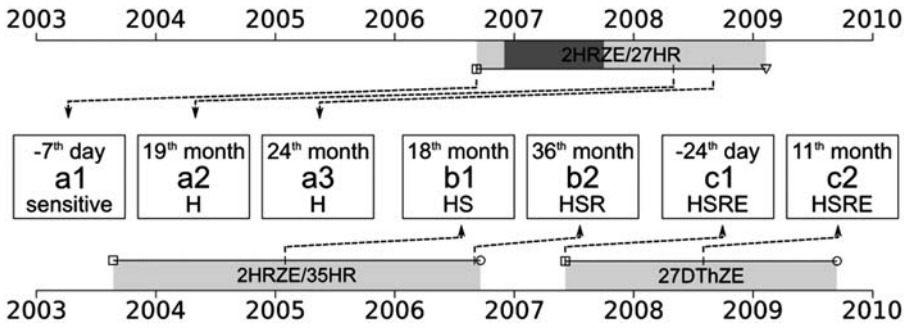

Figure 1. Patients and samples. The treatment durations are indicated as solid lines, which start with a hollow square on the 0th day and end with a triangle (default) or circle (death). The treatment schedules are presented in grayish shadows, and the poor compliance of patient A is indicated by the dark gray shadow. Sample information such as collection time points, designations, and resistance profiles are presented in boxes. Abbreviations: D, Dipasic (isoniazid aminosalicylate); E, ethambutol; H, isoniazid; R, rifampin; S, streptomycin; Th, ethionamide; Z, pyrazinamide.

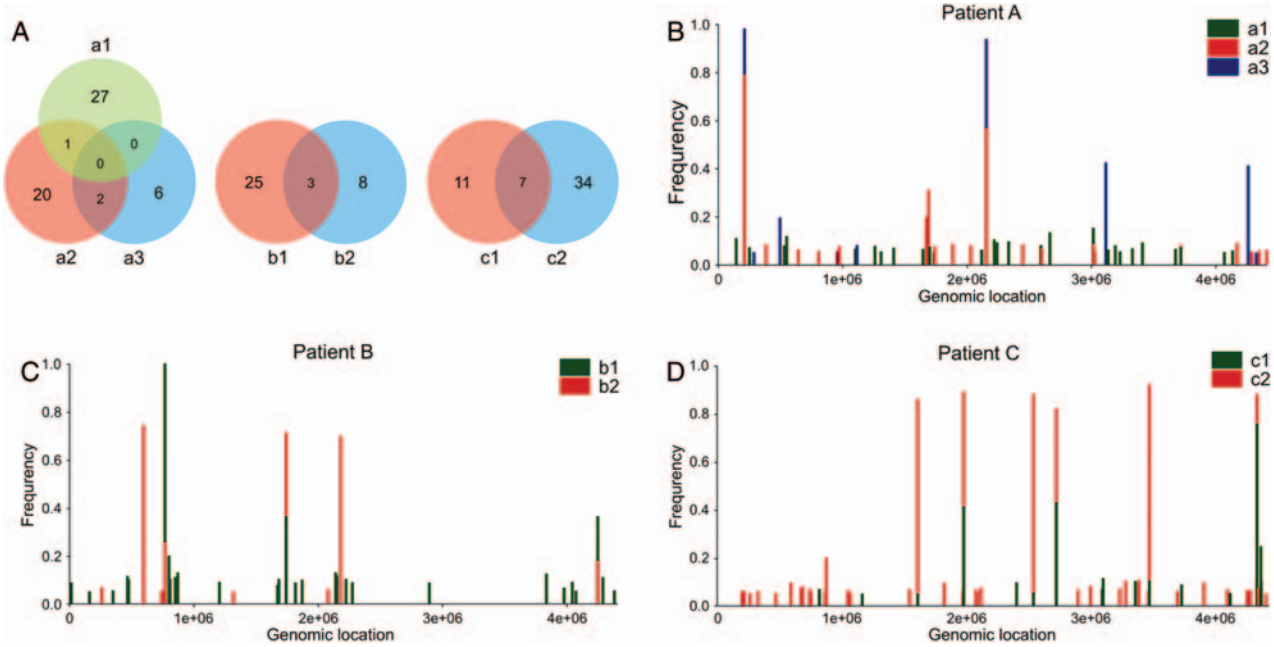

Figure 2. Characteristic SNPs. *A*, Venn diagrams showing the number of characteristic SNPs in each sample. *B–D*, Genomic distribution and frequencies of characteristic SNPs identified in each patient. Different colors represent the serially collected samples. Frequencies are estimated from the mapping result on *M. tuberculosis* H37Rv, using a previously described maximum likelihood approach [17]. Abbreviations: *M. tuberculosis*, Mycobacterium tuberculosis; SNPs, single-nucleotide polymorphisms.

hypothesized that during this prolonged treatment, *M. tuberculosis* might generate and accumulate adaptive mutations that contribute to fitness under antibiotic stress. To characterize genetic variations that evolved within our samples, we analyzed SNPs and their corresponding frequencies in each longitudinal population. After aligning reads to the reference genomes, we employed a likelihood ratio test and a bias filter to differentiate real SNPs from technical errors [17]. Then, by comparing the SNPs between serial samples, we identified characteristic mutations that evolved during treatment.

With frequency threshold at 5%, this analysis yielded 8–41 characteristic SNPs in each sample (Figure 2*A*, or see Supplementary Table 2). These included 6–34 "private" SNPs that were only found in 1 sample and were not found in subsequent or prior samples, suggesting microheterogeneity in the population sampled in the sputum at each time point. Among the total SNPs in subsequent isolates, 82.7% were found at frequencies lower than 20% (Figure 2*B–D*). Sixteen SNPs, including 6 mutations known to be associated with drug resistance, were found at frequencies higher than 50%. These SNPs were mainly (14/16) identified in the second or third sample and were consistent with the emergence of new mutations in these treated populations as a result of selection of mutations conferring drug resistance and other SNPs of adaptive potential.

Evolution of Drug Resistance–Associated Mutations

We observed heteroresistance in all 3 patients with multiple mutations associated with drug resistance to isoniazid, rifampin, and ethionamide, respectively. To understand the evolution of drug resistance, we examined the detailed population structure of these resistant bacteria over time.

In patient A, the initially untreated population was sensitive to all drugs (Figure 3*A*). After 19 months of treatment, the clinical laboratory using standard drug susceptibility testing identified isoniazid resistance. Our sequencing results at this time point indicated that the isoniazid resistance was conferred by as many as 4 independent resistance mutations: 3 in the *katG* gene, including V1A (5.1%), D94N (56.8%), and S315R (19.5%); and a fourth one in the promoter region of *inhA*, C-15T (20.0%). Because the genomic distance between these mutations were longer than the sequencing reads, mutated bases at 1 locus were not likely to bias the detection frequency of the other mutations. The total frequency of these 4 resistance mutations was very close to 100%, suggesting that the treatment of patient A had successfully eliminated sensitive bacteria and selected for a spectrum of resistant mutants that simultaneously existed within this subject. Five months later, the *inhA* C-15T promoter mutation had become undetectable and the frequencies of *katG* S315R and V1A had decreased to 0.6% and 0.8%, respectively. In contrast, the *katG* D94N allele had expanded to 94.2% of the reads, suggesting this mutant had superior fitness

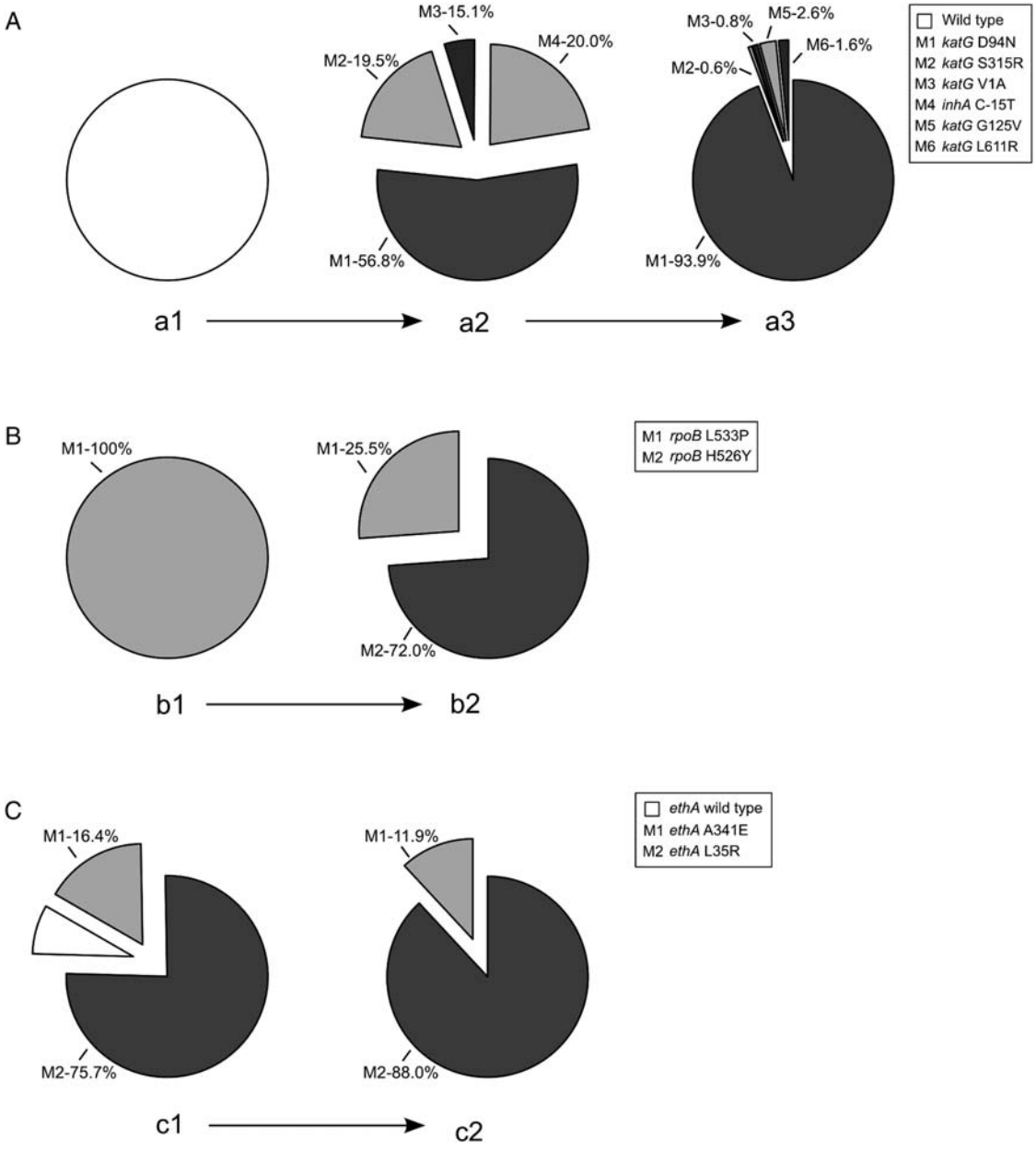

Figure 3. The dynamic heteroresistance in 3 patients. Resistance-conferring mutations are labeled as M1–M6. *A*, Dynamic heteroresistance to isoniazid in patient A. In sample a3, the 4 minor mutations (M2, M3, M5, and M6) are below the frequency threshold of 5% but still detectable. Therefore, all of them are presented for a comprehensive view. *B* and *C*, Unfixed drug resistant mutations in patients B and C, respectively.

compared to the other 3. Because of the high sequencing depth of sample a3, which had about 300-fold coverage on the *katG* gene, 2 other mutations at *katG* G125V (2.6%) and L611R (1.6%) were also detected as candidate mutants that may also have caused isoniazid resistance.

For patient B, an initial sputum sample was not available, but at the 18th month, a mutation in *rpoB* (L533P) was uniformly present in the sample (Figure 3*B*). However, drug-susceptibility testing indicated the sample remained rifampicin susceptible. Consistent with this, the *rpoB* L533P has been previously reported to confer only low-level rifampin resistance [32]. This subject remained on first-line therapy for another 18 months after which sequence analysis revealed that the *rpoB* L533P allele had been replaced primarily by *rpoB* H526Y (76.3%), an allele associated with high-level rifampicin resistance that confers a relatively low fitness cost [4].

In the retreatment case of patient C, unfixed mutations *ethA* L35R (75.7%) and A341E (16.4%) were simultaneously detected in the initial sample, suggesting acquired ethionamide resistance during the previous treatment.

Table 1. Drug Resistant Mutations Detected in the 7 Samples

		A			B		C	
Drug	Gene	a1	a2	a3	b1	b2	c1	c2
H	*katG*	…	**V1A**	**V1A**	S315T	S315T	S315T	S315T
			D94N	**D94N**				
			S315R	**G125V**				
				S315R				
				L611R				
	inhA	…	**C-15T**	…	–	–	…	…
R	*rpoB*	…	…	…	L533P	**H526Y**	S531L	S531L
						L533P		
S	*rpsL*	…	…	…	K43R	K43R	…	…
	rrs	…	…	…	…	…	A513C	A513C
Th	*ethA*	…	…	…	…	…	**L35R**	**L35R**
							A341E	**A341E**
Z	*pncA*	…	…	…	…	…	V180A	V180A
E	*embB*	…	…	…	…	…	D354A	D354A

Unfixed mutations are highlighted in bold.

Abbreviations: E, ethambutol; H, isoniazid; R, rifampin; S, streptomycin; Th, ethionamide; Z, pyrazinamide.

After an 11-month treatment, both mutations remained in the sample with frequencies at 88.0% and 11.9%, respectively (Figure 3*C*). Unlike the dynamic evolution of newly acquired resistance in these patients, the other resistance-conferring mutations were stable fixed SNPs (Table 1). These included known mutations at *katG* S315T, *rpoB* S531L, and *rpsL* K43R, all of which were fixed prior to our initial sample.

Potential Adaptive Mutations

There were mutations in the bacterial populations that increased in frequency during treatment, particularly in the case of patient C, with stable drug-resistance mutations. These results suggest that adaptive evolution continues to occur in alleles not related to drug resistance. We hypothesized that mutations with significant changes in frequencies might be relevant to the population's fitness. To identify these significant events, we employed the Fisher exact test to compare the characteristic SNPs in each serial sample. With E value lower than 1, our process yielded 19 significantly changed mutations, including 15 nonsynonymous, 2 intergenic, and 2 synonymous mutations that emerged through genetic hitchhiking (Table 2).

As expected, this analysis identified the emergence of the 3 drug resistance–associated mutations (*katG* D94N, S315R, and *inhA* C-15T) in patient A and 2 *rpoB* mutations (H526Y and L533P) in patient B as significant events. In addition to these, there were 11 nonsynonymous mutations that showed significant changes. Seven of these were found in patient C and 5 had risen to frequencies over 80%. It was surprising to find that the bacterial populations from patient C were so variable compared with their stable resistance phenotype and corresponding mutations conferring drug resistance. Given that no significant mutation was detected in population a1 and the high prevalence of resistant mutants in the other 6 populations, most of these nonresistance-associated mutations occurred with or after the emergence of resistance-conferring mutations. Therefore, these mutations suggest that the *M. tuberculosis* population within the host may be undergoing adaptive evolution to improve the fitness of the bacteria.

Although the relevance between these potential adaptive mutations and drug resistance was not clear, 7 of them showed similarity to genes related to cell wall biosynthesis or membrane proteins. One selected mutation was observed in *fadD32*, an essential gene in mycolic acid biosynthesis [33, 34]. As the *fadD32* E444G emerged after the selection of *katG* D94N, and the mechanism of isoniazid is to inhibit mycolic acid synthesis, this SNP might be a compensatory mutation that improved the fitness of population a3.

DISCUSSION

Using whole-genome sequencing, we have studied the within-host population diversity and evolution of *M. tuberculosis* during treatment for active disease. In all 7 samples, we found a surprising level of genetic diversity to adapt to antibiotic stress that as many as 4 to 5 different resistant mutants were detected in a single sputum sample. These data suggest that heteroresistance might be a common feature of *M. tuberculosis*

Table 2. Significantly Changed Mutations in Each Patient

Patient (Isolate)	Acquired Resistance	Coordinate[a]	Nucleotide Change	Mutation Frequency (%)	Amino Change		Description
					Gene	Codon	
A (a1–a2–a3)	H	209848	GGC - GAC	0 ↗ 79.1 ↗ 98.3	*lprO*	G322D	Possible liproprotein lprO
		1673425	C - T	0 ↗ 20.0 ↘ 0	*inhA*	-15	Isoniazid-resistant mutation
		1689949	T - C	0 ↗ 31.2 ↘ 0	*Rv1498c*	-29	Probable methyltransferase
		2155169	AGC - CGC	0 ↗ 19.5 ↘ 0.6	*katG*	S315R	Isoniazid-resistant mutation
		2155832	GAC - AAC	0 ↗ 56.8 ↗ 93.9	*katG*	D94N	Isoniazid-resistant mutation
		3116294	CGT - CGC	0 → 0 ↗ 42.5	*Rv2811*	R52R	Conserved hypothetical protein
		4261736	GAG - GGG	0 → 0 ↗ 41.3	*fadD32*	E444G	Probable fatty-acid-CoA ligase fadD32; mycolic acid biosynthesis protein
B (b1–b2)	R	589934	CTG - CAG	0 ↗ 74.3	*Rv0499*	L251Q	Diacylglycerol kinase activity
		761139	CAC - TAC	0 ↗ 72.0	*rpoB*	H526Y	Rifampin-resistant mutation
		761161	CTG - CCG	100.0 ↘ 25.5	*rpoB*	L533P	Rifampin-resistant mutation
		1736605	CGT - CGG	36.4 ↗ 71.2	*ileS*	R29R	Isoleucyl-tRNA synthetase ileS
		2176563	GTT - GGT	0 ↗ 69.8	*Rv1924c*	V123G	Predicted integral membrane protein
C (c1–c2)	-	860184	CTC - CCC	0 ↗ 20.0	*Rv0767c*	L176P	Uncharacterized HTH-type transcriptional regulator
		1596190	GCG - GTG	5.0 ↗ 86.0	*Rv1421*	A71V	Displays ATPase and GTPase activities
		1966553	GTG - GCG	41.4 ↗ 89.1	*Rv1739c*	V362A	Probable sulphate-transport transmembrane protein; ABC transporter
		2527483	ATG - ATA	5.4 ↗ 88.1	*Rv2252*	M165I	Diacylglycerol kinase; phosphatidylinositol mannosides (PIMs) biosynthesis
		2711906	CTG - CCG	43.0 ↗ 82.2	*Rv2414c*	L324P	Conserved hypothetical protein
		3460993	CAC - TAC	10.5 ↗ 92.1	*Rv3092c*	H248Y	Conserved integral membrane protein
		4358154	TAC - TAG	24.7 ↘ 0	*Rv3879c*	Y543*	esx-1 secretion-associated protein espK; hypothetical alanine- and proline-rich protein

Abbreviations: -, No change; *, stop codon; H, isoniazid; R, rifampin.

[a] Coordinates are inferred on reference genome *M. tuberculosis* H37Rv.

population with newly acquired resistance. Although multiple discrete resistant mutants transiently coexist in the same population, ultimately only a single type of mutant was selected. The data from patient B imply that the final selection for resistance mutation might be associated with their resistance level and fitness cost. As we saw, during the 18-month treatment, an initially fixed mutation *rpoB* L533P that reportedly associated with low-level rifampin resistance [35, 36], was outcompeted by another candidate that conferred high-level resistance and relatively low fitness cost [4, 5, 12, 32]. The similar phenomenon was also reported on streptomycin resistance where the high-cost mutations *rpsL* K88T or K88R was replaced by *rpsL* K43R, a highly resistant mutation with no fitness cost [6, 7]. The mechanism of selecting high-resistant and low-cost mutants in vivo and subsequent transmission may contribute to the high prevalence of these drug-resistant mutants in reports from clinical isolates [5].

The in vivo selection of low-cost mutations may occur through competition between different populations of siblings with different drug-resistant mutations. The competition between beneficial mutations within a clonal population, or clonal interference, has been studied in asexual populations like viruses [37], *E. coli* [38], and yeast [39], but has been only rarely reported in *M. tuberculosis*. Clonal interference is usually observed in populations with effective population size (N_e) larger than 2×10^4 [38]. Though the in vivo N_e of *M. tuberculosis* is not readily measurable, and may differ with patients depending on the extent of disease, the density of *M. tuberculosis* in sputum may reach 10^5–10^6 colony-forming units (CFU)/mL [40, 41]. Estimates from nonhuman primate model also reported bacterial burden at about 1.4×10^5 CFU per cm^3 of granuloma [42], or 10^7 CFU per animal [43]. Therefore, we speculate that the heterozygous community of *M. tuberculosis* has sufficient N_e to allow for clonal interference during active infection. It is worthwhile to point out that the heterogeneity of *M. tuberculosis* was usually found in different lesions [44, 45], which meant the dynamics of drug-resistant mutants in our study may also be affected by biased sampling from multiple lesions. Under current design, it is difficult to quantitate the impact of sampling bias or other uncertainties such as genetic drift. Therefore, more evidence is required to support the hypothesis of clonal interference in *M. tuberculosis*.

Mathematic model predicts that as the population size increases, clonal interference results in greater fitness improvement [46]. In case of *M. tuberculosis*, it means that low-cost drug-resistant mutations are more likely to be selected from large populations. This conclusion highlights the importance of sensitive diagnostic tests, which allow for early medical intervention when the bacterial burden and diversity remains low. Nowadays, novel diagnostic test such as Xpert *M. tuberculosis*/RIF have decreased the limit of detection to 131 CFU/mL [47], about 100 times more sensitive than conventional sputum smear microscopy. Strategic application of these sensitive diagnostic tools may have the ability to minimize or prevent tuberculosis cases with acquired drug resistance.

Previous studies indicate that resistant strains sometimes acquire additional mutations outside of the affected gene to improve their fitness through compensatory evolution [4, 48]. However, except for mutations in the *ahpC* promoter, which may compensate for the inactivation of *katG* [14], very few compensatory mutations have been identified. Recently, compensatory mutations for restoring fitness in rifampin-resistant strains with lesions in *rpoB* were identified in the genes *rpoA* and *rpoC* [15]. One of the high-probability compensatory mutations (HCMs) *rpoC* V483G was also detected in our samples from patient C. Because this mutation was fixed before our sampling intervals, the other newly evolved mutations in patient C may act as compensatory mutations for other resistant phenotypes. Considering the different mechanisms of antibiotics, it is possible that there are many compensatory changes in different genes to compensate for drug resistance–associated fitness costs [49]. The dynamic heteroresistance in our samples indicate that before the stabilization of drug resistant mutations, the compensatory mutations are being rapidly sampled and the solutions may be very diverse. This might explain the failure to confirm possible compensatory mutations of KZN MDR/XDR strains in other strains with drug resistance [50]. Despite the high diverseness of compensatory mutations, the mutations found in our study and in KZN strains shared several similarities such as cell wall synthesis and the metabolism of mycolic acids [50]. These diverse mutations in certain pathways such as cell wall synthesis are not unexpected given the number of agents used that interfere with this pathway, and any of them might act as compensatory mutations. These mutations will need to be studied in vitro in genetically defined backgrounds to conclusively show they are genuinely compensatory.

By studying the genetic diversity and dynamics of intrapatient *M. tuberculosis* populations, we have just started to understand the battle between antibiotics and *M. tuberculosis*. Clonal interference between different resistant mutations may play a major role in the selection of low-cost mutations, as well as in the development of compensatory changes. These results suggest that the bacillus has a surprisingly high rate of generating genetic diversity and can efficiently sample many different solutions to the challenge of surviving chemotherapy. A better understanding of these dynamics, and their mechanisms, may allow for the development of treatment strategies to avoid the generation and stable fixation of drug resistance in *M. tuberculosis* strains.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online (http://jid.oxfordjournals.org/). Supplementary materials consist of

data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

Financial support. This work was supported by the grant of International Cooperation Project of Ministry of Science and Technology, China (2010DFA34440), and partially supported by the Intramural Research Division of National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH) and NIH grant D43TW007887.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Almeida Da Silva PE, Palomino JC. Molecular basis and mechanisms of drug resistance in *Mycobacterium tuberculosis*: classical and new drugs. J Antimicrob Chemother **2011**; 66:1417–30.
2. Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB. Tuberculosis drug resistance mutation database. PLoS Med **2009**; 6:e1000002.
3. Böttger EC, Springer B, Pletschette M, Sander P. Fitness of antibiotic-resistant microorganisms and compensatory mutations. Nat Med **1998**; 4:1343–4.
4. Gagneux S, Long CD, Small PM, Van T, Schoolnik GK, Bohannan BJ. The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. Science **2006**; 312:1944–6.
5. Böttger EC, Springer B. Tuberculosis: drug resistance, fitness, and strategies for global control. Eur J Pediatr **2008**; 167:141–8.
6. Mariam SH, Werngren J, Aronsson J, Hoffner S, Andersson DI. Dynamics of antibiotic resistant *Mycobacterium tuberculosis* during long-term infection and antibiotic treatment. PLoS ONE **2011**; 6:e21147.
7. Meacci F, Orru G, Iona E, et al. Drug resistance evolution of a *Mycobacterium tuberculosis* strain from a noncompliant patient. J Clin Microbiol **2005**; 43:3114–20.
8. Rinder H, Mieskes KT, Loscher T. Heteroresistance in *Mycobacterium tuberculosis*. Int J Tuberc Lung Dis **2001**; 5:339–45.
9. Chakravorty S, Aladegbami B, Thoms K, et al. Rapid detection of fluoroquinolone-resistant and heteroresistant *Mycobacterium tuberculosis* by use of sloppy molecular beacons and dual melting-temperature codes in a real-time PCR assay. J Clin Microbiol **2011**; 49:932–40.
10. Hillemann D, Rusch-Gerdes S, Richter E. Feasibility of the GenoType MTBDRsl assay for fluoroquinolone, amikacin-capreomycin, and ethambutol resistance testing of *Mycobacterium tuberculosis* strains and clinical specimens. J Clin Microbiol **2009**; 47:1767–72.
11. Post FA, Willcox PA, Mathema B, et al. Genetic polymorphism in *Mycobacterium tuberculosis* isolates from patients with chronic multidrug-resistant tuberculosis. J Infect Dis **2004**; 190:99.
12. Sander P, Springer B, Prammananan T, et al. Fitness cost of chromosomal drug resistance-conferring mutations. Antimicrob Agents Chemother **2002**; 46:1204–11.
13. Shcherbakov D, Akbergenov R, Matt T, Sander P, Andersson DI, Böttger EC. Directed mutagenesis of *Mycobacterium smegmatis* 16S rRNA to reconstruct the in-vivo evolution of aminoglycoside resistance in *Mycobacterium tuberculosis*. Mol Microbiol **2010**; 77:830–40.
14. Sherman DR, Mdluli K, Hickey MJ, et al. Compensatory *ahpC* gene expression in isoniazid-resistant *Mycobacterium tuberculosis*. Science **1996**; 272:1641–3.
15. Comas I, Borrell S, Roetzer A, et al. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. Nat Genet **2011**; 44:106–10.
16. Brockhurst MA, Colegrave N, Rozen DE. Next-generation sequencing as a tool to study microbial evolution. Mol Ecol **2011**; 20:972–80.
17. Barrick JE, Lenski RE. Genome-wide mutational diversity in an evolving population of *Escherichia coli*. Cold Spring Harb Symp Quant Biol **2009**; 74:119–29.
18. Zagordi O, Klein R, Däumer M, Beerenwinkel N. Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. Nucleic Acids Res **2010**; 38:7400–9.
19. Bull RA, Luciani F, McElroy K, et al. Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. PLoS Pathog **2011**; 7:e1002243.
20. Ioerger TR, Feng Y, Chen X, et al. The non-clonality of drug resistance in Beijing—genotype isolates of *Mycobacterium tuberculosis* from the Western Cape of South Africa. BMC Genomics **2010**; 11:670.
21. Ioerger TR, Koo S, No EG, et al. Genome analysis of multi- and extensively-drug-resistant tuberculosis from KwaZulu-Natal, South Africa. PLoS ONE **2009**; 4:e7778.
22. Schurch AC, Kremer K, Kiers A, et al. The tempo and mode of molecular evolution of *Mycobacterium tuberculosis* at patient-to-patient scale. Infect Genet Evol **2010**; 10:108–14.
23. Saunders NJ, Trivedi UH, Thomson ML, Doig C, Laurenson IF, Blaxter ML. Deep resequencing of serial sputum isolates of *Mycobacterium tuberculosis* during therapeutic failure due to poor compliance reveals stepwise mutation of key resistance genes on an otherwise stable genetic background. J Infect **2011**; 62:212–7.
24. Zhang L, Chen J, Shen X, et al. Highly polymorphic variable-number tandem repeats loci for differentiating Beijing genotype strains of *Mycobacterium tuberculosis* in Shanghai, China. FEMS Microbiol Lett **2008**; 282:22–31.
25. Larsen MH, Biermann K, Tandberg S, Hsu T, Jacobs WR Jr. Genetic manipulation of *Mycobacterium tuberculosis*. Curr Protoc Microbiol **2007**; 6:10A.2.
26. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics **2009**; 25:1754–60.
27. Cole ST, Brosch R, Parkhill J, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature **1998**; 393:537–44.
28. Zhang Y, Chen C, Liu J, et al. Complete genome sequences of *Mycobacterium tuberculosis* strains CCDC5079 and CCDC5080, which belong to the Beijing family. J Bacteriol **2011**; 193:5591–2.
29. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet **2011**; 43:491–8.
30. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics **2009**; 25:2078–9.
31. Gagneux S, DeRiemer K, Van T, et al. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. Proc Natl Acad Sci USA **2006**; 103:2869–73.
32. Ohno H, Koga H, Kohno S, Tashiro T, Hara K. Relationship between rifampin MICs for and *rpoB* mutations of *Mycobacterium tuberculosis* strains isolated in Japan. Antimicrob Agents Chemother **1996**; 40:1053–6.
33. Leger M, Gavalda S, Guillet V, et al. The dual function of the *Mycobacterium tuberculosis* FadD32 required for mycolic acid biosynthesis. Chem Biol **2009**; 16:510–9.
34. Trivedi OA, Arora P, Sridharan V, Tickoo R, Mohanty D, Gokhale RS. Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria. Nature **2004**; 428:441–5.
35. Jenkins C, Bacon J, Allnutt J, et al. Enhanced heterogeneity of *rpoB* in *Mycobacterium tuberculosis* found at low pH. J Antimicrob Chemother **2009**; 63:1118–20.
36. Luo T, Zhao M, Li X, et al. Selection of mutations to detect multidrug-resistant *Mycobacterium tuberculosis* strains in Shanghai, China. Antimicrob Agents Chemother **2010**; 54:1075–81.
37. Miralles R, Gerrish PJ, Moya A, Elena SF. Clonal interference and the evolution of RNA viruses. Science **1999**; 285:1745.

38. Perfeito L, Fernandes L, Mota C, Gordo I. Adaptive mutations in bacteria: high rate and small effects. Science **2007**; 317:813.
39. Kao KC, Sherlock G. Molecular characterization of clonal interference during adaptive evolution in asexual populations of *Saccharomyces cerevisiae*. Nat Genet **2008**; 40:1499–504.
40. Bark C, Okwera A, Joloba M, et al. Time to detection of *Mycobacterium tuberculosis* as an alternative to quantitative cultures. Tuberculosis **2011**; 91:257–9.
41. Joloba M, Johnson J, Namale A, et al. Quantitative sputum bacillary load during rifampin-containing short course chemotherapy in human immunodeficiency virus-infected and non-infected adults with pulmonary tuberculosis. Int J Tuberc Lung Dis **2000**; 4:528–36.
42. Luciw PA, Oslund KL, Yang XW, et al. Stereological analysis of bacterial load and lung lesions in nonhuman primates (*rhesus macaques*) experimentally infected with *Mycobacterium tuberculosis*. Am J Physiol Lung Cell Mol Physiol **2011**; 301:L731–8.
43. Lin PL, Rodgers M, Smith L, et al. Quantitative comparison of active and latent tuberculosis in the cynomolgus macaque model. Infect Immun **2009**; 77:4631–42.
44. Ford CB, Lin PL, Chase MR, et al. Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. Nat Genet **2011**; 43:482–6.
45. Kaplan G, Post FA, Moreira AL, et al. *Mycobacterium tuberculosis* growth at the cavity surface: a microenvironment with failed immunity. Infect Immun **2003**; 71:7099–108.
46. Gerrish PJ, Lenski RE. The fate of competing beneficial mutations in an asexual population. Genetica **1998**; 102:127–44.
47. Blakemore R, Story E, Helb D, et al. Evaluation of the analytical performance of the Xpert MTB/RIF assay. J Clin Microbiol **2010**; 48:2495.
48. Cohen T, Sommers B, Murray M. The effect of drug resistance on the fitness of *Mycobacterium tuberculosis*. Lancet Infect Dis **2003**; 3:13–21.
49. Maisnier-Patin S, Andersson DI. Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. Res Microbiol **2004**; 155:360–9.
50. Motiwala AS, Dai Y, Jones-Lopez EC, et al. Mutations in extensively drug-resistant *Mycobacterium tuberculosis* that do not code for known drug-resistance mechanisms. J Infect Dis **2010**; 201:881–8.