DRUG DISCOVERY AND RESISTANCE

Longitudinal whole genome analysis of pre and post drug treatment Mycobacterium tuberculosis isolates reveals progressive steps to drug resistance

Gargi Datta a, b, Luisa M. Nieto c, Rebecca M. Davidson a, Carolina Mehaffy c, Caroline Pederson d, Karen M. Dobos c, Michael Strong a, b, *

a Center for Genes, Environment and Health, National Jewish Health, Denver, CO 80206, USA
b Computational Bioscience Program, University of Colorado, Denver, School of Medicine, Aurora, CO 80045, USA
c Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, CO 80523, USA
d ATCC, 10801 University Boulevard, Manassas, VA 20110, USA

ARTICLE INFO

Article history:
Received 30 October 2015
Received in revised form 15 February 2016
Accepted 20 February 2016

SUMMARY

Tuberculosis (TB) is one of the leading causes of death due to an infectious disease in the world. Understanding the mechanisms of drug resistance has become pivotal in the detection and treatment of newly emerging resistant TB cases. We have analyzed three pairs of Mycobacterium tuberculosis strains pre- and post-drug treatment to identify mutations involved in the progression of resistance to the drugs rifampicin and isoniazid. In the rifampicin resistant strain, we confirmed a mutation in rpoB (S450L) that is known to confer resistance to rifampicin. We discovered a novel L101R mutation in the katG gene of an isoniazid resistant strain, which may directly contribute to isoniazid resistance due to the proximity of the mutation to the katG isoniazid-activating site. Another isoniazid resistant strain had a rare mutation in the start codon of katG. We also identified a number of mutations in each longitudinal pair, such as toxin–antitoxin mutations that may influence the progression towards resistance or may play a role in compensatory fitness. These findings improve our knowledge of drug resistance progression during therapy and provide a methodology to monitor longitudinal strains using whole genome sequencing, polymorphism comparison, and functional annotation.

* 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Along with HIV, Tuberculosis (TB) is the leading causes of death due to an infectious disease, with an estimated 1.5 million deaths per year [1]. In 2014, the disease infected an estimated 9.6 million people around the world, with about 3.6 million of those people not getting proper treatment [1]. In recent years, inappropriate applications of frontline antibiotic therapies have created an upsurge in the emergence of multiple drug-resistant (MDR) and extensively drug-resistant (XDR) strains of Mycobacterium tuberculosis (Mtbc) [2]. MDR TB is defined as resistance to Isoniazid (INH) and Rifampicin (RIF), and XDR TB is resistant to INH, RIF, any Fluoroquinolone, and any of the three second-line injectables (Amikacin, Capreomycin or Kanamycin) [1]. In 2014, there were an estimated 480,000 MDR TB cases and the World Health Organization received reports of 123,000 of these cases [1]. An estimated 9.7% of MDR TB cases were XDR instances [1]. The current treatment initiatives for MDR TB involve a combination therapy of 8–10 drugs for 18–24 months with serious side effects and result in nearly 30% of the patients experiencing treatment failure [3]. RIF and INH are the two main first line drugs used to treat TB and there has been increasing emergence of strains resistant to both [4]. INH is a bactericide that inhibits dividing bacilli and is used in short-course treatment regimens in TB [5]. Mutations in the gene katG that prohibit activation of INH are the primary cause of resistance to the drug in TB [5]. Mutations in the β subunit of RNA polymerase of Mtbc (rpoB) inhibit binding of RIF to the RNA polymerase, thus introducing resistance [4]. RpoB mutations account for over 95% of strains with RIF resistance [4].

http://dx.doi.org/10.1016/j.tube.2016.02.004
1472-9792/© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Drug resistance in TB is primarily due to inappropriate treatment of initially drug-sensitive disease or the acquisition of already resistant strains [5]. Acquired drug resistance happens during treatment, and is typically a result of single nucleotide polymorphisms (SNPs) in genes that encode target proteins or drug activating enzymes [4,6]. After certain mutation-conferring resistance events occur, Mtb strains are sometimes able to optimize fitness and evolve mutations that confer the highest drug resistance with the lowest fitness cost [7]. Understanding the mechanisms of drug resistance has become important in correctly diagnosing and treating MDR TB. Clinical drug-resistant TB is primarily diagnosed by phenotypic drug-susceptibility testing on slowly-growing Mtb cultures [2]. This process can take several weeks, and the lag can result in improper treatment leading to higher mortality and transmission rates of drug-resistant strains [2]. Current genotypic susceptibility testing techniques for Mtb can only elucidate resistance profiles based on known mutations [8]. Therefore, there is a need to develop efficient strategies to elucidate the mechanisms of drug resistance and generate faster resistance profiles for novel strains using molecular means [2,9].

Whole genome sequencing has been used to identify novel mutations for resistance in bacteria and to elucidate mutation rates in TB [6,10–14]. In this paper, we describe the genomic analyses of three pairs of pre- and post-drug treatment strains of Mtb. These strain pairs are representative of drug resistance evolution in patients due to drug therapy. Two pairs evolved resistance to INH, and the third evolved resistance to RIF. We identify activating enzymes [4,6]. After certain mutation-conferring resistance events with the lowest fitness cost [7]. Understanding the mechanisms of drug resistance has become important in correctly diagnosing and treating MDR TB. Clinical drug-resistant TB is primarily diagnosed by phenotypic drug-susceptibility testing on slowly-growing Mtb cultures [2]. This process can take several weeks, and the lag can result in improper treatment leading to higher mortality and transmission rates of drug-resistant strains [2]. Current genotypic susceptibility testing techniques for Mtb can only elucidate resistance profiles based on known mutations [8]. Therefore, there is a need to develop efficient strategies to elucidate the mechanisms of drug resistance and generate faster resistance profiles for novel strains using molecular means [2,9].

Whole genome sequencing has been used to identify novel mutations for resistance in bacteria and to elucidate mutation rates in TB [6,10–14]. In this paper, we describe the genomic analyses of three pairs of pre- and post-drug treatment strains of Mtb. These strain pairs are representative of drug resistance evolution in patients due to drug therapy. Two pairs evolved resistance to INH, and the third evolved resistance to RIF. We identify mutations in rpoB and katG that likely explain the acquisition of resistance in clinical strains of Mtb. Identification of these polymorphisms, and the adoption of similar genomic analysis pipelines, will aid in the diagnosis and monitoring of drug resistance in new Mtb strains, thus improving treatment of TB. We also identify mutations in genes that may contribute to the fitness or drug tolerance of the strains.

2. Materials and methods

2.1. Mtb strains

All Mtb strains were obtained from patients being treated for active pulmonary TB. Isolates were sent to National Jewish Health, Denver, CO, for confirmation of drug susceptibility data initially obtained by the patient’s primary clinical lab. Clinical isolates C26-RIFs and C22-RIFr of the Beijing genetic subgroup representing one pan susceptible (C26-RIFs) and one RIF resistant (C22-RIFr) isolate were used for this study (details regarding this pair were reported previously) (Supp. table 1) [4]. Clinical isolate C4-INHs presented susceptibility to all drugs tested (RIF, INH, pyrazinamide, and ethambutol) and was obtained from a patient living in Spain prior to initiating anti-tubercular therapy (Supp. table 1). Clinical isolate C1-INHr was later obtained from the same patient after several weeks on frontline antibiotic treatment and was confirmed resistant to INH (tested at 0.2 and 1.0 μg/ml) by the Agar proportion test, but remained susceptible to the other three first line agents (Supp. table 1). Spoligotyping and insertion sequence (IS)6110 restriction fragment length polymorphism (RFLP) pattern confirmed C21-INHs and C27-INHr as a parent–mutant pair belonging to the T genetic subgroup [15]. Clinical isolate C21-INHs also presented susceptibility to all drugs tested (RIF, INH, pyrazinamide, and ethambutol) and was obtained from a patient in San Francisco prior to initiating standard-first line anti-tubercular therapy (Supp. table 1). After several weeks on therapy, Clinical isolate C27-INHr was obtained from the same patient and confirmed resistant to INH by Agar proportion test (tested as above), but remained susceptible to the other three first-line agents (Supp. table 1). Spoligotyping and IS6110 RFLP confirmed C21-INHs and C27-INHr as a parent–mutant pair belonging to the T genetic subgroup [15].

2.2. Preparation of isolates for WGS

Mtb clinical isolates were plated onto 7H11 + OADC media and incubated for 14–28 days at 37 °C; growth on plates was monitored until several isolated colonies appeared. One colony from each strain was selected and inoculated into 100 mL of 7H9 + OADC + 0.1% Tween-80 and allowed to grow for 14–21 days at 37 °C with shaking; from which 3 new 100 mL cultures for each strain were inoculated at an OD600 of 0.05 into 7H9 + OADC + 0.1% Tween-80. Cultures were allowed to grow at 37 °C with shaking until harvested by centrifugation when the OD reads between 0.4 and 0.6 (~48 h). Cell pellets were washed three times with PBS and processed to isolate genomic DNA; all samples were stored at −80 °C or kept on ice [16]. Library preparation was done according to the manufacturer’s protocol and fragments were size selected between 50 and 75bp. Genome sequencing was performed using the SOLID 3 (Life Technologies Corporation, Carlsbad, CA) platform [17].

2.3. katG PCR amplification and sequencing

PCR amplification for Sanger sequence confirmation of the katG gene was devised at Colorado State University and utilized a reaction mix that included forward and reverse primers for katG (125 ng):

Forward 5’ CGA TTT TCG GTG CCG TGC GTT TT 3’
Reverse 5’ CGC CCA GCA ATG CAT GAG CAT TAT 3’

Takara Ex Taq™ Hot Start DNA polymerase (1.25 U) (TAKARA BIO INC), a mixture of deoxynucleoside triphosphates (dNTP) (200 μM each), betaine (1 M) and Mtb gDNA (250 ng).

An initial denaturation step at 94 °C for 5 min followed by a total of 40 cycles of amplification were performed, with template denaturation at 94 °C for 10 s, primer annealing at 59.5 °C for 30 s and extension at 72 °C for 3 min. Then a final extension at 72 °C was done for 6 min. After visualization of the PCR product on an agarose gel (1.5%), post-PCR cleanup and Sanger sequencing occurred at the genomic core of the Proteomics and Metabolomics facilities at Colorado State University.

2.4. Single nucleotide polymorphism detection and annotation

The Lifescope Genome Analysis Software whole genome resequencing pipeline was used with default parameters to map fragmented reads for each Mtb strain to the H37Rv reference genome (Supp. table 1) [18,19]. SNPs and Indels were identified with the mpileup program in samtools version 0.1.18 in conjunction with bcftools version 0.1.17-dev [19–21]. The following parameters filtered SNPs and Indels:

1. Minimum SNP mapping quality score >20
2. Minimum of 20× high quality read depth

These filtering criteria are a standard used in numerous WGS analyses [13,12,6,11,19]. Only SNPs passing the filtering criteria were used for downstream analysis. SNPs and Indels were annotated as intergenic or genic and for amino acid level changes using the ANNOVAR functional gene annotation software [19,22]. Repeat regions (PE/PPE and PE-PGRS gene families) that might cause incorrect read alignments were excluded from further analysis. A fully automated version of this pipeline is available at https://github.com/dattagargi/MutationAnalysisPipeline.
2.5. Drug-resistance conferring SNP analysis

A custom Perl script was used to annotate SNPs in genes associated with Mtb drug resistance and the antibiotic(s) that each polymorphism confers resistance to using data mined from the Tuberculosis Drug Resistance Mutation Database (TBDr eaMDB) [23]. SNPs commonly shared with susceptible strains were considered naturally occurring and non-causal for resistance. SNPs identified in known antibiotic resistance determining genes in phenotypically resistant Mtb strains and not shared with susceptible strains were investigated further to find drug-resistance conferring SNPs. This required mining the TBDr eaMDB and the TBDB databases for resistance-associated SNPs [23–25]. The drug resistance profiles predicted by the resistance-associated SNP genotypes were compared with the phenotypic resistance profiles.

2.6. Comparison of SNPs pre and post drug treatment

We compared all annotated SNPs between the pairs of strains pre and post drug treatment (C26-RIFs vs. C22-RIFr, C4-INHs vs. C1-INHr, and C21-INHs vs. C27-INHr). We identified SNPs in the post-drug treatment resistant strains (C22-RIFr, C1-INHr, and C27-INHr) that were not in their corresponding pre-drug treatment susceptible strains (C26-RIFs, C4-INHs, and C21-INHs respectively). SNP calls in the drug resistant strain that were reference calls in the susceptible strain were considered high quality.

3. Results

The six Mtb strains were grouped according to their drug resistance profile and genotype (Table 1).

3.1. C26-RIFs versus C22-RIFr

There were 48 SNPs in C22-RIFr that were not present in C26-RIFs. Of these, there were 5 base changes causing non-synonymous mutations between the paired isolates (Table 2).

Of the 5 SNPs, one is in rpoB (Rv0667). The SNP, C1349T, resulted in an amino acid change from serine to leucine (Supp. table 3). This polymorphism is not mentioned in TBDreaMDB, but is present in TBDb as a known resistance conferring SNP in H37Rv [4,26,27].

We did a literature survey of the 4 remaining SNPs, and we found a mutation in the antitoxin encoding gene vapB42 [28].

3.2. C4-INHs versus C1-INHr

There were 55 SNPs in C1-INHr that were not present in C4-SUS. Of the 55 SNPs, 8 were base changes causing non-synonymous mutations between the two isolates (Table 2).

Of these 8 SNPs, one is in katG (Rv1908), a gene known to confer resistance to INH [5]. The SNP is at nucleotide position 302 (T302G) causing a change from leucine to arginine (Supp. table 2), and was confirmed using Sanger sequencing. This katG mutation was not found in any strain in TBDB, thus making it a novel mutation.

We did a literature survey of the 7 remaining SNPs, we also found a SNP in mazF5, a toxin belonging to the mazEF family [28].

3.3. C21-INHs versus C27-INHr

There were 17 SNPs in C27-INHr that were not present in C21-INHs. Of the 17 SNPs, there were 14 SNPs causing non-synonymous mutations between the two isolates (Table 2).

Of these 14 SNPs, one is in katG (Rv1908). The SNP, T2C, yields a disruption of the native katG GTG start codon (Supp. table 4), and was confirmed using Sanger sequencing. This mutation is not found in any strain in TBDB, but it was reported in an MDR strain from Russia in 1995 [29].

4. Discussion

We analyzed drug resistance progression to INH and RIF for three pairs of Mtb (Table 1). Of the pair that gained resistance to RIF (C26-RIFs, C22-RIFr), C22-RIFr was the post-treatment resistant strain. Among the unique SNPs identified in strain C22-RIFr (Supp. table 3), we found a mutation at nucleotide position 1349 (S450L) in rpoB, the known target gene for RIF [4,26]. There are several known drug resistance conferring polymorphisms in this gene, including T350I, S388L, S450L, D516F/Y, S531L and E639D [24,25,30–32]. Of these, we found S450L in our drug resistance progression analysis (Figure 1). Since it is a known mutation for drug resistance to RIF, S450L acted as a positive control and validated our genomic pipeline and analysis. Figure 1 depicts a model of the RIF binding site in rpoB, and the position of the S450L mutation with respect to the binding site [26,33].

To understand progression of resistance to INH in Mtb, we analyzed SNPs found in two isolate pairs. In most clinical isolates of Mtb, resistance is due to mutations in katG that may alter mycobacterial catalase–peroxidase activity and at the same time block prodrug activation [5]. Mtb katG is a homodimer of identical 82-kDa subunits [34]. The N-terminal domain of katG contains a functional heme-binding site (Figure 2) [35]. Studies have shown that the dimerization of katG is largely mediated by the first 100 amino acids of the N-terminal domain [35]. There is a lack of information in literature about the actual binding site of INH to katG [36]. Some studies have shown that INH can bind in the heme pocket of the

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Number of SNPs</th>
<th>Number of non-synonymous SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>C26-RIFs versus C22-RIFr</td>
<td>48</td>
<td>5 (10.4%)</td>
</tr>
<tr>
<td>C4-INHs versus C1-INHr</td>
<td>55</td>
<td>8 (14.5%)</td>
</tr>
<tr>
<td>C21-INHs versus C27-INHr</td>
<td>17</td>
<td>14 (82.4%)</td>
</tr>
</tbody>
</table>

Table 1: Details of the six Mtb strains.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Description of drug susceptibility</th>
<th>Mtb genotype</th>
<th>Origin</th>
<th>Pre or post drug treatment</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C26-RIFs</td>
<td>RIFs</td>
<td>Beijing</td>
<td>Costa Rica</td>
<td>Pre</td>
<td>Bisseur et al. [4]</td>
</tr>
<tr>
<td>C22-RIFr</td>
<td>RIFr</td>
<td>Beijing</td>
<td>Costa Rica</td>
<td>Post</td>
<td>Bisseur et al. [4]</td>
</tr>
<tr>
<td>C4-INHs</td>
<td>INHs</td>
<td>Beijing</td>
<td>Spain</td>
<td>Pre</td>
<td>Unpublished</td>
</tr>
<tr>
<td>C1-INHr</td>
<td>INHr</td>
<td>Beijing</td>
<td>Spain</td>
<td>Post</td>
<td>Unpublished</td>
</tr>
<tr>
<td>C21-INHs</td>
<td>INHs</td>
<td>T</td>
<td>USA</td>
<td>Pre</td>
<td>Unpublished</td>
</tr>
<tr>
<td>C27-INHr</td>
<td>INHr</td>
<td>T</td>
<td>USA</td>
<td>Post</td>
<td>Unpublished</td>
</tr>
</tbody>
</table>
katG dimer, and at different binding sites along the heme access channel [36]. In the first isolate pair (C4-INHs, C1-INHr), C4-INHs was the pre-treatment strain susceptible to INH and C1-INHr was the strain post treatment that developed resistance to the drug. The katG mutation found between these two isolates, T302G (L101R), has not been found in previous INH resistance studies, and could contribute to INH drug resistance observed in the patient isolate. This mutation is in the second N-terminal domain of katG [35,37]. This SNP causes a hydrophobic Leucine to mutate to a hydrophilic Arginine. This amino acid change may change the binding energy or protein conformation near the active site, preventing INH from binding. Subsequent katG analysis and phenotypic evaluation demonstrated ambiguous susceptibility results for C1-INHr, possibly suggesting reemergence of the susceptible clone. Figure 2 is a model of the heme-binding site of katG, and the location of the mutation L101R with respect to the heme-binding site [35,37,38].

From the C21-INHs (susceptible) and C27-INHr (post-treatment INH resistant) pair, we identified a mutation at nucleotide position T2C (f-Met1A) in katG in C27-INHr (Supp. table 4). This mutation (GTG to GCG) is most likely responsible for conferring resistance to INH in C27-INHr as it is located at the start codon position of katG, likely disrupting translation initiation at this position. This region is also responsible for dimerization, as shown by Wilming et al. [35]. Previously, Heym et al. reported this mutation in an INH resistant TB strain, and provided evidence that the disruption of the native start codon may result in the use of an alternate upstream start codon, yielding a larger protein, but without native peroxidase activity [29].

Among our findings related to acquired antibiotic resistance, we also found polymorphisms potentially related to Mtb virulence. For example, we identified a mutation in the gene encoding for the toxin mazF5 in C1-INHr (INH-resistant) (Supp. table 2) and in the gene encoding for the antitoxin vupB42 in C22-RIFr (RIF-resistant) (Supp. table 3). Toxin–antitoxin (TA) systems are small genetic modules composed of a protein toxin and a protein antitoxin or a non-coding RNA as an antagonistic antitoxin [39]. Sala et al. and Georgiades et al. have both hypothesized that the persistence induced by active toxins could contribute to the pathogenesis of Mtb, which has been shown to have at least 79 TA systems [28,40]. The mechanisms of these systems as virulence factors have not been fully elucidated [40].
4.1. Conclusion

We utilized comparative genomics to analyze pre- and post-drug treatment paired isolates of *Mtb* and identified both known and novel mutations thought to be responsible for progression of drug resistance. We found a known mutation in *rpoB* that is responsible for acquisition of resistance to Rif, as well as a known but rare mutation in katG affecting the initiation codon. For another strain that gained resistance to INH, we found a novel mutation affecting the N-terminus of KatG protein.

The identification of a candidate mutation involved in drug resistance will give further insight into acquired drug resistance mechanisms to INH. This analysis also improves our understanding of drug resistance acquisition and progression in TB, and paves the way for improving diagnostic monitoring for drug-resistant TB. Finally, this study shows the power of whole genome sequencing and comparative analysis in identifying possible mutations that could be supporting the drug resistant phenotype or affecting fitness.

Acknowledgments

We would like to thank Drs. Marcus Burgos, Philip Knabenbauer and Melissa Karres for their contribution in the clinical strain collection and genomic preparation.

Funding: G.D., R.D. and M.S. were supported by a Colorado Bioscience Grant and the Boettcher Foundation Webb Waring grant. L.N., K.D., C.M. and C.P. were supported by the Colombian Administrative Department of Science, Technology and Innovation COLCIENCIAS with the Francisco Jose de Caldas-512 Scholarship, and the American Type Culture Collection (ATCC) fund #2010-0516-0005.

Competing interests: We declare that we have no conflict of interest.

Ethical approval: Not required.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.tube.2016.02.004.

References


