SKW Tsui 徐國榮
RCY Chan 陳超揚
KM Kam 甘啟文
HS Kwan 關海山
KS Leung 梁廣錫
KP Fung 馮國培
MMY Waye 韋妙宜
PTW Law 羅迪雲
SK Lou 婁紹科
KK Leung 梁家傑

Genome analysis of *Mycobacterium tuberculosis* Beijing family strains

Key Messages

1. The genomes of the three 'totally drug-resistant' (TDR) Beijing/W strains of *Mycobacterium tuberculosis* (MTB) were sequenced and compared with publicly available genome sequences of KwaZulu-Natal MTB.
2. The sequences of the KwaZulu-Natal MTB strains shared a close ancestral relationship with MTB F11, whereas the three Beijing/W isolates formed a separate cluster.
3. The numbers of deletions, truncations, and frame-shift mutations were significantly greater in TDR Beijing/W strains.
4. Some DNA repair genes were defective in TDR Beijing/W strains.
5. Many genes involved in optimal mycobacterial growth were mutated in TDR Beijing/W strains.

Hong Kong Med J 2013;19(Suppl 5):S12-4

The Chinese University of Hong Kong:
School of Biomedical Sciences
SKW Tsui, KP Fung, MMY Waye
Hong Kong Bioinformatics Centre
SKW Tsui, SK Lou, KK Leung
Centre for Microbial Genomics and Proteomics
SKW Tsui
Department of Microbiology
RCY Chan
School of Life Sciences
HS Kwan
Department of Computer Science and Engineering
KS Leung
Faculty of Science
PTW Law
Public Health Laboratory Centre, Hospital Authority
KM Kam

RFCID project number: 08070502

Principal applicant and corresponding author:
Prof Stephen Kwok Wing Tsui
School of Biomedical Sciences, The Chinese University of Hong Kong, Shatin, NT, Hong Kong SAR, China
Tel: (852) 2609 6381
Fax: (852) 2603 7732
Email: kwtsui@cuhk.edu.hk

Introduction

Mycobacterium tuberculosis (MTB) currently affects more than two billion people and causes 1.5 to 2 million deaths every year. Multidrug-resistant and the virtually untreatable extensively drug-resistant MTB strains are estimated to cause 490 000 and 40 000 new cases per year, respectively. A Beijing/W subtype has attracted attention for its global emergence, increased transmissibility, and tendency to develop multidrug resistance. In the Beijing/W strain, alterations in DNA repair genes, *mutT2*, *mutT4*, and *ogt*, result in increased mutation frequencies and better adaptability to stress. Compared with *Mycobacterium bovis* and H37Rv, the Beijing/W subtype has an intact open reading frame of *pks15/1*, which is involved in the biosynthesis of phenolic glycolipids and may contribute to its hypervirulence. The Beijing/W subtype has three unique classes of large sequence polymorphisms (LSPs): four LSPs (RD105, RD149, RD152, and RD207) that are deleted; three LSPs (RD142, RD150, and RD181) that are variably deleted; and 14 LSPs that are encountered in individual isolates.

To understand the genetic factors responsible for the drug-resistant phenotype in the Beijing/W strains, we sequenced the genome of three isolates of MTB Beijing/W subtype, one of which was drug-sensitive and the other two were drug-resistant. The latter were resistant to nearly all commonly used anti-TB drugs, which we termed a 'totally drug-resistant' (TDR) phenotype. We compared the genome sequences of these three isolates with known MTB complete genome sequences having drug-resistant profiles.

Methods

This study was conducted from September 2008 to August 2010. The genomes of the TDR strains, BT1 and BT2, and the drug-sensitive strain BS1 were sequenced using the Roche 454 GS FLX pyrosequencing system. Reference assembly was performed separately using the manufacturer-provided software (GS Mapper) as well as BWA. De novo assembly was also performed using Velvet. The two sets of scaffolds were then compared and combined to form a first draft, using our in-house scripts. Gaps closure was performed using the Sanger approach. Whole-genome sequence comparison and alignment were done in Mauve and Mummer.

Genome sequences were first annotated using an automated subsystem approach. Results were then carefully crosschecked with GLIMMER as well as search outputs from a non-redundant nucleotide database. Additional ontological information was obtained from KEGG. Phylogenetic analysis was conducted by comparing the four Beijing/W isolates to a set of 113 geographically diverse MTB isolates using polymorphic loci found within a set of 89 housekeeping genes. MEGA4 was used to generate a NJ tree.

For SNP analysis, the reference strains used were: H37Rv (RL123456.2), H37Ra (NC_00952.1), KZN4207 (NZ_ACVS00000000), KZN1435 (NC_012943), and KZN605 (NZ_ABGN0000000).[1]

The inter-relationship of mutated genes was studied by submitting the gene list to the DAVID Gene Functional Classification Tool, and also by the STRING software for the prediction of protein-protein interaction network.

Results

The three MTB Beijing/W isolates were obtained from the Tuberculosis Reference Laboratory at the Public Health Laboratory Centre of the Department of Health, Hong Kong Special Administrative Region. They were chosen based on: (1) their drug-resistance phenotypes, and (2) individual differences in RFLP patterns to maximise heterogeneity in genomic variations for comparison. The drug-sensitive strain, BS1, was sensitive to streptomycin, isoniazid, rifampicin, and ethambutol, whereas the two TDR strains, BT1 and BT2, were resistant to these antimicrobials as well as ethionamide, kanamycin, capreomycin, ofloxacin, amikacin, and pyrazinamide. The genomic DNA of the three isolates was extracted, purified, and subjected to high-throughput sequencing using the Roche 454 approach. The sequences of all the strains were then gap-filled and sequenced to completion. The three genomes were fully annotated and their corresponding GenBank files were submitted to the National Center for Biotechnology Information, National Institute of Health, USA. Whole-genome sequence alignment was performed. The genomic sequences of many KwaZulu-Natal (KZN) strains were publicly available. The phylogenetic relationships between the three Beijing/W isolates and the KZN strains, along with all the known MTB genomes were compared. The three KZN strains shared a close ancestral relationship with MTB F11, whereas the three Beijing/W isolates formed a separate cluster.

To identify novel genetic factors associated with the drug-resistance phenotype in the Beijing/W subtype, we identified genes that were frame-shifted, truncated, or lost, as well as non-synonymous SNPs that were specific in drug-resistant strains but not drug-sensitive strains. To reduce the number of false positives, we adopted more stringent criteria by excluding the hypermutable PPE proteins, PE-PGRS proteins, repetitive proteins, hypothetical proteins, proteins with synonymous substitution and mutations in promoter and intergenic regions. The numbers of mutations in Beijing/W drug-resistant strains were significantly greater than in other drug-resistant strains. Many of the mutations, such as those within haemolysin A, the TetR family transcriptional repressor, DNA gyrase subunit A, catalase-peroxidase and DNA-directed RNA polymerase subunit β, were present in our Beijing/W TDR isolates. However, mutations in these genes may differ in terms of amino acid changes and codon numbers when compared with reported mutation hot spots. Surprisingly, the number of deletions, truncations, and frame-shift mutations were significantly greater in TDR strains.

There were 15 genes that were common in the two BT strains. The overlap was small, which reflected the stochastic nature of the emergence of drug resistance or non-clonality (owing to the increased fitness cost associated with drug-resistance mutations). Among the commonly mutated genes in the two strains, only seven were associated with drug resistance.

To explore the relationship of mutated genes, the gene list was submitted to the DAVID Gene Functional Classification Tool.[2] Significant enrichment of the mismatch repair, geraniol degradation, and nicotinate and nicotinamide metabolism pathways was identified. As none of the mutator phenotypes (mutT1, mutT2, and mutT3) in the Beijing/W strains had changes, other genes involved in the mismatch repair of MTB were examined.[3] Mutations were found in three DNA repair-related genes in BT1 and one in BT2. We hypothesised that the two BT strains weakened their mismatch repair machinery to enable more aggressive destructive evolution in order to survive the inhibition caused by a large number of anti-TB drugs.

Genes required for mycobacterial growth are defined by high-density mutagenesis.[4] A total of 614 genes were considered essential for optimal growth in MTB and *M bovis*. Among genes having a non-synonymous mutation in Beijing/W and KZN families, only one in each KZN strain and 12 and 10 in BT1 and BT2, respectively, were identified. The increased mutation rate of genes for optimal growth implies that MTB strains with extreme drug resistance tend to slow down their growth in order to survive. This is in line with the hypothesis that TDR MTB strains are rare because of the high fitness cost of developing drug resistance.

The STRING software was used to study the inter-relationship of genes with non-synonymous mutations.[5] A cluster of lipid metabolism-associated genes were found in the PPI network, and many novel gene products interacting with known drug targets were identified. These gene products may be related to the establishment and maintenance of the drug resistance. These products may modulate or compensate the mutations in the drug-resistant genes.

Discussion

M leprae evolved from other mycobacterium species by reductive evolution. Only 1604 potentially active genes remain, 1439 of which can be found in MTB. The *M leprae* and TDR strains have many similar properties, including slow growth and gene decay by deletion and frame-shift. The dnaQ-mediated proofreading activities of DNA polymerase III is lost in *M leprae*, whereas mutations are detected in the Pol III in BT1. On comparing the list of deleted genes with the genes with non-synonymous mutations in BT1 and BT2, there are significant overlaps in the categories of fatty acid metabolism genes, lipid-related genes, esterase, and membrane proteins. We consider that TDR MTB strains follow the reductive evolution strategy of *M leprae*.

Acknowledgement

This study was supported by the Research Fund for the

Control of Infectious Diseases, Food and Health Bureau, Hong Kong SAR Government (#08070502).

References

1. 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.
2. Huang da W, Sherman BT, Tan Q, et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol 2007;8:R183.
3. Mizrahi V, Andersen SJ. DNA repair in *Mycobacterium tuberculosis*. What have we learnt from the genome sequence? Mol Microbiol 1998;29:1331-9.
4. Sassetti CM, Boyd DH, Rubin EJ. Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol 2003;48:77-84.
5. Szklarczyk D, Franceschini A, Kuhn M, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 2011;39:D561-8.