Whole-Genome Sequences of Two Clinical Isolates of *Mycobacterium tuberculosis* from Kerala, South India

G. K. Madhavilatha,[a] Biljo V. Joseph,[a] Laiza K. Paul,[a] R. Ajay Kumar,[a] Ramkumar Hariharan,[b] and Sathish Mundayoor[a]

Mycobacterium Research Group, Rajiv Gandhi Centre for Biotechnology, Trivandrum, India,[a] and Cancer Research Program, Rajiv Gandhi Centre for Biotechnology, Trivandrum, India[b]

We report the annotated genome sequence of two clinical isolates of *Mycobacterium tuberculosis* isolated from Kerala, India.

India has the highest prevalence of tuberculosis (TB) worldwide, but the genetic diversity of *Mycobacterium tuberculosis* in India still remains largely unknown. Different studies have shown the prevalence of ancestral strains in India (2, 4). Genome sequencing and comparative genomics of multiple strains have been the basis for studying the epidemiology and evolution of bacterial strains, as these techniques have always revealed novel genome-derived markers.

Here we report the whole-genome sequences of two clinical isolates of *Mycobacterium tuberculosis*, RGTB327 and RGTB423, isolated from Kerala, a state in south India. These strains were part of a repository of field strains that had been made as a part of a drug screening program of the laboratory (3). The two strains were isolated from sputum samples of patients M/41 and M/68, respectively.

The sequencing of the strains was done on a 454 platform, and all alignments were done against the reference strain *Mycobacterium tuberculosis* H37Rv (NC_000962.2) using MIRA (1). Data analysis was performed using inGAP (5). For MTB327, 111 M bases were generated, equivalent to 2,18,166 reads, with a median read length of 518. The same protocol was applied to sequence the MTB423 genome, where 136 M bases were generated, corresponding to 2,68,656 reads and a median read length of 518. The MTB327 genome assembly was derived from 137 contigs, while MTB423 came from 160 contigs. Completeness of these circular genomes was further validated by checking for the absence of partial reading frames or broken genes at both ends of each genome.

We used an automated pipeline, PGAAP (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html; accessed 1 March 2012), for genome annotation. This was followed by manual curation to edit ambiguous assignments where extra information was available. Here we describe some of the salient features of the two strains, using *M. tuberculosis* H37Rv as a reference for the comparative analyses.

The *Mycobacterium tuberculosis* MTB327 genome, consisting of 4,380,119 bases, was obtained at 25.3× coverage, and its G+C content is 65.63%. The number of protein coding genes from consensus predictions was 4,056, and all of the tRNA and rRNA genes found in the reference strain were preserved in this strain. Strikingly, we found a 26-kb deletion along with 66 other large deletions. For the 4,406,587-base genome of *Mycobacterium tuberculosis* MTB423, at 30.8× coverage, the average G+C content was 65.62%. Consensus protein coding gene predictions identified 4,032 genes, 45 tRNAs, and 3 rRNAs. Of the 57 large deletions observed in this genome, the longest one is around 7.5 kb.

Nucleotide sequence accession numbers. The annotated whole-genome sequences obtained from the two strains have been deposited in GenBank under the following accession numbers: CP003233 (MTB327) and CP003234 (MTB423).

REFERENCES

1. **Chevreux B, Wetter T, Suhai S.** 1999. Genome sequence assembly using trace signals and additional sequence information, p 45–56. *In* Computer science and biology. Proceedings of the German Conference on Bioinformatics, GCB '99. GCB, Hannover, Germany.
2. **Gutierrez MC, et al.** 2006. Predominance of ancestral lineages of Mycobacterium tuberculosis in India. Emerg. Infect. Dis. **12**:1367–1374.
3. **Joseph BV, et al.** 2009. Drug resistance in Mycobacterium tuberculosis isolates from tuberculosis patients in Kerala, India. Int. J. Tuberc. Lung Dis. **13**:494–499.
4. **Narayanan S, et al.** 2008. Genomic interrogation of ancestral Mycobacterium tuberculosis from south India. Infect. Genet. Evol. **8**:474–483.
5. **Qi J, Zhao F, Buboltz A, Schuster SC.** 2010. inGAP: an integrated next-generation genome analysis pipeline. Bioinformatics **26**:127–129.