Complete Genome Sequences of a *Mycobacterium smegmatis* Laboratory Strain (MC² 155) and Isoniazid-Resistant (4XR1/R2) Mutant Strains

Abhilash Mohan,[a] Jyothi Padiadpu,[a,b] Priyanka Baloni,[a,c] Nagasuma Chandra[a]

Department of Biochemistry, IISc, Bangalore, India[a]; Supercomputer Education and Research Centre, IISc, Bangalore, India[b]; Molecular Biophysics Unit, IISc, Bangalore, India[c]

We report the whole genome sequences of a *Mycobacterium smegmatis* laboratory wild-type strain (MC² 155) and mutants (4XR1, 4XR2) resistant to isoniazid. Compared to *Mycobacterium smegmatis* MC² 155 (NC_008596), a widely used strain in laboratory experiments, the MC² 155, 4XR1, and 4XR2 strains are 60, 128 and 93 bp longer, respectively.

Emergence of drug resistance is a major bottleneck in tuberculosis treatment (1). *Mycobacterium smegmatis* MC² 155, a non-pathogenic, fast-replicating mycobacterium (2), is widely used as a model system to study *Mycobacterium tuberculosis*. *M. smegmatis* MC² 155 and *M. tuberculosis* share significant similarities in their genomes, including genes associated with mycolic-acid rich cell walls (3). Isoniazid (INH), a front-line clinical drug against *M. tuberculosis*, also inhibits *M. smegmatis* MC² 155, albeit with a higher MIC. Here we report the whole-genome sequences of wild-type and drug-resistant strains of *M. smegmatis* MC² 155.

The wild-type *M. smegmatis* MC² 155 (MC² 155) was subjected to increasing doses of INH, and INH-resistant strains (4XR1 and 4XR2) were selected using standard protocols (4, 5). Two colonies at 4-fold MIC were finally selected and sequenced along with MC² 155.

The genomes of MC² 155, 4XR1, and 4XR2 were sequenced using next-generation sequencing technology, at Genotypic Technology Private Limited, Bangalore, India, using the Illumina HiSeq platform. A total of 12,832,698, 13,636,408, and 13,636,408 unpaired reads of an average length of 101 bp for MC² 155, 4XR1, and 4XR2, respectively, were obtained after trimming and clipping using Trimmomatic (v.0.30) (6). A template-based assembly of the genomes was performed using Bowtie2 (7). The MC² 155 draft genome sequence was assembled using the *M. smegmatis* (NC_008596) strain as the template, while the 4XR1 and 4XR2 sequences were assembled by using the MC² 155 strain. The draft genomes of MC² 155, 4XR1, and 4XR2 were 6,988,269 bp (6,790 genes, 6,625 coding sequences [CDS] and 54 RNAs [47 tRNA, 6 rRNA, and 1 ncRNA]), 6,988,337 bp, and 6,988,302 bp (6,791 genes) long, respectively, with a GC content of 67.4%. The genomes were annotated using Rapid Annotation Transfer Tool (RATT) (8), with annotation from NC_008596 being transferred to MC² 155 and the annotations of MC² 155 transferred to 4XR1 and 4XR2.

The preprocessing of the assembled genome was performed using Picard tools (v1.119), while variant calling was performed using Genome Analysis Toolkit (GATK) (v3.1-1) (9). The GATK-

prescribed best practices were followed to filter and accept single nucleotide polymorphisms (SNPs) and insertion and deletions (indels). The SNP and indel annotations for the strains were added using snpEff (10). A total of 133 indels and 90 SNPs were observed in MC² 155. The 4XR1 strain showed 19 indels and 2 SNPs, while the 4XR2 strain showed 14 indels and 2 SNPs.

The genome sequences of these strains will be useful for obtaining a global understanding of the mechanism of drug resistance and also providing new insights into the biology of *M. smegmatis*, which can also be extrapolated to *M. tuberculosis* and may pave the way for exploring different treatment strategies.

Nucleotide sequence accession numbers. The draft genome sequences of MC² 155, 4XR1, and 4XR2 have been deposited at DDBJ/EMBL/GenBank under the accession numbers CP009494 (strain MC2 155), CP009495 (strain INHR1), and CP009496 (strain INHR2).

ACKNOWLEDGMENTS

REFERENCES

1. **Zumla AI, Gillespie SH, Hoelscher M, Philips PP, Cole ST, Abubakar I, McHugh TD, Schito M, Maeurer M, Nunn AJ.** 2014. New antituberculosis drugs, regimens, and adjunct therapies: needs, advances, and future prospects. Lancet Infect Dis 14:327–340. http://dx.doi.org/10.1016/S1473-3099(13)70328-1.
2. **McGuire AM, Weiner B, Park ST, Wapinski I, Raman S, Dolganov G, Peterson M, Riley R, Zucker J, Abeel T, White J, Sisk P, Stolte C, Koehrsen M, Yamamoto RT, Iacobelli-Martinez M, Kidd MJ, Maer AM, Schoolnik GK, Regev A, Galagan J.** 2012. Comparative analysis of *Mycobacterium* and related actinomycetes yields insight into the evolution of *Mycobacterium tuberculosis* pathogenesis. BMC Genomics 13:120. http://dx.doi.org/10.1186/1471-2164-13-120.
3. **Singh AK, Reyrat JM.** 2009. Laboratory maintenance of *Mycobacterium smegmatis*. Curr Protoc Microbiol Chapter 10:Unit10C.1. http://dx.doi.org/10.1002/9780471729259.mc10c01s14.
4. **Hegreness M, Shoresh N, Damian D, Hartl D, Kishony R.** 2008. Accelerated evolution of resistance in multidrug environments. Proc Natl

Acad Sci U. S. A. **105**:13977–13981. http://dx.doi.org/10.1073/pnas.0805965105.

5. **Lázár V, Pal Singh G, Spohn R, Nagy I, Horváth B, Hrtyan M, Busa-Fekete R, Bogos B, Méhi O, Csörgő B, Pósfai G, Fekete G, Szappanos B, Kégl B, Papp B, Pál C.** 2013. Bacterial evolution of antibiotic hypersensitivity. Mol Syst Biol **9**:700. http://dx.doi.org/10.1038/msb.2013.57.

6. **Bolger AM, Lohse M, Usadel B.** 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**:2114–2120. http://dx.doi.org/10.1093/bioinformatics/btu170.

7. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9**:357–359. http://dx.doi.org/10.1038/nmeth.1923.

8. **Otto TD, Dillon GP, Degrave WS, Berriman M.** 2011. RATT: rapid annotation transfer tool. Nucleic Acids Res **39**:e57. http://dx.doi.org/10.1093/nar/gkq1268.

9. **McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA.** 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res **20**:1297–1303. http://dx.doi.org/10.1101/gr.107524.110.

10. **Reumers J, Schymkowitz J, Ferkinghoff-Borg J, Stricher F, Serrano L, Rousseau F.** 2005. SNPeffect: a database mapping molecular phenotypic effects of human non-synonymous coding SNPs. Nucleic Acids Res **33**:D527–D532. http://dx.doi.org/10.1093/nar/gki086.