



Comparative Genomic Analysis of *Mycobacteriaceae* Reveals Horizontal Gene Transfer-Mediated Evolution of the CRISPR-Cas System in the *Mycobacterium tuberculosis* Complex

Anoop Singh,^a Mohita Gaur,^a Vishal Sharma,^a Palak Khanna,^a Ankur Bothra,^b Asani Bhaduri,^c Anupam Kumar Mondal,^b Debasis Dash,^b Yogendra Singh,^a Richa Misra^{a,d}

^aDepartment of Zoology, University of Delhi, Delhi, India

^bCSIR-IGIB, Sukhdev Vihar, New Delhi, India

^cCluster Innovation Centre, University of Delhi, Delhi, India

^dDepartment of Zoology, Sri Venkateswara College, University of Delhi, Delhi, India

ABSTRACT Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) genes are conserved genetic elements in many prokaryotes, including *Mycobacterium tuberculosis*, the causative agent of tuberculosis. Although knowledge of CRISPR locus variability has been utilized in *M. tuberculosis* strain genotyping, its evolutionary path in *Mycobacteriaceae* is not well understood. In this study, we have performed a comparative analysis of 141 mycobacterial genomes and identified the exclusive presence of the CRISPR-Cas type III-A system in *M. tuberculosis* complex (MTBC). Our global phylogenetic analysis of CRISPR repeats and Cas10 proteins offers evidence of horizontal gene transfer (HGT) of the CRISPR-Cas module in the last common ancestor of MTBC and *Mycobacterium canettii* from a *Streptococcus*-like environmental bacterium. Additionally, our results show that the variation of CRISPR-Cas organization in *M. tuberculosis* lineages, especially in the Beijing sublineage of lineage 2, is due to the transposition of insertion sequence IS6110. The direct repeat (DR) region of the CRISPR-Cas locus acts as a hot spot for IS6110 insertion. We show in *M. tuberculosis* H37Rv that the repeat at the 5' end of CRISPR1 of the forward strand is an atypical repeat made up partly of IS-terminal inverted repeat and partly CRISPR DR. By tracing an undetectable spacer sequence in the DR region, the two CRISPR loci could theoretically be joined to reconstruct the ancestral single CRISPR-Cas locus organization, as seen in *M. canettii*. This study retracing the evolutionary events of HGT and IS6110-driven genomic deletions helps us to better understand the strain-specific variations in *M. tuberculosis* lineages.

IMPORTANCE Comparative genomic analysis of prokaryotes has led to a better understanding of the biology of several pathogenic microorganisms. One such clinically important pathogen is *M. tuberculosis*, the leading cause of bacterial infection worldwide. Recent evidence on the functionality of the CRISPR-Cas system in *M. tuberculosis* has brought back focus on these conserved genetic elements, present in many prokaryotes. Our study advances understanding of mycobacterial CRISPR-Cas origin and its diversity among the different species. We provide phylogenetic evidence of acquisition of CRISPR-Cas type III-A in the last common ancestor shared between MTBC and *M. canettii*, by HGT-mediated events. The most likely source of HGT was an environmental *Firmicutes* bacterium. Genomic mapping of the CRISPR loci showed the IS6110 transposition-driven variations in *M. tuberculosis* strains. Thus, this study offers insights into events related to the evolution of CRISPR-Cas in *M. tuberculosis* lineages.

KEYWORDS CRISPR-Cas system, *Mycobacterium*, *Mycobacterium tuberculosis*, *Mycobacterium canettii*, horizontal gene transfer, IS6110, evolution, CRISPR-Cas system, comparative genomics, transposons

Citation Singh A, Gaur M, Sharma V, Khanna P, Bothra A, Bhaduri A, Mondal AK, Dash D, Singh Y, Misra R. 2021. Comparative genomic analysis of *Mycobacteriaceae* reveals horizontal gene transfer-mediated evolution of the CRISPR-Cas system in the *Mycobacterium tuberculosis* complex. mSystems 6:e00934-20. <https://doi.org/10.1128/mSystems.00934-20>.

Editor David W. Cleary, University of Southampton

Copyright © 2021 Singh et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Yogendra Singh, ysinghdu@gmail.com, or Richa Misra, richamisra@svc.ac.in.

Received 15 September 2020

Accepted 31 December 2020

Published 19 January 2021