

First Report on the Draft Genome Sequences of *Corynebacterium diphtheriae* Isolates from India

Balaji Veeraraghavan,^a Shalini Anandan,^a Suresh Kumar Rajamani Sekar,^a Radha Gopi,^a Naveen Kumar Devanga Ragupathi,^a Srilekha Ramesh,^a Valsan Philip Verghese,^b Sophy Korulla,^b Sarah Mathai,^b Lucky Sangal,^c Sudhir Joshi^c

Department of Clinical Microbiology, Christian Medical College, Vellore, Tamil Nadu, India^a; Department of Child Health, Christian Medical College, Vellore, Tamil Nadu, India^b; World Health Organisation, Country office, New Delhi, India^c

We report here the draft genome sequences of five *Corynebacterium diphtheriae* isolates of Indian origin. The *C. diphtheriae* isolates TH1141, TH510, TH1526, TH1337, and TH2031 belong to sequence type ST-50, ST-295, ST-377, ST-405, and ST-405, with an average genome size of 2.5 Mbp.

Received 29 September 2016 Accepted 30 September 2016 Published 23 November 2016

Citation Veeraraghavan B, Anandan S, Rajamani Sekar SK, Gopi R, Devanga Ragupathi NK, Ramesh S, Verghese VP, Korulla S, Mathai S, Sangal L, Joshi S. 2016. First report on the draft genome sequences of *Corynebacterium diphtheriae* isolates from India. *Genome Announc* 4(6):e01316-16. doi:10.1128/genomeA.01316-16.

Copyright © 2016 Veeraraghavan 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 Balaji Veeraraghavan, vbalaji@cmcvellore.ac.in.

Corynebacterium diphtheriae is an important human pathogen and the causative agent of the upper respiratory tract infection known as diphtheria (1). The existence of genetic diversity among the strains and its circulation around the world are a serious concern in recent time (2). To the best of our knowledge, this is the first genome report on the *C. diphtheriae* strains isolated from India.

Five *C. diphtheriae* isolates (TH1141, TH510, TH1526, TH1337, and TH2031) were obtained from the patients diagnosed with diphtheria infection at Christian Medical College & Hospital, South India. The presence of *toxA* and *rpoB* genes, typical of *C. diphtheriae*, was confirmed by real-time PCR (RT-PCR) analysis. All the isolates were found to have intermediate resistance to penicillin. Whole-genome sequencing (WGS) was done to elucidate the genetic diversity among the strains. Genomic DNA was extracted with QIAamp DNA minikit (Qiagen, Germany), and the library fragments were prepared using the Ion Plus fragment library kit (Ion Torrent; Life Technologies, USA). Whole-genome sequencing was performed using the Ion Torrent PGM sequencer (Thermo Scientific Fisher Corp., USA) using 400-bp chemistry. The embedded assembler SPAdes tool (version 5.0.0.0) of the Ion Torrent server (version 5.0.4) was used for *de novo* assembly of the next-generation sequencing (NGS) data. Genome annotation was performed using the Pathosystems Resource Integration Centre (PATRIC) (<http://www.patricbrc.org>) (3) and NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/) (4).

The whole-genome assembly of the *C. diphtheriae* isolates (TH1141, TH510, TH1526, TH1337, and TH2031) resulted in the 68, 27, 48, 111, and 136 contigs (≥ 500 bp) with sequencing coverages of 66 \times , 96 \times , 97 \times , 75 \times , and 106 \times , respectively. The number of coding sequences (CDSs) per genome ranged between 2,240 and 2,385, with an average genome size and G+C content of 2.5 Mbp and 53.5%, respectively. The isolates TH1141, TH510, and TH1526 belong to sequence ST-50, ST-295, and ST-377, and TH1337 and TH2031 belong to ST-405, as predicted by the *in-silico* multilocus sequence type analysis (<https://cge.cbs.dtu.dk/services/MLST/>) (5). Genome annotation by PATRIC predicted a total of 52 tRNAs, seven rRNAs, five antibiotic resistance genes (ARDB), and five virulence genes (VFDB) for the TH1141 genome. The TH510 and TH1526 *C. diphtheriae* genomes harbor 52, seven, five, and four and 53, seven, eight, and seven genes corresponding to the tRNA, rRNA, antibiotic resistance (ARDB), and virulence factor (VFDB), respectively. The TH1337 and TH2031 genomes were found to contain 51, eight, five, and five and 55, seven, five, and five genes related to tRNA, rRNA, antibiotic resistance (ARDB), and virulence factor (VFDB). This study advances our knowledge on the genomic diversity and molecular epidemiology of *C. diphtheriae* in India.

Accession number(s). The draft genome sequences of the five *C. diphtheriae* isolates have been deposited in the DDBL/ENA/GenBank database under the accession numbers listed in [Table 1](#).

TABLE 1 Genome characteristics and accession numbers of *C. diphtheriae* isolates from India

<i>C. diphtheriae</i> isolate	Sequencing coverage (\times)	No. of contigs (>500 bp)	Draft genome size (Mbp)	Sequence type	Accession no.
TH1141	66	68	2.46	ST-50	MDYT00000000
TH510	96	27	2.54	ST-295	MDYV00000000
TH1526	97	48	2.45	ST-377	MDYU00000000
TH1337	75	111	2.44	ST-405	MJAQ00000000
TH2031	106	136	2.49	ST-405	MJAR00000000

ACKNOWLEDGMENT

We thank the institutional review board of Christian Medical College, Vellore, India, for approving the study.

FUNDING INFORMATION

This work, including the efforts of Balaji Veeraraghavan, was funded by World Health Organization (WHO) (2015/561588-0).

REFERENCES

1. Holmes RK. 2000. Biology and molecular epidemiology of diphtheria toxin and the *tox* gene. *J Infect Dis* 181:S156–S167. <http://dx.doi.org/10.1086/315554>.
2. Sangal V, Hoskisson PA. 2016. Evolution, epidemiology and diversity of *Corynebacterium diphtheriae*: new perspectives on an old foe. *Infect Genet Evol* 43:364–370. <http://dx.doi.org/10.1016/j.meegid.2016.06.024>.
3. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, Machi D, Mao C, Nordberg EK, Olson R, Overbeek R, Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Vonstein V, Warren A, Will R, Wilson MJC, Yoo HS, Zhang C, Zhang Y, Sobral BW. 2014. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* 42:D581–D591. <http://dx.doi.org/10.1093/nar/gkt1099>.
4. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. Prokaryotic genome annotation pipeline. *Nucleic Acids Res* 44:6614–6624. <http://dx.doi.org/10.1093/nar/gkw569>.
5. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Ponten T, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* 50:1355–1361. <http://dx.doi.org/10.1128/JCM.06094-11>.