JB Journals.ASM.org

GENOME ANNOUNCEMENT

The Draft Genome Sequence of *Corynebacterium diphtheriae* bv. mitis NCTC 3529 Reveals Significant Diversity between the Primary Disease-Causing Biovars

Vartul Sangal,[a] Nicholas P. Tucker,[a] Andreas Burkovski,[b] and Paul A. Hoskisson[a]

Strathclyde Institute of Pharmacy and Biomedical Science, University of Strathclyde, Glasgow, United Kingdom,[a] and Lehrstuhl für Mikrobiologie, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany[b]

We report the draft genome of the human pathogen *Corynebacterium diphtheriae* bv. mitis NCTC 3529. This is the first *C. diphtheriae* bv. mitis strain to be sequenced and reveals significant differences from the other primary biovar, *C. diphtheriae* bv. gravis.

Corynebacterium diphtheriae is the causative agent of diphtheria, a toxemic localized infection of the upper respiratory tract (6). Currently, four biovars are recognized (*C. diphtheriae* bv. gravis, *C. diphtheriae* bv. mitis, *C. diphtheriae* bv. intermedius, and *C. diphtheriae* bv. belfanti), based on their morphogenic and biochemical properties (5), and yet the molecular and genetic bases for these differences are poorly understood. We undertook the sequencing of *C. diphtheriae* bv. mitis NCTC 3529, a strain isolated prior to mass vaccination in the United Kingdom, to better understand the genetic basis that differentiates this biovar from the other main disease-causing biovar, *C. diphtheriae* bv. gravis.

Genome sequencing of *C. diphtheriae* bv. mitis strain NCTC 3529 was carried out using a whole-genome shotgun sequencing approach performed on a Roche GS-Junior 454 apparatus at the University of Strathclyde Genomics Facility. Using Rapid library-prepared single-ended runs, we obtained 175,879 reads with an average length of 446 bp. The reads were assembled using GS *de novo* Assembler (Roche), which led to a final assembly of 43 contigs of >200 bp. The total size of the assembly was 2,483,675 bp, with a mean contig size of 57,760 bp (an average of 47× coverage) and a G+C content of 53.57%. The contigs were ordered onto *C. diphtheriae* bv. gravis NCTC 13129 (1) using Mauve (4), and the merged sequence was annotated using xBASE (2).

The draft genome of *C. diphtheriae* bv. mitis NCTC 3529 is estimated to have a total of 2,272 protein coding genes. The overall genome size is smaller than that of *C. diphtheriae* bv. gravis by 4,960 bp. Analysis using mGenomeSubtractor (7) indicated that 1,982 coding sequences (CDS) were present in both strains, with 318 CDS being present in *C. diphtheriae* bv. mitis NCTC 3529 that were not present in *C. diphtheriae* bv. gravis NCTC 13129. These differences were largely confined to the presence of transposons and clustered, regularly interspaced, short palindromic repeat (CRISPR) elements. The biochemical tests used to distinguish between the gravis and mitis biovars in the laboratory include the inability of *C. diphtheriae* bv. mitis to ferment starch and reduce nitrates (3, 5). Genomic analysis shows that the lack of starch fermentation in NCTC 3529 is likely to be linked to the putative aldose-1-epimerase (DIP1011) that is present in NCTC 13129 but absent in NCTC 3529. In *C. diphtheriae* bv. mitis NCTC 3529, the region that encodes the molybdopterin biosynthesis machinery and the *narIJHGK* operon (DIP0492 to DIP0507 in *C. diphtheriae* bv. gravis NCTC 13129) is absent and likely to account for the lack of nitrate reduction by this biovar. Overall, the genome sequence of *C. diphtheriae* bv. mitis NCTC 3529 should advance our knowledge of the biology of pathogenic corynebacteria and the molecular basis for strain differentiation.

Nucleotide sequence accession numbers. The *C. diphtheriae* bv. mitis (NCTC 3529) Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. AJGI00000000. The version described in this paper is the first version, AJGI01000000.

ACKNOWLEDGMENTS

Work in the P.A.H. laboratory is supported by Medical Research Scotland (grant 422 FRG) and the University of Strathclyde. A.B. is supported by the Deutsche Forschungsgemeinschaft (SFB796, B5).

REFERENCES

1. **Cerdeno-Tarraga AM, et al.** 2003. The complete genome sequence and analysis of Corynebacterium diphtheriae NCTC13129. Nucleic Acids Res. **31**:6516–6523.
2. **Chaudhuri RR, Pallen MJ.** 2006. xBASE, a collection of online databases for bacterial comparative genomics. Nucleic Acids Res. **34**:D335–D337.
3. **Coyle MB, Nowowiejski DJ, Russell JQ, Groman NB.** 1993. Laboratory review of reference strains of Corynebacterium diphtheriae indicates mistyped intermedius strains. J. Clin. Microbiol. **31**:3060–3062.
4. **Darling A, Mau B, Blattner F, Perna N.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. **14**: 1394–1403.
5. **Efstratiou A, et al.** 2000. Current approaches to the laboratory diagnosis of diphtheria. J. Infect. Dis. **181**:S138–S145.
6. **Hadfield TL, McEvoy P, Polotsky Y, Tzinserling VA, Yakovlev AA.** 2000. The pathology of diphtheria. J. Infect. Dis. **181**(Suppl. 1):S116–S120.
7. **Shao Y, et al.** 2010. mGenomeSubtractor: a web-based tool for parallel in silico subtractive hybridization analysis of multiple bacterial genomes. Nucleic Acids Res. **38**:W194–W200.

Received 2 April 2012 Accepted 9 April 2012

Address correspondence to Paul A. Hoskisson, Paul.hoskisson@strath.ac.uk.

doi:10.1128/JB.00503-12