JB Journals.ASM.org

GENOME ANNOUNCEMENT

Draft Genome Sequence of *Corynebacterium diphtheriae* Biovar Intermedius NCTC 5011

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, Nuremberg, Germany[b]

We report an annotated draft genome of the human pathogen *Corynebacterium diphtheriae* bv. intermedius NCTC 5011. This strain is the first *C. diphtheriae* bv. intermedius strain to be sequenced, and our results provide a useful comparison to the other primary disease-causing biovars, *C. diphtheriae* bv. gravis and *C. diphtheriae* bv. mitis. The sequence has been deposited at DDBJ/EMBL/GenBank with the accession number AJVH01000000.

Prior to the introduction of mass vaccination in the United Kingdom, *Corynebacterium diphtheriae*, which is the etiological agent of diphtheria, was a major cause of human disease, with more than 50,000 cases per year (6). The main mechanism of virulence is through the bacteriophage-carried diphtheria toxin (1). Currently, there are four recognized biovars, *C. diphtheriae* bv. gravis, *C. diphtheriae* bv. mitis, *C. diphtheriae* bv. intermedius, and *C. diphtheriae* bv. belfanti, based on biochemical and morphogenic properties (3, 5). The molecular basis for these differences is not well defined and requires further investigation. To address this, we have sequenced the whole genome of *C. diphtheriae* bv. intermedius NCTC 5011, a strain deposited in the culture collection prior to the introduction of mass vaccination in the United Kingdom and therefore not subject to the evolutionary selective pressure of vaccination.

Sequencing of the *C. diphtheriae* bv. intermedius strain NCTC 5011 whole genome was performed using whole-genome shotgun sequencing on a Roche GS-Junior 454 apparatus at the University of Strathclyde. The reads were assembled using the GS *de novo* Assembler (Roche), which led to a final assembly of 34 contigs of >300 bp. The total size of the assembly was 2.38 Mbp, with a mean contig size of 70 kbp (average of 31-fold coverage) and a G+C content of 53.6%. Contigs were reordered onto the *C. diphtheriae* bv. gravis NCTC 13129 reference genome (1) by using the Mauve program (4) and were annotated using the Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) at NCBI and xbase (2).

The whole genome of *C. diphtheriae* bv. intermedius NCTC 5011 is estimated to have a total of 2,318 coding sequences (CDS). The genome differs from *C. diphtheriae* bv. gravis by 108 kb. Analysis using mGenomesubstractor (8) indicated that 2,050 CDS were present in both strains, with 104 CDS being present in *C. diphtheriae* bv. intermedius NCTC 5011 that were not present in *C. diphtheriae* bv. gravis NCTC 13129. The majority of *C. diphtheriae* bv. intermedius-specific sequences were transposons and restriction-modification systems (type I and type III); additionally, the presence of short repeat regions is suggestive of *cas*/CRISPR systems, indicating that genomic plasticity and barriers to lateral gene transfer may be responsible for the majority of differences observed between *C. diphtheriae* biovars.

The genome sequence of *C. diphtheriae* bv. *intermedius* advances our understanding of the genome and population structure of *C. diphtheriae* and adds to data relating to other recent *C. diphtheriae* sequencing efforts (7, 9).

Nucleotide sequence accession numbers. The results of this *C. diphtheriae* bv. intermedius (NCTC 5011) annotated genome project have been deposited at DDBJ/EMBL/GenBank under accession number AJVH00000000. The version described in this paper is the first version, AJVH01000000.

ACKNOWLEDGMENTS

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. **Mortimer P.** 2010. Diphtheria and the origins of the UK childhood immunisation programme. Microbiol. Today **February**:38–41.
7. **Sangal V, Tucker NP, Burkovski A, Hoskisson PA.** 2012. The draft genome sequence of *Corynebacterium diphtheriae* bv. mitis NCTC 3529 reveals significant diversity between the primary disease-causing biovars. J. Bacteriol. **194**:3269.
8. **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.
9. **Trost E, et al.** 2012. Pan-genomics of *Corynebacterium diphtheriae*: insights into the genomic diversity of pathogenic isolates from cases of classical diphtheria, endocarditis, and pneumonia. J. Bacteriol. **194**:3199–3215.

Received 29 May 2012 **Accepted** 5 June 2012

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

doi:10.1128/JB.00939-12