JB Journals.ASM.org

GENOME ANNOUNCEMENT

Draft Genome Sequence of the Nontoxigenic *Clostridium difficile* Strain CD37

Michael S. M. Brouwer, Elaine Allan, Peter Mullany, and Adam P. Roberts

UCL Eastman Dental Institute, London, United Kingdom

Here we report the draft genome sequence of *Clostridium difficile* strain CD37, the first nontoxigenic strain sequenced. Every sequenced strain of *Clostridium difficile* has been shown to contain multiple different mobile genetic elements. The draft genome sequence of strain CD37 reveals the presence of two putative conjugative transposons.

Clostridium difficile is an anaerobic, spore-forming, Gram-positive bacterium that is an opportunistic pathogen of humans and other mammals (2). The main virulence factors are the toxins A and B and the less-well-characterized binary toxin (18). Interestingly, nontoxigenic *C. difficile* strains (which lack the genes for toxins A and B and the binary toxin) are relatively common, although little is known about their biology (18). *C. difficile* has previously been shown to contain various mobile genetic elements, such as conjugative transposons (5, 6, 17, 19, 20), mobilizable transposons (11), prophages (12), and other elements, such as the IStron (4) and the *C. difficile skin* (*skin*^{Cd}) element (13).

Genome shotgun sequencing of strain CD37, ribotype 009 (5), was performed using an Illumina GAII-X. A total of 7,515,781 72-bp paired-end reads resulted in ~130-fold coverage of the genome in 245 contigs. Initial genome assembly was carried out using the Velvet software suite (21). The assembled genome was compared and mapped against the annotated 630 reference genome (GenBank accession no. AM180355) using xBase (9). Gene prediction was performed using Glimmer (10), tRNA genes were searched for with tRNAScan-SE (16), rRNA genes were searched for with RNAmmer (15), and Protein BLAST (1) was run using the translated coding sequences as a query against the 630 reference sequence. The best result for each BLAST search was imported as the gene annotation. A comparison file with concatenated reference and sequence files was produced using MUMMER (14). Analysis of the genome was performed using ACT (8).

Based on an alignment with strain 630 as the reference sequence, CD37 contains putative conjugative transposons with high nucleotide sequence similarity to CTn*4* and CTn*7* of strain 630 (~96% and ~98%, respectively). These mobile genetic elements are consistently found in *C. difficile* genomes (5, 6, 20). To date, all *C. difficile* genome sequences contain a *skin*^{Cd} element inserted within *sigK*, an RNA polymerase sigma factor shown to be essential for sporulation in *C. difficile*. However, *skin*^{Cd} is absent in the CD37 sequence, confirming previous PCR data (13).

In strain CD37, the PaLoc (encoding toxin A and B) is absent; in its place is 115 bp of noncoding DNA. This has also been observed in the nontoxigenic strain 7322 (3). Carter et al. (7) previously reported a 68-bp linker in place of the binary toxin locus in strain CD37. Our data show that within this 68 bp there is in fact only 16 bp of unique sequence in strain CD37 compared to that in strain 630. The fact that none of the main toxin genes are present in strain CD37 means that it will be a useful resource for workers wishing to investigate the effect of putative virulence factors without the complication of the toxins, which may mask other effects.

Nucleotide sequence accession numbers. This Whole Genome Shotgun project has been deposited in DDBJ/EMBL/GenBank under accession number AHJJ00000000. The version described in this paper is the first version, AHJJ01000000.

ACKNOWLEDGMENTS

This work was funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 223585 and the Medical Research Council (grant no. G0601176).

We are grateful for the excellent help from the sequencing team at UCL Genomics.

REFERENCES

1. **Altschul SF, et al.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.
2. **Bartlett JG.** 2008. Historical perspectives on studies of *Clostridium difficile* and *C. difficile* infection. Clin. Infect. Dis. **46**(Suppl 1):S4–S11.
3. **Braun V, Hundsberger T, Leukel P, Sauerborn M, von Eichel-Streiber C.** 1996. Definition of the single integration site of the pathogenicity locus in *Clostridium difficile*. Gene **181**:29–38.
4. **Braun V, et al.** 2000. A chimeric ribozyme in *Clostridium difficile* combines features of group I introns and insertion elements. Mol. Microbiol. **36**:1447–1459.
5. **Brouwer MS, Warburton PJ, Roberts AP, Mullany P, Allan E.** 2011. Genetic organisation, mobility and predicted functions of genes on integrated, mobile genetic elements in sequenced strains of *Clostridium difficile*. PLoS One **6**:e23014.
6. **Brouwer MS, Roberts AP, Mullany P, Allan E.** 2012. In silico analysis of sequenced strains of *Clostridium difficile* reveals a related set of conjugative transposons carrying a variety of accessory genes. Mob. Genet. Elements **2**:1–5.
7. **Carter GP, et al.** 2007. Binary toxin production in *Clostridium difficile* is regulated by CdtR, a LytTR family response regulator. J. Bacteriol. **189**:7290–7301.
8. **Carver T, et al.** 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinformatics **24**:2672–2676.
9. **Chaudhuri RR, et al.** 2008. xBASE2: a comprehensive resource for comparative bacterial genomics. Nucleic Acids Res. **36**:D543–D546.
10. **Delcher AL, Bratke KA, Powers EC, Salzberg SL.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23**:673–679.
11. **Farrow KA, Lyras D, Rood JI.** 2001. Genomic analysis of the erythromy-

Received 26 January 2012 **Accepted** 3 February 2012

Address correspondence to Adam P. Roberts, adam.roberts@ucl.ac.uk.

doi:10.1128/JB.00122-12

cin resistance element Tn5398 from *Clostridium difficile*. Microbiology **147**:2717–2728.
12. **Goh S, Ong PF, Song KP, Riley TV, Chang BJ.** 2007. The complete genome sequence of *Clostridium difficile* phage phiC2 and comparisons to phiCD119 and inducible prophages of CD630. Microbiology **153**:676–685.
13. **Haraldsen JD, Sonenshein AL.** 2003. Efficient sporulation in *Clostridium difficile* requires disruption of the sigmaK gene. Mol. Microbiol. **48**:811–821.
14. **Kurtz S, et al.** 2004. Versatile and open software for comparing large genomes. Genome Biol. **5**:R12.
15. **Lagesen K, et al.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–3108.
16. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**: 955–964.
17. **Mullany P, et al.** 1990. Genetic analysis of a tetracycline resistance element from *Clostridium difficile* and its conjugal transfer to and from *Bacillus subtilis*. J. Gen. Microbiol. **136**:1343–1349.
18. **Rupnik M, Wilcox MH, Gerding DN.** 2009. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. Nat. Rev. Microbiol. **7**:526–536.
19. **Sebaihia M, et al.** 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nat. Genet. **38**:779–786.
20. **Stabler RA, et al.** 2009. Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biol. **10**:R102.
21. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**:821–829.