C. 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 (skinCd) element (13).

Genome shotgun sequencing of strain CD37, ribotype 009 (5), was performed using an Illumina GAILI-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), RNAse 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 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 CTn4 and CTn7 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 skinCd element inserted within sigK, an RNA polymerase sigma factor shown to be essential for sporulation in C. difficile. However, skinCd 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 nontoxicigenic 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

11. Farrow KA, Lytras D, Rood JJ. 2001. Genomic analysis of the erythromycin-
cin resistance element Tn5398 from *Clostridium difficile*. Microbiology 147:2717–2728.