Microb Health Dis 2019; 1: e148

European Helicobacter & Microbiota Study Group

Draft genome sequences of a *Campylobacter curvus* and three *Campylobacter ureolyticus* strains isolated from human colonic mucosal tissue

M. Gemmell[1], S. Berry[2], I. Mukhopadhya[2], L. Gong[3], R. Hansen[4], and G. L. Hold[3]

[1]Centre for Genomic Research, Institute of Integrative Biology, University of Liverpool, Liverpool, L69 7ZB UK
[2]School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, AB25 2ZD UK
[3]Microbiome Research Centre, St George & Sutherland Clinical School, University of New South Wales, Sydney, NSW 2052, Australia
[4]Department of Paediatric Gastroenterology, Royal Hospital for Children, Glasgow, G51 4TF UK

Corresponding Author: Georgina Hold; e-mail: georgina.hold@unsw.edu.au

Abstract: During a study to assess the human microaerophilic colonic mucosal microbiota, various *Campylobacter* isolates were identified based on phenotypic characterisation and 16S rRNA gene sequence analysis. Here we report the draft genome sequences of four new *Campylobacter* strains isolated from colonic mucosal biopsies – a *Campylobacter curvus* strain isolated from a paediatric patient and three *Campylobacter ureolyticus* strains isolated from adults.

Keywords: *Campylobacter curvus*, *Campylobacter ureolyticus,* Whole genome sequencing, Colonic mucosal isolates.

GENOME ANNOUNCEMENT

Genus *Campylobacter* has been recognized as a causative agent of various animal and human diseases[1,2]. Whilst *Campylobacter jejuni* and *C. coli* are major zoonotic species, causing human health issues via the consumption of contaminated poultry products. Other *Campylobacters* including *C. concisus, C. showae, C. gracilis, C. ureolyticus, C. curvus*, and *C. rectus* are commensal human oral species. The prevalence of human illness associated with oral *Campylobacter* species is considerably lower than zoonotic *Campylobacter* species induced illness; however it is thought that non-jejuni/coli *Campylobacter* illness is heavily underreported due to the lack of effective culture-based detection[3,4]. Whilst recently there has been an increase in comparative genomics and genome biology studies of *C. concisus* and *C. showae*[5-7], there is a paucity of whole genome sequences for other *Campylobacter* species including *C. curvus* and *C. ureolyticus*. *C. ureolyticus* has been reported to surpass *C. coli* as the second most common causative agent of *Campylobacter*-related gastroenteritis in Ireland[8]. *C. curvus* is rarely encountered in the human intestinal

tract but is thought to be underreported. Additional genome sequences of clinical isolates of these species are needed to allow comprehensive understanding of pathogenicity/virulence traits. Here, we report the draft genome sequences of one *C. curvus* and three *C. ureolyticus* strains.

The four *Campylobacter* clinical isolates – *Campylobacter curvus* strain isolated from a paediatric patient undergoing colonoscopy to rule out a diagnosis of inflammatory bowel disease (B23 cloudy PP) as part of the BISCUIT study[9] and three *Campylobacter ureolyticus* strains from two adults undergoing colonoscopy based on a positive faecal occult blood test, who were found to have no evidence of macroscopic or microscopic disease (GH136_SC, GH136_TPP, and GH187_Mit_SmBr). Isolates GH136_SC and GH136_TPP were both isolated from the same individual but they had differing morphology. Culture work was performed as described in Mukhopadhya et al[10], utilising five selective plates and one plain blood agar plate, each incubated in microaerophilic gas conditions generated by Anoxomat® (Mart® Microbiology, Drachten, the Netherlands) and reviewed twice weekly for up to one month. Gram-negative and oxygen sensitive (by virtue of failed subculture in room air) bacterial isolates were identified by sequencing of the 16S rRNA gene. A minimum read length of 400 bp was obtained for attributing bacterial identities, the result of which was searched against the NCBI BLAST database (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

We performed short-read sequencing using the Illumina MiSeq with MiSeq Reagent Kit v3 (Illumina, San Diego, CA, USA) for 600 cycles and 2 × 300 bp paired end reads as described previously[6]. Genomic DNA was extracted from liquid cultures using the Promega Genomic DNA Purification kit. Library construction was performed using the Illumina Nextera XT DNA Library Prep Kit and Nextera XT Index Kit v2. The quality of raw paired-end reads was checked using FastQC.

A total of 573,164 (B23 cloudy PP), 777,723 (GH136_SC), 284,645 (GH136_TPP), and 413,897 (GH187_Mit_SmBr) paired-end reads were generated of read length 300bp. This constitutes coverage equivalent to around 170X (B23 cloudy PP), 300X (GH136_SC), 110X (GH136_TPP), and 150X (GH187_Mit_SmBr). Strains were error corrected and assembled with the default settings of the A5-miseq pipeline[11]. Prior to assembly, low-quality bases were trimmed from the sequence using Trim galore[12]. We used the CLARK sequence classification method (v1.2.3) with a custom database to classify the strains. This confirmed their identity as *C. curvus* (B23 cloudy PP) and *C. ureolyticus* (GH136_SC, GH136_TPP and GH187_Mit_SmBr). We assembled plasmids from publicly available databases (NCBI and DDBJ) using plasmidSPAdes as described previously[7,13]. Genome and plasmid assemblies were annotated with Prokka[14]. The draft genome sequence of B23 cloudy PP was comprised of 28 contigs, GH136_SC, GH136_TPP, and GH187_Mit_SmBr draft genome sequences were comprised of 19, 28, and 281 contigs, respectively. The total genome size of each of the four isolates was 2,017,439 bases (B23 cloudy PP), 1,556,879 bases (GH136_SC), 1,551,312 bases (GH136_TPP), and 1,712,726 bases (GH187_Mit_SmBr). GC content of the isolates was 44.31% (B23 cloudy PP), 29.2% (GH136_SC), 29.17% (GH136_TPP), and 28.99% (GH187_Mit_SmBr). In comparison, type strain *C. curvus* DSM 6644 has a genome size of 1,954,904 and GC content of 44.3% and 8 contigs (Accessed from NCBI genome resource). *C. ureolyticus* RIGS 9880 has a genome size of 1,642 kb and GC content of 29.23% and is a closed genome[15]. Genome completeness estimation was carried out on the new assemblies using the BUSCO tool in combination with the Proteobacteria BUSCO database (16). Of the 221 Proteobacteria BUSCOs, 205 were detected in B23 cloudy PP, 200 in GH136_SC, 200 in GH136_TPP and 195 in GH187_Mit_SmBr. The new genomes contained 2009 coding DNA sequences (CDS; B23 cloudy PP), 1531 (GH136_SC), 1524 (GH136_TPP), and 1667 (GH187_Mit_SmBr). This equated to the following number of unique genes - 2055 (B23 cloudy PP), 1574 (GH136_SC), 1567 (GH136_TPP), and 1710 (GH187_Mit_SmBr). This compares with 1980 genes identified in *C. curvus* DSM 6644 and 1595 genes identified in *C. ureolyticus* RIGS 9880. Prokka also predicted 45 RNA sequences (2 rRNA and 43 tRNA) for B23 cloudy PP, and 42 RNA sequences (2 rRNA and 40 tRNA) for all 3 *C. ureolyticus* isolates.

Plasmid assembly produced two contigs for B23 cloudy PP, one contig for GH136_SC, 15 contigs for GH136_TPP and 45 contigs for GH187_Mit_SmBr. The total length of the contigs were 29,732 bp (B23 cloudy PP), 122,486 bp (GH136_SC), 236,292 bp (GH136_TPP), and 36,340 bp (GH187_Mit_SmBr). The longest contig for each isolate was 19,987 bp (B23 cloudy PP), 122,486 bp (GH136_SC), 83,171 bp (GH136_TPP), and 35,321 bp (GH187_Mit_SmBr). GC content of the contigs was 38.24% (B23 cloudy PP), 27.8% (GH136_SC), 30.45% (GH136_TPP), and 29.3% (GH187_Mit_SmBr). The plasmid contigs for B23 cloudy PP contained 30 coding DNA sequences, whilst the single contig for GH136_SC contained 244 coding sequences. One hundred and twenty-seven coding sequences were detected within GH136_TPP plasmid contigs and 370 coding sequences within GH187_Mit_SmBr plasmid contigs.

We looked for the presence of bacterial secretion systems within the genomes as these virulence factors have been the focus of interest for other *Campylobacter* species[6,7,17]. Proteins be-

longing to a type II secretion system were detected in all four isolates. GH187_Mit_SmBr also had proteins belonging to a type IV secretion system. CRISPR-associated endonuclease genes *cas*1, *cas*2 and helicase *cas*3 were found in all isolates along with RTX toxins consisting of hemolysins (all isolates) and leukotoxins (all isolates except *C. curvus* B23_cloudy_PP). The presence of different pathogenicity factors indicates varying virulence methods are present within different clinical isolates and this variability should be the focus of further study.

NUCLEOTIDE SEQUENCE ACCESSION NUMBERS

The *C. concisus* raw sequencing reads and genome assemblies are freely available from the EMBL-EBI ENA under the study accession number PRJEB34136. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number PRJE34136.

Acknowledgements

The authors thank Centre for Genome Enabled Biology and Medicine, University of Aberdeen for Illumina sequencing. This work was supported by a Fulbright Scholarship to G.L.H., an NHS Grampian Endowment grant fund to I.M. and G.L.H., and a CSO clinical academic fellowship to R.H. (CAF/08/01). R.H. is supported by an NHS Research Scotland Career Researcher Fellowship.

Conflict of interest

The Authors declare that they have no conflict of interests.

REFERENCES

1. Facciolà A, Riso R, Avventuroso E, Visalli G, Delia SA, Laganà P. Campylobacter: from microbiology to prevention. J Prev Med Hyg 2017; 58(2): E79-E92.
2. Sheppard SK, Maiden MC. The evolution of Campylobacter jejuni and Campylobacter coli. Cold Spring Harb Perspect Bio 2015; 7(8): a018119.
3. Nielsen HL, Ejlertsen T, Engberg J, Nielsen H. High incidence of Campylobacter concisus in gastroenteritis in North Jutland, Denmark: a population-based study. Clin Microbiol Infect 2013; 19(5): 445-450.
4. Nielsen HL, Engberg J, Ejlertsen T, Nielsen H. Clinical manifestations of Campylobacter concisus infection in children. Pediatr Infect Dis J 2013; 32(11): 1194-1198.
5. Brunner K, John CM, Phillips NJ, Alber DG, Gemmell MR, Hansen R, Nielsen HL, Hold GL, Baja-Elliott M, Jarvis GA. Novel Campylobacter concisus lipooligosaccharide is a determinant of inflammatory potential and virulence. J Lipid Res 2018; 59(10): 1893-1905.
6. Gemmell MR, Berry S, Mukhopadhya I, Hansen R, Nielsen HL, Bajaj-Elliott M, Nielsen H, Hold GL. Comparative genomics of Campylobacter concisus: analysis of clinical strains reveals genome diversity and pathogenic potential. Emerg Microbes Infect 2018; 7(1): 116.
7. Hsu T, Gemmell MR, Franzosa EA, Berry S, Mukhopadhya I, Hansen R, Michaud M, Nielsen H, Miller WG, Nielsen H, Bajaj-Elliott M, Huttenhower C, Garrett WS, Hold GL. Comparative genomics and genome biology of Campylobacter showae. Emerg Microbes Infect 2019; 8(1): 827-840.
8. Bullman S, Corcoran D, O'Leary J, O'Hare D, Lucey B, Sleator RD. Emerging dynamics of human campylobacteriosis in Southern Ireland. FEMS Immunol Med Microbiol 2011; 63(2): 248-253.
9. Hansen R, Berry SH, Mukhopadhya I, Thomson JM, Saunders KA, Nicholl CE, Bisset WM, Loganathan S, Mahdi G, Kastner-Cole D, Barclay AR, Bishop J, Flynn DM, McGrogan P, Russell RK, El-Omar EM, Hold GL. The microaerophilic microbiota of de-novo paediatric inflammatory bowel disease: the BISCUIT study. PLoS One 2013; 8(3): e58825.
10. Mukhopadhya I, Hansen R, Nicholl CE, Alhaidan YA, Thomson JM, Berry SH, Pattinson C, Stead DA, Russell RK, El-Omar EM, Hold GL. A comprehensive evaluation of colonic mucosal isolates of Sutterella wadsworthensis from inflammatory bowel disease. PLoS One 2011; 6(10): e27076.
11. Coil D, Jospin G, Darling AE. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics 2014; 31(4): 587-589.
12. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 2011; 17(1): 10-12.
13. Antipov D, Hartwick N, Shen M, Raiko M, Lapidus A, Pevzner P. plasmidSPAdes: assembling plasmids from whole genome sequencing data. bioRxiv 2016: 048942.
14. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014; 30(14): 2068-2069.
15. Miller WG, Yee E, On SL, Andersen LP, Bono JL. Complete genome sequence of the Campylobacter ureolyticus clinical isolate RIGS 9880. Genome Announc 2015; 3(6). pii: e01291-15.
16. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 2015; 31(19): 3210-3212.
17. Tay AP, Kaakoush NO, Deshpande NP, Chen Z, Mitchell H, Wilkins MR. Genome sequence of Campylobacter showae UNSWCD, isolated from a patient with Crohn's disease. Genome Announc 2013; 1(1). pii: e00193-12.