MICROBIAL GENOMICS

RESEARCH ARTICLE

Liu *et al.*, *Microbial Genomics* 2020;6

DOI 10.1099/mgen.0.000457

Analysis of complete *Campylobacter concisus* genomes identifies genomospecies features, secretion systems and novel plasmids and their association with severe ulcerative colitis

Fang Liu[1], Siying Chen[1], Laurence Don Wai Luu[1], Seul A. Lee[1], Alfred Chin Yen Tay[2], Ruochen Wu[1], Stephen M. Riordan[3], Ruiting Lan[1], Lu Liu[4] and Li Zhang[1,*]

Abstract

Campylobacter concisus is an emerging enteric pathogen that is associated with several gastrointestinal diseases, such as inflammatory bowel disease (IBD), which includes Crohn's disease (CD) and ulcerative colitis (UC). Currently, only three complete *C. concisus* genomes are available and more complete *C. concisus* genomes are needed in order to better understand the genomic features and pathogenicity of this emerging pathogen. DNA extracted from 22 *C. concisus* strains were subjected to Oxford Nanopore genome sequencing. Complete genome assembly was performed using Nanopore genome data in combination with previously reported short-read Illumina data. Genome features of complete *C. concisus* genomes were analysed using bioinformatic tools. The enteric disease associations of *C. concisus* plasmids were examined using 239 *C. concisus* strains and confirmed using PCRs. Proteomic analysis was used to examine T6SS secreted proteins. We successfully obtained 13 complete *C. concisus* genomes in this study. Analysis of 16 complete *C. concisus* genomes (3 from public databases) identified multiple novel plasmids. pSma1 plasmid was found to be associated with severe UC. Sec-SRP, Tat and T6SS were found to be the main secretion systems in *C. concisus* and proteomic data showed a functional T6SS despite the lack of ClpV. T4SS was found in 25% of complete *C. concisus* genomes. This study also found that GS2 strains had larger genomes and higher GC content than GS1 strains and more often had plasmids. In conclusion, this study provides fundamental genomic data for understanding *C. concisus* plasmids, genomospecies features, evolution, secretion systems and pathogenicity.

DATA SUMMARY

Genome assemblies of 13 *Campylobacter concisus* complete genomes and 7 draft genomes have been deposited in the National Center for Biotechnology Information (NCBI) bacterial genome database. The accession numbers for the genome assemblies are available in Tables 1 and S2 (available in the online version of this article). The assemblies are also available on FigShare (private link: https://figshare.com/s/486820ffc5c4610233e1; DOI: 10.6084/m9.figshare.12279473). Raw data were submitted to the NCBI SRA database under BioProject numbers PRJNA388128 and PRJNA348396.

Received 12 May 2020; Accepted 01 October 2020; Published 28 October 2020

Author affiliations: [1]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, Australia; [2]Helicobacter Research Laboratory, Marshall Centre for Infectious Diseases Research and Training, School of Pathology and Laboratory Medicine, University of Western Australia, Perth, Australia; [3]Gastrointestinal and Liver Unit, Prince of Wales Hospital, University of New South Wales, Sydney, Australia; [4]School of Medical Sciences, University of New South Wales, Sydney, NSW, Australia.

***Correspondence:** Li Zhang, l.zhang@unsw.edu.au

Keywords: *Campylobacter concisus*; *Campylobacters*; inflammatory bowel disease; secretion system; ulcerative colitis.

Abbreviations: *asd*, aspartate-semialdehyde dehydrogenase; CD, Crohn's disease; CDS, coding sequence; GS, genomospecies; HBA, horse blood agar; Hcp, hemolysin-coregulated protein; IBD, inflammatory bowel disease; IPAA, ileal pouch–anal anastomosis; NCBI, National Center for Biotechnology Information; PCR, polymerase chain reaction; Sec, general secretion system; SRP, signal recognition particle; Tat, twin-arginine translocation; T4SS, type IV secretion system; T6SS, type VI secretion system; UC, ulcerative colitis.

Repositories: Genome assembly accession numbers for newly sequenced *C. concisus* complete genomes: CP049274, CP049272-CP049273, CP049270-CP049271, CP049239, CP049238, CP049237, CP049267, CP049266, CP049264-CP049265, CP049275, CP049277, CP049234-CP049236, CP049232-CP049233, and CP049263. Genome assembly accession numbers for newly sequenced *C. concisus* draft genomes: JAAKZE000000000, JAAKZD000000000, JAAKZC000000000, JAAKZB000000000, JAAKZA000000000, JAAKYZ000000000, and JAAKYY000000000.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Three supplementary figures and eight supplementary tables are available with the online version of this article.

INTRODUCTION

Campylobacter concisus is an H_2-requiring Gram-negative bacterium that normally colonizes the human oral cavity [1–3]. Colonization of *C. concisus* at the intestinal tract has been found to be associated with inflammatory bowel disease (IBD) [4–7]. IBD is a chronic inflammatory disease of the gastrointestinal tract, and Crohn's disease (CD) and ulcerative colitis (UC) are the two major forms [8]. In addition to IBD, *C. concisus* has also been linked to other diseases in the gastrointestinal tract, such as diarrhoeic diseases and Barrett's oesophagus [9–12].

Previous studies have shown that the *C. concisus* strains isolated from enteric tissues of patients with IBD originated from their oral cavities [13–15]. *C. concisus* consists of two genomospecies (GS1 and GS2), which differ in their core genomes, housekeeping genes and 23S rRNA gene [16–20]. GS2 *C. concisus* strains were isolated more often from enteric samples of patients with IBD and gastroenteritis, suggesting that they have better abilities than GS1 strains in colonizing the human intestinal tract [21]. Previous studies also demonstrated that some oral GS2 *C. concisus* strains isolated from patients with IBD were invasive to intestinal epithelial cells [15].

The classification of GS alone is unable to differentiate disease-associated virulent *C. concisus* strains from commensal strains. Both GS1 and GS2 consist of *C. concisus* strains isolated from patients with enteric diseases and healthy individuals. Furthermore, some individuals particularly patients with IBD are colonized by multiple different *C. concisus* strains, and while some of these strains may be virulent, the others may be commensal [15]. Given this, molecular markers that can be used to identify *C. concisus* strains that are associated with IBD are needed. We previously sequenced the complete genome of *C. concisus* strain P2CDO4 using PacBio. Analysis of the complete P2CDO4 genome led to the discovery of the pICON plasmid, which further led to the discovery of the *csep1* gene that is associated with CD [22].

Currently there are only three complete *C. concisus* genomes (strains 13826, ATCC 33237 and P2CDO4) available. Strain 13826 was isolated from the faecal sample of a patient with gastroenteritis, strain ATCC 33237 was isolated from the gingival sulcus of an individual with gingivitis and strain P2CDO4 was isolated from the saliva of a patient with relapsed CD [22, 23]. The lack of complete *C. concisus* genomes from patients with UC has limited our ability to potentially identify mobile element-encoded molecular markers that may be associated with this IBD subtype. Furthermore, no complete *C. concisus* genomes from healthy individuals are available, which also limits the comparison of entire genomes between strains isolated from patients with enteric diseases and healthy controls. To address these issues, in this study we used a combination of Illumina and Oxford Nanopore sequencing for hybrid assembly and successfully closed the genomes of 13 *C. concisus* strains isolated from patients with UC, CD and healthy controls. Interestingly, analysis of these complete *C. concisus* genomes resulted in the identification of novel plasmids and their association with severe UC. Furthermore, this study found that GS2 strains had larger genomes and higher GC content than GS1 strains and more often had plasmids . Bacterial secretion systems were also identified in complete *C. concisus* genomes. A type IV secretion system (T4SS) was acquired through different genomic islands and proteomic analysis supports the view that the type VI secretion system (T6SS) identified in *C. concisus* is functional. Furthermore, we were able to compare paired complete genomes of enteric and oral strains.

Impact Statement

Campylobacter concisus is an important emerging enteric pathogen that is associated with inflammatory bowel disease (IBD). IBD is a chronic inflammatory disease of the gastrointestinal tract including both Crohn's disease (CD) and ulcerative colitis (UC). *C. concisus* usually colonizes the human oral cavity as a commensal bacterium. However, its enteric colonization was previously found to be associated with IBD. Strains of *C. concisus* are diverse. Bacterial molecular markers that can differentiate disease-associated virulent *C. concisus* strains from commensal *C. concisus* strains need to be identified. In this study, we successfully sequenced the complete genomes of 13 *C. concisus* strains isolated from patients with IBD and controls. Analysis of fully sequenced *C. concisus* genomes identified genomospecies features, secretion systems and novel plasmids. The novel plasmid pSma1 identified in this study was found to be associated with patients with UC who underwent surgical intervention because of the severity of their disease. This study provides fundamental genomic data for elucidating *C. concisus* adaptation to the enteric environment and enteric pathogenicity. The finding that pSmal1 is associated with severe UC is particularly interesting, suggesting that future studies should be conducted to investigate whether pSmal1 is a disease progression marker and a therapeutic target for UC.

METHODS

Complete genome sequencing of *C. concisus* strains using Oxford Nanopore and hybrid assembly

We previously sequenced 64 draft genomes of *C. concisus* strains isolated from patients with IBD and healthy controls using Illumina MiSeq [19, 22]. In contrast to Illumina sequencing, which generates short reads of less than 300 bp, Oxford Nanopore sequencing produces long-read data, enabling complete genome assembly, particularly when combined with Illumina data [24].

In order to obtain more closed *C. concisus* genomes for analysis, in this study we sequenced the genomes of 22 strains isolated from patients with IBD and healthy controls using Nanopore technology. These strains were chosen randomly from strains whose draft genomes we had previously submitted to public

Table 1. Comparison of the complete *C. concisus* genomes

Strain	Isolation source	Health status	GS	Genome size (bp)	Chromosome size (bp)	Genome GC (%)	Chromosome GC (%)	Core genome GC (%)	No. of plasmids	N50 (kb)	Coverage (×)	CDS	Assembly accession no.
P10CDO-S2*	Oral cavity	CD	1	1932 636	1932636	37.4	37.4	41	–	8.033	109	1881	CP049274
P27CDO-S2*	Oral cavity	CD	1	1834 912	1831320	37.6	37.6	40.9	1	7.325	271	1746	CP049272-CP049273
P26UCO-S2*	Oral cavity	UC	1	1897765	1894099	37.6	37.7	41	1	12.512	36	1842	CP049270-CP049271
P3UCO1*	Oral cavity	UC	1	1800 519	1800519	37.7	37.7	40.9	–	12.107	57	1754	CP049239
P3UCB1*	Intestinal biopsy		1	1831655	1831655	37.6	37.6	40.9	–	10.474	18	1794	CP049238
H1O1*	Oral cavity	Healthy	1	1861371	1861371	37.7	37.7	41	–	6.806	188	1801	CP049237
ATCC33237	Oral cavity	Gingivitis	1	1840041	1840041	37.6	37.6	41	–			1767	CP012541
P1CDO2*	Oral cavity	CD	2	2031332	2031332	39.4	39.4	42.7	–	12.749	11	1933	CP060707
P1CDO3*			2	2051359	2051359	39.4	39.4	42.6	–	6.389	1045	1907	CP049266
P2CDO4	Oral cavity	CD	2	2095236	1975443	39.7	39.7	42.7	1			1971	CP021642-CP021643
P11CDO-S1*	Oral cavity	CD	2	2028838	2025227	39.5	39.5	42.6	1	11.572	38	1908	CP049264-CP049265
P13UCO-S1*	Oral cavity	UC	2	2002415	1998513	39.4	39.4	42.6	1	5.589	47	1913	CP060705 CP060706
P15UCO-S2*	Oral cavity	UC	2	1946644	1943962	39.5	39.5	42.7	2	4.985	421	1844	CP049234-CP049236
H9O-S2*	Oral cavity	Healthy	2	2028901	2025058	39.7	39.7	42.6	1	9.012	109	1894	CP049232-CP049233
H16O-S1*	Oral cavity	Healthy	2	1987364	1987364	39.3	39.3	42.6	–	5.743	126	1888	CP049263
13826	Faeces	Gastroenteritis	2	2099413	2052007	39.4	39.4	42.7	2			2033	CP000792-CP000794

*Complete genome sequenced in this study.
CD, Crohn's disease; CDS, coding sequence; GS, genomospecies; UC, ulcerative colitis.

databases (Table S1, available in the online version of this article) [19, 22].

C. concisus strains were cultured on horse blood agar (HBA) plates and bacterial DNA was extracted with phenol/chloroform followed by ethanol purification as described previously [22]. The 260/280 and 260/230 ratios were determined by Nanodrop. The DNA concentration was quantified with the Qubit dsDNA HS Assay kit (Thermo Fisher Scientific) and measured by a Qubit Fluorometer (Thermo Fisher Scientific). Libraries were prepared with the Nanopore Rapid Barcoding Sequencing kit (SQK-RBK004, Nanopore) according to the manufacturer's instructions. The barcoded library was loaded onto a R9.4 flow cell (FLO-MIN106D, Nanopore) and sequenced on the MinION Sequencing Device (Nanopore) for 48 h. Local basecalling and demultiplexing of fast5 reads were performed using Guppy (v3.2.1). Read statistics were generated using Nanostat (v1.1.2), and genome coverage was determined using Minimap2 (v2.17) and Qualimap (v2.2.1) [25, 26].

For hybrid assembly of complete genomes, Nanopore reads were first filtered with Filtlong (v0.2.0, https://github.com/rrwick/Filtlong) and then assembled with Canu (v1.7.1) and Flye (v2.4.2) [27, 28]. The Filtlong settings utilized were: minimum read length of 1000 bp, keep 90% of the highest quality reads and remove the worst reads until 500 Mbp remained. Nanopore-filtered reads were specified as input for Canu and Flye. Canu/Flye long-read assembly was then used as input into Unicycler (v0.4.7) for hybrid assembly [19, 22, 29]. For strains P1CDO2 and P13UCO-S1, an additional depth filter of 0.35 was applied to remove spurious short linear contigs with low depth. Core genomes were determined using Roary (v3.12.0) [30]. Genomes that revealed a single circularized chromosome contig were defined as closed genomes. Extrachromosomal circularized contigs were defined as plasmids [31].

Draft genome sequencing of seven *C. concisus* strains from healthy individuals

Most of the *C. concisus* draft genomes in the public databases are from patients with IBD and diarrhoeal disease. To increase the number of *C. concisus* genomes of healthy controls, we sequenced the draft genomes of seven *C. concisus* strains isolated from the saliva samples of six healthy adults using Illumina MiSeq; these strains are listed in Table S2 [19, 22].

The criteria for defining a healthy individual and methods for bacterial isolation from saliva samples are described elsewhere [1, 15, 32]. Bacterial DNA extraction and draft genome sequencing were performed according to the procedures described in our previous publications [19, 22]. The sequencing was performed at the University of Western Australia.

The procedures for collecting saliva from healthy individuals were approved by the Ethics Committees at the University of New South Wales and the South East Sydney Local Health District, Australia (HREC/17/POWH/173). A completed consent form regarding saliva collection was obtained from the six healthy adults.

Genome annotation and submission

Genomes were annotated using the National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline [33]. The complete genomes of 13 *C. concisus* strains and the draft genomes of 7 strains sequenced in this study were submitted to the NCBI bacterial genome database. Raw data were submitted under BioProject numbers PRJNA388128 and PRJNA348396.

Analysis of plasmids found in complete *C. concisus* genomes

Our success in completing 13 *C. concisus* genomes in this study provided us with the opportunity to analyse the plasmids in these strains. The presence of plasmids in the 13 complete *C. concisus* genomes was determined using Unicycler and visualized with Bandage [29, 34]. The presence of identified *C. concisus* plasmids in other bacterial species was examined by searching the sequences of their genes and proteins using NCBI nucleotide and protein BLAST, respectively [35].

Secreted proteins and transmembrane proteins encoded by the plasmids were predicted using Phobius [36]. Proteins were also compared with known virulence factors of other bacterial species in the Virulence Factors Database core dataset using protein BLAST, and E values <0.05 were noted [37].

Examination of the association between the newly identified plasmids in *C. concisus* and gastrointestinal diseases

To determine whether the *C. concisus* plasmids found in this study are associated with different types of gastrointestinal disease, their prevalence in a total of 239 strains in the public databases was compared using NCBI nucleotide BLAST [35]. These strains included 58 strains from 29 patients with CD, 62 strains from 28 patients with UC, 64 strains from 42 healthy individuals, 54 strains from 46 patients with diarrhoea and 1 strain from a patient with IBD (unknown if CD or UC) [13, 14, 19, 22, 38–40]. Of the *C. concisus* strains isolated from patients with UC, 40 strains were from 15 patients who have undergone ileal pouch–anal anastomosis (UC-IPAA) due to their severe disease, and they were referred to as severe UC [41].

The remaining six strains in the public databases were excluded from prevalence analysis, including one strain from each patient with gingivitis, pyrexia and exanthema, collagenous colitis, or lymphocytic colitis.

PCR confirmation of pSma1 prevalence in *C. concisus* strains

The above prevalence analysis of newly identified *C. concisus* plasmids in *C. concisus* strains isolated from gastrointestinal diseases and healthy controls identified a plasmid associated with severe UC, which was designated pSma1. The presence of pSma1 in *C. concisus* strains was confirmed by polymerase

chain reaction (PCR) using DNA samples extracted from 71 *C. concisus* strains that are available in our laboratory. Three PCR methods were established, two pairs of primers were designed to specifically target the pSma1 plasmid, and one pair of primers targeting the housekeeping gene aspartate-semialdehyde dehydrogenase (*asd*) was also designed. The PCR reactions amplifying the pSma1 plasmid and the *asd* gene were performed in parallel to ensure that the absence of pSma1 was not due to an inadequate DNA template. The primer sequences and thermocycling conditions are listed in Table S3.

Detection and isolation of pSma1 plasmid

Nanopore sequencing revealed that pSma1 had a high sequencing depth, suggesting high copy numbers. To confirm this, *C. concisus* DNA extracted from strain P15UCO-S2 (a strain that has pSma1) using the phenol/chloroform method was visualized on a 0.8% agarose gel. A band at approximately 1300 bp was excised from the gel, purified using the QIAquick Gel Extraction kit (Qiagen) and subjected to PCRs targeting pSma1 as described above. pSma1 plasmid was also purified from *C. concisus* strain P15UCO-S2 using the Plasmid Miniprep kit (Qiagen). The purified plasmid was also visualized on a 0.8% agarose gel and the band at approximately 1300 bp was excised, purified and subjected to PCRs targeting pSma1.

Examination of the phylogenetic relationship of pSma1 plasmid found in different *C. concisus* strains and in other bacterial species

A phylogenetic tree based on the nucleotide sequences of pSma1 plasmid found in different *C. concisus* strains and in other bacterial species was generated using the maximum-likelihood method implemented in MEGA7 [42]. pSma1-encoded proteins between *C. concisus* strains were compared using MUSCLE [43].

Examination of secretion systems in *C. concisus*

Bacterial secretion systems in *C. concisus* were examined using the 16 complete *C. concisus* genomes. Proteins of different bacterial secretion systems in these *C. concisus* strains were identified using KofamKOALA [44].

Identification of protein secretion through T6SS by proteomic analysis

In order to examine whether *C. concisus* carries a functional T6SS, proteomic analysis was performed to identify secretion of hemolysin-coregulated protein (Hcp), which is a hallmark protein indicating a functional T6SS. *C. concisus* strain P1CDO2, which was found to have T6SS genes in this study, was cultured on HBA plates as described previously [22]. Bacteria were collected and resuspended in heart infusion broth and further cultured for 24 h with constant rotation at 200 r.p.m. [22]. Bacterial culture supernatant was collected by centrifugation and then filtered through a 0.22 μm MILLEX GP filter (Merck Millipore Ltd) to remove residual bacteria. Supernatant was concentrated by 10× using Amicon Ultra 3K columns (Merck Millipore Ltd), loaded onto SDS-polyacrylamide gel, and separated by electrophoresis. Protein bands were excised and digested by trypsin. Digested peptides were separated by liquid chromatography and analysed using a LTQ-FT Ultra mass spectrometer (Thermo Electron) [15]. All tandem mass spectrometer (MS/MS) spectra were searched against the NCBI database using MASCOT (version 2.5.1). Scaffold Q+ (v.4.7.3, Proteome software) was used to validate peptide identities. Mass spectrometry was conducted at the Bioanalytical Mass Spectrometry Facility, University of New South Wales, Australia.

Genome comparison of paired oral and enteric *C. concisus* strains

Complete genomes of *C. concisus* strains P3UCO1 and P3UCB1, paired oral and enteric *C. concisus* strains isolated from the saliva and intestinal biopsies of a patient with UC, were compared using Mauve [45]. Proteins sequences were compared using CLUSTAL Omega [46].

Statistical analysis

Fisher's exact test was used to compare the prevalence of plasmids in different groups of individuals, and in strains isolated from patients and controls. *P* values <0.05 were considered to be statistically significant. Statistical analysis was performed using GraphPad Prism 7.

RESULTS

The complete genomes of GS1 and GS2 *C. concisus* strains differ in some features

Completed genomes were obtained for 13 of the 22 *C. concisus* strains subjected to Nanopore sequencing in this study. Comparison of the features of the 16 complete *C. concisus* genomes (13 from this study and 3 from public database) showed that GS2 strains contained larger genomes than GS1 strains; the genome sizes ranged from 1.95 to 2.10 Mb in GS2 strains and 1.8 to 1.93 Mb in GS1 strains (Table 1). Additionally, GS2 strains had a higher GC content than GS1 strains; the GC content in GS2 strains was 39.4–39.7%, while in GS1 strains it ranged from 37.4–37.7%. Furthermore, the core genome of GS2 strains also had higher GC content than GS1 (42.6–42.7% versus 40.9–41%) (Table 1).

Of the 16 complete *C. concisus* genomes, 8 (50%) contained plasmids. GS2 strains had plasmids more often than GS1 strains; 67% (6/9) of complete GS2 genomes and 29% (2/7) of complete GS1 genomes had plasmids (Table 1).

The accession numbers of complete genomes and draft genomes sequenced in this study are listed in Tables 1 and S2, respectively.

Analysis of novel plasmids in newly complete *C. concisus* genomes

As mentioned above, 6 plasmids were found in 6 of the complete genomes of the 13 *C. concisus* strains sequenced in this study. These plasmids were assigned names and further analysed (Table 2). The plasmids found in strains P26UCO-S2

Table 2. *C. concisus* plasmids identified in this study

Strain	Plasmid	Size (bp)	Depth (×)	GC (%)
P26UCO-S2 P27CDO-S2	pAJ3	3666	50.72	38.4
P11CDO-S1	pSA2	3611	29.12	34.9
P13UCO-S1	pADS1	3902	14.05	35.2
P15UCO-S2	pSma1	1324	64.32	36.9
	pSma2	1358	67.70	35
H9O-S2	pTJ3	3843	14.55	37.9

Sequencing depth was proportional to the chromosomal average. The plasmids found in strains P26UCO-S2 and P27UCO-S2 shared 92% sequence identity, and they were thus considered to be the same plasmid.

and P27UCO-S2 shared 92% sequence identity, and were therefore considered to be the same plasmid. One strain (P15UCO-S2) contained two plasmids. The size of these plasmids varied from 1300 to 3900 bp and their GC content ranged from 34–38%.

The sequencing depth of pSma1 and pSma2 was 64.32 and 67.70×, respectively, relative to that of the chromosome, suggesting that the copy numbers of pSma1 and pSma2 in the prepared DNA sample were between 60 and 70. The remaining four plasmids also had multiple copies, and their depth was between 14.05 and 50.72× (Table 2).

The pSma1 plasmid only encoded two proteins, which were hypothetical proteins with molecular weights of 23.33 (locus tag: G5B99_09655; 199 amino acids) and 7.64 kDa (locus tag: G5B99_09660; 64 amino acids). The pSma2 plasmid only encoded one protein, which was a hypothetical protein with a molecular weight of 17.76 kDa (155 amino acids). The other plasmids encoded three to four proteins. Some of the proteins encoded by these plasmids were secreted proteins (Table S4). No potential virulence factor was found in pSma1, while two potential virulence factors were found in pSA2, but the similarities were low (Table S4).

Of the six plasmids identified in *C. concisus* strains in this study, pSma1 was also found in two strains of *Streptococcus agalactiae* and one strain of *Xanthomonas citri* pv. *citri* in the public databases. The remaining five plasmids were only found in *C. concisus*. The sequence identity of pSma1 shared between *C. concisus* and *S. agalactiae*, *C. concisus* and *X. citri* pv. *citri*, and *S. agalactiae* and *X. citri* pv. *citri* was 98, 97.6 and 98%, respectively.

pSma1 is associated with severe UC

The potential disease association of each of the six plasmids was examined by assessing their prevalence in 239 *C. concisus* strains isolated from patients with different types of gastrointestinal diseases and healthy controls.

The pSma1 plasmid was found in 1 strain from 1 individual with active CD, 12 strains from 4 patients with UC-IPAA, 1 strain from 1 patient with active UC, and 2 strains from 2 healthy controls. The pSma1 plasmid was absent in patients with diarrhoea (Fig. 1a). The prevalence of pSma1 in patients with UC-IPAA was 27% (4/15), which was significantly higher than that in patients with CD (3%, 1/29, P=0.039) and diarrhoea (0/42, P=0.0035) and healthy individuals (5%, 2/42, P=0.036). When assessed for disease associations, the positivity was counted only once in individuals who were colonized by more than one strain that was positive for a particular plasmid.

At the strain level, 30% (12/40) of the *C. concisus* strains isolated from patients with UC-IPAA were positive for the pSma1 plasmid, which was significantly higher than the percentage in strains isolated from patients with CD (2%, 1/58, P=0.0001) and diarrhoea (0/54, P=0.0001) and healthy controls (3%, 2/64, P=0.0002) (Fig. 1b). The prevalence of the pSma1 plasmid in strains isolated from all patients with UC was 21 % (13/62), which was also significantly higher than that in those isolated from patients with diarrhoea (P=0.0002) and healthy individuals (P=0.0021) (Fig. 1b).

PCRs targeting pSma1 were performed on 71 strains available in our laboratory and the prevalence of pSma1 demonstrated by PCR results was consistent with that obtained by BLAST using draft genomes. The other five plasmids did not show disease associations (Table S5). *C. concisus* strains that were positive for plasmids are listed in Table S6.

pSma1 is a small size plasmid with high copy number

The pSma1 had a size of 1324 bp with a relative sequencing depth of 64×, showing that this small plasmid has more than 60 copies. This is supported by the visible DNA band of approximately 1300 bp on a 0.8% agarose gel when the genomic DNA of strain P15UCO-S2 was electrophoresed (Fig. 1d). This band was confirmed to be pSma1 using pSma1-specific PCR (Fig. 1e). The pSma1 plasmid was also purified from *C. concisus* strain P15UCO-S2, which yielded a band at approximately 1300 bp when electrophoresed on agarose gel (Fig. 1f). This band was again confirmed to be the pSma1 plasmid using pSma1-specific PCR (Fig. 1g).

DNA sequences of pSma1 plasmid in *C. concisus* strains isolated from patients with UC are phylogenetically close with unique polymorphisms

Nucleotide sequence alignment of pSma1 plasmid from seven *C. concisus* strains, two *S. agalactiae* strains and one *X. citri* pv. *citri* strain was performed. The pSma1 sequences from strains AAUH-15UCpp, AAUH-16UCo-a, AAUH-16UCf, AAUH-16UCdp, AAUH-16UCdp5, AAUH-20UCo and H9O-S1 were not included because the full-length sequences were unavailable due to draft genomes.

The DNA sequences of pSma1 from seven strains isolated from patients with UC formed one cluster (Fig. 2a), regardless of the geographical differences between the countries

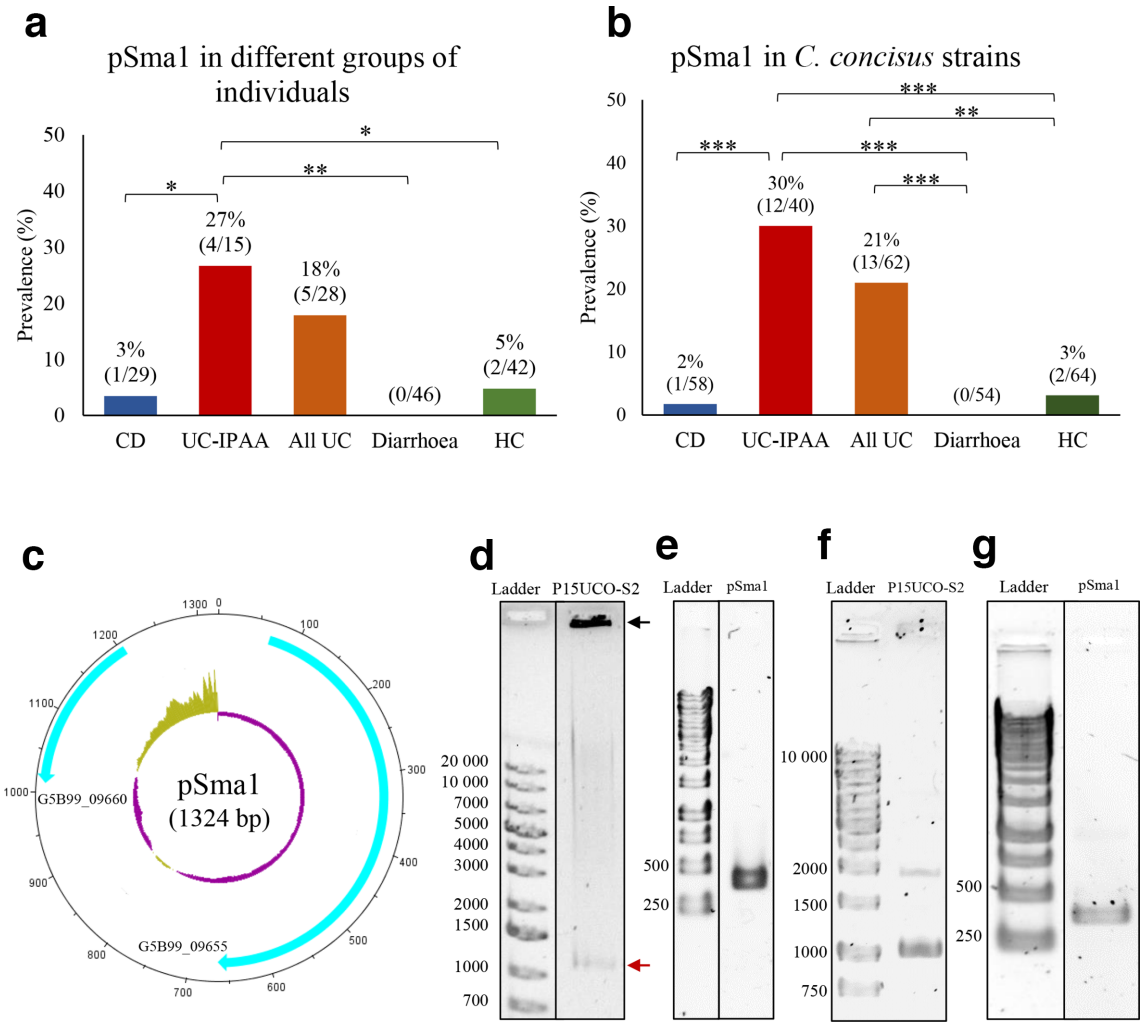


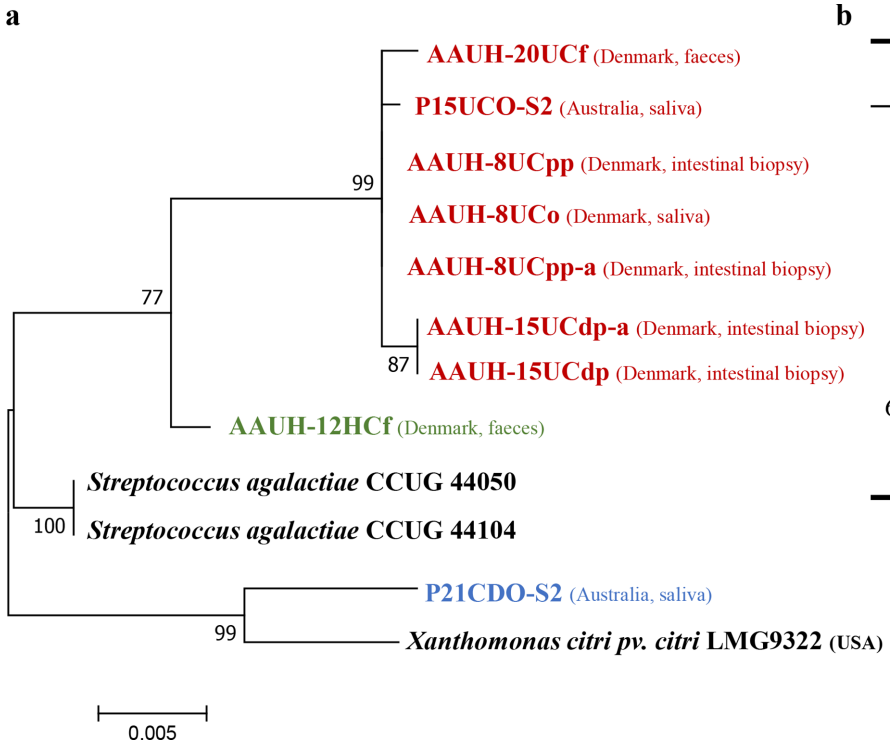
Fig. 1. pSma1 is a small plasmid with a high copy number that is associated with severe ulcerative colitis. (a) The prevalence of pSma1 in patients with UC-IPAA was 27% (4/15), which was statistically higher than that in patients with CD (3%, 1/29, *P*=0.039) and diarrhoea (0/42, *P*=0.0035) and healthy individuals (5%, 2/42, *P*=0.036).(b) The prevalence of *C. concisus* strains isolated from patients with severe UC that were positive for pSma1 was 30% (12/40), which was significantly higher than that in patients with CD (2%, 1/58, *P*=0.0001) and diarrhoea (0/54, *P*=0.0001) and healthy controls (3%, 2/64, *P*=0.0002). (c) Circular representation of pSma1 plasmid. The two hypothetical proteins encoded by pSma1 plasmid are in blue. G+C content below and above average is in purple and yellow, respectively. (d) The pSma1 plasmid was visible at approximately 1300 bp (red arrow) when the genomic DNA of *C. concisus* strain P15UCO-S2 was electrophoresed on 0.8% agarose gel, indicating that it was present at high copy numbers within bacteria cells. The intact genomic DNA is indicated with a black arrow. (e) A PCR targeting pSma1 using DNA extracted from the 1300 bp band on the genomic DNA gel confirmed its identity. (f) The pSma1 plasmid could be purified with a commercially available plasmid extraction kit. (g) A PCR targeting pSma1 using DNA extracted from the 1300 bp band on the purified plasmid gel confirmed its identity. CD, Crohn's disease. UC, ulcerative colitis. IPAA, ileal pouch–anal anastomosis. HC, healthy control. *, statistical significance. **P* <0.05; ***P* <0.005; ****P* <0.001.

of isolation, with five polymorphisms unique to this cluster being identified (Figs 2b and S1).

The amino acid sequences of the two proteins (locus tag G5B99_09655 and G5B99_09660) encoded by pSma1 from *C. concisus*, *S. agalactiae* and *X citri* pv. *citri* were aligned and compared (Figs S2 and S3, respectively). The polymorphic nucleotide at position 658 (Fig. 2b) has resulted in a change to the third amino acid of the protein encoded by locus tag G5B99_09655. The protein encoded by G5B99_09655 from *C. concisus* strains isolated from patients with UC had an asparagine (N) at position 3, while the amino acid was aspartic acid (D) for *C. concisus* strains isolated from healthy individuals or the patient with CD, as well as for *S. agalactiae* strains CCUG44050 and CCUG44104 (Fig. S2). The same protein encoded by *X. citri* pv. *citri* LMG9322 had a shorter length and its first amino acid started from position 4.

C. concisus secretion systems

Examination of the 16 complete *C. concisus* genomes revealed genes encoding proteins from several bacterial secretion



b

No.	UC	Others
410	T	A
473	C	A
596	G	A
617	T	A
658*	T	C/Del
808	T	A

Fig. 2. DNA sequences of pSma1 from *C. concisus* strains isolated from patients with severe ulcerative colitis are phylogenetically distinct with unique polymorphisms. (a) A phylogenetic tree generated based on the nucleotide sequences of pSma1 from *C. concisus*, *S. agalactiae*, and *X. citri* pv. *citri* grouped pSma1 from *C. concisus* strains isolated from patients with UC into a distinct cluster, regardless of the differences between the geographical origins of isolation. (b) Nucleotide sequence comparison of pSma1 from *C. concisus*, *S. agalactiae* and *X. citri* pv. *citri* revealed six polymorphisms unique to pSma1 strains isolated from patients with UC, among which the polymorphism at nucleotide position 658 (*) resulted in a change of amino acid. *C. concisus* strains isolated from patients with UC and CD and healthy controls are in red, blue and green, respectively. The origins of the bacteria strains, where available, are indicated in parentheses. The pSma1 from *C. concisus* strain P15UCO-S2 was used as a reference for nucleotide position. Del, nucleotide deletion. Genomes containing short contigs that partially cover the full length of pSma1 were excluded from sequence comparison.

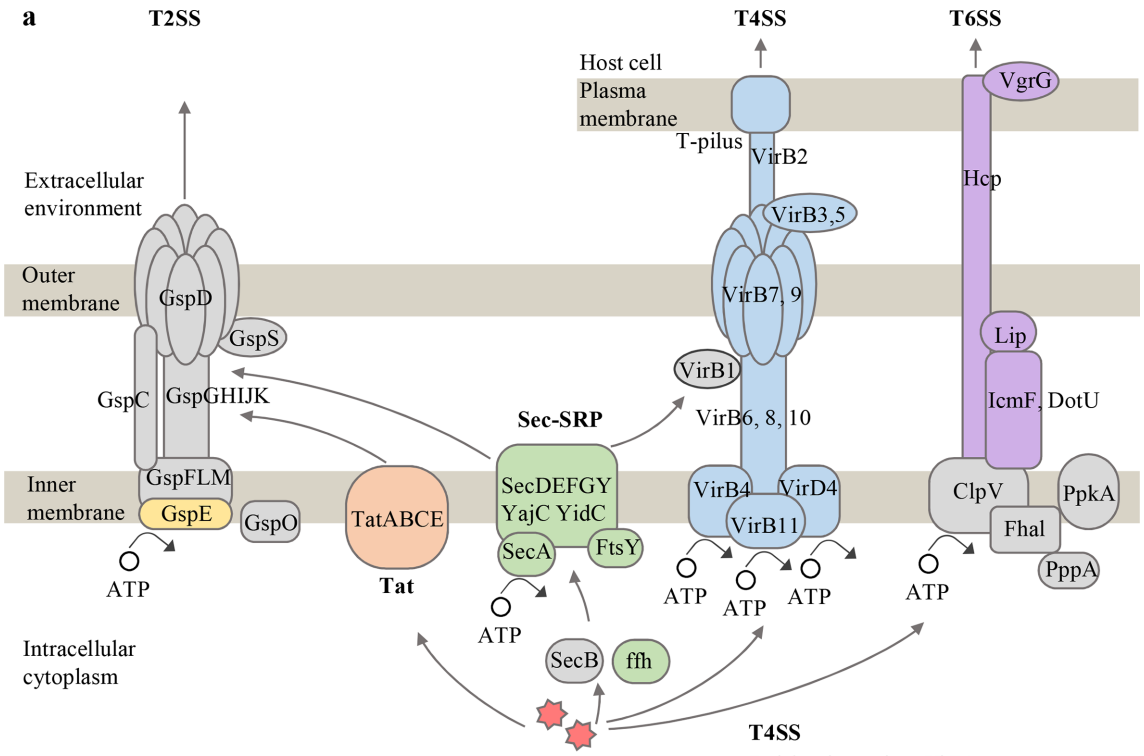
systems, and strain differences were noted for some secretions systems (Fig. 3).

The general secretion system (Sec) has two pathways, the SecA pathway and the signal recognition particle (SRP) pathway. Genes encoding five proteins of the Sec pathways (SecA, Sec D/F, SecE, SecG and SecY) were found in all 16 *C. concisus* strains, but the gene encoding SecB (the chaperone of the Sec system) was absent. Genes encoding the four proteins of the SRP pathway, FtsY, ffh, YajC and YidC, were found in all *C. concisus* strains. Genes encoding the 3 essential components of the twin-arginine translocation (Tat) system, TatA, TatB and TatC, were found in all 16 *C. concisus* strains (Fig. 3).

Genes encoding components of the T2SS were mostly absent in *C. concisus*, with only GspE (an ATPase) being found in 7 of the 16 *C. concisus* strains.

Four of the 16 strains (25%) had genes encoding components of the T4SS. Strains P3UCB1 and P13UCO-S1 contained seven genes encoding the T4SS components, which were located within a previously reported genomic island, CON_PiiA. Strains P1CDO2 and H16O-S1 had nine genes encoding T4SS components, and these genes were located within a novel genomic island identified in this study. We designated this novel genomic island CON_PiiC (Fig. 4). The size of CON_PiiC was 42775 bp, beginning with an integrase. CON_PiiC was located at different positions within the genomes of P1CDO2 and H16O-S1, between 1339776–1382550 bp and 578884–622435 bp, respectively, but the immediate upstream and downstream genes were similar. The T4SS proteins in CON_PiiC were found to share homology with the T4SS proteins from *Brucella melitensis* bv. 1 (Table S7). Additionally, other proteins encoded by CON_PiiC were found to share homology with virulence factors from other bacterial species, such as M protein type 1 from *Streptococcus pyogenes* and toxin B from *Clostridium difficile*.

A T6SS system was found in 14 out of the 16 strains (87.5%, 14/16). These strains had genes encoding Hcp, Lip and IcmF. Strains P3UCO1 and P3UCB1 only had the gene encoding for one T6SS system component. However, genes encoding the ATPase ClpV and several regulatory proteins were not found in any of these *C. concisus* strains (Fig. 3b).



b

Secretion system	Gene	P10CDO-S2; P27CDO-S2; P15UCO-S2 H1O1; ATCC 33237	P1CDO3; P2CDO4; P11CDO-S1; H9O-S2; 13826	P26UCO-S2	P3UCO1	P3UCB1	P13UCO-S1	P1CDO2	H16O-S1
T2SS	GspE								
Tat	TatA; TatB; TatC								
Sec-SRP	SecD/F; SecE; SecG; SecY; YajC; YidC; SecA; FtsY; ffh								
T4SS	VirB2								
	VirB5								
	VirB4, VirB9, VirB11								
	VirB6								
	VirB8; VirB10; VirD4								
T6SS	VgrG								
	Hcp; Lip; IcmF;								
	DotU								

Fig. 3. Secretion systems identified in *C. concisus* complete genomes. (a) The components of different secretion systems that are present in *C. concisus* are highlighted with different colours. The components that are not present in *C. concisus* are in grey. (b) A summary of the secretion systems found in complete *C. concisus* genomes; the colours match the highlighted components in (a). The Sec–SRP and Tat pathways were found in all 16 complete *C. concisus* genomes. The T4SS system was found in 25% of *C. concisus* genomes and the T6SS system was found in 87.5% of *C. concisus* strains. The figure is illustrated according to the bacterial secretion system (map03070) in the KofamKOALA database. OMP, outer-membrane protein. IMP, inner-membrane protein.

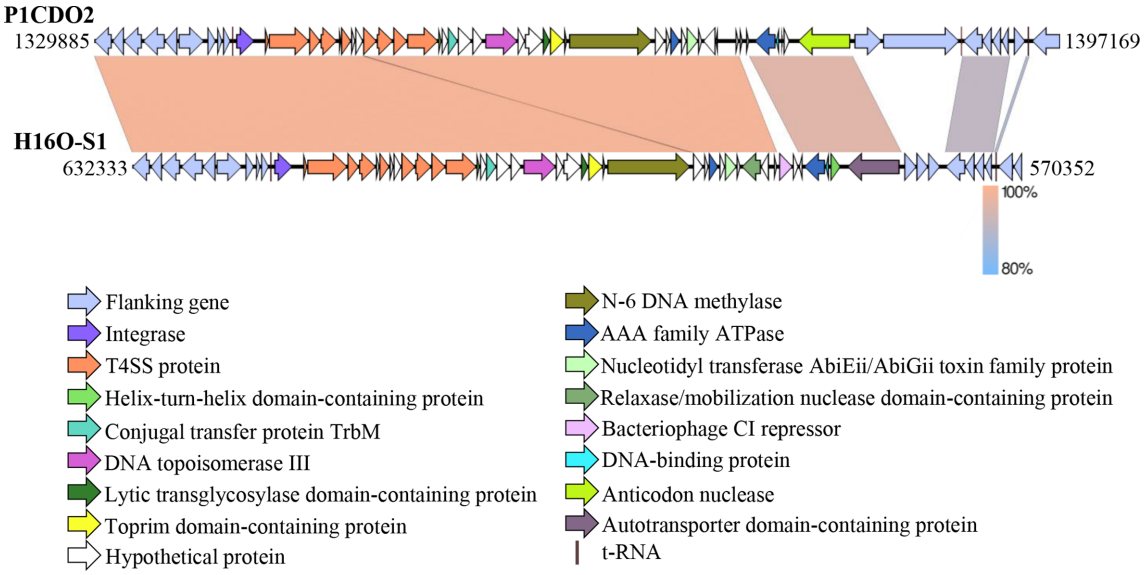


Fig. 4. Genomic island CON_PiiC contains the T4SS system. CON_PiiC, identified in this study, contains genes encoding T4SS components in *C. concisus* strains P1CDO2 and H16O-S1. Components of the T4SS are shaded in orange. CON_PiiC in these two strains shared more than 80% nucleotide identity. The image was generated using EasyFig [60]. Annotations of the proteins encoded byCON_PiiC are in Table S7.

Genes encoding components of T6SS were scattered within the genome. Components of T1SS, T3SS and T5SS were not found in *C. concisus* genomes.

Mass spectrometry identified that the secreted Hcp protein was present in the bacterial culture supernatant of *C. concisus* strain P1CDO2 (identified peptides VHKPFSFTTSLNK and VKGSTQGLISSGASTEASIGNR), supporting the view that T6SS is functional in *C. concisus*.

Comparison of a paired oral and enteric strains isolated from an individual with UC

In this study we were able to compare the complete genome of the enteric strain P3UCB1 isolated from intestinal biopsies of a patient with UC with the oral strain P3UCO1 isolated from the same individual. We found that the two strains were almost identical, except for the CON_PiiA genomic island and another five proteins (Table S8). The CON_PiiA genomic island was present in the enteric strain P3UCB1 but not the oral strain P3UCO1. The CON_PiiA island was previously reported by us [19]. The proteins encoded by G5B98_00305 in strain P3UCO1 and G5B97_00305 in strain P3UCB1 were both annotated as retention module-containing protein, but they only shared 34% sequence identity. Additionally, the protein encoded by G5B98_00315 in P3UCO1 was split into two proteins in strain P3UCB1 (G5B97_00315 and G5B97_00320); this was also seen for the protein encoded by G5B98_01840 in strain P3UCO1, which was divided into two proteins in P3UCB1 (G5B97_02050 and G5B97_02055). Furthermore, there were two small hypothetical proteins in P3UCO1 (G5B98_06400 and G5B98_09010); their corresponding sequences in strain P3UCB1 were non-coding sequences (Table S8).

DISCUSSION AND CONCLUSION

In this study we successfully sequenced the complete genomes of 13 *C. concisus* strains. Analysis of 16 complete *C. concisus* genomes, namely 13 complete *C. concisus* genomes obtained from this study and 3 additional complete *C. concisus* genomes from public databases, led to the identification of multiple novel *C. concisus* plasmids. Further analysis of the prevalence of these *C. concisus* plasmids in 239 *C. concisus* genomes in public databases revealed that the pSma1 plasmid is associated with severe UC. We also analysed the *C. concisus* secretion systems and genomospecies features and compared the complete genomes of paired oral and enteric strains.

Multiple novel plasmids were identified in *C. concisus* strains. The larger numbers of complete *C. concisus* genomes obtained in this study led to the identification of six plasmids that have not been previously described (Table 2). The plasmids found in this study were mostly small with multiple copy numbers. The sizes of naturally occurring plasmids vary substantially, from a few hundred bp to hundreds of kb [47]. The three previously reported *C. concisus* plasmids, pCCON16 and pCCON31 from strain 13826 and pICON from strain P2CDO4 had sizes of 16457, 30949 and 119793 bp, respectively [22]. The plasmids found in this study were small plasmids, ranging from 1324 to 3902 bp. The currently available bioinformatic tools were unable to predict the replication origin of these small plasmids. Naturally occurring plasmids normally have copy numbers between 1 and 30, although they may reach

200 [47]. The plasmids that we identified in this study had a relative sequencing depth of 14.05 to 67.70×, showing that these plasmids are present as multiple copies in their host *C. concisus* strains. The pSma1 plasmid had copy numbers in excess of 60. Both the sequencing data and the electrophoresis consistently showed that this plasmid naturally had high copy numbers (Fig. 1d). We also purified pSma1 plasmid and verified it using PCR.

The plasmids were more frequently present in *C. concisus* GS2 strains than GS1 strains. Of the 16 complete sequenced *C. concisus* genomes, we found that 67% (6/9) of GS2 strains and 29% (2/7) of GS1 strains had plasmids. Plasmids usually confer bacteria with superior adaptation and resistance to adverse environmental factors [48, 49]. Previous studies found that *C. concisus* strains isolated from enteric samples were predominately GS2 strains [15, 21]. It is possible that plasmids in GS2 strains have contributed to the superior ability of these strains in adapting to the enteric environment, although the exact mechanisms require further investigation. pSma1 plasmid was found to be associated with severe UC. Of the six *C. concisus* plasmids identified in this study, only pSma1 was found to be associated with patients with UC who have had IPAA due to severe UC. This interesting finding suggests that future studies should be conducted to evaluate the potential use of pSma1 as a marker for predicting disease prognosis in patients with UC by examining a larger group of patients and controls. Currently, it is not clear whether pSma1 contributes directly to UC pathology or is merely a potential surrogate disease severity marker. The pSma1 plasmid only encodes two hypothetical proteins, which are not similar to known bacterial virulence factors (Table S4). It is well known that bacterial pathogens can acquire virulence factors from plasmids [50]. Further investigation is required to determine whether pSma1-encoded proteins are novel virulence proteins. Another possibility is that the high-copy-number pSma1 plasmid DNA is a virulence factor. Unmethylated CpG motifs are prevalent in bacteria and are known to induce innate immune responses through interaction with Toll-like receptor 9 expressed by immune cells such as macrophages and dendritic cells [51]. Further studies are needed to examine whether pSma1 plasmid can be transported to host cells, for example through bacterial outer-membrane vesicles, in order to shed light on the potential pathogenic role of pSma1.

The pSma1 plasmids found in *C. concisus* strains were more closed related than the pSma1 found in the *S. agalactiae* and *X. citri* pv. *citri* strains. The pSma1 cluster from patients with UC was not due to geographical origin, as there were patients from Denmark and Australia in the same cluster (Fig. 2a). Six nucleotide polymorphisms were found to be specific to pSma1 in *C. concisus* strains isolated from patients with UC and one nucleotide polymorphism resulted in a change of amino acid. These results show that pSma1-associated UC could potentially be distinguished from the others based on its phylogenetic grouping and nucleotide sequences. Nevertheless, the presence of pSma1 in other bacterial species is rare; only two strains *S. agalactiae* and one strain of *X. citri* were found to be positive for pSma1. *S. agalactiae* is a Gram-positive anaerobe and is the only species that belongs to group B *Streptococcus*. *S. agalactiae* can induce severe invasive infection in newborns [52]. *X. citri* is a Gram-negative aerobe and a phytopathogen causing citrus canker [53].

Secretion systems were identified in complete *C. concisus* genomes. The complete *C. concisus* genomes provided us with the opportunity to accurately identify the components of bacterial secretion systems in *C. concisus*. Bacteria use the general secretion systems (sec-SRP) and the Tat system to translocate unfolded and folded proteins, respectively [54]. These systems were found in all 16 complete *C. concisus* genomes, showing that they are ubiquitously carried by *C. concisus* strains. Some *C. concisus* strains (25%) had the T4SS secretion system and, interestingly, they were acquired from different genomic islands. Many human pathogens, such as *Helicobacter pylori*, *Legionella pneumophila* and *Brucella* spp., use T4SS to deliver virulence factors into host cells. The potential pathogenic role of the T4SS system in *C. concisus* requires further investigation. The majority of the complete *C. concisus* genomes carried genes encoding five components of the T6SS, but none of them had the gene encoding for the T6SS-associated ATPase ClpV. Despite the lack of the gene encoding for ClpV, the T6SS may be able to use other ATPases to perform the secretion function. In order to confirm this, we performed proteomics analysis of the supernatant from strain P1CDO2 in which the Hcp protein was detected. Hcp is a structural protein of the T6SS and is also secreted by the T6SS. Secreted Hcp has been used as a hallmark to indicate functional T6SS [55–58]. *Campylobacter jejuni* also has a functional T6SS without ClpV, suggesting that it may be common for *Campylobacter* species to use alternative ATPase instead of ClpV in performing T6SS function [55].

This study also compared the complete genomes of paired oral and enteric *C. concisus* strains. Previous studies suggested that *C. concisus* strains isolated from enteric tissues originate from the oral cavity [15]. However, no studies have examined the complete genomes of paired oral and enteric strains, which may provide insights into genes that may be important for *C. concisus* colonization at different niches of the human gastrointestinal tract. The complete genomes of P3UCO1 and P3UCB1, paired oral and enteric strains from a patient with UC, were indeed highly similar, further supporting the view that translocation of *C. concisus* from the oral cavity to the intestinal tract may cause enteric diseases [59]. The main difference is that the strain isolated from intestinal biopsies (P3UCB1) had the CON_PiiA genomic island, which was missing in the oral P3UCO1 strain. The potential role of the CON_PiiA genomic island in facilitating enteric colonization remains to be investigated.

Another finding from this study is that GS2 strains have larger genomes and higher GC content than GS1 strains. All complete genomes of GS2 *C. concisus* strains were larger than the complete genomes of GS1 strains (Table 1). Although more complete GS2 *C. concisus* strains have plasmids, these were not the reason for the larger genomes of GS2 strains, as the GS2 strains without plasmids also had a genome larger

than the GS1 strain (Table 1). The larger genomes in GS2 strains result from an increased number of coding sequences (CDS), and an increased number of CDS in GS2 strains was also reported in a previous study by Kirk *et al*. using draft *C. concisus* genomes [41]. Additionally, analysis of the complete genomes showed that both the compete genomes and the core genome of GS2 strains had higher GC content than GS1 strains (Table 1), suggesting that GS1 and GS2 are evolving divergently.

In summary, in this study we sequenced the complete genomes of 13 *C. concisus* strains and identified multiple novel plasmids. pSma1 plasmid was found to be associated with severe UC. We also showed that sec-SRP, Tat and T6SS are the main secretion systems in *C. concisus* and provided proteomic analysis data showing that T6SS is functional despite the lack of a gene encoding for ClpV. The T4SS system was found in 25% of *C. concisus* strains, acquired from different genomic islands. Analysis of paired complete oral and enteric *C. concisus* strains showed that they are highly similar, with the main difference being that the enteric strain had the CON_PiiA genomic island. Furthermore, this study found that GS2 strains had larger genomes and higher GC content than GS1 strains and more often had plasmids. In conclusion, this study provides fundamental genomic data for understanding *C. concisus* plasmids, evolution, genomospecies features, secretion systems and pathogenicity.

Data availability statement

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary material.

Funding information
This work is supported by a Faculty Research Grant awarded to L. Z. from the University of New South Wales (grant number PS46772).

Author contributions
F. L., S. C., L. D. W. L., S. A. L. and R. W. performed Oxford Nanopore sequencing. F. L. and S. C. performed bioinformatic analysis. S. C. performed bacterial isolation from saliva samples. S. C., F. L. and R. W. performed bacterial culture and DNA extraction. F. L. and S. C. performed PCR experiments. A. C. Y. T. performed MiSeq sequencing. F. L. and L. Z. wrote the manuscript. All authors have read the manuscript. R. L., L. D. W. L., A. C. Y. T., S. M. R. and L. L. provided critical feedback and helped in editing the manuscript. L. Z. and S. M. R. conceived the project.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement
Procedures for saliva collection from healthy individuals were approved by the Ethics Committees of the University of New South Wales and the South East Sydney Local Health District, Australia (HREC/17/POWH/173). A completed consent form regarding saliva collection was obtained from the six healthy adults.

References

1. Zhang L, Budiman V, Day AS, Mitchell H, Lemberg DA *et al*. Isolation and detection of *Campylobacter concisus* from saliva of healthy individuals and patients with inflammatory bowel disease. *J Clin Microbiol* 2010;48:2965–2967.
2. Yeow M, Liu F, Ma R, Williams TJ, Riordan SM *et al*. Analyses of energy metabolism and stress defence provide insights into *Campylobacter concisus* growth and pathogenicity. *Gut Pathog* 2020;12:13.
3. Lee H, Ma R, Grimm MC, Riordan SM, Lan R *et al*. Examination of the anaerobic growth of *Campylobacter concisus* strains. *Int J Microbiol* 2014;2014:476047
4. Zhang L, Man SM, Day AS, Leach ST, Lemberg DA *et al*. Detection and isolation of *Campylobacter* species other than *C. jejuni* from children with Crohn's disease. *J Clin Microbiol* 2009;47:453–455.
5. Mukhopadhya I, Thomson JM, Hansen R, Berry SH, El-Omar EM *et al*. Detection of *Campylobacter concisus* and other *Campylobacter* species in colonic biopsies from adults with ulcerative colitis. *PLoS One* 2011;6:e21490-e.
6. Mahendran V, Riordan SM, Grimm MC, Tran TAT, Major J *et al*. Prevalence of *Campylobacter* species in adult Crohn's disease and the preferential colonization sites of *Campylobacter* species in the human intestine. *PLoS One* 2011;6:e25417.
7. Kirk KF, Nielsen HL, Thorlacius-Ussing O, Nielsen H. Optimized cultivation of *Campylobacter concisus* from gut mucosal biopsies in inflammatory bowel disease. *Gut Pathog* 2016;8:27.
8. Chandan JS, Thomas T. Inflammatory bowel disease and oral health. *BDJ Team* 2017;4:17083.
9. Lindblom GB, Sjögren E, Hansson-Westerberg J, Kaijser B. *Campylobacter upsaliensis*, *C. sputorum sputorum* and *C. concisus* as common causes of diarrhoea in Swedish children. *Scand J Infect Dis* 1995;27:187–188.
10. Lastovica AJ, le Roux E. Efficient isolation of *Campylobacteria* from stools. *J Clin Microbiol* 2000;38:2798–2800.
11. 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;19:445–450.
12. Macfarlane S, Furrie E, Macfarlane GT, Dillon JF. Microbial colonization of the upper gastrointestinal tract in patients with Barrett's esophagus. *Clin Infect Dis* 2007;45:29–38.
13. Kirk KF, Méric G, Nielsen HL, Pascoe B, Sheppard SK *et al*. Molecular epidemiology and comparative genomics of *Campylobacter concisus* strains from saliva, faeces and gut mucosal biopsies in inflammatory bowel disease. *Sci Rep* 2018;8:1902.
14. Gemmell MR, Berry S, Mukhopadhya I, Hansen R, Nielsen HL *et al*. Comparative genomics of *Campylobacter concisus*: Analysis of clinical strains reveals genome diversity and pathogenic potential. *Emerg Microbes Infect* 2018;7:1–17.
15. Ismail Y, Mahendran V, Octavia S, Day AS, Riordan SM *et al*. Investigation of the enteric pathogenic potential of oral *Campylobacter concisus* strains isolated from patients with inflammatory bowel disease. *PLoS One* 2012;7:e38217.
16. Istivan T. *Molecular Characterisation of Campylobacter concisus: a Potential Etiological Agent of Gastroenteritis in Children*. School of Applied Sciences, RMIT University; 2005.
17. Miller WG, Chapman MH, Yee E, On SLW, McNulty DK *et al*. Multilocus sequence typing methods for the emerging *Campylobacter* Species *C. hyointestinalis*, *C. lanienae*, *C. sputorum*, *C. concisus*, and *C. curvus*. *Front Cell Infect Microbiol* 2012;2:45.
18. Mahendran V, Octavia S, Demirbas OF, Sabrina S, Ma R *et al*. Delineation of genetic relatedness and population structure of oral and enteric *Campylobacter concisus* strains by analysis of housekeeping genes. *Microbiology* 2015;161:1600–1612.
19. Chung HKL, Tay A, Octavia S, Chen J, Liu F *et al*. Genome analysis of *Campylobacter concisus* strains from patients with inflammatory bowel disease and gastroenteritis provides new insights into pathogenicity. *Sci Rep* 2016;6:38442.
20. Nielsen HL, Nielsen H, Torpdahl M. Multilocus sequence typing of *Campylobacter concisus* from Danish diarrheic patients. *Gut Pathog* 2016;8:44.
21. Wang Y, Liu F, Zhang X, Chung HKL, Riordan SM *et al*. *Campylobacter concisus* genomospecies 2 is better adapted to the human gastrointestinal tract as compared with *Campylobacter concisus* genomospecies 1. *Front Physiol* 2017;8:543.

22. Liu F, Ma R, Tay CYA, Octavia S, Lan R *et al*. Genomic analysis of oral *Campylobacter concisus* strains identified a potential bacterial molecular marker associated with active Crohn's disease. *Emerg Microbes Infect* 2018;7:1–14.
23. Tanner ACR, Badger S, Lai C-H, Listgarten MA, Visconti RA *et al*. *Wolinella* gen. nov., *Wolinella succinogenes* (*Vibrio succinogenes* Wolin *et al*.) comb. nov., and description of *Bacteroides gracilis* sp. nov., *Wolinella recta* sp. nov., *Campylobacter concisus* sp. nov., and Eikenella corrodens from humans with periodontal disease. *Int J Syst Bacteriol* 1981;31:432–445.
24. Lu H, Giordano F, Ning Z. Oxford nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics* 2016;14:265–279.
25. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–3100.
26. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 2016;32:btv566–4.
27. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH *et al*. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 2017;27:722–736.
28. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–546.
29. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.
30. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S *et al*. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691–3693.
31. Frost LS, Leplae R, Summers AO, Toussaint A. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* 2005;3:722–732.
32. Mahendran V, Tan YS, Riordan SM, Grimm MC, Day AS *et al*. The prevalence and polymorphisms of zonula occluden toxin gene in multiple *Campylobacter concisus* strains isolated from saliva of patients with inflammatory bowel disease and controls. *PLoS One* 2013;8:e75525.
33. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP *et al*. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 2016;44:6614–6624.
34. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 2015;31:3350–3352.
35. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S *et al*. NCBI BLAST: a better web interface. *Nucleic Acids Res* 2008;36:W5–W9.
36. Käll L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Res* 2007;35:W429–W432.
37. Chen L, Zheng D, Liu B, Yang J, VFDB JQ. Hierarchical and refined dataset for big data analysis-10 years on. *Nucleic Acids Res* 2016;2016:D694–697.
38. Deshpande NP, Kaakoush NO, Wilkins MR, Mitchell HM. Comparative genomics of *Campylobacter concisus* isolates reveals genetic diversity and provides insights into disease association. *BMC Genomics* 2013;14:585.
39. Huq M, Van TTH, Gurtler V, Elshagmani E, Allemailem KS *et al*. The ribosomal RNA operon (rrn) of *Campylobacter concisus* supports molecular typing to genomospecies level. *Gene Rep* 2017;6:8–14.
40. Cornelius AJ, Miller WG, Lastovica AJ, On SLW, French NP *et al*. Complete genome sequence of *Campylobacter concisus* ATCC 33237^T and draft genome Sequences for an additional eight well-characterized *C. concisus* strains. *Genome Announc* 2017;5:e00711–00717.
41. Kirk KF, Méric G, Nielsen HL, Pascoe B, Sheppard SK *et al*. Molecular epidemiology and comparative genomics of *Campylobacter concisus* strains from saliva, faeces and gut mucosal biopsies in inflammatory bowel disease. *Sci Rep* 2018;8:1902.
42. Kumar S, Stecher G, Tamura K. mega7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016;33:1870–1874.
43. Edgar RC. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792–1797.
44. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M *et al*. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *bioRxiv* 2019;602110.
45. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004;14:1394–1403.
46. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. *Mol Syst Biol* 2011;7:539.
47. Espinosa M, Cohen S, Couturier M, Del Solar G, Diaz-Orejas R *et al*. Plasmid replication and copy number control. *The horizontal gene pool: bacterial plasmids and gene spread* 2000:1–47.
48. Utter B, Deutsch DR, Schuch R, Winer BY, Verratti K *et al*. Beyond the chromosome: the prevalence of unique extra-chromosomal bacteriophages with integrated virulence genes in pathogenic *Staphylococcus aureus*. *PLoS One* 2014;9:e100502.
49. Wu Z, Sahin O, Shen Z, Liu P, Miller WG *et al*. Multi-omics approaches to deciphering a hypervirulent strain of *Campylobacter jejuni*. *Genome Biol Evol* 2013;5:2217–2230.
50. Zong Z. Complete sequence of pJIE186-2, a plasmid carrying multiple virulence factors from a sequence type 131 *Escherichia coli* O25 strain. *Antimicrob Agents Chemother* 2013;57:597–600.
51. Cornélie S, Hoebeke J, Schacht A-M, Bertin B, Vicogne J *et al*. Direct evidence that Toll-like receptor 9 (TLR9) functionally binds plasmid DNA by specific cytosine-phosphate-guanine motif recognition. *J Biol Chem* 2004;279:15124–15129.
52. Murayama SY, Seki C, Sakata H, Sunaoshi K, Nakayama E *et al*. Capsular type and antibiotic resistance in *Streptococcus agalactiae* isolates from patients, ranging from newborns to the elderly, with invasive infections. *Antimicrob Agents Chemother* 2009;53:2650–2653.
53. Brunings AM, Gabriel DW. *Xanthomonas citri*: breaking the surface. *Mol Plant Pathol* 2003;4:141–157.
54. Green ER, Mecsas J. Bacterial secretion systems: an overview. *Microbiol Spectr* 2016;4:213–239.
55. Bleumink-Pluym NMC, van Alphen LB, Bouwman LI, Wösten MMSM, van Putten JPM. Identification of a functional type VI secretion system in *Campylobacter jejuni* conferring capsule polysaccharide sensitive cytotoxicity. *PLoS Pathog* 2013;9:e1003393.
56. Ugarte-Ruiz M, Stabler RA, Domínguez L, Porrero MC, Wren BW *et al*. Prevalence of type VI secretion system in Spanish *Campylobacter jejuni* isolates. *Zoonoses Public Health* 2015;62:497–500.
57. Bartonickova L, Sterzenbach T, Nell S, Kops F, Schulze J *et al*. Hcp and VgrG 1 are secreted components of the *Helicobacter hepaticus* type VI secretion system and VgrG 1 increases the bacterial colitogenic potential. *Cell Microbiol* 2013;15:992–1011.
58. Noreen Z, Jobichen C, Abbasi R, Seetharaman J, Sivaraman J *et al*. Structural basis for the pathogenesis of *Campylobacter jejuni* Hcp1, a structural and effector protein of the Type VI Secretion System. *Febs J* 2018;285:4060–4070.
59. Zhang L. Oral *Campylobacter* species: Initiators of a subgroup of inflammatory bowel disease? *World J Gastroenterol* 2015;21:9239–9244.
60. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics* 2011;27:1009–1010.