Genomic analysis identifies *Campylobacter concisus* genomospecies 2 as a novel species and proposes the name *Campylobacter oralis*

Fang Liu[1#], Siying Chen[1#], Joanna M Biazik[2], Christopher Luk[1], Anjaneyaswamy Ravipati[2], Ruiting Lan[1], Stephen M Riordan[3], Li Zhang[1]*

1. School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW 2052, Australia
2. Mark Wainwright Analytical Centre, University of New South Wales, Sydney, NSW 2052, Australia
3. Gastrointestinal and Liver Unit, Prince of Wales Hospital, University of New South Wales, Sydney, NSW 2031, Australia

* Correspondence author: L.Zhang@unsw.edu.au

\# Co-first authors

ABSTRACT

Campylobacter concisus comprises two distinct genomospecies (GS) based on core genome analysis. Here we conducted further analysis and demonstrated that *C. concisus* GS2 strains belong to a novel *Campylobacter* species.

A total of 245 *C. concisus* genomes including 85 GS1 and 160 GS2 strains were analysed. DNA-DNA hybridization (DDH) values between strains of GS1 and GS2 *C. concisus* were from 42.8% to 66%. Average Nucleotide Identity (ANI) between GS1 and GS2 *C. concisus* strains were from 87.8% to 89.7%. The average GC contents of GS1 and GS2 *C. concisus* strains were 37.3% and 39.2% respectively. GS1 and GS2 specific genes have been identified. GS1 and GS2 *C. concisus* strains can also be differentiated based on core genome, 23S rRNA polymorphisms and protein profiles. Their morphology appears different under certain conditions.

In conclusion, ANI and DDH values clearly support that *C. concisus* GS2 as a novel *Campylobacter* species, for which we propose the name *Campylobacter oralis.* The identification of *C. concisus* and *C. oralis* specific genes, along with differences in their protein profiles, morphology, core genome and 23S rRNA phylogeny, further support this classification.

INTRODUCTION

Campylobacter concisus was first isolated by Tanner et al. from human dental specimen in 1981, with strain ATCC 33237 as the type strain [1]. Later studies isolated *C. concisus* from fecal samples and biopsies of patients with gastritis, inflammatory bowel disease (IBD) which includes Crohn's disease and ulcerative colitis, and other diseases [2-8]. In 2016, Chung et al. conducted a comparative genome analysis of 36 *C. concisus* strains and reported that *C. concisus* contained two GS based on the core genome and identified GS specific genes [9]. Based on the core genome, the *C. concisus* reference strain ATCC33237 is a GS1 strain [9]. Kirk et al. in 2018 examined 113 *C. concisus* genomes and reported that the genomes of GS2 *C. concisus* strains contained higher numbers of genes compared to genomes of GS1 *C. concisus* strains [10]. In 2020, Liu et al. analysed 16 complete sequenced *C. concisus* genomes and reported that the genomes of GS2 *C. concisus* strains had larger sizes and higher GC contents than the genomes of GS1 *C. concisus* strains [11]. In 2021, Cornelius et al. analysed 190 *C. concisus* genomes, also confirmed the existence of GS1 and GS2 *C. concisus* [12].

We hypothesize that GS2 *C. concisus* is a novel species and conducted further analysis of 245 *C. concisus* genomes and also investigated methods that may phenotypically differentiate *C. concisus* GS1 and GS2 strains. Our data support that the GS2 *C. concisus* is a novel *Campylobacter* species, and we propose the name *Campylobacter oralis*. We also examined the presence of *csep1* gene, a molecular marker that is associated with active Crohn's disease and found that one of these markers, *csep1* gene, was only present in *C. oralis,* not in *C. concisus*.

MATERIALS AND METHODS

Analysis of genomes of *C. concisus* GS1 and GS2 strains

67 The genomes of 245 *C. concisus* strains, including 85 GS1 and 160 GS2 strains, were obtained
68 from National Center for Biotechnology Information (NCBI) genome database for analysis
69 (Supplementary Table 1).
70
71 ***Phylogenetic analysis based on core genome, 23S and 16S rRNA genes***
72 The core genome of the 245 *C. concisus* strains was generated by using Roary (v3.12.0) without
73 splitting paralogs [13]. The phylogenetic tree of core genome was generated by using FastTree
74 (v2.1.10) based on the approximate-maximum-likelihood method [14]. The sequences of 16S
75 and 23S rRNA gene were used to generate phylogenetic trees based on the maximum-
76 likelihood method (100 bootstrap) by using MEGAcc (v10.1.6), and *Campylobacter jejuni*
77 NCTC113651 was used as an outgroup [15].
78
79 ***DNA-DNA hybridization (DDH), average nucleotide identity (ANI) and GC content***
80 *In silico* DDH and ANI analyses were performed on all 245 *C. concisus* genomes. DDH values
81 were calculated on Genome-to-Genome Distance Calculator (GGDC 2.1
82 https://ggdc.dsmz.de). ANI values were generated using fastANI (v1.3) [16, 17]. The GC
83 contents were determined using SeqKit [18].
84
85 ***Identification of genomospecies-specific genes***
86 The results on the presence and absence of genes among 245 *C. concisus* strains obtained from
87 Roary were further analysed through Scoary (v1.6.16), GS-specific genes were then identified
88 [19]. Redundant genes were excluded until mutually exclusive GS-specific genes were yielded.
89 The presence of each of the candidate GS specific genes in 245 *C. concisus* strains was
90 confirmed by blasting each gene against whole genomes and annotated genes using nucleotide
91 BLAST, and annotated proteins using protein BLAST. In this study, GS specific genes are

92 defined as genes present in all GS1 strains but absent in all GS2 strains, *vice versa*, as described
93 previously [9].
94
95 **Matrix assisted laser desorption ionization-time of flight mass spectrometry (MALDI-**
96 **TOF MS) analysis**
97 A total of 20 *C. concisus* strains (10 GS1 strains and 10 GS2 strains) were used for this analysis
98 (Supplementary Table 1). The complete genomes of these strains were sequenced previously
99 [11]. These strains were cultured on horse blood agar (HBA) plates under anaerobic conditions
100 supplemented with 5% H_2 for 48 hours as described previously [20].
101
102 Bacterial proteins were prepared using ethanol-formic acid extraction procedure. Briefly, a
103 loopful bacteria of each strain was suspended in 300 µL of distilled water, and 900 µL of
104 ethanol was added. After mixing thoroughly, the bacterial cell suspension was centrifuged at
105 13,000 × g for 2 mins and the supernatant was discarded. The residual ethanol was removed
106 by repeating centrifugation. The pellet was dried at room temperature for 5 mins and suspended
107 in 5 to 50 µL of formic acid-water (70:30 [vol/vol]), and an equal volume of ethanol was added.
108 Following centrifugation at 13,000 × g for 2 mins, 1 µL of supernatant was spotted on ground
109 steel MALDI target plate and allowed to dry at room temperature prior to overlaying 1 µL of
110 matrix solution.
111
112 MALDI-TOF MS analysis was performed using the linear positive ion mode on Ultraflex TOF-
113 TOF mass spectrometer (Bruker Daltonik GmbH, Bremen, Germany). External calibrations
114 were performed using The Bruker Bacterial Test Standard (Bruker Daltanics) according to the
115 manufacturer's instructions. The raw spectra data were exported to Biotyper Version 4.0 (Build

116 19) for bacterial identification and FlexAnalysis Version 3.4 for further downstream data
117 analysis.
118
119 **Scanning electron microscope (SEM) imaging of GS1 and GS2 *C. concisus* strains**
120 Six *C. concisus* strains were randomly chosen for SEM examination, including three GS1
121 strains including P26UCO-S2, P10CDO-S2, and P3UCO1; and three GS2 strains including
122 BEO2, P15UCO-S2, and P2CDO4.
123
124 *C. concisus* strains were cultured on HBA plates for 24 hours, followed by cultivation in heart
125 infusion broth containing 5% fetal bovine serum for 12 hours. Bacterial cells were collected
126 and fixed overnight at 4 °C in 2.5% glutaraldehyde in 0.2 M sodium phosphate buffer. Fixed
127 samples were dehydrated using a graded series of ethanol washes (30%, 50%, 70%, 80%, 90%
128 and 2 × 100%) for 5 mins in each solution change. Samples were then air-dried using
129 hexamethyldisilazane evaporative dehydration. Samples were then mounted onto SEM stubs,
130 platinum coated and viewed using a FEI Nova NanoSEM 230 (Oregon, USA) operating at 5kV.
131
132 **Examination of the presence of *csep1* gene, plasmids pSma1 and pICON in GS1 and GS2**
133 ***C. concisus* strains**
134 The *C. concisus csep1* gene, plasmids pSma1 and pICON were previously found to be
135 associated with IBD, including both CD and UC [11, 21]. Their presence in the genomes of 85
136 GS1 and 160 GS2 strains, as well as in other *Campylobacter* species was examined by NCBI
137 nucleotide BLAST [22].
138
139 **Growth characters of GS2 *C. concisus* strains and catalase, oxidase and urease tests**

140 Growth of 10 GS2 *C. concisus* strains with complete sequenced genomes (Supplementary
141 Table 1) under different conditions, including incubation temperature, atmospheric conditions,
142 resistance to bile were examined. Gram staining was performed. Catalase, oxidase and urease
143 tests were also performed.
144
145 **RESULTS**
146 ***C. concisus* GS1 and GS2 strains can be separated based on core genome and 23S rRNA**
147 **gene**
148 The core genome of 245 *C. concisus* strains contained 675 genes, as determined by Roary
149 default setting (the gene is present in more than or equal to 99% of strains). The 85 GS1 *C.*
150 *concisus* strains and 160 GS2 *C. concisus* strains were clearly separated based on the sequences
151 of core genome (Figure 1). The GS2 *C. concisus* strains contained two clusters, most of the
152 strains belonged to cluster 1. The 23S rRNA gene also consistently separated GS1 and GS2 *C.*
153 *concisus* strains (Figure 1). However, the sequences of 16S rRNA genes were unable to
154 differentiate GS1 and GS2 *C. concisus* (Supplementary Figure 1).
155
156 The average GC content of the 85 GS1 *C. concisus* strains was 37.3%, ranging from 34.7 to
157 38.2%. The average GC content of the 160 GS2 *C. concisus* strains was 39.2%, ranging from
158 36.4 to 40.3%.
159
160 **The DDH and ANI values between GS1 and GS2 *C. concisus* strains were below the**
161 **recommended cut off value for defining a bacterial species**
162 The DDH values between each of the 85 GS1 *C. concisus* strains and each of the 160 GS2 *C.*
163 *concisus* strains were all below 70%. The average DDH value between GS1 and GS2 *C.*
164 *concisus* strains was 54.9%, ranging from 42.8 to 66% (Table 1).

165 The average ANI value between each of the 85 GS1 *C. concisus* strains and each of the 160
166 GS2 *C. concisus* strains was 88.7%, ranging from 87.8 to 89.7% (Table 1).
167
168 **The DDH values between all GS1 *C. concisus* strains and the *C. concisus* type strain**
169 **ATCC 33237**
170 The DDH values between the *C. concisus* type strain ATCC 33237 and the 84 remaining GS1
171 strains were all ≥69% (Table 1). The average DDH values between the 84 GS1 *C. concisus*
172 strains and the *C. concisus* type strain ATCC 33237 was 82.9%, ranging from 74.5 to 99.2 %
173 (Table 1).
174
175 The ANI values between the 84 GS1 *C. concisus* strains and the *C. concisus* type strain ATCC
176 33237 ranged from 93.83 to 99.99% (Table 1).
177
178 **The DDH values between all GS2 *C. concisus* strains and the *C. concisus* type strain**
179 **ATCC 33237 were below 70%**
180 The DDH values between the 160 GS2 strains and the *C. concisus* type strain ATCC 33237
181 were all below 70%, ranging from 50.5 to 63.7 %. The ANI values between the 160 GS2 *C.*
182 *concisus* strains and the *C. concisus* type strain ATCC 33237 ranged from 88.41 to 89.15%
183 (Table 1).
184
185 **Both GS1 and GS2 *C. concisus* strains carry GS-specific genes**
186 Both GS1 and GS2 *C. concisus* strains contained their respective specific genes. Four GS1 and
187 three GS2 *C. concisus* specific genes were identified (Table 2).
188
189 **MALDI-TOF MS profiles were different between GS1 and GS2 *C. concisus* strains**

190 On MALDI-TOF MS analysis, protein peaks specific to GS1 or GS2 *C. concisus* strains were
191 identified. Two protein peaks, at m/z 7401 and 7557, were found to be present only in GS1 *C.*
192 *concisus* strains, with the prevalence being 80 and 60% respectively (Table 3). Four protein
193 peaks with m/z 6924, 7424, 8045 and 10359, were present only in GS2 *C. concisus* strains,
194 with the prevalence being 90%, 90%, 50% and 50% respectively (Table 3). Combination of
195 m/z 6924 with any of the other three protein peaks can 100% differentiate GS2 from GS1
196 strains.
197
198 **GS1 and GS2 *C. concisus* strains showed difference in morphology**
199 Electron microscopy examination showed that strains of GS1 and GS2 *C. concisus* differ in
200 morphology. GS1 *C. concisus* strains were straight to curved rods and the shapes of GS2 *C.*
201 *concisus* strains were spiral to curved (Figure 2).
202
203 **The Crohn's disease associated *csep1* gene and plasmids were predominately present in**
204 **GS2 *C. concisus* strains**
205 The Crohn's disease associated *csep1* gene was found in the chromosome of 38.75% (62/160)
206 GS2 *C. concisus* strains and none of the GS1 strains. The Crohn's disease associated pICON
207 plasmid was found in two GS1 strains (2/85, 2.4%), four GS2 strains (4/160, 2.5%) and one
208 strain of *Campylobacter showae* (B32_SW) [23]. The severe ulcerative colitis associated
209 pSma1 plasmid was found in three GS1 strains (3/85, 3.5%) and 13 GS2 strains (13/160, 8.1%)
210 (Supplementary Table 1).
211
212 **DISCUSSION**
213 We previously reported that *C. concisus* consisted of GS1 and GS2 strains based on the
214 difference in the core genome, which was confirmed by other researchers [9-12]. Our findings

215 in this study show that GS2 *C. concisus* is a novel *Campylobacter* species, we propose naming
216 it *C. oralis,* whereas GS1 *C. concisus* solely remains as *C. concisus*.
217
218 The recommended cut off DDH and ANI values for defining bacterial species are 70% DDH
219 and 95% ANI [17, 24-26]. Both the DDH and ANI values between *C. oralis* and *C. concisus*
220 strains were clearly below these criteria (Table 1). Phenotypically, *C. oralis* has a spiral to
221 curved rod shape and *C. concisus* has a straight or curved rod shape under the culture conditions
222 that we have used (Figure 2). MALDI-TOF MS analysis identified *C. oralis* and *C. concisus*
223 specific protein peaks. Our study and the study by On et al. consistently identified four *C. oralis*
224 protein peaks with the m/z 6924, 7424, 8045 and 10359 (Table 3) [27]. Combination of these
225 protein peaks can 100% differentiate *C. oralis* from *C. concisus*. Based on our genome analysis,
226 *C. oralis* and *C. concisus* specific genes were not related to common metabolic pathways,
227 making biochemical differentiation difficult. PCR targeting 23S rRNA gene can also be used
228 as a quick method for differentiating *C. oralis* from *C. concisus* and the sequences of PCR
229 primers were reported in previous studies [28, 29]. The 16S rRNA gene cannot differentiate *C.*
230 *oralis* from *C. concisus* (Supplementary Figure 1).
231
232 We and other researchers previously reported the association between *C. concisus* and IBD [2-
233 5]. In these earlier studies, the identification of *C. concisus* was based on 16S rRNA gene, thus
234 both *C. oralis* and *C. concisus* were detected. Recently, we also reported the associations
235 between *csep1* gene, pICON and pSma1 plasmids with IBD [11, 21]. In this study, we found
236 that Crohn's disease associated *csep1* gene was present in 38.75% of *C. oralis* strains and none
237 of the *C. concisus* strains. The plasmid pICON is rare, we found it in only seven strains of three
238 *Campylobacter* species (four *C. oralis* strains, two *C. concisus* strains and one *C. showae*
239 strain), interestingly all these strains were isolated from patients with Crohn's disease [10, 21,

240 23]. The severe ulcerative colitis associated pSma1 plasmid was found in less than 5% of *C.*
241 *concisus* strains and 10% of *C. oralis* strains. These suggest that *C. oralis* may play a more
242 important role in the development of Crohn's disease than *C. concisus.*
243
244 In conclusion, ANI and DDH values clearly support that *C. concisus* GS2 as a novel
245 *Campylobacter* species, for which we propose the name *Campylobacter oralis.* The
246 identification of *C. concisus* and *C. oralis* specific genes, along with differences in their protein
247 profiles, morphology, core genome and 23S rRNA phylogeny, further support this
248 classification.
249
250 **Description of *C. oralis***
251 *C. oralis* bacterial cells are Gram negative, with a curved to spiral shape, 0.5 – 0.8 × 4 – 8 µm,
252 having rounded ends, motile with darting motility by means of a single polar flagellum (Figure
253 2). Colonies are circular, with convex elevation, translucent, smooth margins, no pigmentation,
254 and no observable haemolysis on HBA plates. The growth and biochemical characteristics of
255 *C. oralis* are listed in Table 4.
256
257 P15UCO-S2 strain is the type strain of *C. oralis*. This strain was isolated from saliva of a patient
258 with ulcerative colitis and its complete genome has been sequenced [11]. The DDH values
259 between the 159 *C. oralis* strains and the *C. oralis* type strain P15UCO-S2 ranged from 66.2
260 to 98.7%. The ANI values between the 159 *C. oralis* strains and the type strain P15UCO-S2
261 ranged from 93.6 to 98.7%. Based on these analyses, we recommend 66.2% DDH and 93.6%
262 ANI with the type strain P15UCO-S2 as the cut off values for identifying *C. oralis*. These
263 values are able to exclude *C. concisus* strains, the DDH and ANI values between the 85 *C.*
264 *concisus* strains and the *C. oralis* type strain P15UCO-S2 were 53.4 – 64.9% and 88.2 – 89.3%

265 respectively. The *C. oralis* type strain has been deposited in the DSMZ under the number DSM
266 115209.
267
268 *C. oralis* was most closely related to *C. concisus* (Supplementary Table 2). *C. oralis* and *C.*
269 *concisus* can be differentiated based on the core genome sequence, 23S rRNA gene, DDH
270 value, ANI value and species-specific genes. *C. oralis* and *C. concisus* show different
271 morphology under electron microscope under certain culture conditions. PCR targeting 23S
272 rRNA gene can be used to differentiate *C. oralis* and *C. concisus* in clinical laboratories.
273 MALDI-TOF MS is a method that can be potentially used to differentiate *C. oralis* and *C.*
274 *concisus*.

275 **TABLES**

276 **Table 1 DDH and ANI values obtained from analyses of GS1 and GS2 *C. concisus***
277 **genomes**

		DDH	ANI
85 GS1 vs 160 GS2 strains	Max	66	89.69
	Min	42.8	87.83
	Mean	54.9	88.69
84 GS1 strains versus ATCC 33237	Max	99.2	99.56
	Min	74.5	94.01
	Mean	82.9	94.74
160 GS2 strains versus ATCC 33237	Max	63.7	89.15
	Min	50.5	88.41
	Mean	57.4	88.73

278 GS: genomospecies. ATCC 33237: *C. concisus* type strain described by Tanner et al [1]

279 **Table 2 GS1 and GS2 *C. concisus* specific genes**

GS	Gene	Annotation	Gene locus	Protein length (AA)
1	*tehA*	Tellurite resistance protein TehA	CVT06_06385	330
	abgT	p-aminobenzoyl-glutamate transport protein	CVT06_04695	508
	pstS	Phosphate-binding protein PstS	CVT06_05910	346
2	*hcp*	Hydroxylamine reductase	CCC13826_1540	550
	aqpZ	Aquaporin Z	CCC13826_1636	231

280 GS: genomospecies. AA: amino acid. GS specific genes: present in all the strains from one
281 GS and absent in all the strains from the other GS.

282 **Table 3 MALDI-TOF MS markers differentiating GS1 and GS2 *C. concisus***

		Average m/z					
Strain ID	**GS**	**6924.389**	**7401.524**	**7424.504**	**7557.485**	**8045.274**	**10359.59**
P10CDO-S2	1		7392.891				
P26UCO-S2	1		7409.682		7558.543		
H25O-S1	1						
P3UCO1	1		7411.752		7556.284		
P19CDO-S1	1				7558.897		
H17O-S1	1		7373.999				
H28O-S1	1		7406.009				
H1O1	1		7404.886		7555.046		
P27CDO-S2	1		7403.849		7557.604		
P28CDO-S1	1		7409.12		7558.533		
P1CDO2	2	6924.038		7426.252		8044.699	
P1CDO3	2	6926.038					10359.56
P15UCO-S2	2	6924.789		7425.441			
P2CDO4	2	6923.513		7424.174		8044.692	10359.5
H14O-S1	2			7424.693		8045.149	10358.19
P16UCO-S2	2	6923.745		7426.193		8045.902	10360.75
H19O-S1	2	6924.265		7422.043			10359.98
P20CDO-S2	2	6921.875		7424.294			
13826	2	6926.661		7423.399		8045.93	
P13UCO-S1	2	6924.579		7424.045			

283 GS: genomospecies

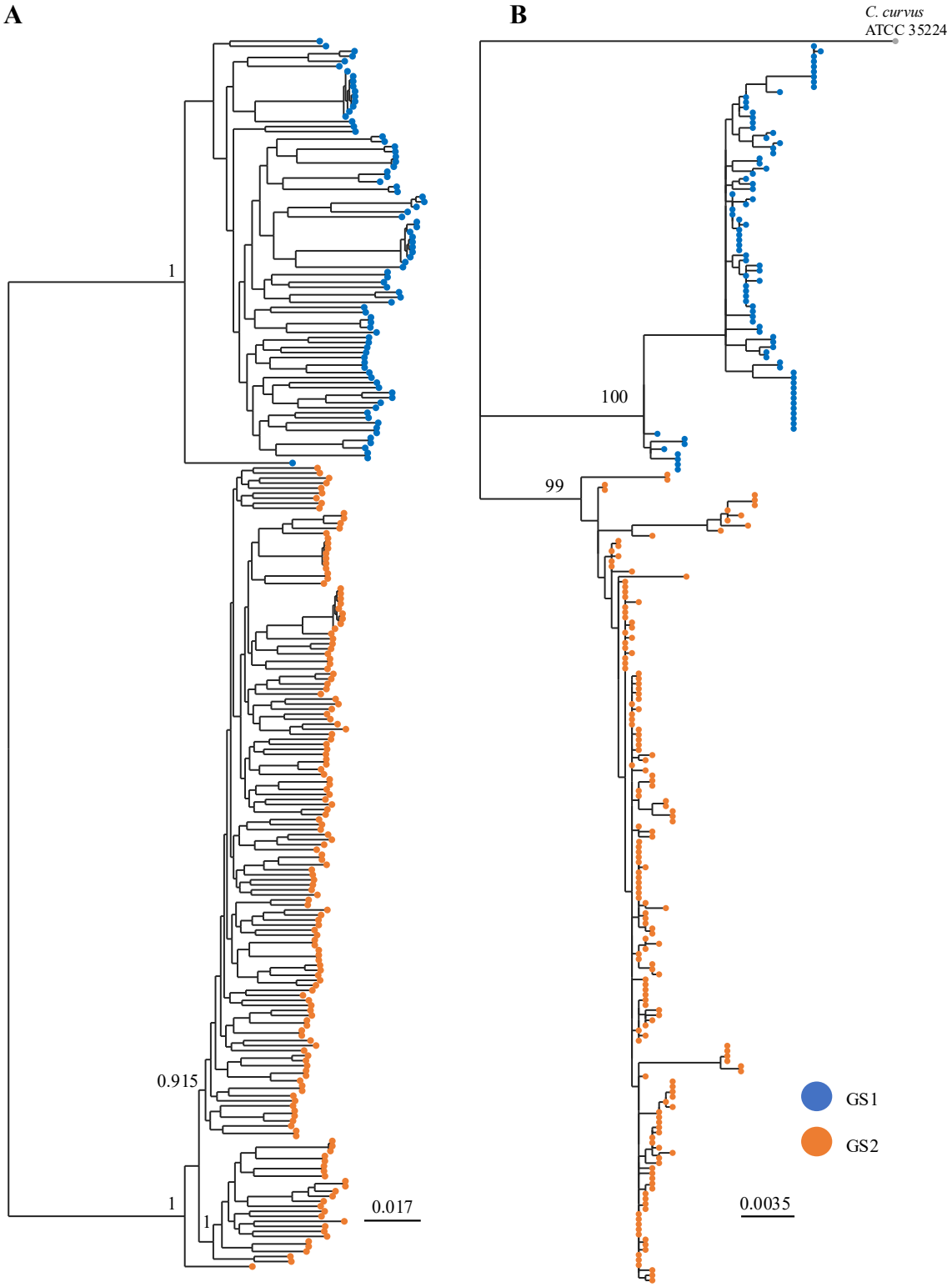
284 **Table 4 The growth and biochemical characters of *C. oralis***

Conditions	**Growth**
37 ºC	Growth
42 ºC	Growth
Microaerobic conditions (generated using Oxford gas-pak)	No growth
Anaerobic conditions (generated using Oxford gas-pak)	Slow growth
Microaerobic conditions with 5 -10% hydrogen	Growth
Anaerobic conditions with 5 -10% hydrogen	Growth
Aerobic conditions containing 10% CO2	No growth
HBA plates containing 1% ox bile	Growth
HBA plates containing 2% ox bile	Growth
MacConkey agar	Growth
HBA plates containing 0.4% Sodium fumarate	Increased growth
HBA plates containing 1% NaCl	Growth
HBA plates containing 2% NaCl	Growth
Catalase	Negative
Oxidase	Positive
Urease	Negative

285 *C. oralis* and *C. concisus* displayed same growth and biochemical characters when same
286 parameters were used.

287 **FIGURES**

Figure 1



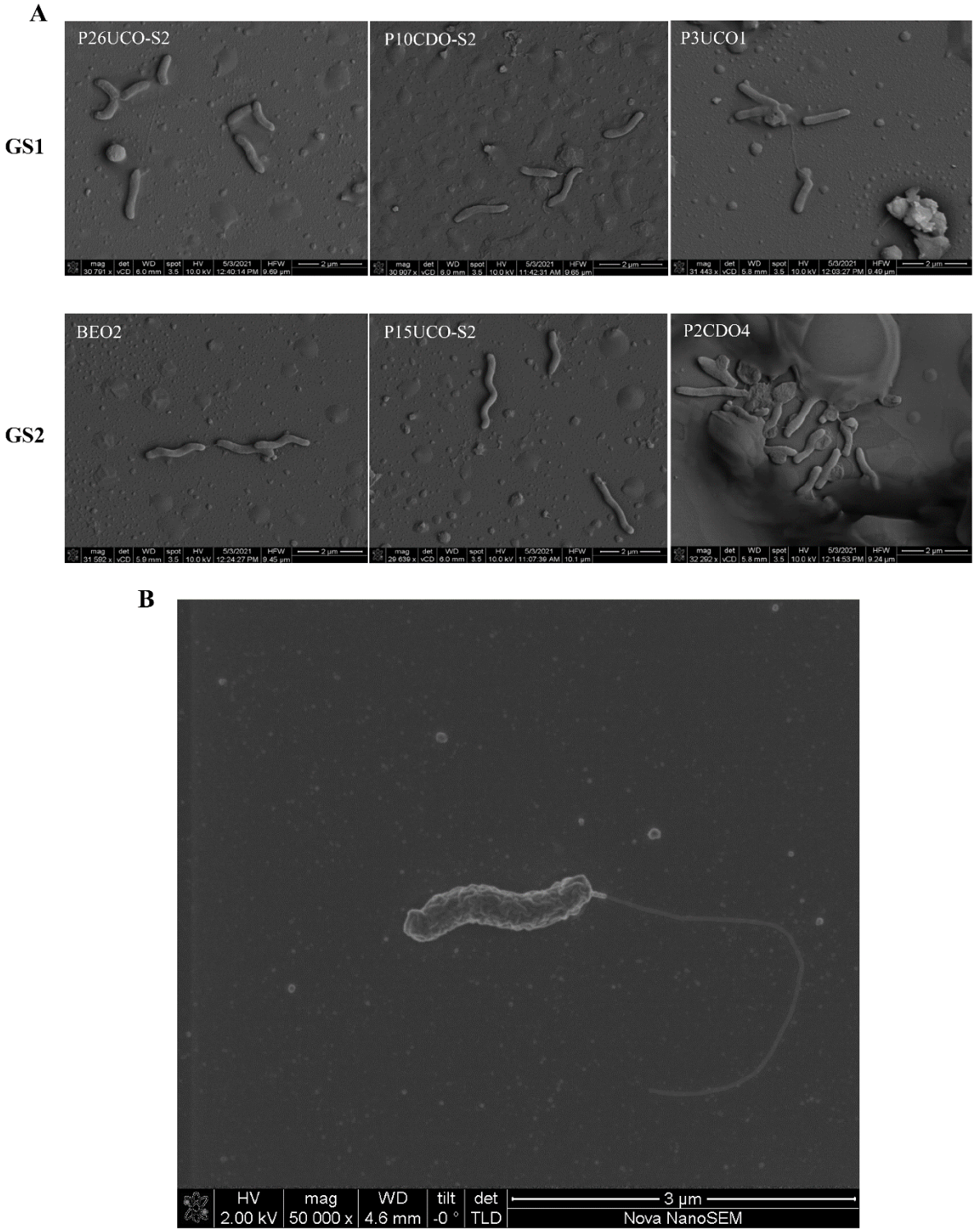
288

289 **Figure 1 Phylogenetic trees of GS1 and GS2 *C. concisus* based on sequences of *C. concisus***

290 **core genome and the 23S rRNA gene.** *C. concisus* core genome and 23S rRNA were able to

291 separate GS1 and GS2 strains. **A:** Core genome was generated by Roary (v3.12.0) and the
292 approximate maximum-likelihood tree was generated by FastTree (v2.1.10). **B:** 23S rRNA
293 maximum-likelihood tree (100 bootstrap) was constructed with MEGAXcc. Strains P21CDO-
294 S2 and H25O-S1 were removed for the 23S rRNA phylogenetic tree due to incomplete 23S
295 rRNA sequences.

Figure 2



296

297 **Figure 2 Scanning electron microscope images showing the different morphologies of GS1**

298 **and GS2 *C. concisus*. A:** *C. concisus* strains were cultured on HBA plates for 24 hours

299 followed by cultivation in heart infusion broth with 5% fetal bovine serum for 12 hours.

300 Bacterial cells were then subjected to scanning electron microscopic examination. Top row:
301 GS1 strains including P26UCO-S2, P10CDO-S2 and P3UCO1. Bottom row: GS2 strains
302 including BEO2, P15UCO-S2 and P2CDO4. Magnification: 30000×, scale bars represent 2
303 µm. **B:** Showing the single polar flagellum of *C. oralis*. Magnification:500000×. GS:
304 genomospecies.

305 **AUTHOR CONTRIBUTION:**

306 Fang Liu: Conducted analysis of *csep1* gene and plasmids. Performed confirmation analysis
307 on GS specific genes.

308 Siying Chen: Performed genome analysis, most of the experiments and prepared the type strain
309 for submission to ATCC.

310 Joanna M Biazik: Generated SEM pictures for Figure 2A using bacterial strains cultured by
311 Siying Chen.

312 Christopher Luk: Conducted catalase, oxidase and urease tests.

313 Anjaneyaswamy Ravipati: Generated data for Table 3 using bacterial strains cultured by Siying
314 Chen.

315 Ruiting Lan: Provided critical discussion during this project.

316 Li Zhang and Stephen M Riordan: Conceived the project and Li Zhang supervised the project.
317 Li Zhang, Fang Liu and Siying Chen played a major role in writing the manuscript. All authors
318 have read the manuscript and provided feedback.

319

320 **ACKNOWLEDGEMENT:**

321 Picture in Figure 2B was generated by Vikneswari Mahendran during her study for a Masters
322 by Research degree in Associate Professor Li Zhang group at the University of New South
323 Wales in 2011.

324

325 **FUNDING SUPPORT:**

326 This work is supported by a Faculty Research Grant awarded to LZ from the University of New
327 South Wales.

328

329 **CONFLICTS OF INTEREST**

The authors declare that there are no conflicts of interest

331 **REFERENCE:**

332 1. Tanner, A.C.R., et al., *Wolinella gen. nov., WoZinella succinogenes (Vibrio*
333 *succinogenes Wolin et al.) comb. nov., and description of Bacteroides gracilis sp. nov.,*
334 *Wolinella recta sp. nov., Campylobacter concisus sp. nov., and Eikenella corrodens*
335 *from Humans with Periodontal Disease.* Int J Syst Bacteriol, 1981(ll): p. 432-445.

336 2. Zhang, L., et al., *Detection and isolation of Campylobacter species other than C. jejuni*
337 *from children with Crohn's disease.* J Clin Microbiol, 2009. **47**(2): p. 453-455.

338 3. Mukhopadhya, I., et al., *Detection of Campylobacter concisus and other*
339 *Campylobacter species in colonic biopsies from adults with ulcerative colitis.* PLOS
340 ONE, 2011. **6**(6): p. e21490-e21490.

341 4. Mahendran, V., et al., *Prevalence of Campylobacter species in adult Crohn's disease*
342 *and the preferential colonization sites of Campylobacter species in the human intestine.*
343 PLOS ONE, 2011. **6**(9): p. e25417.

344 5. Kirk, K.F., et al., *Optimized cultivation of Campylobacter concisus from gut mucosal*
345 *biopsies in inflammatory bowel disease.* Gut Pathog, 2016. **8**: p. 27.

346 6. Lastovica, A.J. and E. Roux, *Efficient isolation of Campylobacteria from stools.* J Clin
347 Microbiol, 2000. **38**(7): p. 2798-2800.

348 7. Macfarlane, S., et al., *Microbial colonization of the upper gastrointestinal tract in*
349 *patients with Barrett's esophagus.* Clin Infect Dis, 2007. **45**(1): p. 29-38.

350 8. Zhang, L., et al., *Isolation and detection of Campylobacter concisus from saliva of*
351 *healthy individuals and patients with inflammatory bowel disease.* J Clin Microbiol,
352 2010. **48**(8): p. 2965-7.

353 9. Chung, H.K., et al., *Genome analysis of Campylobacter concisus strains from patients*
354 *with inflammatory bowel disease and gastroenteritis provides new insights into*
355 *pathogenicity.* Sci Rep, 2016. **6**: p. 38442.

356 10. Kirk, K.F., et al., *Molecular epidemiology and comparative genomics of*
357 *Campylobacter concisus strains from saliva, faeces and gut mucosal biopsies in*
358 *inflammatory bowel disease.* Sci Rep, 2018. **8**(1): p. 1902.
359 11. Liu, F., et al., *Analysis of complete Campylobacter concisus genomes identifies*
360 *genomospecies features, secretion systems and novel plasmids and their association*
361 *with severe ulcerative colitis.* Microb Genom, 2020. **6**(11): p. mgen000457.
362 12. Cornelius, A.J., et al., *Genetic characterisation of Campylobacter concisus: Strategies*
363 *for improved genomospecies discrimination.* Syst Appl Microbiol, 2021. **44**(3): p.
364 126187.
365 13. Page, A.J., et al., *Roary: rapid large-scale prokaryote pan genome analysis.*
366 Bioinformatics, 2015. **31**(22): p. 3691-3693.
367 14. Price, M.N., P.S. Dehal, and A.P. Arkin, *FastTree: computing large minimum evolution*
368 *trees with profiles instead of a distance matrix.* Mol Biol Evol, 2009. **26**(7): p. 1641-
369 1650.
370 15. Kumar, S., et al., *MEGA-CC: computing core of molecular evolutionary genetics*
371 *analysis program for automated and iterative data analysis.* Bioinformatics, 2012.
372 **28**(20): p. 2685-2686.
373 16. Meier-Kolthoff, J.P., et al., *Genome sequence-based species delimitation with*
374 *confidence intervals and improved distance functions.* BMC Bioinformatics, 2013.
375 **14**(1): p. 60.
376 17. Jain, C., et al., *High throughput ANI analysis of 90K prokaryotic genomes reveals clear*
377 *species boundaries.* Nat Commun, 2018. **9**(1): p. 5114.
378 18. Shen, W., et al., *SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file*
379 *manipulation.* PLOS ONE, 2016. **11**(10): p. e0163962.

380 19. Brynildsrud, O., et al., *Rapid scoring of genes in microbial pan-genome-wide*
381 *association studies with Scoary.* Genome Biol, 2016. **17**(1): p. 1-9.
382 20. Lee, H., et al., *Examination of the anaerobic growth of Campylobacter concisus strains.*
383 Int J Microbiol, 2014. **2014**: p. 476047.
384 21. Liu, F., et al., *Genomic analysis of oral Campylobacter concisus strains identified a*
385 *potential bacterial molecular marker associated with active Crohn's disease.* Emerg
386 Microbes Infect, 2018. **7**(1): p. 64.
387 22. Johnson, M., et al., *NCBI BLAST: a better web interface.* Nucleic Acids Res, 2008.
388 **36**(suppl 2): p. W5-W9.
389 23. Hsu, T., et al., *Comparative genomics and genome biology of Campylobacter showae.*
390 Emerg Microbes Infect, 2019. **8**(1): p. 827-840.
391 24. Wayne, L.G., *International Committee on Systematic Bacteriology: Announcement of*
392 *the Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial*
393 *Systematics.* Syst Appl Microbiol, 1988. **10**(2): p. 99-100.
394 25. Goris, J., et al., *DNA-DNA hybridization values and their relationship to whole-genome*
395 *sequence similarities.* Int J Syst Evol Microbiol, 2007. **57**(1): p. 81-91.
396 26. Richter, M. and R. Rossello-Mora, *Shifting the Genomic Gold Standard for the*
397 *Prokaryotic Species Definition.* Proc Natl Acad Sci U S A, 2009. **106**(45): p. 19126-
398 19131.
399 27. On, S.L., et al., *Markers for discriminating Campylobacter concisus genomospecies*
400 *using MALDI-TOF analysis.* CRMS, 2021. **2**: p. 100019.
401 28. Bastyns, K., et al., *Specific detection of Campylobacter concisus by PCR amplification*
402 *of 23S rDNA areas.* Mol Cell Probes, 1995. **9**(4): p. 247-50.

403 29. Wang, Y., et al., *Campylobacter concisus genomospecies 2 is better adapted to the*
404 *human gastrointestinal tract as compared with Campylobacter concisus*
405 *genomospecies 1.* Front Physiol, 2017. **8**: p. 543.
406