Delineation of genetic relatedness and population structure of oral and enteric *Campylobacter concisus* strains by analysis of housekeeping genes

Vikneswari Mahendran,1 Sophie Octavia,1 Omer Faruk Demirbas,1 Sheryl Sabrina,1 Rena Ma,1 Ruiting Lan,1 Stephen M. Riordan,2 Michael C. Grimm3 and Li Zhang1

Correspondence

Li Zhang
l.zhang@unsw.edu.au

1School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, Australia
2Gastrointestinal and Liver Unit, Prince of Wales Hospital, Sydney, Australia
3St George and Sutherland Clinical School, University of New South Wales, Sydney, Australia

*Campylobacter concisus* is an oral bacterium that has been shown to be associated with inflammatory bowel disease (IBD). In this study we examined clusters of oral *C. concisus* strains isolated from patients with IBD and healthy controls by analysing six housekeeping genes. In addition, we investigated the population structure of *C. concisus* strains. Whether oral and enteric strains form distinct clusters based on the sequences of these housekeeping genes was also investigated. The oral *C. concisus* strains were found to contain two genomospecies, which belong to the two genomospecies previously found in enteric *C. concisus* strains. *C. concisus* clusters formed based on the sequences of a single *aspA* gene were the same as that formed by using previously reported MLST schemes. The analysis of combined oral and enteric *C. concisus* strains found that enteric *C. concisus* strains did not form distinct clusters. Genetic structure analysis identified five subpopulations of *C. concisus* and showed that genetic recombination between *C. concisus* strains was common. However, genetic recombination was significantly less in oral strains isolated from patients with IBD than from healthy individuals. Previously reported oral and enteric intestinal epithelial invasive *C. concisus* strains were in cluster II and subpopulation III. Furthermore, this study shows that there are no distinct enteric *C. concisus* strain clusters or subpopulations.

INTRODUCTION

*Campylobacter concisus* is a Gram-negative anaerobic spiral bacterium that also grows under microaerobic conditions in the presence of H₂ gas (Kamma et al., 2000; Lastovica, 2006; Lastovica et al., 2013; Lee et al., 2014; Vandamme et al., 2005). In recent years, *C. concisus* has gained increasing attention as an emerging enteric pathogen. A significantly higher intestinal prevalence of *C. concisus* detected by PCR in patients with inflammatory bowel disease (IBD) as compared with healthy controls was reported (Mahendran et al., 2011; Man et al., 2010a; Mukhopadhya et al., 2011; Zhang et al., 2009). IBD is a chronic inflammation of the gastrointestinal tract, and the two major clinical forms are Crohn’s disease (CD) and ulcerative colitis (UC). In addition to IBD, *C. concisus* has been associated with diarrhoeal disease; several studies have reported the isolation of *C. concisus* from diarrhoeal faecal samples (Lastovica, 2006; Lindblom et al., 1995; Nielsen et al., 2013).

The natural colonization site of *C. concisus* is the human oral cavity. Nearly every human individual is colonized by *C. concisus* in their oral cavity and an individual may be colonized by more than one oral *C. concisus* strain (Ismail et al., 2012; Tanner et al., 1981; Zhang et al., 2009).
We previously examined the genetic relatedness of oral and enteric *C. concisus* strains obtained from patients with IBD and controls and found that enteric *C. concisus* strains from two patients with IBD were genetically either identical to each patient’s own oral *C. concisus* strains or closely related to the oral strains of the other IBD patients (Ismail et al., 2012). These data suggest that *C. concisus* strains colonizing the human intestinal tract originate from the oral *C. concisus* strains. Some oral *C. concisus* strains were shown to have the potential to cause intestinal diseases. For example, some oral *C. concisus* strains were able to invade the human colonic epithelial Caco-2 cells (Ismail et al., 2013; Nielsen et al., 2011). Both *C. concisus* oral and enteric *C. concisus* strains were shown to cause epithelial barrier dysfunction and apoptosis, which are characteristic of diarrhoea-causing pathogens (Nielsen et al., 2011).

*C. concisus* strains are genetically diverse. Several previous studies have used different methods to group *C. concisus* strains based on their genetic differences (Bastyns et al., 1995; Istivan et al., 2004; Kalischuk & Inglis, 2011; On et al., 2013). A study by Aabenhus et al. (2005) examined 57 enteric *C. concisus* strains and five oral *C. concisus* strains using amplified fragment length polymorphism (AFLP) and found that most of the *C. concisus* strains examined belonged to two genomospecies (genomospecies 1 and 2). Most recently, Miller et al. (2012) analysed *C. concisus* strains using MLST analysis, in which they included *C. concisus* strains of different genomospecies from the study by Aabenhus et al. (2005). Miller et al. (2012) found that their MLST method grouped

---

**Table 1. *C. concisus* strains used in the analysis of this study**

<table>
<thead>
<tr>
<th>Strain ID*</th>
<th>Strain information</th>
<th>Reference/source</th>
</tr>
</thead>
<tbody>
<tr>
<td>P4CDO-S2, P4CDO-S3, P4CDO-S4, P7UCB-S1</td>
<td>The 35 oral strains and one enteric strain (P7UCB-S1) were isolated in our previous study (Mahendran et al., 2013)</td>
<td>Mahendran et al. (2013)</td>
</tr>
<tr>
<td>P9CDO-S1, P10CDO-S1, P10CDO-S2, P11CDO-S1, P12CDO-S1, P12CDO-S2, P12CDO-S3, P13UCO-S1, P13UCO-S2, P13UCO-S3, P14UCO-S1, P14UCO-S2, P14UCO-S3, P15UCO-S1, P15UCO-S2, P15UCO-S3, P16UCO-S1, P16UCO-S2, P16UCO-S3, P17UCO-S1, H9O-S1, H9O-S2, H9O-S3, H10O-S1, H10O-S2, H10O-S3, H12O-S1, H12O-S2, H12O-S3, H13O-S1, H14O-S1, H18O-S1</td>
<td>The six housekeeping genes (aspA, gln, tkt, asd, atpA and pgi) in these strains were sequenced in this study</td>
<td></td>
</tr>
<tr>
<td>P1CDO10, P1CDO18, P1CDB1 (UNSWCD)*, P2CDO3†, P2CDO4†, P3UCO1, P3UCR1, P3UCW1, P4CDO1, P5CDO1†, P6CDO1, P7UCO1, P8UCO1†, H1O1, H2O1, H3O1, H4O1, H5O1, H6O1</td>
<td>The 16 oral strains and three enteric strains [P1CDB1 (UNSWCD)], P3UCB1 and P3UCW1] were isolated in our previous studies</td>
<td>Zhang et al. (2009)</td>
</tr>
<tr>
<td>P1CDO10, P1CDO18, P1CDB1 (UNSWCD)*, P2CDO3†, P2CDO4†, P3UCO1, P3UCR1, P3UCW1, P4CDO1, P5CDO1†, P6CDO1, P7UCO1, P8UCO1†, H1O1, H2O1, H3O1, H4O1, H5O1, H6O1</td>
<td>The sequences of six housekeeping genes (aspA, gln, tkt, asd, atpA and pgi) of these strains were obtained from Ismail et al. (2012)</td>
<td>Zhang et al. (2010), Mahendran et al. (2011), Ismail et al. (2012)</td>
</tr>
<tr>
<td>P9CDO-S1, P10CDO-S1, P10CDO-S2, P11CDO-S1, P12CDO-S1, P12CDO-S2, P12CDO-S3, P13UCO-S1, P13UCO-S2, P13UCO-S3, P14UCO-S1, P14UCO-S2, P14UCO-S3, P15UCO-S1, P15UCO-S2, P15UCO-S3, P16UCO-S1, P16UCO-S2, P16UCO-S3, P17UCO-S1, P12CDO-S2, P12CDO-S3, P13UCO-S1, P13UCO-S2, P13UCO-S3, P14UCO-S1, P14UCO-S2, P14UCO-S3, P15UCO-S1, P15UCO-S2, P15UCO-S3, P16UCO-S1, P16UCO-S2, P16UCO-S3, P17UCO-S1, H9O-S1, H9O-S2, H9O-S3, H10O-S1, H10O-S2, H10O-S3, H12O-S1, H12O-S2, H12O-S3, H13O-S1, H14O-S1, H18O-S1</td>
<td>The sequences of six housekeeping genes (aspA, gln, tkt, asd, atpA and pgi) of these strains were obtained from Ismail et al. (2012)</td>
<td></td>
</tr>
<tr>
<td>ATCC 51561, ATCC 51562, UNSW1†, UNSW3†, UNSWCS</td>
<td>The sequences of six housekeeping genes (aspA, gln, tkt, asd, atpA and pgi) of these five enteric strains were obtained from Deshpande et al. (2013)</td>
<td>Deshpande et al. (2013)</td>
</tr>
<tr>
<td>13826</td>
<td>The sequences of six housekeeping genes (aspA, gln, tkt, asd, atpA and pgi) of this enteric strain were obtained from the NCBI database</td>
<td>NCBI database</td>
</tr>
<tr>
<td>Sixty-six <em>C. concisus</em> strains</td>
<td>The <em>aspA</em> gene sequences of these enteric strains were obtained from Miller et al. (2012)</td>
<td>Miller et al. (2012)</td>
</tr>
</tbody>
</table>

*P, isolated from patients with IBD; H, isolated from healthy controls; O, oral strains; B, biopsy strains. †Strains were reported to be invasive to human intestinal cell line Caco-2 cells in previous studies.
C. concisus strains into two clusters that corresponded to gen-
omospecies 1 and 2 reported by Aabenhus et al. (2005).

The studies above have predominantly examined enteric C. concisus strains isolated from diarrhoeal faecal samples (Aabenhus et al., 2005; Matsheka et al., 2002; Miller et al., 2012).

Currently, limited information is available on whether oral C. concisus strains contain different clusters. Previously, we

Fig. 1. Phylogenetic tree based on the sequences of six housekeeping genes (asd, aspA, atpA, glnA, pgi and tkt) illustrating the phylogenetic relationships of oral C. concisus strains analysed in this study. Strains from patients with IBD are coloured red. Strains from healthy controls are coloured blue. 'O' refers to oral strains. Bootstrap values of more than 70 are indicated on the internal branches. The clusters and groups of C. concisus strains are indicated in roman numerals. Campylobacter curvus 525.92 was used as an outgroup. *Strains that were reported to be invasive to Caco-2 cells. Bar, 0.05 changes per nucleotide position.
Housekeeping genes of C. concisus strains
employed an MLST method to examine six housekeeping genes of 16 oral \(C.\) concisus strains (Ismail et al., 2012). These 16 oral \(C.\) concisus strains appeared to form two clusters based on the sequences of the housekeeping genes. However, the number of oral strains examined in that study was low (Ismail et al., 2012). Furthermore, it is not known whether one or both clusters belonged to the genomospecies observed by Miller et al. (2012) as the two studies used a different set of genes for MLST.

In this study we examined 50 oral \(C.\) concisus strains isolated from patients with IBD and healthy controls by analysing the six housekeeping genes that we previously described to investigate the population structure of \(C.\) concisus strains (Ismail et al., 2012). We further determined whether the oral \(C.\) concisus belonged to the same genomospecies as described by Miller et al. (2012).

**METHODS**

\(C.\) concisus strains and housekeeping genes analysed in this study. Fifty oral \(C.\) concisus strains and 10 enteric strains were analysed using the six housekeeping genes scheme that we previously reported (Ismail et al., 2012). The housekeeping genes examined were aspartase \(A\) (\(aspA\)), glutamine synthetase (\(glnA\)), transketolase (\(tkt\)), aspartate semialdehyde dehydrogenase (\(asd\)), ATP synthase \(F_1\) alpha subunit (\(atpA\)) and glucose-6-isomerase (\(pgi\)). The six housekeeping genes of 35 oral \(C.\) concisus strains and one enteric strain were sequenced in this study. The six housekeeping gene sequences of five enteric strains were obtained from the National Center for Biotechnology Information (NCBI) database.

The sequences of the \(aspA\) gene of 65 enteric and one oral \(C.\) concisus strains were obtained from the study by Miller et al. (2012). Details of the \(C.\) concisus strains used in this study for analysis of housekeeping genes are listed in Table 1.

**Bacterial DNA extraction.** \(C.\) concisus DNA was extracted using the Puregene DNA Extraction kit (Gentra) following the manufacturer’s instructions.

**Amplification and sequencing of housekeeping genes.** Six housekeeping genes (\(aspA\), \(glnA\), \(tkt\), \(asd\), \(atpA\) and \(pgi\)) were amplified from 35 oral strains and one enteric \(C.\) concisus strain using previously described PCR primers and thermal cycling conditions (Ismail et al., 2012). Phusion high-fidelity DNA polymerase (New England Biolabs) was used for all PCRs.

The six housekeeping genes for each strain were sequenced using both forward and reverse primers. All sequences were evaluated manually by checking through the sequenced chromatograms of each strain. For some sequences with low-quality bases in the chromatograms, the processes of amplification of housekeeping genes by PCR and sequencing of PCR products were repeated to confirm the ambiguous bases.

**Reconstruction of phylogenetic tree of \(C.\) concisus strains based on housekeeping gene sequences.** Molecular evolutionary genetics analysis software version 5.0 (MEGA 5), available from http://www.megasoftware.net/, was used for sequence alignment and editing (Tamura et al., 2011). Phylogenetic trees were generated using the neighbour-joining distance method in MEGA 5.

The housekeeping genes of the whole genome sequenced \(Campylobacter\) curvus 525.92 were used for outgroup rooting. For phylogenetic trees generated in this study, bootstrap values of >70 are shown in Figs 1–6 and Figs S1–S5 (available in the online Supplementary Material). SplitsTree version 3.2 was used to create a network structure using the uncorrected ‘p’ distance method (Huson & Bryant, 2006).

**Comparison of the MLST scheme with that of Miller et al. in grouping \(C.\) concisus strains.** In this study, we used the MLST method that we previously reported to examine \(C.\) concisus strains by analysis of six housekeeping genes (Ismail et al., 2012). Miller et al. (2012) reported a different MLST scheme employing seven housekeeping genes including \(aspA\), \(atpA\), \(glnA\), citrate \((Si)\)-synthase (\(gltA\)), serine hydroxymethyltransferase (\(glyA\)), dihydroxy-acid dehydratase (\(ilvD\)) and phosphoglucomutase (\(pgm\)) to group \(C.\) concisus strains. To determine whether the \(C.\) concisus clusters designated by our MLST scheme belong to the genomospecies assigned by that of Miller et al. (2012), we compared clusters of 14 oral \(C.\) concisus strains formed by using these two methods. Three housekeeping genes \(aspA\), \(glnA\) and \(atpA\) were used in both MLST schemes. The sequences of \(gltA\), \(glyA\), \(ilvD\) and \(pgm\) from \(C.\) concisus strains were obtained from the NCBI database (Deshpande et al., 2013). The sequences of \(gltA\), \(glyA\), \(ilvD\) and \(pgm\) of an additional seven strains, P2CDO3, P2CDO4, P11CDO-S1, P13UCO-S3, P15UCO-S2, P16UCO-S2 and H1O1, were from our unpublished genome sequence data. The sequences of these four genes were submitted to GenBank.

**Examination of the population structure of \(C.\) concisus.** The program STRUCTURE version 2.2, which utilizes a Bayesian approach for analysis of multilocus sequence data, was used to analyse the population clustering of a strain, assuming that each strain has derived all of its ancestry from only one population (Pritchard et al., 2000). The number of populations, \(K\), was determined under the ‘admixture’ model and in each simulation run; the Markov chain Monte Carlo simulation of 30 000 iterations gave the posterior probability of \(K\) following a burn-in of 10 000 iterations. Different values of \(K\) were run multiple times and the \(K\) value that generated the highest posterior probability was used as the probable number of ancestral populations. The assignment of a strain to a particular population was done under the linkage model.

**GenBank sequence submission.** Sequences of six housekeeping genes (\(aspA\), \(glnA\), \(tkt\), \(asd\), \(atpA\) and \(pgi\)) from 35 oral \(C.\) concisus strains and four housekeeping genes (\(gltA\), \(glyA\), \(ilvD\) and \(pgm\)) from...
Housekeeping genes of C. concisus strains
seven oral *C. concisus* strains were submitted to GenBank. The accession numbers of the sequences of PCR products of housekeeping genes submitted to GenBank were: KP200895–KP201098 and KP996212–KP996239, respectively. The six housekeeping gene sequences of enteric strain P7UCB-S1 were identical to that of a previously submitted oral strain P7UCO1 (Ismail et al., 2012).

**RESULTS**

**Oral C. concisus strains contain two genomospecies**

A phylogenetic tree was reconstructed based on the sequences of six housekeeping genes with a total of 2567 bp from each strain. The 50 oral *C. concisus* strains formed two clusters, with bootstrap values of 99 and 81, respectively, suggesting that oral *C. concisus* strains contain two genomospecies (Fig. 1). Most of the strains belonged to Cluster II (64 %, 32/50) and the remaining strains (36 %, 18/50) formed Cluster I. The proportions of *C. concisus* strains isolated from patients with UC, CD and healthy controls in Cluster II were 37.5 % (12/32), 31.23 % (10/32) and 31.25 % (10/32), respectively, values that were not statistically different from those in Cluster I, which were 16.7 % (3/18), 44.4 % (8/18) and 38.8 % (7/18), respectively (*P* > 0.05).

Multiple oral strains isolated from two individuals were present in the same clusters (patient No. 16 and healthy individual No. 9). However, there were also multiple oral strains isolated from two other individuals (patient No. 4 and healthy individual No. 8) that were found in different clusters. The previously reported invasive oral and enteric strains were all in Cluster II.

In both Cluster II and Cluster I, some strains were closely related, forming small groups with bootstrap values > 90 (Fig. 1). Six such groups in Cluster II (groups I–VI) and three groups in Cluster I (groups I–III) were identified; seven of these groups consisted of strains isolated from patients with IBD only (Fig. 1).

**Oral C. concisus strains divide into two clusters based on the sequences of the aspA gene, which has cluster-associated protein polymorphisms**

To investigate whether a single metabolic gene could determine the separation of oral *C. concisus* strains, six phylogenetic trees were reconstructed based on each of the six housekeeping genes.

The tree generated using the *aspA* gene (528 bp) showed division into two clusters, with bootstrap values of 94 and 99, respectively (Fig. 2). The strains in these two clusters were identical to those in the two clusters formed based on the sequences of the six housekeeping genes in Fig. 1. Comparison of the proteins encoded by the *aspA* gene in each strain identified cluster-associated protein polymorphisms.

At amino acid position 241 (full-length protein position), 91 % (29/32) of strains in Cluster II had glutamic acid (E) and the remaining three strains (3/36, 8 %) had aspartic acid (D). At the same position all strains in Cluster I had aspartic acid (D) (18/18, 100 %). The prevalence of glutamic acid (E) in strains in Cluster II was significantly higher than that in Cluster I (*P* < 0.05). The prevalence of aspartic acid (D) in strains in Cluster I was significantly higher than that in strains of Cluster II (*P* < 0.05).

At amino acid position 385, 47 % (14/32) of strains in Cluster II had glutamine (Q) and the remaining 18 strains (18/32, 56 %) had aspartic acid (D). All strains (18/18, 100 %) in Cluster I had glutamic acid (E). The prevalence of glutamine (Q) in strains in Cluster II was significantly higher than those in Cluster I (*P* < 0.05).

The trees generated based on the *asd* and *glnA* genes divided the strains into two clusters (Figs S1 and S3). However, these two clusters were not identical to the two clusters formed based on the sequences of six housekeeping genes (Fig. 1). The remaining three housekeeping genes did not separate the strains into two clusters (Figs S2, S4 and S5). The inconsistencies between gene trees may be due to recombination between the clusters that have blurred the phylogenetic signal. Most of the branching orders were poorly supported statistically based on bootstrap values, supporting this hypothesis. We further performed a SplitsTree analysis which allows visualization of recombination based on strain relationships as a network structure. Both individual
Housekeeping genes of C. concisus strains
gene trees and the combined gene tree showed an extensive network structure (data not shown).

**The aspA gene divided both oral and enteric *C. concisus* strains into two clusters**

Miller *et al.* (2012) found that 65 enteric strain and one oral *C. concisus* strain formed two clusters based on the sequences of seven housekeeping genes: *aspA*, *atpA*, *glnA*, *gltA*, *glyA*, *ilvD* and *pgm*. Given our finding in the current study that a single *aspA* gene was able to divide oral

**Fig. 4.** Phylogenetic tree based on the sequences of six housekeeping genes (*asd*, *aspA*, *atpA*, *glnA*, *gpi* and *tkt*) illustrating the phylogenetic relationships of oral and enteric *C. concisus* strains analysed in this study. The sequences of strains 51561, 51562, UNSW1, UNSW3 and UNSWCS were retrieved from Deshpande *et al.* (2013). Details of other strains used in this study are in Table 1. Oral strains from patients with IBD are coloured red. Strains from healthy controls are coloured blue. Enteric strains from patients with IBD and gastroenteritis are coloured green. *C. concisus* strain 13826 is the whole genome sequenced strain (GenBank accession no. CP000792.1). Bootstrap values of more than 70 are indicated on the internal branches. The clusters of *C. concisus* strains are indicated in roman numerals. *Campylobacter curvus* 525.92 was used as an outgroup. *Strains that were reported to be invasive to Caco-2 cells. Bar, 0.05 changes per nucleotide position.*

*C. concisus* strains into two clusters, we conducted a further analysis to see whether the *aspA* gene was also able to divide enteric *C. concisus* strains into two clusters. This was done by including the *aspA* gene sequences of 75 enteric strains into the analysis, including 65 enteric strains from the study by Miller *et al.* (2012), three enteric strains from the study by Ismail *et al.* (2012), one enteric strain from this study, five enteric strains from the study by Deshpande *et al.* (2013) and the whole genome sequenced strain *C. concisus* 13826.

The *aspA* gene sequences of 126 *C. concisus* strains (75 enteric and 51 oral) were amplified. The *aspA* gene sequences in strains from Deshpande *et al.* (2013) and our study amplified a total of 528 bp. Strains from Miller *et al.* (2012) amplified a

**Fig. 5.** Phylogenetic tree based on the sequences of seven housekeeping genes (*aspA*, *atpA*, *glnA*, *pgm*, *gltA*, *glyA* and *ilvD*) used by Miller *et al.* (2012) illustrating the phylogenetic relationship of 14 *C. concisus* strains. The sequences of strains 51561, 51562, UNSW1, UNSW3 and UNSWCS were retrieved from Deshpande *et al.* (2013). Details of other strains used in this study are in Table 1. Oral strains from patients with IBD are coloured red. Enteric strains from patients with IBD and gastroenteritis are coloured green. The strain from healthy control is coloured blue. *C. concisus* strain 13826 is the whole genome sequenced strain (GenBank accession no. CP000792.1). Bootstrap values of more than 70 are indicated on the internal branches. *Campylobacter curvus* 525.92 was used as an outgroup. *Bar, 0.05 changes per nucleotide position.*

**Fig. 6.** Phylogenetic tree based on the sequences of six housekeeping genes (*asd*, *aspA*, *atpA*, *glnA*, *pgi* and *tkt*) used by Ismail *et al.* (2012) illustrating the phylogenetic relationships of 14 *C. concisus* strains. The sequences of strains 51561, 51562, UNSW1, UNSW3 and UNSWCS were retrieved from Deshpande *et al.* (2013). Details of other strains used in this study are in Table 1. Oral strains from patients with IBD are coloured red. Enteric strains from patients with IBD and gastroenteritis are coloured green. The strain from healthy control is coloured blue. *C. concisus* strain 13826 is the whole genome sequenced strain (GenBank accession no. CP000792.1). Bootstrap values of more than 70 are indicated on the internal branches. *Campylobacter curvus* 525.92 was used as an outgroup. *Bar, 0.05 changes per nucleotide position.*
total of 477 bp. The *aspA* gene sequences of these 126 *C. concisus* strains shared 438 common nucleotide bases, which were used to generate the phylogenetic tree. The 438 nt bases of the *aspA* gene divided both oral and enteric *C. concisus* strains into two clusters (Fig. 3). Based on sequences of the *aspA* gene, the *C. concisus* strains from Miller *et al.* (2012) formed two clusters; these two clusters contained strains that were identical to those in clusters formed using the MLST scheme of Miller *et al.* (2012) with the exception of strain ST-46.

**Oral and enteric C. concisus strains did not form distinct clusters based on the sequences of six housekeeping genes**

To investigate whether oral and enteric *C. concisus* strains formed distinct clusters based on the sequences of six housekeeping genes, a phylogenetic tree was reconstructed using 50 oral *C. concisus* strains and 10 enteric *C. concisus* strains. These strains formed two clusters. However, the oral and enteric strains did not form distinct clusters (Fig. 4). Seven of the 10 enteric strains were in Cluster II and remaining three strains in Cluster I (Fig. 4). The percentages of enteric strains in Clusters II and I were 18 % (7/39) and 16 % (3/19) respectively, which was not statistically significant (*P* > 0.05) (Fig. 4).

**Comparison of the methods reported by Miller et al. and Ismail et al. in grouping C. concisus strains**

There were 14 strains typed by the MLST schemes of both Miller *et al.* (2012) and us. As shown in Figs 5 and 6, the 14 strains were divided into the same two clusters. Therefore, either MLST scheme can be used for typing of strains to genomospecies level. However, there were differences in branching order of strains within the clusters between the two MLST schemes.

---

**Fig. 7.** Structure analysis of *C. concisus* strains. The five subpopulations are coloured red, green, blue, yellow and pink for subpopulations I, II, III, IV and V, respectively. Oral strains from patients with IBD are coloured red. Enteric strains from patients with IBD and gastroenteritis are coloured green. Strains from healthy controls are coloured blue. Mosaic colours for an isolate indicate a mixed population origin from respective populations of matching colour. The *y*-axis represents the percentage of population assignment, i.e. proportions of ancestry from the *x* subpopulations as colour-coded. *Strains that were reported to be invasive to Caco-2 cells.*
Population structure analyses showed gene recombination between C. concisus strains; oral C. concisus strains from healthy controls showed frequent gene recombination as compared with IBD patients

Analysis of population structure of C. concisus using the software STRUCTURE allocated the 60 C. concisus strains into five subpopulations (I–V) (Fig. 7). Subpopulations I, II, III, IV and V contained four, eight, 31, five and 12 strains, respectively. Oral C. concisus strains from either healthy individuals or patients with IBD did not form distinct subpopulations. Multiple oral strains isolated from an individual were in either the same subpopulation or different subpopulations. The enteric C. concisus strains were separated into four different subpopulations (II–V) which also included the oral C. concisus strains. There were no enteric strains found in subpopulation I. C. concisus strains isolated from intestinal biopsies of three patients with IBD [P7UCB-S1, P1CDB1 (UNSWCD) and P3UCB1] were in the same subpopulation as their oral strains (P7UCO1, P1CDO10 and P3UCO1).

Of the 60 strains, 44 (73 %) possessed mosaic sequence blocks, indicating that they contained ancestral nucleotides from other subpopulations through recombination (Fig. 7). The ancestry of these strains was then estimated as the sum of probabilities from each subpopulation over all polymorphic nucleotides. The proportion of nucleotides from one or more minority subpopulations varied from 3 to 47 %. Of the 44 strains, 28, 10, four and two strains had imports from one, two, three and four other ancestral subpopulations, respectively.

Of the 33 oral strains isolated from patients with IBD, 19 strains (57.6 %) had imported sequences from other subpopulations, which was significantly lower than that in oral strains isolated from healthy controls (94 %, 16/17) ($P<0.001$) (Fig. 7). All four enteric strains isolated from intestinal biopsies of patients with IBD had imported sequences (100 %, 4/4) (Fig. 7).

Subpopulations identified by STRUCTURE generally correlated with the clusters identified by phylogenetic analysis (compare Figs 4 and 7). Thirty-one of the 39 strains in Cluster II were in subpopulation III, and the remaining eight strains (P16UCO-S2, P15UCO-S3, P13UCO-S3, P17CDO-S1, P16UCO-S3, H18O-S1, H8O-S2 and S1561) formed subpopulation II. Strains in Cluster I of the phylogenetic tree were divided into subpopulations I, IV and V.

DISCUSSION

In this study we conducted phylogenetic and subpopulation structure analysis of C. concisus strains based on the sequences of housekeeping genes.

C. concisus is commonly present in the human oral cavity (Zhang et al., 2010). This bacterium was found to colonize the intestinal tract in some individuals and was shown to be associated with IBD (Mahendran et al., 2011; Man et al., 2010b; Mukhopadhyya et al., 2011; Zhang et al., 2009). Previous studies found that enteric C. concisus strains isolated from intestinal biopsies and faecal samples contained two genomospecies (Aabenhus et al., 2005; Kalischuk & Inglis, 2011; Miller et al., 2012). In this study, we examined the clusters formed by 50 oral C. concisus strains based on the similarities of six housekeeping genes using the neighbour-joining distance method, which showed the strains formed two clusters (Fig. 1). The results showed that the oral C. concisus strains comprise two genomospecies.

Miller et al. (2012) verified their MLST method in assigning the two C. concisus genomospecies by using C. concisus strains from Aabenhus et al. (2005). In this study, we found that both our MLST scheme and that of Miller et al. (2012) produced identical results in assigning C. concisus genomospecies (Figs 5 and 6). Taken together, these results show that the oral C. concisus genomospecies that we found in this study belonged to the same two genomospecies previously defined using enteric C. concisus strains isolated from diarrhoeal faecal samples (Aabenhus et al., 2005; Miller et al., 2012).

Accumulated evidence suggests that C. concisus clusters may have clinical implications. The study by Aabenhus et al. (2005) examined 62 C. concisus strains (57 enteric and five oral) using AFLP analysis and found the four enteric strains isolated from patients with bloody stool were all in Cluster 2. In the current study, we found that previously reported invasive oral and enteric strains were all in Cluster II and subpopulation III (Figs 4 and 7).

Whether oral C. concisus strains and enteric strains are of distinct clusters remained unanswered. To investigate this, we generated a phylogenetic tree using both 50 oral C. concisus strains and an additional 10 enteric strains (Fig. 4). The enteric strains were seen within the clusters formed by oral C. concisus strains. These results show that there are no distinct oral or enteric C. concisus clusters. This view is further supported by analysis of the aspA gene in 126 oral and enteric C. concisus strains, which also did not reveal any distinct oral or enteric strain clusters (Fig. 3).

A further interesting finding in this study was that a single metabolic aspA gene was able to divide both oral C. concisus and enteric strains into two clusters. The two C. concisus clusters determined based on the aspA gene contained strains that were identical to the strains in the two clusters determined based on the sequences of the combined six housekeeping genes (Figs 1 and 2). The aspA gene encodes aspartase A (AspA), which is an enzyme that carries out the reversible conversion of aspartate to fumarate and ammonia (Suzuki et al., 1973). In Campylobacter jejuni, addition of fumarate to microaerobic cultures caused an increase in both growth rate and final cell density (Sellars et al., 2002). AspA was shown to be associated with bacterial virulence. Colonization studies showed an aspA mutant in C. jejuni 81-176 colonizing chickens to a significantly lower level than in the wild-type C. jejuni 81-176 (Guccione et al., 2008).
In addition, *aspA* mutant *C. jejuni* from the caecum of chickens were non-culturable by direct plating. This showed that AspA plays a role in the persistence of *C. jejuni* in the avian gut (Guccione et al., 2008).

Analysis of *C. concisus* AspA protein sequences revealed cluster-associated polymorphisms. Cluster II had a significantly higher prevalence of glutamic acid (E) than Cluster I at position 241 (*P* < 0.001). The active site of aspartase from *Bacillus* sp. YM551 has been previously identified and the amino acid at position 241 (glutamic acid, E) was within the active site of the enzyme. This suggests that the polymorphisms of *C. concisus* AspA at the same position (position 241) may affect enzyme activity, which remains to be investigated.

The phylogenetic tree generated using the neighbour-joining tree method is based on nucleotide sequence similarities, which is unable to reveal the exchange of nucleotide sequences between strains. Analysis of the sequences of the six housekeeping genes using the software STRUCTURE revealed that 70 % (35/50) of oral *C. concisus* strains and 90 % (9/10) of enteric *C. concisus* strains contained at least one mosaic sequence block of other *C. concisus* strains, suggesting that genetic recombination between *C. concisus* strains is common. Such genetic recombination between strains is probably one of the factors contributing to the great diversity of *C. concisus*, which was observed in previous studies (Aabenhus et al., 2005; Bastyns et al., 1995; Ismail et al., 2012; Matsheka et al., 2002; Vandamme et al., 1989).

Interestingly, of the strains that do not contain genes from other *C. concisus* strains, only one was from a healthy control (H1O1); the remaining 15 strains (14 oral strains and one enteric strain) were all from patients with IBD, and the prevalence of such oral strains in patients with IBD and healthy controls was statistically significant (Fig. 7). This suggests that patients with IBD are colonized by a group of genetically more stable oral *C. concisus* strains. The mechanisms by which these strains maintain their genetic stability and their clinical significance remain to be investigated.

Analysis of the housekeeping gene sequences of *C. concisus* strains identified five subpopulations (Fig. 7). Cluster II was further divided into two subpopulations. Strains in Cluster I were more diverse and divided into three subpopulations (Fig. 7). Similar to what was observed in the phylogenetic analysis, STRUCTURE did not identify distinct subpopulations that contain only enteric strains. Subpopulation I contained only oral strains. However, this subpopulation had only four strains, and it is therefore uncertain whether this is a subpopulation comprising just oral strains. *C. concisus* strains isolated from intestinal biopsies of patients with IBD and each patient’s own oral *C. concisus* strains were in the same subpopulation, supporting our previous notion that in some patients with IBD, the *C. concisus* strains colonizing the human intestinal tract may originate from the patient’s own oral *C. concisus* strains (Ismail et al., 2012).

In summary, this study found that oral *C. concisus* strains contain two genospecies, which belong to the two genospecies previously found in enteric *C. concisus* strains. *C. concisus* clusters formed based on the sequences of a single *aspA* gene were identical to those formed by MLST schemes previously reported. The analysis of combined oral and enteric *C. concisus* strains found that enteric *C. concisus* strains did not form distinct clusters. Genetic structure analysis identified five subpopulations of *C. concisus* and showed that genetic recombination between *C. concisus* strains was common. However, genetic recombination was significantly less in oral strains isolated from patients with IBD as compared with that from healthy individuals. Previously reported oral and enteric intestinal epithelial invasive *C. concisus* strains were in Cluster II and subpopulation III. Furthermore, this study shows that there are no distinct enteric *C. concisus* strain clusters or subpopulations.

**REFERENCES**


Edited by: T. Msadek