Systematic and Applied Microbiology 44 (2021) 126187

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: http://www.elsevier.com/locate/syapm

Genetic characterisation of *Campylobacter concisus*: Strategies for improved genomospecies discrimination

Angela J. Cornelius[a,*], Mohsina Huq[b], Stephen L.W. On[c], Nigel P. French[d], Olivier Vandenberg[e,f], William G. Miller[g], Albert J. Lastovica[h], Taghrid Istivan[b], Patrick J. Biggs[d]

[a] *Institute of Environmental Science and Research Ltd, P.O. Box 29181, Christchurch 8540, New Zealand*
[b] *School of Science, RMIT University, G.P.O. Box 2476, Bundoora, Victoria 3001, Australia*
[c] *Lincoln University, P.O. Box 85084, Lincoln 7647, New Zealand*
[d] *Massey University, Private Bag 11 222, Palmerston North 4442, New Zealand*
[e] *National Reference Centre for Campylobacter, Laboratoire Hospitalier Universitaire de Bruxelles, 322 rue Haute, 1000 Brussels, Belgium*
[f] *School of Public Health, Campus Erasme – Bâtiment A, Route de Lennik 808 – CP591, Université Libre de Bruxelles, 1070 Bruxelles, Belgium*
[g] *Produce Safety and Microbiology Research Unit, Agricultural Research Service, United States Department of Agriculture, 800 Buchanan Street, Albany, CA 94710, USA*
[h] *University of Western Cape, Private Bag X17, Bellville 7535, South Africa*

ARTICLE INFO

Article history:
Received 7 November 2020
Received in revised form 3 February 2021
Accepted 8 February 2021

Keywords:
Comparative genomic analysis
Average nucleotide identity
Genome BLAST distance phylogeny
Clusters of orthologous groups analysis
Ribosomal RNA operon
Large scale BLAST score ratio

ABSTRACT

Although at least two genetically distinct groups, or genomospecies, have been well documented for *Campylobacter concisus*, no phenotype has yet been identified for their differentiation and thus formal description as separate species. *C. concisus* has been isolated from a variety of sites in the human body, including saliva and stool samples from both healthy and diarrhoeic individuals. We evaluated the ability of a range of whole genome-based tools to distinguish between the two *C. concisus* genomospecies (GS) using a collection of 190 *C. concisus* genomes. Nine genomes from related *Campylobacter* species were included in some analyses to provide context. Analyses incorporating sequence analysis of multiple ribosomal genes generated similar levels of *C. concisus* GS discrimination as genome-wide comparisons. The *C. concisus* genomes formed two groups; GS1 represented by ATCC 33237^T and GS2 by CCUG 19995. The two *C. concisus* GS were separated from the nine genomes of related species. GS1 and GS2 also differed in G+C content with medians of 37.56% and 39.51%, respectively. The groups are consistent with previously established GS and are supported by DNA reassociation results. Average Nucleotide Identity using MUMmer (ANIm) and Genome BLAST Distance Phylogeny generated *in silico* DNA-DNA hybridisation (*is*DDH) (against ATCC 33237^T and CCUG 19995), plus G+C content provides cluster-independent GS discrimination suitable for routine use. Pan-genomic analysis identified genes specific to GS1 and GS2. WGS data and genomic species identification methods support the existence of two GS within *C. concisus*. These data provide genome-level metrics for strain identification to genomospecies level.

Abbreviations: AFLP, amplified fragment length polymorphism; ANIb, average nucleotide identity using BLAST; ANIm, average nucleotide identity using MUMmer; ATCC, American Type Culture Collection; BLAST, basic local alignment search tool; BSR, BLAST score ratio; CCUG, Culture Collection University of Göteborg Sweden; Ccon, *Campylobacter concisus*; Ccur, *Campylobacter curvus*; CDS, coding DNA sequences; Cfet, *Campylobacter fetus*; Cjej, *Campylobacter jejuni*; Cmuc, *Campylobacter mucosalis*; COG, clusters of orthologous groups; Crec, *Campylobacter rectus*; Csho, *Campylobacter showae*; Cspu, *Campylobacter sputorum*; DDH, DNA-DNA hybridisation; DNA, deoxyribonucleic acid; GBDP, genome BLAST distance phylogeny; G+C, guanine (G) plus cytosine (C) (nitrogenous bases); GGDC, genome to genome distance calculator; GS, genomospecies; HSPs, high-scoring segment pairs; *is*DDH, *in silico* DNA-DNA hybridisation; ITS, internal transcribed spacer; LS-BSR, large scale BLAST score ratio; MLST, multi-locus sequence typing; MUM, maximal unique matching; NCBI, National Center for Biotechnology Information; NJ, neighbor joining; Ns, nucleotides of unknown identity (could be adenine [A], cytosine [C], guanine [G] or thymine [T]); rMLST, ribosomal MLST; RNA, ribonucleic acid; rRNA, ribosomal RNA; *rrn*, ribosomal RNA operon; WGS, whole genome sequence or whole genome sequencing.

* Corresponding author.
E-mail address: Angela.Cornelius@esr.cri.nz (A.J. Cornelius).

https://doi.org/10.1016/j.syapm.2021.126187
0723-2020/

Introduction

Campylobacter concisus has been isolated from a variety of sites in the human body including the gingival crevices of patients with gingivitis and periodontitis [1], oesophageal and intestinal biopsies [2–9], blood [10], saliva [3,7,11] and a brain abscess [12]. In addition, *C. concisus* has been isolated from gastroenteritis-associated stool samples [5,7,13,14] as well as stool samples from healthy people [7,15,16].

Determining the role of *C. concisus* in human disease is complex due to its taxonomic structure. Genospecies or genomospecies (GS), which describes genetically distinct groups of strains within a species without an appropriate phenotypic marker required for differentiation [17], were proposed for *C. concisus* based on conventional DNA-DNA hybridisation (DDH) [10,18], and is supported by various molecular methods [3,5,7,18–26]. Differences in pathogenic potential and genetic heterogeneity observed for this species led some researchers to hypothesise that there may be variation in the pathogenic potential of different *C. concisus* GS [13,19,23,26]. Accurate assignation of strains to GS will facilitate improved understanding of the role of each GS in disease. However, many of the genotyping methods applied to date, including DDH and amplified fragment length polymorphism (AFLP), are relatively time consuming.

Whole-genome sequencing (WGS), and methods that use the genome sequence data for strain identification, are becoming increasingly available. Average nucleotide identity (ANI [27–29]) and genome BLAST distance phylogeny (GBDP [30,31]) analyses have been proposed as *in silico* alternatives to DNA-DNA hybridisation for determining the interspecific genomic relatedness of bacteria and the use of these methods has been evaluated for taxa within *Arcobacter, Campylobacter, Helicobacter* and *Wolinella* [32]. Several other WGS-based analyses are also available for phylogenetic investigations. Ribosomal multi-locus sequence typing (rMLST) aims to provide a single analysis that spans intraspecies bacterial typing and identification to the domain level by indexing the sequence variation in 53 ribosomal protein subunit genes [33]. Comparisons at the functional level are possible using the carefully manually curated Clusters of Orthologous Groups of proteins (COG) database (https://www.ncbi.nlm.nih.gov/COG/). This allows predictions of gene functions based on the amino acid sequence similarity between unknown genes and genes that have been studied experimentally [34]. The database currently contains 4,632 COG classified into 26 functional categories [35]. The diversity of the ribosomal RNA (*rrn*) operon, which includes the 5S rRNA, 16S rRNA, 23S rRNA genes and the internal transcribed spacer (ITS) region, has shown promise for evaluating the relationships among *C. concisus* isolates [5]. Relatively few of the described approaches have been used to assign *C. concisus* to a genomospecies level with the exception of sequence analysis of 23S rRNA sequence [3] and *rrn* [5]. However, there are no species boundaries proposed for these analyses and they evaluate sequence similarity in a relatively small proportion of the genome. The performance of other WGS tools, which evaluate larger proportions of the genome, is currently unknown.

A plethora of software programs are available for the identification of core and accessory genomes. The core genome of a bacterial species is responsible for the basic biology and major phenotypic traits of the species [36]. Conversely, the identification of accessory genes has application in characterising metabolic pathways, virulence attributes, adaptation to different environments, host associations and providing molecular fingerprinting targets useful in epidemiological and population studies [37–42]. Two software programs that identify core and accessory genomes are large scale BLAST score ratio (LS-BSR, [43]) and Roary [44].

In this study, we evaluate the ability of a range of genome-based phylogenetic tools to distinguish between the two *C. concisus* GS. We also assess the specificity of GS-specific coding DNA sequences (CDS) identified using Roary and LS-BSR.

Material and methods

Whole genome sequences

Three complete and 160 draft *C. concisus* genomes [5,7,11,13,45,46] were downloaded from GenBank (https://www.ncbi.nlm.nih.gov/genome/?term=campylobacter+concisus) as assembled genomes. An additional 26 assembled draft *C. concisus* genomes, representing strains reported previously [3], were kindly provided by Dr. Li Zhang (University of New South Wales, Australia). All of the available *C. concisus* genomes were obtained from strains isolated from humans. The clinical symptoms, country, year of isolation and characteristics of the genomes are summarised in Table S1 of the Supplementary Material. An additional genome, 10_1_50, was identified as *Campylobacter* sp. in GenBank, and included in the study because it was tentatively identified by us as *C. concisus* GS2 based on G+C content and homology to PCR primers and taxon-specific genes (data not shown).

Genomes from at least one strain of each of the following species were also included in some analyses to provide context: the *Campylobacter* type species *C. fetus* (82-40), the clinically important *C. jejuni* (ATCC 33560^T), a selection of species closely related to *C. concisus* on the basis of 16S rRNA sequence (*C. curvus* [DSM 6644^T], *C. rectus* [ATCC 33238^T], *C. showae* [ATCC 51146^T] and *C. mucosalis* [CCUG 21559 and DSM 21682^T]) as well as two distant species (*C. hominis* [ATCC BAA-381^T] and *C. sputorum* [LMG 8532]). Genomes from type strains were used, where possible. The sources for these genomes are listed in Table S2 of the Supplementary Material.

Contigs of the draft genomes were joined with the addition of 100 or 200 N's between contigs, to avoid genes being annotated across contig boundaries, and then concatenated to form a single sequence per genome. A subset of the genomes, including 49 *C. concisus* and 9 related species, were annotated using Prokka (version 1.8; [47]) to provide a consistent annotation for the comparison.

Comparative genomic analyses

Fig. 1 is a flow diagram summarising the genomes included in each of the following analyses. ANI, using both BLAST (ANIb) and MUMmer (ANIm), was undertaken on 31 genomes (22 *C. concisus* genomes and 9 from related species) using JSpecies (version 1.2.1, website http://imedea.uib-csic.es/jspecies/download.html). ANIm was later performed on an additional 168 *C. concisus* genomes from three recent studies [3,7,11] with representatives of GS1 and GS2, namely strains ATCC 33237^T and CCUG 19995, respectively.

GBDP was performed on the 31 genomes using the genome-to-genome distance calculator (GGDC) version 2.0 available online (http://ggdc.dsmz.de/distcalc2.php) using the recommended BLASTn method. Results were generated using all three formulae. Formula 1 divides the high scoring segment (HSP) length by the total length, formula 2 divides the identities by the HSP length and formula 3 divides the identities by the total length. GBDP was later performed on an additional 168 *C. concisus* genomes from three recent studies [3,7,11] with ATCC 33237^T and CCUG 19995 used as reference genomes for GS1 and GS2, respectively. The programming language R was used to construct heatmaps for the 31-genome ANI and GBDP-generated *in silico* DNA-DNA hybridisation (*is*DDH) results and boxplots for the full set of *C. concisus* genomes for the same two analyses.

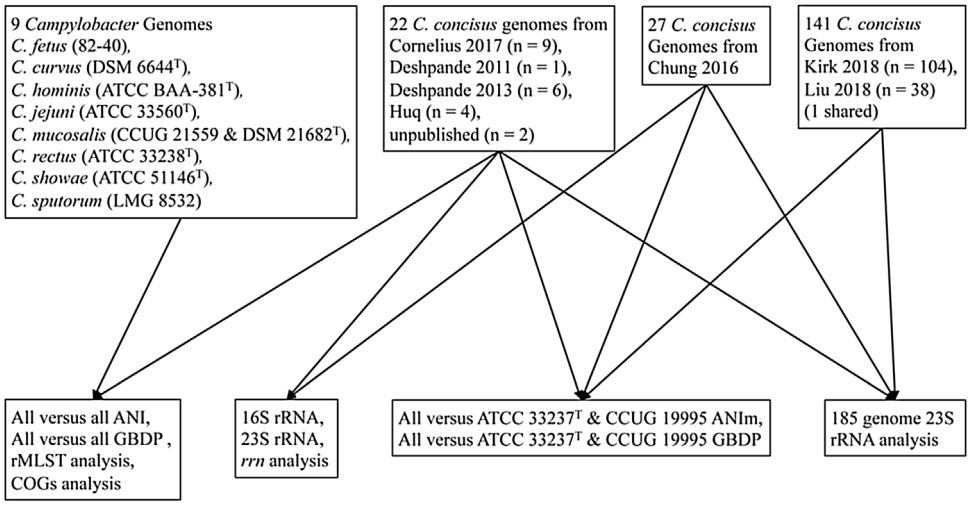

Fig. 1. Flow diagram of the comparative genomic analysis methods used and the genomes analysed by each analysis.

rMLST analysis was undertaken on the 31 genomes using a custom script that extracted the target genes from the Prokka-annotated genomes, primarily on the basis of annotation but also using a *Campylobacter* reference set of rMLST genes as a database for a BLAST search [33]. The gene sequences for each genome were concatenated and a distance matrix generated from the alignment of the concatenated genomes generated using MUltiple Sequence Comparison by Log-Expectation (MUSCLE, version 3.8.31, [48]).

The custom Perl script COGnitorParse2full.pl, available at https://github.com/pjbiggs/CconcisusGenomospecies, was used to compare the amino acid sequences of all protein encoding genes for each of the 31 genomes against a local copy of the COG database (https://www.ncbi.nlm.nih.gov/COG/) using the BLAST algorithm and return a summary of the number of paralogues of each COG observed for each genome. This was then converted to a distance matrix using the custom Perl script table 2Dist2Nex.pl, available at https://github.com/pjbiggs/CconcisusGenomospecies. The matrix outputs of rMLST and COG were converted to NEXUS format [49] and visualised using SplitsTree4 (version 4.12.6, [50]).

The full set of 190 *C. concisus* genomes were evaluated using the genome assembly quality assessment tool QUAST [51] using the online calculator (available at http://cab.cc.spbu.ru/quast/). The Welch two-sample *t*-test in R was used to evaluate the significance of the differences observed between GS1 and GS2 genomes for ANIm, GBDP-generated *is*DDH and the G+C content calculated by QUAST.

Geneious (version R8.1.7, available from http://www.geneious.com/) was used to construct *rrn* operon sequence alignments and consensus sequences as follows. A total of 52 *rrn* sequences were included from 49 *C. concisus* strains; three alleles of the *rrn* operon of *C. concisus* 13826 (*rrn*A, *rrn*B and *rrn*C) and two alleles of *C. concisus* RMIT-JF1 *rrn* operon (*rrn*I and *rrn*II) were included in this analysis. Pairwise distances were computed using MEGA X (Molecular Evolutionary Genetic Analysis across computing platforms, version 10.0.05) [52]. Phylogenetic analysis was performed on the whole *rrn* operon and the 16S and 23S rRNA genes separately. Sequences from *C. fetus* 82-40 and *C. curvus* DSM 6644 were included in the 16S rRNA analysis to provide context. The Lasto205.94 sequence was excluded from the 16S rRNA analysis because only 69% (1005 bp) of the whole 16S rRNA could be extracted from the draft whole genome sequence. A ClustalW (version 2.1) [53] alignment was also generated with a gap opening penalty of 15 and a gap extension penalty of 6.66. The phylogenetic tree was generated based on the sequences of ribosomal RNA (*rrn*) operons. The Neighbor-Joining (NJ) method [54] was used to generate the phylogenetic tree, which was performed using MEGA X [52] with 500 bootstrap replications. Bootstrap values of more than 75 are indicated on the internal branches. The genetic distances were computed using the Maximum Composite Likelihood method [55]. Barrnap v0.9 (https://github.com/tseemann/barrnap) was later used to exact 23S rRNA from the full set of 190 *C. concisus*. The full-length sequences obtained from 185 of the genomes were aligned using MAFFT v7.450 with 'auto' parameters [56,57] and then processed using IQ-TREE [58,59] with the Model Finder Plus [60] option, to find the most likely substitution model. The final phylogeny was built using 1000 ultrafast bootstrap approximation. The Interactive Tree of Life (v 5.7, [61]) was used to visualise the tree. Bootstrap values of more than 75 are indicated on the internal branches.

Identification of possible genomospecies-specific genes

Large-scale BLAST score ratio (LS-BSR, [43]) analyses were was used to evaluate the pan-genome of 17 *C. concisus* genomes (strains with [d] in Table S1 of the Supplementary Material). An in-house script designated listAandB.pl (available at https://github.com/pjbiggs/CconcisusGenomospecies) was used to identify coding sequences (CDS, called centroids by LS-BSR) present (BSR ≥ 0.80) in one GS and absent (BSR < 0.4) from the other. The locus_tag and annotation for GS1- and GS2-specific CDS were retrieved from the GenBank versions of the *C. concisus* ATCC 33237^T and 13826 genomes, respectively. The GS-specific CDS identified from the *C. concisus* LS-BSR were compared to those identified using Roary [3] and the specificity of the CDS were further evaluated by searching a local BLAST database containing all 190 *C. concisus* genomes and the NCBI nucleotide collections (nr/nt). The BLAST searches used the CDS as the query, a word size of 11 and a maximum number of hits of either 1000 or 500. A BSR was calculated for each hit by dividing the bit score for the hit with the bit score generated for the query versus itself. This results in a number between 0 and 1. A hit with a BSR of ≥0.80 was considered to be present and a hit with a BSR of <0.4 was considered effectively absent [62].

Results

Comparative genomic analyses

A flow diagram of all comparative genomic analyses, the genome sources and the genomes that were included in each analysis, is shown in Fig. 1.

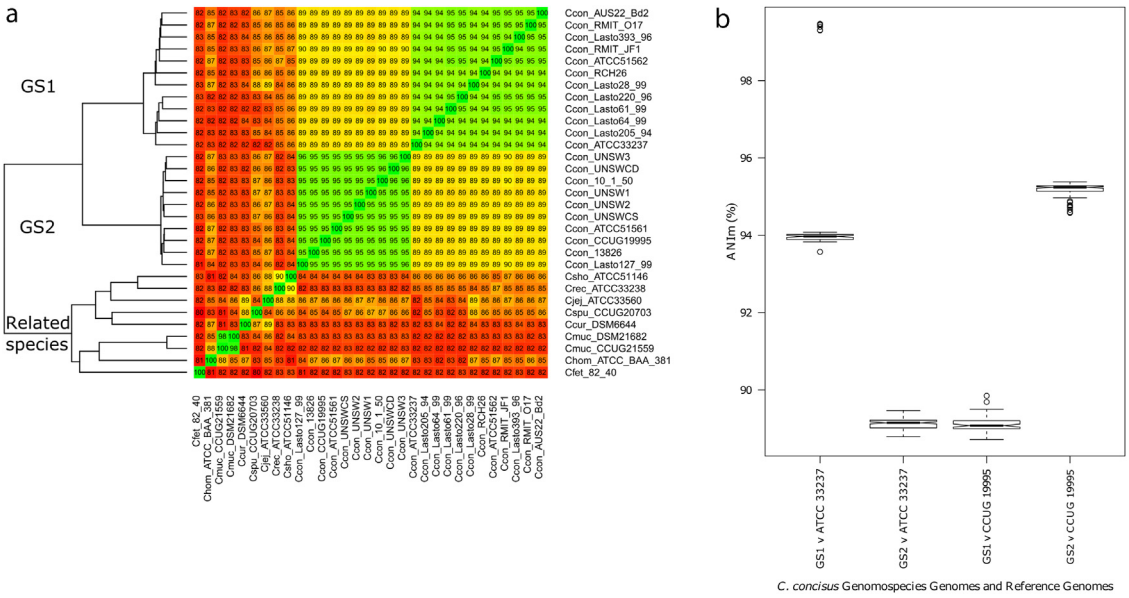

Fig. 2. Average Nucleotide Identity (using MUMmer). Panel a: Heatmap for 22 *C. concisus* genomes and nine non-*C. concisus Campylobacter* genomes. Panel b: Boxplot for 189 *C. concisus* genomes (67 GS1 and 122 GS2) relative to ATCC 32327^{T} and CCUG 19995. The results were rounded to whole numbers to aid readability. *Abbreviations*: Ccon *C. concisus*, Ccur *C. curvus*, Cfet *C. fetus*, Chom *C. hominis*, Cjej *C. jejuni*, Cmuc *C. mucosalis*, Crec *C. rectus*, Csho *C. showae*, Cspu *C. sputorum.*

The *C. concisus* genomes formed two groups, one for each GS, in the ANIm (Fig. 2a) and ANIb heatmaps (Fig. S1 of the Supplementary Material), generated using a selection of 31 genomes (22 *C. concisus* and nine related species). The other *Campylobacter* spp. included in the comparison were separated from the *C. concisus* genomes. Only the two intra-*C. mucosalis* pairwise results exceeded 97%. All of the *C. concisus* GS1 versus GS1 and GS2 versus GS2 (intra-GS, self versus self excluded) pairwise ANIm and ANIb results were between 93% and 96%. All of the *C. concisus* GS1 versus GS2 and GS2 versus GS1 (inter-GS) and the inter-species pairwise results were <91% identity using both ANIb and ANIm. Using ANIm, almost all (89/90, 98.9%) of the *C. concisus* GS2 and a small number (6/132, 4.5%) GS1 intra-GS pairwise ANIm results were at least 95%. The intra-*C. concisus* and intra-*C. mucosalis* ANIb values were similar to those observed with the ANIm analysis however the inter specific ANIb values were noticeably lower than the corresponding ANIm values.

The results for the ANIm analysis of 190 *C. concisus* genomes against representatives of GS1 and GS2, namely ATCC 33237^{T} and CCUG 19995, respectively, are illustrated in Fig. 2b. Eight GS1 genomes were outliers when compared to ATCC 33237^{T}, one (AAUH-3HCo) was lower than the majority of the GS1 genomes and the other seven (AAUH-16UCf, AAUH-16UCf2, AAUH-16UCf3, AAUH-16UCo-a, AAUH-8HCo, AAUH-9HCasc and H15O-S1) were higher. There were no GS2 outliers when compared to ATCC 33237^{T}. Two GS1 genomes (AAUH-12CDo and AAUH-3HCo) were high outliers when compared to CCUG 19995. Twelve GS2 genomes (AAUH-12CDrec-a, AAUH-15HCti, AAUH-19HCf, AAUH-19HCf2, AAUH-3HCce2, AAUH-8UCo, H29O-S1, P13UCO3, P1CDO3, P27CDO-S1, P2CDO3 and P2CDO-S6) were low outliers when compared to CCUG 19995. Seven (10.6%) of the 66 GS1 and 108 (88.5%) of the 122 GS2 intra-GS ANIm results in this analysis were greater than or equal to 95%.

When compared to ATCC 33237^{T}, the 66 GS1 genomes (the ATCC 33237^{T} result was excluded) had a mean score of 94.53% (95% confidence interval 91.14–97.92%) and the 123 GS2 genomes had a mean score of 89.13% (95% confidence interval 88.85–89.40%). The Welch two sample *t*-test for the ATCC 33237^{T} comparison had a *t*-value of 25.84, with 65.45 degrees of freedom and the *p*-value was <0.001. When compared to CCUG 19995, the 67 GS1 genomes had a mean score of 89.13% (95% confidence interval 87.71–89.56%) and the 122 GS2 genomes (the CCUG 19995 result was excluded) had a mean score of 95.17% (95% confidence interval 94.83–95.52%). The Welch two sample *t*-test for the CCUG 19995 comparison had a *t*-value of −198.59, with 114.10 degrees of freedom and the *p*-value was <0.001.

GBDP was initially performed on 31 genomes (22 *C. concisus* and nine related species) using GGDC and heatmaps were generated for the resulting *is*DDH for each of the three formulae. The heatmap for formula 3 is illustrated in Fig. 3a and the heatmaps for formulae 1 and 2 are illustrated in Fig. S2 of the Supplementary Material. All three formulae separated the *C. concisus* genomes into the two GS, with the same component strains observed with ANI. The *C. mucosalis* intra-species pairwise *is*DDH results were >70% for all formulae. No *C. concisus* intra- (GS1 versus GS1 and GS2 versus GS2, self versus self excluded) or inter-GS (GS1 versus GS2 and GS2 versus GS1) pairwise *is*DDH results were >70% using formula 2. All *C. concisus* pairwise GS1 versus GS1 (intra-GS1, self versus self excluded) *is*DDH results using formulae 1 and 3 were >70% and 14 of the 90 (15.6%) *C. concisus* pairwise GS2 versus GS2 (intra-GS2, self versus self excluded) *is*DDH results were <70%. In addition, 14 of the 240 (5.8%) *C. concisus* GS1 versus GS2 and GS2 versus GS1 (inter-GS) pairwise results were >70% using formula 1. Using formula 3, all *C. concisus* inter-GS pairwise results were between 48.9% and 63.2% while all inter-species *is*DDH pairwise results were <19% except between *C. rectus* and *C. showae* where the result was 46%.

GBDP was later performed on the 168 *C. concisus* genomes against ATCC 33237^{T} and CCUG 19995 as representatives of GS1 and GS2, respectively. The formula 3 *is*DDH results were combined with the equivalent results from the 31-genome analysis to generate boxplots for the full set of 190 genomes (Fig. 3b). Ten GS1 genomes were outliers when compared to ATCC 33237^{T}, three (AAUH-11UCdes-a, AAUH-12CDo and AAUH-5CDo) were lower than the majority of GS1 genomes and seven (AAUH-16UCo-a, AAUH-16UCf, AAUH-16UCf2, AAUH-16UCf3, AAUH-8HCo, AAUH-9HCasc and H15O-S1) were higher. One GS2 genome (AAUH-44UCsig-a) was a low outlier when compared to ATCC 33237^{T}. Four GS1 genomes (AAUH-11HCf, AAUH-11UCdes-a, AAUH-12CDo and AAUH-5CDo) and one GS1 genome (AAUH-44UCsig-a) were low outliers when compared to CCUG 19995.

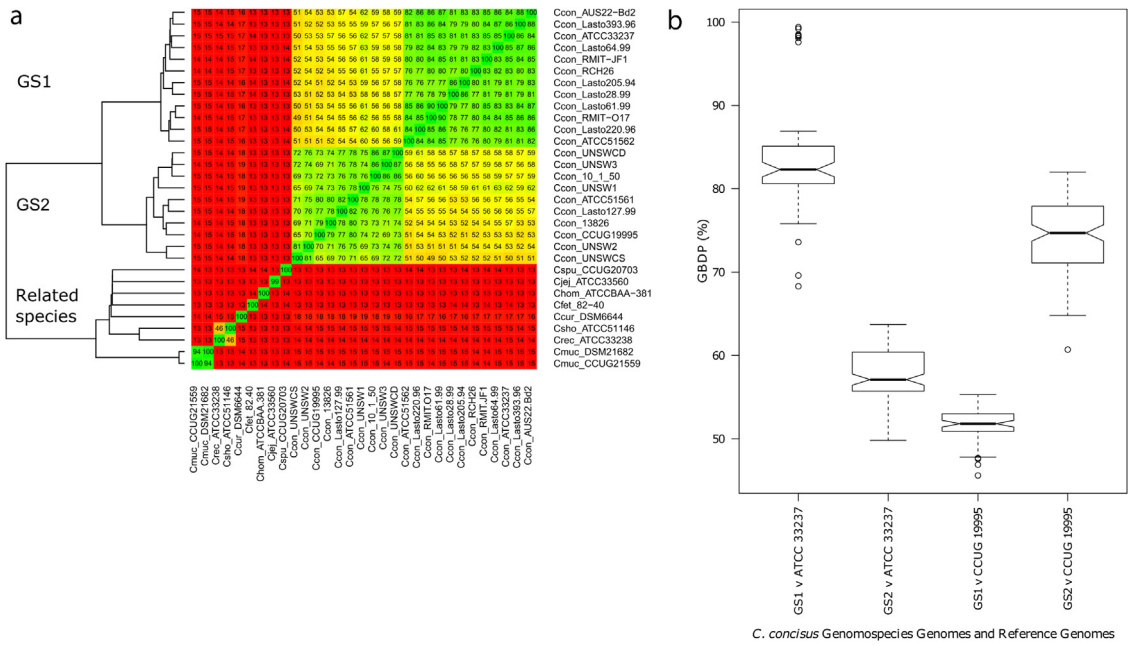

Fig. 3. Genome BLAST Distance Phylogeny (GBDP). Panel a: Heatmap for 22 *C. concisus* genomes and nine non-*C. concisus Campylobacter* genomes. using Formula 3 identities/total length. Panel b: Boxplot for 189 *C. concisus* genomes (67 GS1 and 122 GS2) relative to ATCC 32327^T and CCUG 19995. The results were rounded to whole numbers to aid readability. *Abbreviations*: Ccon *C. concisus*, Ccur *C. curvus*, Cfet *C. fetus*, Chom *C. hominis*, Cjej *C. jejuni*, Cmuc *C. mucosalis*, Crec *C. rectus*, Csho *C. showae*, Cspu *C. sputorum*.

When compared to ATCC 33237^T, the 66 GS1 genomes (the ATCC 33237^T result was excluded) had a mean *is*DDH for formula 3 of 83.40% (95% confidence interval 70.87–95.93%) and the 123 GS2 genomes had a mean *is*DDH of 57.51% (95% confidence interval 49.80–65.22%). The Welch two sample *t*-test for the ATCC 33237^T comparison had a *t*-value of 30.61, with 92.08 degrees of freedom and the *p*-value was <0.001. When compared to CCUG 19995, the 67 GS1 genomes had a mean *is*DDH of 51.77% (95% confidence interval 47.75–55.79%) and the 122 GS2 genomes (the CCUG 19995 result was excluded) had a mean score of 74.15% (95% confidence interval 63.62–84.68%) against ATCC 33237^T and CCUG 19995. The Welch two sample *t*-test for the CCUG 19995 comparison had a *t*-value of −41.74, with 171.56 degrees of freedom and the *p*-value was <0.001.

Of the 21042 nucleotides included in the rMLST analysis, 8506 (40.4%) were variable. The *C. concisus* strains formed two GS, separated from the related species, in the NeighborNet diagrams generated from the rMLST nucleotide sequences of 31 genomes (Fig. 4).

The numbers of genes for each COG functional group that were identified in the 31 genomes is shown in Table S3 (Supplementary Material). Two *C. concisus* GS clusters, with the same constituent strains and separated from the related species, were also observed in the NeighborNet generated from the COG analysis (Fig. 5).

The G+C content for the genome for strain GS2 AAUH-44UCsig-a was 47.88% and the size of the assembled genome was 3.9 Mb which suggested the genome was contaminated so this genome was excluded from the statistical analysis of the G+C content. The mean (and 95% confidence intervals) G+C content from the QUAST analysis for *C. concisus* GS1 ($n = 67$) and GS2 ($n = 122$) were 37.56% (37.28–37.85%) and 39.51% (39.14–39.89), respectively. The G+C content of ATCC 33237^T (GS1) and CCUG 19995 (GS2) was 37.62% and 39.39%, respectively. The Welch two sample *t*-test for the for G+C content had a *t*-value of −82.45, with 164.68 degrees of freedom and the *p*-value was <0.001.

For the phylogenetic tree based on 16S rRNA sequences (Fig. S3 in the Supplementary Material), the 18 sequences from the 17 GS1 genomes (Lasto205.94 was excluded) formed a single cluster with a bootstrap score of 48, but the 33 sequences from the 31 GS2 genomes do not. However, for the phylogenetic tree based on 23S rRNA sequences (Fig. S4 in the Supplementary Material), the 19 sequences from the 18 GS1 genomes formed a single cluster and the 33 sequences from the 31 GS2 genomes formed a second cluster with a bootstrap value of 100. When barrnap was used to extract 23S rRNA genes from the full set of 190 *C. concisus* genomes, one genome (AAUH-59UCpp-a) had no gene extracted and four genomes (AAUH-12CDtra2-a, H22O-S1, H25O-S1 and P21CDO-S2) had fragmented genes. These five genomes were excluded from the IQ-TREE analysis. The 185 *C. concisus* 23S rRNA genes were aligned and the most likely substitution model was TIM3+F+I+G4 (AC = CG, AT = GT and unequal base frequencies inferred from the alignment, invariable sites and a Gamma model with 4 parameters). The 65 GS1 genomes formed a single cluster (blue branches in Fig. S5 in the Supplementary Material well separated from the 120 GS2 genomes (red branches). For the phylogenetic tree generated from the *rrn* of 49 *C. concisus* genomes (Fig. 6), the 19 sequences from the 18 GS1 genomes formed a single cluster and the 33 sequences from the 31 GS2 genomes formed a second cluster, also with a bootstrap value of 100.

Identification of possible genomospecies-specific genes

Thirteen GS1-specific CDS were identified in the *C. concisus* LS-BSR results using the listAandB.pl script compared to nine identified using Roary [3] (Table 1). Similarly, 26 GS2-specific CDS were identified using LS-BSR compared to 14 using Roary (Table 2). The specificity of these CDS were evaluated by calculating BSR from the bit scores generated by BLAST searches of a local BLAST database and the NCBI nucleotide collections (nr/nt). Ten of the 16 GS1-specific CDS had BSR ≥ 0.80 for all 67 GS1 genomes. Five of these 10 GS1-specific CDS (two identified using both LS-BSR and Roary, two identified only using LS-BSR and one identified only using Roary) had no observed non-target hits with BSR ≥ 0.40. Of the 29 GS2-specific CDS, six had BSR ≥ 0.80 for all 123 GS2 genomes. Five of these six GS2-specific CDS (all identified by both LS-BSR and Roary) had no observed non-target hits with BSR ≥ 0.40.

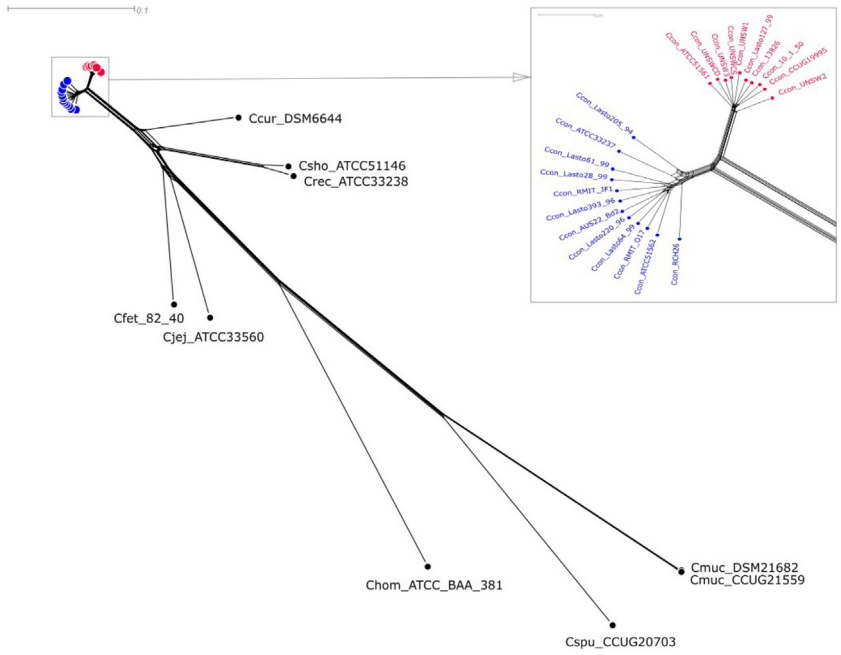

Fig. 4. NeighborNet of Ribosomal Multi-Locus Sequence Typing (rMLST) of 22 *C. concisus* genomes and nine non-*C. concisus Campylobacter* genomes using nucleotide sequences. *Abbreviations*: Ccon *C. concisus*, Ccur *C. curvus*, Cfet *C. fetus*, Chom *C. hominis*, Cjej *C. jejuni*, Cmuc *C. mucosalis*, Crec *C. rectus*, Csho *C. showae*, Cspu *C. sputorum*.

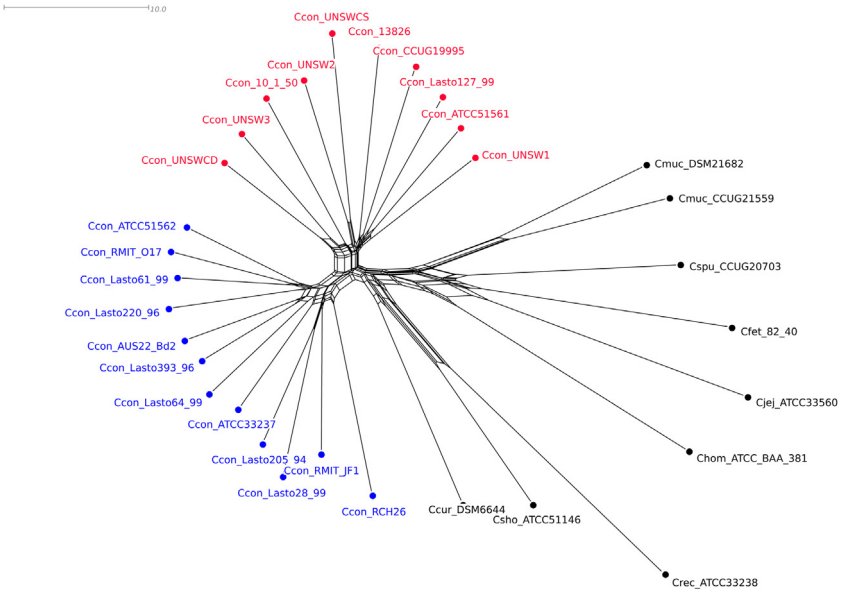

Fig. 5. NeighborNet of Clusters of Orthologous Genes (COG) Analysis of 22 *C. concisus* genomes and non-*C. concisus Campylobacter* genomes. Abbreviations are: Ccon *C. concisus*, Ccur *C. curvus*, Cfet *C. fetus*, Chom *C. hominis*, Cjej *C. jejuni*, Cmuc *C. mucosalis*, Crec *C. rectus*, Csho *C. showae*, Cspu *C. sputorum*.

Discussion

A plethora of tools based on WGS are now available for evaluating the genomic and phylogenetic relatedness of bacterial strains. For ANIb, ANIm, and GBDP-generated *is*DDH, species boundary values have been proposed [28–30,63]. For ANIm, intra-species pairs generally have >96% identity, inter-species pairs generally have <93% identity with an intermediate zone of 93-96% where species circumspection cannot be assured [63]. All *C. concisus* intra-GS ANIb and ANIm results for the 31-genome analyses from this study were between 93% and 96%, the intermediate zone [63]. Conversely, all of the *C. concisus* inter-GS ANIb and ANIm results fall well below the intermediate zone (<91%), with results similar to the *C. rectus/C. showae* inter-species pair, suggesting the two groups are from distinct GS. Richter and Rossello-Mora [29] previously proposed that ANI values of ~95-96% defined a species boundary. Almost all (98.9%) of the *C. concisus* GS2, and a small proportion (4.5%) of the all versus all GS1 intra-GS ANIm results were at least 95%. The mean ANIm score for the 66 GS1 genomes was compared to ATCC 33237^T was 94.53% and the mean ANIm score for 122 GS2 genomes compared to CCUG 19995 was 95.17%. The majority (88.5%) of the *C. concisus* GS1 and a small proportion (10.6%) of the *C. concisus* GS2 intra-GS ANIm scores from the 190-genome analysis were at least 95%. The GS were clearly separated in the boxplot of these results (Fig. 2b). In addition, the very low *p*-values generated by the Welch two-sample *t*-tests (both <0.001) confirm that the two GS have statistically different ANIm scores. These results supporting at least two distinct GS within the 22 *C. concisus* genomes.

Fig. 6. Phylogenetic tree for 49 *C. concisus* based on sequencing of the ribosomal RNA (*rrn*) operon. The neighbour-joining method [55] was used to generate the phylogenetic tree, which was performed using Molecular Evolutionary Genetic Analysis software version 10.0.05 (MEGA X) [53]. The optimal tree with the sum of branch length = 0.18436511 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches [84]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method [55] and are in the units of the number of base substitutions per site. This analysis involved 52 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 6877 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [52]. Bootstrap values of more than 75 are indicated on the internal branches.

Table 1
Comparison of the *C. concisus* GS1-specific CDS identified using Large Scale-BLAST Score Ratio (LS-BSR) and Roary.

Locus_tag	Gene products	COG[a]	Identified by	Length (nt)	Number of GS1 genomes with BSR[b] ≥ 0.8 (%)	Highest non-target hit with BSR[b] ≥ 0.4
CCON33237_1253	Transcriptional regulator, Crp family	COG0664; K	Both	645	67 (100%)[c]	None
CCON33237_1169	Phosphate ABC transporter, periplasmic substrate-binding protein PstS	COG0226; P	Both	873	67 (100%)	None
CCON33237_1170	Phosphate ABC transporter, permease protein PstC	COG0573; P	Both	858	67 (100%)[c]	*C. concisus* GS2 (BSR = 0.52)
CCON33237_1171	Phosphate ABC transporter, permease protein PstA	COG0581; P	Both	858	67 (100%)	*C. concisus* GS2 (BSR = 0.85)
CCON33237_1254	Tellurite-resistance/dicarboxylate transporter, TDT family	COG1275; P	Both	960	67 (100%)	*C. concisus* GS2 (BSR = 0.41)
CCON33237_1252	Putative NADH dehydrogenase	COG0446; S	Roary	1338	67 (100%)[c]	None
CCON33237_1262	Sel1 domain repeat-containing protein	COG0790; S	LS-BSR	444	53 (79%)	None
CCON33237_0317	Hypothetical protein	S	LS-BSR	546	67 (100%)	None
CCON33237_1832	Nitrous oxide reductase accessory protein (NosL)	S	LS-BSR	1122	62 (92.5%)[c]	*C. concisus* GS2 (BSR = 0.61)
CCON33237_0734	Hypothetical protein	T	Roary	747	62 (92.5%)[c]	None
CCON33237_0883	Transporter, AbgT family		Both	1539	67 (100%)[c]	*C. showae* (BSR = 0.41)
CCON33237_1674	Hypothetical protein		LS-BSR	495	64 (95.5%)	*C. showae* (BSR = 0.43)
CCON33237_1772	Hypothetical protein		Roary	159	67 (100%)	*C. concisus* GS2 (BSR = 0.82)
	Hypothetical protein[d]		LS-BSR	162	67 (100%)	None
	Hypothetical protein[d]		LS-BSR	183	47 (70.1%)	None
	Hypothetical protein[d]		LS-BSR	204	64 (95.5%)[c]	None

[a] Clusters of Orthologous Groups (COG) numbers (if available) followed by functional groups where K = Transcription; P = Inorganic ion transport and metabolism; S = Function unknown; and T = Signal transduction mechanisms.
[b] BLAST Score Ratio.
[c] The CDS in at least one genome was fragmented.
[d] No gene annotated on GenBank file, Prokka annotation was hypothetical protein.

GDBP aims to provide an *in silico* result equivalent to wet laboratory DNA-DNA hybridisation, and so also has a proposed species boundary of 70% [30]. For formulae 1 and 3, all GS1 and 15.6% of the GS2 intra-GS pairwise results exceeded 70% but 5.8% of the *C. concisus* inter-GS results also exceeded 70% using formula 1. Only the *C. mucosalis* intra-species results exceed this value using the recommended formula 2. Formula 3 was also found to be optimal when analysing selected genomes in the *Campylobacteraceae* and *Helicobacteraceae* families [32]. Thus, formula 3 was preferred for this dataset and was used for the 190-genome set where the intra-GS means were 83.40% (GS1) and 74.15% (GS2) and the inter-GS means were 57.51% (GS1) and 51.77% (GS2). All GBDP-generated *is*DDH heatmaps (Fig. 3a and Fig. S2) clustered the same *C. concisus* genomes into GS1 and GS2, as observed for ANI, with the related species in a separate branch. When Welch two sample *t*-tests were performed on the formula 3 *is*DDH results for the 190-genome set against ATCC 33237^T and CCUG 19995, the very small *p*-values (both <0.001) confirm that the two GS have statistically different GBDP-generated *is*DDH results. These results add support to at least two GS within *C. concisus*, even if the proposed species boundary isn't consistently achieved.

A total of 38 outlying results were observed in the ANIm and GBDP-generated *is*DDH boxplots however 12 genomes accounted for 27 of the outlying results. Six GS1 genomes (AAUH-16UCf, AAUH-16UCf2, AAUH-16UCf3, AAUH-16UCo-a, AAUH-8HCo, AAUH-9HCasc and H15O-S1) were high outliers for both ANIm and GBDP-generated *is*DDH compared to ATCC 33237^T suggesting that these genomes were more closely genetically related to the type strain that the other GS1 genomes. Three GS1 genomes (AAUH-11UCdes-a and AAUH-12CDo and AAUH-5CDo) were low outliers for GBDP-generated *is*DDH compared to both ATCC 33237^T and CCUG 19995 with AAUH-12CDo also being a high outlier for ANIm compared to CCUG 19995 and AAUH-5CDo had the lowest observed G+C content (37.21%). The GS1 genome AAUH-3HCo was a high outlier for ANIm compared to ATCC 33237^T and a low outlier compared to CCUG 19995 and it also had the highest G+C content of the genomes included in the statistical analysis. The GS2 genome AAUH-44UCsig-a was a low outlier for GBDP-generated *is*DDH compared to both ATCC 33237^T and CCUG 19995. This genome had a G+C content of 47.88% and a size of 3.9 Mb suggesting it was contaminated which may also explain the low GBDP-generated *is*DDH results. The ANIm results for this genome were within the whiskers of the boxplots suggesting that ANIm is more tolerant of contamination than GBDP-generated *is*DDH.

rMLST and COG analyses separated the 22 *C. concisus* genomes into the same two GS generated by ANI and GBDP-generated *is*DDH. rMLST has been proposed to provide phylogenetic information over a very broad taxonomic range [33]; has been shown to provide species-level identification as well as sub-species clustering for the closely related species *C. jejuni* and *C. coli* [64]; and provided data towards the classification of a new *Campylobacter* species [65]. rMLST has also been applied to the classification of new species within the order Enterobacterales [66] and provided clarification of classification within the *Neisseria* genus [67,68]. The distribution of COG in the core- and pan-genomes of four genera (*Bifidobacterium, Lactobacillus, Leuconostoc* and *Enterococcus*) has also been compared [69] although the comparisons did not extend to evaluating the phylogenetic relationship of species within each genus.

rMLST and COG analyses do not have numerical cut-offs for species circumscription but rely on clustering the genomes under investigation to provide evidence of phylogenetic relatedness. The NeighborNets generated from the analysis of the 22 *C. concisus* genomes using these two approaches to phylogenetic inference resulted in the same *C. concisus* GS being observed as those seen in the ANI and GBDP-generated *is*DDH with the type strain ATCC 33237 representing GS1 and CCUG 19995 representing GS2.

Before WGS, the G+C content of bacterial strains was estimated from the physical properties induced in extracted or digested

Table 2
Comparison of the *C. concisus* GS2-specific CDS identified using Large Scale-BLAST Score Ratio (LS-BSR) and Roary.

Locus_tag	Gene products	COG[a]	Identified by	Length (nt)	Number of GS2 genomes with BSR[b] ≥ 0.8 (%)	Highest non-target hit with BSR[b] ≥ 0.4
CCC13826_0436	Oxidoreductase, FAS/FMN-binding	COG1902; C	Both	1023	114 (92.7%)	*C. concisus* GS1 (BSR = 0.55)
CCC13826_1636	Aquaporin Z	COG0580; G	Roary	711	118 (95.9%)	*C. showae* (BSR = 0.82)
CCC13826_1452	Translation initiation inhibitor	COG0251; J	LS-BSR	384	117 (95.1%)[c]	None
CCC13826_0272	DNA-3-methyladenine glycosylase I	COG2818;L	Roary	558	109 (88.6%)	None
CCC13826_1511	Aspartate racemase	COG1794; M	Both	693	123 (100%)[c]	None
CCC13826_1584	Twitching motility protein	COG2805; N, U	Both	1161	123 (100%)[c]	None
CCC13826_1830	FlavoCytochrome c heme subunit	P	LS-BSR	438	123 (100%)[c]	*C. concisus* GS1 (BSR = 0.90)
CCC13826_0895	Periplasmic protein	COG3672; S	Both	627	122 (99.2%)	None
CCC13826_1263	Rhomboid family protein	COG0705; S	Roary	576	121 (98.4%)[c]	None
CCC13826_1451	MukF protein	COG2964; S	LS-BSR	621	108 (87.8%)	None
CCC13826_1702	LemA protein	COG1704; S	Both	564	123 (100%)	None
CCC13826_2180	Beta-lactamase HcpA (Cysteine-rich 28 kDa protein)	COG0790; S	Both	960	118 (95.9%)[c]	*C. concisus* GS1 (BSR = 0.42)
CCC13826_1703	Hypothetical protein	S	LS-BSR	891	120 (97.6%)	None
CCC13826_0721	PAS/PAC sensor signal transduction histidine kinase	T	Both	741	122 (99.2%)	*C. concisus* GS1 (BSR = 0.60)
CCC13826_0177	Na+/H+ antiporter NhaC		Both	1413	123 (100%)	None
CCC13826_0178	Beta-aspartyl peptidase		Both	1131	123 (100%)[c]	None
CCC13826_0466	Luciferase family protein		LS-BSR	1041	67 (54.5%)	None
CCC13826_0565	Hypothetical protein		LS-BSR	516	111 (90.2%)	None
CCC13826_0683	Conserved hypothetical protein		LS-BSR	372	72 (58.5%)	*C. concisus* GS1 (BSR = 0.58)
CCC13826_0780	Hypothetical protein		LS-BSR	519	110 (89.4%)	*C. concisus* GS1 (BSR = 0.52)
CCC13826_1402	Glyoxalase II		Both	834	121 (98.4%)	*C. concisus* GS1 (BSR = 0.78)
CCC13826_1540	Hydroxylamine reductase		Both	1329	122 (99.2%)	*C. curvus* (BSR – 0.47)
CCC13826_1571	Hypothetical protein		LS-BSR	612	86 (69.9%)	None
CCC13826_1698	Hypothetical protein		LS-BSR	540	87 (70.7%)	None
CCC13826_1704	Hypothetical protein		LS-BSR	654	101 (82.1%)	*C. concisus* GS1 (BSR = 0.74)
CCC13826_1813	PQQ enzyme repeat domain protein		LS-BSR	930	118 (95.9%)	*C. concisus* GS1 (BSR = 0.85)
CCC13826_1826	Conserved hypothetical protein		LS-BSR	519	79 (64.2%)	*C. concisus* GS1 (BSR = 0.71)
CCC13826_1877	Conserved hypothetical protein		LS-BSR	540	108 (87.8%)	*C. showae* (BSR = 0.44)
CCC13826_2181	Putative beta-lactamase HcpC (Cysteine-rich protein)		LS-BSR	441	117 (95.1%)	None

[a] Clusters of Orthologous Groups (COG) numbers (if available) followed by functional groups where C = Energy production and conversion; G = Carbohydrate transport and metabolism; J = Translation, ribosomal structure and biogenesis; L = Replication, recombination and repair; M = Cell wall/membrane/envelope biogenesis; N = Cell motility; P = Inorganic ion transport and metabolism; S = Function unknown; and U = Intracellular trafficking, secretion, and vesicular transport.
[b] BLAST Score Ratio.
[c] The CDS in at least one genome was fragmented.

genomic DNA [70] and reported as molar percentages (mol%) [71]. The actual G+C content can now be calculated directly from genome sequences. The G+C content range from the 189 study genomes (37.21-39.97%, AAUH-44UCsig-a excluded) were well within the estimated G+C content range (37-41 mol%) reported for *C. concisus* [71]. Meier-Kolthoff, Klenk and Goker [70] have proposed that G+C differences should be no more than 1% within a species if determined using genome sequences. The median G+C contents of the two *C. concisus* GS described here differ by 2%. The G+C content calculated directly from the ATCC 33237^T (37.62%) and CCUG 19995 (39.39%) genomes were similar to the results observed using thermal denaturation (37.9 mol% and 39.9 mol%, respectively) [10] which also differ by 2%. The Welch two-sample *t*-test of the G+C contents generated a very low *p*-value (<0.001) confirming that the two GS have statistically different G+C contents. These differences in G+C content between GS1 and GS2 are consistent with those reported in the literature [72,73].

Larsen, Cosentino, Lukjancenko, Saputra, Rasmussen, Hasman, Sicheritz-Ponten, Aarestrup, Ussery and Lund [74] compared the ability of SpeciesFinder, rMLST, TaxonomyFinder and KmerFinder to predict species identification of assembled draft or complete genomes from 695 isolates representing the bacterial diversity known at the time. They found that SpeciesFinder, which uses full length 16S rRNA sequence, and rMLST performed poorly and KmerFinder performed the best [74]. rMLST performed somewhat better for 10,407 draft genomes generated from the sequence read archive (SRA) but the method consistently made incorrect identification for a number of closely related species for both the original 695 and subsequent 10,407 groups of genomes [74]. Since Larsen and colleagues generated their own rMLST database, the poor performance of this method in this comparison may relate to the criteria (only hits with at least 95% identity and 95% coverage were considered potential matches) applied by the authors. Genomes can now be uploaded to the PubMLST rMLST website (https://pubmlst.org/rmlst/) where a predicted taxon is provided and a list of the best matches for each gene is generated. rMLST separated the 31 genomes in this study into three main branches corresponding to the two *C. concisus* GS and the related species which was concordant with the ANI, GBDP-generated *is*DDH and COG results.

The phylogenetic tree generated using 16S rRNA gene sequences extracted from 49 *C. concisus* genomes demonstrated that there was little diversity within the *C. concisus* 16S rRNA genes and although the GS1 genomes formed a single cluster, the GS2 genomes did not. Chung, Tay, Octavia, Chen, Liu, Ma, Lan, Riordan, Grimm and Zhang [3] and Huq, Van, Gurtler, Elshagmani, Allemailem, Smooker and Istivan [5] also observed little diversity in the 16S rRNA gene sequences of *C. concisus* isolates. The discriminatory power of the 16S rRNA gene sequence has previously been shown to be poor for some species within the *Campylobacter* genus [75–77]. In addition, WGS-based analyses such as ANI are considered superior to 16S

rRNA sequence and other single gene analyses for studying phylogeny because they are based on a much larger part of the genome, are less susceptible to the effects of horizontal gene transfer, and have better resolution for discriminating both distantly and closely related bacteria [78]. However, analysis based on 23S rRNA and *rrn* operon gene sequences provided better discriminatory power with the GS1 and GS2 genomes forming two clear clusters in the phylogenetic trees generated using gene sequences extracted from 49 *C. concisus* genomes. These results are consistent with the 23S rRNA and both 23S rRNA and *rrn* observations of Chung, Tay, Octavia, Chen, Liu, Ma, Lan, Riordan, Grimm and Zhang [3] and Huq, Van, Gurtler, Elshagmani, Allemailem, Smooker and Istivan [5], respectively, where *C. concisus* genomes formed two clear clusters. In contrast, Gemmell, Berry, Mukhopadhya, Hansen, Nielsen, Bajaj-Elliott, Nielsen and Hold [79] used 88 *C. concisus* genomes and found the 23S rRNA phylogenetic tree formed two clusters with both containing GS1 and GS2 genomes. Our larger 23S rRNA phylogenetic tree of 185 genomes, including 36 of the 88 from the Gemmell study, had clear separation of the GS1 and GS2 sequences consistent with the earlier observation of Chung, Tay, Octavia, Chen, Liu, Ma, Lan, Riordan, Grimm and Zhang [3] and the smaller analysis in this study.

The proposed species boundaries for ANI and GBDP-generated *is*DDH were not consistently exceeded for *C. concisus* intra-GS pairs, however two clusters containing the same strains were consistently produced by all methods except 16S rRNA gene sequencing. Previous studies into the genetic diversity of *C. concisus* have reported significant heterogeneity, even within GS [3,18–21,23–25]. The separation of *C. concisus* genomes into these same GS has also been observed using the Roary core-genome based [3,11] and cgMLST-based [7] phylogenetic trees. Where isolates were shared between studies, the GS assignments were congruent.

The overlap of GS-specific CDS generated by LS-BSR (with listAandB.pl) and Roary demonstrates congruence between the methods. The variation in genomes making up the two pan-genomic analyses is also likely to have affected the CDS identified. Nine genomes (2 GS1 and 7 GS2) were common to both studies, eight genomes were unique to LS-BSR and 27 were unique to Roary. Five GS1-specific CDS were present (BSR $\geq$ 0.80) in all 67 GS1 genomes and had no significant matches in non-target genomes (BSR <0.40) from both local and NCBI nt BLAST databases. Two of these five GS1-specific CDS were identified using both methods, two were identified only by LS-BSR and one was identified only by Roary. Similarly, six GS2-specific CDS were present (BSR $\geq$ 0.80) in all 123 GS2 genomes and had no significant matches in non-target genomes (BSR <0.40) from both local and NCBI nt BLAST databases. All six GS2-specific CDS were identified using both LS-BSR and Roary. These GS-specific CDS could be used to develop PCR assays for differentiating the GS for laboratories not routinely performing WGS. In addition, closer examination of these genes may identify phenotypic differences that can be used to differentiate the two GS and thus allow for circumscription of the GS and description of a new species.

Within the two GS there are no clear clusters of genomes that are observed across multiple analyses suggesting that any groupings below this level are not phylogenetically meaningful. The draft GenBank *Campylobacter* sp. genome 10_1_50, which has only been identified to the genus level and tentatively assigned to *C. concisus* on the basis of sequence similarity with the *C. concisus* GS2 23S rRNA CON2 primer and taxon-specific CDS, clustered within *C. concisus* GS2 for all analyses.

The term genomospecies is used when there are genetic distinct groups within a species while no diagnostic phenotypic difference has yet been identified [17]; previous research using AFLP fingerprinting have indicated up to six such groups [25]. We were unable to culture representatives of these for analysis, but the current study examined 190 genomes and found evidence only for two *C. concisus* genomospecies. Since the AFLP profiles of the proposed additional four genomospecies [25] are relatively simplistic, it is likely their classification was aberrant and the current data strongly support the existence of two genetically distinct groups (GS1 and GS2) only.

The role *C. concisus* GS have in human disease has not yet been established. Some early studies that evaluated *C. concisus* GS reported an association between GS and disease or virulence potential [19,21,23,26]. Later studies found no significant differences between GS and disease [3,8,80] but associations between GS and isolation site were reported [7,80]. More recent studies have also reported associations between disease and GS [81] or GS subgroups [72]. Although assignment of isolates to GS may not provide a clear indication of virulence potential, it does provide more complete taxonomic information that, if collected routinely, may contribute to a more complete picture of the role of *C. concisus* GS in various human diseases. The combination of ANIm and GBDP-generated *is*DDH (against ATCC 33237^T and CCUG 19995), plus G+C content provides cluster-independent GS discrimination where WGS are available [81].

In conclusion, the same constituent *C. concisus* genomes were grouped into GS by the five multigene analyses (ANI, GBDP-generated *is*DDH, rMLST, COG and *rrn*). 23S rRNA sequence analysis also generated the same two GS; however, the 16S rRNA sequence analysis did not have sufficient diversity to clearly differentiate the two GS. The G+C content of the GS1 and GS2 genomes were also clearly different. GS assignments were congruent with previous publications. The assignment of *Campylobacter* sp. 10_1_50 to *C. concisus* GS2 demonstrates that WGS may be a cost-effective method for tentatively assigning partially characterised isolates to a taxonomic group, especially for laboratories not specialising in the phenotypic identification of the genus of an isolate of interest. In order to describe a new species there must be phenotypic and genetic cohesion amongst members of the new species and a diagnostic phenotype that differentiates it from its closest relatives [82]. To date, a routinely applicable, differentiating phenotype that would allow the formal circumscription of individual species from the current *C. concisus* species-complex in accordance with minimal standards [32] has not yet been identified. The comparative genomic analyses provide useful phylogenetic information that can help inform classification and the identification of taxon-specific genes could lead to phenotypic differentiation of the GS facilitating the description of a new species. Correlation with recently described protein markers detected by mass spectrometry [83] would also be of interest. These taxon-specific genes could also be used as the targets for molecular methods such as polymerase chain reaction (PCR) to provide GS discrimination for laboratories that are not performing WGS routinely. The combination of ANIm and GBDP-generated *is*DDH (against ATCC 33237^T and CCUG 19995), plus G+C content provides cluster-independent GS discrimination suitable for routine use for laboratories that perform WGS.

Declarations of interest

None

Authors contribution

Angela J. Cornelius: Conceptualisation, methodology, formal analysis, writing – original draft, visualisation. **Mohsina Huq**: Resources, formal analysis, visualisation, writing – original draft, writing – review & editing. **Stephen L.W. On**: Conceptualisation, writing – review & editing, supervision. **Nigel P. French**: Writing – review & editing, supervision. **Olivier Vandenberg**: Writing –

review & editing, supervision. **William G. Miller**: Resources, writing – review & editing. **Albert J. Lastovica**: Resources, writing – review & editing. **Taghrid Istivan**: Resources, writing – review & editing, supervision. **Patrick J. Biggs**: Conceptualisation, methodology, software, resources, writing – review & editing, visualisation, supervision.

Acknowledgements

Funding: This work was supported by the New Zealand Ministry for Business, Innovation and Employment through ESR Strategic Science Investment Funding.

We are grateful to Dr Li Zhang (University of New South Wales, Australia) for sharing the assembled genomes included in the Chung 2016 paper and to Dr Pierre-Yves Dupont (Institute of Environmental Science and Research, New Zealand) for help with the larger 23S rRNA analysis and valuable critical review of the draft manuscript. Thank you also to Dr Joanne Kingsbury (Institute of Environmental Science and Research, New Zealand) for valuable critical review of the draft manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at https://doi.org/10.1016/j.syapm.2021.126187.

References

[1] Macuch, P.J., Tanner, A.C. (2000) *Campylobacter* species in health, gingivitis, and periodontitis. J. Dent. Res. 79, 785–792.
[2] Blackett, K.L., Siddhi, S.S., Cleary, S., Steed, H., Miller, M.H., Macfarlane, S., Macfarlane, G.T., Dillon, J.F. (2013) Oesophageal bacterial biofilm changes in gastro-oesophageal reflux disease, Barrett's and oesophageal carcinoma: association or causality? Alim. Pharmacol. Ther. 37, 1084–1092.
[3] Chung, H.K., Tay, A., Octavia, S., Chen, J., Liu, F., Ma, R., Lan, R., Riordan, S.M., Grimm, M.C., Zhang, L. (2016) Genome analysis of *Campylobacter concisus* strains from patients with inflammatory bowel disease and gastroenteritis provides new insights into pathogenicity. Sci. Rep. 6, 38442.
[4] Hansen, R., Berry, S.H., Mukhopadhya, I., Thomson, J.M., Saunders, K.A., Nicholl, C.E., Bisset, W.M., Loganathan, S., Mahdi, G., Kastner-Cole, D., Barclay, A.R., Bishop, J., Flynn, D.M., McGrogan, P., Russell, R.K., El-Omar, E.M., Hold, G.L. (2013) The microaerophilic microbiota of *de-novo* paediatric inflammatory bowel disease: the BISCUIT study. PLoS One 8, e58825.
[5] Huq, M., Van, T.T.H., Gurtler, V., Elshagmani, E., Allemailem, K.S., Smooker, P.M., Istivan, T. (2017) The ribosomal RNA operon (*rrn*) of *Campylobacter concisus* supports molecular typing to genomospecies level. Gene Rep. 6, 8–14.
[6] Kaakoush, N.O., Deshpande, N.P., Wilkins, M.R., Tan, C.G., Burgos-Portugal, J.A., Raftery, M.J., Day, A.S., Lemberg, D.A., Mitchell, H. (2011) The pathogenic potential of *Campylobacter concisus* strains associated with chronic intestinal diseases. PLoS One 6, e29045.
[7] Kirk, K.F., Meric, G., Nielsen, H.L., Pascoe, B., Sheppard, S.K., Thorlacius-Ussing, O., Nielsen, H. (2018) Molecular epidemiology and comparative genomics of *Campylobacter concisus* strains from saliva, faeces and gut mucosal biopsies in inflammatory bowel disease. Sci. Rep. 8, 1902.
[8] Mahendran, V., Riordan, S.M., Grimm, M.C., Tran, T.A., Major, J., Kaakoush, N.O., Mitchell, H., Zhang, L. (2011) Prevalence of *Campylobacter* species in adult Crohn's disease and the preferential colonization sites of *Campylobacter* species in the human intestine. PLoS One 6, e25417.
[9] Zhang, L., Man, S.M., Day, A.S., Leach, S.T., Lemberg, D.A., Dutt, S., Stormon, M., Otley, A., O'Loughlin, E.V., Magoffin, A., Ng, P.H., Mitchell, H. (2009) Detection and isolation of *Campylobacter* species other than *C. jejuni* from children with Crohn's disease. J. Clin. Microbiol. 47, 453–455.
[10] Vandamme, P., Falsen, E., Pot, B., Hoste, B., Kersters, K., De Ley, J. (1989) Identification of EF group 22 campylobacters from gastroenteritis cases as *Campylobacter concisus*. J. Clin. Microbiol. 27, 1775–1781.
[11] Liu, F., Ma, R., Tay, C.Y.A., Octavia, S., Lan, R., Chung, H.K.L., Riordan, S.M., Grimm, M.C., Leong, R.W., Tanaka, M.M., Connor, S., Zhang, L. (2018) Genomic analysis of oral *Campylobacter concisus* strains identified a potential bacterial molecular marker associated with active Crohn's disease. Emerg. Microbes Infect. 7, 64.
[12] de Vries, J.J., Arents, N.L., Manson, W.L. (2008) *Campylobacter* species isolated from extra-oro-intestinal abscesses: a report of four cases and literature review. Eur. J. Clin. Microbiol. Infect. Dis. 27, 1119–1123.
[13] Deshpande, N.P., Kaakoush, N.O., Wilkins, M.R., Mitchell, H.M. (2013) Comparative genomics of *Campylobacter concisus* isolates reveals genetic diversity and provides insights into disease association. BMC Genomics 14, 585.
[14] Nielsen, H.L., Engberg, J., Ejlertsen, T., Nielsen, H. (2013) Clinical manifestations of *Campylobacter concisus* infection in children. Pediatr. Infect. Dis. J. 32, 1194–1198.
[15] Engberg, J., On, S.L., Harrington, C.S., Gerner-Smidt, P. (2000) Prevalence of *Campylobacter, Arcobacter, Helicobacter*, and *Sutterella* spp. in human fecal samples as estimated by a reevaluation of isolation methods for Campylobacters. J. Clin. Microbiol. 38, 286–291.
[16] Van Etterijck, R., Breynaert, J., Revets, H., Devreker, T., Vandenplas, Y., Vandamme, P., Lauwers, S. (1996) Isolation of *Campylobacter concisus* from feces of children with and without diarrhea. J. Clin. Microbiol. 34, 2304–2306.
[17] Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, M.I., Moore, L.H., Moore, W.E.C., Murray, R.G.E., Stackebrandt, E., Starr, M.P., Truper, H.G. (1987) Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. 37, 463–464.
[18] Matsheka, M.I., Elisha, B.G., Lastovica, A.L., On, S.L. (2002) Genetic heterogeneity of *Campylobacter concisus* determined by pulsed field gel electrophoresis-based macrorestriction profiling. FEMS Microbiol. Lett. 211, 17–22.
[19] Aabenhus, R., On, S.L., Siemer, B.L., Permin, H., Andersen, L.P. (2005) Delineation of *Campylobacter concisus* genomospecies by amplified fragment length polymorphism analysis and correlation of results with clinical data. J. Clin. Microbiol. 43, 5091–5096.
[20] Engberg, J., Bang, D.D., Aabenhus, R., Aarestrup, F.M., Fussing, V., Gerner-Smidt, P. (2005) *Campylobacter concisus*: an evaluation of certain phenotypic and genotypic characteristics. Clin. Microbiol. Infect. 11, 288–295.
[21] Ismail, Y., Mahendran, V., Octavia, S., Day, A.S., Riordan, S.M., Grimm, M.C., Lan, R., Lemberg, D., Tran, T.A., Zhang, L. (2012) Investigation of the enteric pathogenic potential of oral *Campylobacter concisus* strains isolated from patients with inflammatory bowel disease. PLoS One 7, e38217.
[22] Istivan, T.S., Coloe, P.J., Fry, B.N., Ward, P., Smith, S.C. (2004) Characterization of a haemolytic phospholipase A^2 activity in clinical isolates of *Campylobacter concisus*. J. Med. Microbiol. 53, 483–493.
[23] Kalischuk, L.D., Inglis, G.D. (2011) Comparative genotypic and pathogenic examination of *Campylobacter concisus* isolates from diarrheic and non-diarrheic humans. BMC Microbiol. 11, 53.
[24] Miller, W.G., Chapman, M.H., Yee, E., On, S.L., McNulty, D.K., Lastovica, A.J., Carroll, A.M., McNamara, E.B., Duffy, G., Mandrell, R.E. (2012) Multilocus sequence typing methods for the emerging *Campylobacter* Species *C. hyointestinalis, C. lanienae, C. sputorum, C. concisus*, and *C. curvus*. Front. Cell. Infect. Microbiol. 2, 45.
[25] On, S.L.W., Siemer, B.L., Brandt, S.M., Chung, P., Lastovica, A.J. (2013) Characterisation of *Campylobacter concisus* strains from South Africa using amplified fragment length polymorphism (AFLP) profiling and a genomospecies-specific polymerase chain reaction (PCR) assay: identification of novel genomospecies and correlation with clinical data. Afr. J. Microbiol. Res. 7, 1845–1851.
[26] Istivan, T. 2005 Molecular characterisation of *Campylobacter concisus*: a potential etiological agent of gastroenteritis in children. In: Biotechnology and Environmental Biology, School of Applied Sciences, RMIT University, Melbourne, Australia, pp. , 326.
[27] Konstantinidis, K.T., Ramette, A., Tiedje, J.M. (2006) Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. Appl. Environ. Microbiol. 72, 7286–7293.
[28] Konstantinidis, K.T., Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. U.S.A. 102, 2567–2572.
[29] Richter, M., Rossello-Mora, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. U.S.A. 106, 19126–19131.
[30] Auch, A.F., Klenk, H.P., Goker, M. (2010) Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. Stand. Genomic Sci. 2, 142–148.
[31] Auch, A.F., von Jan, M., Klenk, H.P., Goker, M. (2010) Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand. Genomic Sci. 2, 117–134.
[32] On, S.L.W., Miller, W.G., Houf, K., Fox, J.G., Vandamme, P. (2017) Minimal standards for describing new species belonging to the families *Campylobacteraceae* and *Helicobacteraceae*: *Campylobacter*, *Arcobacter*, *Helicobacter* and *Wolinella* spp. Int. J. Syst. Evol. Microbiol. 67, 5296–5311.
[33] Jolley, K.A., Bliss, C.M., Bennett, J.S., Bratcher, H.B., Brehony, C., Colles, F.M., Wimalarathna, H., Harrison, O.B., Sheppard, S.K., Cody, A.J., Maiden, M.C. (2012) Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology (Reading, England) 158, 1005–1015.
[34] Tatusov, R.L., Koonin, E.V., Lipman, D.J. (1997) A genomic perspective on protein families. Science (New York, N.Y.) 278, 631–637.
[35] Galperin, M.Y., Makarova, K.S., Wolf, Y.I., Koonin, E.V. (2015) Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 43, D261–D269.
[36] Chaudhari, N.M., Gupta, V.K., Dutta, C. (2016) BPGA – an ultra-fast pan-genome analysis pipeline. Sci. Rep. 6, 24373.
[37] Baig, A., McNally, A., Dunn, S., Paszkiewicz, K.H., Corander, J., Manning, G. (2015) Genetic import and phenotype specific alleles associated with hyper-invasion in *Campylobacter jejuni*. BMC Genomics 16, 852.
[38] Benedict, M.N., Henriksen, J.R., Metcalf, W.W., Whitaker, R.J., Price, N.D. (2014) ITEP: an integrated toolkit for exploration of microbial pan-genomes. BMC Genomics 15, 8.
[39] Chen, P.E., Shapiro, B.J. (2015) The advent of genome-wide association studies for bacteria. Curr. Opin. Microbiol. 25, 17–24.
[40] Laing, C., Buchanan, C., Taboada, E.N., Zhang, Y., Kropinski, A., Villegas, A., Thomas, J.E., Gannon, V.P. (2010) Pan-genome sequence analysis using Panseq:

an online tool for the rapid analysis of core and accessory genomic regions. BMC Bioinform. 11, 461.
[41] Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T., Shendure, J. (2015) Large-scale genomic sequencing of extraintestinal pathogenic *Escherichia coli* strains. Genome Res. 25, 119–128.
[42] Thepault, A., Meric, G., Rivoal, K., Pascoe, B., Mageiros, L., Touzain, F., Rose, V., Beven, V., Chemaly, M., Sheppard, S.K. (2017) Genome-wide identification of host-segregating epidemiological markers for source attribution in *Campylobacter jejuni*. Appl. Environ. Microbiol. 83.
[43] Sahl, J.W., Caporaso, J.G., Rasko, D.A., Keim, P. (2014) The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. PeerJ 2, e332.
[44] Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T., Fookes, M., Falush, D., Keane, J.A., Parkhill, J. (2015) Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics (Oxford, England) 31, 3691–3693.
[45] Cornelius, A.J., Miller, W.G., Lastovica, A.J., On, S.L.W., French, N.P., Vandenberg, O., Biggs, P.J. (2017) Complete genome sequence of *Campylobacter concisus* ATCC 33237^T and draft genome sequences for an additional eight well-characterized *C. concisus* strains. Genome Announcements 5.
[46] Deshpande, N.P., Kaakoush, N.O., Mitchell, H., Janitz, K., Raftery, M.J., Li, S.S., Wilkins, M.R. (2011) Sequencing and validation of the genome of a *Campylobacter concisus* reveals intra-species diversity. PLoS One 6, e22170.
[47] Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics (Oxford, England) 30, 2068–2069.
[48] Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
[49] Maddison, D.R., Swofford, D.L., Maddison, W.P. (1997) NEXUS: an extensible file format for systematic information. Syst. Biol. 46, 590–621.
[50] Huson, D.H., Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.
[51] Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics (Oxford, England) 29, 1072–1075.
[52] Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549.
[53] Thompson, J.D., Higgins, D.G., Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.
[54] Saitou, N., Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
[55] Tamura, K., Nei, M., Kumar, S. (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. U.S.A. 101, 11030–11035.
[56] Katoh, K., Misawa, K., Kuma, K., Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066.
[57] Katoh, K., Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
[58] Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S. (2018) UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35, 518–522.
[59] Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.
[60] Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589.
[61] Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B., Bork, P. (2006) Toward automatic reconstruction of a highly resolved tree of life. Science (New York, N.Y.) 311, 1283–1287.
[62] Sahl, J.W., Gillece, J.D., Schupp, J.M., Waddell, V.G., Driebe, E.M., Engelthaler, D.M., Keim, P. (2013) Evolution of a pathogen: a comparative genomics analysis identifies a genetic pathway to pathogenesis in Acinetobacter. PLoS One 8, e54287.
[63] Rossello-Mora, R., Amann, R. (2015) Past and future species definitions for Bacteria and Archaea. Syst. Appl. Microbiol. 38, 209–216.
[64] Read, D.S., Woodcock, D.J., Strachan, N.J., Forbes, K.J., Colles, F.M., Maiden, M.C., Clifton-Hadley, F., Ridley, A., Vidal, A., Rodgers, J., Whiteley, A.S., Sheppard, S.K. (2013) Evidence for phenotypic plasticity among multihost *Campylobacter jejuni* and *C. coli* lineages, obtained using ribosomal multilocus sequence typing and Raman spectroscopy. Appl. Environ. Microbiol. 79, 965–973.
[65] Grange, Z.L., Gartrell, B.D., Biggs, P.J., Nelson, N.J., Anderson, M., French, N.P. (2016) Microbial genomics of a host-associated commensal bacterium in fragmented populations of endangered takahe. Microb. Ecol. 71, 1020–1029.
[66] Facey, P.D., Meric, G., Hitchings, M.D., Pachebat, J.A., Hegarty, M.J., Chen, X., Morgan, L.V., Hoeppner, J.E., Whitten, M.M., Kirk, W.D., Dyson, P.J., Sheppard, S.K., Del Sol, R. (2015) Draft genomes, phylogenetic reconstruction, and comparative genomics of two novel cohabiting bacterial symbionts isolated from *Frankliniella occidentalis*. Genome Biol. Evol. 7, 2188–2202.
[67] Bennett, J.S., Jolley, K.A., Earle, S.G., Corton, C., Bentley, S.D., Parkhill, J., Maiden, M.C. (2012) A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus *Neisseria*. Microbiology (Reading, England) 158, 1570–1580.
[68] Bennett, J.S., Jolley, K.A., Maiden, M.C. (2013) Genome sequence analyses show that *Neisseria oralis* is the same species as '*Neisseria mucosa* var. heidelbergensis'. Int. J. Syst. Evol. Microbiol. 63, 3920–3926.
[69] Lukjancenko, O., Ussery, D.W., Wassenaar, T.M. (2012) Comparative genomics of *Bifidobacterium*, *Lactobacillus* and related probiotic genera. Microb. Ecol. 63, 651–673.
[70] Meier-Kolthoff, J.P., Klenk, H.P., Goker, M. (2014) Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. Int. J. Syst. Evol. Microbiol. 64, 352–356.
[71] Vandamme, P., Dewhirst, F.E., Paster B.J, On, S.L.W., Genus, I. (2005) Campylobacter. In: Brenner, D.J., Krieg, N.R., Staley, J.T. (Eds.), Bergey's Manual of Systematic Bacteriology, Springer, pp. 1147–1160.
[72] Aagaard, M.E.Y., Kirk, K.F., Nielsen, H., Nielsen, H.L. (2021) High genetic diversity in *Campylobacter concisus* isolates from patients with microscopic colitis. Gut Pathog. 13, 3.
[73] Liu, F., Chen, S., Luu, L.D.W., Lee, S.A., Tay, A.C.Y., Wu, R., Riordan, S.M., Lan, R., Liu, L., Zhang, L. (2020) Analysis of complete *Campylobacter concisus* genomes identifies genomospecies features, secretion systems and novel plasmids and their association with severe ulcerative colitis. Microb. Genomics 6.
[74] Larsen, M.V., Cosentino, S., Lukjancenko, O., Saputra, D., Rasmussen, S., Hasman, H., Sicheritz-Ponten, T., Aarestrup, F.M., Ussery, D.W., Lund, O. (2014) Benchmarking of methods for genomic taxonomy. J. Clin. Microbiol. 52, 1529–1539.
[75] Gorkiewicz, G., Feierl, G., Schober, C., Dieber, F., Kofer, J., Zechner, R., Zechner, E.L. (2003) Species-specific identification of campylobacters by partial 16S rRNA gene sequencing. J. Clin. Microbiol. 41, 2537–2546.
[76] Hansson, I., Persson, M., Svensson, L., Engvall, E.O., Johansson, K.E. (2008) Identification of nine sequence types of the 16S rRNA genes of *Campylobacter jejuni* subsp. *jejuni* isolated from broilers. Acta Vet. Scand. 50, 10.
[77] Korczak, B.M., Stieber, R., Emler, S., Burnens, A.P., Frey, J., Kuhnert, P. (2006) Genetic relatedness within the genus *Campylobacter* inferred from *rpoB* sequences. Int. J. Syst. Evol. Microbiol. 56, 937–945.
[78] Vandamme, P., Peeters, C. (2014) Time to revisit polyphasic taxonomy. Antonie van Leeuwenhoek 106, 57–65.
[79] Gemmell, M.R., Berry, S., Mukhopadhya, I., Hansen, R., Nielsen, H.L., Bajaj-Elliott, M., Nielsen, H., Hold, G.L. (2018) Comparative genomics of *Campylobacter concisus*: analysis of clinical strains reveals genome diversity and pathogenic potential. Emerg Microbes Infect. 7, 116.
[80] Wang, Y., Liu, F., Zhang, X., Chung, H.K.L., Riordan, S.M., Grimm, M.C., Zhang, S., Ma, R., Lee, S.A., Zhang, L. (2017) *Campylobacter concisus* genomospecies 2 is better adapted to the human gastrointestinal tract as compared with *Campylobacter concisus* genomospecies 1. Front. Physiol. 8, 543.
[81] Yde Aagaard, M.E., Frahm Kirk, K., Linde Nielsen, H., Harder Tarpgaard, I., Bach Hansen, J., Nielsen, H. (2020) *Campylobacter concisus* is prevalent in the gastrointestinal tract of patients with microscopic colitis. Scand. J. Gastroenterol. 55, 924–930.
[82] Gevers, D., Dawyndt, P., Vandamme, P., Willems, A., Vancanneyt, M., Swings, J., De Vos, P. (2006) Stepping stones towards a new prokaryotic taxonomy. Philos. Trans. R. Soc. Lond. Ser. B: Biol. Sci. 361, 1911–1916.
[83] On, S.L.W., Zhang, J., Cornelius, A.J., Anderson, T.P. (2021) Markers for discriminating *Campylobacter concisus* genomospecies using MALDI-TOF analysis. Curr. Res. Microb. Sci. 2, 100019.
[84] Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.