International Journal of Food Microbiology 425 (2024) 110855

Contents lists available at ScienceDirect

International Journal of Food Microbiology

journal homepage: www.elsevier.com/locate/ijfoodmicro

Host associations of *Campylobacter jejuni* and *Campylobacter coli* isolates carrying the L-fucose or D-glucose utilization cluster

Pjotr S. Middendorf [a,b], Aldert L. Zomer [c,d], Indra L. Bergval [b], Wilma F. Jacobs-Reitsma [b], Heidy M.W. den Besten [a,*], Tjakko Abee [a,*]

[a] *Food Microbiology, Wageningen University and Research, Wageningen, Netherlands*
[b] *National Institute for Public Health and the Environment, Bilthoven, Netherlands*
[c] *Faculty of Veterinary Medicine, Department of Infectious Diseases and Immunology, Utrecht University, Utrecht, Netherlands*
[d] *WHO Collaborating Center for Campylobacter/OIE Reference Laboratory for Campylobacteriosis, Utrecht, Netherlands*

ARTICLE INFO

Keywords:
L-fucose utilization
D-glucose utilization
Phylogeny
Campylobacter jejuni
Campylobacter coli
Genomics

ABSTRACT

Campylobacter was considered asaccharolytic, but is now known to carry saccharide metabolization pathways for L-fucose and D-glucose. We hypothesized that these clusters are beneficial for *Campylobacter* niche adaptation and may help establish human infection.

We investigated the distribution of D-glucose and L-fucose clusters among ~9600 *C. jejuni* and *C. coli* genomes of different isolation sources in the Netherlands, the United Kingdom, the United States of America and Finland. The L-fucose utilization cluster was integrated at the same location in all *C. jejuni* and *C. coli* genomes, and was flanked by the genes *rpoB*, *rpoC*, *rspL*, *repsG* and *fusA,* which are associated with functions in transcription as well as translation and in acquired drug resistance. In contrast, the flanking regions of the D-glucose utilization cluster were variable among the isolates, and integration sites were located within one of the three different 16S—23S ribosomal RNA areas of the *C. jejuni* and *C. coli* genomes. In addition, we investigated whether acquisition of the L-fucose utilization cluster could be due to horizontal gene transfer between the two species and found three isolates for which this was the case: one *C. jejuni* isolate carrying a *C. coli* L-fucose cluster, and two *C. coli* isolates which carried a *C. jejuni* L-fucose cluster. Furthermore, L-fucose utilization cluster alignments revealed multiple frameshift mutations, most of which were commonly found in the non-essential genes for L-fucose metabolism, namely, *Cj0484* and *Cj0489*. These findings support our hypothesis that the L-fucose cluster was integrated multiple times across the *C. coli/C. jejuni* phylogeny.

Notably, association analysis using the *C. jejuni* isolates from the Netherlands showed a significant correlation between human *C. jejuni* isolates and *C. jejuni* isolates carrying the L-fucose utilization cluster. This correlation was even stronger when the Dutch isolates were combined with the isolates from the UK, the USA and Finland. No such correlations were observed for *C. coli* or for the D-glucose cluster for both species. This research provides insight into the spread and host associations of the L-fucose and D-glucose utilization clusters in *C. jejuni* and *C. coli*, and the potential benefits in human infection and/or proliferation in humans, conceivably after transmission from any reservoir.

1. Introduction

Campylobacter is a zoonotic bacterium and is the main cause of bacterial foodborne gastroenteritis worldwide. Most human campylobacteriosis cases are the result of infection with *Campylobacter jejuni* (83.1 %), primarily linked to poultry, or *Campylobacter coli* (10.8 %), primarily linked to pig/swine (Kaakoush et al., 2015; Tack et al., 2019; EFSA, 2021). Therefore, most sampling is done for *C. jejuni,* as it is a much larger problem than *C. coli*, likely due to slaughter processes and individual food preparation practices such as washing chicken (Thames and Theradiyil Sukumaran, 2020). Campylobacteriosis generally involves symptoms like watery or bloody diarrhea accompanied by abdominal pain, nausea and fever, and in rare occasions infection with *Campylobacter* can lead to the development of serious illnesses like the Guillain-Barré syndrome or irritable bowel syndrome (Rees et al., 1995; Allos, 1997; Blaser, 1997). All animals and environmental waters are a

* Corresponding authors.
E-mail addresses: heidy.denbesten@wur.nl (H.M.W. den Besten), tjakko.abee@wur.nl (T. Abee).

https://doi.org/10.1016/j.ijfoodmicro.2024.110855
Received 16 March 2024; Received in revised form 6 July 2024; Accepted 3 August 2024
Available online 5 August 2024

potential source for *Campylobacter* (Hepworth et al., 2011; Wagenaar et al., 2015; Nilsson et al., 2018; Mughini-Gras et al., 2021). However, in most countries, poultry and cattle have been identified as the main sources for *Campylobacter*, which is also reflected in multiple case-control studies that identify the consumption of raw and/or undercooked meat as risk factors for human campylobacteriosis (Doorduyn et al., 2010; EFSA, 2010; Mughini-Gras et al., 2012; Mughini-Gras et al., 2021).

Interestingly, *Campylobacter* isolates are considered fragile organisms, as they grow microaerobically at temperatures between 30 and 45 °C, yet they are able to withstand many stresses during transmission from host to host (Levin, 2007; Silva et al., 2011). *Campylobacter* carries many genes of which their products are involved in the protection and survival of *Campylobacter* in stressful environments, such as starvation and oxidative, osmotic, heat shock, pH and nitrosative stresses (Andersen et al., 2005; Brøndsted et al., 2005; Candon et al., 2007; Bronowski et al., 2014). Due to its natural competence, one way of *Campylobacter* to adapt to new environments is through horizontal gene transfer (HGT) (Golz and Stingl, 2021). DNA-rich environments, such as the animal and human gastro-intestinal tracts, are ideal for the uptake of new DNA, and can result in strain variety of *Campylobacter* isolates (Sheppard and Maiden, 2015; Golz and Stingl, 2021). However, in a study that estimated the molecular clock rate of *Campylobacter*, the authors highlighted that multiple lineages are maintained, implying that large-scale clonal sweeps (such as uptake of resistance genes) may take hundreds of years or more in these species (Calland et al., 2021).

For a long time *Campylobacter* was thought to be asaccharolytic, lacking most key enzymes to metabolize sugars. However, the recent discovery of the L-fucose and D-glucose utilization clusters indicates that most *Campylobacter* isolates are able to metabolize L-fucose, D-arabinose and/or D-glucose (Muraoka and Zhang, 2011; Stahl et al., 2011; Vorwerk et al., 2015; Vegge et al., 2016; Garber et al., 2020). The human gut is an L-fucose rich environment, and intestinal epithelial cells produce fucosylated mucins, making L-fucose metabolism beneficial for *Campylobacter* residing in the human gut. Furthermore, *Campylobacter* lacks fucosidases and it has been shown that in the presence of fucosidase-producing bacteria, like *B. fragilis*, *Campylobacter* uses cross-feeding by exchanging nutrients with *B. fragilis*, resulting in higher invasiveness of epithelial cells (Garber et al., 2020; Luijkx et al., 2020). Not only the human gut, but also the intestine of pigs and chicken are heavily fucosylated. Interestingly, unlike during pig colonization, no competitive advantage was observed for isolates carrying the L-fucose utilization cluster during the colonization of poultry, possibly due to a decreased fucosidase activity in poultry, as chicken fucosylated O-glycan mucin structures are more sulfated and therefore resistant to enzymatic processing (Stahl et al., 2011).

The L-fucose utilization cluster is a genomic element that comprises nine to ten genes, depending on a frameshift in *Cj0489* that is observed in some isolates (*Cj0480c – Cj0489* or *Cj0480c – Cj0489-S + Cj0489-L*), and has been identified in approximately 60 % of the investigated *C. jejuni* and *C. coli* isolates, from here on referred to as fuc+ isolates. Several studies have shown that L-fucose metabolism results in pyruvate and lactate, supporting growth, survival and/or virulence of *Campylobacter* isolates (Stahl et al., 2011; Dwivedi et al., 2016; Garber et al., 2020; Luijkx et al., 2020; Middendorf et al., 2022).

Another, but less commonly encountered sugar metabolic cluster is the *Campylobacter* D-glucose utilization cluster, which comprises of seven genes (*glcP*, *pgi2*, *glk*, *pgl*, *zwf*, *edd* and *eda*) (Vorwerk et al., 2015; Vegge et al., 2016), from here on referred to as gluc+ isolates. It supports growth and enhances survival and biofilm formation in *Campylobacter* (Vorwerk et al., 2015; Vegge et al., 2016). Similar to L-fucose metabolism, D-glucose is metabolized to the end product pyruvate, which can be further metabolized (Vegge et al., 2016). Both the L-fucose utilization cluster and the D-glucose utilization cluster were partly discovered due to the availability of online deposited genomic sequences.

Whole-genome sequencing (WGS) is rapidly becoming the standard for genotyping a wide variety of pathogens. It is a powerful tool to investigate genomic associations with the epidemiology of the micro-organism, as it can additionally be used to identify potential important survival and virulence factors (Franz et al., 2016; Besser et al., 2018). Several public databases contain deposited genomic sequences and the PubMLST database is the largest *Campylobacter* genome database, currently harboring over 60,000 *C. jejuni* and *C. coli* genomes (Jolley et al., 2018).

In this study, we analyzed *C. jejuni* and *C. coli* isolate genomes from four different countries, namely, the Netherlands, the United Kingdom (UK), the United States of America (USA) and Finland. The Dutch dataset provides a balanced snapshot of the *C. coli* and *C. jejuni* population in the Netherlands (Mughini-Gras et al., 2021), while the UK, the USA and Finland datasets reflect the highest deposited genomic sequences counts in PubMLST. We studied the organization and origin of the L-fucose and D-glucose utilization clusters, including their flanking regions and position in the genomes, and found several cases of horizontal gene transfer of the L-fucose utilization cluster. By extrapolating the phylogeny, we studied the presence of L-fucose and D-glucose utilization clusters and found a correlation of isolates carrying the L-fucose utilization cluster and human origin. The obtained results give further insight in the diversity of these clusters and their putative contribution to human infection and/or proliferation of fuc+ *C. jejuni* in humans, conceivably after transmission from any reservoir.

2. Methods

2.1. Isolate collection

Isolate collections were used from four different countries, namely, the Netherlands, the UK, the USA and Finland. Isolates from the UK, the USA and Finland were obtained from the PubMLST database (https://pubmlst.org/) accessed in June 2021.

All *C. jejuni* and *C. coli* WGS data from the PubMLST database were downloaded and accompanied by metadata, namely, isolate ID, isolate name, country of isolation, year of isolation, clonal complex, and species.

2.1.1. The Netherlands

The WGS of the Dutch dataset were provided by the National Institute for Public Health and the Environment (RIVM) and are described in (Mughini-Gras et al., 2021) and show a balanced snapshot of the *C. coli* and *C. jejuni* isolates from difference sources. In the current study, 1057 *C. jejuni* and 349 *C. coli* whole-genome sequenced isolates from the Netherlands were used that were collected in 2014–2019. This set included isolates from human cases ($n = 280$), water ($n = 251$), sheep/goat ($n = 110$), chicken ($n = 256$), turkey ($n = 37$), cattle ($n = 207$), swine (n = 110), wild birds ($n = 61$) and pets ($n = 94$). The isolates from human cases were collected from 13 different medical microbiology laboratories in the Netherlands. Isolates from livestock animals were collected by Wageningen Bioveterinary Research (WBVR) and Wageningen Food Safety Research (WFSR), in collaboration with the RIVM and the Netherlands Food and Consumer Product Safety Authority (NVWA), within the framework of established surveillance programs for zoonotic agents in food-producing animals in the Netherlands during 2014–2019. Isolates of pets were collected from veterinary clinics all across the Netherlands by The Veterinary Microbiological Diagnostic Centre (VMDC) of Utrecht University. Wild bird isolates were collected in June and December 2018 by Wageningen Ecological Research (WER). The water isolates were collected from six different geographic areas of comparable size in the Netherlands.

2.1.2. United Kingdom (UK)

In total, 21,690 *C. jejuni* isolates were deposited in the PubMLST database, however, only a randomly selected subset of 3150 *C. jejuni*

isolates was used in the current study, as the total number was too high for the creation of phylogenetic trees. Random selection was performed by appointing each genome to a number and using these 21,690 numbers as input for the sample() function of base R. These isolates included isolates from human cases ($n = 2407$), water ($n = 1$), sheep/goat ($n = 47$), chicken ($n = 531$), turkey ($n = 18$), cattle ($n = 73$), wild bird (n = 4), goose/duck ($n = 12$) and unknown source ($n = 57$). Furthermore, all 2786 *C. coli* isolates from the PubMLST database were used for the current study. These isolates included isolates from human cases ($n = 1401$), water ($n = 38$), sheep/goat ($n = 148$), chicken ($n = 564$), turkey ($n = 11$), cattle ($n = 67$), swine ($n = 182$), wild bird ($n = 21$), goose/duck ($n = 39$), soil ($n = 35$) and unknown source ($n = 280$). The isolates from the UK were collected between 1980 and 2018.

2.1.3. United States of America (USA)

In total, 16,772 *C. jejuni* isolates were deposited in the PubMLST database, however, due to a large number of isolates being of unknown source, a sub selection was made of 1251 isolates. These isolates included isolates from human cases ($n = 619$), water ($n = 41$), sheep/goat ($n = 22$), chicken ($n = 207$), turkey (n = 6), cattle ($n = 160$), wild bird ($n = 167$) and goose/duck ($n = 29$). The *C. jejuni* isolates from the USA were collected between 1979 and 2020, however, for the majority of the isolates no isolation year was reported. For *C. coli,* all 7400 isolates were selected. However, similarly as for the *C. jejuni* isolates, a large number of isolates were isolated from unknown sources. Therefore, >95 % the unknown sources isolates were removed (for a collection over 300 isolates), leaving 335 *C. coli* isolates. These isolates included isolates from human cases ($n = 26$), chicken ($n = 108$), turkey ($n = 24$), cattle ($n = 31$), swine ($n = 90$), goose/duck (n = 11) and unknown source ($n = 45$). The *C. coli* isolates from the USA were collected between 1979 and 2019, however, for the majority of the isolates no isolation year was reported.

2.1.4. Finland

All 634 *C. jejuni* isolates deposited in the PubMLST database were used. These isolates included isolates from human cases ($n = 104$), water (n = 4), chicken ($n = 124$), cattle (n = 6), wild bird ($n = 359$), goose/duck ($n = 21$) and unknown source (n = 16). Isolates from Finland were collected between 1998 and 2018 and concerned only *C. jejuni* isolates.

2.2. Integration sites of the L-fucose and D-glucose utilization cluster

Integration sites of the L-fucose utilization cluster were studied in all fuc+ isolates of the Dutch dataset. The L-fucose utilization cluster and 10 Kb flanking regions were selected and annotated using Prokka v1.13 (Seemann, 2014). Annotated flanking regions were visualized using the Benchling software (www.benchling.com).

Since Illumina paired end sequencing did not cover the flanking regions of the D-glucose utilization cluster in the Dutch dataset, four randomly selected isolates were used for long-read sequencing, namely, *C. jejuni* 103292-005-103, *C. coli* 18-556, *C. coli* 8230 and *C. jejuni* 18-440 (here named *C. jejuni* 4). Up to 50 Kb flanking regions were selected and annotated using Prokka v1.13, which were visualized using the Benchling software.

Next, all gluc+ isolates deposited in the PubMLST database (UK, USA and Finland) were screened for the presence of these flanking regions and the integration sites of the D-glucose utilization cluster.

For long-read sequencing, DNA was isolated using the Qiagen UltraClean Microbial DNA isolation kit (Qiagen, Venlo, the Netherlands) and sequenced using Oxford Nanopore technology to fully resolve the genome. This was performed as outlined in the genomic DNA ligation protocol (SQK-LSK109), with sequencing on a MinION device using flow cell type R9.4.1 (FLO-MIN106D) (Oxford Nanopore, Oxford, United Kingdom) using "super accurate" basecalling. Reads were filtered with options minimal length of 5000 and keep percentage of 90 % using Filtlong v0.2.1 (https://github.com/rrwick/Filtlong). Reads were assembled using Flye v2.720 using options -nano hq -min overlap 1000 -meta, into a single scaffold and used the option "existing long read assembly" in Unicycler v. 0.4.721. Gene alignments and visualization were done using the Clinker alignment tool (Gilchrist and Chooi, 2021).

2.3. HGT and frameshift analyses of the L-fucose utilization cluster

In silico HGT analyses were performed on all L-fucose utilization cluster of the *C. jejuni* and *C. coli* isolates of the Dutch dataset. The L-fucose utilization cluster were selected in these genomes and aligned using Mafft (Katoh et al., 2002). Phylogenetic trees were built using Fasttree (Price et al., 2009). Branch lengths of the trees were square root transformed in R using ape 5.4.1 (Paradis and Schliep, 2019) to improve visualization of strain differences in phylogenetic trees. Frameshifts were visualized by viewing all L-fucose utilization clusters using the Benchling software. For the outgroup of the tree 8 fuc+ *Campylobacter* isolates from the NCBI database were selected, namely, *C. jejuni doylei* FDAARGOS 295, NCTC11924 and NCTC11951, *Campylobacter insulaenigrae* NCTC12927 and NCTC12928, *Campylobacter upsaliensis* NCTC11540 and NCTC11541, *Campylobacter canadensis* LMG24001.

2.4. Phylogeny & statistical procedures

Genome phylogenetic trees were constructed using Mashtree (Katz et al., 2019). Clades in the phylogenetic tree were partitioned into clusters using Treestruct (Volz et al., 2020), and MLST (Multi-Locus Sequence Typing) types were obtained from PubMLST. The percentage of fuc+ isolates and percentage of human derived isolates was determined per cluster and the correlation between the values per cluster were correlated using Spearman rank and Pearson correlations. By condensing the data to one datapoint per cluster, the effect of oversampled lineages was removed.

3. Results

3.1. Integration sites of the L-fucose and D-glucose utilization clusters in C. jejuni and C. coli

To get a better understanding of the distribution and transfer of the L-fucose and D-glucose utilization clusters between the *C. jejuni* and *C. coli* isolates, we investigated the integration location of these clusters in the genome. For the L-fucose utilization cluster, we analyzed up to 10 Kb flanking regions of this operon in both *C. jejuni* and *C. coli* isolates. In all flanking regions the same genes were observed, namely, *rpoB*, *rpoC* upstream and *rspL*, *repsG*, *fusA* downstream of the cluster (Fig. 1). Protein analysis with Protein blast, uniprot and STRINGdb showed that these genes encode proteins with functions in transcription and translation.

Next, we investigated the flanking regions of the D-glucose cluster (*glcP*, *pgi2*, *glk*, *pgl*, *zwf*, *edd* and *eda*). In the Dutch dataset, which consisted out of 37 gluc+ isolates, the D-glucose utilization cluster was always present on small contigs of the genome sequences, making it difficult to accurately acquire large proportions of the flanking regions. Therefore, four isolates of the Dutch dataset were selected for long-read sequencing, which allows a more in-depth analysis of the flanking regions. Of all gluc+ isolates present in the PubMLST database (from the UK, the USA and Finland), 27 isolates carried large enough contigs to investigate proportions of the flanking regions of the D-glucose utilization cluster. Combined with the Dutch isolates, in total 31 isolates were used for this analysis. Homology analysis of the D-glucose utilization clusters and the 10 Kb flanking regions using the alignment tool Clinker (Gilchrist and Chooi, 2021), selected *C. jejuni and C. coli* isolates were divided in six groups in total based on flanking region similarity (Fig. 2).

Groups 1 and 2 carried, next to the D-glucose utilization cluster (*glcP*, *pgi2*, *glk*, *pgl*, *zwf*, *edd* and *eda*), also a galactose cluster (*galK_1, galK_2, galT_1, galT_2* and *galT_3*). Further analyses of the galactose cluster

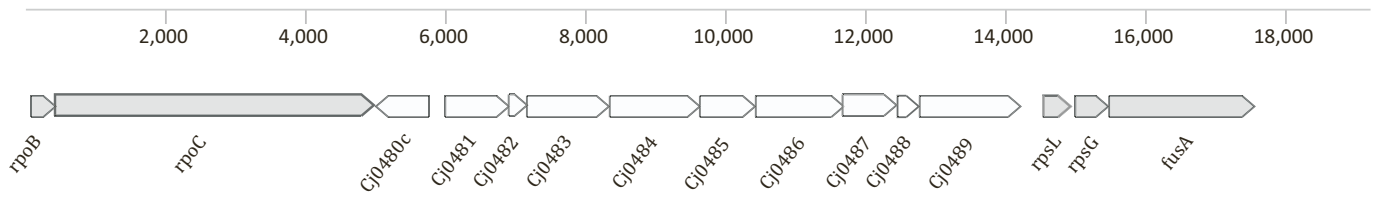

Fig. 1. Flanking regions of the L-fucose utilization cluster (*Cj0480c-Cj0489*) in *C. jejuni* and *C. coli*.

showed that it was only present in three *C. jejuni* isolates and one *C. coli* isolate. For group 1, the integration sites could be identified, and here the cluster was flanked by *Cj0756* (hypothetical gene) and *Cj0734c* (histidine-binding protein). Further analyses highlights that this region is in the internal transcribed spacer (ITS) domain which is between the 23S and 16S ribosomal RNA region (Suppl. Fig. 1A). Group 3 was integrated in the same region as group 1 and was flanked by *Cj0755* (putative iron uptake) or *Cj0753c* (TonB transport gene) and by *Cj0734c*. From group 4, only genes upstream of the cluster could be identified with certainty, as genes downstream were encoding hypothetical proteins. The upstream part of the cluster was formed by *Cj0029* (cytoplasmic L-asparaginase), *Cj0028* (putative single-stranded-DNA-specific exonuclease) and *Cj0027* (CTP synthase), which is even more upstream in the 23S and 16S ribosomal RNA region than groups 1 and 3 (Suppl. Fig. 1B). In the 10 Kb flanking regions of group 5, only hypothetical genes were observed, therefore, 50 Kb flanking regions were analyzed. From these isolates, three different flanking regions were observed around the genomic locations: *Cj0734c*, *Cj0031* (IIS restriction/modification enzyme) and *Cj0431* (putative periplasmic ATP/GTP-binding) (Suppl. Fig. 2). Lastly, group 6, was flanked by *Cj0432c* (UDP-*N*-acetylmuramoylalanine–D-glutamate ligase) and *Cj0431*, which is a 23S and 16S ribosomal RNA region that is upstream of what was found in groups 1 and 3, and downstream of what was found in group 4 (Suppl. Fig. 1C).

Taken together, there are three genomic regions where the glucose cluster was integrated, which was either in the neighborhood of *Cj0734c,* of *Cj0029* or of *Cj0431*. Notably, the cluster that was found in these three genomic regions were all inserted within one of the three copies of 23S and 16S ribosomal RNA in *C. jejuni* and *C. coli* genomes (Suppl. Figs. 1 and 3).

3.2. Phylogeny of C. jejuni and C. coli isolates and horizontal gene transfer between the species

As similar integration sites of either the L-fucose or the D-glucose utilization clusters were observed in both tested species, we investigated the phylogeny of *C. jejuni* and *C. coli* isolates and analyzed whether HGT of the utilization clusters occurred between both species. For these analyses, the Dutch collection was used, as it provided a balanced snapshot of the *C. coli* and *C. jejuni* population in the Netherlands which is well described in literature (Mughini-Gras et al., 2021), unlike the collections that were obtained from PubMLST. In the Dutch collection, we first investigated the presence of the L-fucose and D-glucose utilization clusters using a phylogenetic tree based on the whole genome sequences of both species. The phylogenetic tree showed a clear distinction between *C. jejuni* and *C. coli* isolates (Fig. 3). For *C. coli*, four different clades were observed, while this was less clear for *C. jejuni* isolates. Clonal clusters of fuc+ isolates were observed in especially *C. jejuni*, but large sections within these clonal clusters were also lacking the L-fucose cluster. The D-glucose utilization cluster was sporadically present in the *C. jejuni* isolates (0.2 % of all tested isolates), but more prevalent in *C. coli* isolates (10 % of all tested isolates). Furthermore, despite conceivable introduction at multiple stages in the phylogeny, subsequent clonal spread of the D-glucose utilization cluster, as observed for the L-fucose utilization cluster, was not observed.

Due to the large number of fuc+ isolates, we were able to investigate HGT of the L-fucose utilization cluster between these *C. jejuni* and *C. coli* isolates, using a phylogenetic tree based on the L-fucose utilization cluster (~9 Kb per isolate) of all Dutch *C. jejuni* and *C. coli* isolates. Other *Campylobacter* (non-*jejuni* and non-*coli*) species were included as outgroup for the rooting of the tree (Fig. 4). Clear indications were found of two HGT events within the dataset, namely, one *C. jejuni* isolate carried an L-fucose utilization cluster that was highly identical to the L-fucose utilization clusters typically observed in *C. coli,* and two *C. coli* isolates carried an utilization cluster that was highly identical to the L-fucose utilization clusters typically found in *C. jejuni* (marked in red).

3.3. Distribution of Campylobacter spp. based on isolation source and L-fucose/D-glucose utilization clusters

We further analyzed the presence of the L-fucose and D-glucose utilization clusters using larger databases which we grouped on the basis of country of isolation: the Netherlands, the United Kingdom (UK), the United States of America (USA) and Finland. We included in the analysis 1057 *C. jejuni* and 349 *C. coli* isolates from the Netherlands, 3150 *C. jejuni* and 2786 *C. coli* isolates from the UK, 1251 *C. jejuni* and 335 *C. coli* isolates from the USA and only 634 *C. jejuni* isolates from Finland because *C. coli* isolates from this country were not present in the PubMLST database (Tables 1 and 2). Together, these isolates consisted of 4837 human isolates, 335 water isolates, 327 sheep/goat isolates, 1790 chicken isolates, 96 turkey isolates, 544 cattle isolates, 382 swine isolates, 612 wild bird isolates, 94 pet isolates, 112 goose/duck isolates, 35 soil isolates and 398 isolates of unknown isolation source.

The L-fucose cluster prevalence in *C. jejuni* was for the Netherlands 51 %, for the UK 65 %, for the USA 44 % and for Finland 15 %. For *C. coli*, the L-fucose prevalence was 43 %, 77 % and 91 % for the Netherlands, the UK and the USA, respectively. To investigate the cluster prevalence over time, we split the *C. jejuni* and *C. coli* collection of the UK per year, as this was the largest dataset that included metadata such as year of isolation (Suppl. Fig. 4). For *C. jejuni,* no notable changes in the percentage of fuc+ isolates were observed throughout the years 2001–2018. For *C. coli* an increasing trend in the prevalence of the L-fucose cluster was observed in later years (2011–2018) and was not linked to clonal spread.

The D-glucose cluster prevalence for *C. jejuni* per country was for the Netherlands 0.2 %, for the UK 0.6 %, for the USA 0.7 % and for Finland 9.9 %. For *C. coli,* the D-glucose cluster prevalence was 10 %, 4.3 % and 6.6 % for the Netherlands, the UK and the USA, respectively.

Interestingly, isolates carrying the D-glucose utilization cluster often carried the L-fucose utilization cluster. For *C. jejuni* this was observed in 2/2 isolates (100 %), 16/19 isolates (84 %), 2/9 isolates (22 %) and 8/63 isolates (13 %), for the Netherlands, the UK, the USA and Finland, respectively, and for *C. coli*, 29/35 isolates (83 %), and 120/121 isolates (99 %), and 22/22 (100 %) isolates, for the Netherlands, the UK and the USA, respectively.

Phylogenetic trees were created per country to further investigate the distribution of the L-fucose and D-glucose utilization clusters and possible host associations. For the generation of the country-specific phylogenetic *C. jejuni* trees, 1057 (the Netherlands), 3150 (UK), 1289 (USA) and 634 (Finland) *C. jejuni* isolates were used (Fig. 5, Suppl. Fig. 5–7). Isolates within the trees were clustered into groups using the classification methods Treestruct and MLST. MLST subclassification is

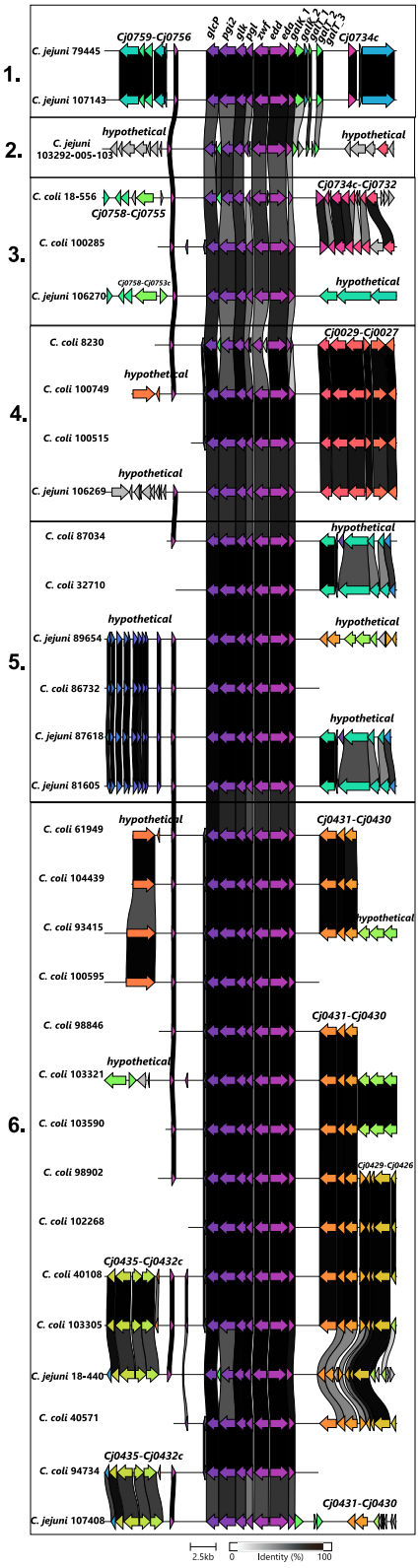

Fig. 2. Flanking regions of the D-glucose utilization cluster in 31 selected *C. jejuni* and *C. coli* isolates. Based on analysis of 10 Kb flanking regions, Isolates were divided into 6 groups based on homology. Colors signify different gene groups based on genetic difference. For *C. jejuni/coli* isolates number 1 to 4, we performed long-read sequencing and for the other isolates, numbers indicate the PubMLST isolate ID, as described in materials and methods.

based on the sequences (alleles) of seven household genes, whereas Treestruct is based on the branch lengths in the phylogenetic trees using the whole genome. Note that there are no universal Treestruct clusters between different phylogenetic trees, as the branch lengths in each tree are different (Volz et al., 2020), therefore, Treestruct groups are tree-specific. Only the Dutch dataset was the result of a surveillance study in a defined period of two years and therefore little to no sampling bias expected, we mainly focused on the Dutch dataset, as in that dataset there was no unknown sampling bias.

In the Dutch dataset, large clusters of fuc+ isolates (> 40 isolates per cluster) were observed, with apparent clonal expansion, while only two isolates carried the D-glucose utilization cluster. Similarly for the PubMLST datasets of the UK (> 100 isolates) and the USA (> 50 isolates per cluster), also large fuc+ isolate clusters were observed, while the D-glucose utilization cluster was much more rare. Interestingly, in the Finland dataset, large clusters of either fuc+ or gluc+ isolates (>12 isolates per cluster) were present, indicating clonal expansion. Notably, the Finland dataset consisted of 56 % wild bird isolates, while the other countries only had between 0.1 % and 13 % wild bird isolates. In all tested countries, gluc+ isolates were mostly wild bird isolates, highlighting a clear link between gluc+ isolates and wild bird isolates. However, the implications of this link requires further studies.

3.4. Correlation between fuc+ isolates and human isolates in C. jejuni

Studies have shown that fuc+ isolates show increased survival and growth in the presence of L-fucose, however, no link to source specific isolates was made (Garber et al., 2020; Middendorf et al., 2022; Stahl et al., 2011; Middendorf et al., 2024). In the current study when using the Dutch dataset, large quantities of human fuc+ isolates (67 %) were observed. Therefore, we hypothesized that fuc+ isolates are more likely to survive and proliferate in the human host after transmission from an animal or environmental reservoir. This was investigated by dividing the phylogenetic trees in groups using the classification methods Treestruct and MLST, and calculating the correlation between the percentage of fuc+ isolates and the percentage of human isolates over all clusters to alleviate oversampling biases for specific clones. For the Dutch dataset, isolates were partitioned into 22 different Treestruct groups and in 16 MLST groups. In this dataset a significant correlation was observed between human isolates and fuc+ isolates, with the Treestruct (Spearman *p*-value = 0.0072 and Pearson p-value = 0.0177) and MLST (Spearman p-value = 0.0400 and Pearson p-value = 0.0384) classification methods (Suppl. Table 1), suggesting that fuc+ isolates have advantages to survive and proliferate in the human hosts.

Although the PubMLST datasets (UK, USA and Finland) potentially had unknown sampling biases, we screened whether a correlation was also observed in these countries. For the UK 40 Treestruct groups (>35 isolates per group) and 15 MLST groups (>35 isolates per group), for the USA 18 Treestruct groups (>25 isolates per group) and 12 MLST groups (>25 isolates per group) and for Finland 18 Treestruct groups and 7 MLST groups (>25 isolates per group). However, no significant correlations were observed.

Notably, after combining the *C. jejuni* datasets of the four countries, which resulted in 6092 *C. jejuni* isolates, a significant correlation with both the Treestruct (Spearman p-value = 0.0017 and Pearson p-value <0.0001) and MLST classification (Spearman p-value = 0.034 and Pearson p-value = ns) methods was observed between human isolates and the presence of the L-fucose utilization cluster.

3.5. L-fucose and D-glucose cluster distribution in C. coli

Similar analyses were performed with the phylogenetic trees of *C. coli*, for which 349, 2786 and 335 *C. coli* isolates were used for the Netherlands, the UK and the USA, respectively (Fig. 6 and Suppl. Fig. 8–9).

In the Dutch dataset, large clusters of fuc+ isolates (> 15 isolates per

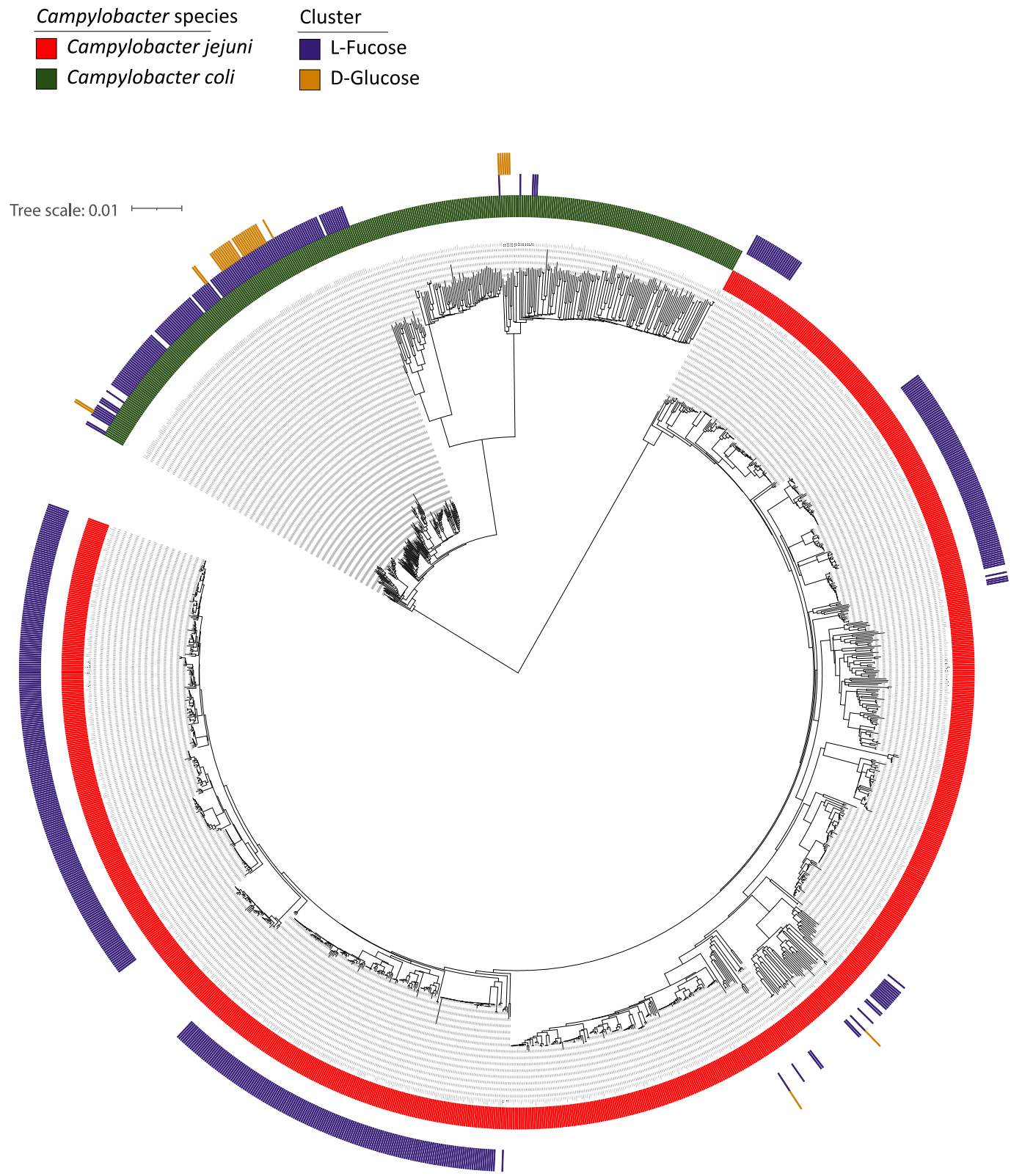

Fig. 3. Phylogenetic tree using 1057 *C. jejuni* (red) and 349 *C. coli* (green) isolates from the Netherlands, based on the whole genome. Isolates were collected between 2014 and 2019 and consisted of multiple sources, including human, water, sheep/goat, chicken, turkey, cattle, swine, wild bird and pet isolates (Mughini-Gras et al., 2021).

cluster) were observed in livestock associated isolates, with apparent clonal expansion, while the D-glucose utilization cluster was only sporadically present. Interestingly, the Dutch dataset included a high number (n = 175, 50 %) of water isolates, which were present in multiple clades, and these water isolates rarely carried the L-fucose utilization cluster ($n = 3$). However, unlike what was observed in animal host environments, none of these integrations resulted in clonal fuc+ isolate clusters, suggesting that there was no benefit for carrying this

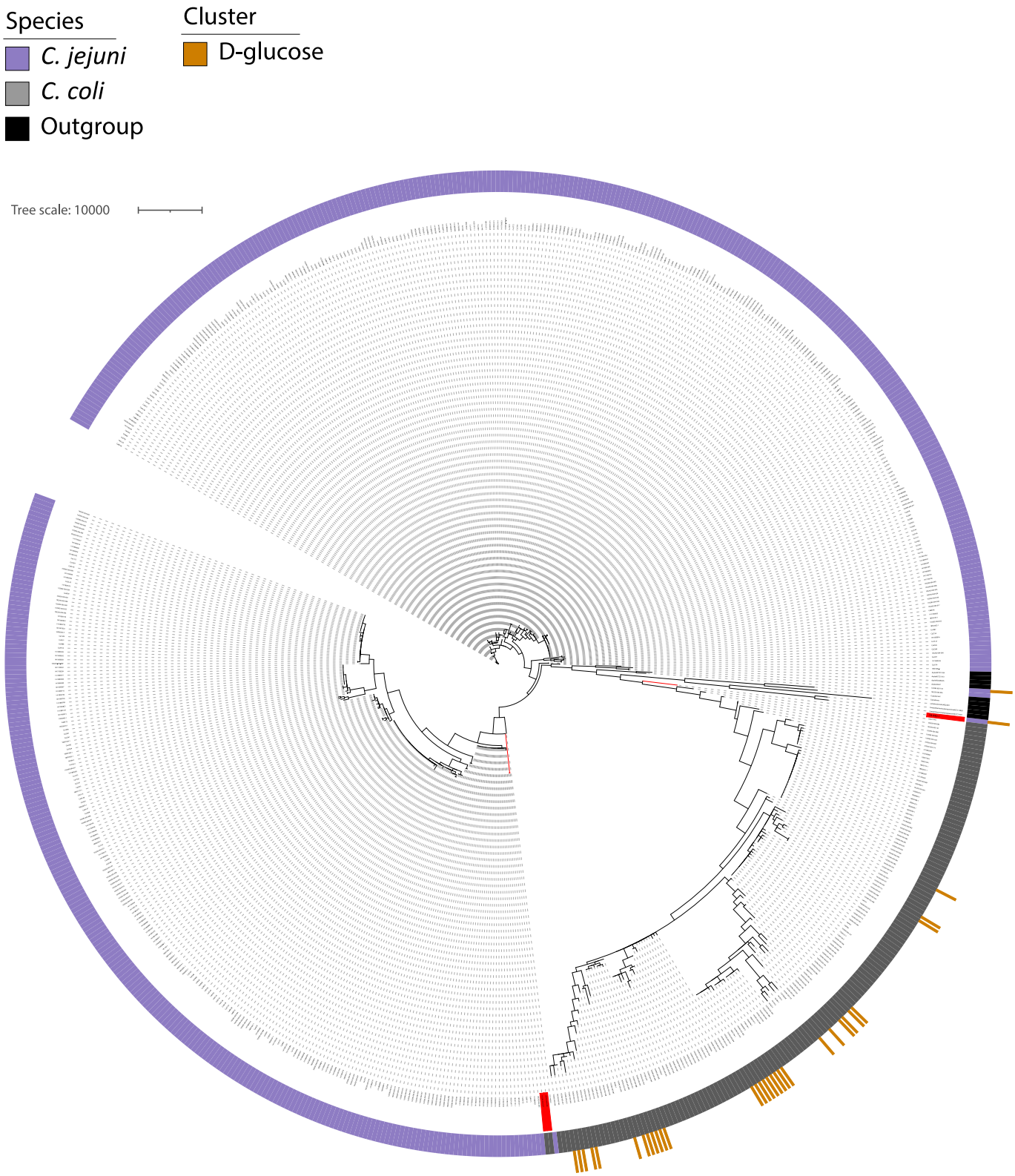

Fig. 4. Phylogenetic tree of *C. jejuni* and *C. coli* isolates from the Netherlands, based on 685 sequences of the L-fucose utilization cluster (all fuc+ isolates). Two species were included, namely *C. jejuni* (purple) and *C. coli* (grey). The outgroup (black, fuc+) consisted of 8 isolates: *C. jejuni doylei* FDAARGOS 295, NCTC11924 and NCTC11951, *C. insulaenigrae* NCTC12927 and NCTC12928, *C. upsaliensis* NCTC11540 and NCTC11541, and *C. canadensis* LMG24001. Isolates marked in red highlight events of HGT.

Table 1
C. jejuni isolate collection used in the current study. Total number of isolates and percentages of fuc+ isolates, gluc+ isolates and fuc+/gluc+ isolates per source are displayed.

Source	Netherlands	UK*	USA*	Finland
	Nr. *C. jejuni*	Nr. *C. jejuni*	Nr. *C. jejuni*	Nr. *C. jejuni*
	Total (%fuc+/% gluc+/%both)	Total (% fuc+/% gluc+/% both)	Total (% fuc+/% gluc+/% both)	Total (% fuc+/% gluc+/% both)
Human	272 (67 %/0 %/0 %)	2407 (68 %/0.2 %/0.1 %)	619 (52 %/0 %/0 %)	104 (19 %/0 %/0 %)
Water	76 (16 %/1.3 %/1.3 %)	1 (0 %/0 %/0 %)	41 (41 %/2.4 %/0 %)	4 (0 %/0 %/0 %)
Sheep/ Goat	85 (49 %/0 %0 %)	47 (40 %/2 %/0 %)	22 (100 %/0 %/0 %)	–
Chicken	241 (56 %/0 %/0 %)	531 (56 %/0 %/0 %)	207 (40 %/0 %/0 %)	124 (1.6 %/0 %/0 %)
Turkey	37 (49 %/0 %/0 %)	18 (78 %/0 %/0 %)	6 (17 %/0 %/0 %)	
Cattle	196 (53 %/0 %/0 %)	73 (44 %/0 %/0 %)	160 (47 %/0 %/0 %)	6 (67 %/0 %/0 %)
Swine	10 (40 %/0 %/0 %)	–		
Wild Bird	46 (17 %/6.7 %/6.7 %)	4 (100 %/0 %/0 %)	167 (22 %/4.8 %/1.2 %)	359 (18 %/18 %/2.2 %)
Pet	94 (32 %/0 %/0 %)	–		
Goose/ duck	–	12 (8 %/0 %/0 %)	29 (0 %/0 %/0 %)	21 (9.5 %/0 %/0 %)
Soil	–	–	–	
Unknown	–	57 (68 %/23 %/23 %)	–	16 (0 %/0 %/0 %)
Total	1057 (51 %/0.2 %/0.2 %)	3150 (65 %/0.6 %/0.5 %)	1251 (44 %/0.7 %/0.2 %)	634 (15 %/9.9 %/1.3 %)

Table 2
C. coli isolate collection used in the current study. Total number of isolates and precentages of fuc+ isolates, gluc+ isolates and isolates with both clusters per source are displayed.

Source	Netherlands	UK*	USA*
	Nr. *C. coli*	Nr. *C. coli*	Nr. *C. coli*
	Total (%fuc+/% gluc+/%both)	Total (%fuc+/% gluc+/%both)	Total (%fuc+/% gluc+/%both)
Human	8 (100 %/13 %/13 %)	1401 (81 %/1.9 %/1.9 %)	26 (77 %/0 %/0 %)
Water	175 (2.9 %/2.9 %/0 %)	38 (7.9 %/0 %/0 %)	–
Sheep/ Goat	25 (88 %/20 %20 %)	148 (99 %/7.4 %/7.4 %)	–
Chicken	15 (47 %/0 %/0 %)	564 (56 %/1.2 %/1.2 %)	108 (100 %/1.9 %/1.9 %)
Turkey	–	11 (55 %/0 %/0 %)	24 (92 %/0 %/0 %)
Cattle	11 (100 %/0 %/0 %)	67 (99 %/0 %/0 %)	31 (97 %/0 %/0 %)
Swine	100 (95 %/23 %/22 %)	182 (92 %/36 %/36 %)	90 (80 %/18/18 %)
Wild Bird	15 (13 %/2.2 %/2.2 %)	21 (19 %/0 %/0 %)	–
Pet	–	–	–
Goose/ duck	–	39 (18 %/0 %/0 %)	11 (100 %/0 %/0 %)
Soil	–	35 (89 %/17 %/17 %)	–
Unknown	–	280 (89 %/1.8 %/1.8 %)	45 (91 %/8.9 %/8.9 %)
Total	349 (43 %/10 %/8.3 %)	2786 (77 %/4.3 %/4.3 %)	335 (91 %/6.6 %/6.6 %)

cluster in a water environment or, alternatively, that it is challenging to isolate these isolates in a water environment. In the *C. coli* PubMLST dataset (the UK and the USA), large clusters of fuc+ isolates (> 100 isolates per cluster for the UK and > 15 isolates per cluster for the USA) were observed, with apparent clonal expansion. In this dataset, also several clusters of gluc+ isolates were observed, however, on a much smaller (< 14 isolates per cluster) scale than compared to the fuc+ clusters.

Next, similarly as for *C. jejuni*, we investigated whether the isolates carrying the L-fucose utilization clusters are more likely to proliferate in the human host after transmission from an animal or environmental reservoir. For the *C. coli* phylogenetic trees, isolates were only divided into different Treestruct complexes, as MLST complexes are very limited for *C. coli*. Based on branch lengths, each phylogenetic group was divided into 7 Treestruct groups, which were used for the statistical correlation analysis. However, no significant correlations were observed in any of the *C. coli* datasets (Suppl. Table 1).

4. Discussion

We investigated two carbohydrate utilization clusters that have previously been associated with growth, survival and biofilm formation of *Campylobacter*; the L-fucose and the D-glucose utilization clusters (Muraoka and Zhang, 2011; Stahl et al., 2011; Vorwerk et al., 2015; Dwivedi et al., 2016; Vegge et al., 2016). The sugars L-fucose and D-glucose are present in many environments, including the human and animal gastro-intestinal tracts. Previous studies have shown a role for fucose utilization in colonization of chicken and piglets, and in adhesion and invasion efficacy of Caco-2 cells (Middendorf et al., 2024). To date, several studies have been performed to investigate the transmission of *Campylobacter* between animals and humans (Kaakoush et al., 2015; Mughini-Gras et al., 2021), however, attributes that promote this transmission were rarely studied. In the current study we performed an in silico analysis to investigate the distribution of *C. jejuni* and *C. coli* isolates carrying the L-fucose and D-glucose utilization cluster and correlated the presence of the L-fucose utilization cluster in *Campylobacter* isolates to human host isolates, pointing to possible roles in transmission and/or virulence in the human host.

Previous studies highlighted that there are at least two versions of the L-fucose utilization cluster, one with an intact cluster (*Cj0480c – Cj0489)* and one with a frameshift in the *Cj0489* gene (*Cj0480c – Cj0489-L)* (Muraoka and Zhang, 2011; Middendorf et al., 2022). Here we found that all intact and truncated L-fucose utilization clusters were integrated in the exact same location of the genome in both *C. jejuni* and *C. coli* isolates (Fig. 1). The L-fucose utilization cluster was flanked upstream by the genes *rpoB* and *rpoC* while downstream by *rspL*, *repsG*, and *fusA*. This is a genomic area that is typically targeted by antibiotics such as rifampicin and streptomycin (Huang et al., 2003; Goldstein, 2014), and resistance is almost exclusively acquired by point mutations or indels (Comas et al., 2012; Lo et al., 2018; Godfroid et al., 2020). To gain a better understanding about the L-fucose cluster and how it evolved, we subsequently aligned all *C. jejuni* and *C. coli* fucose utilization clusters of the dataset from the Netherlands to identify mutations within the cluster (Suppl. Fig. 10). Interestingly, only 36 % of the fuc+ *C. jejuni* isolates did not carry any frameshift mutation in the cluster; in the remaining strains common frameshift mutations were found in the aldehyde dehydrogenase *Cj0489* (39 %), the unknown transporter *Cj0484* (14 %) and the amidohydrolase *Cj0487* (8 %) (Suppl. Fig. 10). Studies where knockout assays of each gene of the L-fucose utilization clusters were performed showed that knockouts of *Cj0489* and *Cj0484* did not result in a loss of function, because these knockouts showed the same phenotype as those with a fully intact cluster. In contrast, *Cj0487* knockouts showed no growth stimulation in the presence of L-fucose, highlighting that *Cj0487* is an essential gene for L-fucose metabolism (Stahl et al., 2011; Dwivedi et al., 2016; Garber et al., 2020). Notably, these knockout isolates did still show chemotaxis towards L-fucose (Stahl et al., 2011; Dwivedi et al.,

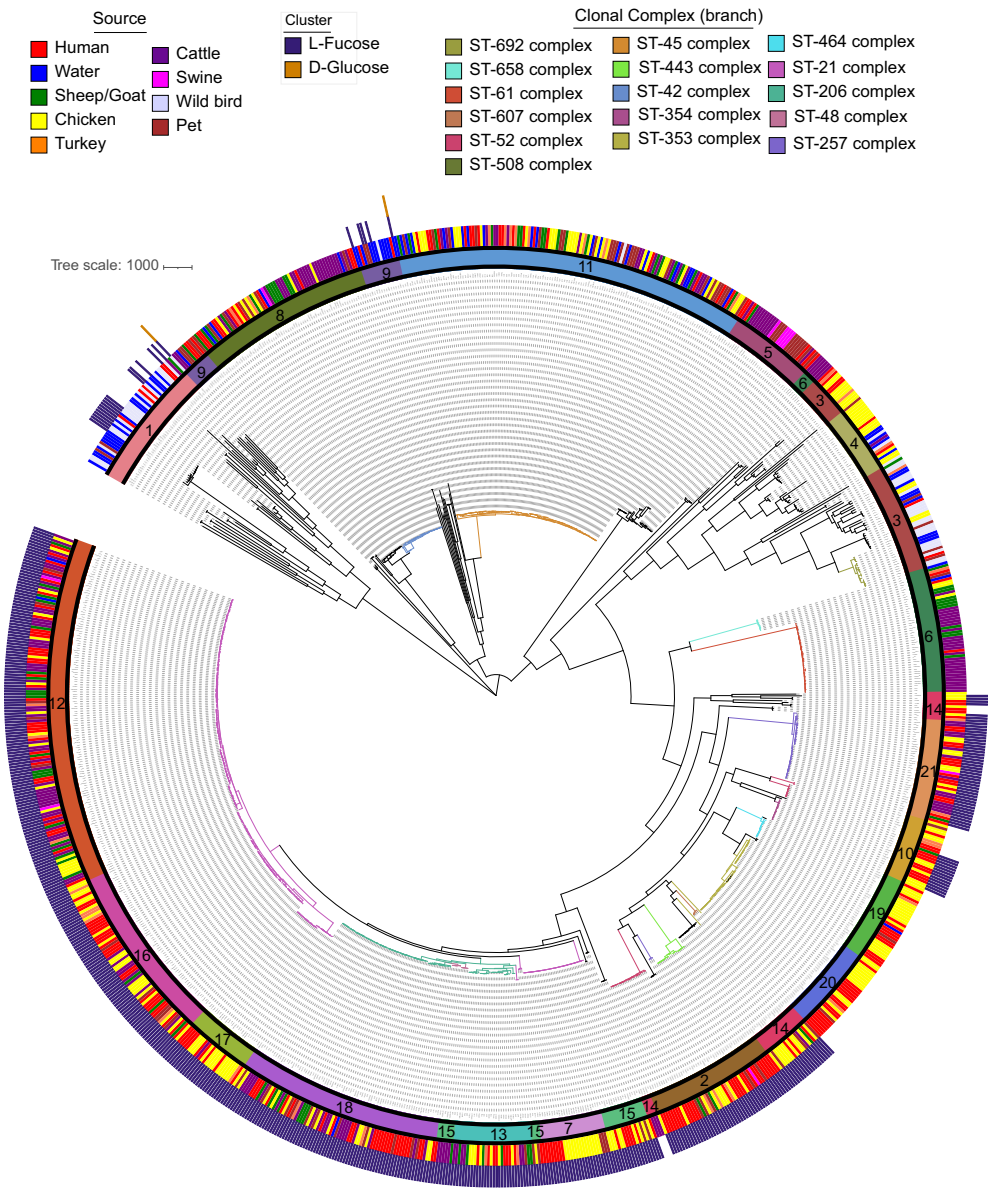

Fig. 5. Phylogenetic tree using 1057 full genomes of *C. jejuni* isolates from the Netherlands. From outside to inside: the outer (orange) ring indicate gluc+ isolates, the purple ring indicates fuc+ isolates, the multicolored ring indicates the isolation source per isolate and, lastly, the inner ring (multicolored with numbers) indicates the different Treestruct groups. Branch colors indicate the different clonal complexes (MLST groups).

2016). In the current study, all isolates that carried the *Cj0487* mutation also carried a mutation in Cj0489. Furthermore, all these isolates were either chicken or human isolates and belonged to the same clonal cluster (Suppl. Fig. 11). Since information on performance of such *C. jejuni* double mutants is not available, additional studies with selected isolates are required to determine the effect of L-fucose on metabolism, growth, chemotaxis and virulence. Unlike *C. jejuni*, *C. coli* isolates did not carry any mutation in the L-fucose utilization cluster. Comparing the source of fuc+ *C. jejuni* and *C. coli* isolates, much higher quantities of fuc+ isolates were found in livestock associated *C. coli* isolates, with over 80 % of the *C. coli* swine isolates carrying the L-fucose cluster. These results highlight a selection pressure of carrying an intact L-fucose utilization cluster in *C. coli*, possibly for the colonization and/or long-term survival in the animal host.

The integration of the D-glucose utilization cluster was much more diverse. We found evidence that the cluster was integrated in at least three different genomic locations, all within the three different 16S—23S ribosomal RNA regions, which we highlighted by viewing the genome of *C. jejuni* NCTC11168 (accession: NC_002163) (Suppl. Figs. 1 and 3). There could be several reasons for this observation, such as the finding that rearrangement can take place in well conserved regions like the 16S—23S areas (Page et al., 2020). Alternatively, due to faulty assemblies, possible rearrangements within the 16S—23S areas are observed in genome sequences, yet are not actually existing (Page et al., 2020). On the other hand, the high sequence identity of the flanking regions of the D-glucose utilization cluster allows for efficient integration of DNA from the environment, as *Campylobacter* is genetically competent.

Interestingly, some gluc+ *C. jejuni* and *C. coli* isolates carried several galactose utilization genes next to the D-glucose utilization cluster. Although the number of gluc+ *C. jejuni* and *C. coli* isolates in which the flanking regions could be identified was low (10 *C. jejuni* and 21 *C. coli* isolates), this suggested that this version of the D-glucose utilization cluster with neighboring the galactose cluster, has a different origin than the other analyzed D-glucose utilization clusters and was most likely introduced from other bacterial species to *Campylobacter*. However, as

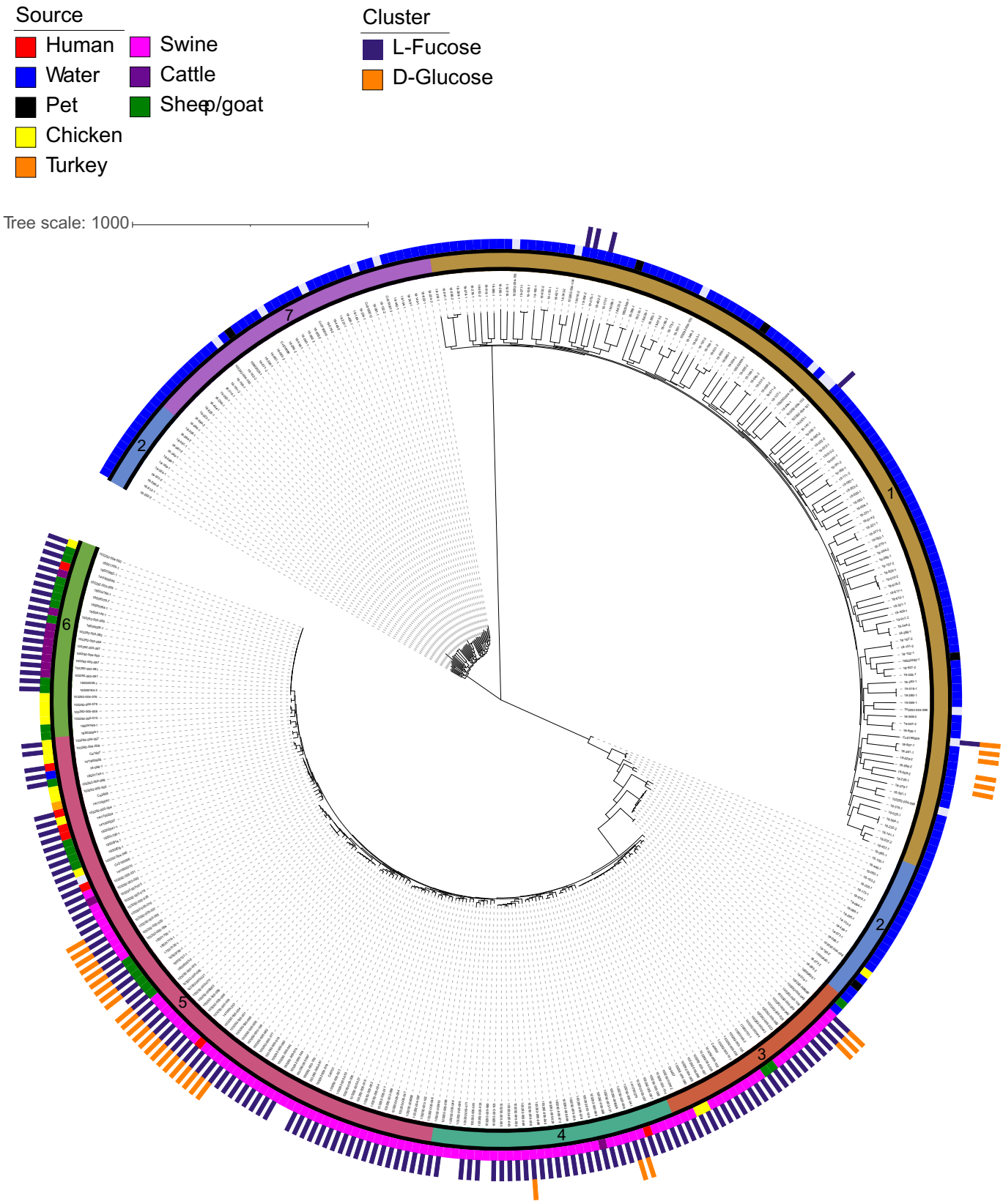

Fig. 6. Phylogenetic tree using 349 full genomes of *C. coli* isolates from the Netherlands. From outside to inside: the outer (orange) ring indicate gluc+ isolates, the purple ring indicate fuc+ isolates, The multicolored ring indicate the isolation source per isolate and, lastly, the inner ring (multicolored with numbers) indicates the different Treestruct groups.

the current dataset of glucose integration sites is very limited, more long-read sequencing procedures have to be performed to obtain a better insight in the origin of the galactose cluster. The function of the observed *gal* genes are putative, however, studies in different kinds of bacteria have shown that several of the genes were linked to the LPS synthesis, galactose metabolism (via the Leloir pathway) or synthesis of polysaccharides (Houng et al., 1990; Maskell et al., 1991; Kanipes et al., 2008).

The influence of L-fucose on the virulence and colonization of *Campylobacter* has been studied in piglets and chicken models, and in human cell lines, using wildtype (wt) and selected mutants in the L-fucose utilization cluster (Stahl et al., 2011; Luijkx et al., 2020). In addition, L-fucose pre-activated *C. jejuni* NCTC11168 cells recently showed increased invasion of Caco-2 epithelial cells and binding to fibronectin, highlighting a possible advantage of fuc+ strains in transmission across fucose-rich rich environments, such as from the animal intestine to the human gut (Middendorf et al., 2024). In the current study we investigated whether isolates carrying the L-fucose utilization cluster are more prevalent in human isolates by using correlation analyses on selected datasets from four countries. The dataset from the Netherlands was obtained from a country-wide study that was performed over a time period of two years. In the study design, elimination of issues related to differential recall bias, selection bias and misclassification were performed (Mughini-Gras et al., 2021). The other *C. jejuni* collections were obtained from the PubMLST database, in which the exact origin of isolates and type of sampling plans are unknown. Furthermore, possible selection bias in these datasets, for example towards disease symptoms or outbreaks, cannot be excluded. It was therefore not surprising that only a significant correlation was found between the percentage of human isolates and the percentage of fuc+ isolates in the *C. jejuni* dataset from the Netherlands. In the intestine of poultry, fuc+ isolates do not have a competitive colonization advantage over *fucP* mutant isolates, possibly due to inaccessibility of the fucosylated O-glycan mucin structures from chicken that are highly sulfated in contrast to mucin structures in humans and pigs (Stahl et al., 2011; Luis et al., 2022). Interestingly, 80 % of the campylobacteriosis cases in the Netherlands are linked to the poultry reservoir as a whole (Doorduyn et al., 2010; Mughini-Gras et al., 2012; EFSA, 2021; Mughini-Gras et al., 2021). Combining all information, it is likely that *C. jejuni* isolates carrying the L-fucose utilization cluster have an advantage in transmission from poultry towards humans by increased survival and/or proliferation in the human gut. Next to the observed correlation between fuc+ isolates and human isolates in the Dutch *C. jejuni* dataset, a correlation was also evident in the combined dataset, suggesting potential benefits in human infection and/or proliferation of fuc+ *C. jejuni* in the human gut after transmission from animal or environmental reservoirs world-wide.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijfoodmicro.2024.110855.

CRediT authorship contribution statement

Pjotr S. Middendorf: Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Aldert L. Zomer:** Writing – review & editing, Supervision, Software, Conceptualization. **Indra L. Bergval:** Writing – review & editing, Conceptualization. **Wilma F. Jacobs-Reitsma:** Writing – review & editing, Supervision, Conceptualization. **Heidy M.W. den Besten:** Writing – review & editing, Supervision, Conceptualization. **Tjakko Abee:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

None.

Data availability

Data will be made available on request.

Acknowledgements

This research was financially supported by the Graduate School VLAG (Wageningen University & Research, Wageningen, The Netherlands).

AZ acknowledges financial support by the Netherlands Organisation for Health Research and Development (ZonMw) with grant number 50-52200-98-316 (project name: "DEPiCT – Discerning Environmental Pathways of *Campylobacter* Transmission").

All genomes of the Dutch dataset were obtained via the Dutch National Institute for Public Health and the Environment (RIVM).

References

Allos, B.M., 1997. Association between *Campylobacter* infection and Guillain-Barré syndrome. J. Infect. Dis. 176 (Supplement_2), S125–S128. https://doi.org/10.1086/513783.

Andersen, M.T., Brøndsted, L., Pearson, B.M., Mulholland, F., Parker, M., Pin, C., et al., 2005. Diverse roles for HspR in *Campylobacter jejuni* revealed by the proteome, transcriptome and phenotypic characterization of an hspR mutant. Microbiology 151 (3), 905–915. https://doi.org/10.1099/mic.0.27513-0.

Besser, J., Carleton, H.A., Gerner-Smidt, P., Lindsey, R.L., Trees, E., 2018. Next-generation sequencing technologies and their application to the study and control of bacterial infections. Clin. Microbiol. Infect. 24 (4), 335–341. https://doi.org/10.1016/j.cmi.2017.10.013.

Blaser, M.J., 1997. Epidemiologic and clinical features of *Campylobacter jejuni* infections. J. Infect. Dis. 176 (Supplement_2), S103–S105. https://doi.org/10.1086/513780.

Brøndsted, L., Andersen, M.T., Parker, M., Jørgensen, K., Ingmer, H., 2005. The HtrA protease of *Campylobacter jejuni* is required for heat and oxygen tolerance and for optimal interaction with human epithelial cells. Appl. Environ. Microbiol. 71 (6), 3205–3212. https://doi.org/10.1128/AEM.71.6.3205-3212.2005.

Bronowski, C., James, C.E., Winstanley, C., 2014. Role of environmental survival in transmission of *Campylobacter jejuni*. FEMS Microbiol. Lett. 356 (1), 8–19. https://doi.org/10.1111/1574-6968.12488.

Calland, J.K., Pascoe, B., Bayliss, S.C., Mourkas, E., Berthenet, E., Thorpe, H.A., et al., 2021. Quantifying bacterial evolution in the wild: a birthday problem for *Campylobacter* lineages. PLoS Genet. 17 (9), e1009829 https://doi.org/10.1371/journal.pgen.1009829.

Candon, H.L., Allan, B.J., Fraley, C.D., Gaynor, E.C., 2007. Polyphosphate kinase 1 is a pathogenesis determinant in *Campylobacter jejuni*. J. Bacteriol. 189 (22), 8099–8108. https://doi.org/10.1128/JB.01037-07.

Comas, I., Borrell, S., Roetzer, A., Rose, G., Malla, B., Kato-Maeda, M., et al., 2012. Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. Nat. Genet. 44 (1), 106–110. https://doi.org/10.1038/ng.1038.

Doorduyn, Y., Van Den Brandhof, W., Van Duynhoven, Y., Breukink, B., Wagenaar, J., Van Pelt, W., 2010. Risk factors for indigenous *Campylobacter jejuni* and *Campylobacter coli* infections in the Netherlands: a case-control study. Epidemiol. Infect. 138 (10), 1391–1404. https://doi.org/10.1017/S095026881000052X.

Dwivedi, R., Nothaft, H., Garber, J., Xin Kin, L., Stahl, M., Flint, A., et al., 2016. L-fucose influences chemotaxis and biofilm formation in *Campylobacter jejuni*. Mol. Microbiol. 101 (4), 575–589. https://doi.org/10.1111/mmi.13409.

EFSA, 2010. Scientific opinion on quantification of the risk posed by broiler meat to human campylobacteriosis in the EU. EFSA J. 8 (1), 1437.

EFSA, 2021. The European Union one health 2019 zoonoses report. EFSA J. 19 (2).

Franz, E., Gras, L.M., Dallman, T., 2016. Significance of whole genome sequencing for surveillance, source attribution and microbial risk assessment of foodborne pathogens. Curr. Opin. Food Sci. 8, 74–79. https://doi.org/10.1016/j.cofs.2016.04.004.

Garber, J.M., Nothaft, H., Pluvinage, B., Stahl, M., Bian, X., Porfirio, S., et al., 2020. The gastrointestinal pathogen *Campylobacter jejuni* metabolizes sugars with potential help from commensal *Bacteroides vulgatus*. Communications Biology 3 (1), 1–11. https://doi.org/10.1038/s42003-019-0727-5.

Gilchrist, C.L., Chooi, Y.-H., 2021. Clinker & clustermap.js: automatic generation of gene cluster comparison figures. Bioinformatics 37 (16), 2473–2475. https://doi.org/10.1093/bioinformatics/btab007.

Godfroid, M., Dagan, T., Merker, M., Kohl, T.A., Diel, R., Maurer, F.P., et al., 2020. Insertion and deletion evolution reflects antibiotics selection pressure in a *Mycobacterium tuberculosis* outbreak. PLoS Pathog. 16 (9), e1008357 https://doi.org/10.1371/journal.ppat.1008357.

Goldstein, B.P., 2014. Resistance to rifampicin: a review. J. Antibiot. 67 (9), 625–630. https://doi.org/10.1038/ja.2014.107.

Golz, J.C., Stingl, K., 2021. Natural competence and horizontal gene transfer in *Campylobacter*. In: Fighting Campylobacter Infections. Springer, pp. 265–292.

Hepworth, P.J., Ashelford, K.E., Hinds, J., Gould, K.A., Witney, A.A., Williams, N.J., et al., 2011. Genomic variations define divergence of water/wildlife-associated

Campylobacter jejuni niche specialists from common clonal complexes. Environ. Microbiol. 13 (6), 1549–1560. https://doi.org/10.1111/j.1462-2920.2011.02461.x.
Houng, H., Kopecko, D.J., Baron, L.S., 1990. Molecular cloning and physical and functional characterization of the *Salmonella typhimurium* and *Salmonella typhi* galactose utilization operons. J. Bacteriol. 172 (8), 4392–4398. https://doi.org/10.1128/jb.172.8.4392-4398.1990.
Huang, H.-N., Han, J.-X., Wang, J.-H., Song, C.-Z., Liang, H., Zhang, Z.-L., 2003. rpsL gene analysis associated with streptomycin resistance in *Mycobacterium* tuberculosis. *Yi Chuan xue bao=*. Acta Genet. Sin. 30 (4), 376–381.
Jolley, K.A., Bray, J.E., Maiden, M.C., 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Research 3. https://doi.org/10.12688/wellcomeopenres.14826.1.
Kaakoush, N.O., Castaño-Rodríguez, N., Mitchell, H.M., Man, S.M., 2015. Global epidemiology of *Campylobacter* infection. Clin. Microbiol. Rev. 28 (3), 687–720. https://doi.org/10.1128/CMR.00006-15.
Kanipes, M.I., Tan, X., Akelaitis, A., Li, J., Rockabrand, D., Guerry, P., et al., 2008. Genetic analysis of lipooligosaccharide core biosynthesis in *Campylobacter jejuni* 81-176. J. Bacteriol. 190 (5), 1568–1574. https://doi.org/10.1128/JB.01696-07.
Katoh, K., Misawa, K., Kuma, K.i., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30 (14), 3059–3066.
Katz, L.S., Griswold, T., Morrison, S.S., Caravas, J.A., Zhang, S., den Bakker, H.C., et al., 2019. Mashtree: a rapid comparison of whole genome sequence files. Journal of Open Source Software 4 (44). https://doi.org/10.21105/joss.01762.
Lefebure, T., Pavinski Bitar, P.D., Suzuki, H., Stanhope, M.J., 2010. Evolutionary dynamics of complete *Campylobacter* pan-genomes and the bacterial species concept. Genome Biol. Evol. 2, 646–655. https://doi.org/10.1093/gbe/evq048.
Levin, R.E., 2007. *Campylobacter jejuni*: a review of its characteristics, pathogenicity, ecology, distribution, subspecies characterization and molecular methods of detection. Food Biotechnol. 21 (4), 271–347. https://doi.org/10.1080/08905430701536565.
Lo, S.W., Kumar, N., Wheeler, N.E., 2018. Breaking the code of antibiotic resistance. Nat. Rev. Microbiol. 16 (5), 262. https://doi.org/10.1038/nrmicro.2018.33.
Luijkx, Y.M., Bleumink, N.M., Jiang, J., Overkleeft, H.S., Wösten, M.M., Strijbis, K., et al., 2020. *Bacteroides fragilis* fucosidases facilitate growth and invasion of *Campylobacter jejuni* in the presence of mucins. Cell. Microbiol. 22 (12), e13252 https://doi.org/10.1111/cmi.13252.
Luis, A.S., Baslé, A., Byrne, D.P., Wright, G.S., London, J.A., Jin, C., et al., 2022. Sulfated glycan recognition by carbohydrate sulfatases of the human gut microbiota. Nat. Chem. Biol. 18 (8), 841–849. https://doi.org/10.1038/s41589-022-01039-x.
Maskell, D., Szabo, M., Butler, P., Williams, A., Moxon, E., 1991. Molecular analysis of a complex locus from *Haemophilus influenzae* invoked in phase-variable lipopolysaccharide biosynthesis. Mol. Microbiol. 5 (5), 1013–1022. https://doi.org/10.1111/j.1365-2958.1991.tb01874.x.
Middendorf, P.S., Jacobs-Reitsma, W.F., Zomer, A.L., den Besten, H.M.W., Abee, T., 2022. Comparative analysis of L-fucose utilization and its impact on growth and survival of *Campylobacter* isolates. Front. Microbiol. 1410 https://doi.org/10.3389/fmicb.2022.872207.
Middendorf, P.S., Wijnands, L.M., Boeren, S., Zomer, A.L., Jacobs-Reitsma, W.F., Den Besten, H.M.W., Abee, A., 2024. Activation of the L-fucose utilization cluster in *Campylobacter jejuni*induces proteomic changes and enhances Caco-2 cell invasion and fibronectin binding. Heliyon 10 (6), e34996. https://doi.org/10.1016/j.heliyon.2024.e34996.
Mughini-Gras, L., Smid, J.H., Wagenaar, J.A., De Boer, A.G., Havelaar, A.H., Friesema, I. H., et al., 2012. Risk factors for campylobacteriosis of chicken, ruminant, and environmental origin: A combined case-control and source attribution analysis. PLoS One, e42599. https://doi.org/10.1371/journal.pone.0042599.
Mughini-Gras, L., Pijnacker, R., Coipan, C., Mulder, A.C., Veludo, A.F., de Rijk, S., et al., 2021. Sources and transmission routes of campylobacteriosis: a combined analysis of genome and exposure data. J. Infect. 82 (2), 216–226. https://doi.org/10.1016/j.jinf.2020.09.039.
Muraoka, W.T., Zhang, Q., 2011. Phenotypic and genotypic evidence for L-fucose utilization by *Campylobacter jejuni*. J. Bacteriol. 193 (5), 1065–1075. https://doi.org/10.1128/JB.01252-10.
Nilsson, A., Johansson, C., Skarp, A., Kaden, R., Bertilsson, S., Rautelin, H., 2018. Survival of *Campylobacter jejuni* and *Campylobacter coli* water isolates in lake and well water. Apmis 126 (9), 762–770. https://doi.org/10.1111/apm.12879.
Page, A.J., Ainsworth, E.V., Langridge, G.C., 2020. socru: typing of genome-level order and orientation around ribosomal operons in bacteria. Microbial Genomics 6 (7). https://doi.org/10.1099/mgen.0.000396.
Paradis, E., Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35 (3), 526–528. https://doi.org/10.1093/bioinformatics/bty633.
Price, M.N., Dehal, P.S., Arkin, A.P., 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. 26 (7), 1641–1650. https://doi.org/10.1093/molbev/msp077.
Rees, J.H., Soudain, S.E., Gregson, N.A., Hughes, R.A., 1995. *Campylobacter jejuni* infection and Guillain–Barré syndrome. N. Engl. J. Med. 333 (21), 1374–1379. https://doi.org/10.1056/NEJM199511233332102.
Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30 (14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
Sheppard, S.K., Maiden, M.C., 2015. The evolution of *Campylobacter jejuni* and *Campylobacter coli*. Cold Spring Harb. Perspect. Biol. 7 (8), a018119 https://doi.org/10.1101/cshperspect.a018119.
Silva, J., Leite, D., Fernandes, M., Mena, C., Gibbs, P.A., Teixeira, P., 2011. *Campylobacter* spp. as a foodborne pathogen: a review. Front. Microbiol. 2, 200. https://doi.org/10.3389/fmicb.2011.00200.
Stahl, M., Friis, L.M., Nothaft, H., Liu, X., Li, J., Szymanski, C.M., et al., 2011. L-fucose utilization provides *Campylobacter jejuni* with a competitive advantage. Proc. Natl. Acad. Sci. 108 (17), 7194–7199. https://doi.org/10.1073/pnas.1014125108.
Tack, D.M., Marder, E.P., Griffin, P.M., Cieslak, P.R., Dunn, J., Hurd, S., et al., 2019. Preliminary incidence and trends of infections with pathogens transmitted commonly through food—Foodborne Diseases Active Surveillance Network, 10 US sites, 2015–2018. Morb. Mortal. Wkly Rep. 68 (16), 369. https://doi.org/10.15585/mmwr.mm6816a2.
Thames, H.T., Theradiyil Sukumaran, A., 2020. A review of Salmonella and Campylobacter in broiler meat: emerging challenges and food safety measures. Foods 9 (6), 776.
Vegge, C.S., Jansen van Rensburg, M.J., Rasmussen, J.J., Maiden, M.C., Johnsen, L.G., Danielsen, M., et al., 2016. Glucose metabolism via the Entner-Doudoroff pathway in *Campylobacter*: a rare trait that enhances survival and promotes biofilm formation in some isolates. Front. Microbiol. 7, 1877. https://doi.org/10.3389/fmicb.2016.01877.
Volz, E.M., Carsten, W., Grad, Y.H., Frost, S.D., Dennis, A.M., Didelot, X., 2020. Identification of hidden population structure in time-scaled phylogenies. Syst. Biol. 69 (5), 884–896. https://doi.org/10.1093/sysbio/syaa009.
Vorwerk, H., Huber, C., Mohr, J., Bunk, B., Bhuju, S., Wensel, O., et al., 2015. A transferable plasticity region in *Campylobacter coli* allows isolates of an otherwise non-glycolytic food-borne pathogen to catabolize glucose. Mol. Microbiol. 98 (5), 809–830. https://doi.org/10.1111/mmi.13159.
Wagenaar, J.A., Newell, D.G., Kalupahana, R.S., Mughini-Gras, L., 2015. *Campylobacter*: Animal reservoirs, human infections, and options for control. In: Zoonoses-infections Affecting Humans and Animals. Springer, pp. 159–177.