Molecular and *In Silico* typing of the lipooligosaccharide biosynthesis gene cluster in *Campylobacter jejuni* and *Campylobacter coli*

Amber Hameed[1], Julian Ketley[2], Alex Woodacre[1], Lee R. Machado[1*], Gemma L. Marsden[3]

[1] Centre for Physical Activity and Life Sciences, University of Northampton, Northampton, UK

[2] Department of Genetics and Genome Biology, University of Leicester, Leicester, UK

[3] Healthcare Infection Society, London, UK

* Corresponding author

E-mail: lee.machado@northampton.ac.uk (LRM)

Abstract

The extensive genetic variation in the lipooligosaccharide (LOS) core biosynthesis gene cluster has led to the development of a classification system; with 8 classes (I-VIII) for *Campylobacter coli* (*C. coli*) LOS region and with 23 classes (A-W) or four groups (1-4) for *Campylobacter jejuni* (*C. jejuni*) LOS region. PCR based LOS locus type identification for *C. jejuni* clinical isolates from a UK hospital as well as *in silico* LOS locus analysis for *C. jejuni* and *C. coli* genome sequences from GenBank was carried out to determine the frequencies of various LOS genotypes in *C. jejuni* and *C. coli*. Analysis of LOS gene content in 60 clinical *C. jejuni* isolates and 703 *C. jejuni* genome sequences revealed that class B (Group 1) was the most abundant LOS class in *C. jejuni*. The hierarchy of *C. jejuni* LOS group prevalence (group 1 > group 2 > group 3 > group 4) as well as the hierarchy of the frequency of *C. jejuni* LOS classes present within the group 1 (B > C > A > R > M > V), group 2 (H/P > O > E > W), group 3 (F > K > S) and group 4 (G > L) was identified. *In silico* analysis of LOS gene content in 564 *C. coli* genome sequences showed class III as the most abundant LOS locus type in *C. coli*. *In silico* analysis of LOS gene content also identified three novel LOS types of *C. jejuni* and previously unknown LOS biosynthesis genes in *C. coli* LOS locus types I, II, III, V and VIII. This study provides *C. jejuni* and *C. coli* LOS loci class frequencies in a smaller collection of *C. jejuni* clinical isolates as well as within the larger, worldwide database of *C. jejuni* and *C. coli*.

Introduction

Campylobacter is a foodborne enteropathogen which transmits to humans mainly by consumption of *Campylobacter* contaminated dairy products and raw or partially cooked meat [1, 2]. Most *Campylobacter* species, including *C. jejuni*, *C. coli*, *C. lari*, *C. fetus*, *C. upsaliensis, C. hypointestinalis*, *C. helveticus*, *C. lanienae*, and *C. mucosalis* are found in warm-blooded animals [2-7]. However, some *Campylobacter* species, such as *C. fetus* and *C. geochelonis* can also occur in cold-blooded reptiles (lizard, tortoise, and snake) [8, 9]. Chicken is the main reservoir of *Campylobacter* through intestinal colonisation after hatching, usually at the age of 2-5 weeks [10-12]. Chickens contaminated with *Campylobacter* (with approximately 10^9 CFU/g caecal contents) are considered as a major source of *Campylobacter* transmission to humans [13, 14]. *Campylobacter* isolates can also be present as environmental contamination in non-livestock niches, which may be agricultural or non-agricultural [15-18].

The estimated number of cases of *Campylobacter* infection is approximately 96 million per year worldwide [19]. *Campylobacter* infection is characterised by an acute, self-limiting gastroenteritis in humans which causes various clinical symptoms including watery or bloody diarrhoea, abdominal pain, headache, fever, chills, and dysentery [20-22]. In some cases, neuronal disorders Guillain–Barré syndrome (GBS) and Miller Fisher syndrome (MFS), Reiter's arthritis, and irritable bowel syndrome can also occur postinfection [23, 24].

Lipooligosaccharide (LOS) is an integral component of the outer cell membrane of *Campylobacter* and is synthesised by a cluster of lipooligosaccharide biosynthesis genes [25-27]. Each LOS biosynthesis gene present in this cluster produces an individual enzyme either for monosaccharide biosynthesis or addition of a particular monosaccharide to the LOS structure [28-31]. The LOS inner core biosynthesis genes

are less variable and flank a highly variable central region of LOS outer core (OC) biosynthesis genes [27, 28, 32]. At one flank are *waaC, waaM,* and *lgtF* (also known as *cj1135*), and at the other, *waaV, waaF, gmhA, waaE, waaD, gmhB,* and *cyf* (also known as *cj1153*); these IC biosynthesis genes occur in the same order in almost all *C. jejuni* and *C. coli* strains. However, the OC biosynthesis gene cluster that extends from *lgtF* and *waaV* varies extensively among *C. jejuni* and *C. coli* strains and this region can acquire new genes by horizontal gene transfer during infection inside an animal host [33, 34]. Insertion or deletion events in this gene region give rise to a new LOS locus organisation or type in *Campylobacter* strains, which can be variable both in gene content and gene organisation [29, 35]. Disruption in resident LOS biosynthesis genes and allelic variation can also contribute to new LOS types [29]. A *C. jejuni* LOS class comprises a specific organisation of LOS genes named alphabetically; altogether 23 *C. jejuni* LOS classes (A through W) were previously described [27, 30, 32]. An updated and simplified classification system with categorisation of the previously described 23 *C. jejuni* LOS classes into four LOS groups (1–4) was presented in order to better understand the prevalence of *C. jejuni* LOS groups and group-related LOS classes [36]. Group 1 includes all those LOS locus types, A, B, C, R, M and V, which contain the sialic acid biosynthesis genes (*neuA1*, *neuB1*, *neuC1* and *cst*-II/*cst-III*) whereas the other three groups had LOS loci with no sialic acid biosynthesis genes. Five classes (E, H, O, P and W) in LOS group 2, eight classes (D, F, K, Q, N, I, J, and S) in LOS group 3 and four classes (L, G, T, and U) in LOS group 4 were assimilated [36]. Variations in the LOS OC biosynthesis gene content cause modifications in the LOS OC structure [37, 38]. The LOS structures with variable OC epitopes help *Campylobacter* evade the host immune system as they mimic the human gangliosides [36, 38-42]. For this reason, antibodies produced

against the LOS structural epitopes not only bind to LOS structures, but also to human gangliosides [33, 38, 39, 43]. The cross-reactivity of anti-LOS antibodies with human gangliosides is a critical contributory factor that leads to the development of GBS or MFS in humans [23, 24, 44]. Thus, the expression of variable cell surface LOS structures due to variable gene content in the LOS locus in *C. jejuni* is considered an important virulence factor and may have direct connection with the progression of different neural disorders [37, 45]. For example, *C. jejuni* strains with LOS locus class A and variable human ganglioside mimics (GM1a, GM1b, GD1a, and GD1b) may help *C. jejuni* to trigger GBS post-infection in *Campylobacter* infected patients [38, 40, 41, 46]. In contrast, *C. jejuni* strains with LOS class B and corresponding GQ1b-like LOS structures are suggested to be associated with MFS in *Campylobacter* infected patients [38, 47]. Based on the strong relationship between the variable LOS synthesis region and *C. jejuni* virulence, the current study aimed to determine the frequency of *C. jejuni* LOS genotypes from whole genome sequences present within GenBank (n = 703), as well as in *C. jejuni* clinical isolates (n=60) from a UK hospital in order to further estimate the extent of gene variation in the *C. jejuni* LOS biosynthesis gene region and identify any novel *C. jejuni* LOS locus types. We previously analysed data from the literature relevant to the abundance of *C. jejuni* LOS classes in various geographical areas of the world to provide an overview of *C. jejuni* LOS genotype predominance [36]. The present study found that the prevalence of *C. jejuni* LOS group 1 classes (M, R, & V), group 2 class W, group 3 classes (K, Q, N, I, S & J) and all classes of group 4 (L, G, T & U) has never been identified before, which may be because of the relatively small size of collections involved in the previous studies. This study has for the first time presented the distribution of all, previously known LOS classes of *C. jejuni* using publicly accessible GenBank database.

Eight previously established *C. coli* LOS classes have been described I-VIII [32]. The relationship between different *C. coli* LOS locus structures and virulence is not fully established [48]. This is due to the presence of a wide variety in *C. coli* LOS biosynthesis locus types and limited knowledge of how this results in *C. coli* LOS structures at molecular level [32, 49]. As a result, we were interested in identifying the LOS locus types which dominate and are frequently present in *C. coli* isolates. To achieve this, we determined the frequency of *C. coli* LOS types present in GenBank which represents a global database of *C. coli* genome sequence deposits (n=564). A few studies have been carried out previously to estimate the distribution of *C. coli* LOS classes present within the collections of livestock associated *C. coli* strains (n=33) and agriculture related *C. coli* strains (n=261) [32, 50]. This study presents an up-to-date picture of LOS genotype predominance in *C. coli* and identifies previously unreported *C. coli* LOS biosynthesis genes. The overall aim of this work was to provide a hierarchy of the frequency of LOS classes and identify prevalence of any novel LOS genes in *C. jejuni* and *C. coli.* This will aid investigation of increasingly complex levels of LOS variation and the role such variation plays in *Campylobacter* infection.

Results

Identification of frequency of *C. jejuni* LOS locus classes in GenBank database

With the routine upload of whole genome sequences of *Campylobacter* spp. in Genbank we aimed to determine the frequencies of *C. jejuni* LOS locus classes and LOS groups in a collection of 703 *C. jejuni* GenBank sequences by *in silico* analysis (Fig 1 and supplementary S1 Table online). [Insert Figure 1]

Fig 1. A Circos plot showing the distribution of *C. jejuni* LOS locus classes (A-W), subclasses (A1, A2, B1, B2) and LOS groups (1-4) in the online *C. jejuni* sequence database.
Each segment of the inner circle specifies the total number of *C. jejuni* strains sequences (703) extracted for the LOS classification. The frequency of *C. jejuni* isolates classified for each particular LOS class/group is mentioned in numbers (n out of 703) on the top of each inner circle segment and represented with ribbon width. The frequency of a *C. jejuni* LOS class/group in percentage (% of overall frequency-100) is shown with each outer circle segment and represented by the orange or coloured bars. Ribbons link each *C. jejuni* LOS class to its related LOS group.

58% (n=400 of 703) of *C. jejuni* sequences belonged to the LOS group 1. The LOS classes A, B, and C adopted a hierarchy with class B1 (n=125; 18%) > C (n=109; 16%) > A1 (n=68; 10%) > A2 (n=44; 6%) > B2 (n=36; 5%), were the most common LOS group 1 related classes. Other group 1 related classes including R (n=10; 1.4%), M (n=7; 1%), and V (n=1; <1%) were rare. 30% (n=214) of sequences were positive for either class E (n=16; 2%), class H (n=72; 10%), class O (n= 63; 9%), class P (n=44; 6%) or class W (n=13; 2%) and therefore, belonged to LOS group 2. 10% (n=73) of *C. jejuni* sequences were positive for LOS group 3 classes including D (n=2; <1%), F (n=36; 5%), K (n=16; 2%), Q (n=1; <1%), N (n=1; <1%), I (n=3; <1%), S (n=6; 1%) and J (n=4; <1%). Only 2% of strains (n=10 and n=4 positive for class G and L, respectively) belonged to group 4.

Identification of frequency of *C. jejuni* LOS locus classes among *C. jejuni* clinical isolates

We were interested in examining whether Genbank LOS sequences were representative of LOS sequences present more locally. To address this, gDNA from 60 clinical *C. jejuni* isolates was extracted and the origin of DNA only from *C. jejuni* strains was confirmed by performing PCR reactions with *waaM* and *waaV* LOS gene specific control primers. Each DNA sample, positive for *waaM* and *waaV* LOS genes, was assigned with a LOS class based on PCR and Sanger sequencing results (supplementary S5 Table online). 50 of 60 *C. jejuni* isolates were typeable while

remaining 10 were non-typeable. 6 of 50 (12%) classified *C. jejuni* strains (54386, S2, 92691, 118973, 118715 and 93133Y) were PCR positive for more than one LOS class. The frequency of *C. jejuni* LOS locus classes (A through W), subclasses (A1, A2, B1, B2) and LOS groups (1-4) prevalent in a *C. jejuni* clinical strains collection was determined (Fig 2). 62% (n=31) of *C. jejuni* strains belonged to group 1 LOS classes. This included A1 (n=3; 6%), A2 (n=4; 8%), B1 (n=3; 6%), B2 (n=8; 16%), C (n=10; 20%), and a mixed ABC class (n=3; 6%). No *C. jejuni* strain was positive for other group 1 related classes (M, R, V). 32% (n=16) of *C. jejuni* strains were positive for group 2 classes. E (n=1; 2%), H (n=3; 6%), O (n=1; 2%), P (n=8; 16%) or a mix HP (n=3; 6%) were identified. Only 6% (n=3) *C. jejuni* strains were assigned to the LOS group 3 related class F. No *C. jejuni* strain was associated with other group 3 classes (D, K, Q, N, I, S, J). In addition, *C. jejuni* strains with the LOS group 4 classes, L, G, T, & U, were absent from the clinical *C. jejuni* isolates. [Insert Figure 2].

Fig 2. A Circos plot showing the distribution of *C. jejuni* LOS locus classes (A-W), subclasses (A1, A2, B1, B2) and LOS groups (1-4) from clinical isolates.
Each segment of the inner circle specifies the total number of classified *C. jejuni* isolates (50) used for the PCR based typing assay. The frequency of *C. jejuni* isolates classified for each particular LOS class/group is mentioned in numbers (n out of 50) on the top of each inner circle segment and represented with ribbon width. The frequency of a *C. jejuni* LOS class/group in percentage (% of overall frequency-100) is mentioned with each outer circle segment and represented by the orange or coloured bars. Ribbon ends link each *C. jejuni* LOS class to its related LOS group.

A comparison of *C. jejuni* LOS class and group frequencies, identified in both collections of clinical *C. jejuni* isolates and online *C. jejuni* sequences reveal that the frequencies of different classes were comparable between the two sources except class P (Fig 3). [Insert Figure 3]

Fig 3. A comparison of *C. jejuni* LOS biosynthesis locus class and group frequencies found in the collections of GenBank *C. jejuni* sequences (n=703) and *C. jejuni* clinical, typed strains (n=50)

Identification of novel LOS biosynthesis locus types in *C. jejuni*

C. jejuni 1336 (Accession no: CM000854.1), *C. jejuni* 414 (Accession no: ADGM01000014.1) and *C. jejuni* CFSAN054107 (Accession no: CP028185.1) were found with novel LOS gene organisations or LOS types which were designated as class X, class Y and class Z respectively. The LOS biosynthesis region present between previously known LOS genes (*cj1135*, ORF17 and *waaV*) contained 13 LOS biosynthesis genes in *C. jejuni* 1336, 5 in *C. jejuni* 414 and 5 in *C. jejuni* CFSAN054107 (Fig 4). [Insert Figure 4]

Fig 4. The genetic organisation of *C. jejuni* 1336, *C. jejuni* 414, and *C. jejuni* CFSAN05410 that contain novel LOS gene content

13 LOS biosynthesis genes in *C. jejuni* 1336, 5 LOS biosynthesis genes in *C. jejuni* 414 and 5 LOS biosynthesis genes in *C. jejuni* CFSAN054107 occurred between previously known LOS genes (*cj1135*, ORF17, and *waaV*). Green arrows: previously reported LOS genes in *C. jejuni* strains; Pink arrows: previously known, variable LOS genes that had similarity to the LOS biosynthesis genes of other *C. jejuni* strains; Blue arrows: LOS genes that had similarity to the CPS biosynthesis genes of other *C. jejuni* strains; Purple arrows: LOS genes that had similarity to the LOS biosynthesis genes of *C. coli* strains. The direction of arrow represents the direction of gene transcription. Black star: Gene with an unknown function.

Functions of 12 of 13 *C. jejuni* 1336 LOS genes are not known while one of them encodes an aminotransferase (WbdK). Functions of 4 of 5 *C. jejuni* 414 LOS genes are unknown while one of them encodes FkbM family methyltransferase. In *C. jejuni* CFSAN054107, 4 of 5 LOS genes encode glycosyltranferases and the remaining one encodes a methlytransferase (data from GenBank database). Further, the origin of these LOS novel genes was predicted by blast searching each gene against all sequences available in GenBank. 4 LOS genes in *C. jejuni* 1336 locus had >99% similarity with the capsular polysaccharide biosynthesis (CPS) genes of other *C. jejuni* strains, suggesting a possible gene transfer of CPS loci from other *C. jejuni* strains to the *C. jejuni* 1336 LOS locus. Six LOS genes of *C. jejuni* 1336 and one gene of *C.*

jejuni 414 had no identity with the previously known *C. jejuni* LOS genes. Instead, they had >99% similarity with various LOS biosynthesis genes of *C. coli*, suggesting interspecies gene recombination events.

Identification of frequency of *C. coli* LOS locus types in GenBank database

The frequency of different *C. coli* LOS locus types (Fig 5) present in the online, publicly available GenBank database was identified by *in silico* analysis of 564 *C. coli* sequences (supplementary S3 Table online). *C. coli* LOS class III (41%; n = 229) was the most abundant in the GenBank database followed by class VIII (18%; n = 103), class I (13%; n = 70), class II (8%; n = 47), class VII (7%; n = 42), class IV (5% (n = 27), class VI (4%; n =25) and class V (4%; n = 21). [Insert Figure 5]

Fig 5. Distribution of *C. coli* LOS locus classes within the online *C. coli* sequences GenBank database.
The number of *C. coli* strains associated with each *C. coli* LOS class in 2D column chart and corresponding percentages of *C. coli* strains in Pie chart, represent the frequency of *C. coli* LOS locus classes within the global collection of *C. coli* isolates.

Identification of previously unknown LOS biosynthesis genes in *C. coli*

Genes at the LOS biosynthesis cluster's one flank (*waaC*, *waaM*, *lgtF*) and at the other (*waaV*, *waaF*, *gmhA*, *waaE*, *waaD*, *gmhB*) are present with the same order or organisation in almost all *Campylobacter* strains and are involved in the biosynthesis of the LOS inner core [30, 49]. LOS class W (*C. jejuni* M1) [32] and class E (*C. jejuni* 81116) contain two genes between *waaF* and *gmhA* while class B (*C. jejuni* 81-176) contains a single gene between *waaF* and *gmhA*. Similarly, previously unreported LOS biosynthesis genes, localised between *waaF* and *gmhA* in the distal end of five

C. coli LOS locus types (I, II, III, V, VIII) were found in this study. Each *C. coli* LOS type had insertion of one LOS biosynthesis gene between *waaF* and *gmhA* (Fig 6, also given with *C. coli* reference strains in supplementary S4 Table online). These LOS genes are present at the same position in all five types of *C. coli* LOS locus but vary in size and nucleotide composition. The class I gene has a maximum size of ~1.4 kb and is distantly related to those genes which occur at the same position in other *C. coli* LOS locus types (II, III, V & VIII). The class II gene has similarity (~91%) to genes present in other classes (III, V and VIII). Type III and VIII genes are found to be identical (100%) and the class V gene is partially (~51%) similar to these identical genes. The presence of these previously unreported LOS biosynthesis genes was confirmed in 436 *C. coli* GenBank sequences (grey coloured columns in supplementary S3 Table online) within the LOS type I (n= 63), II (n=43), III (n=229), V (n=3), and VIII (n=98). [Insert Figure 6]

Fig 6. Genomic organisation of previously unreported *C. coli* LOS core genes and their relative similarity.
The position of novel gene content between *waaF* and *gmhA* in *C. coli* LOS types (I, II, III, V and VIII). The sizes of these genes, obtained from the GenBank database, are given in kb. Each dotted line links two genes to represent the similarity between them in terms of query cover score (non-bold) and Megablast identity score (bold).

The Class I gene encodes a β-Kdo transferase, class II gene encodes a phosphoheptose isomerase and genes in other classes produce glycosyltranferases for LOS synthesis (Data extracted from GenBank; supplementary S4 Table online).

Association of *C. jejuni* and *C. coli* LOS loci distribution to *Campylobacter* sources

The source of microbe isolation is usually specified with each sequence recorded in GenBank. The prevalence of *C. jejuni* and *C. coli* LOS genotypes in different

322 *Campylobacter* niches was estimated (Fig 7) by examining the online published
323 sources (also given in supplementary Tables S1 and S3 online) of these *C. jejuni* and
324 *C. coli* strains. *C. jejuni* and *C. coli* strains, isolated from faecal samples, were not
325 included in this analysis as the actual source was unknown. The frequency of LOS
326 genotypes within the pool of human *C. jejuni* isolates with online sequence data was
327 comparable to the frequency of LOS genotypes identified within the collection of *C.*
328 *jejuni* clinical isolates. For example, *C. jejuni* strains with LOS class F were common
329 in our clinical isolates in comparison to other group 3 LOS classes (K, Q, N, I, J, S)
330 and similarly, most abundantly isolated from humans prior to submit their genome
331 sequences in GenBank. *C. jejuni* LOS group 1 associated *C. jejuni* strains were found
332 mostly in humans [B (n=33; 8%) > C (n=29; 7%) > A (n=24; 6%)] and chickens [B
333 (n=23; 5%) > A (n=14; 3%) > C (n=9; 2%)]. Humans, chickens, and the animal farm
334 environment were the common isolation sources for *C. jejuni* and almost every *C.*
335 *jejuni* LOS class was associated with at least one of these sources. Moreover, the
336 most predominant LOS class III related *C. coli* strains were largely isolated from
337 animal farm environments (n=165; 37%), as well as from humans (n=20, 5%) and
338 chickens (n=16, 4%), indicating that these are the common niches for type III LOS
339 locus containing *C. coli* strains. The second most prevalent LOS class VIII was also
340 frequent in the environment (n=31, 7%) and at similar levels to isolates from humans
341 (n=26, 6%) and chickens (n=25, 6%). These results indicated that chicken and
342 humans are the most common hosts for *C. jejuni* while animal farm environments
343 (particularly farm water and soil) is the most common niche for *C. coli*. [Insert Figure
344 7]

345
346 **Fig 7. Frequency of *C. jejuni* and *C. coli* LOS locus classes in different *Campylobacter* sources**
347 Environment involves animal farm soil and animal farm water as *Campylobacter* sources.

348

Discussion

By analysing the distribution of *C. jejuni* LOS locus types, we determined that the frequency of LOS classes within the LOS group 1 was similar in both collections of *C. jejuni* online sequences [class B (23%) > class C (16%) ≥ class A (16%)] and *C. jejuni* clinical isolates [class B (22%) > class C (20%) > class A (14%)]. The LOS class B was the most common class in *C. jejuni* isolates from a clinical cohort as well as in the *C. jejuni* GenBank database. It has been reported previously in other studies [29, 47, 51-53) as the most common LOS class in humans as well as poultry *C. jejuni* isolates. The current study highlights that class C is the second most abundant LOS locus class in *C. jejuni*. Many other studies have described the LOS class C as the major class of *C. jejuni* LOS biosynthesis locus [54-55]. However, in comparison to the high prevalence of LOS class C (42%) in clinical isolates in Sweden [55], a very small number of clinical strains (2%) in Bangladesh had association with LOS locus C [47]. Most of the GBS-related *C. jejuni* strains in Bangladesh and China possessed the LOS locus class A rather than class C [47, 56, 57], suggesting that *C. jejuni* LOS class distribution is likely to vary geographically. High frequencies of classes B and C may be present due to their ability to encode heterogenous ganglioside mimics, which is advantageous for pathogenesis [58, 59]. The frequency of sialic acid biosynthesis LOS loci (A, B & C) in the online *C. jejuni* GenBank database as well as among the clinical *C. jejuni* isolates was high, although the former collection contained other *C. jejuni* sources (e.g., animals, birds and farm soil) in addition to human. It demonstrates that LOS group 1 containing LOS locus classes A, B, and C are commonly present in every type of *C. jejuni* source population. We have described this earlier in our literature review [36]. The low rate of GBS and MFS in *Campylobacter* infected patients [41, 46] despite high predominance of GBS/MFS associated LOS classes (A,

B, and C) in the clinical cohort supports the notion that some other factors in addition to LOS structures significantly contribute to the development of these neural diseases [35, 36, 47]. Other LOS classes from group 1 (R, V and M), also contain genes for the biosynthesis of sialylated LOS structures [30, 32, 42], but they were absent from the clinical isolates, and this may reflect the fact that, even in a large repository, sequences belonging to these classes represented only 2.5% of the online database. The reasons for poor distribution of these classes remain unclear. 16% of LOS typed *C. jejuni* clinical strains had the LOS group 2 related class P and 10% of analysed *C. jejuni* sequences belonged to the LOS group 2 related class H, marking the class P in our local clinical *C. jejuni* collection and class H in the online *C. jejuni* sequence database as the most predominant LOS group 2 classes. These contrasting results might be explained because both LOS class P and H share almost similar gene content except for two LOS biosynthesis genes (Orf 26' and Orf28) [57] and therefore only a few studies [47,52] have ever considered these classes as two separate classes. LOS group 2 appeared as the second most abundant group among a small population of *C. jejuni* from the enteritis cases, although these types of loci lack the sialic acid biosynthesis genes and likely produce non-sialylated LOS structures [30, 60]. Therefore, LOS sialylation appears to be advantageous when present, but may not be absolutely critical for Campylobacter survival in animal hosts. Ganglioside-like structures other than GM1 or GQ1b were infrequent in LOS locus E associated *C. jejuni* strains as evidenced using serological assays [40]. However, the prevalence was low and structural assays were not employed. Therefore, it currently remains unclear whether ganglioside mimicry is produced in any non-group 1 strains. Within the LOS group 3, class F was the most prevalent class among *C. jejuni* clinical isolates (6%) and in the online sequence database (5%). Class K (2%) was the second most

399 common class of group 3, whereas other LOS classes (I, S, J, D, Q, and N) were less
400 frequent (≤1%) in GenBank database. *C. jejuni* sequences in a very small number
401 (<3%) belonged to group 4. In contrast, no group 3 related classes (except for class
402 F) and group 4 related classes were identified from *C. jejuni* clinical isolates, which
403 may be because of the relatively small size of the collection. The hierarchy of LOS
404 group prevalence was group 1 > group 2 > group 3 > group 4 in both types of
405 collections of *C. jejuni* isolates.
406 Six *C. jejuni* strains were found positive for more than two LOS classes using PCR,
407 which occur due to co-infection in patients with multiple *C. jejuni* strains. The co-
408 infection occurrence with multiple *C. jejuni* strains has been previously observed in
409 GBS patients [61]. Another reason could be the occurrence of LOS gene
410 recombination during infection as it has been observed previously [33, 34]. The *in-*
411 *silico* prediction for the presence of six *C. coli* LOS genes in *C. jejuni* 1336 and one *C.*
412 *coli* LOS gene in *C. jejuni* 414 highlight the occurrence of interspecies genes
413 recombination events. *C. jejuni* does not only harbour the genes from *C. coli*, but *C.*
414 *coli* can also uptake and acquire *C. jejuni* DNA, especially when they are present in
415 the same niche [18]. *C. jejuni* and *C. coli* share 71% of LOS biosynthesis genes and
416 65% of CPS biosynthesis genes because of recombination events [32].
417 LOS locus type III was abundant (41%; n=229 of 564) in the online GenBank database
418 of *C. coli* sequences. The high frequency of class III (28%; n=72 of 261) within the
419 agriculture-associated *C. coli* strains has been reported in a previous study [50]. The
420 reasons behind the high prevalence of LOS class III in *C. coli* are yet to be
421 investigated. *C. coli* strain 76339 contains the sialic acid biosynthesis genes (*cst-V*,
422 *neuA*, *neuB*, & *neuC*) and sialic acids in its LOS structure [49, 62]. In the current study,
423 all *C. coli* LOS sequences extracted from GenBank were scrutinised for the presence

of these sialic acid biosynthesis genes and only *C. coli* RM4661 (Accession no: CP007181.1) was found to contain sialic acid synthesis genes (*cst*, *neuB* & *neuC*) in the LOS locus. However, *C. coli* LOS locus classes with *cst* alleles were identified in another study where XV-XXIV had *cst-II*; XIII and XIV had *cst-III* and IX, XXV, XXVI had *cst-V*. All these classes were positive for *neuABC* [49, 50]. Other LOS classes of *C. coli* including II, III, XXVII - XXXV had different alleles of *cst* (*cst-IV* and *cst-VI*) and only a small fraction of these classes (6.44% and 4.51% of strains positive with *cst-IV* and *cst-VI* respectively) had *neuABC* positioned outside of LOS locus [50]. Despite having genes (sialyltransferases) associated with the LOS sialylation, *C. coli* strains with ganglioside mimicry have not been reported so far [48-50, 63]. The most common LOS class (III) and the second most common class (VIII) linked *C. coli* strains were largely isolated from humans, which is concordant with a previous study, where half (57%) of the clinical isolates belonged to class III, VIII and II [64]. All *C. coli* classes tended to come from humans, chickens and the farm environment, suggesting that animal farm water and soil, in addition to chickens, are the primary sources of *C. coli* transmission to humans. This agrees with a previous study that reported agriculture associated *C. coli* as an emerging human pathogen [17, 18]. In this study, previously unreported LOS genes were identified in the *C. coli* LOS biosynthesis locus types I, II, III, V and VIII. The identified LOS biosynthesis genes in *C. coli* vary at the sequence and functional level among *C. coli* LOS locus classes (based on data available in GenBank). It highlights that other region of LOS biosynthesis locus also vary amongst *C. coli* strains in addition to the central region of LOS locus that can impact the LOS structure in *C. coli*. The sequence level variation in these genes and their possible putative functions have been determined *in silico* but require further functional characterisation.

The genome sequences deposited in GenBank may not be absolutely correct due to occurrence of errors at the experimental or sequence data analysis stages [65], and therefore can produce false-positive results. In this study, variation in LOS biosynthesis locus was analysed at the base sequence level to determine the presence of clustered whole LOS genes rather than finding the modifications between sequence bases. The identification of several LOS genes within a single genomic sequence reduces the possibility of false-positive results but does not eliminate it.

In conclusion, this work compares the frequency of various *C. jejuni* LOS locus classes in local versus global collections of *C. jejuni* isolates and provides an overview of predominance of various LOS biosynthesis gene clusters in the populations of *C. jejuni* and *C. coli*.

Materials and methods

Collection of Bacterial Strains and their growth

In a 12-month period from November 2015-2016, *C. jejuni* isolates (n=60; supplementary S5 Table online) from anonymised clinical samples from Northampton General Hospital, UK were collected by swabbing cultured Charcoal-Cefoperazone-Deoxycholate Agar (CCDA) plates. Amines and charcoal swabs (Thermo Fisher Scientific) were used for collection and transportation of *Campylobacter* isolates. Bacterial isolates were cultured again within 24 hours of collection and grown on MHA plates at 37 °C for 24-48 hours under a microaerobic atmosphere of 5% O_2, 10% CO_2 and 85% N_2. The microaerobic environment was provided by either using CampyGen sachets (Oxoid Limited) in 2.5 L air-tight jars or BOC gas mixture (2% H_2, 5% O_2, 10% CO_2 and 83% N_2) in a Whitley G2 workstation (Don Whitley Scientific).

DNA extraction from *C. jejuni* clinical strains

For the extraction of genomic DNA (gDNA) from *C. jejuni* isolates, the protocol provided by the DNeasy Blood and Tissue kit manufacturer (Qiagen) was followed.

Designing of LOS class specific primers

25 *C. jejuni* LOS class specific PCR primer pairs including *waaM* and *waaV* LOS gene specific control primers (supplementary S2 Table online) were designed using Clone Manager Professional Suite (Version 8; Scientific & Educational Software, Morrisville, USA) and purchased from the Eurofins Genomics (Ebersberg, Germany). Each LOS class specific primer pair spanned the junction of two adjacent LOS genes and were specific to two LOS biosynthesis genes rather than a single gene.

PCR for LOS typing

PCR master mix (20 µL) was prepared after mixing the template DNA (~50 ng) with forward and reverse primers (10µM each), MyTaq™ red DNA polymerase (0.25 µL containing 1.25 units; Bioline Reagents Ltd) and 5X MyTaq™ red Reaction Buffer (5 µL containing 5 mM dNTPs and 15 mM $MgCl_2$; Bioline Regents Ltd). Each reaction was carried out in a thermocyler (TECHNE) using appropriate cycling conditions: 1 cycle of initial template denaturation (95 °C; 5 min) followed by 35 cycles of template amplification (template denaturation at 95 °C for 30 sec, primer annealing at optimised temperature for 35 sec, primer extension at 72 °C for 30 sec per 1 kb of expected PCR product size) and one cycle of final extension (72 °C for 5 min). PCR products were resolved by electrophoresis on 1% (w/v; dissolved in 1X TAE buffer) agarose gel, which were stained with SYBR® safe stain (10,000X; Invitrogen) and visualised under UV light in G: Box (Syngene). PCRs with gDNA of reference *C. jejuni* strains (as

indicated in supplementary S5 Table online) were used as positive controls and PCR with gDNA of *C. jejuni* 11168Δ32-52 (a mutant strain of *C. jejuni* NCTC11168 lacking LOS biosynthesis region from gene *cj1132*-*cj1152*) was used as a negative control in all PCR reactions.

Sanger sequencing of PCR products

The protocol provided with the Eurofins Mix2Seq kit was followed for Sanger sequencing PCR products. According to the protocol, 15 µL purified PCR product (1-15 ng/µL) was mixed with 2 µL of either forward or reverse primer stock solution (10 pmol/µL). 17 µL DNA/primer mix was pipetted into a Mix2Seq tube and sent to Eurofins, Wolverhampton, U.K. for sequencing. Sequence data, obtained online in fasta format, was analysed using Clone Manager Professional Suite.

Bioinformatic analysis

The genome sequences of 703 *C. jejuni* strains (125 complete; 578 draft) and 564 *C. coli* strains (22 complete; 542 draft) available on the 1st March 2018 were obtained from GenBank and scrutinised by alignment using Megablast (https://blast.ncbi.nlm.nih.gov). The information related to the reference strains with previously defined LOS types (subject sequences) has been given in supplementary Tables S2 and S4 while for query sequences, used in this analysis, has been given in supplementary Tables S1 and S3. A LOS gene was considered present if ≥80% of the query sequence was effectively mapped to the reference LOS gene sequence and had ≥80% nucleotide identity with the reference LOS gene sequence. Subsequently, based on the presence or absence of distinct LOS genes, combination of LOS genes or a LOS class was identified, and assigned to a particular *C. jejuni* or *C. coli* query sequence. Megablast detects a query by a single accession number for a complete

genome sequence but incorporates multiple sub-accession numbers for a draft genome sequence. Therefore, to reduce the burden of query sequences, those draft sequence contigs were identified which contained the LOS biosynthesis gene sequence. For this purpose, all sequenced contigs associated with each draft sequence were aligned against the commonly present two LOS gene (*waaC* and *waaF*) sequences using Megablast. Subsequently, contigs with LOS sequence were used further for LOS classification analysis.

Ethics statement

Campylobacter isolates from anonymised clinical samples were collected by the Swab method under the sterile conditions from already cultured plates. No experimentation was carried out on humans during this research and this research did not involve the use of human tissues, fluids or DNA samples.

Acknowledgments

We would like to thank Andrea O'Connell at Northampton General Hospital for providing the clinical isolates.

Author contributions

Conceptualization: Julian Ketley, Gemma L. Marsden

Data curation: Amber Hameed

Formal analysis: Amber Hameed, Lee R. Machado, Alex Woodacre

Investigation: Amber Hameed, Lee R. Machado, Alex Woodacre

Methodology: Lee R. Machado, Alex Woodacre, Gemma L. Marsden

Project administration: Lee R. Machado

Supervision: Lee R. Machado, Alex Woodacre

Validation: Lee R. Machado, Amber Hameed, Alex Woodacre

Visualization: Lee R. Machado, Alex Woodacre

Writing – original draft: Amber Hameed

Writing – review & editing: Julian Ketley, Amber Hameed, Lee R. Machado, Alex Woodacre.

References

1. Robinson DA. Infective dose of *Campylobacter jejuni* in milk. Br Med J. (Clinical research ed). 1981;282(6276):1584.
2. Rahimi E, Kazemeini HR, Safaei S, Allahbakhshi K, Momeni M, Riahi M. Detection and identification of *Campylobacter* spp. from retail raw chicken, turkey, sheep and goat meat in Ahvaz, Iran. African J Microbiol Res. 2010;4(15):1620–3.
3. Roberts L, Lawson GH, Rowland AC. Experimental infection of neonatal pigs with *Campylobacter sputorum* subspecies *mucosalis* with special reference to the oral cavity. Vet Microbiol, 1980;5(3):249–255.
4. Gebhart CJ, Edmonds P, Ward GE, Kurtz HJ, Brenner DONJ. "*Campylobacter hyointestinalis"* sp. nov.: a New Species of *Campylobacter* Found in the Intestines of Pigs and Other Animals. J Clin Microbiol. 1985;21(5):715–720.
5. Inglis GD, Kalischuk LD, Busz HW, Kastelic JP. Colonization of cattle intestines by *Campylobacter jejuni* and *Campylobacter lanienae*. Appl Environ Microbiol. 2005;71(9):5145–53.
6. Workman SN, Mathison GE, Marc C, Lavoie MC. Pet Dogs and Chicken Meat as Reservoirs of *Campylobacter* spp . in Barbados Pet Dogs and Chicken Meat

as Reservoirs of *Campylobacter* spp . in Barbados. J Clin Microbiol. 2005;43(6):2642–50.

7. Weis AM, Miller WA, Byrne BA, Chouicha N, Boyce WM, Townsend AK. Prevalence and pathogenic potential of *Campylobacter* isolates from free-living, human-commensal American crows. Appl Environ Microbiol. 2014;80(5):1639–44.
8. Wang CM, Wu ZY, Shia WY, Jhou YJ, Tung KC, Shyu CL. Complete genome sequence of *Campylobacter fetus subsp. testudinum* strain Pet-3, isolated from a lizard (*Hydrosaurus pustulatus*). Genome Announc. 2015;3(1):1–14
9. Piccirillo A, Niero G, Calleros L, Pérez R, Naya H, Iraola G. *Campylobacter geochelonis* sp. nov. isolated from the western Hermann's tortoise (Testudo hermanni hermanni). Int J Syst Evol Microbiol. 2016;66(9):3468–76.
10. Neill SD, Campbell JN, Greene JA. *Campylobacter* species in broiler chickens. Avian Pathol. 1984;13(4):777–85.
11. Humphrey TJ, Henley A, Lanning DG. The colonization of broiler chickens with *Campylobacter jejuni*: some epidemiological investigations. Epidemiol Infect. 1993;110(3): 601–607.
12. Berndtson E, Danielsson-Tham ML, Engvall A. *Campylobacter* incidence on a chicken farm and the spread of *Campylobacter* during the slaughter process. Int J Food Microbiol. 1996;32(1-2):35–47.
13. Atanassova V, Ring C. Prevalence of *Campylobacter* spp. in poultry and poultry meat in Germany. Int J Food Microbiol. 1999;51(2-3):187–90.
14. Newell DG, Fearnley C. Sources of *Campylobacter* colonization in broiler chickens. Appl Environ Microbiol. 2003;69(8):4343–51.

15. Champion OL, Gaunt MW, Gundogdu O, Elmi A, Witney AA, Hinds J, et al. Comparative phylogenomics of the food-borne pathogen *Campylobacter jejuni* reveals genetic markers predictive of infection source. PNAS. 2005;102(44):16043–16048.
16. Wilson DJ, Gabriel E, Leatherbarrow AJH, Cheesbrough J, Gee S, Bolton E, et al. Tracing the source of campylobacteriosis. PLoS Genet. 2008;4(9).
17. Sheppard SK, Dallas JF, Wilson DJ, Strachan NJC, McCarthy ND, Jolley KA, et al. Evolution of an agriculture-associated disease causing *Campylobacter coli* clade: Evidence from national surveillance data in Scotland. PLoS One. 2010;5(12):1–9.
18. Sheppard SK, Didelot X, Jolley KA, Darling AE, Pascoe B, Meric G, et al. Progressive genome-wide introgression in agricultural *Campylobacter coli*. Mol Ecol. 2013;22(4):1051–64.
19. Havelaar AH, Kirk MD, Torgerson PR, Gibb HJ, Hald T, Lake RJ, et al. World Health Organization Global Estimates and Regional Comparisons of the Burden of Foodborne Disease in 2010. PLoS Med. 2015;12(12):e1001923.
20. van Spreeuwel JP, Duursma GC, Meijer CJ, Bax R, Rosekrans PC, Lindeman J. *Campylobacter* colitis: histological immunohistochemical and ultrastructural findings. Gut. 1985;26(9):945–51.
21. Black RE, Levine MM, Clements ML, Hughes TP, Blaser MJ. Experimental *Campylobacter jejuni* infection in humans. J Infect Dis. 1988;157(3):472–9.

22. Perkins DJ, Newstead GL. (1994) *Campylobacter jejuni* enterocolitis causing peritonitis, ileitis and intestinal obstruction. The ANZ J Surg. 1994;64(1):55–58.
23. Endtz HP, Ang CW, Van Den Braak N, Duim B, Rigter A, Price LJ, et al. Molecular characterization of *Campylobacter jejuni* from patients with Guillain-Barré and Miller Fisher syndromes. J Clin Microbiol. 2000;38(6):2297–301.
24. McCarthy N, Giesecke J. Incidence of Guillain-Barré syndrome following infection with *Campylobacter jejuni*. Am J Epidemiol. 2001;153(6):610–4.
25. Mandrell RE, McLaughlin R, Kwaik YA, Lesse A, Yamasaki R, Gibson B, et al. Lipooligosaccharides (LOS) of some *Haemophilus* species mimic human glycosphingolipids, and some LOS are sialylated. Infect Immun. 1992;60(4):1322–1328.
26. Moran AP. Structure and conserved characteristics of *Campylobacter jejuni* lipopolysaccharides. J Infect Dis. 1997;176:S115-21.
27. Gilbert M, Karwaski MF, Bernatchez S, Young NM, Taboada E, Michniewicz J, et al. The genetic bases for the variation in the lipo-oligosaccharide of the mucosal pathogen, *Campylobacter jejuni*. Biosynthesis of sialylated ganglioside mimics in the core oligosaccharide. J Biol Chem. 2002;277(1):327–37.
28. Karlyshev A, Ketley J, Wren B. The glycome. FEMS Microbiol Rev. 2005;29(2):377–90.
29. Parker CT, Horn ST, Gilbert M, Miller WG, Woodward DL, Mandrell RE. Comparison of *Campylobacter jejuni* lipooligosaccharide biosynthesis loci from a variety of sources. J Clin Microbiol. 2005;43(6):2771–81.
30. Parker CT, Gilbert M, Yuki N, Endtz HP, Mandrell RE. Characterization of lipooligosaccharide-biosynthetic loci of *Campylobacter jejuni* reveals new

lipooligosaccharide classes: Evidence of mosaic organizations. J Bacteriol. 2008;190(16):5681–9.

31. Iwata T, Chiku K, Amano K ichi, Kusumoto M, Ohnishi-Kameyama M, Ono H, et al. Effects of Lipooligosaccharide Inner Core Truncation on Bile Resistance and Chick Colonization by *Campylobacter jejuni*. PLoS One. 2013;8(2).
32. Richards VP, Lefébure T, Pavinski Bitar PD, Stanhope MJ. Comparative characterization of the virulence gene clusters (lipooligosaccharide [LOS] and capsular polysaccharide [CPS]) for *Campylobacter coli*, *Campylobacter jejuni* subsp. *jejuni* and related *Campylobacter* species. Infect Genet Evol. 2013;14(1):200–13.
33. Gilbert M, Godschalk PCR, Karwaski MF, Ang CW, Van Belkum A, Li J, et al. Evidence for Acquisition of the Lipooligosaccharide Biosynthesis Locus in *Campylobacter jejuni* GB11, a Strain Isolated from a Patient with Guillain-Barré Syndrome, by Horizontal Exchange. Infect Immun. 2004;72(2):1162–5.
34. Phongsisay V, Perera VN, Fry BN. Exchange of lipooligosaccharide synthesis genes creates potential Guillain-Barré syndrome-inducible strains of *Campylobacter jejuni*. Infect Immun. 2006;74:1368–1372.
35. Revez J, Hänninen ML. Lipooligosaccharide locus classes are associated with certain *Campylobacter jejuni* multilocus sequence types. Eur J Clin Microbiol Infect Dis. 2012;31:2203–9.
36. Hameed A, Woodacre A, Machado LR, Marsden GL. An Updated Classification System and Review of the Lipooligosaccharide Biosynthesis Gene Locus in *Campylobacter jejuni*. Front Microbiol. 2020;11:677.
37. Müller J, Meyer B, Hänel I, Hotzel H. Comparison of lipooligosaccharide biosynthesis genes of *Campylobacter jejuni* strains with varying abilities to

colonize the chicken gut and to invade Caco-2 cells. J Med Microbiol. 2007;56(12):1589–1594.

38. Godschalk PCR, Kuijf ML, Li J, St. Michael F, Ang CW, Jacobs BC, et al. Structural characterization of *Campylobacter jejuni* lipooligosaccharide outer cores associated with Guillain- Barré Syndrome and Miller Fisher syndromes. Infect Immun. 2007;75(3):1245–54.
39. Linton D, Gilbert M, Hitchen PG, Dell A, Morris HR, Wakarchuk WW, et al. Phase variation of a β-1,3 galactosyltransferase involved in generation of the ganglioside GM1-like lipo-oligosaccharide of *Campylobacter jejuni*. Mol Microbiol. 2000;37(3):501–14.
40. Godschalk PCR, Heikema AP, Gilbert M, Komagamine T, Wim Ang C, Glerum J, et al. The crucial role of *Campylobacter jejuni* genes in anti-ganglioside antibody induction in Guillain- Barré Syndrome. J Clin Invest. 2004;114(11):1659–65.
41. Mortensen NP, Kuijf ML, Ang CW, Schiellerup P, Krogfelt KA, Jacobs BC, et al. Sialylation of *Campylobacter jejuni* lipo-oligosaccharides is associated with severe gastro-enteritis and reactive arthritis. Microbes Infect. 2009;11(12):988–94.
42. Houliston RS, Vinogradov E, Dzieciatkowska M, Li J, St Michael F, Karwaski MF, et al. Lipooligosaccharide of *Campylobacter jejuni*: Similarity with multiple types of mammalian glycans beyond gangliosides. J Biol Chem. 2011;286(14):12361–70.
43. Hameed A. Human Immunity against *Campylobacter* Infection. 2019;19(6):1–17.

687 44. Nachamkin I, Allos BM, Ho T. *Campylobacter* species and Guillain-Barré
688 syndrome. Clin Microbiol Rev. 1998;11(3):555–67.
689 45. Wachira VK, Peixoto HM, and de Oliveira MRF. Systematic review of factors
690 associated with the development of Guillian-Barré syndrome 2007- 2017: what
691 has changed? Trop. Med. Int. Health 24. 2019;24(2):132–142.
692 46. Nachamkin I, Liu J, Li M, Ung H, Moran AP, Prendergast MM, et al.
693 *Campylobacter jejuni* from patients with Guillain-Barré syndrome preferentially
694 expresses a GD1a-like epitope. Infect Immun. 2002;70(9):5299–303.
695 47. Islam Z, van Belkum A, Wagenaar JA, Cody AJ, de Boer AG, Sarker SK, et al.
696 Comparative population structure analysis of *Campylobacter jejuni* from human
697 and poultry origin in Bangladesh. Eur J Clin Microbiol Infect Dis.
698 2014;33(12):2173–81.
699 48. Van Belkum A, Jacobs B, Van Beek E, Louwen R, Van Rijs W, Debruyne L, et
700 al. Can *Campylobacter coli* induce Guillain-Barré syndrome? Eur J Clin
701 Microbiol Infect Dis. 2009;28(5):557–60.
702 49. Kolehmainen A, Rossi M, Stupak J, Li J, Gilbert M, Wakarchuk W. Genetics
703 behind the biosynthesis of nonulosonic acid containing lipooligosaccharides in
704 *Campylobacter coli*. J Bacteriol, 2019;201(8):e00759-18.
705 50. Culebro A, Machado MP, Carriço JA, Rossi M. Origin, evolution , and
706 distribution of the molecular machinery for biosynthesis of sialylated
707 lipooligosaccharide structures in *Campylobacter coli*. Sci Rep, 2018;8:3028.
708 51. Quiñones B, Parker CT, Janda JM, Miller WG, Mandrell RE. Detection and
709 genotyping of Arcobacter and *Campylobacter* isolates from retail chicken
710 samples by use of DNA oligonucleotide arrays. Appl Environ Microbiol.
711 2007;73(11):3645–3655.

52. Habib I, Louwen R, Uyttendaele M, Houf K, Vandenberg O, Nieuwenhuis EE, et al. Correlation between genotypic diversity, lipooligosaccharide gene locus class variation, and caco-2 cell invasion potential of *Campylobacter jejuni* isolates from chicken meat and humans: Contribution to virulotyping. Appl Environ Microbiol. 2009;75(13):4277–88.
53. Islam Z, Sarker SK, Jahan I, Farzana KS, Ahmed D, Faruque ASG, et al. Capsular genotype and lipooligosaccharide locus class distribution in *Campylobacter jejuni* from young children with diarrhea and asymptomatic carriers in Bangladesh. Eur J Clin Microbiol Infect Dis. European Journal of Clinical Microbiology & Infectious Diseases; 2018;37(4):723–8.
54. Ellström P, Feodoroff B, Hänninen ML, Rautelin H. Characterization of clinical *Campylobacter jejuni* isolates with special emphasis on lipooligosaccharide locus class, putative virulence factors and host response. Int J Med Microbiol. Elsevier GmbH.; 2013;303(3):134–9.
55. Ellström P, Hansson I, Nilsson A, Rautelin H, Olsson Engvall E. Lipooligosaccharide locus classes and putative virulence genes among chicken and human *Campylobacter jejuni* isolates. BMC Microbiol. BMC Microbiology; 2016;16(1):1–6.
56. Islam Z, van Belkum A, Wagenaar JA, Cody AJ, de Boer AG, Tabor H, et al. Comparative genotyping of *Campylobacter jejuni* strains from patients with Guillain-Barré syndrome. PLoS One. 2009;4(9):1–8.
57. Jiang H, Zhang MJ, Liu RC, Tian XY, Gu YX, Zhang JZ. Characteristics of lipo-oligosaccharide loci of *Campylobacter jejuni* isolates associated with Guillain-Barré syndrome from Hebei, China. Int J Mol Sci. 2010;11(3):1155–61.

58. Guerry P, Szymanski CM, Prendergast MM, Hickey TE, Ewing CP, Pattarini DL, et al. Phase Variation of *Campylobacter jejuni* 81-176 Lipooligosaccharide Affects Ganglioside Mimicry and Invasiveness In Vitro. Infect Immun. 2002;70:787–93.
59. St Michael F, Szymanski CM, Li J, Chan KH, Khieu NH, Larocque S, et al. The structures of the lipooligosaccharide and capsule polysaccharide of *Campylobacter jejuni* genome sequenced strain NCTC 11168. Eur J Biochem. 2002;269(21):5119-5136.
60. Poly F, Read TD, Chen YH, Monteiro MA, Serichantalergs O, Pootong P, et al. Characterization of two *Campylobacter jejuni* strains for use in volunteer experimental-infection studies. Infect Immun. 2008;76(12):5655–67.
61. Godschalk PCR, Gilbert M, Jacobs BC, Kramers T, Tio-Gillen AP, Ang CW, et al. Co-infection with two different *Campylobacter jejuni* strains in a patient with the Guillain-Barré syndrome. Microbes Infect. 2006;8(1):248–53.
62. Skarp-de Haan CPA, Culebro A, Schott T, Revez J, Schweda EKH, Hänninen ML, et al. Comparative genomics of unintrogressed *Campylobacter coli* clades 2 and 3. BMC Genomics. 2014;15(1).
63. Funakoshi K, Koga M, Takahashi M, Hirata K, Yuki N. *Campylobacter coli* enteritis and Guillain–Barré syndrome: No evidence of molecular mimicry and serological relationship. J Neurol Sci. 2006;1(2):163–168.
64. Culebro A, Revez J, Pascoe B, Friedmann Y, Hitchings MD, Stupak J, et al. Large sequence diversity within the biosynthesis locus and common biochemical features of *Campylobacter coli* lipooligosaccharides. J Bacteriol. 2016;198(20):2829–40.

65. Pertsemlidis A, Fondon JW. Having a BLAST with bioinformatics (and avoiding BLASTphemy). Genome Biol.2001;2(10):1–10.

Data availability statement

All datasets generated or analysed during this study are available in the PURE Digital Repository https://pure.northampton.ac.uk/ at doi:10.24339/4685036f-6122-4e76-8b70-6110bb0ba382 and manuscript related Supplementary information files.

Supporting information

S1 Table. LOS types of *C. jejuni* complete (n=125) and draft sequences (n=578)

S2 Table. Primers for the identification of *C. jejuni* LOS locus classes

S3 Table. LOS types of *C. coli* complete (n=22) and draft sequences (n=542)

S4 Table. Summary of *C. coli* reference strains used to define LOS classes (Richards et al., 2013) in this study

S5 Table. Summary of *C. jejuni* LOS locus typing and PCR products' sequencing results