1 Title: Large sequence diversity within biosynthesis locus and common biochemical features of
2 *Campylobacter coli* lipooligosaccharides

3 Running title: *Campylobacter coli* LOS

4 **Alejandra Culebro[1]*, Joana Revez[1], Ben Pascoe[2,4], Yasmin Friedmann[2], Matthew D Hitchings[2],**
5 **Jacek Stupak[3], Samuel K Sheppard[2,4], Jianjun Li[3], Mirko Rossi[1]**

6 [1]Department of Food Hygiene and Environmental Health, Faculty of Veterinary Medicine, University
7 of Helsinki, Agnes Sjöbergin katu 2, Helsinki, FI-00014, Finland.

8 [2]College of Medicine, Institute of Life Science, Swansea University, Singleton Park, Swansea SA2
9 8PP, United Kingdom.

10 [3]Institute for Biological Sciences, National Research Council, Ottawa, Ontario K1A 0R6, Canada.

11 [4]Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom

12 * Corresponding author: e-mail: aleja.culebro@helsinki.fi; Phone: +3580294157428; Mobile:
13 +3580504486622

14 Keywords: *Campylobacter*; LOS; glycoconjugate; Qui3*p*NAcyl;

15 Word count: Abstract: 227; Main text (excluding table and figure legends, references): 4907.

16 Depositories (where applicable): The whole genome sequences of *C. coli* are publicly available on the
17 RAST server (http://rast.nmpdr.org) with guest account (login and password 'guest') under IDs: 195.91,
18 195.96-195.119, 195.124-195.126, 195.128-195.130, 195.133, 195.134, 6666666.94320

19 Abbreviations LOS, lipooligosaccharides; RAST, Rapid Annotation using Subsystem Technology;
20 GOs, groups of orthologues; EA-OTLC-MS, electrophoresis-assisted open-tubular liquid
21 chromatography-electrospray mass spectrometry; ESI, electrospray ionization; oligosaccharide (OS);
22 GlcN, 2-amino-2-deoxy-D-glucose; GlcN3N, β-1'-6 linked 3-diamino-2, 3-dideoxy-D-glucopyranose;
23 PEtn, phosphoethanolamine; Hep, L-glycero-D-manno-heptose; Kdo, 3-deoxy-D-manno-octulosonic
24 residue; Qui*p*3NAcyl, 3-acylamino-3,6-dideoxy-D-glucose; HexNac, hexosamine; deoxyHex,
25 deoxyhexose; Hex, hexose; Qui*p*3NAc, 3-acetamido-3,6-dideoxy-D-glucose; LPS, lipopolysaccharide.

26 **ABSTRACT**

27 Despite the importance of lipooligosaccharides (LOS) in the pathogenicity of campylobacteriosis, little
28 is known about the genetic and phenotypic diversity of LOS in *C. coli*. In this study, we investigated
29 the distribution of LOS locus classes among a large collection of unrelated *C. coli* isolates sampled
30 from several different host species. Furthermore, we paired *C. coli* genomic information and LOS
31 chemical composition for the first time to investigate possible associations between LOS locus classes
32 sequence diversity and biochemical heterogeneity. After identifying three new LOS locus classes, only
33 85% of the 144 isolates tested were assigned to a class, suggesting higher genetic diversity than
34 previously thought. This genetic diversity is at the basis of a completely unexplored LOS structure
35 heterogeneity. Mass spectrometry analysis of the LOS of nine isolates, representing four different LOS
36 classes, identified two features distinguishing *C. coli* LOS from *C. jejuni*'s. GlcN-GlcN disaccharides
37 were present in the lipid A backbone in contrast to the GlcN3N-GlcN backbone observed in *C. jejuni*.
38 Moreover, despite that many of the genes putatively involved in Qui3*p*NAcyl were apparently absent
39 from the genomes of various isolates, this rare sugar was found in the outer core of all *C. coli*.
40 Therefore, regardless the high genetic diversity of LOS biosynthesis locus in *C. coli*, we identified
41 species-specific phenotypic features of *C. coli* LOS which might explain differences between *C. jejuni*
42 and *C. coli* in terms of population dynamics and host adaptation.

43

44 **IMPORTANCE**

45 Despite the importance of *C. coli* to human health and its controversial role as a causative agent of the
46 Guillain–Barré syndrome, little is known about the genetic and phenotypic diversity of *C. coli* LOS.
47 Therefore, we paired *C. coli* genomic information and LOS chemical composition for the first time to
48 address this paucity of information. We identified two species-specific phenotypic features of *C. coli*
49 LOS, which might contribute to elucidating the reasons behind the differences between *C. jejuni* and *C.*
50 *coli* in terms of population dynamics and host adaptation.

51 **INTRODUCTION**

52 Campylobacteriosis is the most common bacterial food-borne disease in developed countries, with over
53 200,000 human cases reported annually in the European Union alone (1). The true burden of the
54 disease in the population is likely underestimated, as many infections result in mild gastroenteritis (1).
55 Approximately ~80% of reported infections are caused by *Campylobacter jejuni* and 7-18% of cases
56 are attributed to *C. coli*. Therefore, *C. coli* is among the five most important bacterial aetiological
57 agents of human gastroenteritis (2, 3).

58

59 As in other Gram-negative bacteria, *Campylobacter* spp. cell surface glycoconjugates, including
60 lipooligosaccharides (LOS), play an important role in serum and bile resistance, resistance to
61 phagocytic killing, adhesion, invasion, and survival in host cells (4-8). Current knowledge on LOS
62 diversity has been based primarily on work in *C. jejuni* and its role in promoting severe clinical
63 symptoms (9-12). *C. jejuni* LOS is a potent TLR4 agonist and the subsequent immune response is
64 affected by changes in LOS structure and composition (10-14). Additionally, due to molecular mimicry
65 between human gangliosides and certain LOS structures, *C. jejuni* has been identified as one of the
66 causative agents of the Guillain–Barré syndrome (GBS) (15). Contrarily, the little knowledge on *C. coli*
67 LOS variability has limited our understanding of the pathogenesis of GBS in patients infected with *C.*
68 *coli*, as it remains unclear whether *C. coli* is able to mimic human ganglioside structures (16-18).
69 Valuable insights into the genetic origins of significant strain variable traits have been gained by
70 studying the effect that *C. jejuni* LOS genotypes have on phenotype (19-24). However, so far, only two
71 studies have addressed the variation in gene composition in *C. coli* LOS biosynthesis locus. Until now,
72 nine genetic classes composed of a variable combination of 10 to 20 genes have been described in *C.*
73 *coli* (25, 26), but no chemical analysis of their LOS structures was executed. A couple of decades ago
74 the LOS structure of a single *C. coli* strain was described (27). Additionally, three other studies have

75 explored the chemical composition of *C. coli* LOS in a few strains (28-30), but no genetic information
76 of the strains is available.
77
78 In this study, we investigated the diversity and distribution of LOS locus classes among a large
79 collection of unrelated *C. coli* isolates sampled from several different host species. We expanded the
80 current *C. coli* LOS classification by describing three additional LOS locus classes (25, 26). Moreover,
81 by analysing genomic data with the LOS chemical composition of selected isolates, we identified
82 possible associations between gene content in the LOS biosynthesis locus and observed differences in
83 LOS phenotype. Despite the extensive introgression between *C. coli* and *C. jejuni* (31, 32), only
84 negligible levels of recombination were detected in LOS biosynthesis genes, which might explain the
85 distinctive species-specific chemical features observed herein.

86 **METHODS**

87 **Bacterial isolates, cultivation, and DNA extraction.** In total, 144 *C. coli* isolates, including 90
88 isolated from swine, 34 from humans, 18 from poultry, and two from wild birds, were chosen for LOS
89 locus screening. The selection comprised 133 *C. coli* isolates from previous studies collected between
90 1996 and 2012 from Finnish human patients, chicken and pigs reared in Finland, and wild birds
91 sampled in Helsinki region (25, 33-39). This collection was supplemented with 11 *C. coli* isolates from
92 the Campynet (CNET) collection (hosted by DSMZ GmbH, https://www.dsmz.de/). Isolate selection
93 was based on genotype (PFGE, MLST), host, country of origin, and year of isolation to encompass the
94 largest possible diversity. Cultivation and DNA isolation were carried out as previously described (25),
95 unless otherwise stated.
96
97 **PCR.** The length of LOS biosynthesis loci was determined by amplifying the region between
98 orthologue 10 (LOS biosynthesis glycosyltransferase, *waaV*) and orthologue 16 (uncharacterized
99 glycosyltransferase) (ID numbers according to Richards and colleagues (26)). PCR reactions were set

100 up as follows: 25 µl reactions containing 0.5 U Phusion high-fidelity (Thermo Scientific), 200 µM of
101 each dNTP (Thermo Scientific), 0.4 µM of each primer (ORF3F2 and waaV; Table 1), 1 X Phusion GC
102 buffer (Thermo Scientific), 700 µM of $MgCl_2$ (Thermo Scientific), and 50 ng of template. Cycling
103 conditions were as follows: one cycle at 98 ºC for 30 s followed by 30 cycles of denaturation at 98 ºC
104 for 10 s, annealing at 62.4 ºC for 30 s, extension at 72 ºC for 6 min, and a final elongation at 72 ºC for 6
105 minutes. The size of the LOS locus was estimated by gel electrophoresis with 1 kb-plus (Thermo
106 Scientific) and long-range (Thermo Scientific) molecular weight markers. Specific primers for each
107 class, based on the previously described *C. coli* LOS locus classes (I to IX), were designed (25, 26).
108 Primer pairs and their amplicon size for each LOS class are shown in Table 1, and a graphic
109 representation of the primers annealing positions within the LOS locus is shown in Supplementary
110 Figure 2. Since global alignment using progressiveMauve (40) revealed that LOS locus class IV and V
111 (26) differ by only 3 single nucleotide polymorphism (which resulted in fragmentation of orthologue
112 1959 in class V), hereafter the two LOS locus classes are considered as a single class named IV/V. The
113 specificity of each primer pair was verified *in silico*. All primers were designed on specific features
114 characterizing each LOS locus class using, when possible, multiple sequence alignments of
115 homologous sequences to improve sensibility and specificity. A preliminary gradient PCR was
116 performed for each primer pair to select the most stringent conditions to minimize artefacts.
117 Additionally, same results were obtained when primers of PCR-2 to -12 were tested on both genomic
118 DNA or as a nested PCR using PCR-1 as template. PCRs were carried out in a semi-high-throughput
119 manner, thus isolates were classified into a LOS class based on the results of all PCRs (Table 1).
120 Isolates with unexpected LOS size, negative to all tested orthologues, or with unexpected combinations
121 of orthologues, were classified as untypable.
122
123 **Genome sequencing and annotation.** For ascertaining the LOS locus classes, 35 isolates were chosen
124 for genome sequencing (Supplementary Table 1) using either HiSeq or MiSeq. For HiSeq, NGS library

125 preparation, enrichment, sequencing, and sequence analyses were performed by the Institute for
126 Molecular Medicine Finland (FIMM Technology Centre, University of Helsinki, Finland). MiSeq
127 sequencing was performed by Institute of Life Science, Swansea University (Swansea, United
128 Kingdom). Reads were filtered and assembled using SPAdes Assembler v. 3.3.0 (41). Primary
129 annotation of all the genomes was performed using Rapid Annotation using Subsystem Technology
130 (RAST) (42). Sequences were manually curated using Artemis (43) and LOS locus classes were
131 aligned and compared with ACT (44). The whole genome sequences of *C. coli* are publicly available
132 on the RAST server (http://rast.nmpdr.org) with guest account (login and password 'guest') under IDs:
133 195.91, 195.96-195.119, 195.124-195.126, 195.128-195.130, 195.133, 195.134, 6666666.94320
134
135 **Orthologue clustering and phylogenetic analysis.** A database including all the translated coding
136 sequences of *C. jejuni* and *C. coli* LOS biosynthesis was assembled using Richards and colleagues (26)
137 orthologues nomenclature. Reciprocal all-versus-all BLASTp search was performed (threshold $E \leq 1e-$
138 10) (45) and orthologous groups were determined by orthAgogue and MCL (ignoring E-values, percent
139 match length $\geq 80\%$ and inflation value of 5 (46, 47)). The groups of orthologues (GOs) were then
140 aligned using MUSCLE and back-translated to nucleotide sequence using Translatorx perl script (48-
141 50). Maximum likelihood phylogenetic reconstruction of each GO was performed in MEGA6.06 (51)
142 using Kimura-2 as nucleotide substitution model and a discrete Gamma distribution (4 categories) to
143 model evolutionary rate differences among sites. A total of 100 bootstrap runs were performed and
144 summarized in a 95% consensus tree.
145
146 **LOS silver staining.** LOS profiles were assessed by silver staining as described earlier (52), with some
147 modifications. In brief, the absorbance of the biomass obtained from a 16 h Nutrient broth n°2 (Oxoid)
148 culture (100 rpm, microaerobic atmosphere, 37 °C) was adjusted to an OD_{600} of 0.5. Cells were
149 digested with 20 mg/ml proteinase K (Thermo Scientific), and incubated at 55 °C for 1 h followed by

150 boiling for 10 min. Samples were then diluted 1: 5 in loading buffer, and ran in 15% SDS-PAGE gels.
151 Gels were silver stained for visualization (53).
152
153 **CE-MS and EA-OTLC-MS analyses.** Biomass was produced in broth as indicated above and LOS
154 was prepared with the rapid method applying microwave irradiation as previously described (54). In
155 short, the lyophilized biomass was suspended in 50 µl of 20 mM ammonium acetate buffer (pH 7.5)
156 containing DNase (100 µg/ml) and RNase (200 µg/ml) and heated by direct microwave irradiation.
157 Proteinase K was then added to a final concentration of 60 µg/ml and heated under the same conditions.
158 Solutions were allowed to cool at room temperature and subsequently dried using a Speed Vac
159 (vacuum centrifuge concentrator; Savant). LOS samples were washed three times with methanol (100
160 µl) with vigorous stirring. Insoluble residues were collected by centrifugation and resuspended in 30 µl
161 water for electrophoresis-assisted open-tubular liquid chromatography-electrospray MS (EA-OTLC-
162 MS) analysis. A sheath solution (isopropanol-methanol, 2:1) was delivered at a flow rate of 1.0
163 µL/minute. Separation was performed using 30 mM morpholine in deionized water, pH 9.0. A
164 separation voltage of 20 kV, together with a pressure of 500 mbar, was applied for the EA-OTLC-MS
165 analysis. The electrospray ionization (ESI) voltage applied on the sprayer was set at −5.2 kV. Data
166 acquisition was performed for an m/z range of 600 to 2000 at a 2s/spectrum scan rate.
167
168 **Statistical analysis.** Fisher's exact test was used to assess host-LOS locus class association. P values
169 equal to or less than 0.05 were considered significant.

170 **RESULTS**

171 **PCR typing method for *C. coli* LOS locus diversity.** We explored the genetic diversity of the LOS
172 biosynthesis loci in 144 *C. coli* isolates (Supplementary table 1) using a PCR typing scheme based on
173 published LOS locus class definitions (25, 26). Isolates were classified into putative LOS locus classes
174 according to their PCR-profile and LOS locus size as described in Table 1. The LOS PCR typing

scheme was validated by genome sequencing of 35 isolates (isolates marked in yellow in Supplementary table 1). Typing results are summarised in Table 2. We were able to classify 68% of the isolates into one of the nine previously published LOS locus classes (25, 26). Most of the isolates were assigned to LOS locus class II (17%) with the remaining isolates assigned to LOS classes IV/V (15%), III (13%), VI (13%), VIII (7%), I (2%), VII (1%), and IX (0.7%). The final 46 (out of 144, ~32%) isolates remained untypable by this method.

Six untypable isolates, with a LOS locus length of ~11.5 kbp, were sequenced (45, 63, 114, 125, 149, and 153). All isolates belong to a novel LOS locus class X. This new class shares 12 (out of 15) orthologues with other LOS locus classes (see below), and is characterised by the presence of three unique genes (Supplemental Fig. 2). A blastp search of the NCBI database, revealed sequence similarity with: (i) hypothetical protein of *Helicobacter* sp. MIT 05-5293 (e-value $1e^{-98}$; identity 45%); (ii) hypothetical protein of *Helicobacter hepaticus* (e-value $3e^{-108}$; identity 53%); (iii) UDP-N-acetylglucosamine 2-epimerase of *H. hepaticus* (e-value $3e^{-165}$; identity 63%). Following this finding, primers were designed (Table 1) for LOS locus class X which further identified 15% of the isolates (Table 2). The genomes of isolates 138 and 99, which have a similar LOS size to class X but a different PCR profile (Supplementary Table 1) were also sequenced. Analysis of these genomes revealed two additional LOS locus classes, defined as XI (isolate 138) and XII (isolate 99). In total, we were able to assign a LOS locus class to 85% of the isolates in our collection by incorporating these additional classes. LOS profile diversity was high, suggesting that further LOS locus classes may be described in the future.

Origin of the novel LOS locus classes X, XI, and XII. As in *C. jejuni, C. coli* exhibits a mosaic LOS loci (22) with several classes containing similar orthologous loci. LOS locus classes X and XI are very similar to each other, diverging only at a single locus (1967 vs 1920; Fig. 1). Additionally, these two

200 classes also have similarity in gene content and organisation to LOS locus classes I, III, IV/V, VI, and
201 VII (Fig. 1). To infer evolutionary relationships between these classes, phylogenetic analyses were
202 performed for each shared GOs. Phylogenetic reconstruction revealed LOS class I and LOS class III as
203 the two possible origins for the region encompassing orthologue 16 to orthologue 1668 in LOS locus
204 class X (Fig. 1). Specifically, in the phylogenetic tree of orthologues 16, 1850, and 1668, *C. coli*
205 isolates 45, 63, and 114 are monophyletic with strains from LOS locus class III, while *C. coli* isolates
206 125 and 149 formed a separate clade with LOS locus class I strains (Supplemental Fig. 1A, B, and C).
207 Orthologues 8 and 1821 in LOS class X and both IV/V and VI share the same origin. Contrarily, the
208 origin of the region including orthologues 1967, 1742, and 1743 is less clear. In the phylogenetic tree
209 of orthologue 1967 (Supplemental Fig. 1D), *C. coli* isolates 63 and 114 are grouped with LOS locus
210 class VI isolates, while the other strains form a separate clades. In addition, the star-like phylogeny
211 inferred for orthologues 1742 and 1743 hampered any kind of conclusion. These results suggest that
212 extensive recombination and gene reorganisation between LOS locus classes took place, masking the
213 origin of common shared loci. Excepting for orthologue 1920, LOS locus class XI orthologues are
214 closely related to those found in LOS locus class X (Supplemental Fig. 1). LOS locus class XII shares
215 orthologues with LOS locus classes I, IV/V, VII, and IX. Yet, in our phylogenetic analysis LOS locus
216 class XII orthologues are distantly related to those found in other LOS classes, forming a separate
217 branch in the phylogenetic trees. Additionally, LOS locus class XII is characterized by the presence of
218 a set of unique genes having the best BLASTp hit against NCBI nr with: (i) methyltransferase type 12
219 of *H. hepaticus* (e-value $6e^{-75}$; identity 58%); (ii) hypothetical protein of *Anaerovibrio lipolyticus* (e-
220 value $5e^{-102}$; identity 65%); (iii) phosphoserine phosphatase of *Helicobacter* sp. MIT 05-5293 (e-value
221 $3e^{-92}$; identity 63%) (Fig. 1). Proposed functions for each ORF of the herein newly identified LOS locus
222 classes are described in Supplemental Table 2.
223

224 **Cluster analysis of the LOS locus classes.** Both species share a total of 19 LOS orthologues (26) and
225 with previous evidence of introgression between *C. coli* and *C. jejuni* in mind (31, 32) we attempted to
226 quantify the level of interspecies recombination in *C. coli* LOS diversity. We compared individual gene
227 descriptions of the LOS loci rather than the original gene family ontologies used by Richards and
228 colleagues (26). Out of the 19 shared orthologues, 16 gene locus descriptions split into species-specific
229 clusters while only three were common in both species (orthologues 10, 16 and 1821). Interspecies
230 gene transfer was investigated by comparing the topology of individual gene trees with the overall
231 population structure (25). Evidence of interspecies gene transfer was only observed for orthologue 10
232 (26) (lipooligosaccharide biosynthesis glycosyltransferase, *waaV*) where all *C. coli* loci of LOS locus
233 class II formed a monophyletic clade with *C. jejuni* genes (Fig. 2). Thus, interspecies recombination is
234 likely to have a limited effect on the LOS loci diversity observed in *C. coli*.
235
236 **Host-LOS locus class association.** The non-random distribution of LOS locus classes between hosts
237 was investigated further by supplementing our isolate collection with Richards and colleagues data
238 (26). The distribution of LOS locus classes by source of isolation is represented in Figure 3. All LOS
239 locus classes, except class XII, were present among strains isolated from humans. More than half
240 (57%) of the clinical isolates were LOS locus classes II, III, and VIII, while LOS locus classes VI, VII,
241 and X were less commonly found in clinical cases. Most pig isolates were of LOS locus class X, but
242 also frequently found among LOS locus classes II, III, IV/V, and VI. Only one pig isolate belonged to
243 LOS locus class VIII and no pig strain was from classes I, IX, or XII. Poultry isolates were also found
244 among all LOS locus classes, except for classes VII, IX, and XII. Most poultry isolates were classified
245 as LOS locus class II.
246 There was a positive association ($p < 0.05$) of class VIII to human clinical infections, while class VI
247 was negatively associated with clinical cases. Swine was positively associated with classes VI and X,
248 but negatively associated with classes I and VIII. Poultry was positively associated only with LOS

249 locus class I. Bovine and wild-bird isolates were underrepresented in the dataset. However, some
250 association was observed in bovine (class IV/V) and wild bird isolates (class XII). Isolates classified as
251 LOS locus classes II and III were equally distributed among humans, pigs, and poultry.
252
253 **Chemical analysis of *C. coli* LOS composition.** The LOS phenotype of nine selected isolates was
254 investigated. This selection included strains from classes overrepresented in clinical isolates, II and
255 VIII, as well as isolates from two of the newly described LOS classes (X and XI) and which are
256 uncommon in clinical isolates. Silver staining SDS-PAGE gels of LOS extracts provided migration
257 profiles for the selected isolates (Fig. 4A). A complimentary mass spectroscopy approach was used
258 (CE-MS and EA-OTLC-MS) to explore inter- and intra-LOS class structural diversity. Example spectra
259 is shown in Supplemental Fig. 3. The oligosaccharide (OS) composition of each of the nine isolates
260 was predicted based on the fragment ions and components of the previously reported *C. coli* OS (27).
261 Size and composition of the lipid A group was defined for each glycoform by tandem mass
262 spectrometry. For example, the fragment ion at m/z 1063.2 (doubly charged ion) in *C. coli* 137
263 (Supplemental Fig. 3), which was produced from the glycoform detected as triply charged ion at m/z
264 1422.8, corresponds to a lipid A with a 2-amino-2-deoxy-D-glucose (GlcN) disaccharide backbone
265 carrying negative charged groups, PPEtn and PPEtn, substituted by six fatty acid chains and with a
266 calculated mass of ~2128 Da. Additionally, the fragment ion at m/z 1001.7 corresponds to a second
267 lower mass lipid A species (~2006 Da) as it carries P and PPEtn instead. All analyzed *C. coli* isolates
268 exhibited a hexa-acylated lipid A containing four tetradecanoic (14:0) and two hexadecanoic (16:0)
269 acid chains, modified with two phosphate residues (55-57). Only GlcN disaccharides were detected in
270 *C. coli* isolates, in contrast to the hybrid backbone of β-1'-6 linked 3-diamino-2, 3-dideoxy-D-
271 glucopyranose (GlcN3N) and GlcN observed in *C. jejuni* (55, 57). Thus, *C. coli* synthesizes a lipid A
272 with two ester- and two amide-linked acyl chains, while *C. jejuni* has a lipid A containing mainly three
273 amide-linked acyl chains and one ester-linked acyl chain. The lower mass lipid A was detected in all

274 samples, while LOS locus class II isolates (except for isolate 65, Supplemental Fig. 3) had an additional
275 lipid A species as exemplify by strain 137 in the Supplemental Fig. 3.
276 Like in *C. jejuni, C. coli* exhibited a conserved inner core consisting of two L-glycero-D-manno-
277 heptose (Hep) residues attached to a 3-deoxy-D-manno-octulosonic residue (Kdo) which is linked to
278 the lipid A through a Kdo linker (20, 57). In the variable outer core region at least one predicted
279 Qui*p*3NAcyl residue (where Qui*p*3NAc represents 3-acylamino-3,6-dideoxy-D-glucose in which the
280 N-acyl residue was a 3-hydroxybutanoyl) was detected in all isolates. Although more than one OS was
281 detected by MS in all isolates (Fig. 4B), only isolates from LOS locus classes X and XI exhibited
282 visible high-M_r and low-M_r LOS on SDS-PAGE (Fig. 4A). Intra-LOS class diversity was observed in
283 both LOS class II and class X. Isolate 65 displayed a LOS composition like other LOS class II isolates
284 but with the addition of two hexosamines (HexNac) and one deoxyhexose (deoxyHex), and absence of
285 PEtn residues (Fig. 4B). Likewise, isolates 45 and 63 shared similar LOS composition, with the
286 exception of a variable Qui*p*3NAcyl residue in isolate 63. In contrast, isolate 114 exhibited a very
287 different LOS composition compared with other isolates of the same class, including the presence of a
288 third Hep and a deoxyHex as well as a reduced number of hexoses (Hex) (Fig. 4B). The LOS of
289 isolates 38, 45, and 138 have similar core size and proposed composition, yet they are classified into
290 three different LOS locus classes. However, our biochemical analysis is not able to identified
291 saccharide sequence, stereochemistry, absolute configuration (D or L), anomeric configurations (α or
292 β), and linkage positions. Thus, further studies would be required to determine whether these three
293 different LOS classes indeed produce the same LOS structure.
294 **Genetic and phenotypic diversity within *C. coli* LOS class II.** The four strains with LOS locus class
295 II shared 99.64% DNA sequence similarity and from 99.39% to 99.98% pairwise alignment identity.
296 Isolate 65 was the most dissimilar among strains with LOS locus class II due to large fragments
297 deletions. Deletions resulted in shorter 2400 and 2473 orthologues, as one pseudogene (Fig. 5).
298 Orthologues 2470 and 2471 were also truncated as one pseudogene (re-annotated as 2470-1), as

299 evidenced by isolate 151. The remainder of the class II isolates had an insertion of 68 nt in 2470-1,
300 disrupting the orthologue (Fig. 5). Despite the differences observed in orthologue 2470-1 isolates 73,
301 137, and 151 were predicted to have identical LOS chemical compositions.
302
303 Amino acid sequences of orthologues 6, 1541, 1501, 2472, and 10 were identical (100%) in all four
304 class II strains, while orthologues 9004 and 16 exhibited a single amino acid difference in isolate 65.
305 All isolates, with the exception of 65, exhibited differences in the C-terminal of orthologue 1715 and
306 were variable in the number of Hep and/or PEtn residues observed. However, no GC homopolymeric
307 tracts or other possible genetic signals associated with phase variation were identified within the LOS
308 loci.
309
310 **Genetic and phenotypic diversity within *C. coli* LOS locus class X.** In LOS locus class X the overall
311 sequence identity among strains was 99.31%, with percentage identity ranging from 98.96% to 99.94%
312 in pairwise alignments, with strain 45 being the most distantly related. Although some minor gaps were
313 observed, single point mutations were largely responsible for the diversity observed at nucleotide level.
314 The largest insertion (69 nt) was seen in strain 63 between orthologues 2 and 3. Between strains, 100%
315 amino acid identity was observed in orthologues 16, 8, and 2, while one or two amino acid substitutions
316 were present in orthologues 1668, 1, 1821, 1967, and 1743. The most prominent difference was
317 observed in orthologue 1742 in the form of a deleted A base at position 668, resulting in premature
318 translational termination in isolates 114 and 63. Furthermore, several single amino acid substitutions
319 were detected in orthologue 1742 in strain 45, while 100% identity was observed between isolates 63
320 and 114. In spite of dissimilar LOS composition, the only difference observed within the LOS locus
321 between isolates 63 and 114 was in eight amino acids at the C-terminal of orthologue 3.
322
323 **DISCUSSION**

324 *Campylobacter* LOS is a fundamental feature involved in the pathogenesis of gastroenteritis and post-
325 infection sequelae (10-14, 58, 59). However, despite the burden imposed by *C. coli* and the importance
326 of this structure in campylobacteriosis, little is known about the LOS diversity in this species (26-29,
327 60). Therefore, we sought to contribute to the paucity of information by investigating the variability
328 and distribution of *C. coli* LOS locus genetic classes in a large collection of isolates and by coupling
329 genomic and LOS chemical composition data for the first time.
330 We developed a PCR methodology which was able to classify 85% of the isolates into a LOS class (25,
331 26). Among them, we described three additional LOS locus classes, named X, XI, and XII, which
332 accounted for 17% of the isolates in our collection. The remaining untypable isolates (15%) suggests
333 that further new classes will likely be described in the future and that *C. coli* LOS biosynthesis is more
334 diverse than previously observed (26).
335 This genetic diversity is at the basis of a completely unexplored LOS structure heterogeneity which
336 might contribute substantially to the population dynamics of *C. coli*, including host specificity. We
337 combined our 144 isolates with 33 *C. coli* previously studied (26) to investigate the non-random
338 distribution of LOS locus classes among different hosts. All hosts were significantly associated with at
339 least one LOS locus class. In particular, isolates possessing LOS locus classes VI and X were
340 predominantly isolated from swine, which have very high prevalence of *C. coli* (up to 99%) (61). Both
341 of these classes were rarely detected in human isolates, which is supported by a previous source
342 attribution study in Scotland in which pigs are a relatively unimportant source of *C. coli* human
343 infections (61). The majority of human cases in our study were assigned to LOS locus classes II or III,
344 which were also found in swine and poultry isolates. However, human isolates were overrepresented
345 among LOS locus class VIII, which was rarely detected in the sources included in this study. This
346 indicates the presence of other, unknown potential reservoirs contributing to human infections, which
347 corroborates with a previous study where 54% of human *C. coli* strains were attributed to other sources
348 than poultry and pig (61). In opposition to previous findings (26), we did not observe partitioning

349 between bovine and poultry sourced strains, and LOS locus classes previously shown to be associated
350 with bovine hosts were populated by isolates of poultry and swine origin. Due to the limited number of
351 isolates available from alternative sources, the host-LOS class associations found in this study may not
352 necessarily represent the true *C. coli* population structure in various hosts. However, our findings
353 suggest that generalist isolates possessing LOS locus class II and III might be more successful at
354 colonizing multiple species and, as seen in generalist lineages of *C. jejuni* ST-45 and ST-21 clonal
355 complexes, being largely responsible for human infections (32).

356

357 Mosaic *C. coli* LOS classes appear to have arisen by the insertion and/or deletion of genes or gene
358 cassettes through homologous recombination, as previously described in *C. jejuni* (22). In spite of
359 substantial genome-wide introgression between agricultural *C. coli* and *C. jejuni* (25, 31), very limited
360 interspecies recombination was detected among LOS biosynthesis loci. Only orthologue 10 (*waaV*) in
361 *C. coli* LOS locus class II may have originated as result of recombination with *C. jejuni*. These results
362 confirmed previous studies (31), and are supported by the species-specific features detected in the
363 chemical composition of *C. coli* LOS.

364

365 GlcN disaccharide backbones, which is the most common structure among members of the family
366 *Enterobacteriaceae* (57), were predicted in the lipid A of all analysed *C. coli* strains. This result is in
367 contrast to the hybrid GlcN3N-GlcN backbone observed in *C. jejuni*. The genes *gnnA* and *gnnB,*
368 located outside the LOS biosynthesis locus, are associated with the synthesis of GlcN3N-substituted
369 lipid A (9, 62). Inactivation of either of these genes in *C. jejuni* resulted in the substitution of an *N*-
370 linked with an *O*-linked acyl chain and an increased LOS biological activity in humans (9). *C. coli*
371 contains in a similar genomic location both genes, having approximately 70% BLASTp score ratios
372 against *C. jejuni* orthologues (9). Yet, *C. coli gnnA* and *gnnB* are separated by a putative cobalamin
373 independent methionine synthase II in the same gene orientation. We suggest therefore three possible

374 explanations for the absence of GlcN3N in *C. coli* lipid A backbone: (i) single or multiple mutations in
375 the putative active sites of GnnA and GnnB have rendered one or both enzymes inactive, as observed in
376 functional studies in other bacteria (62, 63); (ii) *gnnB-gnnA* operon transcription might be hampered by
377 the presence of the putative methionine synthase II (9)*;* (iii) GnnA and GnnB may be involved in the
378 biosynthesis of alternative glycoconjugates in *C. coli* (62). Nevertheless, the substitution of an *N*-linked
379 with an *O*-linked acyl chain in *C. coli* might have an impact in host-bacterial interaction and adaptation
380 (9).

381

382 A second species-specific feature, common among all our analysed isolates, was the presence of at least
383 one putative Qui*p*3NAcyl residue. Qui*p*3N is an unusual deoxysugar, which has been observed in the
384 O-antigen of various Gram negative bacteria and in the S-layer of glycoprotein glycans of some Gram
385 positives (64-66). Although rarely studied, Qui*p*3N has also been found in the OS of LOS class E, H,
386 and P isolates in *C. jejuni* exclusively as an *N*-acetyl derivative (Qui*p*3NAc) (54, 67-69). Conversely,
387 Qui*p*3N has only been reported in *C. coli* as an *N*-acyl derivative with two possible substituents; 3-
388 hydroxybutanoyl or 3-hydroxy-2, 3-dimethyl-5-oxoprolyl (30). The presence of Qui*p*3NAcyl in *C. coli*
389 was first described by Seltmann and Beer (30), and later on it was reported in several *C. coli* (28).
390 However, the molecular basis behind the biosynthesis of this sugar and associated glycoconjugate in *C.*
391 *coli* remains unknown. The dTDP-D-Qui*p*3NAc biosynthesis pathway has, to our knowledge, only
392 been described in the Gram positive *Thermoanaerobacterium thermosaccharolyticum* (70). This
393 pathway involves five enzymes; a thymydylyltransferase (RmlA), a 4, 6-dehydratase (RmlB), a 3, 4-
394 isomerase (QdtA), a transaminase (QdtB), and a transacetylase (QdtC). Genome comparison of *T.*
395 *thermosaccharolyticum* and *C. coli* identified homologs of *rmlA* (GO 1743)*, rmlB* (GO 1742), *qdtA*
396 (GOs 1920 and 1967), and *qdtB* (GO 8) in a subset of strains. However, no homologue for *qdtC* was
397 found in *C. coli*. This may be expected as *C. coli* Qui*p*3N is an *N*-acyl derivative instead of the *N*-acetyl
398 derivative found in *T. thermosaccharolyticum* (27, 30). Moreover, these results are in agreement with

399 previous studies in which *C. jejuni* isolates carrying the aforementioned orthologues in the LOS locus
400 have been found to express Qui*p*3NAc in their LOS (26, 54, 67-69). Despite the presence of this sugar
401 in all *C. coli* investigated in this study, as described above, the putative dTDP-D-Qui*p*3NAc
402 biosynthesis genes are only present in a subset of strains all belonging to LOS classes IV/V, VI, VII, X,
403 and XI (Supplemental Fig. 2). Furthermore, truncation of orthologue 1742 due to a single base deletion
404 should have resulted in the loss of Qui*p*3NAcyl in isolates 114 and 63, which was not the case. Cross
405 talk between different glycosylation pathways have been previously observed in *C. jejuni* (67, 71).
406 Thus, due to Qui*p*3NAcyl being predicted to be ubiquitously found in *C. coli* LOS structures, we
407 hypothesize that the synthesis of this residue might be carried out by genes in conserved glycosylation
408 pathways. Because of structural similarity between Qui*p*3NAc and bacillosamine precursors, it is
409 tempting to speculate that the *pgl* system may play a role in the biosynthesis of Qui*p*3NAc in *C. coli*.
410

411 In all *C. coli*, phenotypic variation was observed affecting at least one sugar residue, as strains exhibit
412 different numbers of Hep, Qui*p*3NAcyl, HexNac, or PEtn (Fig 4B). Phenotypic variation in *C. jejuni*
413 has been mainly associated with phase variation of genes containing repeats of GC homopolymeric
414 tracts (23). However, no GC tracts were detected in the LOS locus of the chemically analysed *C. coli*
415 isolates. Further inspection of all the LOS locus sequences generated in this and previous studies (25,
416 26) revealed that G-tracts are uncommon in *C. coli* LOS. Only isolates from LOS class IV/V and VI
417 had G-tracts longer than 5 bases in their LOS biosynthesis locus. It is therefore unlikely that the
418 observed phenotypic variation in our analysed samples was caused by slipped strand mispairing due to
419 homopolymeric tracts within the LOS locus. These data suggest that other mechanisms, such as post-
420 transcriptional regulation or epigenetic methylation of DNA, might be responsible for phenotypic
421 variation in LOS composition in *C. coli*.
422 Among LOS locus class II isolates, strain 65 exhibited the most divergent composition. Orthologue
423 1715 (*wlaTB*) has been associated with a HexNac residue in *C. jejuni* 81116 (67) and the diversity

424 observed in the C-terminal of this orthologue may be responsible for the absence of HexNAc residues
425 in isolates 73, 137, and 151. However, further research is required to confirm the exact role of 1715 in
426 LOS biosynthesis. Similarly to LOS locus class II, strains with LOS locus class X isolates minor
427 genetic dissimilarities resulted in major differences in LOS chemical composition.

428 Isolates 65 and 114 also contained a deoxyHex residue in the LOS. No orthologues potentially involved
429 in deoxyHex synthesis were identified within the LOS region in isolates 65, suggesting that genes
430 outside the LOS locus may play a bigger role in LOS biosynthesis than previously thought.
431 Deoxyhexoses, such as 6-deoxy-β-l-altrose, fucose, or rhamnose have been frequently detected in the
432 O-chain of the lipopolysaccharide (LPS) of several Gram-negative species (72, 73). Nevertheless, in
433 the genus *Campylobacter*, these sugars have been described as components of *C. jejuni* capsule (74)
434 and *C. fetus* LPS (75).

435

436 In conclusion, the genetic and biochemical diversity of *C. coli* is greater than expected. *C. coli* LOS is
437 characterised by a lipid A consisting of GlcN-GlcN disaccharides and an outer core substituted with at
438 least one Qui*p*3NAcyl residue. Our results hint at cross talk between different glycosylation pathways,
439 which has not been generally considered to play a role in LOS diversity. The relevance of these
440 characteristic features for the ecology and virulence of *C. coli* is yet to be explored.

441

442 **ACKNOWLEDGEMENTS**

443 The authors wish to thank Ann-Katrin Llarena for her comments and Marja-Liisa Hänninen for
444 providing the strains. This research project was supported by the University of Helsinki research grant
445 n. 313/51/2013. AC was supported by the Microbiology and Biotechnology graduate program (MBDP)
446 from the University of Helsinki. SS and BP were supported by the Biotechnology and Biological
447 Sciences Research Council (BBSRC) grant BB/I02464X/1, and the Medical Research Council (MRC)
448 grants 473 MR/M501608/1 and MR/L015080/1.

449 **REFERENCES**

450 1. **EFSA, and ECDC.** 2015. The European Union summary report on trends and sources of zoonoses,
451 zoonotic agents and food-borne outbreaks in 2013. EFSAJournal. **13:**3991. doi:
452 10.2903/j.efsa.2015.3991.

453 2. **Gillespie IA, O'Brien SJ, Frost AF, Adak GK, Horby P, Swan AV, Painter MJ, Neal KR.** 2002.
454 A case-case comparison of *Campylobacter coli* and *Campylobacter jejuni* infection: A tool for
455 generating hypotheses. Emerg Infect Dis. **8:**937. doi: 10.3201/eid0809.010187..

456 3. **Gürtler M, Alter T, Kasimir S, Fehlhaber K.** 2005. The importance of *Campylobacter coli* in
457 human campylobacteriosis: prevalence and genetic characterization. Epidemiol Infect. **133:**1081. doi:
458 10.1017/S0950268805004164.

459 4. **Young, KT, Davis LM, DiRita VJ.** 2007. *Campylobacter jejuni:* molecular biology and
460 pathogenesis. Nature Rev Microbiol. **5:**665-79.

461 5. **Karlyshev AV, Ketley JM, Wren BW.** 2005. The *Campylobacter jejuni* glycome. FEMS
462 Microbiol. Rev. **29:**377-390. doi: 10.1016/j.fmrre.2005.01.003.

463 6. **Iwata T, Chiku K, Amano K, Kusumoto M, Ohnishi-Kameyama M, Ono H, Akiba M.** 2013.
464 Effects of lipooligosaccharide inner core truncation on bile resistance and chick colonization by
465 *Campylobacter jejuni*. PLoS ONE. **8:**e56900.

466 7. **Javed MA, Cawthraw SA, Baig A, Li J, McNally A, Oldfield NJ, Newell DG, Manninga G.**
467 2012. Cj1136 is required for lipooligosaccharide biosynthesis, hyperinvasion, and chick colonization
468 by *Campylobacter jejuni*. Infect Immun. **80:**2361-2370.

469 8. **Naito M, Frirdich E, Fields JA, Pryjma M, Li J, Cameron A, Gilbert M, Thompson SA,**
470 **Gaynor EC.** 2010. Effects of sequential *Campylobacter jejuni* 81-176 lipooligosaccharide core
471 truncations on biofilm formation, stress survival, and pathogenesis. J Bacteriol. **192:**2182-2192.

472 9. **van Mourik AL, Steeghs L, van Laar J, Meiring HD, Hamstra HJ, van Putten JP, Wösten**
473 **MM.** 2010. Altered linkage of hydroxyacyl chains in lipid A of *Campylobacter jejuni* reduces TLR4
474 activation and antimicrobial resistance. J Biol Chem. **285:**15828-15836. doi:
475 10.1074/jbc.M110.102061.

476 10. **Stephenson HN, John CM, Naz N, Gundogdu O, Dorrell N, Wren BW, Jarvis GA, Bajaj-**
477 **Elliott M.** 2013. *Campylobacter jejuni* lipooligosaccharide sialylation, phosphorylation, and
478 amide/ester linkage modifications fine-tune human toll-like receptor 4 activation. J Biol Chem.
479 **288:**19661-19672. doi: 10.1074/jbc.M113.468298.

480 11. **Kuijf ML, Samsom JN, van Rijs W, Bax M, Huizinga R, Heikema AP, van Doorn PA, van**
481 **Belkum A, van Kooyk Y, Burgers PC, Luider TM, Endtz HP, Nieuwenhuis EE, Jacobs BC.** 2010.
482 TLR4-mediated sensing of *Campylobacter jejuni* by dendritic cells is determined by sialylation. J
483 Immunol. **185:**748-755. doi: 10.4049/jimmunol.0903014.

484 12. **Bax M, Kuijf ML, Heikema AP, van Rijs W, Bruijns SC, García-Vallejo JJ, Crocker PR,**
485 **Jacobs BC, van Vliet SJ, van Kooyk Y.** 2011. *Campylobacter jejuni* lipooligosaccharides modulate
486 dendritic cell-mediated T cell polarization in a sialic acid linkage-dependent manner. Infect Immun.
487 **79:**2681-2689. doi: 10.1128/IAI.00009-11.

488 13. **Huizinga R, van Rijs W, Bajramovic JJ, Kuijf ML, Laman JD, Samsom JN, Jacobs BC.** 2013.
489 Sialylation of *Campylobacter jejuni* endotoxin promotes dendritic cell–mediated B cell responses
490 through CD14-dependent production of IFN-ß and TNF-a. J Immunol. **191:**5636-5645. doi:
491 10.4049/jimmunol.1301536.

492 14. **Huizinga R, Easton AS, Donachie AM, Guthrie J, van Rijs W, Heikema A, Boon L, Samsom**
493 **JN, Jacobs BC, Willison HJ, Goodyear CS.** 2012. Sialylation of *Campylobacter jejuni* lipo-
494 oligosaccharides: Impact on phagocytosis and cytokine production in mice. PLoS ONE. **7:**e34416.

495 15. **Yuki N, Susuki K, Koga M, Nishimoto Y, Odaka M, Hirata K, Taguchi K, Miyatake T,**
496 **Furukawa K, Kobata T, Yamada M.** 2004. Carbohydrate mimicry between human ganglioside GM1
497 and *Campylobacter jejuni* lipooligosaccharide causes Guillain–Barré syndrome. Proc Natl Acad Sci
498 USA. **101:**11404-11409. doi: 10.1073/pnas.0402391101.

499 16. **Funakoshi K, Koga M, Takahashi M, Hirata K, Yuki N.** 2006. *Campylobacter coli* enteritis and
500 Guillain–Barré syndrome: No evidence of molecular mimicry and serological relationship. J Neurol
501 Sci. **246:**163-168.

502 17. **Bersudskya M, Rosenbergb P, Rudenskyc R, Wirguin I.** 2000. Lipopolysaccharides of a
503 *Campylobacter coli* isolate from a patient with Guillain-Barre syndrome display ganglioside mimicry.
504 Neuromuscular Disord. **10:**182-186.

505 18. **van Belkum A, Jacobs B, van Beek E, Louwen R, van Rijs W, Debruyne L, Gilbert M, Li J,**
506 **Jansz A, Mégraud F, Endtz H.** 2009. Can *Campylobacter coli* induce Guillain-Barré syndrome? Eur J
507 Clin Microbiol Infect Dis. **28:**557-560. doi: 10.1007/s10096-008-0661-9.

508 19. **Gilbert M, Karwaski M, Bernatchez S, Young NM, Taboada E, Michniewicz J, Cunningham**
509 **A, Wakarchuk WW.** 2002. The genetic bases for the variation in the lipo-oligosaccharide of the
510 mucosal pathogen, *Campylobacter jejuni*. J Biol Chem. **277:**327-337. doi: 10.1074/jbc.M108452200.

511 20. **Dzieciatkowska M, Brochu D, van Belkum A, Heikema AP, Yuki N, Houliston RS, Richards**
512 **JC, Gilbert M, Li J.** 2007. Mass spectrometric analysis of intact lipooligosaccharide: Direct evidence

513 for O-acetylated sialic acids and discovery of O-linked glycine expressed by *Campylobacter jejuni*.
514 Biochemistry. **46:**14704-14714.

515 21. **Chiu CP, Watts AG, Lairson LL, Gilbert M, Lim D, Wakarchuk WW, Withers SG,**
516 **Strynadka NC.** 2004. Structural analysis of the sialyltransferase CstII from *Campylobacter jejuni* in
517 complex with a substrate analog. Nat Struct Mol Biol. **11:**163-170.

518 22. **Parker CT, Gilbert M, Yuki N, Endtz HP, Mandrell RE.** 2008. Characterization of
519 lipooligosaccharide-biosynthetic loci of *Campylobacter jejuni* reveals new lipooligosaccharide classes:
520 evidence of mosaic organizations. J Bacteriol. **190:**5681-5689.

521 23. **Linton D, Gilbert M, Hitchen PG, Dell A, Morris HR, Wakarchuk WW, Gregson NA, Wren**
522 **BW.** 2000. Phase variation of a β-1,3 galactosyltransferase involved in generation of the ganglioside
523 GM1-like lipo-oligosaccharide of *Campylobacter jejuni*. Mol Microbiol. **37:**501-514. doi:
524 10.1046/j.1365-2958.2000.02020.x.

525 24. **Revez J, Hänninen M.** 2012. Lipooligosaccharide locus classes are associated with certain
526 *Campylobacter jejuni* multilocus sequence types. Eur J Clin Microbiol Infect Dis. **31:**2203-2209.

527 25. **Skarp-de Haan CP, Culebro A, Schott T, Revez J, Schweda E, Hänninen M, Rossi M.** 2014.
528 Comparative genomics of unintrogressed *Campylobacter coli* clades 2 and 3. BMC Genomics. **15:**129.

529 26. **Richards VP, Lefébure T, Pavinski Bitar PD, Stanhope MJ.** 2013. Comparative characterization
530 of the virulence gene clusters (lipooligosaccharide [LOS] and capsular polysaccharide [CPS]) for
531 *Campylobacter coli, Campylobacter jejuni* subsp. *jejuni* and related *Campylobacte*r species. Infect
532 Genet Evol. **14:**200-213. doi: 10.1016/j.meegid.2012.12.010.

533 27. **Aspinall GO, McDonald AG, Pang H, Kurjanczyk LA, Penner JL.** 1993. Lipopolysaccharide of
534 *Campylobacter coli* serotype O:30. Fractionation and structure of liberated core oligosaccharide. J Biol
535 Chem. **268:**6263-6268.

536 28. **Beer W, Adam M, Seltmann G.** 1986. Monosaccharide composition of lipopolysaccharides from
537 *Campylobacter jejuni* and *Campylobacter coli*. J Basic Microbiol. **26:**201-204. doi:
538 10.1002/jobm.3620260405.

539 29. **Naess V, Hofstad T.** 1984. Chemical studies of partially hydrolysed lipopolysaccharides from four
540 strains of *Campylobacter jejuni* and two strains of *Campylobacter coli*. J Gen Microbiol. **130:**2783-
541 2789.

542 30. **Seltmann G, Beer W.** 1985.
543 Vorkommen von 3-Amino-3,6-didesoxy-D-glucose in einem Lipopolysaccharid von *Campylobacter*
544 *coli* . J. Basic Microbiol. **25:**551-552.

545 31. **Sheppard SK, Didelot X, Jolley KA, Darling AE, Pascoe B, Meric G, Kelly DJ, Cody A, Colles**
546 **FM, Strachan NJ, Ogden ID, Forbes K, French NP, Carter P, Miller WG, McCarthy ND, Owen**
547 **R, Litrup E, Egholm M, Affourtit JP, Bentley SD, Parkhill J, Maiden MC, Falush D.** 2013.
548 Progressive genome-wide introgression in agricultural *Campylobacter coli*. Mol Ecol. **22:**1051-1064.
549 doi: 10.1111/mec.12162.

550 32. **Sheppard SK, Maiden MC.** 2015. The evolution of *Campylobacter jejuni* and *Campylobacter*
551 *coli*. Cold Spring Harb Perspect Biol. **7:**. doi: 10.1101/cshperspect.a018119.

552 33. **Olkkola SH, Juntunen P, Heiska H, Hyytiäinen H, Hänninen M.** 2010. Mutations in the *rpsL*
553 gene are involved in streptomycin resistance in *Campylobacter coli*. Microb Drug Resist. **16:**105. doi:
554 1089/mdr.2009.0128.

555 34. **Juntunen P, Olkkola S, Hänninen M.** 2011. Longitudinal on-farm study of the development of
556 antimicrobial resistance in *Campylobacter coli* from pigs before and after danofloxacin and tylosin
557 treatments. Vet Microbiol. **150:**322-330.

558 35. **Juntunen P, Heiska H, Olkkola S, Myllyniemi AL, Hänninen ML.** 2010. Antimicrobial
559 resistance in *Campylobacter coli* selected by tylosin treatment at a pig farm. Vet. Microbiol. **146:**90-97.

560 36. **Lehtopolku M, Kotilainen P, Haanperä-Heikkinen M, Nakari U, Hänninen M, Huovinen P,**
561 **Siitonen A, Eerola E, Jalava J, Hakanen AJ.** 2011. Ribosomal mutations as the main cause of
562 macrolide resistance in *Campylobacter jejuni* and *Campylobacter coli*. Antimicrob Agents Chemother.
563 **55:**5939-5941. doi: 10.1128/AAC.00314-11.

564 37. **Kärenlampi R, Rautelin H, Schönberg-Norio D, Paulin L, Hänninen M.** 2007. Longitudinal
565 study of Finnish *Campylobacter jejuni and C. coli* isolates from humans, using multilocus sequence
566 typing, including comparison with epidemiological data and isolates from poultry and cattle. Appl
567 Environ Microbiol. **73:**148-155. doi: 10.1128/AEM.01488-06.

568 38. **Llarena AK, Skarp-de Haan CP, Rossi M, Hänninen M.** 2015. Characterization of the
569 *Campylobacter jejuni* population in the barnacle geese reservoir. Zoonoses Public Health. **62:**209-221.
570 doi: 10.1111/zph.12141.

571 39. **Hänninen M, Pajarre S, Klossner M, Rautelin H.** 1998. Typing of human *Campylobacter*
572 *jejuni* isolates in Finland by pulsed-field gel electrophoresis. J Clin Microbiol. **36:**1787-1789.

573 40. **Darling AE, Mau B, Perna NT.** 2010. progressiveMauve: multiple genome alignment with gene
574 gain, loss and rearrangement. PLoS One. **5:**e11147.

575 41. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin A, Kulikov AS, Lesin VM,**
576 **Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev**

577 **MA, Pevzner PA.** 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell
578 sequencing. J Comput Biol. **19:**455-477. doi: 10.1089/cmb.2012.0021.

579 42. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass
580 EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK,
581 Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V,
582 Wilke A, Zagnitko O.** 2008. The RAST Server: Rapid Annotations using Subsystems Technology.
583 BMC Genomics. **9:**1-15. doi: 10.1186/1471-2164-9-75.

584 43. **Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M, Barrell B.** 2000.
585 Artemis: sequence visualization and annotation. Bioinformatics. **16:**944-945. doi:
586 10.1093/bioinformatics/16.10.944.

587 44. **Carver T, Berriman M, Tivey A, Patel C, Böhme U, Barrell BG, Parkhill J, Rajandream MA.**
588 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database.
589 Bioinformatics. **24:**2672-2676. doi: 10.1093/bioinformatics/btn529.

590 45. **Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997.
591 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic
592 Acids Res. **25:**3389-3402. doi: 10.1093/nar/25.17.3389.

593 46. **Enright AJ, Van Dongen S, Ouzounis CA.** 2002. An efficient algorithm for large-scale detection
594 of protein families. Nucleic Acids Res. **30:**1575-1584. doi: 10.1093/nar/30.7.1575.

595 47. **Ekseth OK, Kuiper M, Mironov V.** 2013. orthAgogue: an agile tool for the rapid prediction of
596 orthology relations. Bioinformatics. doi: 10.1093/bioinformatics/btt582.

597 48. **Edgar RC, Sjolander K.** 2004. A comparison of scoring functions for protein sequence profile
598 alignment. Bioinformatics. **20:**1301-1308.

599 49. **Edgar RC.** 2004. MUSCLE: a multiple sequence alignment method with reduced time and space
600 complexity. BMC Bioinformatics. **5:**1-19. doi: 10.1186/1471-2105-5-113.

601 50. **Abascal F, Zardoya R, Telford MJ.** 2010. TranslatorX: multiple alignment of nucleotide
602 sequences guided by amino acid translations. Nucleic Acids Res. doi: 10.1093/nar/gkq291.

603 51. **Tamura K, Stecher G, Peterson D, Filipski A, Kumar S.** 2013. MEGA6: Molecular
604 Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. doi: 10.1093/molbev/mst197.

605 52. **Revez J, Rossi M, Ellström P, de Haan CP, Rautelin H, Hänninen M.** 2011. Finnish
606 *Campylobacter jejuni s*trains of multilocus sequence type ST-22 complex have two lineages with
607 different characteristics. PLoS ONE. **6:**e26880.

608 53. **Tsai C, Frasch CE.** 1982. A sensitive silver stain for detecting lipopolysaccharides in
609 polyacrylamide gels. Anal Biochem. **119:**115-119.

610 54. **Dzieciatkowska M, Liu X, Heikema AP, Houliston RS, van Belkum A, Schweda EK, Gilbert**
611 **M, Richards JC, Li J.** 2008. Rapid method for sensitive screening of oligosaccharide epitopes in the
612 lipooligosaccharide from *Campylobacter jejuni* strains isolated from Guillain-Barré syndrome and
613 Miller Fisher syndrome patients. J Clin Microbiol. **46:**3429-3436. doi: 10.1128/JCM.00681-08.

614 55. **Moran AP, Zähringer U, Seydel U, D. Scholz, Stütz P, Rietschel ET.** 1991. Structural analysis
615 of the lipid A component of *Campylobacter jejuni* CCUG 10936 (serotype O:2) lipopolysaccharide.
616 Eur J Biochem. **198:**459-469. doi: 10.1111/j.1432-1033.1991.tb16036.x.

617 56. **Moran AP, Rietschel ET, Kosunen TU, Zähringer U.** 1991. Chemical characterization of
618 *Campylobacter jejuni* lipopolysaccharides containing N-acetylneuraminic acid and 2,3-diamino-2,3-
619 dideoxy-D-glucose. J Bacteriol. **173:**618-626.

57. **Moran AP.** 1997. Structure and conserved characteristics of *Campylobacter jejuni* lipopolysaccharides. J Infect Dis. **176:**S115-S121. doi: 10.1086/513781.

58. **Stahl M, Ries J, Vermeulen J, Yang H, Sham HP, Crowley SM, Badayeva Y, Turvey SE, Gaynor EC, Li X.** 2014. A novel mouse model of *Campylobacter jejuni* gastroenteritis reveals key pro-inflammatory and tissue protective roles for toll-like receptor signaling during Infection. PLoS Pathog. **10:**e1004264. doi: doi: 10.1371/journal.ppat.1004264.

59. **Bereswill S, Fischer A, Plickert R, Haag L, Otto B, Kühl AA, Dashti JI, Zautner AE, Muñoz M, Loddenkemper C, Groß U, Göbel UB, Heimesaat MM.** 2011. Novel murine infection models provide deep insights into the "ménage à trois" of *Campylobacter jejuni*, microbiota and host innate immunity. PLoS ONE. **6:**e20953.

60. **Aspinall GO, McDonald AG, Pang H, Kurjanczyk LA, Penner JL.** 1993. An antigenic polysaccharide from *Campylobacter coli* serotype O:30. Structure of a teichoic acid-like antigenic polysaccharide associated with the lipopolysaccharide. J Biol Chem. **268:**18321-18329.

61. **Roux F, Sproston E, Rotariu O, MacRae M, Sheppard SK, Bessell P, Smith-Palmer A, Cowden J, Maiden MC, Forbes K, Strachan NJ.** 2013. Elucidating the aetiology of human *Campylobacter coli* infections. PLoS ONE. **8:**e64504.

62. **Sweet CR, Ribeiro AA, Raetz CR.** 2004. Oxidation and transamination of the 3″-Position of UDP-N-acetylglucosamine by enzymes from *Acidithiobacillus ferrooxidans*. J Biol Chem. **279:**25400-25410. doi: 10.1074/jbc.M400596200.

63. **Jansonius JN.** 1998. Structure, evolution and action of vitamin B6-dependent enzymes. Curr Opin Struct Biol. **8:**759-769.

641 64. **Altman E, Schäffer C, Brisson JR, Messner P.** 1995. Characterization of the glycan structure of a
642 major glycopeptide from the surface layer glycoprotein of *Clostridium thermosaccharolyticum* E207-
643 71. Eur J Biochem. **229:**308-315. doi: 10.1111/j.1432-1033.1995.0308l.x.

644 65. **Ovchinnikova OG, Rozalski A, Liu B, Knirel YA.** 2013. O-antigens of bacteria of the genus
645 *Providencia*: Structure, serology, genetics, and biosynthesis. Biochemistry (Mosc). **78:**798-817. doi:
646 10.1134/S0006297913070110.

647 66. **Veremeichenko S, Zdorovenko GM.** 2004. Structure and properties of the lipopolysaccharide of
648 *Pseudomonas fluorescens* IMV 2366 (biovar III)]. Microbiology. **73:**312-319.

649 67. **Holden KM, Gilbert M, Coloe PJ, Li J, Fry BN..** 2012. The role of WlaRG, WlaTB and WlaTC
650 in lipooligosaccharide synthesis by *Campylobacter jejuni* strain 81116. Microb Pathog. **52:**344-352.
651 doi: 10.1016/j.micpath.2012.03.004.

652 68. **Godschalk PC, Gilbert M, Jacobs BC, Kramers T, Tio-Gillen AP, Ang CW, Van den Braak**
653 **N, Li J, Verbrugh HA, Van Belkum A, Endtz HP.** 2006. Co-infection with two different
654 *Campylobacter jejuni* strains in a patient with the Guillain–Barré syndrome. Microb Infect. **8:**248-253.

655 69. **Aspinall GO, Lynch CM, Pang H, Shaver RT, Moran AP.** 1995. Chemical structures of the core
656 region of *Campylobacter jejuni* O:3 lipopolysaccharide and an associated polysaccharide. Eur J
657 Biochem. **231:**570-578. doi: 10.1111/j.1432-1033.1995.tb20734.x.

658 70. **Pföstl A, Zayni S, Hofinger A, Kosma P, Schäffer C, Messner P.** 2008. Biosynthesis of dTDP-3-
659 acetamido-3,6-dideoxy-α-D-glucose. Biochem J. **410:**187-194. doi: 10.1042/BJ20071044.

660 71. **Bernatchez S, Szymanski CM, Ishiyama N, Li J, Jarrell HC, Lau PC, Berghuis AM, Young**
661 **NM, Wakarchuk WW.** 2005. A single bifunctional UDP-GlcNAc/Glc 4-epimerase supports the

662 synthesis of three cell surface glycoconjugates in *Campylobacter jejuni*. J Biol Chem. **280:**4792-4802.
663 doi: 10.1074/jbc.M407767200.

664 72. **Ma B, Simala-Grant JL, Taylor DE.** 2006. Fucosylation in prokaryotes and eukaryotes.
665 Glycobiology. **16:**158R-184R. doi: 10.1093/glycob/cwl040.

666 73. **Knirel Y.** 2011. Knirel, Y., 2011. Structure of O-antigens. in: Knirel, Y.A., Valvano,
667 M.A. (Eds.), Bacterial Lipopolysaccharides, Springer, Vienna, Austria, 41–115., p. 41-115. *In* Y. A.
668 Knirel and M. A. Valvano (eds.), Bacterial Lipopolysaccharides. Springer, Vienna, Austria.

669 74. **Hanniffy OM, Shashkov AS, Moran AP, Prendergast MM, Senchenkova SN, Knirel YA,**
670 **Savage AV.** 1999. Chemical structure of a polysaccharide from *Campylobacter jejuni* 176.83 (serotype
671 O:41) containing only furanose sugars. Carbohydr Res. **319:**124-132. doi:

672 75. **Moran AP, O'Malley DT, Kosunen TU, Helander IM.** 1994. Biochemical characterization of
673 *Campylobacter fetus* lipopolysaccharides. Infect Immun. **62:**3922-3929.

674

675 **FIGURE LEGENDS**

676 **Figure 1.** LOS locus classes related to X, XI, and XII. Arrows represent ORFs. Genes coloured white
677 are common to all LOS classes. Genes coloured green are present in class I and/or III. Genes coloured
678 blue are present in classes IV/V and VI. Grey genes are common among classes X and XI. The orange
679 genes are particular of the class XII. Striped genes are fragmented. Lines connect closely related
680 orthologues. Strains are identified if more than one origin was observed in the LOS locus class (see
681 text). Gene size is not drawn to scale.

682 **Figure 2.** Consensus cladogram representing the evolutionary relationship among orthologues
683 belonging to GO 10 (nomenclature from Richard *et al*. 26). *C. jejuni* strains are highlighted in green.

684 *C. coli* with the exception of LOS locus class II strains are shown in red. *C. coli* LOS locus class II
685 strains are highlighted in yellow. The 95% bootstrap consensus tree was built from 100 replicates.
686 Strains LOS locus class is indicated after the strain's ID.

687 **Figure 3.** Host-LOS locus class association. Circos diagram shows the distribution of LOS locus
688 classes of *C. coli* strains isolated from different hosts, from both our collection and those from Richards
689 and colleagues (26). Ribbon ends represent links between host and LOS locus class while the width of
690 the ribbon correlates with the percentage of strains belonging to a specific LOS locus class in a certain
691 host. Segments in the outer ring indicate the percentage of strains representing a certain LOS locus
692 class or host while the inner ring indicates the number of strains. Human strains are shown in orange,
693 bovine in red, poultry in green, and swine in cyan.

694 **Figure 4.** *C.coli* LOS biochemical profiles. A) Silver-stained LOS. B) Proposed chemical composition
695 based on MS and MS/MS results analysis of intact LOS (Supplemental Figure 3).

696 **Figure 5.** Comparison of nucleotide sequence of LOS locus class II strains 151 and 65. Genes coloured
697 white are common to all LOS classes. Genes coloured blue are present in LOS locus classes IV/V, VI,
698 and VII. Yellow coloured genes are particular to LOS locus class II. Lines between orthologues
699 represent sequence similarity.

700

701

Table 1. List of primers used in the present study and expected sizes of the amplicons.

PCR	Primers	Sequence	LOS locus class										
			I	II	III	IV/V	VI	VII	VIII	IX	X	XI	XII
1*	ORF3-F2	AAA AGC TTG TGG CTG GTG GCC TGA TCA	7.1	9.9	7.2	12.6	13.2	15.3	18.2	7.1	11.5	11.4	11.1
	waaV-R	AAG AGC TTT GCA AAG CTG TAT AAA TCA GAC											
2	2209-L	TTC AGG TGT TTA TGA TTT GTT TC	+ **(355)**	-	-	-	-	-	-	-	-	-	-
	2209-R	GCT TGT GCC TTT GGT ATA AGG											
3	CstIV-F	TTC CCA GCA GCT ATA AAT GGA	-	+ **(190)**	-	-	-	-	-	-	-	-	-
	CstIV-R	TTT CAT CTC CAA AAT CCA TGC											
4	1541-L	TGG CAA YTA TGG TTT CAA GG	-	+ **(327)**	-	+ **(327)**	+ **(327)**	+ **(327)**	-	-	-	-	-
	1541-R	TGC YCT TTC AAA AGC AAA AAA TTC											
5	1210-L	AAT TTT GCG TGG AAT GCT TG	-	-	+ **(337)**	-	-	-	-	-	-	-	-
	1210-R	GCT GAA GGC AAT TGA TGA TG											
6	1790-L	CCY TAA AYA CYG CTT TTR AAA AC	-	-	-	+ **(328)**	+ **(328)**	+ **(328)**	-	-	-	-	+ **(328)**
	1790-R	TGC GTA TCT TGT TGA TTR CAC											
7	1920-L	CCA AGC CAG ATT TTC CAA GA	-	-	-	+ **(229)**	-	+ **(229)**	-	-	-	+ **(229)**	+ **(229)**
	1920-R	TCG TTA TAG AAA TCA CTT GCC AAT											
8	2344-L	AAA GAA AGA GAA GCC AAA GGA G	-	-	-	-	-	+ **(348)**	-	-	-	-	-
	2344-R	TCT TGG TTT AAT TTT CGC ATA TTC											
9	1790R	TGC GTA TCT TGT TGA TTR CAC	-	-	-	+ **(2252)**	-	+ **(4933)**	-	-	-	-	-
	1920L	CCA AGC CAG ATT TTC CAA GA											
10	38_3454	ACG CCT AGC GTG TAA ACC AT	-	-	-	-	-	-	+ **(1046)**	-	-	-	-
	38_2031	ATC GTC CTA TAG CTA CGG GTG A											
11	CstV-F	TTC CTT TGC AAC ACG AAA TAA	-	-	-	-	-	-	-	+ **(449)**	-	-	-
	CstV-R	GTT TTG GAG CTA GCG GAA TA											
12	45_8	GTG CTT GAG CGC AAT CTT CT	-	-	-	-	-	-	-	-	+ **(1036)**	+ **(1036)**	-
	45_1	GAG GGG CCT TAT GGA GCA AA											

* the amplicons of this PCR are expressed in kb, while all others are in bp.

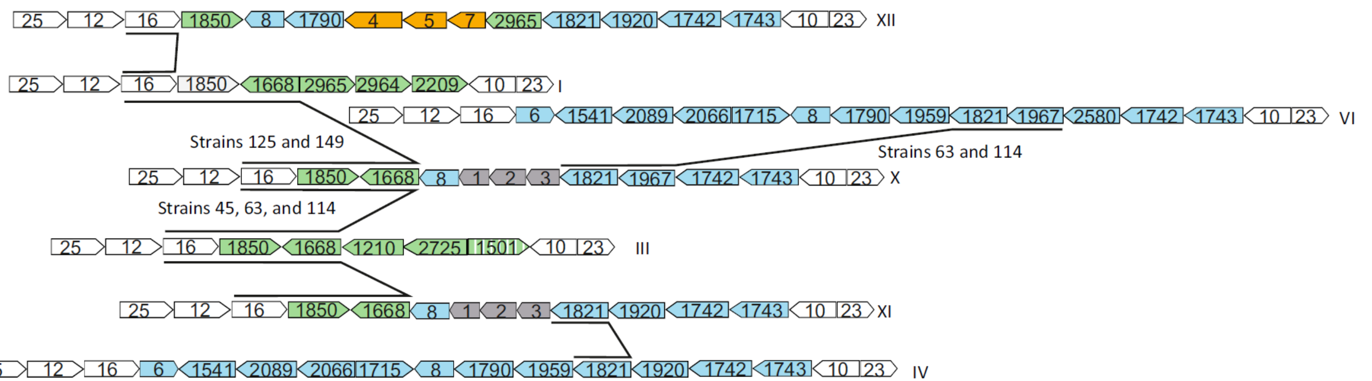
706 TABLE 2. Distribution of LOS classes among hosts

LOS class	Total (%)	Human	Swine	Poultry	Wild birds
I	3	2	0	1	0
II	24 (17)	7	13	4	0
III	18 (13)	4	13	0	1
IV/V	22 (15)	3	16	3	0
VI	18 (13)	1	15	2	0
VII	2 (1)	1	1	0	0
VIII	10 (7)	7	1	2	0
IX	1	1	0	0	0
X	22 (15)	3	18	1	0
XI	1	0	1	0	0
XII	1	0	0	0	1
Untypable	22 (15)	5	12	5	0

707

708

709



710

711 **Figure 1.** LOS locus classes related to X, XI, and XII. Arrows represent ORFs. Genes coloured white
712 are common to all LOS classes. Genes coloured green are present in class I and/or III. Genes coloured
713 blue are present in classes IV/V and VI. Grey genes are common among classes X and XI. The orange
714 genes are particular of the class XII. Striped genes are fragmented. Lines connect closely related
715 orthologues. Strains are identified if more than one origin was observed in the LOS locus class (see
716 text). Gene size is not drawn to scale.

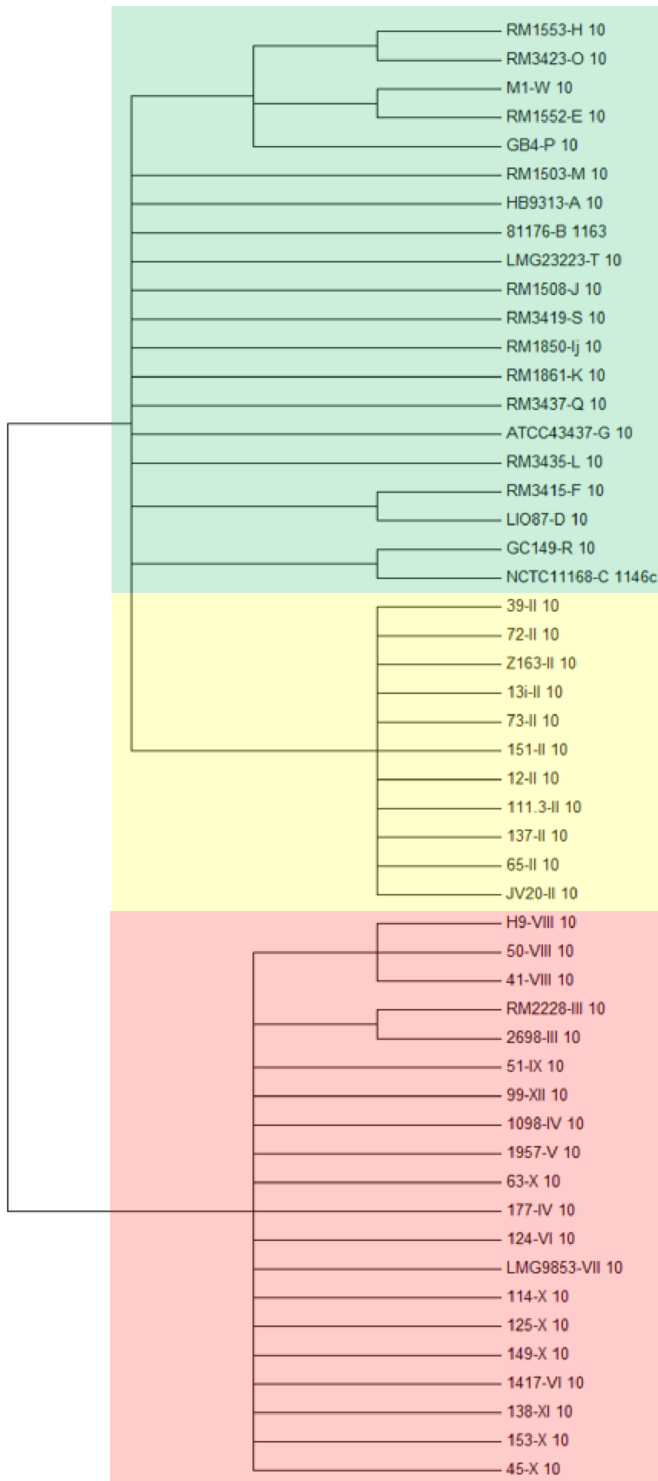
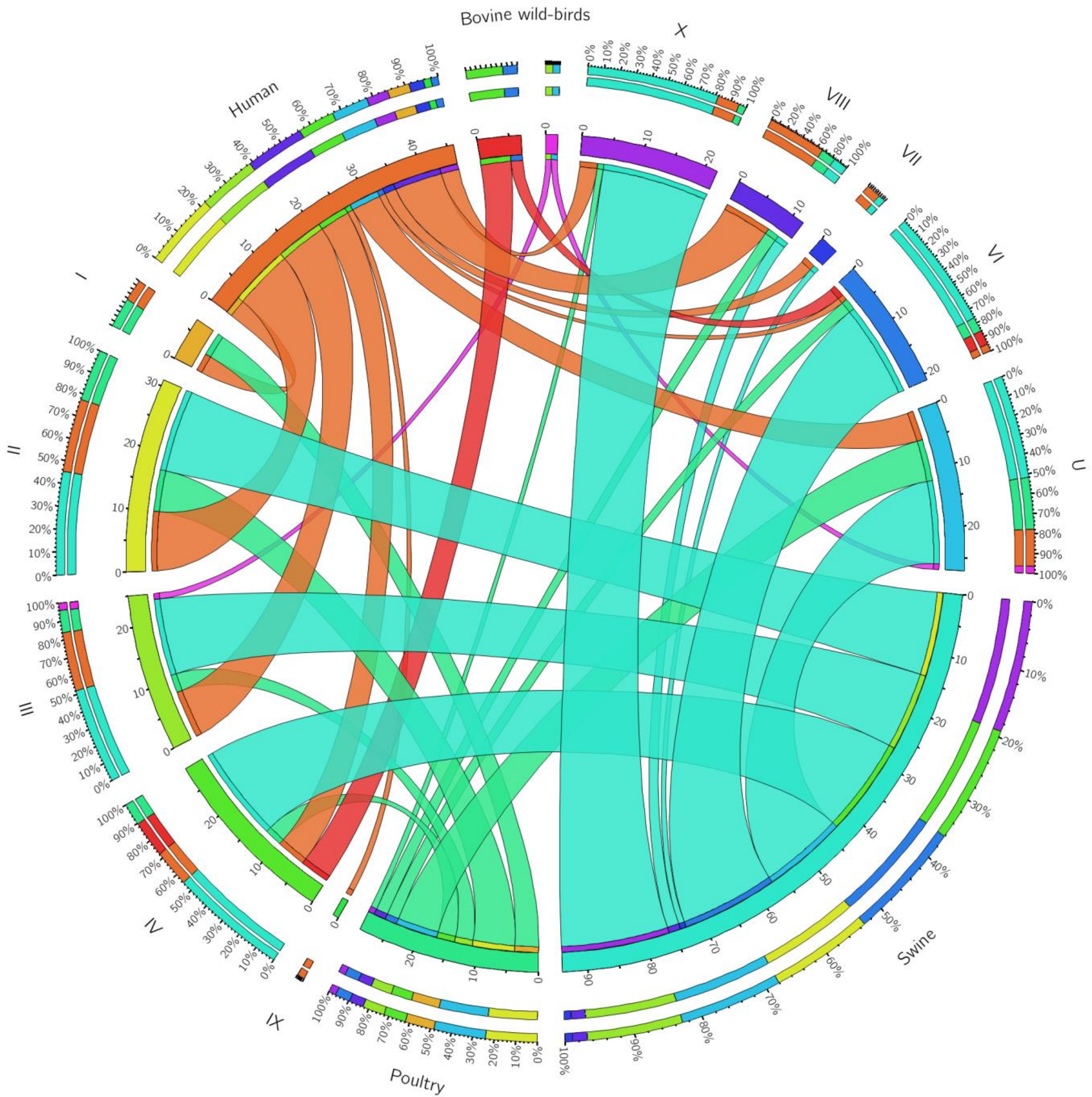
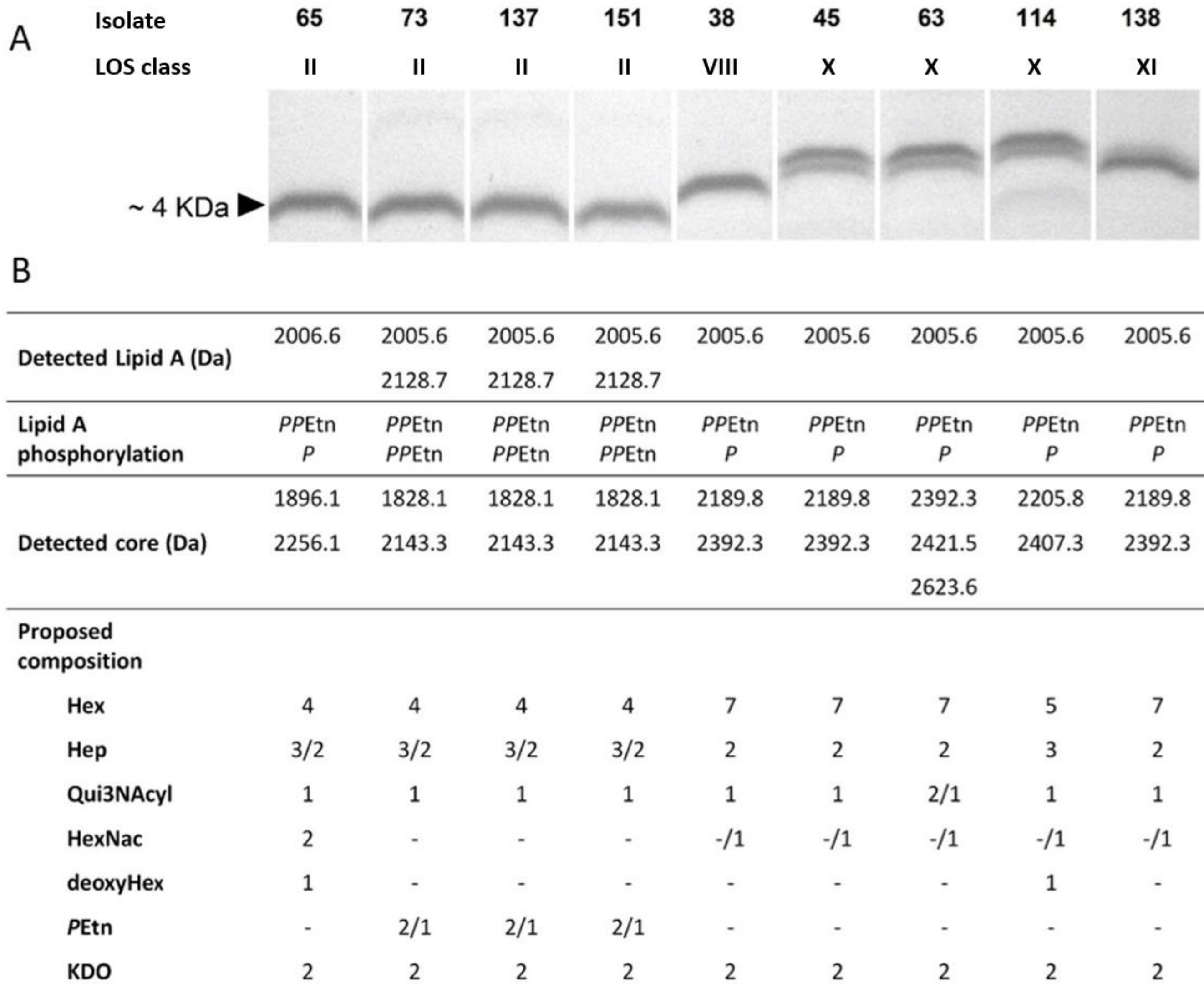


Figure 2. Consensus cladogram representing the evolutionary relationship among orthologues belonging to GO 10 (nomenclature from Richard *et al*. 26). *C. jejuni* strains are highlighted in green. *C. coli* with the exception of LOS locus class II strains are shown in red. *C. coli* LOS locus class II strains are highlighted in yellow. The 95% bootstrap consensus tree was built from 100 replicates. Strains LOS locus class is indicated after the strain's ID.



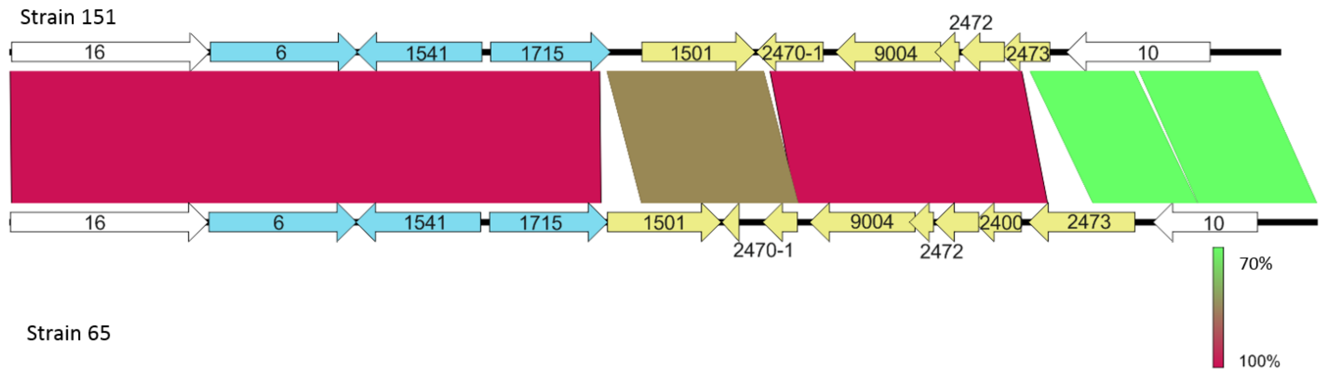
723

724 **Figure 3.** Host-LOS locus class association. Circos diagram shows the distribution of LOS locus
725 classes of *C. coli* strains isolated from different hosts, from both our collection and those from Richards
726 and colleagues (26). Ribbon ends represent links between host and LOS locus class while the width of
727 the ribbon correlates with the percentage of strains belonging to a specific LOS locus class in a certain
728 host. Segments in the outer ring indicate the percentage of strains representing a certain LOS locus
729 class or host while the inner ring indicates the number of strains. Human strains are shown in orange,
730 bovine in red, poultry in green, and swine in cyan.



	65	73	137	151	38	45	63	114	138
Detected Lipid A (Da)	2006.6	2005.6 2128.7	2005.6 2128.7	2005.6 2128.7	2005.6	2005.6	2005.6	2005.6	2005.6
Lipid A phosphorylation	*PP*Etn *P*	*PP*Etn *PP*Etn	*PP*Etn *PP*Etn	*PP*Etn *PP*Etn	*PP*Etn *P*	*PP*Etn *P*	*PP*Etn *P*	*PP*Etn *P*	*PP*Etn *P*
Detected core (Da)	1896.1 2256.1	1828.1 2143.3	1828.1 2143.3	1828.1 2143.3	2189.8 2392.3	2189.8 2392.3	2392.3 2421.5 2623.6	2205.8 2407.3	2189.8 2392.3
Proposed composition									
Hex	4	4	4	4	7	7	7	5	7
Hep	3/2	3/2	3/2	3/2	2	2	2	3	2
Qui3NAcyl	1	1	1	1	1	1	2/1	1	1
HexNac	2	-	-	-	-/1	-/1	-/1	-/1	-/1
deoxyHex	1	-	-	-	-	-	-	1	-
***P*Etn**	-	2/1	2/1	2/1	-	-	-	-	-
KDO	2	2	2	2	2	2	2	2	2

Figure 4. *C.coli* LOS biochemical profiles. A**)** Silver-stained LOS. **B)** Proposed chemical composition based on MS and MS/MS results analysis of intact LOS (Supplemental Figure 3).



736

737 **Figure 5.** Comparison of nucleotide sequence of LOS locus class II strains 151 and 65. Genes coloured

738 white are common to all LOS classes. Genes coloured blue are present in LOS locus classes IV/V, VI,

739 and VII. Yellow coloured genes are particular to LOS locus class II. Lines between orthologues

740 represent sequence similarity.

741