

1 **Genomic epidemiology of *Campylobacter jejuni* associated with** 2 **asymptomatic pediatric infection in the Peruvian Amazon**

3

4 Ben Pascoe^{1,2*}, Francesca Schiaffino^{3,4}, Susan Murray^{5,6}, Guillaume Méric^{1#}, Sion C. Bayliss¹,
5 Matthew D. Hitchings⁵, Evangelos Mourkas¹, Jessica K. Calland¹, Rosa Burga⁷, Pablo
6 Peñataro Yori^{8,9}, Keith A. Jolley¹⁰, Kerry K. Cooper¹¹, Craig T. Parker¹², Maribel Paredes
7 Olortegui⁹, Margaret N. Kosek^{8,9*} and Samuel K. Sheppard^{1,2,10}

8

9 ¹The Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Bath,
10 UK; ²Chiang Mai University, Chiang Mai, Thailand; ³Department of International Health, Johns
11 Hopkins Bloomberg School of Public Health, Baltimore, USA; ⁴Faculty of Veterinary Medicine,
12 Universidad Peruana Cayetano Heredia, Lima, Peru; ⁵Swansea University Medical School, Swansea
13 University, Singleton Park, Swansea, UK; ⁶Department of Medical Biochemistry and Microbiology,
14 Science for Life Laboratories, Uppsala University, Uppsala, Sweden; ⁷Bacteriology Department, Naval
15 Medical Research Unit-6 (NAMRU-6), Iquitos, Peru; ⁸The Division of Infectious Diseases and
16 International Health, University of Virginia, Charlottesville, VA, USA; ⁹Asociacion Benefica Prisma,
17 Loreto, Peru; ¹⁰Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS,
18 UK; ¹¹School of Animal and Comparative Biomedical Sciences, University of Arizona, Tucson,
19 Arizona, USA; ¹²Produce Safety and Microbiology Research Unit, Agricultural Research Service, US
20 Department of Agriculture, Albany, CA, USA.

21

22 #Present address: Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute,
23 75 Commercial Rd, Melbourne 3004, Victoria, Australia.

24

25 *Address correspondence to: Ben Pascoe (b.pascoe@bath.ac.uk), Margaret Kosek
26 (MNK2N@hscmail.mcc.virginia.edu).

27

28 **Short title:** *C. jejuni* genotypes associated with disease severity in Peru

29 **Keywords:** Gastroenteritis; Campylobacteriosis; *Campylobacter jejuni*; Asymptomatic
30 carriage; Infant growth stunting; Source attribution.

31 **Abbreviations:** LMIC: low- to middle- income country; GEMS: The Global Enteric
32 Multicenter Study; MAL-ED: Malnutrition and Enteric Disease Study; MLST: multi-locus
33 sequence typing; ST: sequence type; CC: clonal complex; LOS: lipooligosaccharide; AMR:
34 antimicrobial resistance; CPS: capsular polysaccharide

35 **Abstract**

36 *Campylobacter* is the leading bacterial cause of gastroenteritis worldwide and its incidence is
37 especially high in low- and middle-income countries (LMIC). Disease epidemiology in LMICs
38 is different compared to high income countries like the USA or in Europe. Children in LMICs
39 commonly have repeated and chronic infections even in the absence of symptoms, which can
40 lead to deficits in early childhood development. In this study, we sequenced and characterized
41 *C. jejuni* (n=62) from a longitudinal cohort study of children under the age of 5 with and
42 without diarrheal symptoms, and contextualized them within a global *C. jejuni* genome
43 collection. Epidemiological differences in disease presentation were reflected in the genomes,
44 specifically by the absence of some of the most common global disease-causing lineages. As
45 in many other countries, poultry-associated strains were a major source of human infection but
46 almost half of local disease cases (15 of 31) were attributable to genotypes that are rare outside
47 of Peru. Asymptomatic infection was not limited to a single (or few) human adapted lineages
48 but resulted from phylogenetically divergent strains suggesting an important role for host
49 factors in the cryptic epidemiology of campylobacteriosis in LMICs.

50

51 **Author summary**

52 *Campylobacter* is the leading bacterial cause of gastroenteritis worldwide and despite high
53 incidence in low- and middle-income countries (LMICs), where infection can be fatal, culture
54 based isolation is rare and the genotypes responsible for disease have not broadly been
55 identified. The epidemiology of disease is different to that in high income countries, where
56 sporadic infection associated with contaminated food consumption typically leads to acute
57 gastroenteritis. In some LMICs infection is endemic among children and common
58 asymptomatic carriage is associated with malnutrition, attenuated growth in early childhood,
59 and poor cognitive and physical development. Here, we sequenced the genomes of isolates
60 sampled from children in the Peruvian Amazon to investigate genotypes associated with
61 varying disease severity and the source of infection. Among the common globally circulating
62 genotypes and local genotypes rarely seen before, no single lineage was responsible for
63 symptomatic or asymptomatic infection – suggesting an important role for host factors.
64 However, consistent with other countries, poultry-associated strains were a major source of
65 infection. This genomic surveillance approach, that integrates microbial ecology with
66 population based studies in humans and animals, has considerable potential for describing
67 cryptic epidemiology in LMICs and will inform work to improve infant health worldwide.

68

69 **Introduction**

70 The World Health Organization ranks diarrheal disease as the second most common cause of
71 mortality among children under five years of age in low- and middle-income countries
72 (LMICs), accounting for 10.6 million annual deaths in this age group [1,2]. *Campylobacter* is
73 the most common cause of bacterial gastroenteritis in Europe and the USA, with even higher
74 incidence in LMICs (up to 85% of children infected before 12 months [3]). However,
75 *Campylobacter* infection is largely overlooked in LMICs for several reasons. Infection is
76 thought to be sporadic so outbreaks are seldom recorded. *Campylobacter* are also more difficult
77 to grow in the laboratory than many common enteric pathogens, so it is often not cultured even
78 when present. These factors conspire such that the people at the greatest risk are the least
79 studied.

80
81 In high-income countries, human campylobacteriosis is readily diagnosed as a disease
82 associated with consumption of contaminated food, especially poultry [4,5], but the extremely
83 high incidence in LMICs suggests different epidemiology. High exposure rates [6,7] and
84 apparent endemism among young children [8–10] are a major concern, particularly as frequent
85 or chronic (re)infection is linked to significant morbidity, growth faltering, cognitive
86 impairment, and even death [11,12]. However, there is also evidence of common asymptomatic
87 carriage among children in LMICs [7], a phenomenon that is not well understood. International
88 studies have begun to quantify the causes of enteric infection in children [13–16] but
89 campylobacteriosis surveillance programs remain uncommon and the strains responsible for
90 disease are seldom characterized in LMICs [11,17–23]. Understanding the true disease burden
91 requires not only incidence data, but also knowledge of variation in disease symptoms and the
92 genotypes associated with asymptomatic and severe infection.

93

94 DNA-sequence-based strain characterization, typically of isolates from developed countries,
95 has revealed considerable diversity within the major disease-causing *Campylobacter* species
96 (*C. jejuni* and *C. coli*). This has allowed identification of the genotypes, and in some cases
97 genes, linked with variation in disease symptoms and the source of infecting strains. For
98 example, the identification of host-associated genetic variation [24] and the extent to which
99 this segregates by host (host generalist and specialist genotypes) [25–27], means that human
100 infection can be attributed to a specific reservoir source, when there is no human-to-human
101 transmission [24,25,27–29]. Furthermore, in some cases it is possible to link particular
102 genotypes to common disease sequelae [30–32] or severe infections [33–35], and identify
103 locally [36–38] and globally distributed strains [39,40].

104

105 Among the most fundamental challenges in LMICs is to understand if disease severity and
106 asymptomatic carriage are dictated by host factors, such as malnutrition [12], or the source and
107 genotype of the infecting strain. In this study we address this as part of ongoing surveillance in
108 Santa Clara, a semi-rural community near Iquitos in the Peruvian Amazon (**Figure 1A**). *C.*
109 *jejuni* were isolated from individuals with varying disease severity, from no symptoms to
110 severe infection, and the genomes were sequenced and contextualized within a global reference
111 collection. Both, locally and globally disseminated genotypes were isolated from Peruvian
112 children with a range of disease symptoms. Comparative genomics of isolates from
113 symptomatic and asymptomatic individuals identified signatures of local diversification but
114 little evidence of genetic elements specifically responsible for severe disease. Household
115 crowding, poor sanitation, consumption of contaminated water and cohabitation with animals
116 remain potential risks for local transmission, but poultry were revealed as an important

117 infection reservoir based on source attribution analysis. This study provides a basis for
118 considering complex transmission networks in LMICs and highlights the role of globally
119 transmitted *Campylobacter* lineages.

120

121 **Methods**

122 ***Sampling and cohort information***

123 Samples collected as part of a cohort study from Iquitos, in the Peruvian Amazon, between
124 2002 and 2006. In this age-stratified sample set of 442 children aged 0-5 years [7,13–15,41,42],
125 children were visited 3 times weekly to form a continuous symptom history of childhood
126 illnesses. Stool samples were collected quarterly from all children and in cases in which
127 diarrhea was detected (92.3% of episodes detected by surveillance had a sample collected;
128 **Table S1**). Fecal samples were swabbed into Cary-Blair transport media, suspended in PBS,
129 filtered through a 0.45 µm membrane and placed on a Columbia Blood Agar base (Oxoid)
130 supplemented by 5% defibrinated sheep's blood for 30 minutes prior to removal and streaking
131 of filtrate. The Johns Hopkins Institutional Review Board provided ethical approval for the
132 MAL-ED study in addition to respective partner institutions for each site, including Asociacion
133 Benefica PRISMA, and the Regional Health Department of Loreto, Peru. Written consent was
134 obtained from all participants.

135

136 ***Bacterial isolate genome sequencing***

137 Genomic DNA was extracted from 62 *C. jejuni* isolates and sequenced using an Illumina MiSeq
138 benchtop sequencer (California, USA). Nextera XT libraries (Illumina, California, USA) were
139 prepared and short paired-end reads (250 bp) were assembled *de novo* using Velvet (version
140 1.2.08) [43] with VelvetOptimiser (version 2.2.4). The average number of contiguous
141 sequences (contigs) was 262 (range: 53–701) for an average total assembled sequence size of
142 1.55 Mbp (range: 1.37–1.70). The average N50 contig length (L50) was 14,577 (range: 3,794-
143 55,912) and the average GC content was 30.8 % (range: 30.5-31.6). Short read data are
144 available on the NCBI SRA, associated with BioProject PRJNA350267. Assembled genomes

145 and supplementary material are available from FigShare (doi:10.6084/m9.figshare.10352375;
146 individual accession numbers and assembled genome statistics in **Table S2**). Isolates were
147 compared to a global reference dataset representing the genetic diversity of the species (n=164
148 isolates from eight countries and three continents) (**Table S3**)[26,36,44–47].

149

150 *Diarrheal disease severity*

151 As part of the ongoing surveillance efforts, a questionnaire was completed three times per week
152 to record diarrheal symptoms for all members of the cohort [7,13,14], generating a continual
153 illness record for the surveillance period. *Campylobacter* isolated from patients that did not
154 display any symptoms two days before or after collection of the stool sample were considered
155 asymptomatic. Diarrhea was defined by three or more semi-liquid stools reported over a 24-
156 hour period, with episodes separated by at least three symptom-free days. Diarrheal severity
157 symptoms were catalogued and details recorded of any symptom, including the number of
158 diarrheal episodes, hematochezia (blood in the stool), fever, incidence of vomiting and anorexia
159 (**Table S1**)[48].

160

161 *Core genome genealogies*

162 A reference pan-genome file was constructed by combining open reading frames identified by
163 RAST [49,50] in all the Peruvian isolates and the *C. jejuni* NCTC 11168 reference strain to
164 maintain locus nomenclature [51]. Gene orthologues ($\geq 70\%$ sequence similarity) were
165 identified and duplicates removed (size: 2,045,739 bp; **Supplementary file S1**). Two
166 alignment files were constructed from concatenated gene sequences of all core genes (found in
167 $\geq 95\%$ isolates) from the reference pan-genome list using MAFFT [52] on a gene-by-gene basis
168 [53,54]: one for the Peruvian isolates only (size: 772,794 bp; **Supplementary file S2**); and a

169 second alignment containing the Peruvian isolates plus the global reference collection (size:
170 720,853 bp; **Supplementary file S3**). Maximum-likelihood phylogenies were constructed in
171 IQ-TREE (version 1.6.8) using the GTR+F+I+G4 substitution model and ultra-fast
172 bootstrapping (1,000 bootstraps)[55,56]; and visualized on Microreact [57]: Peru only
173 (<https://microreact.org/project/CampyPeruOnly>); Peru and the global reference dataset
174 (<https://microreact.org/project/CampyPeruContext>).

175

176 *Molecular typing and diversity estimates*

177 Isolate genomes were archived in BIGSdb and MLST sequence types (STs) derived through
178 BLAST comparison with the pubMLST database [58–60]. Capsule polysaccharide (CPS) and
179 lipooligosaccharide (LOS) locus types of each *C. jejuni* isolate were characterized from their
180 raw sequence data: short read sequences were mapped to known capsule and LOS locus types
181 using BLAST as previously described [61,62]. Simpson’s index of diversity (with 95%
182 confidence limits) was calculated for sequence types in the Peruvian and global reference
183 datasets using the equation:

$$184 \quad D = 1 - \frac{\sum n(n-1)}{N(N-1)}$$

185 Where n is the number of isolates of each sequence type and N is the total number of isolates
186 [55,63].

187

188 *Accessory genome characterization*

189 The reference pan-genome list contained 2,348 genes, of which 1,321 genes were shared by all
190 isolates ($\geq 95\%$) and defined as the core genome (**Table S4**). The accessory genomes of each
191 isolate was characterized, including detection of antimicrobial resistance genes, putative
192 virulence factors and known plasmid genes using ABRICATE (version 0.9.8) and the CARD, NCBI,

193 ResFinder, VfDB and PlasmidFinder databases (10th September, 2019 update; **Table S5** and
194 summarized in **Table S6**) [64–69]. Pairwise core and accessory genome distances were
195 compared using PopPunk (version 1.1.4). PopPUNK uses pairwise nucleotide k-mer
196 comparisons to distinguish shared sequence and gene content to identify divergence of the
197 accessory genome in relation to the core genome. A two-component Gaussian mixture model
198 was used to construct a network to define clusters (Components: 43; Density: 0.1059;
199 Transitivity: 0.8716; Score: 0.7793) [70].

200

201 *Source attribution*

202 Sequence type (ST) and clonal complex (CC) ecological association were assigned based on
203 previous publication and the relative abundance of STs among different host/sources within
204 pubMLST (**Table S7**) [26,58]. Probabilistic assignment of the source host of infection was
205 estimated using Structure v2.3.4, a Bayesian model-based clustering method designed to infer
206 population structure and assign individuals to populations using multilocus genotype data
207 [27,28,36,71,72]. In the absence of contemporaneous reservoir samples from Peru, we used a
208 random selection of MLST profiles from pubMLST (n=1,229; ~300 isolates per putative source
209 reservoir; **Table S8**). A global genotype collection can be used for reservoir comparison as it
210 is known that host-associated genetic variation transcends phylogeographic signatures [27].
211 MLST profiles of known provenance were used to train the model (from 13 countries - 98%
212 European; collected from 1996-2018). Isolates were grouped by source reservoir: chicken
213 (denoting chicken carcass, meat or broiler environments), ruminant (cattle, sheep or goat feces,
214 offal, or meat), wild birds (including starlings, ducks and geese) or other animal (as listed in
215 pubMLST).

216

217 Self-assignment of a random subset of the comparison data set was conducted by removing a
218 third of the isolates from each candidate population (n=388). Structure was run for 10,000
219 iterations following a burn-in period of 10,000 iterations using the no admixture model to
220 assign individuals to putative populations. The assignment probability for each source was
221 calculated for each isolate individually and isolates attributed to the putative origin population
222 with the greatest attribution probability. We report an average self-assignment score of 61%
223 (range 56.5-63.6%) following five independent estimations, consistent with other studies
224 [27,28,73,74].

225 **Results**

226 *Globally circulating disease genotypes are found in the Peruvian Amazon*

227 We sequenced and characterized a collection of *C. jejuni* isolates (n=62) from a longitudinal
228 cohort study of children under the age of 5 years sampled from diarrheal episodes and stools
229 collected by protocol in the absence of diarrheal illness (**Figure 1A**). Isolate genotypes were
230 compared with all genomes deposited in the pubMLST database (97,012 profiles, data accessed
231 17th February, 2020) and ranked according to how frequently they were found associated with
232 human disease (**Figure 1B**). Nearly half of the isolates (n=29, 47 %) were from common
233 lineages, isolated many times before and recorded in pubMLST (>50 MLST profiles; **Figure**
234 **1B; Table S7**). Symptomatic (n=16; 52 % of disease isolates) and asymptomatic (n=12; 43 %
235 of carriage isolates) isolates belonged to nine STs (eight CCs), including ST-353 (n=13), ST-
236 45 (n=4), ST-354 (n=3), ST-607 (n=2), ST-460 (n=2), ST21 (CC21, n=1), ST50 (CC21, n=1),
237 52 (n=1) and ST-403 (n=1) (**Tables S6**). Of these common globally-distributed STs,
238 represented by three or more isolates, only ST-45 was associated with disease - with 75 % of
239 isolates (3 of 4) leading to symptomatic infection.

240

241 *Proliferation of globally rare genotypes in Peruvian Amazon children*

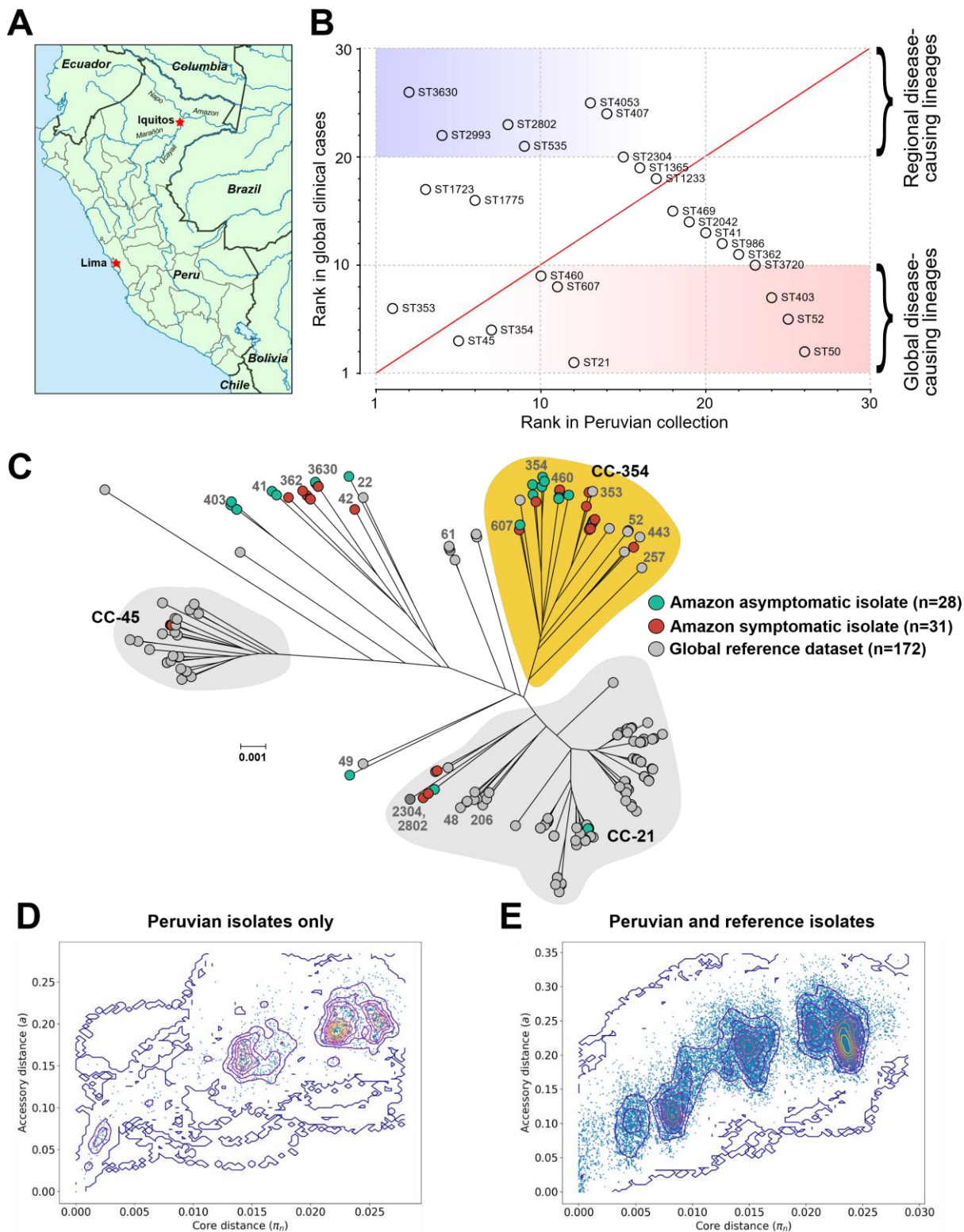
242 The remaining 33 isolates (53 %) belonged to STs that are uncommon in the pubMLST
243 database (<50 MLST profiles; **Figure 1B; Table S7**). This suggests that certain lineages that
244 are rare in the UK and the USA may be more common among children in the Peruvian Amazon.
245 Symptomatic (n=15; 48 % of disease isolates) and asymptomatic (n=16; 57 % of carriage
246 isolates) isolates belonged to 17 STs (15 CCs), including ST-3630 (n=6), ST-1723 (n=5), ST-
247 2993 (n=4), ST-1775 (n=3), ST-2802 (n=2), ST-535 (n=2), ST-362 (n=1), ST-3720 (n=1), ST-
248 407 (n=1), ST-41 (n=1), ST-469 (n=1), ST-1233 (n=1), ST-1365 (n=1), ST-2042 (n=1), ST-

249 2304 (n=1), ST-4053 (n=1) and ST-986 (n=1). Four of these rare STs were represented by three
250 or more isolates: ST-3630 (4 of 6) and ST-2993 (CC362, 4 of 4) were predominantly
251 symptomatic; while ST-1723 (CC354, 4 of 5) and ST-1775 (CC403, 3 of 3) were
252 predominantly asymptomatic (**Table S6**).

253

254 All *C. jejuni* genomes (n=62) were compared to a global reference dataset representing known
255 genetic diversity within *C. jejuni* (n=164 isolates from eight countries and three continents)
256 using a maximum-likelihood phylogenetic tree (**Figure 1C**). Peruvian pediatric isolates did not
257 cluster clearly by geography or disease severity. There was evidence that *C. jejuni* from
258 children in the Peruvian Amazon represented a genetically diverse population. Specifically,
259 there were 26 STs (19 CCs) among the Peruvian isolate collection, with a Simpson's diversity
260 index of 0.904 (95% CI: 0.863-0.946), compared to 50 STs (15 CCs) among the global
261 collection of genomes (Simpson's diversity index = 0.534, 95% CI: 0.453-0.615).

262



263

264 **Figure 1. (A)** Location of study site in Santa Clara, near Iquitos in Peru. **(B)** Sequence types
265 (STs) of isolates collected from children in the Peruvian Amazon ranked according to the
266 frequency in our local dataset and how often they have been sampled from human disease
267 isolates (data from pubMLST; <https://pubmlst.org/>). **(C)** Population structure of *C. jejuni*
268 isolates used in this study. All core (present in $\geq 95\%$ of isolates) genes from the reference pan-
269 genome list (2,348 genes) were used to build alignments of the Peruvian isolates (n=62)
270 contextualized with 172 previously published genomes representing the known genetic
271 diversity in *C. jejuni* (n=234, alignment: 720,853 bp. A maximum-likelihood phylogeny was
272 constructed with IQ-TREE, using a GTR model and ultrafast bootstrapping (1,000 bootstraps;
273 version 1.6.8) [55,56]. Scale bar represents genetic distance of 0.001. Leaves from
274 asymptomatic Peruvian isolates are colored green; symptomatic Peruvian isolates are red; and
275 isolates from the reference dataset are grey. Common STs and clonal complexes (CC), based
276 on four or more shared alleles in seven MLST housekeeping genes, are annotated [60].
277 Interactive visualization is available on Microreact [57]:
278 <https://microreact.org/project/CampyPeruContext>. **(D)** Pairwise core and accessory genome
279 distances were compared using PopPunk for the Peruvian pediatric genomes only and **(E)** with
280 the global reference dataset (version 1.1.4) [70].

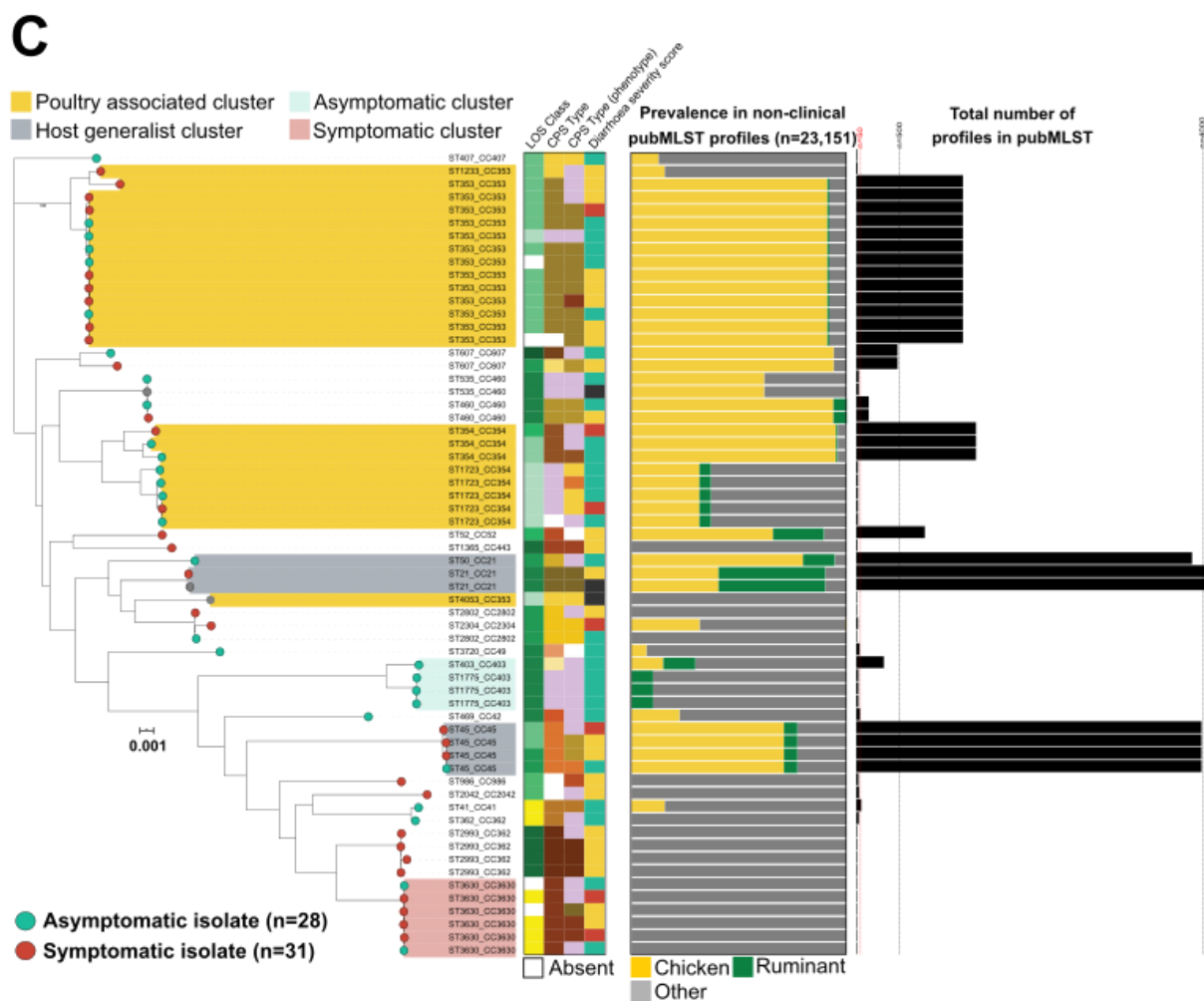
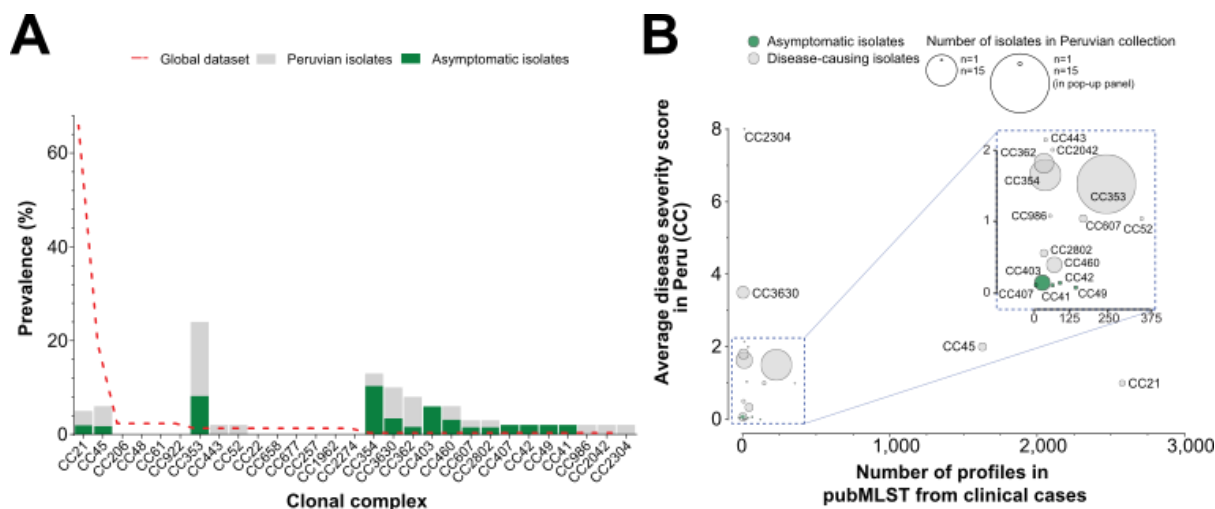
281

282

283 *Peruvian Amazon pediatric isolates have a local gene pool*

284 While there were more STs in the Peruvian collection, there were fewer deep branching
285 lineages compared to the global reference collection (**Figure 1DE**). This is not surprising as
286 there were fewer samples in total and they came from a specific region and source (children).
287 Discontinuous distribution of pairwise genomic distances in the Peruvian pediatric dataset is
288 indicative of multiple genetically distinct clusters that are diverging in both core sequences and
289 accessory gene content. Visualization of this clustering using the t-distributed stochastic
290 neighbor embedding (t-SNE) projection of accessory distances tightly grouped the Peruvian
291 isolates from the Amazon, while isolates from host generalist lineages in the global reference
292 dataset (absent from the Peru dataset) were more loosely clustered (**Figure S1**). This provided
293 evidence of increased horizontal gene transfer (HGT) among Peruvian isolates, compared to
294 global isolate collection.

295



296

297

298 **Figure 2. (A)** Frequency of clonal complexes (CCs) identified among isolates collected from
299 children in the Peruvian Amazon (grey bars) and the global reference dataset (red dotted line).
300 Asymptomatic isolates are colored in green. **(B)** Average severity score of CCs represented by
301 3 or more genomes in our local dataset and how often they have previously been sampled from
302 human disease (data from pubMLST; <https://pubmlst.org/>). Circle diameter represents how
303 frequently they were sampled in our Peruvian Amazon pediatric collection. **(C)** A maximum-
304 likelihood phylogeny was constructed with IQ-TREE, using a GTR model and ultrafast
305 bootstrapping (1,000 bootstraps; version 1.6.8) [55,56] from an alignment of the Peruvian
306 isolates only (n=62, alignment: 772,794 bp. Scale bar represents genetic distance of 0.001.
307 Leaves from asymptomatic isolates are colored green and symptomatic isolates are red. The
308 tree is annotated with lipooligosaccharide classes, capsular types and disease severity scores.
309 Colored bar charts indicate the frequency with which the corresponding sequence type has been
310 isolated from non-human hosts in pubMLST. Black bars indicate the overall frequency that the
311 corresponding ST profile has been sampled before. Interactive visualization is available on
312 Microreact [57]: <https://microreact.org/project/CampyPeruOnly>.
313

314 *Lineages associated with asymptomatic infection in Peruvian Amazon pediatric cases*

315 Asymptomatic isolates and symptomatic isolates represented 17 STs (14 CCs) and 16 STs (14
316 CCs) respectively. Only 9 STs (8 CCs) contained a mixture of both disease etiologies. Of these
317 common global STs represented by three or more isolates, only ST-45 was consistently
318 associated with disease symptoms, with 75 % of isolates (3 of 4) leading to symptomatic
319 infection (**Figure 2AB; Table S1**). Four rare STs: ST-3630 (4 of 6) and ST-2993 (CC362, 4
320 of 4) were predominantly symptomatic; while ST-1723 (CC354, 4 of 5) and ST-1775 (CC403,
321 3 of 3) were predominantly asymptomatic (**Figure 2AB; Table S1**).

322

323 *Regional differences in accessory genome content*

324 There was no difference in the mean genome size between symptomatic and asymptomatic
325 isolates, but significant difference between the Peruvian Amazon pediatric population and the
326 global reference dataset (ANOVA with Tukey's multiple comparisons test, p-value <0.0001;
327 **Figure S1AB**). This can partially be explained by a lack of isolates in the Peruvian pediatric
328 collection from host generalist lineages, which tend to have larger genomes (ST-21 and ST-45

329 CCs; **Figure S1AB**), consistent with genome reduction being associated with increased host
330 specialization [75,76]. As is typical of *Campylobacter* [35,44,54], the isolate collection
331 included a large accessory genome (**Table S4**), with a little over half (56 %) the genes identified
332 in the genomes of our 62 isolates from Peruvian children considered to be core (1,321 of 2,348
333 genes present in 95% of isolates). A large proportion of the accessory genome (446 genes, 43
334 % of the 1,027 accessory genes present in between 0 and 95 % of isolates) were present in less
335 than 15 % of isolates.

336

337 Using the reference pan-genome list, genes that were core in the reference dataset were also
338 present in the Peruvian pediatric dataset (average prevalence: 97.7 %) (**Figure S1C; Table**
339 **S9**). All 29 of the NCTC11168 genes that were absent from Peruvian Amazon isolates
340 (prevalence less than 5%) were found among genomes of isolates in the reference dataset
341 (average prevalence: 43.0 %), with 21 specifically from the lipooligosaccharide (LOS) and
342 capsular polysaccharide (CPS) loci. The LOS and CPS loci are highly variable in gene content
343 [77–80] and this variability is reflected in the diversity of LOS and capsule types for the
344 Peruvian isolates (n=14 LOS types; n=21 capsule types; **Figure S2; Table S6**). The most
345 common LOS class locus was class H in 14 strains and 12/14 of these strains were poultry
346 specialists and 10/14 strains were from symptomatic cases. LOS class B was present in 11
347 strains and only 2/11 were from symptomatic cases. There were four strains with LOS class A
348 and all were from cases with symptomatic etiology and also possessed the HS:41 CPS locus.
349 The most common CPS Penner type was HS:3 (n=10) and 70% of these strains were from
350 symptomatic cases and all ten had LOS class H (**Table S6**).

351

352 ***Poultry is the predominant source of infection in Peruvian Amazon children***

353 STs were attributed to a putative host source based on their predominant sampling source in a
354 global collection on pubMLST (**Table S7**). Isolates from poultry specialist lineages, including
355 the globally disseminated ST-353, ST-354, ST-607 and ST-460, were the most common source
356 of infection (n=32; **Figure 2C, Table S7**). Isolates from rare lineages, scarcely found outside
357 human clinical cases (ST-3630, ST-2993, ST-2802, ST-986, ST41, ST362 and ST2402) were
358 associated with the most severe symptoms. Poultry specialist and clinical specialist STs had
359 average community diarrhea severity scores of 1.57 (n=30, max: 8) and 2.13 (n=16, max: 13),
360 respectively. No isolates from ruminant-associated lineages caused any disease symptoms in
361 this sample population, however the total number of isolates that putatively were from a
362 ruminant background was small (n=5). Few isolates were isolated from the common generalist
363 STs that dominate clinical collections in developed countries: ST-21 clonal complex (n=3) and
364 ST-45 clonal complex (n=4). Quantitative source attribution estimated that 78.4 % (n=5, range
365 56.5 – 87.1 %) of the *C. jejuni* isolates emerged from chickens based on 5 different probability
366 estimates (**Figure S3**).

367

368 **Discussion**

369 Chronic diarrhea and malnutrition are major threats to children's health worldwide. However,
370 despite the high incidence of campylobacteriosis and reported differences in disease
371 epidemiology, there is limited understanding *Campylobacter* in LMIC's. By linking sequence
372 data with detailed clinical records from the Peruvian Amazon pediatric cohort study we were
373 able to show that variation in disease presentation was reflected in bacterial genomes,
374 specifically the source and distribution (local and global) of infecting *C. jejuni* strains.

375

376 The Peruvian Amazon pediatric isolate collection comprised a diverse assemblage of STs,
377 including common disease-causing lineages and regional STs, that have rarely been sampled
378 in Europe and the USA [47,81]. Globalization of industrialized agriculture has dispersed
379 livestock worldwide [82], broadening the geographical distribution of *C. jejuni*. We found
380 evidence of this pervasive spread with two of the three most common strains isolated in the
381 Peruvian Amazon belonging to the poultry-associated ST-353 and ST-354 complexes [47].
382 Quantitative source attribution also implicated chicken as the most likely source of infection,
383 consistent with comparable studies in Europe (**Figure S3**) [27–29,73,83].

384

385 In contrast to the profusion of poultry-associated lineages, there was a striking paucity of host
386 generalist ST-21 and ST-45 clonal complexes [40] that are among the most common disease-
387 causing lineages in Europe and North America. This has previously been observed in another
388 LMICs, with very few ST-21 complex isolates cultured in surveys from Africa, SE Asia and
389 South America [84–88]. Ruminant specialist lineages were also rare among the Peruvian
390 pediatric samples (6.1 %) and the most common cattle associated lineage (ST-61 complex [25])

391 was completely absent. This is clear evidence of different epidemiology in LMICs and
392 potentially suggests different routes to human infection.

393

394 Asymptomatic *Campylobacter* carriage represents an alternative epidemiological context to
395 that which has been the basis for most clinical studies [7,89,90]. *C. jejuni* is typically thought
396 to cause transient infection with little opportunity for human-to-human transmission. This
397 means that the human is an evolutionary dead end and the bacterium is unlikely to adapt to the
398 human host. The high prevalence, regular reinfection and prolonged colonization periods in the
399 Peruvian Amazon cohort study (and likely other LMICs) provide greater opportunity for
400 human-to-human spread and adaptation to the host [91,92]. Some studies have attempted to
401 identify signatures of human tropism, or even adaptation [93,94] and it remains possible that
402 the some of the Peruvian STs that are rarely isolated from non-human infections (**Table S7**)
403 could provide evidence of human adaptation.

404

405 One such candidate for human tropism in the Peruvian Amazon is the ST-403 complex (**Table**
406 **S7**) [76]. None of the four ST-403 isolates we sampled were associated with diarrheal
407 symptoms (**Table S1**), and according to many interpretations, attenuated virulence is often
408 associated with long-term transmission [95]. This ST has also been sampled from human
409 infections in the Dutch Antilles [96] and is a poor colonizer of avian hosts, typically lacking a
410 gene cluster (*Cj1158-1159-1160*; **Figure S1C**; **Table S9**) [76] known to be important in
411 chicken colonization [97]. However, not only was this gene cluster common in the Peruvian
412 Amazon pediatric *C. jejuni* data but also there was no clear phylogenetic distinction between
413 symptomatic and asymptomatic isolates, with multiple clonal complexes linked to
414 asymptomatic carriage. While it remains possible that analysis of larger datasets will identify

415 human adapted genomic signatures, our study suggests that host factors, such as cohabitation
416 and poor sanitation, rather than the circulation of asymptomatic lineages, may be responsible
417 for repeated or long-term infection.

418

419 While disease severity is not explained by specific lineage associations it remains possible that
420 specific molecular variations mediate virulence in the Peruvian Amazon cohort. The intimate
421 interaction of LOS and CPS with the host immune system means that the underlying genes are
422 a useful target for identifying genomic variation associated with asymptomatic carriage [61,98–
423 100]. Hypervariable genes that are common in the reference dataset included several from the
424 class C LOS and HS:2 CPS gene clusters (21 of 29 genes absent in ≥ 95 % Peruvian Amazon
425 isolates), which are absent from the Peruvian Amazon pediatric isolates [62,101]. The LOS
426 locus can be involved in the synthesis of LOS structures that mimic gangliosides, which play
427 a role in the onset of several *Campylobacter* disease sequelae, including post-infectious
428 neuropathies [76–80]. Although, there were no reports of these post-infectious neuropathies in
429 any of these cases, there were 15 Peruvian isolates possessing LOS classes (A or B) that have
430 been shown to be associated with Guillain-Barré and Miller syndromes [102–104]. Among
431 these, all of the strains with LOS class A (n=4) were from symptomatic cases, while only 2 of
432 11 strains possessing LOS class B were from symptomatic cases. It should be noted that strains
433 possessing LOS class B are not characterized by low virulence with strain 81-176 considered
434 to be a highly virulent *C. jejuni* strain. Similarly, LOS classes that produce non-sialylated LOS
435 also came from cases with differential etiology with 10 of 14 strains possessing class H from
436 symptomatic cases and one of seven class K strains from symptomatic cases (**Table S6**).

437

438 Peruvian Amazon isolates were likely to have retained the ability to glycosylate flagella
439 through genes contained in the O-linked glycosylation gene cluster (*Cj1293-1342c*), with each
440 gene present in on average 73% (range 33.3 – 100%) of Peruvian Amazon isolates (**Table S9**).
441 Large portions of the capsular polysaccharide (CPS) gene cluster appear absent from our local
442 Peruvian Amazon isolates (*Cj1421c- Cj1441c*), however the flanking regions involved in
443 capsule assembly and transport are highly conserved in our isolates (*kps* genes; **Table S9**)[77–
444 80,105]. These differences are important to characterize and take into account during vaccine
445 development for *Campylobacter*.

446

447 In conclusion, by contextualizing *C. jejuni* genomes from Peruvian Amazon children within a
448 global reference collection and linking them to clinical data on varying disease symptoms and
449 severity, we were able to identify local and globally distributed genotypes and determine the
450 major source of infection (poultry). Furthermore, we show that common asymptomatic carriage
451 is not the result of a single (or few) human adapted lineages suggesting an important role for
452 host factor in long-term infections. Genomic surveillance integrating microbial ecology with
453 population based studies in humans and animals, has considerable potential for describing
454 cryptic epidemiology and untangling complex disease transmission networks in LMICs where
455 interventions to reduce diarrheal disease are urgently needed.

456

457 **Supplementary materials** (<https://doi.org/10.6084/m9.figshare.10352375>)

458 **Supplementary Table S1:** Isolate list and disease severity scores

459 **Supplementary Table S2:** Assembly metrics and accession numbers

460 **Supplementary Table S3:** Global reference dataset details

461 **Supplementary Table S4:** Reference pan-genome gene presence

462 **Supplementary Table S5:** ABRICATE summary

463 **Supplementary Table S6:** Genome characterization

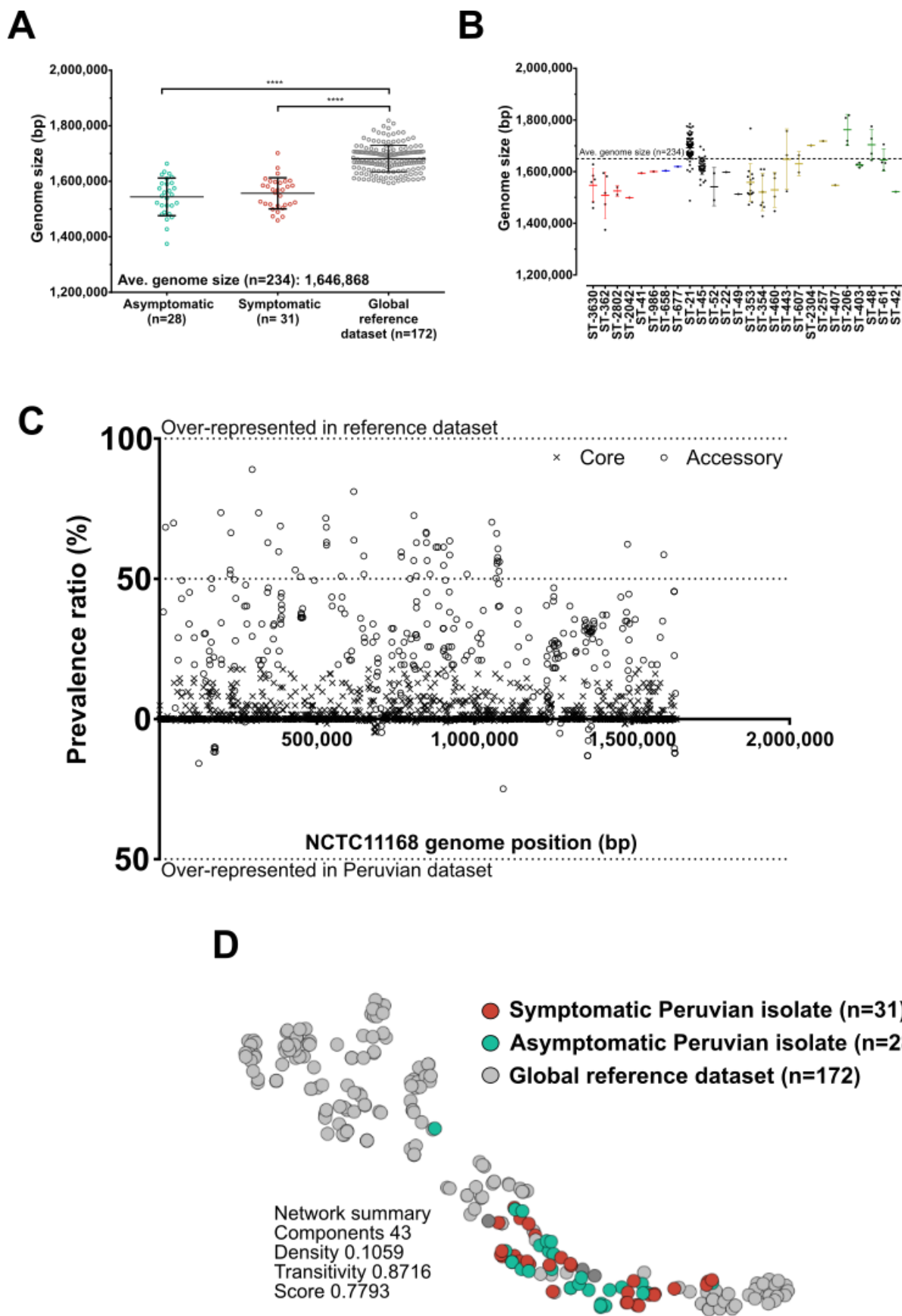
464 **Supplementary Table S7:** ST summary of pubMSLT

465 **Supplementary Table S8:** Source attribution dataset

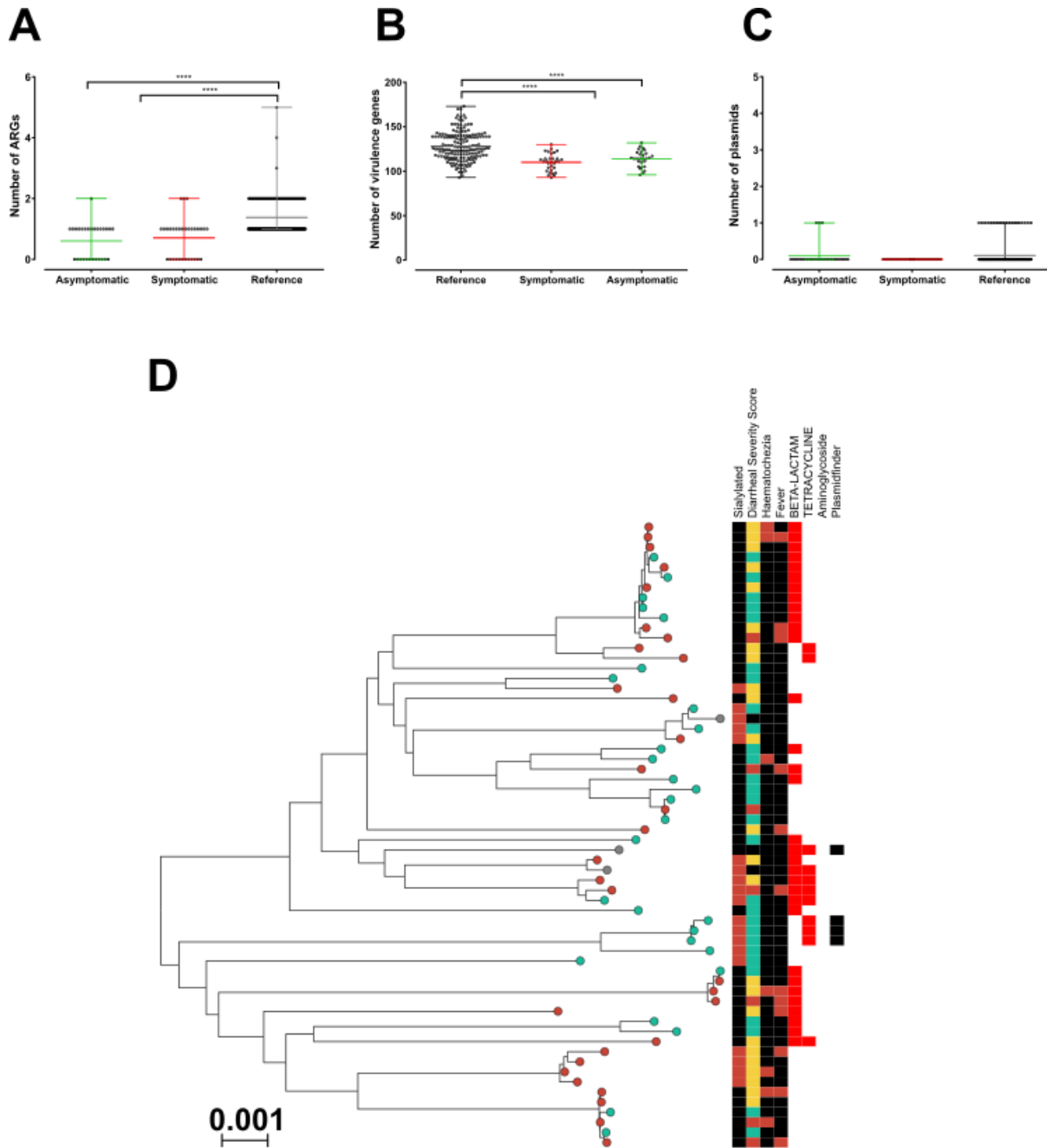
466 **Supplementary Table S9:** Comparison of NCTC11168 gene presence

467

468

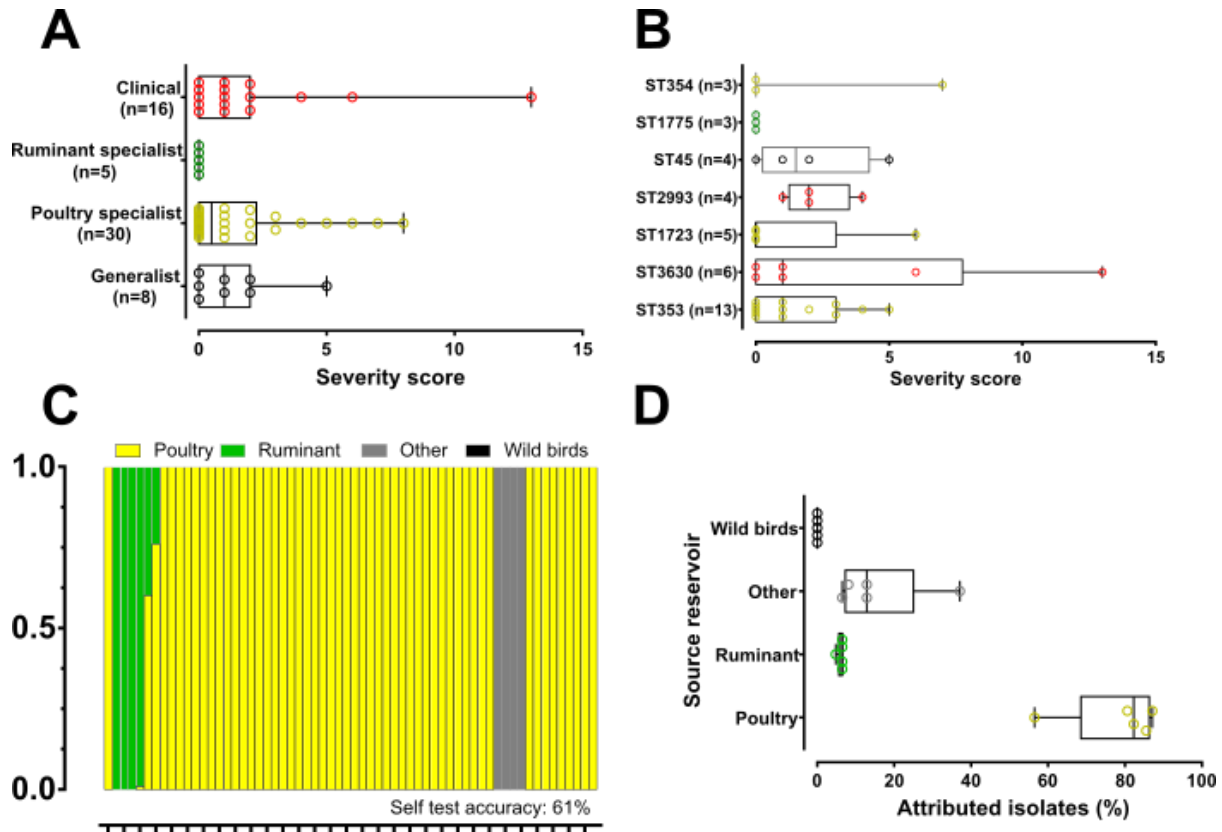


470 **Supplementary figure S1:** Genome size comparisons between **(A)** Asymptomatic (green) and
471 symptomatic (red) Peruvian isolate genomes with the reference dataset (grey); and **(B)** all
472 sequence types (ST) represented by 3 or more genomes in the dataset. Dotted line indicates the
473 average genome size for all isolates in the dataset (1,646,868 bp). **(C)** Relative presence of all
474 NCTC 11168 genes (n=1,623) in the Peruvian and reference datasets. Genes core and accessory
475 in the reference dataset are indicated by (x) and (o), respectively. Genes present more often in
476 one dataset compared to the other appear further from the mid-line. **(D)** Pairwise core and
477 accessory genome distances were compared using PopPunk for the Peruvian genomes and full
478 dataset (version 1.1.4) [70]. Clustering visualized using the t-distributed stochastic neighbor
479 embedding (t-SNE) projection of accessory distances in microreact.
480
481



482
483
484
485
486
487
488
489
490
491
492

Supplementary figure S2: Number of (A) antimicrobial resistance genes (ARGs), (B) virulence genes and (C) predicted plasmids per isolate estimated using ABRICATE (version 0.9.8; [69]). (D) Maximum-likelihood phylogeny of the Peruvian isolates only. The tree is annotated with disease severity scores, the onset of specific symptoms (hematochezia and fever), presence of AMR genes (beta-lactams, tetracyclines or aminoglycosides), identified plasmids and sialylation prediction.



493
494

495 **Supplementary figure S3:** Average disease severity score by **(A)** isolate host ecology and **(B)**
496 sequence type (represented by 3 or more isolates). **(C)** Representative source attribution of
497 Peruvian pediatric isolates using the Bayesian clustering algorithm STRUCTURE (version
498 v2.3.4, [71]). Each isolate is represented by a vertical bar colored by the estimated probability
499 that it originated from putative source reservoirs (yellow: chicken; green: ruminant; black: wild
500 bird and grey: other). **(D)** Summary box plots of predicted attribution of 62 Peruvian pediatric
501 isolates following 5 independent estimations.

502

503 **Supplementary file 1:** Pan-genome

504 **Supplementary file 2:** Alignment – Peru isolates only

505 **Supplementary file 3:** Alignment – Peru plus context isolates.

506

507 **Contributors**

508 Conceptualization: BP and SKS. Data Curation and Investigation: FS, RB, PY, MPO and MK
509 led collection of the isolates. BP, SM and MDH sequenced the isolates. Formal Analysis: BP,
510 FS, SM, SCB, GM, EM, JKC, KKC and CTP. Resources: BP, KAJ, MCJM and SKS. Original
511 Draft Preparation: BP, CTP, MK, SKS. All authors read and approved the final manuscript.

512

513 **Conflict of interest:** All authors declare that they have no conflict of interest.

514

515 **Acknowledgements**

516 SKS is funded by the Medical Research Council (MR/L015080/1) and BBSRC
517 (BB/P504750/1). Isolate sampling and collection were funded by the National Institute of
518 Health (K01-TW05717) and MK received further support from the Bill & Melinda Gates
519 Foundation (OPP1066146). Whole genome sequencing of the isolates was funded by an
520 EPSRC Bridging the Gaps award (Swansea University). All high-performance computing was
521 performed on MRC CLIMB, funded by the Medical Research Council (MR/L015080/1).
522 Sequencing of global context isolates were previously funded by the Wellcome Trust
523 (088786/C/09/Z). This publication made use of the PubMLST website (<http://pubmlst.org/>)
524 developed by Keith Jolley and Martin Maiden at the University of Oxford and funded by the
525 Wellcome Trust. No funding agency had any role in the study design, data collection and
526 analysis, decision to publish, or preparation of the manuscript.

527 **References**

- 528 1. World Health Organization. Communicable Diseases Cluster. Removing obstacles to
529 healthy development: report on infectious diseases. [Internet]. World Health
530 Organization; 1999. Available: <https://apps.who.int/iris/handle/10665/65847>
- 531 2. Högberg U. The World Health Report 2005: “Make every mother and child count” —
532 including Africans. *Scandinavian Journal of Public Health*. Scand J Public Health; 2005.
533 pp. 409–411. doi:10.1080/14034940500217037
- 534 3. Amour C, Gratz J, Mduma E, Svensen E, Rogawski ET, McGrath M, et al.
535 Epidemiology and Impact of *Campylobacter* Infection in Children in 8 Low-Resource
536 Settings: Results From the MAL-ED Study. *Clin Infect Dis*. 2016;63: 1171–1179.
537 doi:10.1093/cid/ciw542
- 538 4. Sheppard SK, Dallas JF, Strachan NJC, MacRae M, McCarthy ND, Wilson DJ, et al.
539 *Campylobacter* genotyping to determine the source of human infection. *Clin Infect Dis*.
540 2009;48: 1072–1078. doi:10.1086/597402
- 541 5. Nichols GL, Richardson JF, Sheppard SK, Lane C, Sarran C. *Campylobacter*
542 epidemiology: a descriptive study reviewing 1 million cases in England and Wales
543 between 1989 and 2011. *BMJ Open*. 2012;2: e001179. doi:10.1136/bmjopen-2012-
544 001179
- 545 6. Martin PM V, Mathiot J, Ipero J, Georges AJ, Georges-Courbot MC. Antibody response
546 to *Campylobacter coli* in children during intestinal infection and carriage. *J Clin*
547 *Microbiol*. 1988;26: 1421–1424.
- 548 7. Lee G, Pan W, Penataro Yori P, Paredes Olortegui M, Tilley D, Gregory M, et al.
549 Symptomatic and Asymptomatic *Campylobacter* Infections Associated with Reduced
550 Growth in Peruvian Children. *PLoS Negl Trop Dis*. 2013;7.
551 doi:10.1371/journal.pntd.0002036
- 552 8. Liu J, Platts-Mills JA, Juma J, Kabir F, Nkeze J, Okoi C, et al. Use of quantitative
553 molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of
554 the GEMS case-control study. *Lancet*. 2016;388: 1291–1301. doi:10.1016/S0140-
555 6736(16)31529-X
- 556 9. Lanata CF, Fischer-Walker CL, Olascoaga AC, Torres CX, Aryee MJ, Black RE. Global
557 Causes of Diarrheal Disease Mortality in Children <5 Years of Age: A Systematic
558 Review. *PLoS One*. 2013;8. doi:10.1371/journal.pone.0072788
- 559 10. Kaakoush NO, Castaño-Rodríguez N, Mitchell HM, Man SM. Global epidemiology of
560 *campylobacter* infection. *Clin Microbiol Rev*. 2015;28: 687–720.
561 doi:10.1128/CMR.00006-15
- 562 11. Coker AO, Isokpehi RD, Thomas BN, Amisu KO, Larry Obi C. Human
563 campylobacteriosis in developing countries. *Emerging Infectious Diseases*. Centers for
564 Disease Control and Prevention (CDC); 2002. pp. 237–243.
565 doi:10.3201/eid0803.010233
- 566 12. Reed RP, Friedland IR, Wegerhoff FO, Khoosal M. *Campylobacter* bacteremia in
567 children. *Pediatr Infect Dis J*. 1996;15: 345–348. doi:10.1097/00006454-199604000-
568 00012
- 569 13. Platts-Mills JA, Babji S, Bodhidatta L, Gratz J, Haque R, Havt A, et al. Pathogen-
570 specific burdens of community diarrhoea in developing countries: A multisite birth
571 cohort study (MAL-ED). *Lancet Glob Heal*. 2015;3: e564–e575. doi:10.1016/S2214-
572 109X(15)00151-5
- 573 14. Miller M, Acosta AM, Chavez CB, Flores JT, Olotegui MP, Pinedo SR, et al. The MAL-

- 574 ED study: A multinational and multidisciplinary approach to understand the relationship
575 between enteric pathogens, malnutrition, gut physiology, physical growth, cognitive
576 development, and immune responses in infants and children up to 2 years of age in
577 resource-poor environments. *Clin Infect Dis.* 2014;59: S193–S206.
578 doi:10.1093/cid/ciu653
- 579 15. Kotloff KL, Nasrin D, Blackwelder WC, Wu Y, Farag T, Panchalingham S, et al. The
580 incidence, aetiology, and adverse clinical consequences of less severe diarrhoeal
581 episodes among infants and children residing in low-income and middle-income
582 countries: a 12-month case-control study as a follow-on to the Global Enteric
583 Multicenter Study (GEMS). *Lancet Glob Heal.* 2019;7: e568–e584. doi:10.1016/S2214-
584 109X(19)30076-2
- 585 16. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, et al.
586 Burden and aetiology of diarrhoeal disease in infants and young children in developing
587 countries (the Global Enteric Multicenter Study, GEMS): A prospective, case-control
588 study. *Lancet.* 2013;382: 209–222. doi:10.1016/S0140-6736(13)60844-2
- 589 17. Pazzaglia G, Bourgeois a L, el Diwany K, Nour N, Badran N, Hablas R. *Campylobacter*
590 diarrhoea and an association of recent disease with asymptomatic shedding in Egyptian
591 children. *Epidemiol Infect.* 1991;106: 77–82. doi:10.1017/S0950268800056466
- 592 18. Georges-Courbot MC, Beraud-Cassel AM, Gouandjika I, Georges AJ. Prospective study
593 of enteric *Campylobacter* infections in children from birth to 6 months in the Central
594 African Republic. *J Clin Microbiol.* 1987;25: 836–839.
- 595 19. Figueroa G, Galeno H, Troncoso M, Toledo S, Soto V. Prospective study of
596 *Campylobacter jejuni* infection in Chilean infants evaluated by culture and serology. *J*
597 *Clin Microbiol.* 1989;27: 1040–1044.
- 598 20. Ani EA, Takahashi T, Shonekan RAO. *Campylobacter jejuni* antibodies in Nigerian
599 children. *J Clin Microbiol.* 1988;26: 605–606.
- 600 21. Calva J, Lopez-Vidal A, Ruiz-Palacios G, Ramos A, Bojalil R. Cohort study of intestinal
601 infection with *Campylobacter* in Mexican children. *Lancet.* 1988;331: 503–506.
602 doi:10.1016/S0140-6736(88)91297-4
- 603 22. Rao MR, Naficy AB, Savarino SJ, Abu-Elyazeed R, Wierzba TF, Peruski LF, et al.
604 Pathogenicity and convalescent excretion of *Campylobacter* in rural Egyptian children.
605 *Am J Epidemiol.* 2001;154: 166–173. doi:10.1093/aje/154.2.166
- 606 23. Poocharoen L, Bruin CW, Sirisanthana V, Vannareumol P, Leechanachai P, Sukhavat
607 K. The relative importance of various enteropathogens as a cause of diarrhoea in
608 hospitalized children in Chiang Mai, Thailand. *J Diarrhoeal Dis Res.* 1986;4: 10–15.
- 609 24. Sheppard SK, Guttman DS, Fitzgerald JR. Population genomics of bacterial host
610 adaptation. *Nat Rev Genet.* 2018;19: 549–565. doi:10.1038/s41576-018-0032-z
- 611 25. Mourkas E, Taylor AJA, Méric G, Bayliss SCS, Pascoe B, Mageiros L, et al.
612 Agricultural intensification and the evolution of host specialism in the enteric pathogen
613 *Campylobacter jejuni*. *Proc Natl Acad Sci.* 2020;
- 614 26. Sheppard SK, Cheng L, Méric G, De Haan CPA, Llarena AK, Marttinen P, et al. Cryptic
615 ecology among host generalist *Campylobacter jejuni* in domestic animals. *Mol Ecol.*
616 2014; doi:10.1111/mec.12742
- 617 27. Sheppard SK, Colles F, Richardson J, Cody AJ, Elson R, Lawson A, et al. Host
618 association of *Campylobacter* genotypes transcends geographic variation. *Appl Environ*
619 *Microbiol.* 2010;76: 5269–77. doi:10.1128/AEM.00124-10
- 620 28. Thépault A, Méric G, Rivoal K, Pascoe B, Mageiros L, Touzain F, et al. Genome-wide
621 identification of host-segregating epidemiological markers for source attribution in

- 622 *Campylobacter jejuni*. Appl Environ Microbiol. 2017;83. doi:10.1128/AEM.03085-16
- 623 29. Sheppard SK, Dallas JF, MacRae M, McCarthy ND, Sproston EL, Gormley FJ, et al.
- 624 *Campylobacter* genotypes from food animals, environmental sources and clinical
- 625 disease in Scotland 2005/6. Int J Food Microbiol. 2009;134: 96–103.
- 626 doi:10.1016/j.ijfoodmicro.2009.02.010
- 627 30. Revez J, Rossi M, Ellström P, de Haan C, Rautelin H, Hänninen M-L. Finnish
- 628 *Campylobacter jejuni* Strains of Multilocus Sequence Type ST-22 Complex Have Two
- 629 Lineages with Different Characteristics. Bereswill S, editor. PLoS One. 2011;6: e26880.
- 630 doi:10.1371/journal.pone.0026880
- 631 31. Heikema AP, Islam Z, Horst-Kreft D, Huizinga R, Jacobs BC, Wagenaar JA, et al.
- 632 *Campylobacter jejuni* capsular genotypes are related to Guillain-Barré syndrome. Clin
- 633 Microbiol Infect. 2015;21. doi:10.1016/j.cmi.2015.05.031
- 634 32. Nielsen LN, Sheppard SK, McCarthy ND, Maiden MCJ, Ingmer H, Krogfelt KA. MLST
- 635 clustering of *Campylobacter jejuni* isolates from patients with gastroenteritis, reactive
- 636 arthritis and Guillain-Barre syndrome. J Appl Microbiol. 2010;108: 591–599.
- 637 doi:10.1111/j.1365-2672.2009.04444.x
- 638 33. Unicomb LE, O'Reilly LC, Kirk MD, Stafford RJ, Smith H V., Becker NG, et al. Risk
- 639 factors for infection with *Campylobacter jejuni* *flaA* genotypes. Epidemiol Infect.
- 640 2008;136: 1480–1491. doi:10.1017/S0950268807000246
- 641 34. Sahin O, Fitzgerald C, Stroika S, Zhao S, Sippy RJ, Kwan P, et al. Molecular evidence
- 642 for zoonotic transmission of an emergent, highly pathogenic *Campylobacter jejuni* clone
- 643 in the United States. J Clin Microbiol. 2012;50: 680–7. doi:10.1128/JCM.06167-11
- 644 35. Kirk KF, Méric G, Nielsen HL, Pascoe B, Sheppard SK, Thorlacius-Ussing O, et al.
- 645 Molecular epidemiology and comparative genomics of *Campylobacter concisus* strains
- 646 from saliva, faeces and gut mucosal biopsies in inflammatory bowel disease. Sci Rep.
- 647 2018;8. doi:10.1038/s41598-018-20135-4
- 648 36. Pascoe B, Méric G, Yahara K, Wimalarathna H, Murray S, Hitchings MD, et al. Local
- 649 genes for local bacteria: Evidence of allopatry in the genomes of transatlantic
- 650 *Campylobacter* populations. Mol Ecol. 2017;26: 4497–4508. doi:10.1111/mec.14176
- 651 37. de Haan CPA, Kivisto R, Hakkinen M, Rautelin H, Hanninen ML. Decreasing Trend of
- 652 Overlapping Multilocus Sequence Types between Human and Chicken *Campylobacter*
- 653 *jejuni* Isolates over a Decade in Finland. Appl Environ Microbiol. 2010;76: 5228–5236.
- 654 doi:10.1128/AEM.00581-10
- 655 38. Asakura H, Brüggemann H, Sheppard SK, Ekawa T, Meyer TF, Yamamoto S, et al.
- 656 Molecular Evidence for the Thriving of *Campylobacter jejuni* ST-4526 in Japan.
- 657 Bereswill S, editor. PLoS One. 2012;7: e48394. doi:10.1371/journal.pone.0048394
- 658 39. Llarena AK, Zhang J, Vehkala M, Välimäki N, Hakkinen M, Hänninen ML, et al.
- 659 Monomorphic genotypes within a generalist lineage of *Campylobacter jejuni* show signs
- 660 of global dispersion. Microb genomics. 2016;2: e000088. doi:10.1099/mgen.0.000088
- 661 40. Méric G, McNally A, Pessia A, Mourkas E, Pascoe B, Mageiros L, et al. Convergent
- 662 Amino Acid Signatures in Polyphyletic *Campylobacter jejuni* Subpopulations Suggest
- 663 Human Niche Tropism. Genome Biol Evol. 2018;10: 763–774. doi:10.1093/gbe/evy026
- 664 41. Schiaffino F, Colston JM, Paredes-Olortegui M, François R, Pisanic N, Burga R, et al.
- 665 Antibiotic resistance of *Campylobacter* species in a pediatric cohort study. Antimicrob
- 666 Agents Chemother. 2019;63. doi:10.1128/AAC.01911-18
- 667 42. Rojas JD, Reynolds ND, Pike BL, Espinoza NM, Kuroiwa J, Jani V, et al. Distribution
- 668 of Capsular Types of *Campylobacter jejuni* Isolates from Symptomatic and
- 669 Asymptomatic Children in Peru. Am J Trop Med Hyg. 2019;101: 541–548.

- 670 doi:10.4269/ajtmh.18-0994
- 671 43. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de
672 Bruijn graphs. *Genome Res.* 2008; doi:10.1101/gr.074492.107
- 673 44. Pascoe B, Méric G, Murray S, Yahara K, Mageiros L, Bowen R, et al. Enhanced biofilm
674 formation and multi-host transmission evolve from divergent genetic backgrounds in
675 *Campylobacter jejuni*. *Environ Microbiol.* 2015;17: 4779–4789. doi:10.1111/1462-
676 2920.13051
- 677 45. Sheppard SK, Didelot X, Méric G, Torralbo A, Jolley KA, Kelly DJ, et al. Genome-
678 wide association study identifies vitamin B5 biosynthesis as a host specificity factor in
679 *Campylobacter*. *Proc Natl Acad Sci.* 2013;110: 11923–11927.
680 doi:10.1073/pnas.1305559110
- 681 46. Yahara K, Méric G, Taylor AJ, de Vries SPW, Murray S, Pascoe B, et al. Genome-wide
682 association of functional traits linked with *Campylobacter jejuni* survival from farm to
683 fork. *Environ Microbiol.* 2017;19. doi:10.1111/1462-2920.13628
- 684 47. Cody AJ, McCarthy NM, Wimalarathna HL, Colles FM, Clark L, Bowler ICJWJW, et
685 al. A Longitudinal 6-Year Study of the Molecular Epidemiology of Clinical
686 *Campylobacter* Isolates in Oxfordshire, United Kingdom. *J Clin Microbiol.* 2012;50:
687 3193–3201. doi:10.1128/JCM.01086-12
- 688 48. Lee G, Yori PP, Olortegui MP, Caulfield LE, Sack DA, Fischer-Walker C, et al. An
689 instrument for the assessment of diarrhoeal severity based on a longitudinal community-
690 based study. *BMJ Open.* 2014;4. doi:10.1136/bmjopen-2014-004816
- 691 49. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST
692 Server: rapid annotations using subsystems technology. *BMC Genomics.* 2008;9: 75.
693 doi:10.1186/1471-2164-9-75
- 694 50. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the
695 Rapid Annotation of microbial genomes using Subsystems Technology (RAST).
696 *Nucleic Acids Res.* 2014;42: D206–14. doi:10.1093/nar/gkt1226
- 697 51. Méric G, Yahara K, Mageiros L, Pascoe B, Maiden MCJ, Jolley KA, et al. A reference
698 pan-genome approach to comparative bacterial genomics: Identification of novel
699 epidemiological markers in pathogenic *Campylobacter*. *PLoS One.* 2014;9.
700 doi:10.1371/journal.pone.0092798
- 701 52. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7:
702 Improvements in Performance and Usability. *Mol Biol Evol.* 2013;30: 772–780.
703 doi:10.1093/molbev/mst010
- 704 53. Sheppard SK, Jolley KA, Maiden MCJ. A gene-by-gene approach to bacterial
705 population genomics: Whole genome MLST of *Campylobacter*. *Genes (Basel).* 2012;3:
706 261–277. doi:10.3390/genes3020261
- 707 54. Méric G, Yahara K, Mageiros L, Pascoe B, Maiden MCJ, Jolley KA, et al. A reference
708 pan-genome approach to comparative bacterial genomics: Identification of novel
709 epidemiological markers in pathogenic *Campylobacter*. *PLoS One.* 2014;9.
710 doi:10.1371/journal.pone.0092798
- 711 55. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective
712 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol.*
713 2015;32: 268–274. doi:10.1093/molbev/msu300
- 714 56. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving
715 the Ultrafast Bootstrap Approximation. *Mol Biol Evol.* 2018;35: 518–522.
716 doi:10.1093/molbev/msx281
- 717 57. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, et al. Microreact:

- 718 visualizing and sharing data for genomic epidemiology and phylogeography. *Microb*
719 *genomics*. 2016;2: e000093. doi:10.1099/mgen.0.000093
- 720 58. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb
721 software, the PubMLST.org website and their applications [version 1; referees: 2
722 approved]. *Wellcome Open Res*. 2018;3. doi:10.12688/wellcomeopenres.14826.1
- 723 59. Jolley KA, Maiden MCJ. BIGSdb: Scalable analysis of bacterial genome variation at the
724 population level. *BMC Bioinformatics*. 2010;11: 595. doi:10.1186/1471-2105-11-595
- 725 60. Dingle KE, Colles FM, Falush D, Maiden MCJ. Sequence typing and comparison of
726 population biology of *Campylobacter coli* and *Campylobacter jejuni*. *J Clin Microbiol*.
727 2005;43: 340–347. doi:10.1128/JCM.43.1.340-347.2005
- 728 61. Parker CT, Gilbert M, Yuki N, Endtz HP, Mandrell RE. Characterization of
729 lipooligosaccharide-biosynthetic loci of *Campylobacter jejuni* reveals new
730 lipooligosaccharide classes: Evidence of mosaic organizations. *J Bacteriol*. 2008;190:
731 5681–5689. doi:10.1128/JB.00254-08
- 732 62. Culebro A, Revez J, Pascoe B, Friedmann Y, Hitchings MDMD, Stupak J, et al. Large
733 sequence diversity within the biosynthesis locus and common biochemical features of
734 *Campylobacter coli* lipooligosaccharides. DiRita VJ, editor. *J Bacteriol*. 2016;198:
735 2829–40. doi:10.1128/JB.00347-16
- 736 63. Grundmann H, Hori S, Tanner G. Determining confidence intervals when measuring
737 genetic diversity and the discriminatory abilities of typing methods for microorganisms.
738 *J Clin Microbiol*. 2001;39: 4190–4192. doi:10.1128/JCM.39.11.4190-4192.2001
- 739 64. Carattoli A, Zankari E, Garcíá-Fernández A, Larsen MV, Lund O, Villa L, et al. *In Silico*
740 detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence
741 typing. *Antimicrob Agents Chemother*. 2014;58: 3895–3903. doi:10.1128/AAC.02412-
742 14
- 743 65. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: Hierarchical and refined dataset
744 for big data analysis - 10 years on. *Nucleic Acids Res*. 2016;44: D694–D697.
745 doi:10.1093/nar/gkv1239
- 746 66. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, et al. Using the
747 NCBI AMRFinder Tool to Determine Antimicrobial Resistance Genotype-Phenotype
748 Correlations Within a Collection of NARMS Isolates. *bioRxiv*. 2019; 550707.
749 doi:10.1101/550707
- 750 67. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al.
751 Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother*.
752 2012;67: 2640–2644. doi:10.1093/jac/dks261
- 753 68. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al.
754 CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic
755 resistance database. *Nucleic Acids Res*. 2019; doi:10.1093/nar/gkz935
- 756 69. Seemann T. ABRicate: Mass screening of contigs for antimicrobial and virulence genes
757 [Internet]. GitHub repository. 2018. Available: <https://github.com/tseemann/abricate>
- 758 70. Lees JA, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, Weiser JN, et al. Fast and
759 flexible bacterial genomic epidemiology with PopPUNK. *Genome Res*. 2019;29: 304–
760 316. doi:10.1101/gr.241455.118
- 761 71. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using
762 multilocus genotype data. *Genetics*. 2000;
- 763 72. Dearlove BL, Cody AJ, Pascoe B, Méric G, Wilson DJ, Sheppard SK. Rapid host
764 switching in generalist *Campylobacter* strains erodes the signal for tracing human
765 infections. *ISME J*. 2016;10. doi:10.1038/ismej.2015.149

- 766 73. Cody AJ, Maiden MC, Strachan NJ, McCarthy ND. A systematic review of source
767 attribution of human campylobacteriosis using multilocus sequence typing.
768 Eurosurveillance. 2019;24. doi:10.2807/1560-7917.ES.2019.24.43.1800696
- 769 74. Baily JL, Méric G, Bayliss S, Foster G, Moss SE, Watson E, et al. Evidence of land-sea
770 transfer of the zoonotic pathogen *Campylobacter* to a wildlife marine sentinel species.
771 Mol Ecol. 2015;24. doi:10.1111/mec.13001
- 772 75. Weinert LA, Chaudhuri RR, Wang J, Peters SE, Corander J, Jombart T, et al. Genomic
773 signatures of human and animal disease in the zoonotic pathogen *Streptococcus suis*.
774 Nat Commun. 2015;6: 6740. doi:10.1038/ncomms7740
- 775 76. Morley L, McNally A, Paszkiewicz K, Corander J, Méric G, Sheppard SK, et al. Gene
776 loss and lineage-specific restriction-modification systems associated with niche
777 differentiation in the *Campylobacter jejuni* sequence type 403 clonal complex. Appl
778 Environ Microbiol. 2015; doi:10.1128/AEM.00546-15
- 779 77. Karlyshev A V., Champion OL, Churcher C, Brisson JR, Jarrell HC, Gilbert M, et al.
780 Analysis of *Campylobacter jejuni* capsular loci reveals multiple mechanisms for the
781 generation of structural diversity and the ability to form complex heptoses. Mol
782 Microbiol. 2005;55: 90–103. doi:10.1111/j.1365-2958.2004.04374.x
- 783 78. Parker CT, Gilbert M, Yuki N, Endtz HP, Mandrell RE. Characterization of
784 lipooligosaccharide-biosynthetic loci of *Campylobacter jejuni* reveals new
785 lipooligosaccharide classes: Evidence of mosaic organizations. J Bacteriol. 2008;190:
786 5681–5689. doi:10.1128/JB.00254-08
- 787 79. Parker CT, Horn ST, Gilbert M, Miller WG, Woodward DL, Mandrell RE. Comparison
788 of *Campylobacter jejuni* lipooligosaccharide biosynthesis loci from a variety of sources.
789 J Clin Microbiol. 2005;43: 2771–2781. doi:10.1128/JCM.43.6.2771-2781.2005
- 790 80. Poly F, Serichantalergs O, Kuroiwa J, Pootong P, Mason C, Guerry P, et al. Updated
791 *Campylobacter jejuni* Capsule PCR Multiplex Typing System and Its Application to
792 Clinical Isolates from South and Southeast Asia. Skurnik M, editor. PLoS One. 2015;10:
793 e0144349. doi:10.1371/journal.pone.0144349
- 794 81. Dunn SJ, Pascoe B, Turton J, Fleming V, Diggle M, Sheppard SK, et al. Genomic
795 epidemiology of clinical *Campylobacter spp.* at a single health trust site. Microb
796 Genomics. 2018; doi:10.1099/mgen.0.000227
- 797 82. Mottet A, Tempio G. Global poultry production: Current state and future outlook and
798 challenges. World's Poultry Science Journal. Cambridge University Press; 2017. pp.
799 245–256. doi:10.1017/S0043933917000071
- 800 83. Wilson DJ, Gabriel E, Leatherbarrow AJH, Cheesbrough J, Gee S, Bolton E, et al.
801 Tracing the Source of Campylobacteriosis. Guttman DS, editor. PLoS Genet. 2008;4:
802 e1000203. doi:10.1371/journal.pgen.1000203
- 803 84. Prachantasena S, Charunontakorn P, Muangnoicharoen S, Hankla L, Techawal N,
804 Chaveerach P, et al. Distribution and genetic profiles of *Campylobacter* in commercial
805 broiler production from breeder to slaughter in Thailand. PLoS One. 2016;11.
806 doi:10.1371/journal.pone.0149585
- 807 85. Ngulukun S, Oboegbulem S, Klein G. Multilocus sequence typing of *Campylobacter*
808 *jejuni* and *Campylobacter coli* isolates from poultry, cattle and humans in Nigeria. J
809 Appl Microbiol. 2016;121: 561–568. doi:10.1111/jam.13185
- 810 86. Duong VT, Tuyen HT, Van Minh P, Campbell JI, Le Phuc H, Nhu TDH, et al. No
811 Clinical benefit of empirical antimicrobial therapy for pediatric diarrhea in a high-usage,
812 high-resistance setting. Clin Infect Dis. 2018;66: 504–511. doi:10.1093/cid/cix844
- 813 87. Mason J, Iturriza-Gomara M, O'Brien SJ, Ngwira BM, Dove W, Maiden MCJ, et al.

- 814 Campylobacter Infection in Children in Malawi Is Common and Is Frequently
815 Associated with Enteric Virus Co-Infections. Hold GL, editor. PLoS One. 2013;8:
816 e59663. doi:10.1371/journal.pone.0059663
- 817 88. de Vries SPW, Vurayai M, Holmes M, Gupta S, Bateman M, Goldfarb D, et al.
818 Phylogenetic analyses and antimicrobial resistance profiles of *Campylobacter spp.* from
819 diarrhoeal patients and chickens in Botswana. Chang Y-F, editor. PLoS One. 2018;13:
820 e0194481. doi:10.1371/journal.pone.0194481
- 821 89. Acheson D, Allos BM. Campylobacter jejuni Infections: Update on Emerging Issues
822 and Trends. Clin Infect Dis. 2001;32: 1201–1206. doi:10.1086/319760
- 823 90. Toledo Z, Simaluiza RJ, Astudillo X, Fernández H. Occurrence and antimicrobial
824 susceptibility of thermophilic *Campylobacter* species isolated from healthy children
825 attending municipal care centers in Southern Ecuador. Rev Inst Med Trop Sao Paulo.
826 2017;59. doi:10.1590/S1678-9946201759077
- 827 91. Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. Within-host evolution of
828 bacterial pathogens. Nature Reviews Microbiology. Nature Publishing Group; 2016. pp.
829 150–162. doi:10.1038/nrmicro.2015.13
- 830 92. Martins NE, Faria VG, Teixeira L, Magalhães S, Sucena É. Host Adaptation Is
831 Contingent upon the Infection Route Taken by Pathogens. PLoS Pathog. 2013;9.
832 doi:10.1371/journal.ppat.1003601
- 833 93. Buchanan CJ, Webb AL, Mutschall SK, Kruczkiewicz P, Barker DOR, Hetman BM, et
834 al. A genome-wide association study to identify diagnostic markers for human
835 pathogenic *Campylobacter jejuni* strains. Front Microbiol. 2017;
836 doi:10.3389/fmicb.2017.01224
- 837 94. Thépault A, Rose V, Quesne S, Poezevara T, Béven V, Hirchaud E, et al. Ruminant and
838 chicken: Important sources of campylobacteriosis in France despite a variation of source
839 attribution in 2009 and 2015. Sci Rep. 2018;
- 840 95. Cressler CE, McLeod D V., Rozins C, Van Den Hoogen J, Day T. The adaptive
841 evolution of virulence: A review of theoretical predictions and empirical tests.
842 Parasitology. Cambridge University Press; 2016. pp. 915–930.
843 doi:10.1017/S003118201500092X
- 844 96. Duim B, Godschalk PCR, Van Den Braak N, Dingle KE, Dijkstra JR, Leyde E, et al.
845 Molecular Evidence for Dissemination of Unique *Campylobacter jejuni* Clones in
846 Curaçao, Netherlands Antilles. J Clin Microbiol. 2003;41: 5593–5597.
847 doi:10.1128/JCM.41.12.5593-5597.2003
- 848 97. Taveirne ME, Theriot CM, Livny J, DiRita VJ. The Complete *Campylobacter jejuni*
849 Transcriptome during Colonization of a Natural Host Determined by RNAseq. PLoS
850 One. 2013;8. doi:10.1371/journal.pone.0073586
- 851 98. Guerry P, Ewing CP, Hickey TE, Prendergast MM, Moran AP. Sialylation of
852 lipooligosaccharide cores affects immunogenicity and serum resistance of
853 *Campylobacter jejuni*. Infect Immun. 2000;68: 6656–62. Available:
854 <http://www.ncbi.nlm.nih.gov/pubmed/11083778>
- 855 99. Zebian N, Merckx-Jacques A, Pittock PP, Houle S, Dozois CM, Lajoie GA, et al.
856 Comprehensive analysis of flagellin glycosylation in *Campylobacter jejuni* NCTC
857 11168 reveals incorporation of legionaminic acid and its importance for host
858 colonization. Glycobiology. 2016;26: 386–397. doi:10.1093/glycob/cwv104
- 859 100. Guerry P, Ewing CP, Hickey TE, Prendergast MM, Moran AP. Sialylation of
860 lipooligosaccharide cores affects immunogenicity and serum resistance of
861 *Campylobacter jejuni*. Infect Immun. 2000;68: 6656–62. doi:10.1128/IAI.68.12.6656-

- 862 6662.2000
863 101. Pascoe B, Williams LK, Calland JK, Meric G, Hitchings MD, Dyer M, et al.
864 Domestication of *Campylobacter jejuni* NCTC 11168. Microb genomics. 2019;5.
865 doi:10.1099/mgen.0.000279
866 102. Gilbert M, Karwaski MF, Bernatchez S, Young NM, Taboada E, Michniewicz J, et al.
867 The genetic bases for the variation in the lipo-oligosaccharide of the mucosal pathogen,
868 *Campylobacter jejuni*. Biosynthesis of sialylated ganglioside mimics in the core
869 oligosaccharide. J Biol Chem. 2002;277: 327–337. doi:10.1074/jbc.M108452200
870 103. Houliston RS, Vinogradov E, Dzieciatkowska M, Li J, St Michael F, Karwaski MF, et
871 al. Lipooligosaccharide of *Campylobacter jejuni*: Similarity with multiple types of
872 mammalian glycans beyond gangliosides. J Biol Chem. 2011;286: 12361–12370.
873 doi:10.1074/jbc.M110.181750
874 104. Godschalk PCR, Kuijf ML, Li J, St. Michael F, Ang CW, Jacobs BC, et al. Structural
875 characterization of *Campylobacter jejuni* lipooligosaccharide outer cores associated
876 with Guillain-Barré and Miller Fisher syndromes. Infect Immun. 2007;75: 1245–1254.
877 doi:10.1128/IAI.00872-06
878 105. Guerry P, Poly F, Riddle M, Maue AC, Chen Y-H, Monteiro MA, et al. CELLULAR
879 AND INFECTION MICROBIOLOGY *Campylobacter* polysaccharide capsules:
880 virulence and vaccines. 2012; doi:10.3389/fcimb.2012.00007
881