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

Ben Pascoe[1,2*], Francesca Schiaffino[3,4], Susan Murray[5,6], Guillaume Méric[1#], Sion C. Bayliss[1], Matthew D. Hitchings[5], Evangelos Mourkas[1], Jessica K. Calland[1], Rosa Burga[7], Pablo Peñataro Yori[8,9], Keith A. Jolley[10], Kerry K. Cooper[11], Craig T. Parker[12], Maribel Paredes Olortegui[9], Margaret N. Kosek[8,9*] and Samuel K. Sheppard[1,2,10]

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

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

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

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

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

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

Abstract

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

Author summary

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

Introduction

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

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

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

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

Methods

Sampling and cohort information

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

Bacterial isolate genome sequencing

Genomic DNA was extracted from 62 *C. jejuni* isolates and sequenced using an Illumina MiSeq benchtop sequencer (California, USA). Nextera XT libraries (Illumina, California, USA) were prepared and short paired-end reads (250 bp) were assembled *de novo* using Velvet (version 1.2.08) [43] with VelvetOptimiser (version 2.2.4). The average number of contiguous sequences (contigs) was 262 (range: 53–701) for an average total assembled sequence size of 1.55 Mbp (range: 1.37–1.70). The average N50 contig length (L50) was 14,577 (range: 3,794-55,912) and the average GC content was 30.8 % (range: 30.5-31.6). Short read data are available on the NCBI SRA, associated with BioProject PRJNA350267. Assembled genomes and supplementary material are available from FigShare (doi:10.6084/m9.figshare.10352375; individual accession numbers and assembled genome statistics in **Table S2**). Isolates were compared to a global reference dataset representing the genetic diversity of the species (n=164 isolates from eight countries and three continents) (**Table S3**)[26,36,44–47].

Diarrheal disease severity

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

Core genome genealogies

A reference pan-genome file was constructed by combining open reading frames identified by RAST [49,50] in all the Peruvian isolates and the *C. jejuni* NCTC 11168 reference strain to maintain locus nomenclature [51]. Gene orthologues (≥70% sequence similarity) were identified and duplicates removed (size: 2,045,739 bp; **Supplementary file S1**). Two alignment files were constructed from concatenated gene sequences of all core genes (found in ≥95 % isolates) from the reference pan-genome list using MAFFT [52] on a gene-by-gene basis [53,54]: one for the Peruvian isolates only (size: 772,794 bp; **Supplementary file S2**); and a second alignment containing the Peruvian isolates plus the global reference collection (size: 720,853 bp; **Supplementary file S3**). Maximum-likelihood phylogenies were constructed in IQ-TREE (version 1.6.8) using the GTR+F+I+G4 substitution model and ultra-fast bootstrapping (1,000 bootstraps)[55,56]; and visualized on Microreact [57]: Peru only (https://microreact.org/project/CampyPeruOnly); Peru and the global reference dataset (https://microreact.org/project/CampyPeruContext).

Molecular typing and diversity estimates

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

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

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

Accessory genome characterization

The reference pan-genome list contained 2,348 genes, of which 1,321 genes were shared by all isolates (≥95 %) and defined as the core genome (**Table S4**). The accessory genomes of each isolate was characterized, including detection of antimicrobial resistance genes, putative virulence factors and known plasmid genes using ABRICATE (version 0.9.8) and the CARD, NCBI, ResFinder, VfDB and PlasmidFinder databases (10th September, 2019 update; **Table S5** and summarized in **Table S6**) [64–69]. Pairwise core and accessory genome distances were compared using PopPunk (version 1.1.4). PopPUNK uses pairwise nucleotide k-mer comparisons to distinguish shared sequence and gene content to identify divergence of the accessory genome in relation to the core genome. A two-component Gaussian mixture model was used to construct a network to define clusters (Components: 43; Density: 0.1059; Transitivity: 0.8716; Score: 0.7793) [70].

Source attribution

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

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

Results

Globally circulating disease genotypes are found in the Peruvian Amazon

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

Proliferation of globally rare genotypes in Peruvian Amazon children

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

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

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

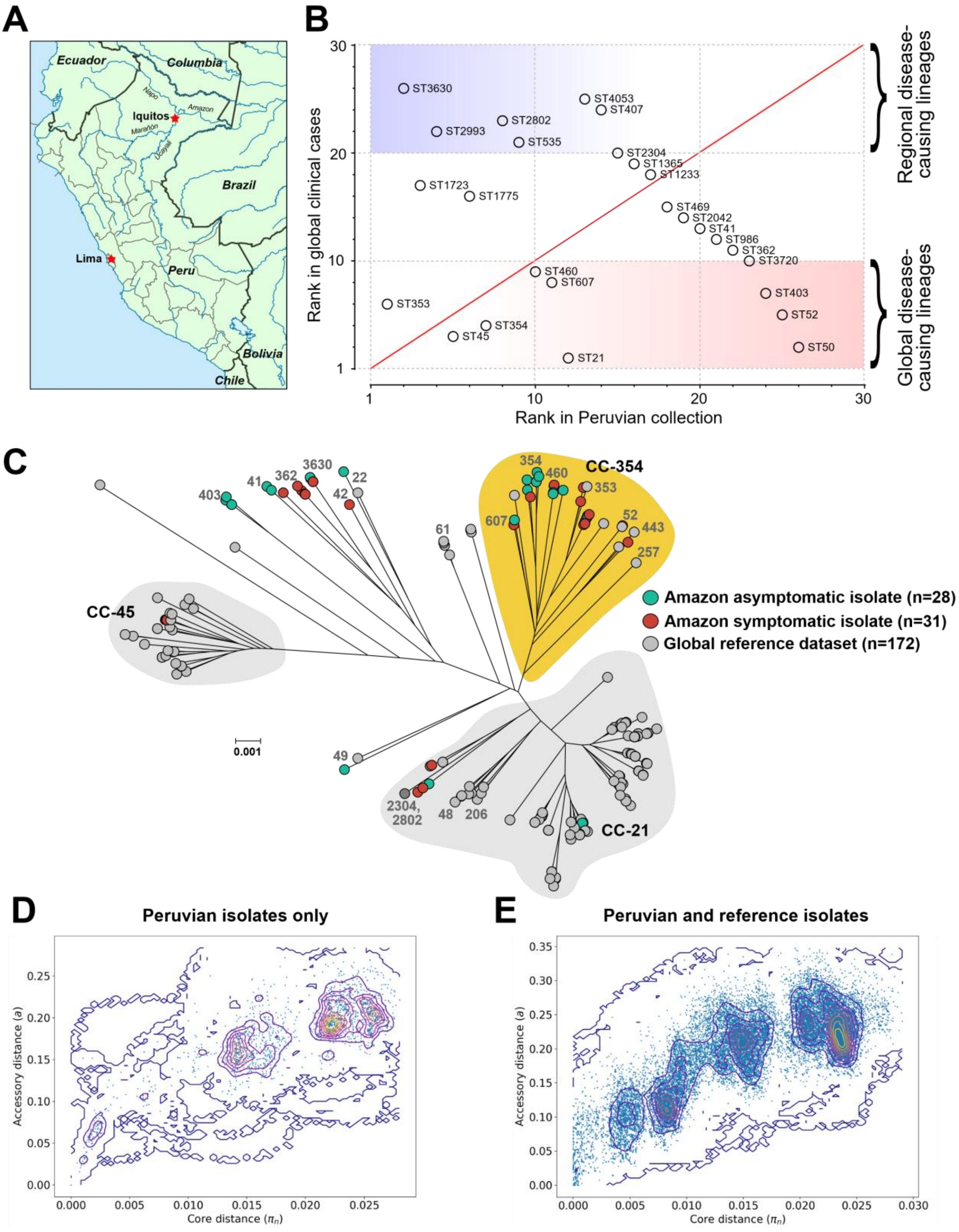

A
Ecuador
Columbia
Napo
Amazon
Iquitos
Marañón
Ucayali
Brazil
Lima
Peru
Bolivia
Chile
B
Rank in global clinical cases
Rank in Peruvian collection
ST3630
ST4053
ST407
ST2802
ST2993
ST535
ST2304
ST1365
ST1233
ST1723
ST1775
ST469
ST2042
ST41
ST986
ST362
ST3720
ST460
ST607
ST403
ST353
ST52
ST354
ST45
ST50
ST21
Regional disease-causing lineages
Global disease-causing lineages
C
CC-354
CC-45
CC-21
354
460
353
607
52
443
257
3630
362
41
22
42
403
61
49
2304, 2802
48
206
0.001
Amazon asymptomatic isolate (n=28)
Amazon symptomatic isolate (n=31)
Global reference dataset (n=172)
D
Peruvian isolates only
E
Peruvian and reference isolates
Accessory distance (a)
Core distance (πn)

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

Peruvian Amazon pediatric isolates have a local gene pool

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

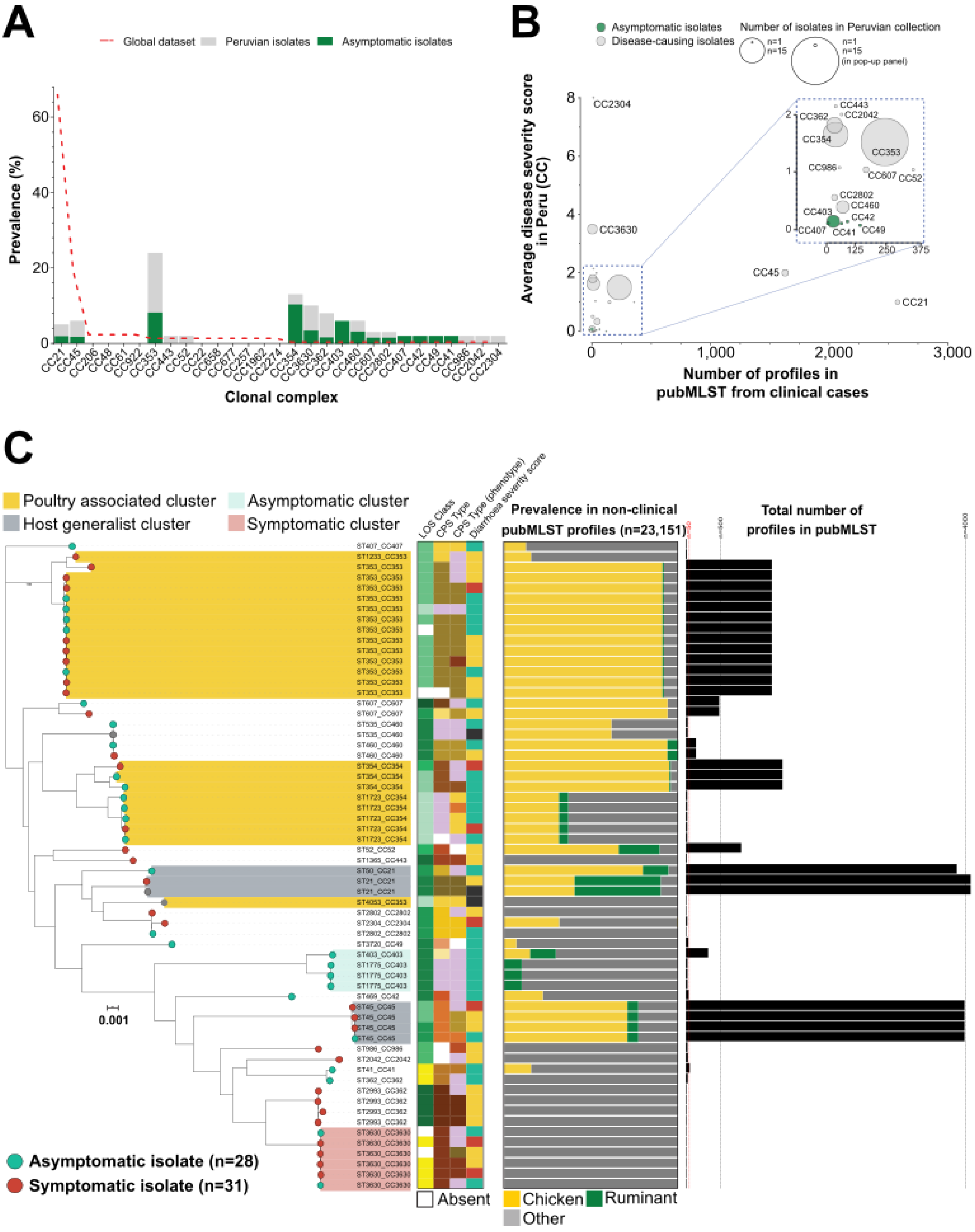

A
Global dataset
Peruvian isolates
Asymptomatic isolates
Prevalence (%)
0
20
40
60
CC21
CC45
CC206
CC48
CC61
CC922
CC353
CC443
CC52
CC22
CC658
CC677
CC257
CC1962
CC2274
CC354
CC3630
CC362
CC403
CC460
CC607
CC2802
CC407
CC42
CC49
CC41
CC986
CC2042
CC2304
Clonal complex
B
Asymptomatic isolates
Disease-causing isolates
Number of isolates in Peruvian collection
n=1
n=15
n=1
n=15
(in pop-up panel)
Average disease severity score in Peru (CC)
0
2
4
6
8
CC2304
CC3630
CC45
CC21
CC443
CC362
CC2042
CC354
CC353
CC986
CC607
CC52
CC2802
CC460
CC403
CC42
CC407
CC41
CC49
0
1
2
0
125
250
375
0
1,000
2,000
3,000
Number of profiles in pubMLST from clinical cases
C
Poultry associated cluster
Asymptomatic cluster
Host generalist cluster
Symptomatic cluster
LOS Class
CPS Type
CPS Type (phenotype)
Diarrhoea severity score
Prevalence in non-clinical pubMLST profiles (n=23,151)
Total number of profiles in pubMLST
0.001
Asymptomatic isolate (n=28)
Symptomatic isolate (n=31)
Absent
Chicken
Ruminant
Other
ST407_CC407
ST1233_CC353
ST353_CC353
ST607_CC607
ST535_CC460
ST460_CC460
ST354_CC354
ST1723_CC354
ST52_CC52
ST1365_CC443
ST50_CC21
ST21_CC21
ST4033_CC353
ST2802_CC2802
ST2304_CC2304
ST3720_CC49
ST403_CC403
ST1775_CC403
ST469_CC42
ST45_CC45
ST986_CC986
ST2042_CC2042
ST41_CC41
ST362_CC362
ST2993_CC362
ST3630_CC3630

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

Lineages associated with asymptomatic infection in Peruvian Amazon pediatric cases

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

Regional differences in accessory genome content

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

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

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

Poultry is the predominant source of infection in Peruvian Amazon children

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

Discussion

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

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

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

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

One such candidate for human tropism in the Peruvian Amazon is the ST-403 complex (**Table S7**) [76]. None of the four ST-403 isolates we sampled were associated with diarrheal symptoms (**Table S1**), and according to many interpretations, attenuated virulence is often associated with long-term transmission [95]. This ST has also been sampled from human infections in the Dutch Antilles [96] and is a poor colonizer of avian hosts, typically lacking a gene cluster (*Cj1158-1159-1160*; **Figure S1C**; **Table S9**) [76] known to be important in chicken colonization [97]. However, not only was this gene cluster common in the Peruvian Amazon pediatric *C. jejuni* data but also there was no clear phylogenetic distinction between symptomatic and asymptomatic isolates, with multiple clonal complexes linked to asymptomatic carriage. While it remains possible that analysis of larger datasets will identify human adapted genomic signatures, our study suggests that host factors, such as cohabitation and poor sanitation, rather than the circulation of asymptomatic lineages, may be responsible for repeated or long-term infection.

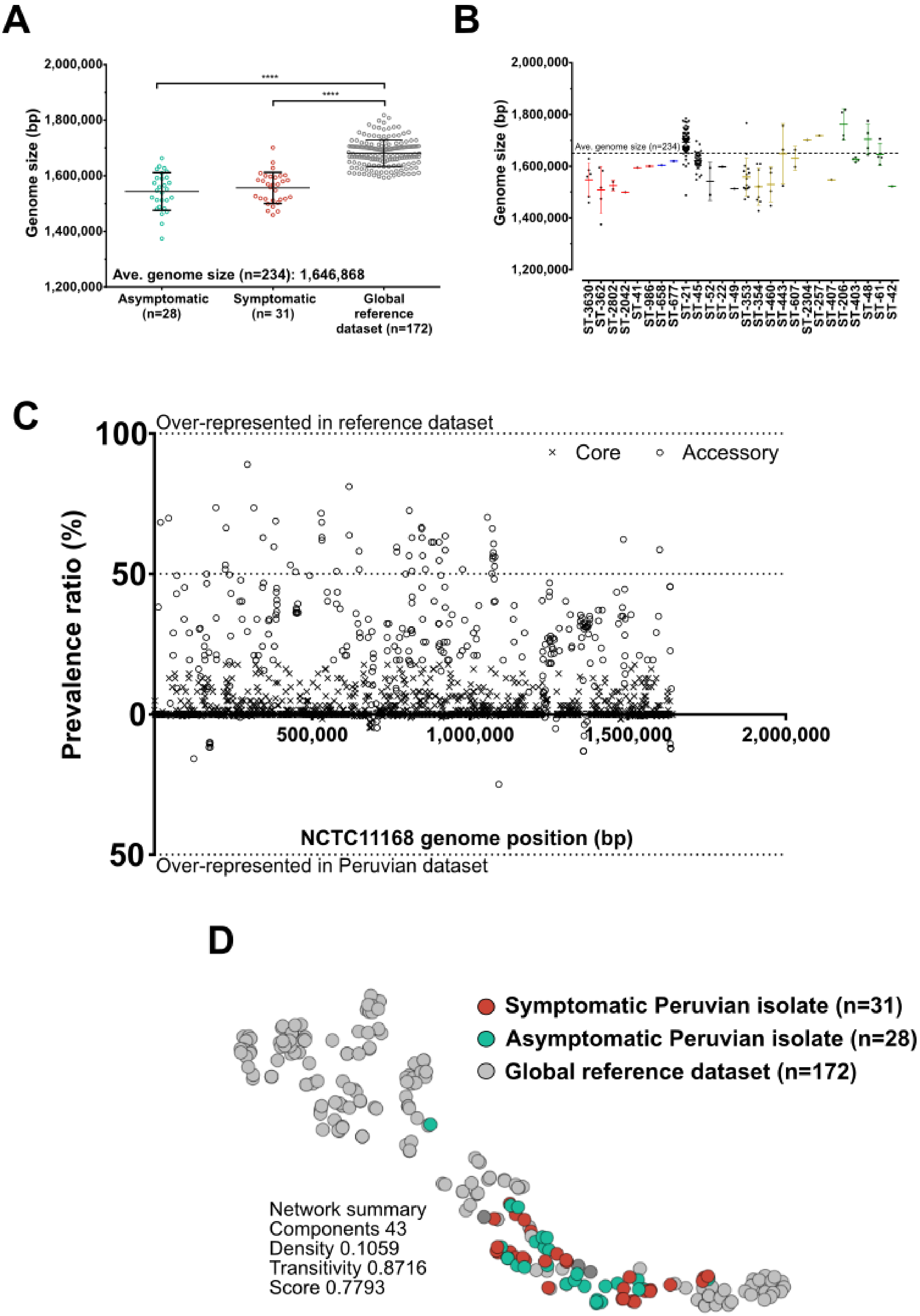
While disease severity is not explained by specific lineage associations it remains possible that specific molecular variations mediate virulence in the Peruvian Amazon cohort. The intimate interaction of LOS and CPS with the host immune system means that the underling genes are a useful target for identifying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

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

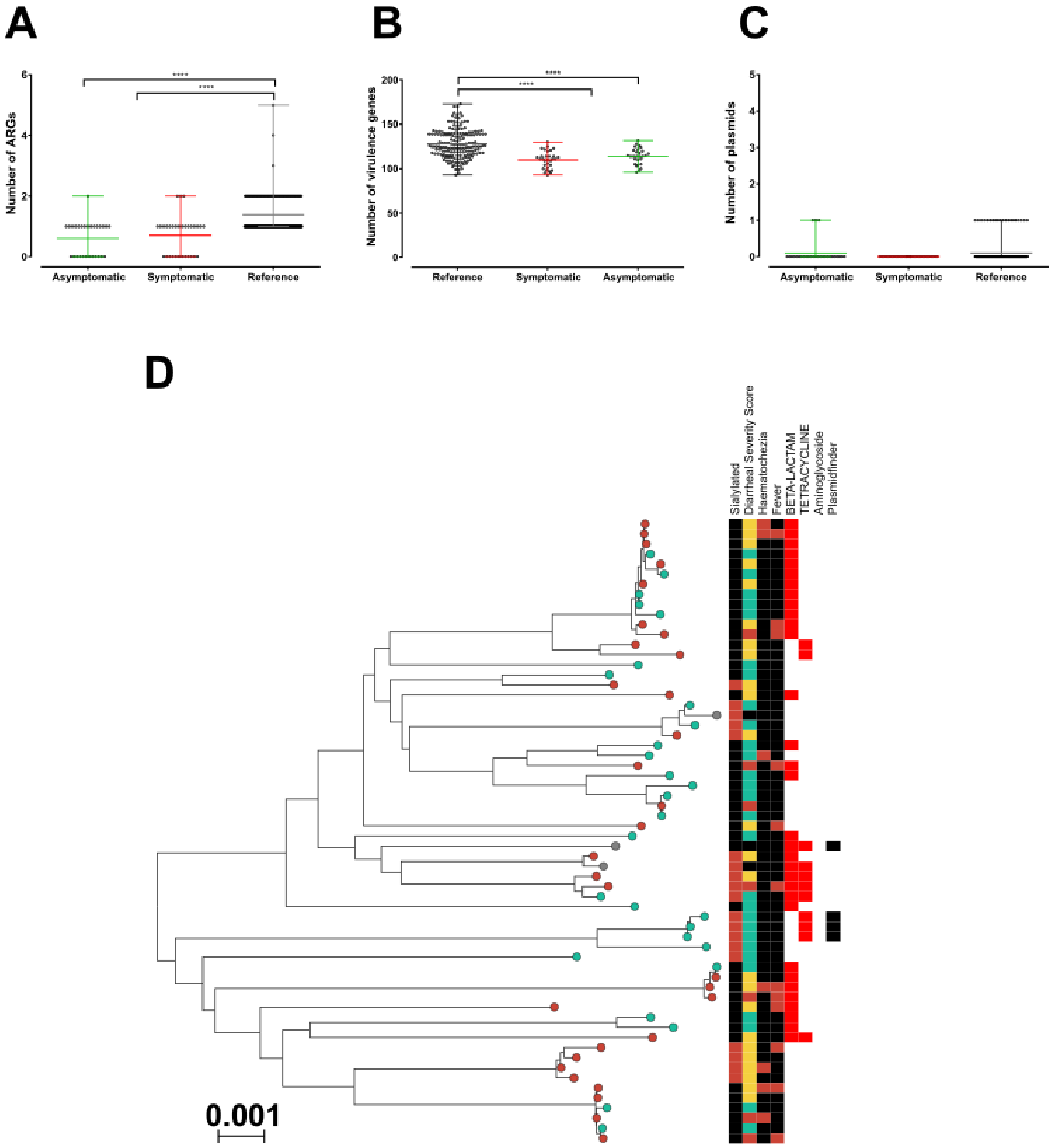
In conclusion, by contextualizing *C. jejuni* genomes from Peruvian Amazon children within a global reference collection and linking them to clinical data on varying disease symptoms and severity, we were able to identify local and globally distributed genotypes and determine the major source of infection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

Supplementary materials (https://doi.org/10.6084/m9.figshare.10352375)

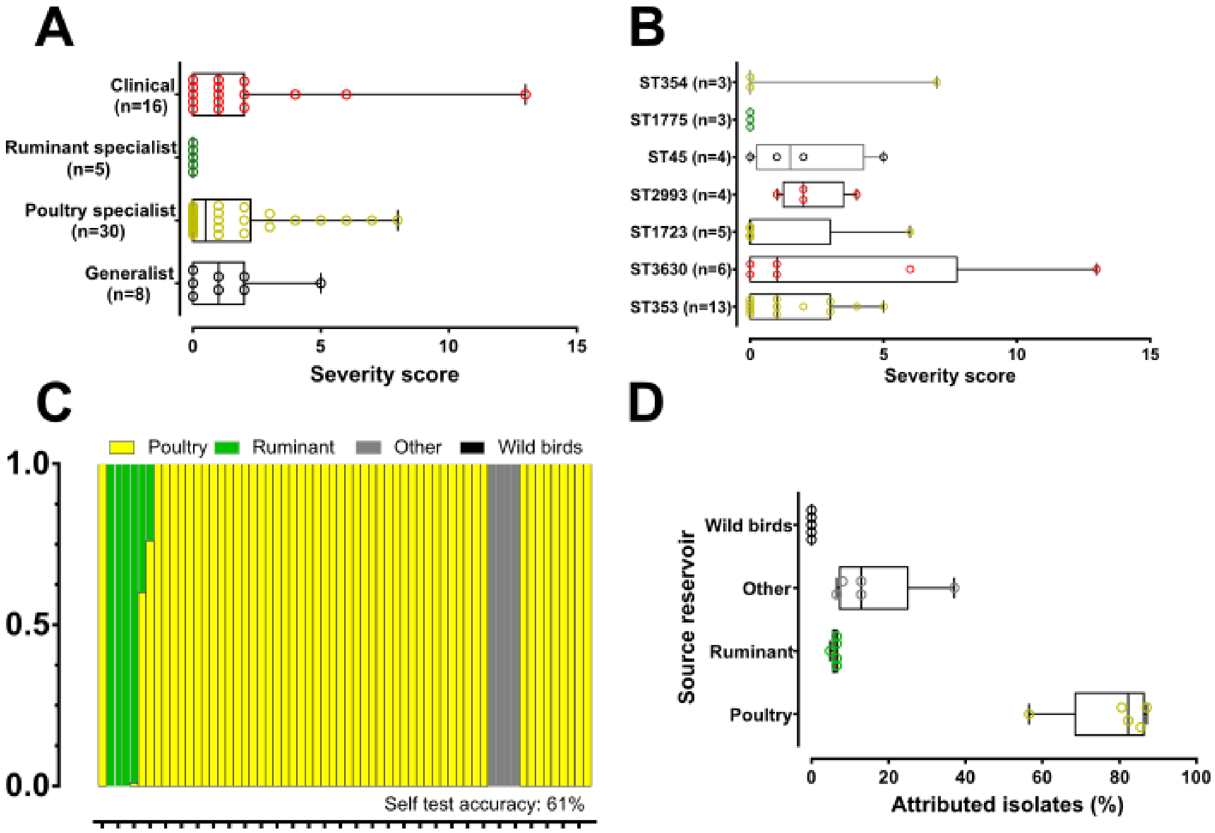
Supplementary Table S1: Isolate list and disease severity scores
Supplementary Table S2: Assembly metrics and accession numbers
Supplementary Table S3: Global reference dataset details
Supplementary Table S4: Reference pan-genome gene presence
Supplementary Table S5: ABRICATE summary
Supplementary Table S6: Genome characterization
Supplementary Table S7: ST summary of pubMSLT
Supplementary Table S8: Source attribution dataset
Supplementary Table S9: Comparison of NCTC11168 gene presence


A
2,000,000
1,800,000
1,600,000
1,400,000
1,200,000
Genome size (bp)
****
****
Ave. genome size (n=234): 1,646,868
Asymptomatic (n=28)
Symptomatic (n= 31)
Global reference dataset (n=172)
B
2,000,000
1,800,000
1,600,000
1,400,000
1,200,000
Genome size (bp)
Ave. genome size (n=234)
ST-3630
ST-362
ST-2802
ST-2042
ST-41
ST-986
ST-658
ST-677
ST-21
ST-45
ST-52
ST-22
ST-49
ST-353
ST-354
ST-460
ST-443
ST-607
ST-2304
ST-257
ST-407
ST-206
ST-403
ST-48
ST-61
ST-42
C
100
50
0
50
Over-represented in reference dataset
× Core
○ Accessory
Prevalence ratio (%)
500,000
1,000,000
1,500,000
2,000,000
NCTC11168 genome position (bp)
Over-represented in Peruvian dataset
D
Symptomatic Peruvian isolate (n=31)
Asymptomatic Peruvian isolate (n=28)
Global reference dataset (n=172)
Network summary
Components 43
Density 0.1059
Transitivity 0.8716
Score 0.7793

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



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.



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

Supplementary file 1: Pan-genome
Supplementary file 2: Alignment – Peru isolates only
Supplementary file 3: Alignment – Peru plus context isolates.

Contributors

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

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

Acknowledgements

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

References

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

ED study: A multinational and multidisciplinary approach to understand the relationship between enteric pathogens, malnutrition, gut physiology, physical growth, cognitive development, and immune responses in infants and children up to 2 years of age in resource-poor environments. Clin Infect Dis. 2014;59: S193–S206. doi:10.1093/cid/ciu653

15. Kotloff KL, Nasrin D, Blackwelder WC, Wu Y, Farag T, Panchalingham S, et al. The incidence, aetiology, and adverse clinical consequences of less severe diarrhoeal episodes among infants and children residing in low-income and middle-income countries: a 12-month case-control study as a follow-on to the Global Enteric Multicenter Study (GEMS). Lancet Glob Heal. 2019;7: e568–e584. doi:10.1016/S2214-109X(19)30076-2

16. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): A prospective, case-control study. Lancet. 2013;382: 209–222. doi:10.1016/S0140-6736(13)60844-2

17. Pazzaglia G, Bourgeois a L, el Diwany K, Nour N, Badran N, Hablas R. *Campylobacter* diarrhoea and an association of recent disease with asymptomatic shedding in Egyptian children. Epidemiol Infect. 1991;106: 77–82. doi:10.1017/S0950268800056466

18. Georges-Courbot MC, Beraud-Cassel AM, Gouandjika I, Georges AJ. Prospective study of enteric *Campylobacter* infections in children from birth to 6 months in the Central African Republic. J Clin Microbiol. 1987;25: 836–839.

19. Figueroa G, Galeno H, Troncoso M, Toledo S, Soto V. Prospective study of *Campylobacter jejuni* infection in Chilean infants evaluated by culture and serology. J Clin Microbiol. 1989;27: 1040–1044.

20. Ani EA, Takahashi T, Shonekan RAO. *Campylobacter jejuni* antibodies in Nigerian children. J Clin Microbiol. 1988;26: 605–606.

21. Calva J, Lopez-Vidal A, Ruiz-Palacios G, Ramos A, Bojalil R. Cohort study of intestinal infection with *Campylobacter* in Mexican children. Lancet. 1988;331: 503–506. doi:10.1016/S0140-6736(88)91297-4

22. Rao MR, Naficy AB, Savarino SJ, Abu-Elyazeed R, Wierzba TF, Peruski LF, et al. Pathogenicity and convalescent excretion of *Campylobacter* in rural Egyptian children. Am J Epidemiol. 2001;154: 166–173. doi:10.1093/aje/154.2.166

23. Poocharoen L, Bruin CW, Sirisanthana V, Vannareumol P, Leechanachai P, Sukhavat K. The relative importance of various enteropathogens as a cause of diarrhoea in hospitalized children in Chiang Mai, Thailand. J Diarrhoeal Dis Res. 1986;4: 10–15.

24. Sheppard SK, Guttman DS, Fitzgerald JR. Population genomics of bacterial host adaptation. Nat Rev Genet. 2018;19: 549–565. doi:10.1038/s41576-018-0032-z

25. Mourkas E, Taylor AJA, Méric G, Bayliss SCS, Pascoe B, Mageiros L, et al. Agricultural intensification and the evolution of host specialism in the enteric pathogen *Campylobacter jejuni*. Proc Natl Acad Sci. 2020;

26. Sheppard SK, Cheng L, Méric G, De Haan CPA, Llarena AK, Marttinen P, et al. Cryptic ecology among host generalist *Campylobacter jejuni* in domestic animals. Mol Ecol. 2014; doi:10.1111/mec.12742

27. Sheppard SK, Colles F, Richardson J, Cody AJ, Elson R, Lawson A, et al. Host association of *Campylobacter* genotypes transcends geographic variation. Appl Environ Microbiol. 2010;76: 5269–77. doi:10.1128/AEM.00124-10

28. Thépault A, Méric G, Rivoal K, Pascoe B, Mageiros L, Touzain F, et al. Genome-wide identification of host-segregating epidemiological markers for source attribution in

Campylobacter jejuni. Appl Environ Microbiol. 2017;83. doi:10.1128/AEM.03085-16

29. Sheppard SK, Dallas JF, MacRae M, McCarthy ND, Sproston EL, Gormley FJ, et al. *Campylobacter* genotypes from food animals, environmental sources and clinical disease in Scotland 2005/6. Int J Food Microbiol. 2009;134: 96–103. doi:10.1016/j.ijfoodmicro.2009.02.010

30. Revez J, Rossi M, Ellström P, de Haan C, Rautelin H, Hänninen M-L. Finnish *Campylobacter jejuni* Strains of Multilocus Sequence Type ST-22 Complex Have Two Lineages with Different Characteristics. Bereswill S, editor. PLoS One. 2011;6: e26880. doi:10.1371/journal.pone.0026880

31. Heikema AP, Islam Z, Horst-Kreft D, Huizinga R, Jacobs BC, Wagenaar JA, et al. *Campylobacter jejuni* capsular genotypes are related to Guillain-Barré syndrome. Clin Microbiol Infect. 2015;21. doi:10.1016/j.cmi.2015.05.031

32. Nielsen LN, Sheppard SK, McCarthy ND, Maiden MCJ, Ingmer H, Krogfelt KA. MLST clustering of *Campylobacter jejuni* isolates from patients with gastroenteritis, reactive arthritis and Guillain-Barre syndrome. J Appl Microbiol. 2010;108: 591–599. doi:10.1111/j.1365-2672.2009.04444.x

33. Unicomb LE, O'Reilly LC, Kirk MD, Stafford RJ, Smith H V., Becker NG, et al. Risk factors for infection with *Campylobacter jejuni flaA* genotypes. Epidemiol Infect. 2008;136: 1480–1491. doi:10.1017/S0950268807000246

34. Sahin O, Fitzgerald C, Stroika S, Zhao S, Sippy RJ, Kwan P, et al. Molecular evidence for zoonotic transmission of an emergent, highly pathogenic *Campylobacter jejuni* clone in the United States. J Clin Microbiol. 2012;50: 680–7. doi:10.1128/JCM.06167-11

35. Kirk KF, Méric G, Nielsen HL, Pascoe B, Sheppard SK, Thorlacius-Ussing O, et al. Molecular epidemiology and comparative genomics of *Campylobacter concisus* strains from saliva, faeces and gut mucosal biopsies in inflammatory bowel disease. Sci Rep. 2018;8. doi:10.1038/s41598-018-20135-4

36. Pascoe B, M?ric G, Yahara K, Wimalarathna H, Murray S, Hitchings MD, et al. Local genes for local bacteria: Evidence of allopatry in the genomes of transatlantic *Campylobacter* populations. Mol Ecol. 2017;26: 4497–4508. doi:10.1111/mec.14176

37. de Haan CPA, Kivisto R, Hakkinen M, Rautelin H, Hanninen ML. Decreasing Trend of Overlapping Multilocus Sequence Types between Human and Chicken *Campylobacter jejuni* Isolates over a Decade in Finland. Appl Environ Microbiol. 2010;76: 5228–5236. doi:10.1128/AEM.00581-10

38. Asakura H, Brüggemann H, Sheppard SK, Ekawa T, Meyer TF, Yamamoto S, et al. Molecular Evidence for the Thriving of *Campylobacter jejuni* ST-4526 in Japan. Bereswill S, editor. PLoS One. 2012;7: e48394. doi:10.1371/journal.pone.0048394

39. Llarena AK, Zhang J, Vehkala M, Välimäki N, Hakkinen M, Hänninen ML, et al. Monomorphic genotypes within a generalist lineage of *Campylobacter jejuni* show signs of global dispersion. Microb genomics. 2016;2: e000088. doi:10.1099/mgen.0.000088

40. Méric G, McNally A, Pessia A, Mourkas E, Pascoe B, Mageiros L, et al. Convergent Amino Acid Signatures in Polyphyletic *Campylobacter jejuni* Subpopulations Suggest Human Niche Tropism. Genome Biol Evol. 2018;10: 763–774. doi:10.1093/gbe/evy026

41. Schiaffino F, Colston JM, Paredes-Olortegui M, François R, Pisanic N, Burga R, et al. Antibiotic resistance of *Campylobacter* species in a pediatric cohort study. Antimicrob Agents Chemother. 2019;63. doi:10.1128/AAC.01911-18

42. Rojas JD, Reynolds ND, Pike BL, Espinoza NM, Kuroiwa J, Jani V, et al. Distribution of Capsular Types of *Campylobacter jejuni* Isolates from Symptomatic and Asymptomatic Children in Peru. Am J Trop Med Hyg. 2019;101: 541–548.

doi:10.4269/ajtmh.18-0994

43. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008; doi:10.1101/gr.074492.107

44. Pascoe B, Méric G, Murray S, Yahara K, Mageiros L, Bowen R, et al. Enhanced biofilm formation and multi-host transmission evolve from divergent genetic backgrounds in *Campylobacter jejuni*. Environ Microbiol. 2015;17: 4779–4789. doi:10.1111/1462-2920.13051

45. Sheppard SK, Didelot X, Meric G, Torralbo A, Jolley KA, Kelly DJ, et al. Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in *Campylobacter*. Proc Natl Acad Sci. 2013;110: 11923–11927. doi:10.1073/pnas.1305559110

46. Yahara K, Méric G, Taylor AJ, de Vries SPW, Murray S, Pascoe B, et al. Genome-wide association of functional traits linked with *Campylobacter jejuni* survival from farm to fork. Environ Microbiol. 2017;19. doi:10.1111/1462-2920.13628

47. Cody AJ, McCarthy NM, Wimalarathna HL, Colles FM, Clark L, Bowler ICJWJW, et al. A Longitudinal 6-Year Study of the Molecular Epidemiology of Clinical *Campylobacter* Isolates in Oxfordshire, United Kingdom. J Clin Microbiol. 2012;50: 3193–3201. doi:10.1128/JCM.01086-12

48. Lee G, Yori PP, Olortegui MP, Caulfield LE, Sack DA, Fischer-Walker C, et al. An instrument for the assessment of diarrhoeal severity based on a longitudinal community-based study. BMJ Open. 2014;4. doi:10.1136/bmjopen-2014-004816

49. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9: 75. doi:10.1186/1471-2164-9-75

50. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 2014;42: D206-14. doi:10.1093/nar/gkt1226

51. Méric G, Yahara K, Mageiros L, Pascoe B, Maiden MCJ, Jolley KA, et al. A reference pan-genome approach to comparative bacterial genomics: Identification of novel epidemiological markers in pathogenic *Campylobacter*. PLoS One. 2014;9. doi:10.1371/journal.pone.0092798

52. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013;30: 772–780. doi:10.1093/molbev/mst010

53. Sheppard SK, Jolley KA, Maiden MCJ. A gene-by-gene approach to bacterial population genomics: Whole genome MLST of *Campylobacter*. Genes (Basel). 2012;3: 261–277. doi:10.3390/genes3020261

54. Méric G, Yahara K, Mageiros L, Pascoe B, Maiden MCJ, Jolley KA, et al. A reference pan-genome approach to comparative bacterial genomics: Identification of novel epidemiological markers in pathogenic *Campylobacter*. PLoS One. 2014;9. doi:10.1371/journal.pone.0092798

55. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2015;32: 268–274. doi:10.1093/molbev/msu300

56. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol. 2018;35: 518–522. doi:10.1093/molbev/msx281

57. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, et al. Microreact:

visualizing and sharing data for genomic epidemiology and phylogeography. Microb genomics. 2016;2: e000093. doi:10.1099/mgen.0.000093

58. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications [version 1; referees: 2 approved]. Wellcome Open Res. 2018;3. doi:10.12688/wellcomeopenres.14826.1

59. Jolley KA, Maiden MCJ. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11: 595. doi:10.1186/1471-2105-11-595

60. Dingle KE, Colles FM, Falush D, Maiden MCJ. Sequence typing and comparison of population biology of *Campylobacter coli* and *Campylobacter jejuni*. J Clin Microbiol. 2005;43: 340–347. doi:10.1128/JCM.43.1.340-347.2005

61. Parker CT, Gilbert M, Yuki N, Endtz HP, Mandrell RE. Characterization of lipooligosaccharide-biosynthetic loci of *Campylobacter jejuni* reveals new lipooligosaccharide classes: Evidence of mosaic organizations. J Bacteriol. 2008;190: 5681–5689. doi:10.1128/JB.00254-08

62. Culebro A, Revez J, Pascoe B, Friedmann Y, Hitchings MDMD, Stupak J, et al. Large sequence diversity within the biosynthesis locus and common biochemical features of *Campylobacter coli* lipooligosaccharides. DiRita VJ, editor. J Bacteriol. 2016;198: 2829–40. doi:10.1128/JB.00347-16

63. Grundmann H, Hori S, Tanner G. Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. J Clin Microbiol. 2001;39: 4190–4192. doi:10.1128/JCM.39.11.4190-4192.2001

64. Carattoli A, Zankari E, Garciá-Fernández A, Larsen MV, Lund O, Villa L, et al. *In Silico* detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014;58: 3895–3903. doi:10.1128/AAC.02412-14

65. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: Hierarchical and refined dataset for big data analysis - 10 years on. Nucleic Acids Res. 2016;44: D694–D697. doi:10.1093/nar/gkv1239

66. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, et al. Using the NCBI AMRFinder Tool to Determine Antimicrobial Resistance Genotype-Phenotype Correlations Within a Collection of NARMS Isolates. bioRxiv. 2019; 550707. doi:10.1101/550707

67. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012;67: 2640–2644. doi:10.1093/jac/dks261

68. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2019; doi:10.1093/nar/gkz935

69. Seemann T. ABRicate: Mass screening of contigs for antimicrobial and virulence genes [Internet]. GitHub repository. 2018. Available: https://github.com/tseemann/abricate

70. Lees JA, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, Weiser JN, et al. Fast and flexible bacterial genomic epidemiology with PopPUNK. Genome Res. 2019;29: 304–316. doi:10.1101/gr.241455.118

71. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;

72. Dearlove BL, Cody AJ, Pascoe B, Méric G, Wilson DJ, Sheppard SK. Rapid host switching in generalist *Campylobacter* strains erodes the signal for tracing human infections. ISME J. 2016;10. doi:10.1038/ismej.2015.149

73. Cody AJ, Maiden MC, Strachan NJ, McCarthy ND. A systematic review of source attribution of human campylobacteriosis using multilocus sequence typing. Eurosurveillance. 2019;24. doi:10.2807/1560-7917.ES.2019.24.43.1800696

74. Baily JL, Méric G, Bayliss S, Foster G, Moss SE, Watson E, et al. Evidence of land-sea transfer of the zoonotic pathogen *Campylobacter* to a wildlife marine sentinel species. Mol Ecol. 2015;24. doi:10.1111/mec.13001

75. Weinert LA, Chaudhuri RR, Wang J, Peters SE, Corander J, Jombart T, et al. Genomic signatures of human and animal disease in the zoonotic pathogen *Streptococcus suis*. Nat Commun. 2015;6: 6740. doi:10.1038/ncomms7740

76. Morley L, McNally A, Paszkiewicz K, Corander J, Méric G, Sheppard SK, et al. Gene loss and lineage-specific restriction-modification systems associated with niche differentiation in the *Campylobacter jejuni* sequence type 403 clonal complex. Appl Environ Microbiol. 2015; doi:10.1128/AEM.00546-15

77. Karlyshev A V., Champion OL, Churcher C, Brisson JR, Jarrell HC, Gilbert M, et al. Analysis of *Campylobacter jejuni* capsular loci reveals multiple mechanisms for the generation of structural diversity and the ability to form complex heptoses. Mol Microbiol. 2005;55: 90–103. doi:10.1111/j.1365-2958.2004.04374.x

78. Parker CT, Gilbert M, Yuki N, Endtz HP, Mandrell RE. Characterization of lipooligosaccharide-biosynthetic loci of *Campylobacter jejuni* reveals new lipooligosaccharide classes: Evidence of mosaic organizations. J Bacteriol. 2008;190: 5681–5689. doi:10.1128/JB.00254-08

79. Parker CT, Horn ST, Gilbert M, Miller WG, Woodward DL, Mandrell RE. Comparison of *Campylobacter jejuni* lipooligosaccharide biosynthesis loci from a variety of sources. J Clin Microbiol. 2005;43: 2771–2781. doi:10.1128/JCM.43.6.2771-2781.2005

80. Poly F, Serichantalergs O, Kuroiwa J, Pootong P, Mason C, Guerry P, et al. Updated *Campylobacter jejuni* Capsule PCR Multiplex Typing System and Its Application to Clinical Isolates from South and Southeast Asia. Skurnik M, editor. PLoS One. 2015;10: e0144349. doi:10.1371/journal.pone.0144349

81. Dunn SJ, Pascoe B, Turton J, Fleming V, Diggle M, Sheppard SK, et al. Genomic epidemiology of clinical *Campylobacter spp*. at a single health trust site. Microb Genomics. 2018; doi:10.1099/mgen.0.000227

82. Mottet A, Tempio G. Global poultry production: Current state and future outlook and challenges. World's Poultry Science Journal. Cambridge University Press; 2017. pp. 245–256. doi:10.1017/S0043933917000071

83. Wilson DJ, Gabriel E, Leatherbarrow AJH, Cheesbrough J, Gee S, Bolton E, et al. Tracing the Source of Campylobacteriosis. Guttman DS, editor. PLoS Genet. 2008;4: e1000203. doi:10.1371/journal.pgen.1000203

84. Prachantasena S, Charununtakorn P, Muangnoicharoen S, Hankla L, Techawal N, Chaveerach P, et al. Distribution and genetic profiles of *Campylobacter* in commercial broiler production from breeder to slaughter in Thailand. PLoS One. 2016;11. doi:10.1371/journal.pone.0149585

85. Ngulukun S, Oboegbulem S, Klein G. Multilocus sequence typing of *Campylobacter jejuni* and *Campylobacter coli* isolates from poultry, cattle and humans in Nigeria. J Appl Microbiol. 2016;121: 561–568. doi:10.1111/jam.13185

86. Duong VT, Tuyen HT, Van Minh P, Campbell JI, Le Phuc H, Nhu TDH, et al. No Clinical benefit of empirical antimicrobial therapy for pediatric diarrhea in a high-usage, high-resistance setting. Clin Infect Dis. 2018;66: 504–511. doi:10.1093/cid/cix844

87. Mason J, Iturriza-Gomara M, O'Brien SJ, Ngwira BM, Dove W, Maiden MCJ, et al.

Campylobacter Infection in Children in Malawi Is Common and Is Frequently Associated with Enteric Virus Co-Infections. Hold GL, editor. PLoS One. 2013;8: e59663. doi:10.1371/journal.pone.0059663

88. de Vries SPW, Vurayai M, Holmes M, Gupta S, Bateman M, Goldfarb D, et al. Phylogenetic analyses and antimicrobial resistance profiles of *Campylobacter spp*. from diarrhoeal patients and chickens in Botswana. Chang Y-F, editor. PLoS One. 2018;13: e0194481. doi:10.1371/journal.pone.0194481

89. Acheson D, Allos BM. Campylobacter jejuni Infections: Update on Emerging Issues and Trends. Clin Infect Dis. 2001;32: 1201–1206. doi:10.1086/319760

90. Toledo Z, Simaluiza RJ, Astudillo X, Fernández H. Occurrence and antimicrobial susceptibility of thermophilic *Campylobacter* species isolated from healthy children attending municipal care centers in Southern Ecuador. Rev Inst Med Trop Sao Paulo. 2017;59. doi:10.1590/S1678-9946201759077

91. Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. Within-host evolution of bacterial pathogens. Nature Reviews Microbiology. Nature Publishing Group; 2016. pp. 150–162. doi:10.1038/nrmicro.2015.13

92. Martins NE, Faria VG, Teixeira L, Magalhães S, Sucena É. Host Adaptation Is Contingent upon the Infection Route Taken by Pathogens. PLoS Pathog. 2013;9. doi:10.1371/journal.ppat.1003601

93. Buchanan CJ, Webb AL, Mutschall SK, Kruczkiewicz P, Barker DOR, Hetman BM, et al. A genome-wide association study to identify diagnostic markers for human pathogenic *Campylobacter jejuni* strains. Front Microbiol. 2017; doi:10.3389/fmicb.2017.01224

94. Thépault A, Rose V, Quesne S, Poezevara T, Béven V, Hirchaud E, et al. Ruminant and chicken: Important sources of campylobacteriosis in France despite a variation of source attribution in 2009 and 2015. Sci Rep. 2018;

95. Cressler CE, McLeod D V., Rozins C, Van Den Hoogen J, Day T. The adaptive evolution of virulence: A review of theoretical predictions and empirical tests. Parasitology. Cambridge University Press; 2016. pp. 915–930. doi:10.1017/S003118201500092X

96. Duim B, Godschalk PCR, Van Den Braak N, Dingle KE, Dijkstra JR, Leyde E, et al. Molecular Evidence for Dissemination of Unique *Campylobacter jejuni* Clones in Curaçao, Netherlands Antilles. J Clin Microbiol. 2003;41: 5593–5597. doi:10.1128/JCM.41.12.5593-5597.2003

97. Taveirne ME, Theriot CM, Livny J, DiRita VJ. The Complete *Campylobacter jejuni* Transcriptome during Colonization of a Natural Host Determined by RNAseq. PLoS One. 2013;8. doi:10.1371/journal.pone.0073586

98. Guerry P, Ewing CP, Hickey TE, Prendergast MM, Moran AP. Sialylation of lipooligosaccharide cores affects immunogenicity and serum resistance of *Campylobacter jejuni*. Infect Immun. 2000;68: 6656–62. Available: http://www.ncbi.nlm.nih.gov/pubmed/11083778

99. Zebian N, Merkx-Jacques A, Pittock PP, Houle S, Dozois CM, Lajoie GA, et al. Comprehensive analysis of flagellin glycosylation in *Campylobacter jejuni* NCTC 11168 reveals incorporation of legionaminic acid and its importance for host colonization. Glycobiology. 2016;26: 386–397. doi:10.1093/glycob/cwv104

100. Guerry P, Ewing CP, Hickey TE, Prendergast MM, Moran AP. Sialylation of lipooligosaccharide cores affects immunogenicity and serum resistance of *Campylobacter jejuni*. Infect Immun. 2000;68: 6656–62. doi:10.1128/IAI.68.12.6656-

6662.2000

101. Pascoe B, Williams LK, Calland JK, Meric G, Hitchings MD, Dyer M, et al. Domestication of *Campylobacter jejuni* NCTC 11168. Microb genomics. 2019;5. doi:10.1099/mgen.0.000279

102. Gilbert M, Karwaski MF, Bernatchez S, Young NM, Taboada E, Michniewicz J, et al. The genetic bases for the variation in the lipo-oligosaccharide of the mucosal pathogen, *Campylobacter jejuni*. Biosynthesis of sialylated ganglioside mimics in the core oligosaccharide. J Biol Chem. 2002;277: 327–337. doi:10.1074/jbc.M108452200

103. Houliston RS, Vinogradov E, Dzieciatkowska M, Li J, St Michael F, Karwaski MF, et al. Lipooligosaccharide of *Campylobacter jejuni*: Similarity with multiple types of mammalian glycans beyond gangliosides. J Biol Chem. 2011;286: 12361–12370. doi:10.1074/jbc.M110.181750

104. Godschalk PCR, Kuijf ML, Li J, St. Michael F, Ang CW, Jacobs BC, et al. Structural characterization of *Campylobacter jejuni* lipooligosaccharide outer cores associated with Guillain-Barré and Miller Fisher syndromes. Infect Immun. 2007;75: 1245–1254. doi:10.1128/IAI.00872-06

105. Guerry P, Poly F, Riddle M, Maue AC, Chen Y-H, Monteiro MA, et al. CELLULAR AND INFECTION MICROBIOLOGY *Campylobacter* polysaccharide capsules: virulence and vaccines. 2012; doi:10.3389/fcimb.2012.00007