

RESEARCH ARTICLE

Shotgun metagenomics of fecal samples from children in Peru reveals frequent complex co-infections with multiple *Campylobacter* species

Craig T. Parker¹, Francesca Schiaffino^{2,3}, Steven Huynh¹, Maribel Paredes Olortegui⁴, Pablo Peñataro Yori^{2,4}, Paul F. Garcia Bardales⁴, Tackeshy Pinedo Vasquez⁴, Greisi E. Curico Huansi⁴, Katia Manzanares Villanueva⁴, Wagner V. Shapiama Lopez⁴, Kerry K. Cooper^{5*}, Margaret N. Kosek^{2,4*}

1 Agricultural Research Service, U.S. Department of Agriculture, Produce Safety and Microbiology Research Unit, Albany, California, United States of America, **2** Division of Infectious Diseases, University of Virginia, Charlottesville, Virginia, United States of America, **3** Faculty of Veterinary Medicine, Universidad Peruana Cayetano Heredia, Lima, Peru, **4** Asociacion Benefica Prisma, Iquitos, Peru, **5** School of Animal and Comparative Biomedical Sciences, University of Arizona, Tucson, Arizona, United States of America

* kcooper@arizona.edu (KKC); mkosek@virginia.edu (MNK)



OPEN ACCESS

Citation: Parker CT, Schiaffino F, Huynh S, Paredes Olortegui M, Peñataro Yori P, Garcia Bardales PF, et al. (2022) Shotgun metagenomics of fecal samples from children in Peru reveals frequent complex co-infections with multiple *Campylobacter* species. PLoS Negl Trop Dis 16(10): e0010815. <https://doi.org/10.1371/journal.pntd.0010815>

Editor: Adly M. M. Abd-Alla, International Atomic Energy Agency, AUSTRIA

Received: May 16, 2022

Accepted: September 13, 2022

Published: October 4, 2022

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: Sequence related files are available from the NCBI database (BioProject PRJNA837236). All other relevant data are available within the manuscript and its [supporting information](#).

Funding: Funding for this study was provided by the Bill and Melinda Gates Foundation (OPP1066146 and OPP1152146 to MNK), the National Institutes of Health of the United States (R01AI158576 and R21AI163801 to MNK and

Abstract

Campylobacter spp. are a major cause of bacterial diarrhea worldwide and are associated with high rates of mortality and linear growth faltering in children living in low- to middle-income countries (LMICs). *Campylobacter jejuni* and *Campylobacter coli* are most often the causative agents of enteric disease among children in LMICs. However, previous work on a collection of stool samples from children under 2 years of age, living in a low resource community in Peru with either acute diarrheal disease or asymptomatic, were found to be qPCR positive for *Campylobacter* species but qPCR negative for *C. jejuni* and *C. coli*. The goal of this study was to determine if whole-genome shotgun metagenomic sequencing (WSMS) could identify the *Campylobacter* species within these samples. The *Campylobacter* species identified in these stool samples included *C. jejuni*, *C. coli*, *C. upsaliensis*, *C. concisus*, and the potential new species of *Campylobacter*, "*Candidatus Campylobacter infans*". Moreover, WSMS results demonstrate that over 65% of the samples represented co-infections with multiple *Campylobacter* species present in a single stool sample, a novel finding in human populations.

Author summary

Analysis of shotgun metagenomic data obtained from fecal samples of children living in a low resource tropical community of Peru revealed multiple *Campylobacter* species. Co-infections with more than one *Campylobacter* species within the same sample was a common finding. A potential new species of *Campylobacter* was also detected within these samples.

CTP; D43TW010913 to MNK). This research was also supported in part by USDA-ARS CRIS project 2030-42000-055-00D to CTP. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Campylobacter is a Gram-negative fastidious organism that requires specific atmospheric and nutritional requirements for successful culture and isolation. In high income countries, campylobacteriosis is one of the most common forms of bacterial enteritis and it is often stated that more than 95% of infections are caused by either *Campylobacter jejuni* or *Campylobacter coli* [1]. However, in low- and middle-income countries (LMICs), where campylobacteriosis is an endemic disease associated with gastroenteritis, environmental enteropathy, and growth stunting in children, disease has been shown to be associated with other *Campylobacter* species besides just *C. jejuni* and *C. coli* [2–6]. Due to challenges that limit culturing of *Campylobacter* isolates these types of infections can often be overlooked, particularly in LMICs where they are most prevalent [7].

To overcome inadequate culturing of pathogens, the Centers for Disease Control and Prevention (CDC) of the United States has highlighted the importance of culture independent diagnostic technologies (CIDTs) for the detection and quantification of foodborne bacterial illnesses [8–10]. Several CIDTs for *Campylobacter* species have been developed that are either antigen-based or nucleic acid amplification-based [11]. Indeed, large-scale epidemiologic studies have relied on CIDTs including quantitative polymerase chain reaction (qPCR), either in stand-alone assays or associated with array card technologies (BioFire, BD Diagnostics/Enteric Bacterial Panel, Applied Biosystems), to determine infection and colonization in human samples [7,12–14]. Yet, these CIDTs target *C. jejuni* and/or *C. coli* rather than all *Campylobacter* species.

With the rapid expansion, increasing portability, and decreasing costs of next-generation sequencing (NGS), a potential sensitive alternative to PCR-based diagnostics is shotgun metagenomics. This CIDT enables the resolution of multiple *Campylobacter* species that are rarely cultured and for which specific probes have not been developed, as well as other pathogenic and non-pathogenic bacterial, viral, parasitic, and fungal populations present in a sample [15]. Additionally, it has the potential to overcome the major CIDT issue of not providing a pathogen genome for source traceback or outbreak investigations for public health agencies without reflex culturing [16,17]. Determining the genetic complexity and polyclonality of *Campylobacter* species in fecal samples from children living in highly endemic settings is a requisite for understanding transmission dynamics and attaining accurate source attribution.

Previous studies occurring in Peru have identified stool samples positive for *Campylobacter* by 16S rRNA gene-based PCR probes, and further differentiated for *C. jejuni* and *C. coli* via qPCR using the *cadF* gene. In many cases *C. jejuni* and *C. coli* were also successfully cultured from these samples. However, culturing unidentified *Campylobacter* species from stool samples positive for *Campylobacter* by 16S rRNA PCR but negative for *C. jejuni* and *C. coli* by *cadF* qPCR has yet to be consistently successful [5,13]. Thus, in this study, we utilize whole-genome shotgun metagenomic sequencing (WGMS), as a CIDT, to detect and identify the presence of other *Campylobacter* species besides *C. jejuni* and *C. coli* in fecal samples from children in Peru.

Methods

Ethics statement

Samples used in this study were collected as a study approved by the Institutional Review Boards of Asociacion Benefica Prisma (Lima, Peru) and Johns Hopkins Bloomberg School of Public Health (Baltimore, MD, USA). Consent to participate in the study was obtained from

the parents or legal guardians of children. Participants of both studies consented for further use of biological specimens.

Biological samples

Archived fecal samples were derived from children enrolled in a community-based cohort study in Iquitos, Loreto, Peru. Between 2009 and 2018, 303 children were enrolled within 17 days of birth and followed for five years. Details of enrollment and surveillance procedures have been previously described [18]. Children were visited twice weekly to create a continuous daily record of early life childhood illness. Stool was collected monthly and in children experiencing diarrhea defined as >3 unformed stools in a 24-hour period. For this study we randomly chose 50 fecal samples for shotgun metagenomic sequencing out of 124 fecal samples that had previously shown to be positive for *Campylobacter* (16S rRNA gene PCR) but were negative for the *cadF* gene associated with the major thermotolerant species, *C. jejuni* and *C. coli*. These fecal samples were part of a pool of 439 symptomatic and asymptomatic fecal samples for which selection criteria has been published previously [5]. These archived samples had been stored at -80°C after initial collection.

Nucleic acid diagnostics

Fecal DNA was extracted from 0.2 grams of feces using the QIAamp DNA Stool Mini Kit (Qiagen, Carlsbad, CA), according to the manufacturer's instructions. A negative control consisting of RNA and DNA free water was used for each extraction set. All samples were processed using a Taqman based multiplex qPCR assay to detect *Campylobacter* spp. (16S rRNA gene), *Campylobacter jejuni* and/or *Campylobacter coli* (*cadF* gene), and *Shigella* sp. (*ipaH* gene), using the primers and probes specified in S1 Table. The final assay consisted of a 25 µL final reaction mixture with 12.5 µL of Environmental Master Mix (2X) (Applied Biosystems, Foster City, CA), 0.5 µL of forward and reverse primers (0.2 µM each), 0.25 µL of probes (0.1 µM each), 1 µL of DNA template and 6.75 µL of nuclease-free water (Ambion, Thermo Fisher Scientific, Waltham, MA, USA). The assays were performed on a QuantStudio 7 Flex (Applied Biosystems, Foster City, CA) using the following cycling conditions: 95°C for 10 minutes followed by 45 cycles of 95°C for 15 seconds and 60°C for 1 minutes. DNA from previously confirmed *C. coli* and *C. jejuni* were used as positive controls. Template-free controls (nuclease-free water) were included as negative controls. A cut-off cycle threshold of 38 was used to determine positivity. A sample positive for both the 16S rRNA and *cadF* genes was interpreted as positive for either *C. coli* or *C. jejuni*.

Shotgun metagenomic DNA sequencing

Depending on the sample, 20 to 1,500 ng of DNA extracted from fecal samples was sheared in a 50 µL screwcap microtube using a Covaris M220 instrument (Covaris, Woburn, MA) at 50 peak power, 20 duty factor, 20°C, 200 cycles per burst and 25–31 seconds duration. Illumina sequencing libraries were prepared using the KAPA High-Throughput Library Preparation Kit with Standard PCR Amplification Module (Kapa Biosystems, Wilmington, MA), following the manufacturer's instructions except for the following changes: libraries were prepared using 1/4th volumes for all steps except size selected to ~500–1500 bp following double-sided size selection protocols with modified volumes of 41 µL PEG/NaCl solution and 5 µL AMPure XP beads. Indexing of each individual sample was done using standard desalted TruSeq HT dual index adapters from Integrated DNA Technologies (Coralville, IA) at 0.75 µM final concentration, and a total of 3–18 PCR cycles were used to minimize bias. Libraries were quantified using the KAPA Library Quantification Kit, except with 10 µL volume and 90 s annealing/

extension PCR, and then pooled and normalized to 4 nM. Pooled libraries were re-quantified by ddPCR on a QX200 system (Bio-Rad, Hercules, CA), using the Illumina TruSeq ddPCR Library Quantification Kit and following the manufacturer's protocols, except with an extended 2-min annealing/extension time. Libraries were sequenced on an Illumina MiSeq instrument (v2 500-cycle kit; Illumina, San Diego, CA) per the manufacturer's protocols. Average sequence read lengths for each sample were >188 nucleotides. Short read data are available at NCBI SRA and are associated with BioProject PRJNA834762.

Detection of *Campylobacter* species from shotgun metagenomics

Identification of *Campylobacter* species was performed using the Map to Reference command within Geneious Prime (v2021.2.2; Biomatters, Ltd., Auckland, New Zealand). Illumina paired-end reads >80 nt generated from an individual fecal sample were simultaneously mapped to 28 *Campylobacter* chromosomes, including: [1] *Candidatus Campylobacter infans* str. 19S00001 (CP049075.1), [2] *C. avium* str. LMG 24591 (CP022347.1), *C. canadensis* str. LMG 24001 (CP035946.1), *C. coli* str. 14983A (CP017025.1), *C. coli* plasmid pCC14983A-1 (CP017026.1), *C. concisus* str. ATCC 33237 (CP012541.1), *C. corcagiensis* str. LMG 27932 (CP053842.1), *C. curvus* str. ATCC 35224 (CP053826.1), *C. fetus* str. NCTC 10354 (CP043435.1), *C. gracilis* str. ATCC 33236 (CP012196.1), *C. helveticus* str. ATCC 51209 (CP020478.1), *C. hepaticus* str. HV10 (CP031611.1), *C. hominis* str. ATCC BAA-381 (CP000776.1), *C. hyointestinalis* str. CHY5 (CP053828.1), *C. iguaniorum* str. RM11343 (CP015577.1), *C. insulaenigrae* str. NCTC 12927 (CP007770.1), *C. jejuni* str. NCTC 11168 (AL111168.1), *C. lanienae* str. NCTC 13004 (CP015578.1), *C. lari* str. RM2100 (CP000932.1), *C. mucosalis* str. ATCC 43264 (CP053831.1), *C. pinnipediorum* str. RM17261 (CP012547.1), *C. rectus* str. ATCC 33238 (CP012543.1), *C. showae* str. ATCC 51146 (CP012544.1), *C. sputorum* str. LMG 7795 (CP043427.1), *C. subantarcticus* str. LMG 24377 (CP007773.1), *C. upsaliensis* str. NCTC 11541 (LR134372.1), *C. ureolyticus* str. RIGS 9880 (CP012195.1), *C. volucrus* str. LMG 24380 (CP043428.1), and *C. vulpis* str. 251/13 (CP041617). In total, 44 shotgun metagenomic samples with the number of reads ranging from 1,404 to 2,975,084 reads were mapped to each of the *Campylobacter* references using the low sensitivity settings (<10% mismatch between read and references), and the total number of reads mapped to each reference genome were determined. Reads mapping to genomic loci conserved between *Campylobacter* species (>90% sequence similarity between species; rRNA loci, antimicrobial genes, transposon genes, IS elements, and bacteriophage genes) were removed from the read mapping counts. These reads were determined to be non-confirmatory reads for only one species and could be mapped to multiple *Campylobacter* species or other bacterial genera. Reads were determined to be species confirmatory when BLASTn against the NCBI nucleotide (nr/nt) database resulted in the identification of a single *Campylobacter* species as the highest match, and the read was also mapped to the same *Campylobacter* species in Geneious. Additionally, the species confirmatory read should have a BLASTn >95% DNA identity across >85% of the read for a particular *Campylobacter* species. Metagenomic samples were scored as positive for a specific *Campylobacter* species when at least one species confirmatory read mapped to the genome. When <50 reads mapped to a particular genome in a sample, all reads were analyzed via BLASTn to determine if they were species confirmatory reads.

Shotgun metagenomic data analysis

Illumina reads for an individual sample were run through the MetaWRAP pipeline [19] to characterize the entire microbial population of the sample. Initially the reads were quality trimmed and index removal using Trim Galore!(v0.6.5) and human DNA contamination was

removed by mapping reads to the human GRCh37/hg19 reference genome using BMTagger [20]. Next, trimmed and cleaned reads were assembled using the metaSPAdes assembler (v3.12.0) [21], and the quality of the initial assemblies assessed using QUAST [22]. The taxonomic abundance of the initial assembled contigs and all the sequence reads were determined using Kraken2 (v2.0.8) [23,24] against the standard Kraken2 database and visualized using KronaTools [25]. Next, the metaSPAdes assembly and trimmed and cleaned reads were used to bin the contigs using both Metabat2 [26] and CONCOCT [27] programs. The initial bins were refined, screened for completeness and contamination using CheckM software (v1.0.12) [28], and consolidated into bins with at least 50% completeness and less than 10% contamination. The bins were then visualized using Blobology software [29], re-assembled using the initial refined and consolidated bins, and screened for completeness and contamination using CheckM again. Finally, the taxonomy of each contig in the different bins were assigned using Taxator-tk software [30] and annotated using Prokka [31]. Taxonomic reports generated by Kraken2 were exported as a biom file using Kraken-biom [32] and the alpha diversity, beta diversity and taxonomic profiles of the stool samples were determined using R software (v4.1.2) with the phyloseq [33], microbiome [34], and vegan [35] packages.

Results

Metagenomic detection of *Campylobacter* species

From the 50 archived fecal samples analyzed, we were able to successfully sequence and obtain data from 44 samples. Total DNA recovery from the 44 samples was quite variable and resulted in total reads per sample between 1,410 and 2,975,084 paired-reads, while overall 84% of samples had more than 150,000 reads (Table 1). The variability of DNA from the samples might be due to DNA degradation following long-term storage of the fecal samples. Nevertheless, we decided to analyze all samples with reads since previous WGMS results of infant stool samples had identified a target organism (*C. infans*) at >80% of all reads [36]. Mapping the sample reads against the genomes from 28 different *Campylobacter* species identified 41 out of 44 samples (93%) that were positive for *Campylobacter* reads. There were three samples in which no *Campylobacter* species was detected (97077, 118488, 130169). Despite negative qPCR results for *C. jejuni* and *C. coli* in these 44 samples, reads that matched *C. jejuni* (28/44 (63.6%)) and *C. coli* (6/44 (13.6%)) were identified in a majority of samples. Other *Campylobacter* species reads were identified in 28 samples with *C. infans* (26/44 (59.1%)), *C. upsaliensis* (4/44 (9.1%)), *C. concisus* (3/44 (6.8%)), *C. helveticus* (1/44 (2.3%)), and *C. curvus* (1/44 (2.3%)) (Table 1). *C. concisus*, *C. curvus*, *C. helveticus* and *C. upsaliensis* were only identified in diarrheal fecal samples from children that were 10 months or more in age (Fig 1).

Factors affecting species specificity of metagenomic reads

Metagenomic reads mapping to conserved genomic regions can affect specificity issues in the detection of *Campylobacter* species. The reads that mapped to the rRNA loci possessed >90% identity with multiple *Campylobacter* species due to sequence conservation of all rRNA genes. Upon BLASTn analysis, many of these mapped reads had higher scores to other genera including *Clostridium*, *Streptococcus*, and *Enterococcus*. Among the reference DNA sequences, the *C. coli* plasmid, pCC14983A-1 (CP017026) was included. This plasmid possesses transposon genes, a tetracycline resistance gene, and type VI secretion system genes that are common in plasmids and chromosomal loci from multiple *Campylobacter* species and other genera. There were 39 samples that had reads that mapped to the plasmid, and most reads mapped to the tetracycline resistance gene. Additionally, many of *Campylobacter* reference sequences possessed prophage, IS elements, and transposons that are shared among various *Campylobacter* species

Table 1. The number of reads detected for *Campylobacter infans*, *C. jejuni*, *C. coli*, *C. upsaliensis*, and *C. helveticus*. *C. concisus* and *C. curvus* in fecal samples from 44 children under two years of age living in Iquitos, Peru that were positive for *Campylobacter* by a Genus specific 16S assay but negative for the *cadF* target.

Sample	Total reads	<i>C. infans</i> ¹	<i>C. jejuni</i> ¹	<i>C. coli</i> ¹	<i>C. upsaliensis</i> ¹	<i>C. helveticus</i> ¹	<i>C. concisus</i> ¹	<i>C. curvus</i>
13323	575,288	52 (8kb)	1,059 (281kb)	164 (27kb)			3	
13812	925,310		15		8			
15704	1,387,764	19	34		98 (10kb)	89 (6kb)	462 (83kb)	30
18102	2,975,084	72 (9kb)	66 (11kb)					
18196	8,798	5	96*					
19879	1,506,736	71,326 (1.6mb)						
22997	252,136		121	12				
25500	153,672	11,399 (488kb)	43		37			
28671	496,196		317 (39kb)	598 (86kb)				
31204	1,267,842	806 (96kb)	13		502 (63kb)			
34891	221,308	77 (9kb)	12					
37185	301,380		299 (58kb)	20				
41480	517,620		5					
52141	553,782		1,544 (258Kb)	125 (24kb)				
57052	1,253,590		139 (10kb)					
57847	1,714,028		21					
63792	392,858	4	4					
65146	14,002				8			
76987	6,862	4						
78740	18,192	4	41*					
97077	950,608							
97770	565,420	861 (154kb)					31	
97915	948,314	2						
101367	591,096		160					
108338	1,047,414	685 (68kb)	1,734 (295kb)					
118291	1,210,340	26	24					
118488	711,132							
124035	372,880		981 (179kb)					
124473	214,924	3	31					
130169	640,924							
132208	205,794	6	4					
147132	1,195,212	170 (4kb)	4166 (633kb)	404 (60kb)				
147161	1,102,270	218*						
147355	428,732	5	6					
147585	1,838,554	94 (7kb)						
148072	6,784	32						
148216	2,454,082	18	471*					
148541	2,658,540	1	446 (21kb)					
150322	246,382	105 (19kb)						
150354	1,412,860	32	104 (7kb)		6630 (350kb)	272 (16kb)		
150410	1,281,136	1	1109 (54kb)	114 (9kb)				
150583	1,410	2						
150687	1,637,790	53,637 (1.5mb)			357 (41kb)		14	
151303	11,598	12	7					

*Over representation of reads mapping at one locus, likely due to PCR amplification during library preparation.

¹Numbers in the parentheses represent total length along a *Campylobacter* genome with >50 mapped reads. These lengths do not consider the coverage of the assemblies.

<https://doi.org/10.1371/journal.pntd.0010815.t001>

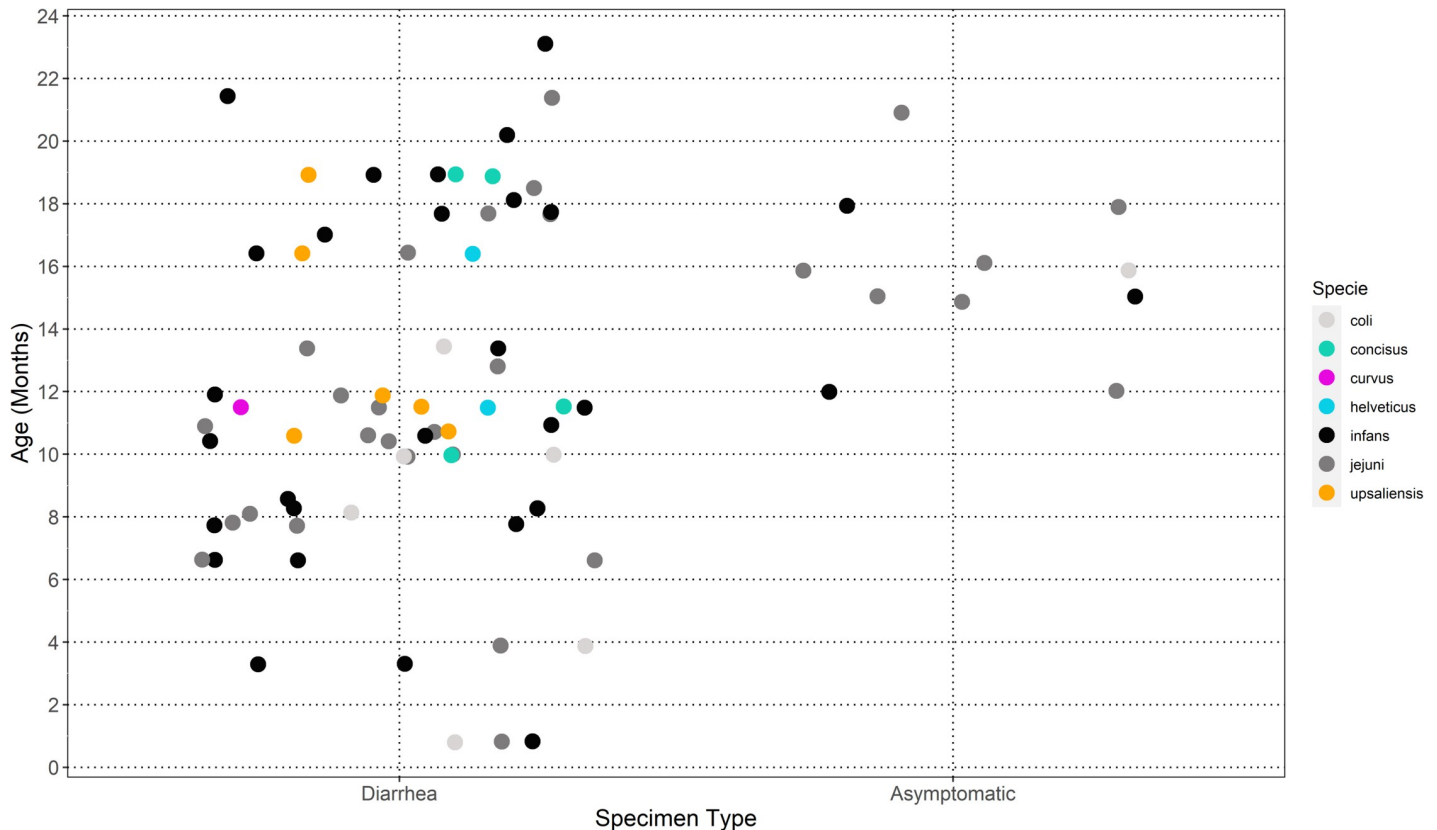


Fig 1. Campylobacter species in symptomatic (diarrhea) and asymptomatic (monthly) fecal samples from children under two years of age. *Campylobacter concisus*, *Campylobacter curvus*, *Campylobacter helveticus*, and *Campylobacter upsaliensis* are only found in diarrheal fecal samples from children over 10 months of age.

<https://doi.org/10.1371/journal.pntd.0010815.g001>

and other bacterial genera. These regions were often the only sites in certain *Campylobacter* reference sequences where reads mapped. For example, reads mapped to *tet(M)* gene and its associated mobile element of *Campylobacter lanienae* (CP015578) (S1A Fig). Reads that mapped to any of these regions in a *Campylobacter* genome were determined to be non-confirmatory reads for only one species and were eliminated from further analysis (S2 Table).

Based on the high prevalence of *C. infans* in these samples and the previous findings of *C. infans* in stool samples from infants in various LMIC [36], the 26 *C. infans* positive samples were examined in detail. Fourteen of the 26 *C. infans* positive samples had between 1 and 32 reads mapped to the *C. infans* genome. BLASTn analysis of all 176 total reads from these 14 samples against the NCBI nr/nt nucleotide database resulted in identification of *C. infans* strain 19S00001 as the only BLASTn match, or the highest scoring match. Additionally, the reads from the 14 samples did not represent common sequence reads between these different samples. Twelve of the 26 *C. infans* positive samples had >50 paired-reads. Three samples 25500, 150687 (S1B Fig, as an example) and 19879 from diarrheal infants provided over 10,000 paired-reads mapping to the *C. infans* genome, resulting in approximately 1x, 5x and 9x coverages, respectively.

Co-infection with multiple *Campylobacter* spp

The metagenomics analysis by reference assembly in Geneious detected 27 samples (65.9%) of analyzed stool with more than a single species of *Campylobacter* present in the same stool

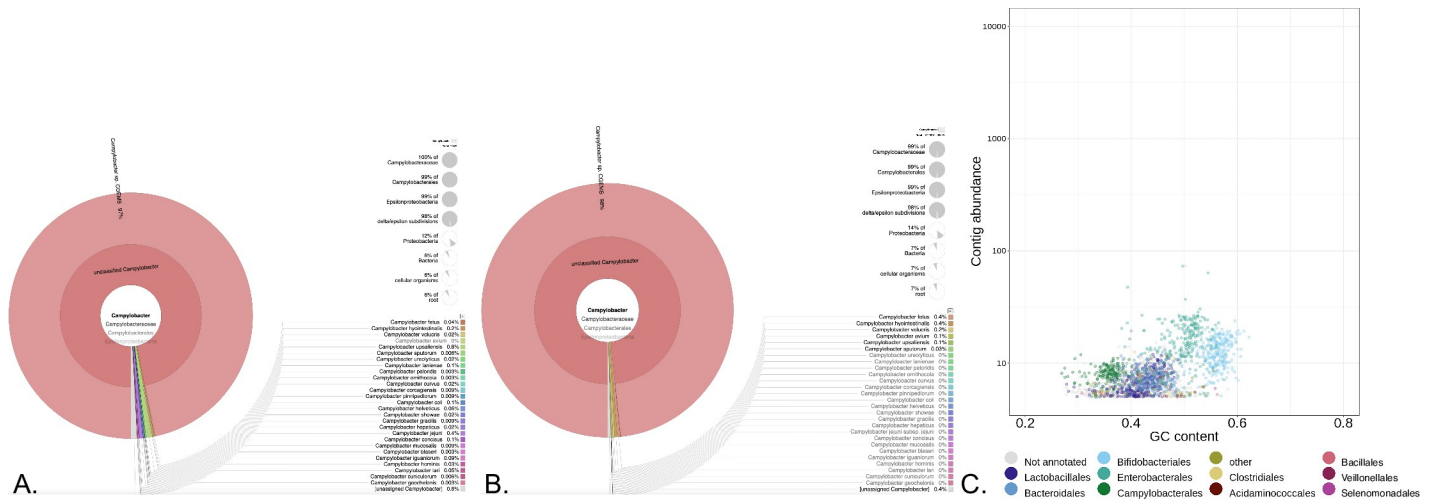


Fig 3. MetaWRAP analysis of sample #150687 that had lower amount of sequence reads associated with *Campylobacter*, but had a higher overall number of sequence reads generated compared to other samples in the study. Standard pipeline analysis was still able use shotgun metagenomics to determine the presence of *C. infans* in the stool sample and identify multiple *Campylobacter* species present in the stool sample. **A.** Kronogram used to represent/display that 6% of total sequencing reads and 97% of the *Campylobacter* specific sequencing reads are matching *C. infans*, and provides a breakdown of the number of sequence reads generated for that sample the match *Campylobacter* species at the different taxonomic levels. **B.** Kronogram used to represent/display that when present in high enough concentrations as for this sample, identified *Campylobacter* species can be partially or completely genome assembled using a standard analysis pipeline and provides a breakdown of the number of assembled reads generated for that sample the match *Campylobacter* species at the different taxonomic levels. For this sample, 7% of all the assembled reads and 98% of the *Campylobacter* contigs for this sample were identified as *C. infans*. **C.** Shotgun metagenomic analysis pipelines can also rely on software to bin reads/contigs based on various parameters such as GC content of the data. Blobplot of all the assembled contigs that were binned from the sample based on percentage of GC content at the Order level of taxonomic identification, which demonstrates *Campylobacteriales* was binned during the process of the shotgun metagenomic analysis even with high levels of sequence reads not associated with *Campylobacter* and many additional bins generated from the stool sample, thus many different aspects of metagenomic analysis can identify *Campylobacter* species from stool samples. The *Campylobacter* bin contained 150 contigs that equals 1,676,492 bp or 95.6% of the *C. infans* genome. However, there were 171 contigs overall assembled to various *Campylobacter* species equaling 1,781,027 bp from the sample including 11 additional *C. infans* contigs that were not binned for a total *C. infans* genome of 1,740,218 bp or 99.2% of the complete genome.

<https://doi.org/10.1371/journal.pntd.0010815.g003>

MetaWRAP analysis also identified other organisms in each of these samples including potential diarrheal pathogens and intestinal flora. An advantage of shotgun metagenomics is the identification of the fecal microbiome in the samples from Peruvian children with *Campylobacter* species, thus allowing additional analysis of the fecal microbiome like the impact of high *Campylobacter* abundance or number of *Campylobacter* species on the alpha and/or beta diversity of the fecal microbiome (S2 Fig). For infant diarrheal sample 25500, 7% of the gut microbiome was identified as *Escherichia coli* and 32% was *Bifidobacterium sp.* For both samples 150687 and 19879, *E. coli* reads were the larger portion of the gut microbiome at 41% and 37% and *Bifidobacterium sp.* was 16% and 0%, respectively. Besides *E. coli*, 67% of reads for sample 19879 were from the family *Enterobacteriaceae* including 9% reads from *Klebsiella sp.* An additional advantage of shotgun metagenomics is that each of these dominate bacteria in the fecal microbiome from the different samples can be genome assembled and binned during the pipeline analysis, as was demonstrated in the majority of the fecal samples in this study (S3 Fig). Having these bins then allows for further studies such as the role these commensals or other enteric pathogens have in interacting with different *Campylobacter* species in the infants intestinal tract.

Discussion

We have previously demonstrated that stools from which *Campylobacter* is identified using qPCR, but *C. jejuni* and *C. coli* are not identified, have an elevated risk of watery diarrhea suggesting that at least some of these less well-known *Campylobacter* species result in clinical

Table 2. MetaWRAP analysis of stool samples from 44 fecal samples from children under two years of age living in Iquitos, Peru which were positive by PCR for the Campylobacter 16S assay but negative for the *cadF* gene.

Sample ID	Percentage of All Reads to <i>Campylobacter</i>	Percentage of All <i>Campylobacter</i> Reads				Number of <i>Campylobacter</i> contigs ¹	Number of <i>C. infans</i> contigs ²	Number of <i>C. jejuni</i> contigs ²
		<i>C. infans</i>	<i>C. jejuni</i>	<i>C. coli</i>	<i>C. upsaliensis</i>			
13323	0.7%	3%	32%	3%	0.3%	1 (1,074 bp)	0	1 (1,074 bp)
13812	0.004%	6%	24%	0%	41%	0	0	0
15704	0.1%	3%	2%	0%	10%	4 (9,380 bp)	0	0
18102	0.03%	42%	24%	2%	1%	0	0	0
18196	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
19879	6.0%	98%	0.3%	0.05%	0.3%	99 (856,787 bp)	99 (856,787 bp)	0
22997	0.2%	3%	38%	0%	3%	0	0	0
25500	23%	98%	0.3%	0.09%	0.3%	15 (20,789 bp)	15 (20,789 bp)	0
28671	0.5%	0%	19%	59%	0.7%	0	0	0
31204	0.2%	56%	2%	0.2%	36%	3 (5,689)	1 (1,209 bp)	1 (3,198 bp)
34891	0.4%	23%	2%	4%	12%	0	0	0
37185	0.2%	0%	33%	4%	0%	0	0	0
41480	0.004%	20%	20%	0%	10%	0	0	0
52141	3%	0%	39%	4%	0.4%	2 (2,047 bp)	0	0
57052	0.04%	2%	14%	3%	0%	0	0	0
57847	0.009%	4%	7%	6%	0%	0	0	0
63792	0.002%	50%	25%	0%	0%	0	0	0
65146	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
76987	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
78740	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
97077	0.04%	2%	19%	7%	3%	1 (1,154 bp)	0	0
97770	0.4%	64%	6%	0.4%	0.1%	4 (9,603)	2 (6,546 bp)	0
97915	0.01%	4%	4%	4%	0%	0	0	0
101367	0.2%	0.3%	74%	1%	0%	3 (4,677 bp)	0	1 (1,391 bp)
108338	0.5%	80%	3%	1%	0.6%	3 (6,989 bp)	1 (1,129 bp)	1 (3,509 bp)
118291	0.008%	34%	12%	7%	7%	0	0	0
118488	0%	0%	0%	0%	0%	0	0	0
124035	8%	0%	49%	6%	0.3%	2 (3,381 bp)	0	0
124473	0.3%	4%	42%	4%	4%	0	0	0
130169	0.01%	0%	13%	10%	0%	0	0	0
132208	0.04%	25%	0%	0%	0%	1 (1,561 bp)	0	0
147132	0.7%	0.5%	33%	4%	0.2%	13 (17,707 bp)	0	6 (9,464 bp)
147161	0.9%	5%	4%	7%	12%	143 (220,649 bp)	6 (10,651 bp)	3 (5,303 bp)
147355	0.008%	15%	23%	8%	0%	0	0	0
147585	0.02%	42%	10%	8%	3%	1 (1,134 bp)	0	0
148072	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
148216	0.04%	4%	3%	0%	0%	0	0	0
148541	0.03%	0.4%	42%	0.4%	0.4%	0	0	0
150322	0.2%	63%	4%	2%	1%	0	0	0
150354	0.7%	0%	0.4%	0.1%	96%	0	0	0
150410	0.2%	0%	30%	5%	0%	0	0	0
150583	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
150687	6.0%	97%	0.4%	0.1%	0.8%	171 (1,781,027 bp)	161 (1,740,218 bp)	0

(Continued)

Table 2. (Continued)

Sample ID	Percentage of All Reads to <i>Campylobacter</i>	Percentage of All <i>Campylobacter</i> Reads				Number of <i>Campylobacter</i> contigs ¹	Number of <i>C. infans</i> contigs ²	Number of <i>C. jejuni</i> contigs ²
		<i>C. infans</i>	<i>C. jejuni</i>	<i>C. coli</i>	<i>C. upsaliensis</i>			
151303	N/A	N/A	N/A	N/A	N/A	N/A	N/A	

¹Numbers in the parentheses represent total amount of *Campylobacter* sequence data from the sample that are assembled into contigs.

²Numbers in the parentheses represent the total amount of base pairs of the genome assembled into contigs for those *Campylobacter* species from that sample.

<https://doi.org/10.1371/journal.pntd.0010815.t002>

diarrhea in LMICs [5]. Herein, we demonstrated a method that identified *Campylobacter* species by mapping WSMS reads to the genomes of 27 different *Campylobacter* species using the reference assembler within Geneious Prime software, although, other assemblers could be used. Indeed, the metaWRAP pipeline retrieved similar results, identifying multiple *Campylobacter* species. The WSMS of stool samples that were previously identified as positive for *Campylobacter* by CIDT via 16S qPCR demonstrated how this second global CIDT process can broadly identify *Campylobacter* species and co-infections with multiple *Campylobacters*. We demonstrate that mapping the sample reads against the *Campylobacter* species genomes identified 41 out of 44 samples positive for *Campylobacter*. As previously observed [36], co-infections with multiple *Campylobacter* species, including *C. infans* and other atypical *Campylobacter* species including *C. upsaliensis* and *C. concisus* was demonstrated.

In this proof-of-principle study, we selected a group of archived samples from children who had tested positive for *Campylobacter* with a non-selective PCR probe for all *Campylobacter* species, but were negative for the *cadF* gene, one of the most common targets used to identify *C. jejuni* and *C. coli* from other *Campylobacter* species, particularly in LMICs. To our surprise, the majority of the samples contained *C. jejuni* and/or *C. coli*, despite the fact that the metagenomics and qPCR assays were run on the same extraction. This demonstrates that PCR using standard primers for the detection of the *cadF* gene is less sensitive than previously recognized.

Whole-genome shotgun metagenomic sequencing (WSMS) as a supportive, species defining CIDT strategy provides more than just confirmation of a *Campylobacter* species. First, our results demonstrate that WSMS can identify *Campylobacter* species even when the sample provides limited sequencing reads (as low as 1,400 reads), but the ability to identify the *Campylobacter* species depends on the amount of these organisms within the microbiome. When levels of *Campylobacter* were high, as seen for sample 150687 (S1B Fig), many distinct reads from around the genome were mapped. In fact, for samples with high read counts to a specific *Campylobacter* species, remapping the reads to diverse variable regions of the species from multiple strains, such as lipooligosaccharide and capsule biosynthesis loci [37–39], will provide additional details to the strains in the samples. At lower levels, where fewer reads were mapped, the reads discriminatory power needs to be addressed through BLASTn analysis and elimination of mapped reads that were associated with non-discriminatory regions like the rRNA loci or transposons. Species-non-confirmatory reads, such as reads mapping to rRNA loci, can map to more than one genome, and in some cases different types of bacteria. In cases where there are few reads that identify a species, species-non-confirmatory reads would require a BLASTn search, which would identify different species and reduce the level of identification accordingly. When there are species-confirmatory reads and non-confirmatory reads, the non-confirmatory reads may help validate one of species but should be used with caution to only confirm based on the species-confirmatory reads. Our study indicates that removal of bacterial conserved sequences such as rRNA loci is required to reduce false species calls. This has been

described previously [40]. Finally, despite identifying *Campylobacter* species reads, samples with very few total sequence reads were not very useful for species confirmation or additional analysis since there are few confirmatory reads mapped to the pathogens being detected or to other organisms of the microbiome. Instead, samples that provide a few million reads would be more useful.

WSMS requires only a single sequencing method and given the breadth of species in the *Campylobacter* genus, a relatively efficient approach to unequivocal species identification. PCR and qPCR require distinct primer pairs for each organism to be detected and possible changes to conditions. Also, PCR/qPCR requires additional validation when altering the detectable species. With 27 possible *Campylobacter* species PCR/qPCR specific primer design quickly becomes impractical. Even designing a single 16S gene amplicon-based sequencing method requires determination and validation of new PCR conditions and primers to ensure each *Campylobacter* species may be distinguished. Using shotgun metagenomics, we utilize a method that works when there is DNA and can detect targeted organisms when there are enough reads. In this study, we were able to assemble a *C. infans* genome at over 95% complete for source tracking and/or transmission dynamics for one of our samples, which is useful given the nearly universal inability to culture this pathogen and lack of standardized culture protocols.

Furthermore, even validated primers like the *cadF* gene primers for *C. jejuni* and/or *C. coli* can produce false negatives, as demonstrated by the results of this study where over 60% of the samples were actually positive for *C. jejuni* and/or *C. coli* after a negative *cadF* qPCR. Third, WSMS allows multiple *Campylobacter* species, including different strains of the same species to be identified from a single sample. Many of the species within the *Campylobacter* genus are more than 10% different, so most reads only have the potential to map to the genome of one species. There are some regions that are more conserved (<10% different) between species and reads mapping to these regions are non-confirmatory. Finally, WSMS provides sequence information regarding other members of the microbiota, identify other enteric pathogens, and co-infections between *Campylobacter* and these other enteric bacterial organisms.

The most notable finding of this study was the number of samples which contained concurrent infections with multiple species of *Campylobacter*, which has recently been observed in LMIC settings [36,41]. This suggests that prolonged and persisting mixed infections are occurring in this population. The infant population under study is chronically undernourished with a high level of intestinal inflammation, which may facilitate prolonged carriage, such as the increased rate of isolation of *C. concisus* in patients with inflammatory bowel disease [42]. Chronic infection, in some cases over years, have been reported in immunocompromised hosts [43], and even in hosts with no evident compromised immune system as a result of microbial factors, such as a cell invasion protein [44]. Given that *Campylobacter* is closely related to *Helicobacter*, the ability of the microbe to persist in the host as a result of microbial factors is also plausible. Lastly, changes in the microbiome in the chronically infected population may be more permissive for the establishment and maintenance of *Campylobacter*. We have previously established that specific taxa in this cohort are associated with *Campylobacter* infection [44] and identified microbes that are associated with an increased risk of infection (*Ruminococcus gnavus* (ASV23), *Dialister* (ASV26), *Prevotella* (ASV204 and ASV275)) and taxa that are associated with a decreased risk of *Campylobacter* infection (*Bacteroides ovatus* (ASV40), *Ruminococcus toraues* (ASV242), *Bacteroides* (ASV27) and *Lachnospiraceae*).

C. infans was identifiable in 59.1% (26/44) fecal samples that were determined to have *Campylobacter* by a genus specific PCR assay, and that did not amplify the *cadF* gene. This potential new species of *Campylobacter* has only been isolated once from a patient in Europe, yet it was shown to be prevalent in fecal samples from children in Southeast Asia and Africa [36, 45, 46].

There is still limited information associated with the presence of this species and the occurrence of gastrointestinal diseases. Further studies are needed to inform its potential pathogenicity. Given the co-infection with *C. infans* and *C. concisus* in three infants, we hypothesize that *C. infans* presence may be associated with previously described decompartmentalization of the gastrointestinal tract with an increased presence of oropharyngeal flora in the stool of undernourished children [47].

Polyclonal pathogen populations facilitate the acquisition of antibiotic resistance in the infected host through the selection of resistant strains [48]. Our data suggest that co-infection of multiple *Campylobacter* species in settings of intense transmission is common and may facilitate the development of resistance and propagation of multidrug resistant *Campylobacter* in response to clinical therapeutics in LMIC at a higher rate than in the US or Europe where monoclonal *Campylobacter* infections are likely the rule. The emergence of multidrug resistant *Campylobacter* is often considered attributable principally to antibiotic usage in livestock and poultry production. However, our findings could suggest that in LMIC contexts, human antibiotic use may also play an important role but further investigation is needed to draw a definitive conclusion.

Conclusion

WSMS was able to efficiently identify *Campylobacter* species other than *C. jejuni* and *C. coli*. Other non-*jejuni* and non-*coli* *Campylobacter* species are prevalent among children in Peru, of which *Candidatus* “*C. infans*” was the most frequently identified *Campylobacter* species, although *C. upsaliensis*, *C. concisus*, *C. helveticus* and *C. curvus* were also identified. Children in Peru infected with *Campylobacter* often have multiple species present in the intestinal tract at one time.

Supporting information

S1 Table. Primers and probes for the detection of *Campylobacter* spp., *Campylobacter jejuni/coli*, and *Shigella* spp.

(DOCX)

S2 Table. Nondiscriminatory Loci/Genes Identified During Analysis^a. These loci and genes by themselves are not discriminatory for *Campylobacter* species and/or Epsilonbacteria.

(DOCX)

S1 Fig. The whole-genome sequencing shotgun metagenomic reads reference from sample 150687 assembled to *Campylobacter* genomes. A. A portion of the *C. lariena* genome (from nucleotides 275,000–294,500) possessing an example of a non-confirmatory genomic region, a mobile element. The metagenomic sequencing reads (gray rectangles below the genome map) only map across a portion of the mobile element that contains the *tet(M)* gene. The absence of reads mapping in the adjacent genomic regions is representative of the rest of the genome. B. Reads (black rectangles below the genome map) mapped to the entire *C. infans* genome showing read coverage of ~5X. Genomic regions containing no reads correspond to loci, such as bacteriophage, restriction modification genes, that are variably present among strains within a *Campylobacter* species.

(DOCX)

S2 Fig. Fecal microbiome analysis of Peruvian children used in this study from the whole-genome sequencing shotgun metagenomic reads, all samples were rarefied to 50,000 sequence reads for all the alpha and beta diversity analysis. A. Shannon diversity of the fecal samples of children based on the number of *Campylobacter* species present in the stool sample,

overall, no significant difference (p-value = (0.24). B. Bray-Curtis PCoA plot of fecal microbiome diversity based on the number of *Campylobacter* species present in the stool sample, overall, no significant difference ($R^2 = 0.1948$, p-value = 0.273). C. Taxonomic barplot of the abundance of the top 12 genera present in the stool samples from the children in this study, 34/44 (77.3%) samples are represented as remaining 10 samples did not have enough reads. D. Shannon diversity of the fecal samples of children based on the overall abundance of *Campylobacter* present in the stool sample, overall, no significant difference (p-value = (0.733). E. Bray-Curtis PCoA plot of fecal microbiome diversity based on the abundance of *Campylobacter* present in the stool sample, overall, no significant difference ($R^2 = 0.0653$, p-value = 0.149). High abundance $\geq 0.5\%$ of all sequence reads, Low abundance $\leq 0.5\%$ of all sequence reads. (DOCX)

S3 Fig. Blobplots from the binning aspect of analysis from the MetaWRAP pipeline that indicates the phylum for each bin for 32/44 (72.7%) of the stool samples, as the remaining 12 samples did not have enough sequence data to either use the MetaWRAP pipeline or generate bins. The blobplots provide overview of the number of total contigs present in each bin for a sample that were assembled for each of the samples using metaSPAdes assembler in the MetaWRAP pipeline. Thus, demonstrating even with a lower number of sequence reads for a stool sample, there is the potential for whole-genome sequencing shotgun metagenomics to result in partial or complete assemble of pathogen genomes for source tracking and transmission dynamics studies in low- and middle-income countries. (DOCX)

S1 Data. Age (months), sex and specimen type of samples included in the analysis. (CSV)

Author Contributions

Conceptualization: Craig T. Parker, Francesca Schiaffino, Maribel Paredes Olortegui, Greisi E. Curico Huansi, Kerry K. Cooper, Margaret N. Kosek.

Data curation: Craig T. Parker, Francesca Schiaffino, Steven Huynh, Maribel Paredes Olortegui, Pablo Peñataro Yori, Paul F. Garcia Bardales, Tackeshy Pinedo Vasquez, Greisi E. Curico Huansi, Katia Manzanares Villanueva, Wagner V. Shapiama Lopez, Kerry K. Cooper, Margaret N. Kosek.

Formal analysis: Craig T. Parker, Francesca Schiaffino, Pablo Peñataro Yori, Kerry K. Cooper, Margaret N. Kosek.

Funding acquisition: Craig T. Parker, Maribel Paredes Olortegui, Pablo Peñataro Yori, Margaret N. Kosek.

Investigation: Craig T. Parker, Francesca Schiaffino, Steven Huynh, Maribel Paredes Olortegui, Pablo Peñataro Yori, Paul F. Garcia Bardales, Tackeshy Pinedo Vasquez, Greisi E. Curico Huansi, Katia Manzanares Villanueva, Wagner V. Shapiama Lopez, Kerry K. Cooper, Margaret N. Kosek.

Methodology: Craig T. Parker, Steven Huynh, Maribel Paredes Olortegui, Kerry K. Cooper, Margaret N. Kosek.

Project administration: Pablo Peñataro Yori, Margaret N. Kosek.

Resources: Margaret N. Kosek.

Supervision: Craig T. Parker, Maribel Paredes Olortegui, Pablo Peñataro Yori, Margaret N. Kosek.

Validation: Craig T. Parker.

Visualization: Craig T. Parker, Francesca Schiaffino.

Writing – original draft: Craig T. Parker, Francesca Schiaffino, Margaret N. Kosek.

Writing – review & editing: Craig T. Parker, Francesca Schiaffino, Steven Huynh, Maribel Paredes Olortegui, Pablo Peñataro Yori, Paul F. Garcia Bardales, Tackeshy Pinedo Vasquez, Greisi E. Curico Huansi, Katia Manzanares Villanueva, Wagner V. Shapiama Lopez, Kerry K. Cooper, Margaret N. Kosek.

References

1. Butzler JP. Campylobacter, from obscurity to celebrity. *Clin Microbiol Infect*. 2004; 10(10):868–76. <https://doi.org/10.1111/j.1469-0691.2004.00983.x> PMID: 15373879
2. 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(1):e2036. <https://doi.org/10.1371/journal.pntd.0002036> PMID: 23383356
3. Lee G, Paredes Olortegui M, Penataro Yori P, Black RE, Caulfield L, Banda Chavez C, et al. Effects of Shigella-, Campylobacter- and ETEC-associated diarrhea on childhood growth. *Pediatr Infect Dis J*. 2014; 33(10):1004–9. <https://doi.org/10.1097/INF.0000000000000351> PMID: 25361185
4. 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(9):1171–9. <https://doi.org/10.1093/cid/ciw542> PMID: 27501842
5. Francois R, Yori PP, Rouhani S, Sigvas Salas M, Paredes Olortegui M, Rengifo Trigoso D, et al. The other Campylobacters: Not innocent bystanders in endemic diarrhea and dysentery in children in low-income settings. *PLoS Negl Trop Dis*. 2018; 12(2):e0006200. <https://doi.org/10.1371/journal.pntd.0006200> PMID: 29415075
6. Lastovica AJ. Emerging Campylobacter spp.: the tip of the iceberg. *Clinical Microbiology Newsletter*. 2006; 28(7):49–56.
7. Platts-Mills JA, Liu J, Gratz J, Mduma E, Amour C, Swai N, et al. Detection of Campylobacter in stool and determination of significance by culture, enzyme immunoassay, and PCR in developing countries. *J Clin Microbiol*. 2014; 52(4):1074–80. <https://doi.org/10.1128/JCM.02935-13> PMID: 24452175
8. Iwamoto M, Huang JY, Cronquist AB, Medus C, Hurd S, Zansky S, et al. Bacterial enteric infections detected by culture-independent diagnostic tests—FoodNet, United States, 2012–2014. *MMWR Morb Mortal Wkly Rep*. 2015; 64(9):252–7. PMID: 25763878
9. Langley G, Besser J, Iwamoto M, Lessa FC, Cronquist A, Skoff TH, et al. Effect of Culture-Independent Diagnostic Tests on Future Emerging Infections Program Surveillance. *Emerg Infect Dis*. 2015; 21(9):1582–8. <https://doi.org/10.3201/eid2109.150570> PMID: 26291736
10. Huang JY, Henao OL, Griffin PM, Vugia DJ, Cronquist AB, Hurd S, et al. Infection with Pathogens Transmitted Commonly Through Food and the Effect of Increasing Use of Culture-Independent Diagnostic Tests on Surveillance—Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 2012–2015. *MMWR Morb Mortal Wkly Rep*. 2016; 65(14):368–71. <https://doi.org/10.15585/mmwr.mm6514a2> PMID: 27077946
11. Janda JM, Abbott SA. Culture-independent diagnostic testing: have we opened Pandora's box for good? *Diagn Micr Infect Dis*. 2014; 80(3)(3):171–6. <https://doi.org/10.1016/j.diagmicrobio.2014.08.001> PMID: 25200256
12. 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. *The Lancet*. 2016; 388(10051):1291–301. [https://doi.org/10.1016/S0140-6736\(16\)31529-X](https://doi.org/10.1016/S0140-6736(16)31529-X) PMID: 27673470
13. Platts-Mills JA, Liu J, Rogawski ET, Kabir F, Lertsethtakarn P, Sigvas M, et al. Use of quantitative molecular diagnostic methods to assess the aetiology, burden, and clinical characteristics of diarrhoea in children in low-resource settings: a reanalysis of the MAL-ED cohort study. *Lancet Glob Health*. 2018.
14. Berenger B, Chui L, Reimer AR, Allen V, Alexander D, Domingo MC, et al. Canadian Public Health Laboratory Network position statement: Non-culture based diagnostics for gastroenteritis and implications

- for public health investigations. *Can Commun Dis Rep*. 2017; 43(12):279–81. <https://doi.org/10.14745/ccdr.v43i12a06> PMID: 29770061
15. McAdam AJ. Unforeseen Consequences: Culture-Independent Diagnostic Tests and Epidemiologic Tracking of Foodborne Pathogens. *J Clin Microbiol*. 2017; 55(7):1978–9. <https://doi.org/10.1128/JCM.00678-17> PMID: 28468857
 16. Shea S, Kubota KA, Maguire H, Gladbach S, Woron A, Atkinson-Dunn R, et al. Clinical Microbiology Laboratories' Adoption of Culture-Independent Diagnostic Tests Is a Threat to Foodborne-Disease Surveillance in the United States. *J Clin Microbiol*. 2017; 55(1):10–9. <https://doi.org/10.1128/JCM.01624-16> PMID: 27795338
 17. Marder EP, Cieslak PR, Cronquist AB, Dunn J, Lathrop S, Rabatsky-Ehr T, et al. Incidence and Trends of Infections with Pathogens Transmitted Commonly Through Food and the Effect of Increasing Use of Culture-Independent Diagnostic Tests on Surveillance—Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 2013–2016. *MMWR Morb Mortal Wkly Rep*. 2017; 66(15):397–403. <https://doi.org/10.15585/mmwr.mm6615a1> PMID: 28426643
 18. Investigators M-EN. 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 Suppl 4:S193–206.
 19. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome*. 2018; 6(1):158. <https://doi.org/10.1186/s40168-018-0541-1> PMID: 30219103
 20. Rotmistrovsky K, Agarwala R. BMTagger: best match tagger for removing human reads from metagenomics datasets. 2011 [Available from: <http://ftp.ncbi.nlm.nih.gov/pub/agarwala/bmtagger/>].
 21. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res*. 2017; 27(5):824–34. <https://doi.org/10.1101/gr.213959.116> PMID: 28298430
 22. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013; 29(8):1072–5. <https://doi.org/10.1093/bioinformatics/btt086> PMID: 23422339
 23. Lu J, Salzberg SL. Ultrafast and accurate 16S rRNA microbial community analysis using Kraken 2. *Microbiome*. 2020; 8(1):124. <https://doi.org/10.1186/s40168-020-00900-2> PMID: 32859275
 24. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*. 2019; 20(1):257. <https://doi.org/10.1186/s13059-019-1891-0> PMID: 31779668
 25. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*. 2011; 12:385. <https://doi.org/10.1186/1471-2105-12-385> PMID: 21961884
 26. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*. 2019; 7:e7359. <https://doi.org/10.7717/peerj.7359> PMID: 31388474
 27. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. *Nat Methods*. 2014; 11(11):1144–6. <https://doi.org/10.1038/nmeth.3103> PMID: 25218180
 28. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*. 2015; 25(7):1043–55. <https://doi.org/10.1101/gr.186072.114> PMID: 25977477
 29. Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M. Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Front Genet*. 2013; 4:237. <https://doi.org/10.3389/fgene.2013.00237> PMID: 24348509
 30. Droge J, Gregor I, McHardy AC. Taxator-tk: precise taxonomic assignment of metagenomes by fast approximation of evolutionary neighborhoods. *Bioinformatics*. 2015; 31(6):817–24. <https://doi.org/10.1093/bioinformatics/btu745> PMID: 25388150
 31. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014; 30(14):2068–9. <https://doi.org/10.1093/bioinformatics/btu153> PMID: 24642063
 32. Dabdoub S. kraken-biom: Enabling interoperative format conversion for Kraken results (Version 1.2) [Software] 2016 [Available from: <https://github.com/smdabdoub/kraken-biom>].
 33. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013; 8(4):e61217. <https://doi.org/10.1371/journal.pone.0061217> PMID: 23630581
 34. Lahti L, Shetty S, et al. Tools for microbiome analysis in R. 2017 [Available from: <http://microbiome.github.com/microbiome>].

35. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. R package version 2.5–7 2020 [Available from: <https://CRAN.R-project.org/package=vegan>].
36. Bian X, Garber JM, Cooper KK, Huynh S, Jones J, Mills MK, et al. Campylobacter Abundance in Breastfed Infants and Identification of a New Species in the Global Enterics Multicenter Study. *mSphere*. 2020; 5(1). <https://doi.org/10.1128/mSphere.00735-19> PMID: 31941810
37. Heikema AP, Strepis N, Horst-Kreft D, Huynh S, Zomer A, Kelly DJ, et al. Biomolecule sulphation and novel methylations related to Guillain-Barre syndrome-associated Campylobacter jejuni serotype HS:19. *Microb Genom*. 2021; 7(11).
38. Parker CT, Gilbert M, Yuki N, Endtz HP, Mandrell RE. Characterization of lipooligosaccharide-biosynthetic loci of Campylobacter jejuni reveals new lipooligosaccharide classes: evidence of mosaic organizations. *J Bacteriol*. 2008; 190(16):5681–9. <https://doi.org/10.1128/JB.00254-08> PMID: 18556784
39. 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. *PLoS One*. 2015; 10(12):e0144349. <https://doi.org/10.1371/journal.pone.0144349> PMID: 26630669
40. Peterson CL, Alexander D, Chen JC, Adam H, Walker M, Ali J, et al. Clinical Metagenomics Is Increasingly Accurate and Affordable to Detect Enteric Bacterial Pathogens in Stool. *Microorganisms*. 2022; 10(2).
41. Terefe Y, Deblais L, Ghanem M, Helmy YA, Mummied B, Chen D, et al. Co-occurrence of *Campylobacter* Species in Children From Eastern Ethiopia, and Their Association With Environmental Enteric Dysfunction, Diarrhea, and Host Microbiome. *Front Public Health*. 2020; 8:99.
42. Kumar S, Kumar A. Microbial pathogenesis in inflammatory bowel diseases. *Microb Pathog*. 2022; 163:105383. <https://doi.org/10.1016/j.micpath.2021.105383> PMID: 34974120
43. Barker CR, Painset A, Swift C, Jenkins C, Godbole G, Maiden MCJ, et al. Microevolution of *Campylobacter jejuni* during long-term infection in an immunocompromised host. *Sci Rep*. 2020; 10(1):10109.
44. Rouhani S, Griffin NW, Yori PP, Olortegui MP, Sigvas Salas M, Rengifo Trigoso D, et al. Gut Microbiota Features Associated With *Campylobacter* Burden and Postnatal Linear Growth Deficits in a Peruvian Birth Cohort. *Clin Infect Dis*. 2020; 71(4):1000–7.
45. Duim B, van der Graaf-van Bloois L, Timmerman A, Wagenaar JA, Flipse J, Wallinga J, et al. Complete Genome Sequence of a Clinical Campylobacter Isolate Identical to a Novel Campylobacter Species. *Microbiol Resour Announc*. 2021; 10(7). <https://doi.org/10.1128/MRA.00721-20> PMID: 33602730
46. Flipse J, Duim B, Wallinga JA, De Wijkerslooth LRH, Graaf-Van Bloois LVD, Timmerman AJ, et al. A Case of Persistent Diarrhea in a Man with the Molecular Detection of Various Campylobacter species and the First Isolation of candidatus Campylobacter infans. *Pathogens*. 2020; 9(12):1003.
47. Vonaesch P, Morien E, Andrianonimadana L, Sanke H, Mbecko JR, Huus KE, et al. Stunted childhood growth is associated with decompartmentalization of the gastrointestinal tract and overgrowth of oropharyngeal taxa. *Proc Natl Acad Sci U S A*. 2018; 115(36):E8489–E98. <https://doi.org/10.1073/pnas.1806573115> PMID: 30126990
48. Caballero JD, Wheatley RM, Kapel N, López-Causapé C, Van der Schalk T, Quinn A, et al. Polyclonal pathogen populations accelerate the evolution of antibiotic resistance in patients. *bioRxiv*. 2021:2021.12.10.472119.