Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genome-resolved metagenomics of Campylobacter at a wildlife–livestock–human interface in Uganda reveals novel species

David Hayman

D.T.S.Hayman@massey.ac.nz

Massey University https://orcid.org/0000-0003-0087-3015

Valter Almeida

Massey University https://orcid.org/0000-0001-8380-3391

Matthew Knox

Massey University

Kim Handley

The University of Auckland https://orcid.org/0000-0003-0531-3009

Anne C. Midwinter

Massey University https://orcid.org/0000-0001-5731-3233

Patrick Biggs

Massey University https://orcid.org/0000-0002-0285-4101

Agata Dziegiel

Quadram Institute Bioscience https://orcid.org/0000-0002-6148-1771

Gladys Kalema-Zikusoka

Conservation Through Public Health https://orcid.org/0000-0003-3473-3205

Stephen Rubanga

Conservation Through Public Health

Alex Ngabirano

Mubare Biodiversity Conservation (MBC)

Willy Valdivia-Granda

Orion Integrated Biosciences Inc.

Hayley MacGregor

University of Sussex

Nigel French

Massey University https://orcid.org/0000-0002-6334-0657

James Lloyd-Smith

University of California, Los Angeles https://orcid.org/0000-0001-7941-502X

Alison Mather

Quadram Institute https://orcid.org/0000-0001-6513-3515

Article

Keywords:

Posted Date: June 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9625640/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations:
There is **NO** Competing Interest.

Informed consent was obtained from human study participants where samples were collected prospectively as part of the study in accordance with the approved study protocols. Anonymised residual diagnostic samples were used for approved secondary testing, and individual consent was not required for these samples under the relevant ethics approvals. The study was approved by the Mbarara University of Science and Technology Research Ethics Committee, the Uganda National Council for Science and Technology, and the Massey University Human Ethics Committee.

Genome-resolved metagenomics of *Campylobacter* at a wildlife–livestock–human interface in Uganda reveals novel species

Valter Almeida, PhD[a] · Matthew A. Knox, PhD[a] · Kim M. Handley, PhD[b] · Anne C. Midwinter, PhD[a] · Prof Patrick J. Biggs, PhD[a,c,d] · Agata H. Dziegiel, PhD[e,f] · Gladys Kalema-Zikusoka, MS[g] · Stephen Rubanga[g] · Alex Ngabirano[h] · Willy A. Valdivia-Granda, PhD[i] · Hayley MacGregor, PhD[j] · Dist Prof Nigel P. French, PhD[a] · Prof Alison E. Mather, PhD[e,f,k] · Prof James O. Lloyd-Smith, PhD[l,m] · Prof David T. S. Hayman, PhD[a]

a, Tāwharau Ora - School of Veterinary Science, Massey University, Palmerston North, New Zealand. **b,** School of Biological Sciences, University of Auckland, Auckland, New Zealand. **c,** School of Food Technology and Natural Sciences, Massey University, Palmerston North, New Zealand. **d,** New Zealand Food Safety Science and Research Centre, Massey University, Palmerston North, New Zealand. **e,** Microbes and Food Safety, Quadram Institute Bioscience, Norwich, UK. **f,** Centre for Microbial Interactions, Norwich, UK. **g,** Conservation Through Public Health, Entebbe, Uganda. **h,** Mubare Biodiversity Conservation, Kanungu, Uganda. **i,** Orion Integrated Biosciences, Manhattan, Kansas, USA. **j,** Health & Nutrition Cluster, Institute of Development Studies, University of Sussex, Brighton, East Sussex, UK. **k,** Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, UK. **l,** Department of Ecology & Evolutionary Biology, University of California, Los Angeles, California, USA. **m,** Department of Computational Medicine, University of California, Los Angeles, California, USA.

Abstract

Campylobacter species are leading causes of gastroenteritis globally, but their diversity and transmission dynamics at wildlife–livestock–human interfaces are poorly understood. We investigated *Campylobacter* diversity across sympatric mountain gorillas, livestock, and humans in Uganda. We analysed 553 faecal samples from humans (n=197), mountain gorillas (n=200), cattle (n=83), and goats (n=73). Human samples were collected from healthcare patients and community volunteers. We used metagenomic shotgun sequencing to reconstruct *Campylobacter* metagenome-assembled genomes (MAGs). We assembled 44 MAGs representing seven *Campylobacter* species, including three novel species from gorilla gut microbiomes. The others belonged to provisionally designated species *C.* sp017646085 (cattle) and *C.* sp900539255 (humans) and known *C. vicugnae* (goats) and *Candidatus* C. infans (humans). *C.* sp900539255 was significantly enriched in healthcare-based human samples (p=0·001), whereas *C. infans* was common and showed no abundance differences between healthcare and community-based human samples, suggesting limited clinical significance.

Introduction

Campylobacter species are among the leading causes of bacterial gastroenteritis and diarrhoea worldwide, with *C. jejuni* and *C. coli* being the most prevalent aetiological agents in human infections.[1,2] Several other *Campylobacter* species, including *C. fetus*,[3] *C. lari*,[4] *C. hyointestinalis*,[5] *C. concisus*,[6] *C. showae*,[7] *C. upsaliensis*,[8] *C. lanienae*,[9] and *C. vicugnae*,[10] have been implicated in human disease. The frequent occurrence of sporadic human infections is mainly linked to contaminated food, water, and the widespread presence of *Campylobacter* as a zoonotic pathogen in both livestock and wildlife populations.[2]

Diarrhoea continues to pose a significant public health challenge, contributing substantially to childhood morbidity and mortality, particularly in countries with limited resources.[11] Diarrhoeal disease mortality rates vary significantly across regions, with striking disparities between high-income areas and some countries in Africa.[11] In high-income regions, fewer than one child under five per 100 000 dies from diarrhoea, whereas in sub-Saharan Africa, the rate surpasses 130 deaths per 100 000.[11] In Uganda, diarrhoea is one of the top five causes of under-five mortality, responsible for over 140 000 deaths annually, accounting for 7.1% of all under-five deaths in the country.[12]

In Africa, studies across several countries indicate that *Campylobacter* species are highly prevalent, particularly among children.[1] However, the presence of *Campylobacter* is not always associated with diarrhoea symptoms, even when the most common pathogenic species, *C. jejuni* and *C. coli*, are detected.[1,13] For instance, a study in Madagascar reported high rates of *Campylobacter* in children without diarrhoea symptoms.[14] These findings suggest that members of the genus are endemic in some African populations and can colonise the gut without always causing symptomatic disease.[15] In 2020, a novel species was

identified in infants less than one year old;[15] *C. jejuni* and *C. coli* were prevalent among these infants, but the second most abundant *Campylobacter* was a new species, named *Candidatus* Campylobacter infans (*C. infans*).[15] Subsequent detection of *C. infans* in children under two years of age in Peru was not associated with diarrhoea symptoms.[16] However, a recent study in Ethiopia using species-specific quantitative PCR identified *C. infans* as the predominant colonisers of the infant gut, associated with increased odds of diarrhoea and fever.[17] The reconstruction of complete or near-complete genomes using high-throughput sequencing and advanced computational approaches (e.g., MAGs) provides a complementary framework to characterise the diversity of *Campylobacter* species and identify novel taxa, facilitating a more comprehensive understanding of their abundance in the gut microbiome and their roles in diarrhoeal disease.[18]

The majority of human infectious diseases in humans have animal origins. However, *Campylobacter* research involving non-human primates (NHPs), our closest relatives, remains limited. When investigated, *C. upsaliensis*, *C. jejuni*, and several novel *Campylobacter* species have been detected, some associated with disease in both captive[19] and wild NHPs.[20] An increase in *Campylobacter* detection in mountain gorillas in Uganda has been suggested to be potential anthropozoonotic transmission due to rising ecotourism in the region,[21] however, others speculate that the high prevalence of *Campylobacter* spp. in gorilla faecal samples primarily reflects a previously uncharacterised, non-pathogenic component of their intestinal flora.[20]

As human encroachment into habitats such as tropical forests increases and urban development accelerates, our understanding of how the associated environmental and socioeconomic changes influence pathogen transmission and emergence among wildlife, livestock, and human populations remains limited. Given that zoonotic diseases are an increasingly significant concern, particularly in regions where humans, wildlife, and domestic

animals share overlapping habitats and interact closely, our study aimed to investigate interspecies transmission of *Campylobacter* species by analysing their presence in faecal samples from habituated and unhabituated mountain gorillas, livestock, and humans living at a wildlife-livestock-human interface in Buhoma, Uganda. We employed metagenomic sequencing to quantify microbial abundance, reconstruct MAGs, and infer evolutionary relationships, offering a high-resolution view of *Campylobacter* species diversity in underrepresented ecological settings.

Methods

Ethical and regulatory approvals

This study received ethical approval from the Massey University Human Ethics Committee (Southern A Application-14-44), Massey University, Palmerston North, Aotearoa New Zealand and the Mbarara University of Science and Technology Research Ethics Committee (via Conservation Through Public Health [CTPH]), Mbarara, Uganda.

Study area and sample collection

Fieldwork was conducted in and around Bwindi Impenetrable National Park (BINP), Buhoma, southwestern Uganda, a region of high biodiversity and zoonotic disease emergence risk due to close human-livestock-wildlife contact. BINP is home to around 43% of the world's remaining approximately thousand mountain gorillas (*Gorilla beringei beringei*). Communities surrounding the park engage in subsistence farming and livestock rearing adjacent to forest margins.

Between June and August 2018, faecal samples from six host groups, habituated mountain gorillas, wild (non-habituated) gorillas, domestic cattle and goats, anonymous human patients from local health centres, and human community members near BINP were collected and

placed in RNAlater® (Sigma, MO, USA) and stored at -20°C. Samples were shipped to New Zealand and stored at the Hopkirk Research Institute, Massey University, until analysis.

Literature search

We searched PubMed from Jan 1, 2000, to Jan 31, 2026, without language restrictions, using combinations of terms related to *Campylobacter* diversity, wildlife–livestock–human interfaces, metagenomics, and clinical infection (appendix p 3).

DNA extraction and amplification

Approximately 150 mg of faeces was processed for nucleic acid extraction followed by DNA processing, quantification and library preparation (appendix p 3).

Genome sequencing

Two sequencing approaches were employed: individual and pooled samples.

Individual samples: A total of 553 DNA samples were sequenced at Orion Integrated Biosciences (KS, USA) using Nextera XT libraries on an Illumina® MiSeq platform. Sequencing data were used for read mapping to MAGs to estimate individual-level abundances and for taxonomic classification and estimation of *Campylobacter* abundances.

Pooled samples: Samples were pooled by host species/group (14–20 samples per pool) and sequenced at the Australian Genome Research Facility (AGRF, Melbourne, Australia) using the xGen™ cfDNA & FFPE DNA Library Preparation Kit (IDT, IA, USA) on an Illumina® NovaSeq platform (2 × 150 bp). The pooled datasets were used for MAG assembly, coverage estimation, taxonomic classification, phylogenetic analysis, and functional annotation.

Bioinformatics pipeline

A comprehensive bioinformatics pipeline was used for metagenomic data analysis; a full list of software, versions, scripts, and parameters is provided in the appendix (p 3). Additional publicly available genomes were used for comparative analyses.

Results

Study location, host populations, and sampling overview

We analysed 553 faecal samples collected in Bwindi Impenetrable National Park (BINP) and the adjacent Buhoma community, Uganda, from gorillas (n=200; 100 habituated and 100 non-habituated), livestock (n=156; 83 cattle and 73 goats), and humans (n=197). Human samples were collected either from healthcare facilities (clinical, n=99; age 5–69 years, median 25 years; collected anonymously) or the community (n=98; age 19–70 years, median 35 years; 81 male, 19 female). Seven community-based participants reported diarrhoea.

Diversity and phylogenetic structure of *Campylobacter* in host gut microbiomes

A total of 44 *Campylobacter* MAGs were reconstructed from metagenomic data (19 gorilla, 17 human, two cattle, six goat). Six MAGs were excluded because they did not meet quality parameters, resulting in 38 MAGs retained for downstream analyses. Species-level assignment identified *C. infans* and *C.* sp900539255 in humans, *C.* sp017646085 in cattle, and *C. vicugnae* in goats. Three gorilla-derived MAGs could not be assigned to any known *Campylobacter* species using GTDB-Tk, indicating putative novel taxa. Following dereplication, 23 unique MAGs were retained (13 human, four gorilla, five goat, one cattle).

Phylogenetic analysis incorporating study-derived and publicly available genomes revealed extensive unrecognised diversity across multiple clades (figure 1). Clade 1 comprised a previously undescribed human-associated *Campylobacter* (*C.* sp934718785) and an NHP-

associated species (*C.* sp006864365). Clade 2 included three novel species identified in this study, two gorilla-associated (sp1 and sp2) and one human-associated (*C.* sp900539255), together with two additional human-associated species from public datasets. Clade 3 comprised a novel gorilla-associated *Campylobacter* (sp3) and a human-derived species (*C.* sp900753405). Clade 4 consisted of *C. infans* genomes recovered from humans and gorillas. High average nucleotide identity (ANI) supported close genomic relatedness between *C. infans* strains from human and gorilla hosts; the human-derived MAG from this study shared 98·3% ANI with a human-derived reference genome and 98·0% ANI with a gorilla-derived genome (figure 2). Collectively, these findings indicate host-associated diversification and shared evolutionary trajectories of *Campylobacter* across humans and NHPs in multiple clades. Livestock-derived MAGs clustered within a distinct clade, with the goat-derived MAG grouping with *C. vicugnae* and the cattle-derived MAG grouping with *C.* sp017646085 (figure 1).

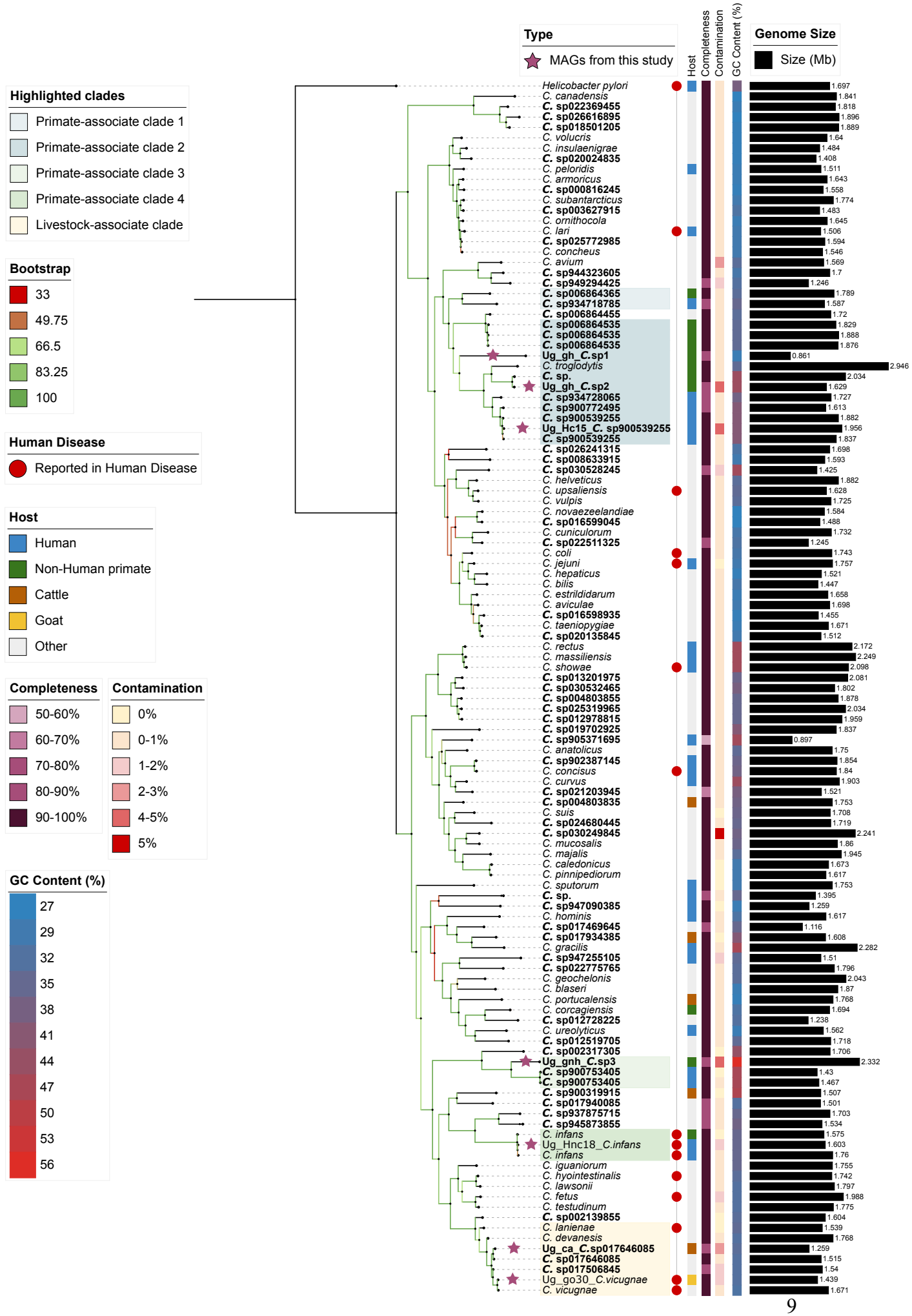

Highlighted clades
Primate-associate clade 1
Primate-associate clade 2
Primate-associate clade 3
Primate-associate clade 4
Livestock-associate clade
Bootstrap
33
49.75
66.5
83.25
100
Human Disease
Reported in Human Disease
Host
Human
Non-Human primate
Cattle
Goat
Other
Completeness
50-60%
60-70%
70-80%
80-90%
90-100%
Contamination
0%
0-1%
1-2%
2-3%
4-5%
5%
GC Content (%)
27
29
32
35
38
41
44
47
50
53
56
Type
MAGs from this study
Host
Completeness
Contamination
GC Content (%)
Genome Size
Size (Mb)
Helicobacter pylori
C. canadensis
C. sp022369455
C. sp026616895
C. sp018501205
C. volucris
C. insulaenigrae
C. sp020024835
C. peloridis
C. armoricus
C. sp000816245
C. subantarcticus
C. sp003627915
C. ornithocola
C. lari
C. sp025772985
C. concheus
C. avium
C. sp944323605
C. sp949294425
C. sp006864365
C. sp934718785
C. sp006864455
C. sp006864535
C. sp006864535
C. sp006864535
Ug_gh_C.sp1
C. troglodytis
C. sp.
Ug_gh_C.sp2
C. sp934728065
C. sp900772495
C. sp900539255
Ug_Hc15_C. sp900539255
C. sp900539255
C. sp026241315
C. sp008633915
C. sp030528245
C. helveticus
C. upsaliensis
C. vulpis
C. novaezeelandiae
C. sp016599045
C. cuniculorum
C. sp022511325
C. coli
C. jejuni
C. hepaticus
C. bilis
C. estrildidarum
C. aviculae
C. sp016598935
C. taeniopygiae
C. sp020135845
C. rectus
C. massiliensis
C. showae
C. sp013201975
C. sp030532465
C. sp004803855
C. sp025319965
C. sp012978815
C. sp019702925
C. sp905371695
C. anatolicus
C. sp902387145
C. concisus
C. curvus
C. sp021203945
C. sp004803835
C. suis
C. sp024680445
C. sp030249845
C. mucosalis
C. majalis
C. caledonicus
C. pinnipediorum
C. sputorum
C. sp.
C. sp947090385
C. hominis
C. sp017469645
C. sp017934385
C. gracilis
C. sp947255105
C. sp022775765
C. geochelonis
C. blaseri
C. portucalensis
C. corcagiensis
C. sp012728225
C. ureolyticus
C. sp012519705
C. sp002317305
Ug_gnh_C.sp3
C. sp900753405
C. sp900753405
C. sp900319915
C. sp017940085
C. sp937875715
C. sp945873855
C. infans
Ug_Hnc18_C.infans
C. infans
C. iguaniorum
C. hyointestinalis
C. lawsonii
C. fetus
C. testudinum
C. sp002139855
C. lanienae
C. devanesis
Ug_ca_C.sp017646085
C. sp017646085
C. sp017506845
Ug_go30_C.vicugnae
C. vicugnae
1.697
1.841
1.818
1.896
1.889
1.64
1.484
1.408
1.511
1.643
1.558
1.774
1.483
1.645
1.506
1.594
1.546
1.569
1.7
1.246
1.789
1.587
1.72
1.829
1.888
1.876
0.861
2.946
2.034
1.629
1.727
1.613
1.882
1.956
1.837
1.698
1.593
1.425
1.882
1.628
1.725
1.584
1.488
1.732
1.245
1.743
1.757
1.521
1.447
1.658
1.698
1.455
1.671
1.512
2.172
2.249
2.098
2.081
1.802
1.878
2.034
1.959
1.837
0.897
1.75
1.854
1.84
1.903
1.521
1.753
1.708
1.719
2.241
1.86
1.945
1.673
1.617
1.753
1.395
1.259
1.617
1.116
1.608
2.282
1.51
1.796
2.043
1.87
1.768
1.694
1.238
1.562
1.718
1.706
2.332
1.43
1.467
1.507
1.501
1.703
1.534
1.575
1.603
1.76
1.755
1.742
1.797
1.988
1.775
1.604
1.539
1.768
1.259
1.515
1.54
1.439
1.671

Figure 1 Phylogenetic relationships and general characteristics of study-derived and publicly available *Campylobacter* genomes.

A phylogenetic tree was inferred from a concatenated alignment of GTDB single-copy marker genes, with 1000 ultrafast bootstrap repetitions to assess branch support. Expected amino acid substitutions per site across the concatenated conserved marker alignment. The tree depicts the placement of *Campylobacter* MAGs from seven species assembled in this study (indicated by stars) and their relationships to publicly available *Campylobacter* genomes, including both described (italic) and undescribed (bold) species. Clades 2, 3, and 4 include duplicate public genomes to illustrate diversity within primate-associated *Campylobacter* species. The *C. lanienae*-like species clade, highlighted by a yellow background, contains the livestock-derived *Campylobacter* species identified in this study.

Campylobacter species abundance and classification sensitivity of metagenomic workflows for emerging taxa

To assess host-associated abundance patterns, metagenomic reads were mapped to dereplicated MAGs. Sequence coverage-based analyses revealed distinct host-group–associated patterns (appendix p 6). Gorilla-derived MAGs were detected almost exclusively in gorilla samples, with livestock-associated MAGs showing overlapping coverage across cattle and goat samples. In human samples, *C. infans* was more abundant in pooled community-based samples, while *C.* sp900539255 predominated in healthcare-based samples.

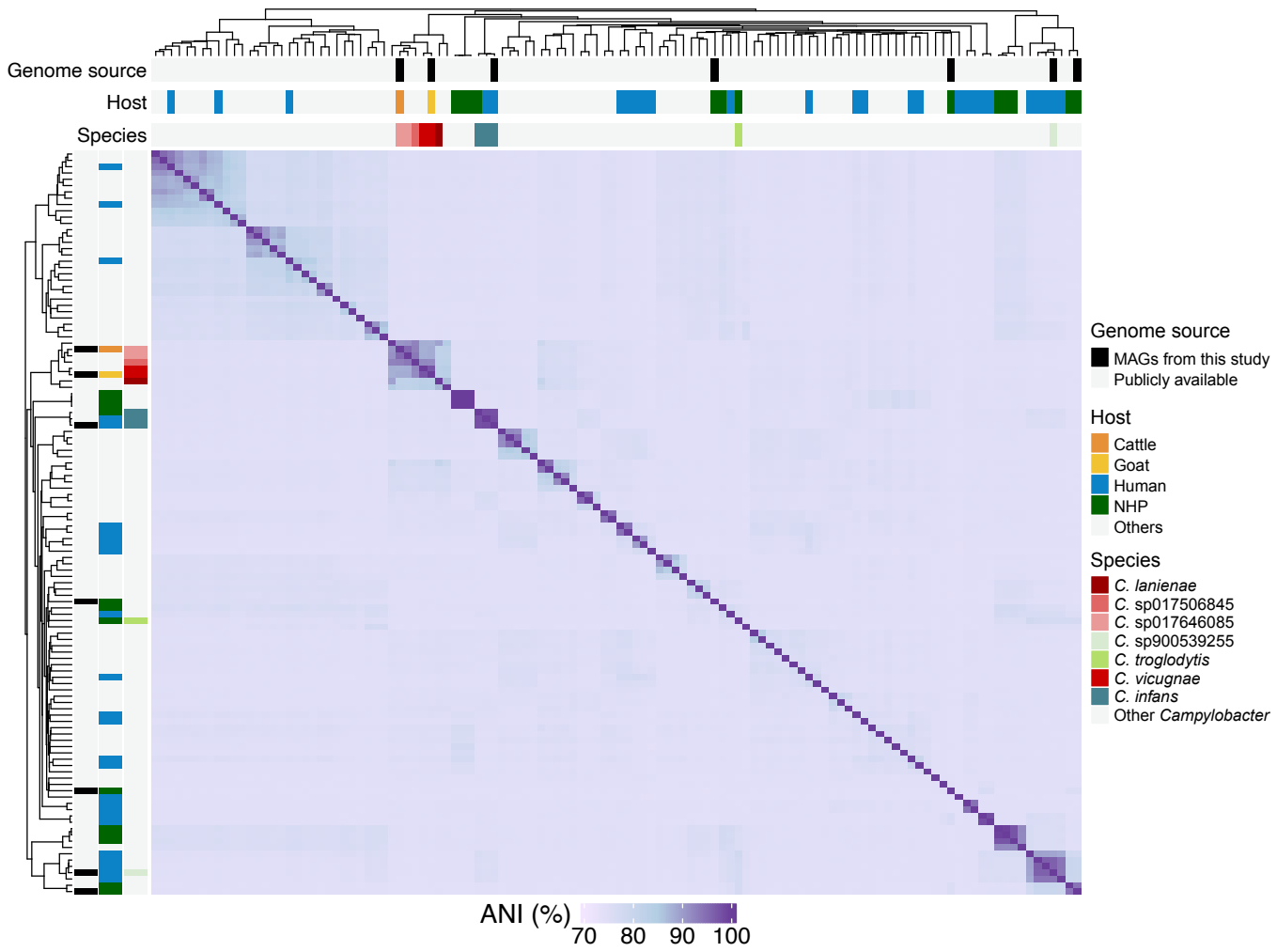


Figure 2 Average nucleotide identity and coverage heatmap of *Campylobacter* MAGs across host-associated samples.

C. infans genomes from human and gorilla hosts showed high ANI values, suggesting cross-species transmission or shared evolutionary history. Other primate-derived *Campylobacter* species showed more variable ANI values across hosts. The livestock-associated species also exhibited genomic similarity to *Campylobacter* taxa previously associated with human disease. High average nucleotide identity (ANI) supported close genomic relatedness between *C. infans* strains from human and gorilla hosts; the human-derived MAG from this study shared 98·3% ANI with a human-derived *C. infans* reference genome from the Netherlands (GTDB ID GCF_013416015.1) and 98·0% ANI with a gorilla-derived *C. infans* genome from Gabon (GTDB ID GCA_963591955.1).

Given the high computational and sequencing demands of MAG-based classification, to evaluate the performance of lower-resource taxonomic workflows, we compared MAG-based results with *k*-mer–based classification of unassembled reads. Among human samples, *k*-mer

283 classification detected *C. infans* in agreement with MAG-based results and additionally
284 assigned reads to *C. jejuni*, despite the absence of assembled *C. jejuni* MAGs (appendix pp 7-
285 9). Notably, *C.* sp900539255 was not detected by the *k*-mer approach. In livestock samples,
286 *k*-mer-based assignments were partially concordant with MAG-based taxonomy. While *C.*
287 *vicugnae* assignments in goats were consistent, cattle-derived reads were assigned to multiple
288 species not detected by MAG-based analyses, including *C. devanensis*, *C. porcelli*, and *C.*
289 *lanienae* (figure 3A). Phylogenetic analyses indicated that these taxa clustered within the
290 same livestock-associated clade as *C.* sp017646085 (figure 1), suggesting shared genomic
291 features rather than true species-level presence. Genomic characteristics of *C. devanensis*
292 were more strongly represented in *C.* sp017646085 from cattle-derived samples, whereas *C.*
293 *vicugnae* showed stronger representation in goat-derived samples (Figure 3B). In gorilla
294 samples, all MAG-based identified taxa represented novel species, and *k*-mer classification
295 failed to resolve distinct taxonomic profiles, instead assigning a proportion of reads to *C.*
296 *jejuni* (appendix pp 7-9).
297

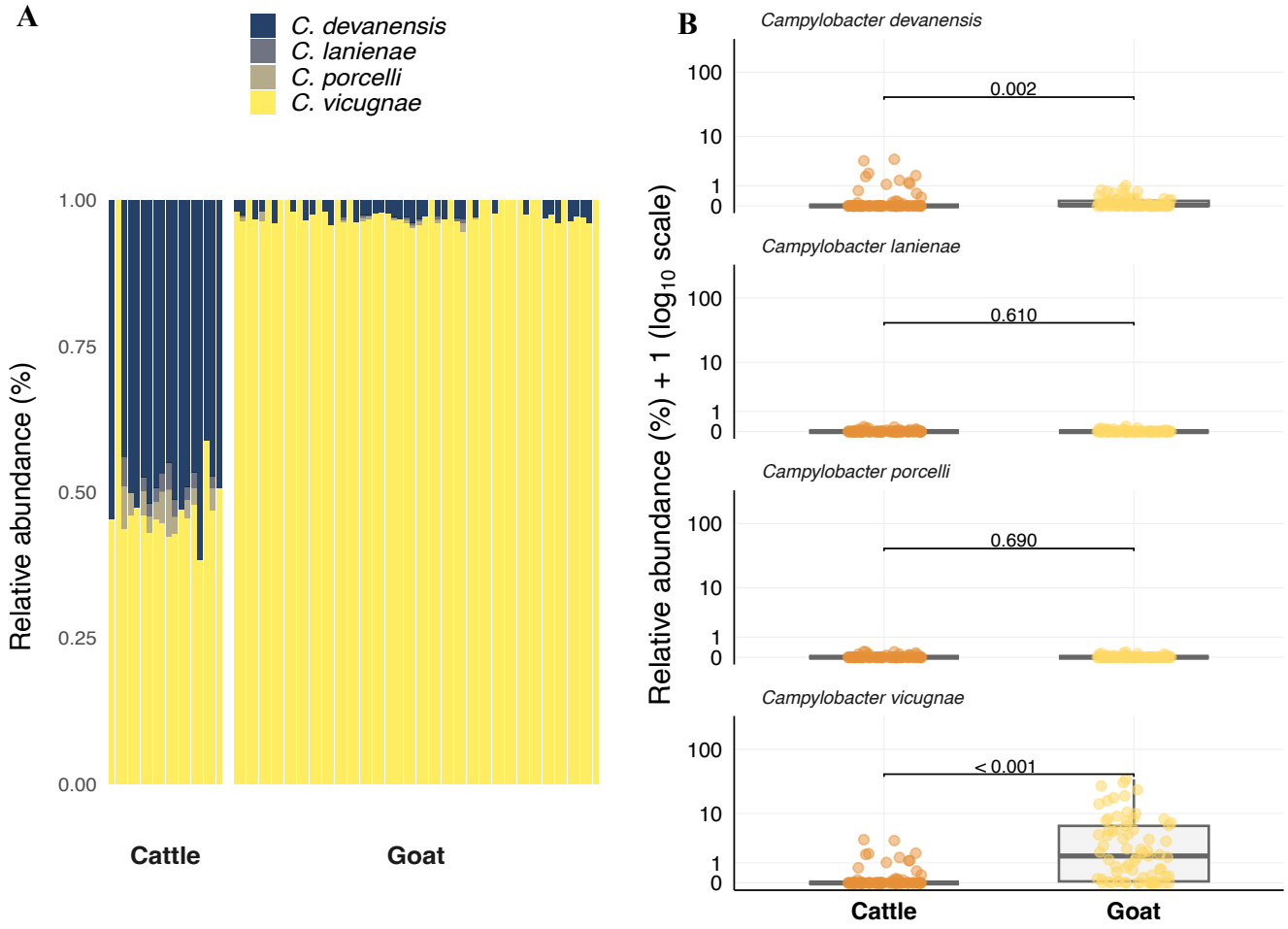


Figure 3 Comparative *k*-mer-based analysis of *Campylobacter* species prevalence and abundance in livestock gut microbiomes.

A) The dominant *Campylobacter* community in livestock samples differed between cattle and goats. Comparison of taxonomic assignments using a *k*-mer-based approach revealed partial discrepancies relative to MAG-based classification. Of the four species detected by the *k*-mer-based approach, only *C. vicugnae* was recovered through MAG assembly. Conversely, the *k*-mer-based approach failed to detect *C.* sp017646085, which was identified from cattle-derived MAGs. B) Comparison of the abundance of key livestock-associated *Campylobacter* taxa based on *k*-mer-based classification. *C. lanienae* and *C. porcelli*, which were not recovered via the assembly-based strategy, showed similar distribution patterns across both host groups (p=0·61 and p=0·69, respectively). In contrast, *C. vicugnae* was significantly more prevalent and abundant in goats (Fisher's exact test, p<0·001; Wilcoxon W=2295·5), consistent with findings from the MAG-based taxonomic classification. Conversely, *C. devanensis*—which was not identified using the MAG-based approach—was more frequently detected and abundant in cattle (Fisher's exact test, p<0·002; Wilcoxon W=1018·5). These findings indicate that, although reference-dependent *k*-mer-based tools can detect broad *Campylobacter* species, they may lack the resolution and database coverage needed to distinguish them from the novel cattle-derived

taxon *C.* sp017646085, which was identified through genome assembly from samples collected approximately six years before the most recent database update used in this study.

To evaluate whether reads initially assigned to *C. jejuni* by *k*-mer-based classification could instead be explained by the locally assembled *C. infans* and *C.* sp900539255 MAGs, we applied a sequential two-step mapping strategy (appendix pp 10-14). Host-filtered reads were first mapped to a combined *C. infans* and *C.* sp900539255 MAG index and only reads that failed to align to either MAG were subsequently remapped to the *C. jejuni* reference genome. This procedure reduced the total number of reads mapped to *C. jejuni* from 123,837 to 10,170 (92% reduction). To confirm that this reduction reflects genuine differences in sequence identity rather than an artefact of mapping order, we performed a competitive mapping analysis in which all host-filtered reads were simultaneously aligned to a combined index containing *C. infans*, *C.* sp900539255, and *C. jejuni* sequences in a single mapping pass, with each read assigned to the reference achieving the highest alignment score (appendix pp 10-14). Under this approach, only 7,606 reads were assigned to *C. jejuni*, a 94% reduction from the original 123,837, and a further 25% reduction compared to the sequential strategy. Alignment score (AS) analysis confirmed that in high-abundance samples (>500 reads assigned to *C. infans* and *C.* sp900539255 under competitive mapping), reads aligned with substantially better mean AS to the locally assembled MAGs (range −7 to −46) than to *C. jejuni* (range −50 to −106), with AS closer to zero indicating better alignment quality ($p<0{\cdot}001$).

To further assess potential *C. jejuni* presence, the 10,170 reads assigned to *C. jejuni* under two-step mapping strategy were assembled. The resulting contigs showed a mean sequence identity of 96% to *C. jejuni*, and analysis of the 23S rRNA gene (3,875 bp) demonstrated 95·11% nucleotide identity to the closest reference sequence (*Campylobacter jejuni*), corresponding to approximately 190 nucleotide differences, with similarly low identities

(94·14–94·99%) observed across multiple other *Campylobacter* species. These values fall well below accepted species-level thresholds, indicating that this sequence represents a distinct *Campylobacter* species rather than *C. jejuni*. Together, these findings demonstrate that reliance on *k*-mer-based taxonomic tools substantially overestimates the prevalence of well-characterised species such as *C. jejuni* while obscuring novel diversity, particularly in underexplored settings where reference databases remain incomplete.

Potential clinical relevance of *Campylobacter* species in human gut microbiomes

To assess associations between *Campylobacter* species detected in human gut microbiomes and sampling setting (healthcare or community-based), reads from individual human metagenomes were mapped using two approaches: mapping to the highest-quality MAG per species and mapping to all unique MAGs for each species. Using both approaches, *C. infans* abundance did not differ significantly between healthcare-based (mean 0·281) and community-based samples (mean 0·417) ($p=0·052$; figure 4A; appendix p 15). One individual reporting severe diarrhoea exhibited among the highest abundances of both *C. infans* and *C.* sp900539255. In contrast, significantly more reads from healthcare-based samples aligned to *C.* sp900539255 than from community-based samples, both when mapping was performed against the highest-quality MAG ($p<0·001$) (figure 4B) and when reads were mapped to all unique MAGs ($p<0·007$) (appendix p 15). Given the observed association of *C.* sp900539255 with healthcare-seeking individuals, we next assessed whether commonly used species-specific diagnostic primers could detect this novel species. We examined the compatibility of *cpn60* gene primers[17] across publicly available *Campylobacter* reference genomes and human-derived *Campylobacter* MAGs. This analysis allowed us to determine whether novel species might be misclassified or missed by standard detection methods. All reference genomes matched their respective species-specific primer sets exclusively, and a study-derived *C. infans* MAG also matched its species-specific primers,

whereas *C.* sp900539255 MAGs did not closely match any available species-specific primer pairs. These results suggest that C. sp900539255 is likely to go undetected in standard diagnostic assays. To assess the prevalence and abundance of *C. infans* across different age groups, 192 human-derived samples with available age data were analysed; 73 showed detectable relative abundances of the species (figure 4 C,D). Logistic regression indicated no significant association between age and its presence (p=0·324). However, age was a significant predictor of its relative abundance (β=0·071, SE=0·016, t=4·38, p<0·001). Although outliers (shown in red) had some influence, the association remained significant after their removal (p<0·01; figure 4D).

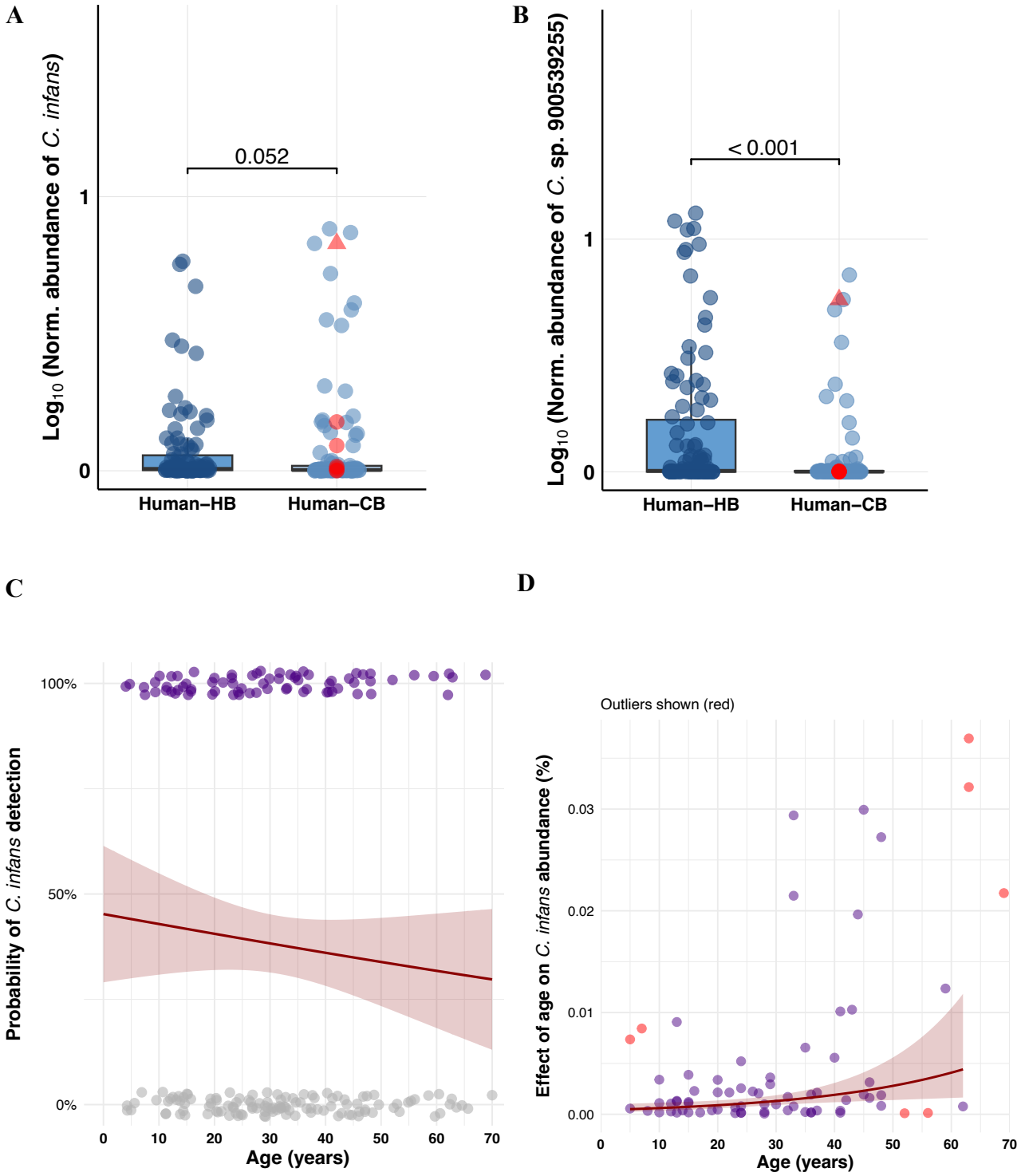


Figure 4 Abundance of *Campylobacter* species across human healthcare and community-based samples and association of *C. infans* with age.

Campylobacter abundance was compared between human healthcare and community-based gut microbiomes. Normalised abundance values were calculated for each sample and $\log_{10}(x+1)$-transformed for visualisation abundance derived from reads mapped to the highest-quality MAG. Community-based samples from individuals reporting diarrhoea are highlighted in red. An individual reporting severe diarrhoea is indicated by a triangle. A) *C. infans* abundance did not differ between the two human groups. B) By contrast, *C.* sp900539255 showed

higher abundance in samples from healthcare-seeking individuals. C) *C. infans* was detected across a wide age range, with the highest abundances observed in individuals aged >30 years. Although prevalence declined with age, no significant association was observed between age and presence. D) Among positive samples, age significantly predicted relative abundance, which increased with age. This association remained significant after exclusion of outliers (red dots; p<0·01).

Genomic divergence of human-associated *Campylobacter* species

Comparative functional analysis, incorporating both study-generated and publicly available genomes, identified differences in metabolic pathway representation between *C. infans* and *C.* sp900539255. Of 178 annotated pathways, 45 were shared, three were unique to *C. infans*, and seven were unique to *C.* sp900539255. Seventeen pathways differed significantly between species (appendix pp 16-17), including sulphur and nitrogen metabolism pathways enriched in *C.* sp900539255 (p<0·001). Differences were also observed in genes encoding glycoside hydrolase family 73 (GH73), which were detected in 16 of 18 *C.* sp900539255 genomes but in none of the *C. infans* genomes (p<0·001). By contrast, pathways involved in polyphenolic compound degradation were identified in 11 of 21 *C. infans* genomes and were absent from *C.* sp900539255 (p<0·001; appendix p 18).

Prevalence of antimicrobial resistance and virulence genes

To characterise the prevalence of antimicrobial resistance and virulence determinants, we screened all genomes analysed in this study, including publicly available genomes, for known resistance and virulence-associated genes. The β-lactam resistance genes $bla_{\text{OXA-471_1}}$, $bla_{\text{OXA-193_1}}$ and $bla_{\text{OXA-489_1}}$ were detected in 11 of 21 *C. infans* genomes (appendix p 19). Additionally, *tet(W)*$_{_5}$, which confers tetracycline resistance, was present in a single publicly available *C. infans* genome. No other antimicrobial resistance genes were detected in the genomes of the other species analysed in this study. The virulence gene *cheY* was detected in

most *C. infans* and *C. vicugnae* genomes, whereas *fliN* was detected in a single *C.* sp900539255 genome (appendix p 20).

Discussion

Campylobacter species are a leading cause of human bacterial gastroenteritis worldwide.[1,2] In Uganda, diarrhoeal disease remains a major contributor to under-five mortality.[12] Although *Campylobacter* species are endemic across Africa, particularly among children, detection of *C. jejuni* and *C. coli* does not consistently correlate with symptoms, highlighting the complexity of asymptomatic carriage and disease.[1,13,14,22] In this study, overall *Campylobacter* prevalence and abundance did not differ between healthcare-based and community-based human samples. However, MAG-based and further genomic analysis revealed that *C.* sp900539255 was significantly enriched in healthcare-seeking individuals, whereas *C. infans* was prevalent in both groups regardless of healthcare-seeking status. Although human campylobacteriosis is traditionally attributed to livestock reservoirs, particularly in rural settings in Africa,[2,23] phylogenetic analyses indicated that taxa detected at this wildlife–livestock–human interface were more closely related to NHP–associated species than to livestock-derived *Campylobacter*. These findings suggest that human–wildlife microbial exchange may contribute to transmission dynamics of novel and emerging *Campylobacter* species, extending beyond conventional agricultural pathways.

Interpretation of *Campylobacter* ecology was sensitive to the analytical approach used. Comparison of MAG-based profiling with reference-dependent *k*-mer classification showed that *k*-mer-based methods did not detect novel taxa recovered through genome assembly, instead assigning a proportion of the reads to well-characterised species represented in reference databases. Such discrepancies are expected in underrepresented settings, where classifier performance is influenced by reference database completeness and the presence of

novel taxa.[24] This methodological limitation is important because MAG-based workflows require high sequencing depth and substantial computational resources, which may be inaccessible in settings such as those studied here.[25] In this study, most reads initially mapped to a *C. jejuni* reference were better explained by mapping to locally assembled *C. infans* and *C.* sp900539255 MAGs, indicating that *C. jejuni* is unlikely to be present at appreciable abundance in this dataset. Misassignment of reads to *C. jejuni* across host species illustrates how incomplete reference databases can lead to overestimation of well-characterised taxa while obscuring local species diversity, particularly in underrepresented settings. Together, these findings highlight the value of MAG-based approaches[18] for resolving species-level diversity and support the need for improved reference databases and computationally accessible genomic surveillance in underrepresented regions.

Building on these methodological considerations, genome-resolved profiling of human gut metagenomes showed contrasting connections between *Campylobacter* taxa and healthcare-seeking behaviour. The enrichment of *C.* sp900539255 in healthcare-based samples suggests a closer association with symptomatic infection or increased disease severity, although this could not be formally evaluated due to limited clinical metadata. One individual with severe diarrhoea showed high abundances of both *C.* sp900539255 and *C. infans*, indicating possible co-occurrence in pronounced gastrointestinal disturbance. By contrast, the clinical relevance of *C. infans* remains uncertain.[16,26] In this study, *C. infans* was detected across both healthcare-based and community-based samples, with abundance increasing with age, suggesting age-related differences in colonisation dynamics rather than infection risk. These patterns also highlight limitations of reference-dependent detection, as *C.* sp900539255 was not detected by *k*-mer classification and showed poor compatibility with existing *cpn60* primer sets, suggesting that *C.* sp900539255 has been underdetected in studies relying on targeted PCR or reference-based classification. Together, these findings underscore the need

for species-level resolution to accurately assess clinical relevance and transmission dynamics of *Campylobacter* in underrepresented populations.

Although clinical associations of *C.* sp900539255 with symptoms or disease in healthcare-seeking individuals could not be directly assessed, comparative genomic analyses provided initial insights into functional divergence between *C.* sp900539255 and *C. infans*. While their metabolic repertoires overlapped substantially, *C.* sp900539255 uniquely encoded sulphur- and nitrogen-associated pathways, including thiosulphate–polysulphide reductase and nitrous oxide reduction, whereas *C. infans* harboured pathways involved in polyphenolic compound degradation.[27] *C.* sp900539255 also showed enrichment of GH73 enzymes, which are involved in glycan processing and include peptidoglycan hydrolases that have been associated with host cell invasion and virulence.[28,29] These differences suggest divergent ecological strategies, with *C.* sp900539255 displaying metabolic features more consistent with host-associated stress adaptation observed in pathogenic bacteria.[30–33] Together, these features suggest ecological strategies that may favour persistence of *C.* sp900539255 in host-associated niches of potential clinical relevance, whereas *C. infans* appears more consistent with broader gut adaptation. However, pathogenicity is multifactorial, and future studies integrating clinical metadata and longitudinal sampling will be required to clarify the clinical significance of these traits.

Antimicrobial resistance and virulence-associated genes were detected across multiple *Campylobacter* species. Notably, over half of the *C. infans* genomes carried β-lactam resistance genes. This finding suggests a potential contribution to the gut resistome, with implications for horizontal gene transfer to pathogenic bacteria.[34] The widespread presence of the chemotaxis regulator *cheY* in *C. infans* and *C. vicugnae* further indicates genetic determinants associated with host colonisation and motility, traits common among pathogenic *Campylobacter* species.[35] Together, these findings indicate that the functional repertoire of

Campylobacter species extends beyond established pathogens such as *C. jejuni* and *C. coli*, challenging simple pathogen–commensal distinctions and highlighting the importance of comprehensive genomic surveillance to accurately assess the public health relevance of diverse *Campylobacter* taxa.

Collectively, these findings highlight that the high diversity of *Campylobacter* at the wildlife–livestock–human interface in this underrepresented region of Africa includes novel taxa with potential clinical relevance. This observation reinforces the need for cautious interpretation of diagnostic and metagenomic data generated using reference-dependent, *k*mer-based approaches, which may underestimate local species diversity. Our results support the expansion of reference genome databases and the wider adoption of genome-resolved approaches to improve detection, functional characterisation, and assessment of virulence and antimicrobial resistance in *Campylobacter* and other potentially enteric bacteria, thereby strengthening diagnostic accuracy and genomic surveillance capacity in African settings.

Author contributions

VA, DTSH, JOL-S, and HM conceptualised and designed the study. VA, DTSH, MAK and KH developed the methodology. DTSH and KH supervised the analyses. VA, KH, and MAK curated the data. VA performed the bioinformatic analysis under the supervision of KH. VA and DTSH conducted the investigation. DTSH obtained funding, while NF and AEM secured additional funding for VA's doctoral visit. GZ, SR, AN, and WV contributed resources and samples. VA and DTSH wrote the first draft of the manuscript. All authors contributed to the intellectual content and to the review and editing of the final draft. VA, KH, ACM, PJB, AHD, NF, AEM, JOL-S and DTSH contributed to the interpretation of the data, and critically reviewed and revised the article. VA, DTSH, KH, and MAK accessed and verified the underlying data. VA and DTSH had final responsibility for the decision to submit for

publication. All authors had full access to all the data in the study and approved the final version of the manuscript.

Competing interests

None.

Data sharing

The sequencing data generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) under the BioProject accession code PRJNA1213876. All custom code and scripts used for the bioinformatic analyses are publicly available on GitHub in the repository titled *Campylobacter*-at-a-wildlife-livestock-human-interface-in-Uganda (GitHub).

Acknowledgements

The study was supported by the Royal Society Te Aparangi (Royal Society of New Zealand) (contracts: RDF-MAU1701; MAU1503) (DTSH), the Percival Carmine Chair in Epidemiology and Public Health (DTSH/VA), Massey University Research Fund, and a McGeorge Research Grant (DTSH). VA's doctoral work was funded by the Percival Carmine Chair and the Massey University Tāwharau Ora - School of Veterinary Science (SoVS) doctoral fund. A doctoral visit for VA was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) (grant number BB/V018221/1) in partnership with NF and AEM. AEM is supported by the BBSRC through the BBSRC Institute Strategic Programme Microbes and Food Safety BB/X011011/1 and its constituent project BBS/E/QU/230002C (Theme 3, Flexible capabilities to reduce food safety threats and respond to national needs); AHD was supported by BBSRC Norwich Research Park Biosciences Doctoral Training Partnership (Grant number BB/T008717/1) as a Joint Funded

Studentship with the Food Standards Agency (FSA); the work was also supported in part by BBSRC grant BB/V018221/1. Genomic courses were provided by Genomics Aotearoa, and New Zealand eScience Infrastructure (NeSI) / Research and Education Advanced Network New Zealand (REANNZ) / Massey University provided supercomputing resources.

References

1. Kaakoush NO, Castaño-Rodríguez N, Mitchell HM, Man SM. Global epidemiology of *Campylobacter* infection. *Clin Microbiol Rev.* 2015; **28:** 687–720.
2. Burnham PM, Hendrixson DR. *Campylobacter jejuni*: collective components promoting a successful enteric lifestyle. *Nat Rev Microbiol*. 2018; **16:** 551–65.
3. Wagenaar JA, Van Bergen MAP, Blaser MJ, Tauxe RV, Newell DG, Van Putten JPM. *Campylobacter fetus* infections in humans: exposure and disease. *Clin Infect Dis*. 2014; **58:** 1579–86.
4. Gourmelon M, Boukerb AM, Nabi N, Banerji S, Joensen KG, Serghine J, et al. Genomic diversity of *Campylobacter lari* group isolates from Europe and Australia in a One Health context. *Appl Environ Microbiol*. 2022; **88:** e0136822.
5. Edmonds P, Patton CM, Griffin PM, Barrett TJ, Schmid GP, Baker CN, et al. *Campylobacter hyointestinalis* associated with human gastrointestinal disease in the United States. *J Clin Microbiol*. 1987; **25:** 685–91.
6. Kaakoush NO, Mitchell HM. *Campylobacter concisus* - A new player in intestinal disease. *Front Cell Infect Microbiol*. 2012; **2:** 4.
7. Hsu T, Gemmell MR, Franzosa EA, Berry S, Mukhopadhya I, Hansen R, et al. Comparative genomics and genome biology of *Campylobacter showae*. *Emerg Microbes Infect*. 2019; **8:** 827–40.
8. Bourke B, Chan VL, Sherman P. *Campylobacter upsaliensis*: waiting in the wings. *Clin Microbiol Rev*. 1998; **11:** 440–49.

9. Lévesque S, Lemay F, Bekal S, Frost EH, Michaud S. First reported case of *Campylobacter lanienae* enteritis in a human. *JMM Case Rep*. 2016; **3:** e005045.
10. Jehanne Q, Bénéjat L, Azzi Martin L, Korolik V, Ducournau A, Aptel J, et al. First isolation of *Campylobacter vicugnae* sp. nov. in humans suffering from gastroenteritis. *Microbiol Spectr*. 2024; **12:** e0152324.
11. Kyu HH, Vongpradith A, Dominguez RMV, Ma J, Albertson SB, Novotney A, et al. Global, regional, and national age-sex-specific burden of diarrhoeal diseases, their risk factors, and aetiologies, 1990–2021, for 204 countries and territories: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet Infect Dis*. 2025; **25:** 519–36.
12. Nantege R, Kajoba D, Ddamulira C, Ndoboli F, Ndungutse D. Prevalence and factors associated with diarrheal diseases among children below five years in selected slum settlements in Entebbe municipality, Wakiso district, Uganda. *BMC Pediatr*. 2022; **22:** 394.
13. 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:** 1074–80.
14. Randremanana RV, Randrianirina F, Sabatier P, Rakotonirina HC, Randriamanantena A, Razanajatovo IM, et al. *Campylobacter* infection in a cohort of rural children in Moramanga, Madagascar. *BMC Infect Dis*. 2014; **14:** 372.
15. 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:** e00735-19.
16. Garcia Bardales PF, Schiaffino F, Huynh S, Olortegui MP, Yori PP, Vasquez TP, et al. "*Candidatus* Campylobacter infans" detection is not associated with diarrhea in children under the age of 2 in Peru. *PLoS Negl Trop Dis*. 2022; **16:** e0010869.

17. Ojeda A, Deblais L, Mummed B, Brhane M, Hassen KA, Ahmedo BU, et al. Determinants of *Campylobacter* species diversity in infants and association with family members, livestock, and household environments in rural Eastern Ethiopia. *Gut Pathog*. 2025; **17:** 51.

18. Kim N, Ma J, Kim W, Kim J, Belenky P, Lee I. Genome-resolved metagenomics: a game changer for microbiome medicine. *Exp Mol Med*. 2024; **56:** 1501–12.

19. Clayton JB, Danzeisen JL, Johnson TJ, Trent AM, Hayer SS, Murphy T, et al. Characterization of *Campylobacter jejuni*, *Campylobacter upsaliensis*, and a novel *Campylobacter* sp. in a captive non-human primate zoological collection. *J Med Primatol*. 2019; **48:** 114–22.

20. Whittier CA, Cranfield MR, Stoskopf MK. Real-time PCR detection of *Campylobacter* spp. in free-ranging mountain gorillas (*Gorilla beringei beringei*). *J Wildl Dis*. 2010; **46:** 791–802.

21. Nizeyi JB, Innocent RB, Erume J, Kalema GRNN, Cranfield MR, Graczyk TK. Campylobacteriosis, salmonellosis, and shigellosis in free-ranging human-habituated mountain gorillas of Uganda. *J Wildl Dis*. 2001; **37:** 239–44.

22. 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. *PLoS One*. 2013; **8:** e59663.

23. Singh N, Thystrup CAN, Hassen BM, Bhandari M, Rajashekara G, Hald TM, et al. Transmission pathways of *Campylobacter jejuni* between humans and livestock in rural Ethiopia are highly complex and interdependent. *Gut Pathog*. 2025; **17:** 26.

24. McIntyre ABR, Ounit R, Afshinnekoo E, Prill RJ, Hénaff E, Alexander N, et al. Comprehensive benchmarking and ensemble approaches for metagenomic classifiers. *Genome Biol*. 2017; **18:** 182.

25. Lynn HM, Gordon JI. Sequential co-assembly reduces computational resources and errors in metagenome-assembled genomes. *Cell Rep Methods*. 2025; **5:** 101005.

26. Flipse J, Duim B, Wallinga JA, de Wijkerslooth LRH, van der Graaf-Van Bloois L, 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:** 1003.
27. Balzerani F, Hinojosa-Nogueira D, Cendoya X, Blasco T, Pérez-Burillo S, Apaolaza I, et al. Prediction of degradation pathways of phenolic compounds in the human gut microbiota through enzyme promiscuity methods. *NPJ Syst Biol Appl*. 2022; **8:** 24.
28. Ronholm J, Wang L, Hayashi I, Sugai M, Zhang Z, Cao X, et al. The *Listeria monocytogenes* serotype 4b autolysin IspC has N-acetylglucosaminidase activity. *Glycobiology*. 2012; 22(10).
29. Lipski A, Hervé M, Lombard V, Nurizzo D, Mengin-Lecreulx D, Bourne Y, et al. Structural and biochemical characterization of the β-N-acetylglucosaminidase from *Thermotoga maritima*: toward rationalization of mechanistic knowledge in the GH73 family. *Glycobiology*. 2014; **25:** 319–30.
30. Rahman MA, Glasgow JN, Nadeem S, Reddy VP, Sevalkar RR, Lancaster JR, et al. The role of host-generated H_2S in microbial pathogenesis: new perspectives on tuberculosis. *Front Cell Infect Microbiol*. 2020; **10:** 586923.
31. Lensmire JM, Hammer ND. Nutrient sulfur acquisition strategies employed by bacterial pathogens. *Curr Opin Microbiol*. 2019; **47:** 52–58.
32. Kies PJ, Hammer ND. A resourceful race: Bacterial scavenging of host sulfur metabolism during colonization. *Infect Immun*. 2022; **90:** e0057921.
33. Rodríguez-Daza MC, Pulido-Mateos EC, Lupien-Meilleur J, Guyonnet D, Desjardins Y, Roy D. Polyphenol-mediated gut microbiota modulation: Toward prebiotics and further. *Front Nutr*. 2021; **8:** 689456.
34. Forsberg KJ, Reyes A, Wang B, Selleck EM, Sommer MOA, Dantas G. The shared antibiotic resistome of soil bacteria and human pathogens. *Science*. 2012; **337:** 1107–11.

678 35. Kreling V, Falcone FH, Kehrenberg C, Hensel A. *Campylobacter* sp.: Pathogenicity
679 factors and prevention methods—new molecular targets for innovative antivirulence
680 drugs? *Appl Microbiol Biotechnol*. 2020; **104:** 10409–36.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AlmeidaetalNatureCommsSupplementary.pdf
- OneDrive20260506.zip