

## Infant diarrheal disease in rhesus macaques impedes microbiome maturation and is linked to uncultured *Campylobacter* species

Nicholas S. Rhoades<sup>1,2</sup>, Isaac R. Cinco<sup>2</sup>, Sara M. Hendrickson<sup>3</sup>, Kamm Prongay<sup>4</sup>, Andrew J. Haertel <sup>4</sup>, Gilberto E. Flores <sup>5</sup>, Mark K. Slifka <sup>3</sup> & Ilhem Messaoudi <sup>2</sup>✉

Diarrheal diseases remain one of the leading causes of death for children under 5 globally, disproportionately impacting those living in low- and middle-income countries (LMIC). *Campylobacter spp.*, a zoonotic pathogen, is one of the leading causes of food-borne infection in humans. Yet to be cultured *Campylobacter spp.* contribute to the total burden in diarrheal disease in children living in LMIC thus hampering interventions. We performed microbiome profiling and metagenomic genome assembly on samples collected from over 100 infant rhesus macaques longitudinally and during cases of clinical diarrhea within the first year of life. Acute diarrhea was associated with long-lasting taxonomic and functional shifts of the infant gut microbiome indicative of microbiome immaturity. We constructed 36 *Campylobacter* metagenomic assembled genomes (MAGs), many of which fell within 4 yet to be cultured species. Finally, we compared the uncultured *Campylobacter* MAGs assembled from infant macaques with publicly available human metagenomes to show that these uncultured species are also found in human fecal samples from LMIC. These data highlight the importance of unculturable *Campylobacter spp.* as an important target for reducing disease burden in LMIC children.

<sup>1</sup>Department of Molecular biology and Biochemistry, University of California Irvine, Irvine, CA, USA. <sup>2</sup>Department of Microbiology, Immunology and Molecular Genetics, College of Medicine, University of Kentucky, Lexington, KY, USA. <sup>3</sup>Division of Neuroscience, Oregon National Primate Research Center, Portland, OR, USA. <sup>4</sup>Division of Animal Resources and Research Support, Oregon National Primate Research Center, Oregon Health and Science University West Campus, Portland, OR, USA. <sup>5</sup>Department of Biology, California State University, Northridge, Northridge, CA, USA. ✉email: [ilhem.messaoudi@uky.edu](mailto:ilhem.messaoudi@uky.edu)

Diarrheal diseases are the second leading cause of death in children under five years of age with over 500,000 fatalities every year that disproportionately impact low- and middle-income countries (LMIC)<sup>1,2</sup>. A wide range of enteric pathogens (*Rotavirus*, *Norovirus*, *Campylobacter*, *Shigella*, *Giardia*, etc.) are responsible for diarrheal diseases in children. However, these known pathogens only account for 40–50% of cases in LMIC leaving a substantial number of cases that cannot be attributed to a specific pathogen<sup>3,4</sup>. It also remains unclear why children who grow up in the same environment have differing susceptibility to enteric pathogens. One potential mechanism is the diversity of the gut microbiome which is highly dynamic in early life. It is well-established that the gut microbiome modulates host susceptibility to enteric pathogens through production of anti-microbial products, competition for niche space, and education of the immune system<sup>5–8</sup>. However, the role of specific components of the commensal gut microbiome in regulating susceptibility to and severity of diarrheal disease remains poorly understood due to lack of a clear definition of a “healthy” microbiome, a lack of longitudinal studies that include pre-symptom samples, and difficulty in collecting samples in resource-limited settings where diarrheal diseases are prevalent. A better understanding of the role of the gut microbial community would pave the way to the development of better therapeutic interventions and vaccination.

We have recently described an infant rhesus macaque model of environmentally acquired diarrheal disease and enteric dysfunction that is primarily driven by *Campylobacteriosis*<sup>9–11</sup>. Our macaque model accurately recapitulates many of the key features of diarrheal diseases in resource limited settings including stunted developmental trajectory of the gut microbiome<sup>9</sup>, markers of local and systemic immune dysfunction<sup>10</sup>, growth faltering, and histopathological changes in the small and large intestine<sup>11</sup>. In this study, we utilize microbiome profiling and metagenomic genome assembly on longitudinal samples collected from over 100 infant macaques across 5 pre-determined time-points during their first year of life. Moreover, over the course of the study, 34 animals developed diarrhea, and in those instances, additional samples were collected at the time of entry and exit from the clinic and subjected to the same analyses. As observed in humans<sup>12–14</sup>, diarrhea was associated with delayed maturation of the infant gut microbiome, specifically by the reduction of diversity in the gut microbial community which persisted after resolution of symptoms. Interestingly, infants that later developed diarrhea harbored a distinct microbiome from that of infants that remained asymptomatic as early as at 1 month of age. One of the distinguishing features was higher relative abundance of *Campylobacter* at the genus level that was evident before disease onset.

*Campylobacter* is one of the four leading causes of diarrheal diseases, and the primary cause of foodborne infections in humans globally, with the majority of infections being caused by the species *Campylobacter jejuni* (*C. jejuni*) and *Campylobacter coli* (*C. coli*)<sup>15</sup>. *Campylobacteriosis* can lead to significant morbidity and mortality in children under five<sup>1,2,16</sup> especially those living in LMIC<sup>17–20</sup>. More importantly, children in LMIC are more likely to experience subclinical *Campylobacter* colonization, which has been associated with growth stunting, aberrant inflammation, and dysbiosis<sup>21,22</sup>. Despite its global impact on public health, our understanding of the complex etiology and true diversity within the *Campylobacter* genus remains incomplete. Over the past decade more than a dozen named species have been added to the genus and associated with multiple clinical manifestations including ulcerative colitis, periodontitis, diarrhea, and cancer<sup>23–25</sup>. Moreover, recent studies using culture-independent techniques such as qPCR and metagenomic genome assembly have identified additional non-*coli/jejuni* *Campylobacter* species

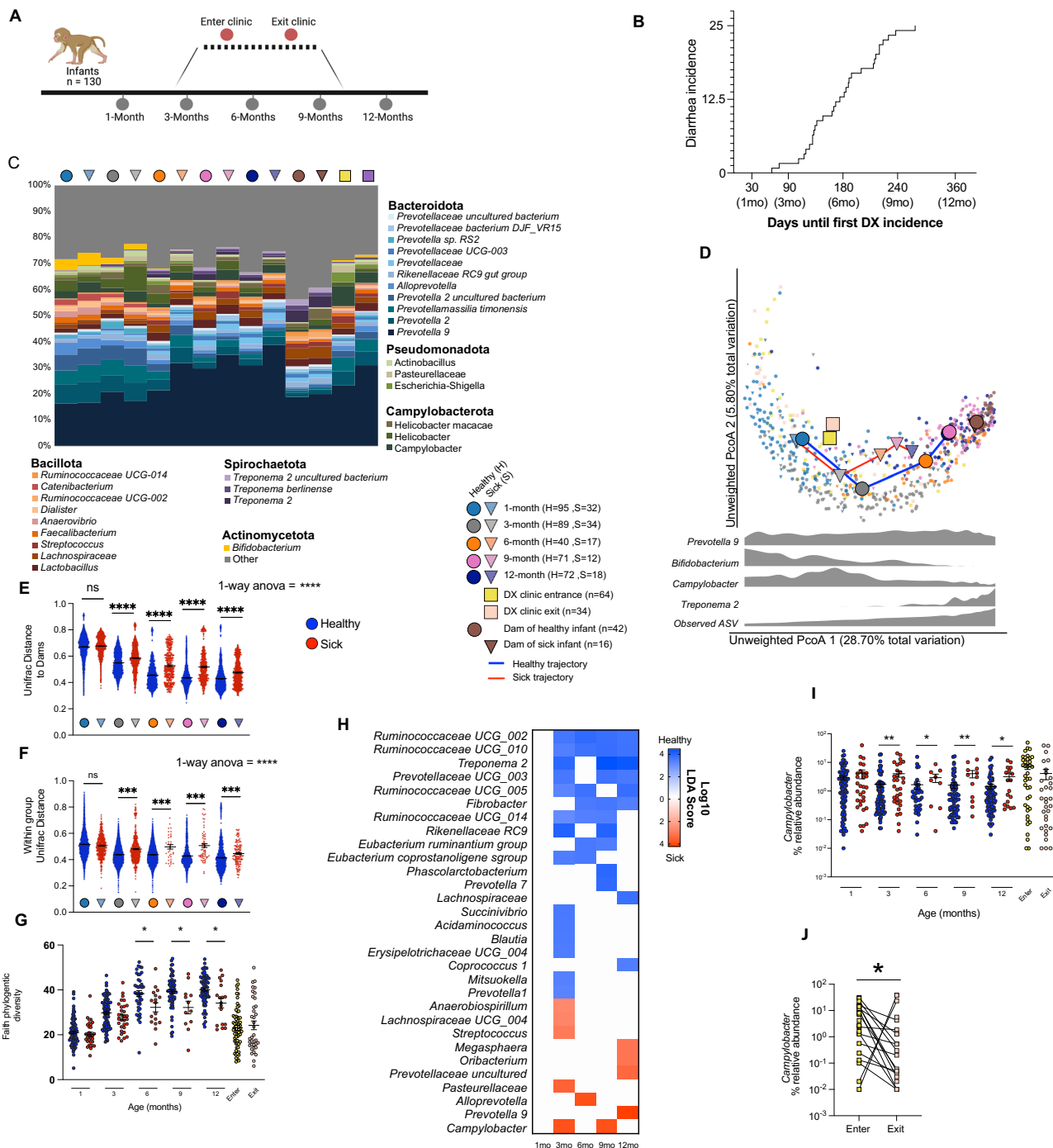
that had been missed by traditional culture assays, notably *Candidatus Campylobacter infans* (*C. infans*) which was initially assembled from a child experiencing diarrhea in India and has since been identified in samples from 6 additional LMIC countries<sup>26–28</sup>. In this study, we constructed 36 *Campylobacter* metagenomic assembled genomes (MAGs), 15 of which belong to uncultured species that could be classified into 4 uncultured metagenomic groups (UMGs). These uncultured *Campylobacter* species had unique metabolic and pathogenicity profiles, and the majority were previously reported in human samples. These data will aid in the efforts aimed at the isolation and cultivation of undescribed *Campylobacter* and provides genomic insights into their impact on human health.

## Results

**Diarrhea disrupts the taxonomic development of the infant macaque gut microbiome.** 16S rRNA gene amplicon sequencing was used to profile the fecal microbiome of 130 infant rhesus macaques, 34 of which developed clinical diarrhea during the first year of life (26.4%). Rectal swabs were collected at 1, 3, 6, 9 and 12 months of age (Fig. 1A). Additional swabs were collected any time an infant was admitted to the clinic for diarrhea and when the infant was released after treatment (Fig. 1A). The majority of infants in this study experienced their first case of clinical diarrhea between 3 and 9 months of age while only two of 34 infants experienced clinical diarrhea prior to 3-months (Fig. 1B, Table S1). We first examined the maturation of gut microbial communities of healthy infants that did not experience an episode of diarrhea. While *Prevotella* was highly abundant at all the timepoints, loss of milk degrading bacteria such as *Bifidobacterium* and concomitant increase in fiber degraders like *Treponema* was observed over the 12 months of life (Fig. 1C, D, Table S2). Additionally, the microbiome of infants that remained healthy progressively became closer to that of adult dams and the inter-group variability was decreased over the 12-month period (Fig. 1D–F).

However, disruptions in this developmental trajectory were observed in infants that developed diarrhea (none of the samples included in this analysis were obtained during acute disease). Specifically, a stunting in the trajectory was evident at 3-months of age, prior to most infants experiencing their first episode of diarrhea (Fig. 1D). These changes included increased dissimilarity from the dam gut microbiome (Fig. 1E) and increased within group dissimilarity (Fig. 1F) compared to healthy infants. This delay in maturation became more pronounced with increasing age (Fig. 1D, E). Moreover, the microbiome of sick infants was less phylogenetically diverse at the 6- to 12-month timepoints compared to that of infants that remained asymptomatic (Fig. 1G). Diversity of the microbial community was further reduced at the time of acute diarrhea as expected with an enteric infection (Fig. 1G). Although the diversity of the microbial community was significantly increased after treatment and exit from the clinic, it remained below that observed in healthy infants (Fig. 1G).

Taxonomically, we found no significant difference at the 16s amplicon level at the 1-month time point between animals that remained healthy and those that developed diarrhea (Fig. 1H). However, several significant differences were identified at the 3-month time point (when only 2 of 34 animals experienced their first episode) and beyond (Fig. 1H, Table S3). These changes included a sustained increase in known short chain fatty acid producers and fiber degraders in healthy infants, notably *Ruminococcaceae*, *Fibrobacter* and *Treponema* in infants that remained healthy (Fig. 1H, Table S3). On the other hand, the gut microbiomes of infants that developed diarrhea were enriched in



**Fig. 1 Diarrhea stunts the development of the infant microbiome.** **A** Study timeline and sample collection. The dotted line between 3 and 9 months denotes the period in which most of the infants entered and exited the clinic for treatment and sample collection during their first case of diarrhea. **B** Kaplan Meier curve of diarrhea incidence within the study population indicating the first incidence of clinical diarrhea for every animal that experienced symptoms over the course of the study (first 12 month of life). **C** Stacked bar plot of the top 30 most abundant taxa across all samples, each bar represents the average relative abundance of each taxon within the indicated group. **D** Principal coordinate analysis (PCoA) of fecal microbiome unweighted UniFrac distance (16S amplicon) colored by host status and timepoint. Small points represent individual samples while larger points represent the centroid for a given timepoint/host status. Solid lines connecting centroids illustrate the developmental trajectory of both healthy and sick infants. Density plots of key taxa and observed ASVs are in grey below the plot. **E-G** Dot plot of **(E)** UniFrac distance between dams and each infant timepoint/health status, **F** within group UniFrac distance for each infant group, and **(G)** Faith's phylogenetic diversity, in healthy and sick infants. **H** Heatmap of select differentially abundant taxa at longitudinal time-points as determined by LEfSe (Log<sub>10</sub> LDA score > 2). **I** Longitudinal dot plot of *Campylobacter* relative abundance, in healthy and sick infants. **J** Relative abundance of *Campylobacter* at first clinic entry and exit. Significance for panels **(E)** and **(F)** was determined using one-way ANOVA \*\*\*\**p* < 0.0001, with Holm-Sidak's multiple comparison test, \*\*\*\**p* < 0.0001, error bars = SEM. Significance for panels **(G)** and **(I)** was determined using unpaired *T*-test at each time-point, \**p* < 0.05, \*\**p* < 0.01, \*\*\**p* < 0.001, error bars = SEM. Significance for panels **(J)** was determined using paired *T*-test, \**p* < 0.05. DX Diarrhea.

*Prevotella*, *Streptococcus*, and *Campylobacter* species (Fig. 1H, Table S3). Indeed, a significantly higher relative abundance of *Campylobacter* was observed at timepoints 3- to 12-months in infants that developed diarrhea (Fig. 1I). The relative abundance of *Campylobacter* was significantly higher when sick infants entered the clinic with acute diarrhea compared to their exit (Fig. 1J). Although, the relative abundance of *Campylobacter* decreased at discharge, it remained higher than that observed in healthy infants (Fig. 1J). Of note, *Campylobacter* relative abundance assessed by 16S rRNA amplicon sequencing correlated significantly with *Campylobacter* colony forming units as measured by qPCR (Fig. S1A). Together these data suggest that diarrhea stunts the development of the infant gut microbiome resulting in a more heterogeneous but less diverse community with increased *Campylobacter* burden.

We also profiled the gut mycobiome and its maturation over the first year of life and how it is impacted by diarrheal disease using internal transcribed sequence (ITS) sequencing using a subset of samples (Fig. S1B). As established for the prokaryotic microbiome, the gut mycobiome became more diverse over time (Fig. S1B–D). While we observed no differences between the two groups at the 1-month timepoint, animals with a history of diarrhea exhibited a less diverse mycobiome than age-matched healthy infants during acute disease and at 12-months of age (Fig. S1D). Taxonomically, the macaque mycobiome is poorly defined, leading to a large portion of sequences being classified as “unidentified fungi” (Fig. S1B). Nevertheless, differentially abundant taxa (*Rhizophyidum* and *Metschnikowia continentalis*) were identified during acute disease and at 12-months of age (Fig. S1E).

A majority of the infants in this study were vaccinated against *C. coli* as previously described<sup>29</sup>. While *C. coli* vaccination reduced overall diarrheal disease burden, it did not impact the composition of the fecal microbiome<sup>29</sup>. However, since *Campylobacter spp.* were associated with diarrheal disease in our study population, we conducted additional analyses to confirm that the trends observed were not associated with vaccination. Microbiome stunting, reduced community richness, and increased *Campylobacter* abundance were still observed when samples were subsetted based on vaccination status (Fig. S2A–J).

**Role of maternal microbiome in infant susceptibility to diarrheal diseases.** The role of the maternal microbiome in modulating infant susceptibility to diarrheal diseases remains poorly defined. Therefore, we compared the microbiomes of dams that gave birth to infants that later developed diarrhea to those that gave birth to infants that remained asymptomatic at ~1 month after delivery. At the 16S rRNA amplicon level, minor differences were noted in the overall composition of the gut microbiome between these two groups (Fig. S3A, B). However, higher resolution taxonomic data generated using shotgun metagenomics indicated that the microbiome of dams who gave birth to infants that later experienced diarrhea harbored increased abundance of both *Escherichia coli* (*E. coli*), *Bifidobacterium pseudocatenulatum* (*B. pseudocatenulatum*), and *C. jejuni* (Fig. S3C, D). Additionally, the gut microbiome of dams whose infants later developed diarrhea was enriched in nitrate reduction and hexitol fermentation pathways while that of dams whose infants remained healthy was enriched in homolactic fermentation (Fig. S3E, F).

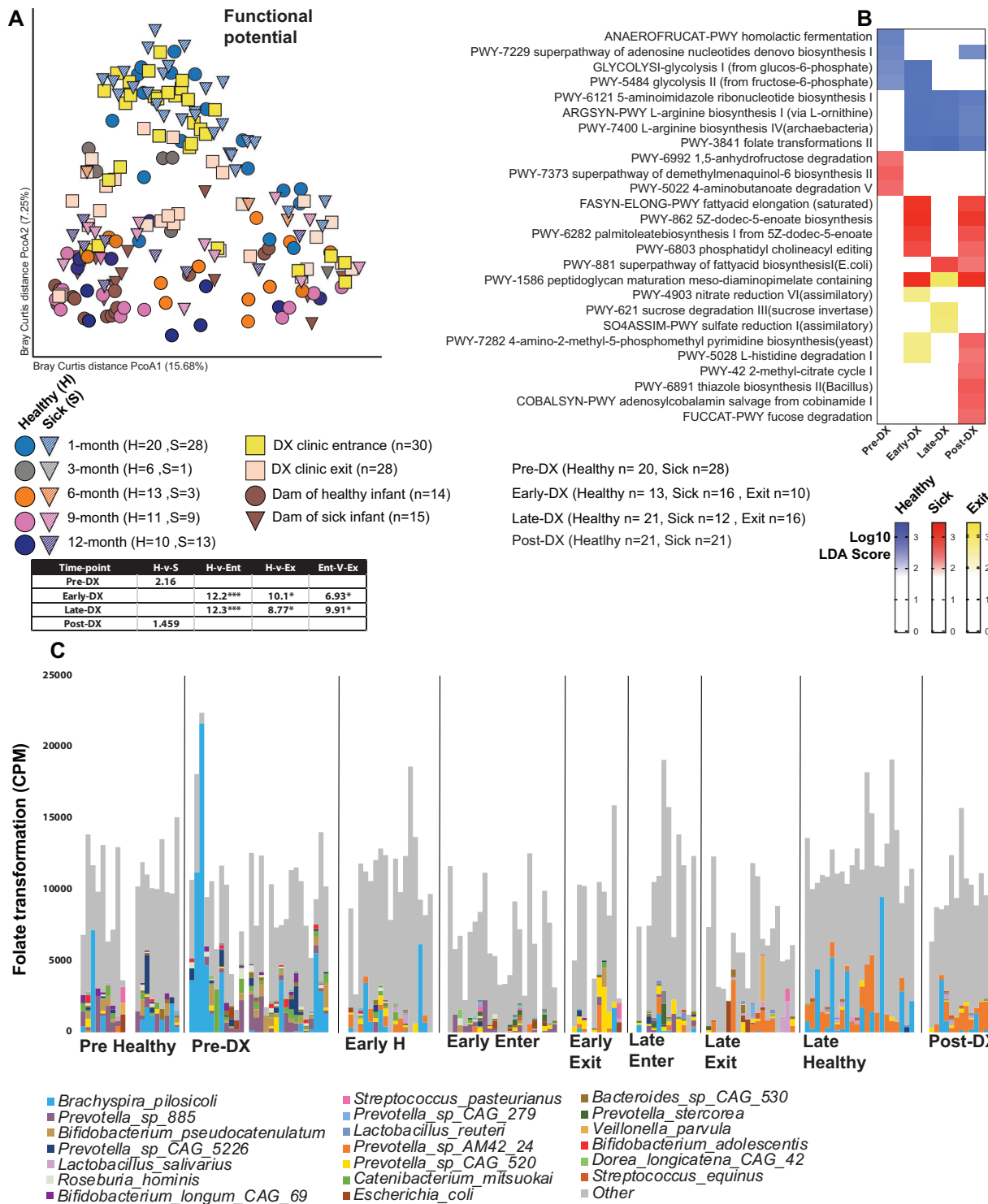
**Diarrheal disease disrupts the functional maturation of the infant macaque gut microbiome.** We next used shotgun metagenomics to define the development of the infant gut microbiome at the functional and taxonomic levels and determine how diarrhea disrupts these processes. As seen with 16S amplicon data,

age was a strong predictor of gut microbiome functional potential and taxonomic composition (Fig. S4A, B). Specifically, at early time-points the microbiome was enriched in pathways for the degradation of sugars commonly found in breastmilk including fucose degradation, Bifidobacterium shunt (1-month) and 1,5-anhydro-fructose degradation (3-months) (Fig. S4C). At later time points anabolic pathways such as L-isoleucine biosynthesis (6-months), L-tryptophan biosynthesis and folate transformation (9-months) as well as pathways associated with the production of short-chain fatty acids such as pyruvate fermentation to isobutanol (9-months) were increased (Fig. S4C).

Next, we explored potential predictors of diarrheal disease and assessed the impact of diarrhea on the functional potential of the gut microbiome (Fig. 2A). Although overall functional capacity did not differ between infants that remained asymptomatic and those that developed diarrhea at 1 month of age (prior to any diarrhea incident; pre-DX) (Fig. 2A), several metabolic pathways were differentially abundant between these two groups at this early time point. Specifically, glycolysis, and homolactic fermentation pathways were more abundant in infants that remained asymptomatic while anhydro-fructose and amino-butanate degradation pathways were enriched in infants that later developed diarrhea (Fig. 2B, Table S4). To assess the impact of acute diarrhea and ensuing treatment, we grouped animals into early (3–6 months; Early-DX) and late onset (6.1–11 months; Late-DX) of disease. At both time frames, the microbiomes of healthy infants were enriched in L-arginine biosynthesis and folate transformation (Fig. 2B, Table S4). Acute diarrhea at 3–6 months of age led to increased abundance of fatty acid elongation, palmitoleate biosynthesis, and phosphatidylcholine editing pathways (Fig. 2B, Table S4). After clinic exit, the microbiomes of these infants were enriched in nitrate reduction and histidine degradation pathways (Fig. 2B, Table S4). The microbiomes of animals that developed diarrhea between 6 and 11 months of age were enriched in fatty acid elongation during acute disease and sulfate reduction and sucrose degradation upon exiting the clinic (Fig. 2B, Table S4).

At 9 or 12 months of age (at least 1 month after the last diarrhea episode; Post-DX) the microbiomes of animals that experienced diarrhea were enriched in multiple pathways including thiazole biosynthesis, methyl-citrate cycle, palmitoleate biosynthesis, and fucose degradation (Fig. 2B). Since the pathway for folate transformation was enriched in healthy infants across multiple time-points, we attempted to identify the species contributing to this pathway. This analysis revealed that *Bracyspira pilosicoli* was a major contributor to folate transformation in healthy infants early in life and replaced by multiple *Prevotella* species namely AM42\_24 and CAG\_520 later in life (Fig. 2C).

**Disruption in the taxonomic maturation of the gut microbiome due to diarrheal disease.** Taxonomically, we observed a loss of all *Bifidobacterium* species and concomitant increase of *Treponema* at ~6-months of age which coincided with when most infant macaques are weaned (Fig. S4B, D). Additionally, we found that *Prevotella* species were dynamic at the species level, with *Prevotella copri* (*P. copri*) only becoming dominant at ~6-month of age, while uncultured *Prevotella sp\_885* was only observed in young infants and *Prevotella CAG\_5226* was only abundant in the dams (Fig. S4D). At 1 month of age (pre), although overall taxonomic composition did not differ between animals that remained asymptomatic and those that developed diarrhea (Fig. 3A), multiple species were differentially abundant between the two groups (Fig. 3B). Of note, relative abundance of *Ruminococcus CAG\_624* was increased in infants that remained asymptomatic while relative abundance of *Anaerostipes hadrus*

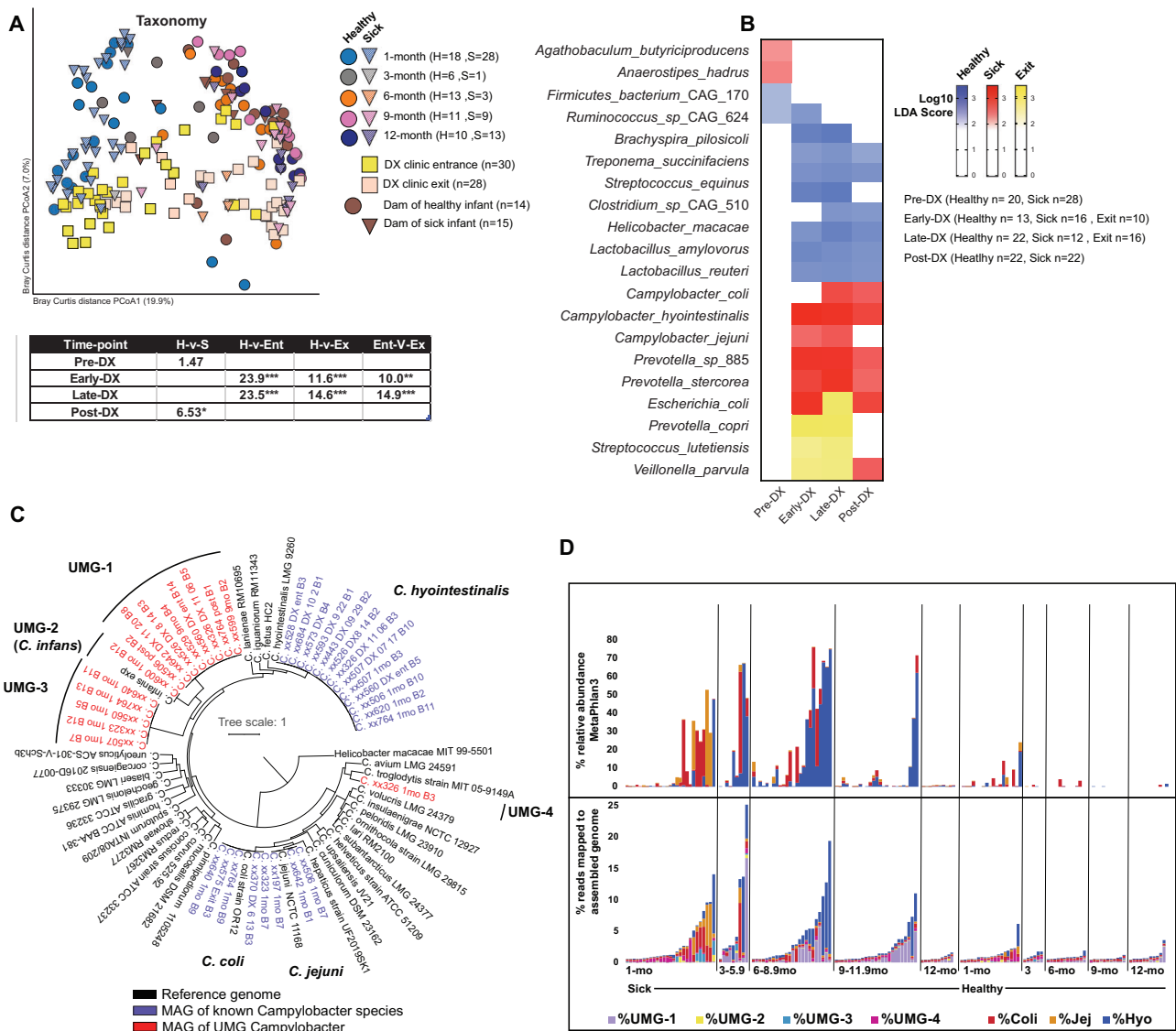


**Fig. 2 Impact of acute and distant diarrhea on the functional potential of the gut microbiome. A** PCoA of Bray-Curtis dissimilarity built on the abundance of all functional genes annotated using HUMAnN3 and the Uniref90 database colored by host status and timepoint. Significance and variance by host status for each time-point determined by PERMANOVA and reported in the included table. **B** Heatmap of select differentially abundant functional pathway at each time-point as determined by LEfSe (Log<sub>10</sub> LDA score > 2). **C** Bar-plot of folate transformation pathway abundance, stratified by bacterial species to which that pathway could be assigned. Each stacked bar represents an individual sample at the indicated time-points. DX Diarrhea.

was increased in infants that later developed diarrhea (Fig. 3B, Table S5).

We then assessed the impact of acute diarrhea and ensuing treatment in infants that developed diarrhea 3–6 months (early) and 6.1–11 months (late) of age. In both timeframes, the microbiomes of infants that remained asymptomatic were

enriched in *Lactobacillus*, *Treponema*, *Bracyspira* species as well as *Helicobacter macacae* (Fig. 3B, Table S5). In contrast, acute diarrhea was associated with increased abundance of multiple *Campylobacter* and *Prevotella* species (Fig. 3B, Table S5). After exit from clinic, the relative abundance of *P. copri*, *Streptococcus lutetuensis*, and *Veillonella parvula* was increased (Fig. 3B,



**Fig. 3** Impact of diarrhea on the species level composition of the gut microbiome. **A** PCoA of Bray-Curtis dissimilarity built on the species level taxonomy generated using MetaPhlan and colored by host status and timepoint. **B** Heatmap of select differentially abundant species pathway at each time-point as determined by LefSe (Log<sub>10</sub> LDA score > 2). **C** Core gene phylogenetic tree built using 60 randomly selected conserved cross-genus gene families (PGfams) for metagenomically assembled *Campylobacter* genomes (MAGs) and 32 known *Campylobacter* species. Black = Reference genome, Blue = MAG similar to known species, Red = MAG not similar to any known species. **D** Stacked bars of *Campylobacter* relative abundance as measured by Metaphlan3 (top) and mapping to MAGs and reference genomes (bottom). Each vertical bar represents a single sample at the denoted timepoint/health status. For “Sick” animals samples collected upon exit from the clinic were not included. DX Diarrhea.

Table S5). At 12-months of age (at least 1-month after the last exit from clinic), many differences that were observed during acute disease persisted, most notably the increased abundance of *Campylobacter* persisted (Fig. 3B, Table S5).

**Metagenomic genome assembly reveals a diverse set of *Campylobacter* species associated with diarrhea.** Given the increased relative abundance of *Campylobacter* in infants who experienced diarrhea and the recent reports of the critical role that unrecognized *Campylobacter* species play in childhood diarrheal disease and growth/cognitive stunting<sup>27,28,30</sup>, we utilized metagenomic genome assembly to explore the diversity of enteric *Campylobacter* species within infant macaques. We assembled 36 high quality (>80% complete with <2% contamination) *Campylobacter* metagenomic assembled genomes (MAGs). Despite the inclusion of 58 samples from healthy infants, all *Campylobacter* MAGs

originated from samples collected from infants that developed diarrhea. A phylogenetic tree based on the alignment of 50 conserved single copy genes showed that most MAG belonged to known species including 13 *Campylobacter hyointestinalis* (*C. hyointestinalis*) genomes, 5 *C. coli* genomes, and 3 *C. jejuni* genomes (Fig. 3C, Tables S6, S7). The remaining 15 *Campylobacter* MAGs were divided into 4 distinct groups (UMGs) with 8 genomes falling into “UMG-1”, 1 genome in “UMG-2”, 5 genomes in “UMG-3”, and 1 in “UMG-4” (Fig. 3C, Table S6). UMG 2 was closely related to the recently identified MAG *C. infans* that was assembled from a child with acute severe diarrhea in the global enterics multi-center study (GEMS)<sup>27</sup>. UMGs 1 and 3 were closely related to each other and to *C. hyointestinalis* (Fig. 3C, Table S6). UMG-4 was distinct and most closely related to *Campylobacter troglodytis* (*C. troglodytis*) which was isolated from chimpanzees<sup>31</sup>.

Next, we determined if these uncultured species could be detected using a traditional shotgun metagenomics taxonomic

classifier MetaPhlan3<sup>32</sup>. This method was able to assign a relative abundance to the 3 known *Campylobacter* species (*C. hyointestinalis*, *C. coli*, and *C. jejuni*) but was unable to quantify the 4 *Campylobacter* UMG (Fig. 3D). We then quantified reads that map to the assembled *Campylobacter* MAGs including the 4 UMGs using BowTie2 (Fig. 3D). This analysis reveals that the 4 *Campylobacter* UMGs were prevalent in sick infant macaques (Fig. 3D). These analyses revealed that infant macaques are colonized by a diverse set of both known and unknown *Campylobacter* species that are more abundant in infants that will or have already experienced an episode of diarrhea.

**Genomic diversity of “Unknown” *Campylobacter* species.** To explore the genomic diversity of the *Campylobacter* MAGs identified in this study relative to published *Campylobacter* MAG, we performed a pangenomic analysis using Anvi'o<sup>33</sup>. We included genomes from 18 reference *Campylobacter* species, 11 recently described MAGs and genomes from rhesus macaques<sup>9,10,34</sup>, and 11 human MAGs<sup>35,36</sup> including *C. infans*<sup>27</sup>. All human MAGs fell within UMG-2, -3 and -4 based on average nucleotide identity and several were closely related to MAGs assembled from infants that developed diarrhea in this study, highlighting the relevance of the rhesus macaque model for investigating these uncultured *Campylobacter* species (Fig. 4A–C, Table S6).

Moreover, these human MAGs were identified from samples obtained from LMIC where infant diarrhea is more prevalent<sup>35,36</sup> (Fig. 4D). We identified a total of 18,076 gene clusters (GCs), with 7099 GCs being found in only a single genome (Singletons), 3078 GCs being found in 2–3 genomes (Accessory), and 69 GCs that were shared across all 77 genomes (Fig. 4B). We also identified 408 GCs that were present in at least 90% of all genomes. This expanded core included GCs essential for amino acid transport (*hisM*, *hisC*, *abcC*) and metabolism (*pykF*, *pgk*, *gltA*) (Fig. 4B).

**Motility and adhesion factors in yet to be cultured *Campylobacter* MAG.** Next, we investigated the presence of known virulence factors in the *Campylobacter* UMG identified in this study. Using Anvi'o we assigned Clusters of Orthologous Groups (COG) annotations to GC's and conducted a functional enrichment analysis. *Campylobacter* are highly motile due to a single polar flagellum at both poles of the cell body which allows them to burrow through host mucosa and access the epithelial surface where they attach and invade. Therefore, we first examined GC's that played a role in motility, chemotaxis, and adhesion to host cells. Most genes associated with flagellar assembly (all *fli*, *fli*, *mot* genes) were highly conserved across all genomes except for UMG-3 (Fig. S5A). Additionally, MAGs within UMG-3 lacked genes associated with chemotaxis that were highly conserved across all other species (Fig. S5A). Next, we examined additional key host cell adhesion and invasion genes (*pldA*, *ciaB*, *cadF*). While *pldA*, which is important for colonization of the avian cecum<sup>37</sup>, was only found in UMG-3 and -4, *cadF* and *ciaB*, which bind host epithelial fibronectin and facilitate internalization, were found in UMG-1 and -2. Finally, UMG-3 only harbored *ciaB* (Fig. S5A).

At the amino acid level, the *cadF* and *ciaB* sequences from rhesus MAGs of *C. jejuni*, *C. coli*, and *C. hyointestinalis* were nearly identical to their respective human reference genomes (Fig. S5B, C). On the other hand, the *cadF* sequences of UMG-1 and -2 were phylogenetically distinct from any other known *Campylobacter* species, with one rhesus macaque MAG *cadF* sequence nearly identical to that of *C. infans* and another human MAG (Fig. S5B). A similar pattern was observed for *ciaB* in UMG-2 (Fig. S5C). However, the *ciaB* amino acid sequences of

UMG-3 from humans and rhesus macaques were distinct (Fig. S5C). These data show that the *Campylobacter* UMG, with the exception of UMG-3, have the genetic capacity for motility and chemotaxis as well as a varied ability to attach and invade the host epithelia.

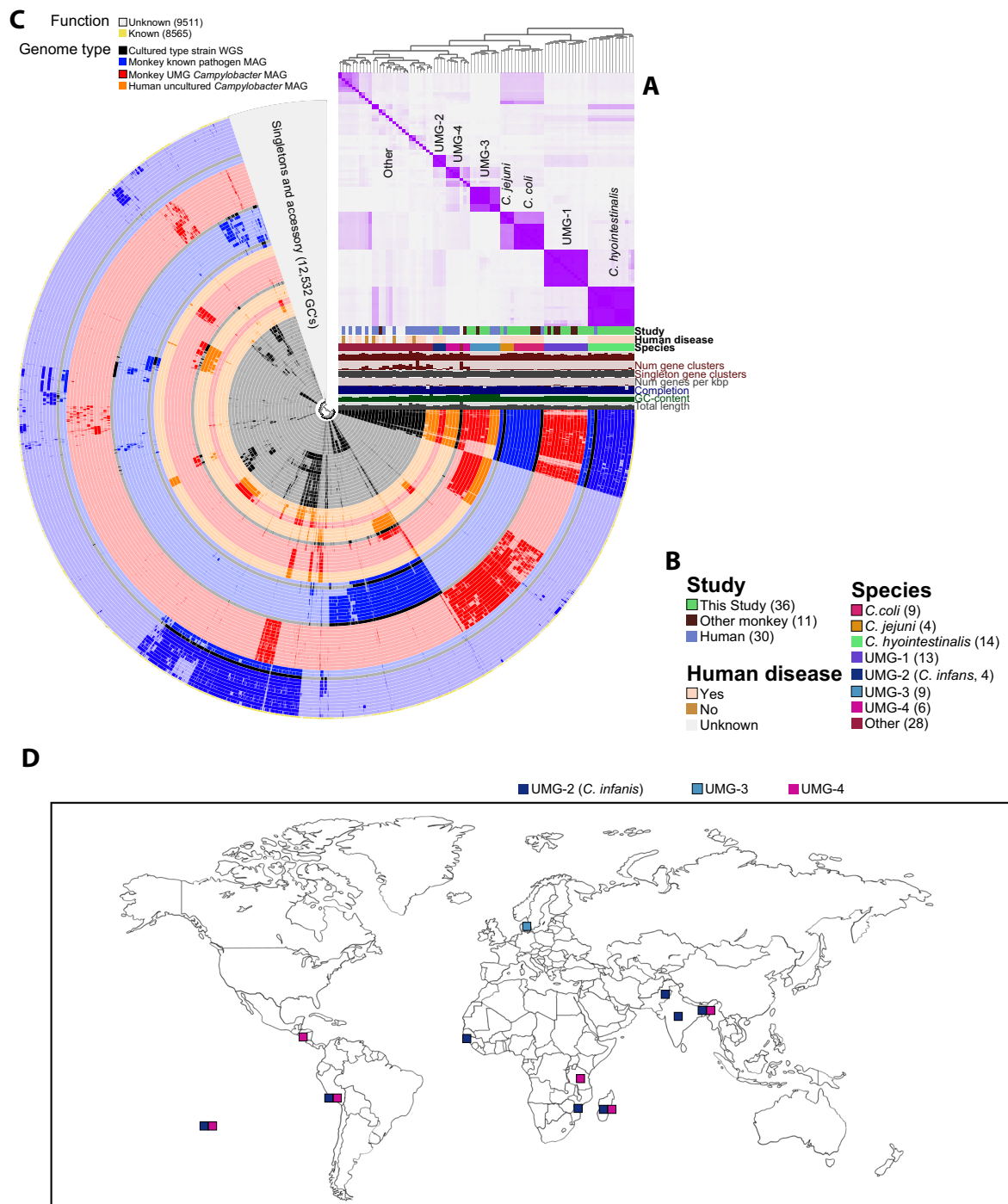
**Additional virulence factors in yet to be cultured *Campylobacter*.** Other traits of *Campylobacter* pathogenesis include iron acquisition<sup>38</sup>, toxin production, and antimicrobial resistance<sup>39</sup>. Using data from our functional enrichment analysis and annotations generated using the Pathosystems Resource Integration Center (PATRIC), we further probed the UMGs for virulence factors associated with these processes. Enterochelin transport system (*ceuA-D*), is an iron acquisition system that is highly conserved in *C. coli* and *jejuni*. This transport system was absent in all 4 *Campylobacter* UMG (Fig. 5A). However, the UMG obtained from both humans and rhesus macaques possessed other genes for iron acquisition including *FeoA*, *HemH*, and *HugZ* (Fig. 5A).

Additionally, we examined the *Liv* operon, a key set of genes involved in the import of the branch chain amino acids leucine, isoleucine, and valine, that is essential for the colonization of the avian caecum<sup>40</sup>. This operon was present in all *C. coli* and *C. jejuni* genomes and all *C. hyointestinalis* except for the *LivJ* gene (Fig. 5A). Interestingly, these genes were absent in all *Campylobacter* MAGs except for those within UMG-2 (*C. infans*) which, like *C. hyointestinalis*, only lacked the *LivJ* gene (Fig. 5A). Finally, we examined a subset of oxidative stress genes important for survival of *Campylobacter* in microaerophilic conditions including catalase and superoxide dismutase<sup>41</sup>. All known and UMG *Campylobacter* species had at least one of these oxidative stress genes suggesting some capacity to neutralize reactive oxygen species (Fig. 5A).

While *Campylobacter* produces a multitude of toxins, only cytotoxin distending toxin (CDT) has been studied in detail<sup>15</sup>. All 3 subunits, encoded by *cdtA-C* are required for toxicity and host cell death<sup>42</sup>. Interestingly, the representative genomes of both *C. coli* and *C. hyointestinalis* selected for this analysis lack *cdtA* and *cdtC* while all rhesus macaque isolates, and MAGs encode all 3 genes (Fig. 5B, Table S8). All 3 subunits were present in both the reference genome and rhesus macaque MAGs of *C. jejuni* (Fig. 5B). Finally, while the complete *cdt* gene set was present in 9 of 27 other known non-*Coli/Jejuni* *Campylobacter* species, these genes were absent in all UMG (Fig. 5B).

The *Campylobacter* Multidrug Efflux system (CME) plays an essential role in the resistance of *C. coli* and *jejuni* to a multitude of antimicrobial agents<sup>43</sup>. CME efflux pumps are composed of two distinct systems, *cmeABC* encode the primary efflux pump while *cmeDEF* encodes a secondary efflux pump that provides additional resistance but not to the level of *cmeABC*<sup>44</sup>. The transcription of both efflux pumps is under regulation of *cmeR*. Both *C. coli* and *C. jejuni* genomes encode all 7 genes (Fig. 5C). Although *C. hyointestinalis* MAGs from rhesus macaques encode the *cmeABC* system, the human reference genome of *C. hyointestinalis* is void of all *cme* genes (Fig. 5C). With the exception of MAGs within UMG-4, which encoded the *cmeDEF* genes, CMW genes were absent in all UMG (Fig. 5C).

Finally, we examined key genes involved in respiration and more specifically the electron transport chain. *Campylobacter* utilize unique and essential electron donors and acceptors such as Formate and Fumarate. Key mutations in the *nuoC-E* genes of respiratory Complex I (NADH ubiquinone oxidoreductase) found in *C. coli*, *C. jejuni*, and many other cultured *Campylobacter* facilitate the use of flavodoxin and menaquinone rather than NADH and ubiquinone as electron carriers to generate

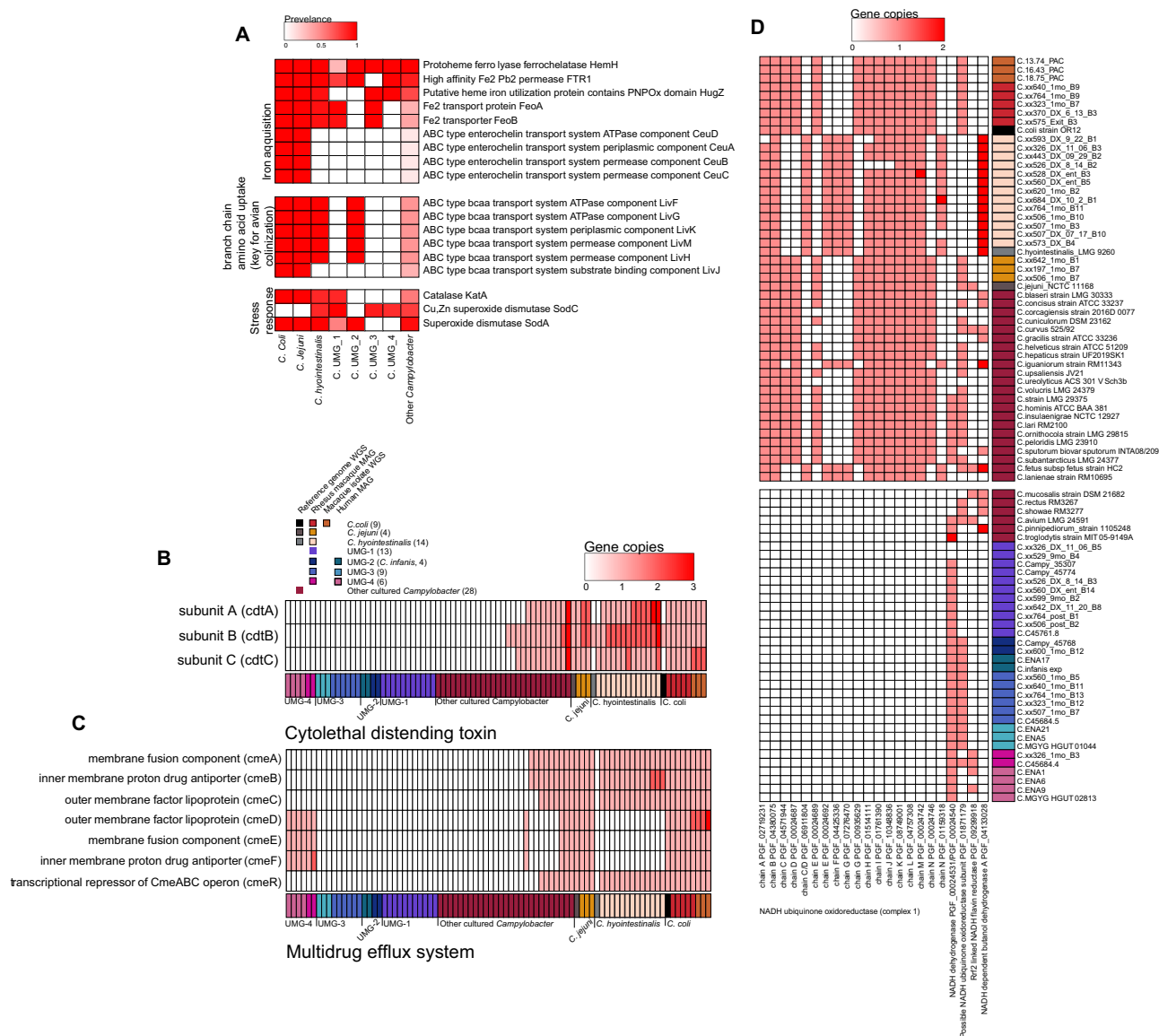


**Fig. 4 A diverse set of *Campylobacter* are associated with diarrheal disease in infant macaques revealed by genome assembly. A** Heatmap of average nucleotide identity (ANI) between all *Campylobacter* genomes in the pangenomic analysis. **B** Metadata and information about genomes included in analysis. **C** Pangenome of 77 *Campylobacter* genomes generated using Anvi'o. Each concentric circle represents a bacterial genome: reference genome (Black), MAGs assembled from rhesus macaques that are similar to a known species (Blue), MAGs assembled from rhesus macaques that are not similar to a known species (UMG) (Red), MAGs assembled from human metagenomes that are similar to our *Campylobacter* UMG genomes (Orange). Blank areas in each circle indicate the absence of a particular gene cluster (GC) in that genome. Outer ring denotes GCs with an annotated function COG function (Yellow). Genomes are ordered by ANI, as depicted by the red heatmap in panel (A). **D** Global map illustrating the distribution of UMGs 2-4 in human data for UMG-2 derived from PCR screening by Bian et al.<sup>27</sup>. Additional distribution information for UMGs 2-4 based on locations of samples collected from which human UMG MAGs were assembled in by Almedia et. al and Nayfach et al.<sup>35,36</sup>.

membrane potential<sup>45,46</sup>. This pattern can be used to distinguish between *C. coli/jejuni* and *C. hyointestinalis/fetus* which use the canonical NADH and ubiquinone as their electron transporters. MAGs from rhesus macaques and genomes from *C. coli/jejuni* and *C. hyointestinalis* isolates recapitulated this pattern (Fig. 5D).

In contrast, all MAGs belonging to UMGs 1-4 lacked all genes needed to assemble respiratory Complex I (NADH ubiquinone oxidoreductase) (Fig. 5D). This was also observed in reference genomes from *C. showae*, *C. mucosalis*, *C. rectus*, *C. avium*, and *C. troglodytis* (Fig. 5D).





**Fig. 5 Iron acquisition, toxin production, antimicrobial resistance and energy generation in Campylobacter UMGs. A** Prevalence heatmap of genes necessary for iron acquisition, branch-chain amino acid uptake, and response to oxidative stress, with the color of each box showing the prevalence of that gene within the noted genome group. **B–D** Prevalence heatmaps for cytolethal distending toxin (**B**), *Campylobacter* multidrug efflux genes (**C**), and electron transport chain complex I (**D**) split by individual genome.

**Discussion**

It is widely accepted that the gut microbiome plays a key role in modulating host susceptibility to enteric disease<sup>3,4,12–14</sup>. However, microbiome studies are challenging to conduct in humans in part due to the substantive individual variability of the human gut microbiome and the difficulties of longitudinal sampling before and after disease. In this study, we leveraged an infant rhesus macaque model to define the maturation/evolution of the gut microbiome throughout the first year of life as well as during and at the end of acute diarrhea episodes.

Data presented in this manuscript recapitulate the pattern of microbiome stunting that has been observed in humans with diarrheal disease<sup>14,47</sup>. This disruption was associated with increased intra-group variability and a reduced diversity of the gut community. The divergence between microbial communities of animals that remained asymptomatic and those that developed diarrhea was evident as early as 1–3 month of age before any clinical episode. Moreover, relative abundance of *Campylobacter* was increased as early as 3-months of age in the microbiome of

infants that later developed diarrhea. These findings suggest that subclinical colonization with pathogens may disrupt the maturation of the microbiome and precede clinical disease<sup>3,13,48</sup>.

In addition to the prokaryotic microbiome, we also profiled the gut mycobiome of infant macaques. We report that the diversity of the mycobiome increased with age, a process that was disrupted by diarrhea, as evidenced by the abundance in rhizophidium and *metschnikowia continentalis* at 12-months of age. In humans, dysbiosis of the gut mycobiome has been reported in individuals with irritable bowel syndrome (IBS) and following antibiotic treatments<sup>49,50</sup>. Further studies are needed to determine if the mycobiome is truly disrupted by diarrheal disease or simply a co-variate of microbial dysbiosis.

We also found that the gut microbiome of dams that gave birth to infants who experienced diarrhea was both functionally and taxonomically distinct prior to the infants becoming sick, including increased pathogen burden (*C. jejuni* and *E. coli*). Our data align with recent findings from human cohorts that demonstrated that the maternal microbiome can modulate the

infant microbiome composition, immune development, and infant growth<sup>51–53</sup>. However, the mechanism by which the maternal microbiome influences protection and/or vulnerability to disease in the infant deserves further attention as the maternal microbiome is dynamic during the perinatal period.

Similar to what has been shown in humans, the microbiome of infant macaques that were still nursing was enriched in pathways associated with milk-oligosaccharide degradation such as fucose degradation<sup>54</sup>. Additionally, microbial pathways for the production of short-chain fatty acids were increased in post-weaned infants in agreement with observational data from humans<sup>54,55</sup>. As noted with taxonomic stunting, diarrhea also disrupted the functional maturation of the infant gut microbiome leading to a reduction in the abundance of folate transformation pathways and an increase in palmitoleate biosynthesis pathway among others. Folate, or vitamin B9, is among the essential vitamins that the host relies on the commensal microbes to synthesize and transform prior to absorption<sup>56</sup>. In line with this observation, reads mapping to some organisms known to be critical in this pathway such as several *Bifidobacterium* species (*B. pseudocatenulatum*, *B. adolescentis*, *B. longum*) were detected<sup>57</sup>. Together, these observations indicate that diarrheal disease disrupts the colonic microbiome's ability to transform folate, in line with histopathological finding from infant macaques with EED<sup>11</sup>.

In agreement with our previous findings, we show that diarrhea was associated with increased abundance of the immunomodulatory palmitoleic acid production pathways<sup>9,58</sup>. Finally, we found that a history of diarrhea resulted in a gut microbiome enriched in pathways for the utilization of alternative electron acceptors such as Nitrate and Sulfate. Host inflammation has been shown to increase Nitrogen electron acceptor and antibiotic use can disrupt redox states within the gut<sup>59</sup>. It is possible that these shifts in redox dynamics prolong microbial dysbiosis and exacerbate inflammatory outcomes.

Taxonomically, we report decreased abundance of beneficial microbes (i.e., *Prevotella*, *Treponema succinifaciens* and *H. macacae*) in infants that experienced diarrhea across the entire timeline. In humans, operational taxonomic units of *Prevotella* have been reported to be overrepresented healthy infants and adults in countries outside of the United States of America<sup>60</sup> in agreement with our earlier report that the infant macaque microbiome was more closely related to that of children in LMIC<sup>9</sup>. Furthermore, *H. macacae* has previously been shown to have a potentially antagonistic association with *Campylobacter* likely through physical competition for resources in the colonic mucosa<sup>9,61</sup>. *Lactobacillus*, a prominent genus in the infant gut, was also reduced following diarrhea<sup>62</sup>. *Lactobacillus* and *Treponema* may play a similar role to *H. macacae* in humans<sup>63,64</sup>. In contrast, *Campylobacter spp.* were among the major taxa that were more abundant in infants that later experienced diarrhea. Due to the self-limiting nature of *Campylobacter* infection, it is probable that, following acute diarrhea, the relative abundance of these bacteria will decrease overtime. However, sick individuals may become asymptomatic carriers resulting in future re-occurrence or spread of the disease. In addition, total *Campylobacter* burden is higher than that of *C. jejuni* and *C. coli* combined in LMIC countries<sup>26–28</sup>, suggesting an important role for other *Campylobacter* species. We and others have identified multiple MAG *Campylobacter* species from both rhesus macaques and humans that have yet to be cultured<sup>9,10,27</sup>. Indeed, several studies have suggested an important role for these unidentified *Campylobacter* species in growth and cognitive stunting in children<sup>28,65,66</sup>. Here we identified 4 genetically distinct *Campylobacter* clusters (UMG) including 15 MAGs assembled in this study and 7 previously assembled MAGs<sup>9,10,34</sup>. We combined the MAGs detected in the infant macaque microbiome with 10

MAGs assembled from human samples that could not be classified to determine their relatedness<sup>27,35,36</sup>.

The MAGs that we identified as UMG-2 shared significant genetic homology and likely belong to the same group as the recently identified candidate species *C. infans* that was assembled from the metagenome of a child experiencing severe diarrhea in India<sup>27</sup>. UMGs 3 and 4 contained genomes assembled from both human and rhesus macaques but their relationship to other known *Campylobacter* species had not been previously investigated. Using reported data sets<sup>27,35,36</sup>, we found that MAG groups UMG-2 and UMG-4 were largely distributed in LMIC countries, suggesting a broader impact of uncultured *Campylobacter*. We investigated the genomic diversity of these unidentified *Campylobacter* MAGs, focusing on known virulence factors and metabolic processes. One caveat to this analysis is the fact that MAGs are often of lower quality than whole genomes sequences (WGS) generated from isolates. Indeed, MAGs are rarely assembled into a single genomic contig and therefore more likely to miss genes that are difficult to assemble from a mixed sample. Therefore, we focused our analysis on large groups of genes that encoded shared functions, notably flagellar assembly, adhesion, components of an efflux pump, iron acquisition and cellular respiration.

UMG-3 lacked all genes for chemotaxis and flagellar assembly suggesting that they are non-motile. This gene group is critical for the ability of pathogenic *Campylobacter* to invade the host mucosa and contact the host epithelia as well as secreting non-flagellar proteins that impact virulence<sup>67,68</sup>. Interestingly *C. gracilis*, and *C. hominis* also lack a flagellum but still cause periodontal disease and colonize the guts of children in the LMIC countries respectively<sup>69–72</sup>. Additional analyses showed that the amino acid sequences of *cadF* and *ciaB* in rhesus MAG were nearly identical to those of human reference genomes suggesting a broad ability of *Campylobacter* to colonize multiple species. Iron acquisition in *C. jejuni/coli* is primarily facilitated by a siderophore enterochelin scavenger system encoded by the *ceu* genes<sup>73</sup>. This system was found in both the human reference genome as well as *C. jejuni/coli* MAGs assembled from rhesus macaques but was absent in the genomes of nearly all other *Campylobacter* species including the 4 UMGs. However, the *Campylobacter* UMG contained alternative genes for iron acquisition (*hugZ*, *frt1*, *feoA/B*)<sup>74–76</sup>. These findings suggest that our unknown UMGs use alternative iron acquisition methods which may impact their pathogenicity. Additionally, UMG-2/*C. infans* encoded genes in the *liv* operon, a branch chain amino acid uptake system that is essential for avian colonization<sup>40</sup>. The presence of these genes may indicate that UMG-2/*C. infans* may have an avian reservoir which provide an explanation for its global distribution.

UMGs lacked genes to produce cytolethal distending toxin (CDT), which has been shown to cleave host DNA resulting in cell death<sup>77</sup>. All three essential components were present in the *C. coli* and *C. hyointestinalis* genomes assembled from rhesus macaque samples. CDT is a key component of *Campylobacter* pathogenicity and its absence in UMGs suggest that they may have additional yet to be described toxins or modes of causing disease. Another key virulence factor is the ability to resist toxic compounds such as bile salts, heavy metals, or antimicrobial agents. This is primarily accomplished in *C. jejuni/coli* via are the *cme* Multidrug efflux pumps *cmeABC* and to a lesser extent *cmeDEF*<sup>44,78</sup> that confer resistance to multiple structurally unrelated antimicrobial compounds including rifampin, cephalothin, cefoperazone<sup>79,80</sup>. The *cme* genes were absent in the UMGs identified in this study with the exception of UMG-4, which encoded *cmeDEF*. This finding may explain why these organisms cannot be isolated on standard *Campylobacter* selective

media. However, since the CME pumps are the primary mechanism by which *C. jejuni/coli* tolerate bile salts<sup>81</sup>, it is likely that the *Campylobacter* UMG encode a different efflux pump that could be challenging to identify without first cultivating these UMG species.

*C. jejuni/coli* harbor key mutations in the *nuoC-E* genes of respiratory Complex I (NADH ubiquinone oxidoreductase) that facilitate the use of flavodoxin and menaquinone rather than NADH and ubiquinone as electron carriers to generate membrane potential<sup>45,46</sup>. Similar isoforms of *nuoC-E* were detected in 19 additional reference *Campylobacter* species. Isoforms of *nuoC-E* genes necessary for the use of NADH and ubiquinone were only detected in *C. hyointestinalis*, *C. fetus*, *C. iguaniorum*, and *C. lanienae*. Surprisingly, UMGs 1–4 lacked the genes for a Complex I all together. Complex I genes are also absent in *C. avium*, *C. showae*, and *C. troglodytis*. The metabolic implications of this finding are unclear highlighting the need for additional studies.

Despite these findings, this study comes with some key limitations. Firstly, while our use of an infant macaque model allows for the control of many variables which typically confound microbiome studies, this model is not guaranteed to reflect clinical findings in LMIC infants. While similar to gut microbiome of human the macaque microbiome has its own unique constituent such as *H. macacae* which are not found in humans. Secondly, this study is limited to molecular methods such as metagenomic genome assembly. While these findings provide strong evidence for pathogenic, metabolic and physiological characteristic of the described UMG *Campylobacter*, these bacteria remain uncultured and therefore lack phenotypic support for our findings.

In summary, this study highlights the need to better understand the role of uncultured, *Campylobacter* species in childhood diarrheal disease burden. While our primary data are derived from infant macaques, the recent sharp increase in publicly available MAGs allowed us to trace these MAGs to human samples obtained from LMIC. This comparative genomic analysis paves the way to address this critical gap in our knowledge.

## Materials and methods

**Animal cohorts.** All rhesus macaque studies were overseen and approved by the OHSU/ONPRC Institutional Animal Care and Use Committees (IACUC) per the National Institutes of Health guide for the care and use of laboratory animals. Animals were housed per the standards established by the US Federal Animal Welfare Act and *The Guide for the Care and Use of Laboratory Animals*. We have complied with all relevant ethical regulations for animal use. All animals were tested for simian viruses (Simian Immunodeficiency Virus, Simian Retrovirus 2, Macacine herpesvirus 1, and Simian T lymphotropic virus) and received a tuberculin test semi-annually.

Infant rhesus macaques are exclusively breastfed for the first 3 months, after which, they are introduced to solid foods and are generally weaned by about 7 months of age. Outdoor-housed rhesus macaques are fed twice daily with Lab Diet, Monkey Diet 5038 (Ralston Purina, St Louis, MO, USA), containing at most 15% crude protein, 5% crude fat, 6% crude fiber, 9% ash, and 12% moisture and after, supplemented daily with fresh fruit and municipal water.

A total of 130 infants were used in this study. Of these 130 infants, 25 were born to dams vaccinated with 40 µg of H<sub>2</sub>O<sub>2</sub> inactivated *Campylobacter Coli* by intramuscular administration twice during their pregnancy and were themselves subsequently vaccinated with the same vaccine as the dams at 1, 3, and 12 months of age; 28 infants were born to unvaccinated dams and were themselves vaccinated at 1, 3, and 12 months of age; finally,

50 infants were born to unvaccinated dams and were themselves vaccinated at 1, 3, and 5 months of age. An additional 27 infants were born to unvaccinated dams and did not receive any vaccines themselves. Details of the vaccination are included in a separate study<sup>29</sup>. We also utilized fecal samples from the dams of these infants ( $n = 103$ ).

For all infants, fecal swabs were obtained at 1, 3, 6, 9 and 12 months of age as well as upon entry and exit from clinic for diarrhea treatment which consisted of oral hydration, antibiotics, and probiotics administered on a case-by-case basis. Each sample was screened for *Campylobacter coli* and *jejuni* as well as *Shigella flexneri* and *dysenteriae* by microbial culture.

**16S amplicon sequencing.** Total DNA was extracted from rectal swabs using the DNeasy Powersoil Pro Kit (Qiagen, Valencia, CA, USA). The hypervariable V4-V5 region of the 16S rRNA gene was amplified using PCR primers (515F/926R with the forward primers including a 12-bp barcode). PCR reactions were conducted in duplicate and contained 12.5 ml GoTaq master mix, 9.5 ml nuclease-free H<sub>2</sub>O, 1 ml template DNA, and 1 ml 10 µM primer mix. Thermal cycling parameters were 94 °C for 5 min, 35 cycles of 94 °C for 20 s, 50 °C for 20 s, 72 °C for 30 s, followed by 72 °C for 5 min. PCR products were purified using a MinElute 96 UF PCR Purification Kit (Qiagen, Valencia, CA, USA). Libraries were sequenced (2 × 300 bases) using an Illumina MiSeq.

Raw FASTQ 16S rRNA gene amplicon sequences were uploaded and processed using the QIIME2 analysis pipeline<sup>82</sup>. Briefly, sequences were demultiplexed and the quality filtered using DADA2<sup>83</sup>, which filters chimeric sequences and generates an amplicon sequence variant (ASV) table equivalent to an operational taxonomic unit (OTU) table at 100% sequence similarity. Sequence variants were then aligned using MAFFT<sup>84</sup> and a phylogenetic tree was constructed using FastTree2<sup>85</sup>. Taxonomy was assigned to sequence variants using q2-feature-classifier against the SILVA database (release 138)<sup>86</sup>. To prevent sequencing depth bias, samples were rarified to 13,781 sequences per sample before alpha and beta diversity analysis. QIIME 2 was also used to generate the following alpha diversity metrics: richness (as observed ASV), Shannon evenness, and phylogenetic diversity. Beta diversity was estimated in QIIME 2 using weighted and unweighted UniFrac distances<sup>87</sup>.

**Shotgun metagenomic library preparation and analysis.** Shotgun metagenomic libraries were prepared from 100 ng of gDNA using the iGenomx RIPTIDE (iGenomx, South San Francisco CA) per iGenomx recommended protocol and sequenced on an Illumina HiSeq 4000 2 × 100. Raw demultiplexed reads were quality filtered using Trimmomatic<sup>88</sup>, and potential host reads were removed by aligning trimmed reads to the *Macaca mulata* genome (Mmul 8.0.1) using BowTie2<sup>89</sup>. Trimmed and decontaminated reads were then annotated using the HUMAnN3 pipeline using default settings with the UniRef90 database and assigned to Metacyc pathways. Functional annotations were normalized using copies per million (CPM) reads before statistical analysis<sup>90–92</sup>. Species-level taxonomy was assigned to quality-controlled short reads using Metaphlan3<sup>32</sup>.

**Metagenomic genome assembly and analysis.** Trimmed and decontaminated reads were assembled into contigs using metaSPAdes with default parameters<sup>93</sup>. Assembled contigs <1 kb were also binned into putative metagenomically assembled genomes (MAGs) using MetaBat<sup>94</sup>. Genome completeness/contamination was tested using CheckM<sup>95</sup>, and all bins with a completeness > 80% and contamination < 2% were annotated using PATRIC<sup>96</sup>. The taxonomy of draft genomes was determined using PATRIC's

similar genome finder. For all assembled genomes classified as *Campylobacter* a core gene codon tree was built using PATRICs Phylogenetic tree building tool which used RAXML<sup>97</sup> to align protein sequences of 50 single copy genes found in all genomes included in the analysis.

Pangenomic analysis was conducted using Anvi'o-V7<sup>33</sup>. Assembled fasta files for each genome were first converted to db files using the anvi-script-FASTA-to-contigs-db command, which uses Prodigal<sup>98</sup> to call open reading frames. Each db file was then annotated against the COG database<sup>99</sup> using the anvi-run-ncbi-cogs command. After generating the genome storage file with the anvi-gen-genomes-storage command, the anvi-pan-genome command was used with parameters outlined in the Anvi'o pangenomics tutorial (<https://merenlab.org/2016/11/08/pangenomics-v2/>). The pangenome was visualized and aesthetics were modified using the *anvi-display-pan* command. GCs found in only a single genome were collapsed into the Singletons bin. GCs found in three or less genomes without a clear pattern, i.e., 3 genomes from 3 different species were collapsed into the Accessory bin to improve the visualization of more abundant GC's. To calculate ANI in Anvi'o, the *anvi-compute-ani* command, which utilizes PyANI<sup>100</sup>, was used. To identify functions (i.e., COG annotations) that were differentially distributed among the phylogroups, we used the *anvi-compute-functional-enrichment-in-pan* command as previously described<sup>101</sup>. Taxonomy of these MAGs were further classified using the Genome Taxonomy Database Tool Kit (GTDB-Tk v2.1.1) and GTDB release 214<sup>102,103</sup>.

**Statistics and reproducibility.** PERMANOVAs were performed using the R package Vegan<sup>104</sup> function ADONIS. 1-way, non-parametric Kruskal-Wallis ANOVA were implemented using PRISM (V8) to generate *p*-values and utilizing the Dunns post-hoc-test when the initial ANOVA was significant. The linear discriminant analysis effect size (LEfSe) algorithm was used to identify differentially abundant taxa and pathways between groups with a logarithmic Linear discriminant analysis (LDA) score cutoff of 2<sup>105</sup>.

**Reporting summary.** Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The datasets supporting the conclusions of this article are available on NCBF's Sequence Read Archive under project numbers PRJNA954012, PRJNA896946, and PRJNA1017507. Supplemental data sets can be found on the online repository Mendeley Data under <https://doi.org/10.17632/8w9r67phvt.2> (Supplemental Tables) and <https://doi.org/10.17632/7hszwny74w.2> (Supplementary Data 1).

### Code availability

No custom code was developed as a part of this study.

Received: 4 July 2023; Accepted: 12 December 2023;

Published online: 05 January 2024

### References

- Liu, L. et al. Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. *Lancet* **388**, 3027–3035 (2016).
- Clasen, T. F. et al. Interventions to improve water quality for preventing diarrhoea. *Cochrane Database Syst Rev*. CD004794 (2015).
- Platts-Mills, J. A. et al. Pathogen-specific burdens of community diarrhoea in developing countries: a multisite birth cohort study (MAL-ED). *Lancet Glob. Health* **3**, e564–e575 (2015).
- Kotloff, K. L. et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* **382**, 209–222 (2013).
- Morrison, D. J. & Preston, T. Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes* **7**, 189–200 (2016).
- Peng, L., Li, Z. R., Green, R. S., Holzman, I. R. & Lin, J. Butyrate enhances the intestinal barrier by facilitating tight junction assembly via activation of AMP-activated protein kinase in Caco-2 cell monolayers. *J. Nutr.* **139**, 1619–1625 (2009).
- den Besten, G. et al. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J. Lipid Res.* **54**, 2325–2340 (2013).
- D'Souza, W. N. et al. Differing roles for short chain fatty acids and GPR43 agonism in the regulation of intestinal barrier function and immune responses. *PLoS ONE* **12**, e0180190 (2017).
- Rhoades, N. et al. Maturation of the infant rhesus macaque gut microbiome and its role in the development of diarrheal disease. *Genome Biol.* **20**, 173 (2019).
- Rhoades, N. S. et al. Growth faltering regardless of chronic diarrhea is associated with mucosal immune dysfunction and microbial dysbiosis in the gut lumen. *Mucosal Immunol.* **14**, 1113–1126 (2021).
- Hendrickson, S. M. et al. Reduced infant rhesus macaque growth rates due to environmental enteric dysfunction and association with histopathology in the large intestine. *Nat. Commun.* **13**, 234 (2022).
- Rouhani, S. et al. Diarrhea as a potential cause and consequence of reduced gut microbial diversity among undernourished children in Peru. *Clin. Infect. Dis.* **71**, 989–999 (2020).
- Rouhani, S. et al. Gut microbiota features associated with campylobacter burden and postnatal linear growth deficits in a Peruvian birth cohort. *Clin. Infect. Dis.* **71**, 1000–1007(2020).
- Subramanian, S. et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature* **510**, 417–421 (2014).
- Igwaran, A. & Okoh, A. I. Human campylobacteriosis: a public health concern of global importance. *Heliyon* **5**, e02814 (2019).
- Liu, L. et al. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet* **379**, 2151–2161 (2012).
- Brown, K. H., Khatun, M. & Ahmed, G. Relationship of the xylose absorption status of children in Bangladesh to their absorption of macronutrients from local diets. *Am. J. Clin. Nutr.* **34**, 1540–1547 (1981).
- Ferdous, F. et al. Severity of diarrhea and malnutrition among under five-year-old children in rural Bangladesh. *Am. J. Trop. Med. Hyg.* **89**, 223–228 (2013).
- Manary, M. J. et al. Perturbed zinc homeostasis in rural 3–5-year-old Malawian children is associated with abnormalities in intestinal permeability attributed to tropical enteropathy. *Pediatr. Res.* **67**, 671–675 (2010).
- Moore, S. R. et al. Prolonged episodes of acute diarrhea reduce growth and increase risk of persistent diarrhea in children. *Gastroenterology* **139**, 1156–1164 (2010).
- Pruss-Ustun, A. et al. Burden of disease from inadequate water, sanitation and hygiene for selected adverse health outcomes: an updated analysis with a focus on low- and middle-income countries. *Int. J. Hyg. Environ. Health.* **222**, 765–777 (2019).
- Brouwer, A. F., Masters, N. B. & Eisenberg, J. N. S. Quantitative microbial risk assessment and infectious disease transmission modeling of waterborne enteric pathogens. *Curr. Environ. Health Rep.* **5**, 293–304 (2018).
- Castano-Rodriguez, N., Kaakoush, N. O., Lee, W. S. & Mitchell, H. M. Dual role of *Helicobacter* and *Campylobacter* species in IBD: a systematic review and meta-analysis. *Gut* **66**, 235–249 (2017).
- Mukhopadhyay, I. et al. Detection of *Campylobacter concisus* and other *Campylobacter* species in colonic biopsies from adults with ulcerative colitis. *PLoS ONE* **6**, e21490 (2011).
- Warren, R. L. et al. Co-occurrence of anaerobic bacteria in colorectal carcinomas. *Microbiome* **1**, 16 (2013).
- Platts-Mills, J. A. et al. Detection of *Campylobacter* in stool and determination of significance by culture, enzyme immunoassay, and PCR in developing countries. *J. Clin. Microbiol.* **52**, 1074–1080 (2014).
- Bian, X. et al. *Campylobacter* abundance in breastfed infants and identification of a new species in the global enterics multicenter study. *mSphere* **5**, e00735–19 (2020).
- Francois, R. et al. The other *Campylobacter*s: not innocent bystanders in endemic diarrhea and dysentery in children in low-income settings. *PLoS Negl. Trop. Dis.* **12**, e0006200 (2018).
- Hendrickson, S. M. et al. *Campylobacter* vaccination reduces diarrheal disease and infant growth stunting among rhesus macaques. *Nat. Commun.* **14**, 3806 (2023).

30. Platts-Mills, J. A. & Kosek, M. Update on the burden of *Campylobacter* in developing countries. *Curr. Opin. Infect. Dis.* **27**, 444–450 (2014).
31. Kaur, T. et al. *Campylobacter troglodytis* sp. nov., isolated from feces of human-habituated wild chimpanzees (*Pan troglodytes schweinfurthii*) in Tanzania. *Appl. Environ. Microbiol.* **77**, 2366–2373 (2011).
32. Beghini, F. et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife* **10**, e65088 (2021).
33. Eren, A. M. et al. Community-led, integrated, reproducible multi-omics with anvio. *Nat. Microbiol.* **6**, 3–6 (2021).
34. Quintel, B. K. et al. Vaccine-mediated protection against *Campylobacter*-associated enteric disease. *Sci. Adv.* **6**, eaba4511 (2020).
35. Almeida, A. et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat. Biotechnol.* **39**, 105–114 (2021).
36. Nayfach, S., Shi, Z. J., Seshadri, R., Pollard, K. S. & Kyrpides, N. C. New insights from uncultivated genomes of the global human gut microbiome. *Nature* **568**, 505–510 (2019).
37. Ziprin, R. L. et al. Role of *Campylobacter jejuni* potential virulence genes in cecal colonization. *Avian Dis.* **45**, 549–557 (2001).
38. Skaar, E. P. The battle for iron between bacterial pathogens and their vertebrate hosts. *PLoS Pathog.* **6**, e1000949 (2010).
39. Kreling, V., Falcone, F. H., Kehrenberg, C. & Hensel, A. *Campylobacter* sp.: pathogenicity factors and prevention methods—new molecular targets for innovative antivirulence drugs? *Appl. Microbiol. Biotechnol.* **104**, 10409–10436 (2020).
40. Ribardo, D. A. & Hendrixson, D. R. Analysis of the LIV system of *Campylobacter jejuni* reveals alternative roles for LivJ and LivK in commensalism beyond branched-chain amino acid transport. *J. Bacteriol.* **193**, 6233–6243 (2011).
41. Kim, J. C., Oh, E., Kim, J. & Jeon, B. Regulation of oxidative stress resistance in *Campylobacter jejuni*, a microaerophilic foodborne pathogen. *Front. Microbiol.* **6**, 751 (2015).
42. Mendez-Olvera, E. T., Bustos-Martinez, J. A., Lopez-Vidal, Y., Verdugo-Rodriguez, A. & Martinez-Gomez, D. Cytolethal distending toxin from *Campylobacter jejuni* requires the cytoskeleton for toxic activity. *Jundishapur. J. Microbiol.* **9**, e35591 (2016).
43. Lin, J., Michel, L. O. & Zhang, Q. CmeABC functions as a multidrug efflux system in *Campylobacter jejuni*. *Antimicrob. Agents Chemother.* **46**, 2124–2131 (2002).
44. Akiba, M., Lin, J., Barton, Y. W. & Zhang, Q. Interaction of CmeABC and CmeDEF in conferring antimicrobial resistance and maintaining cell viability in *Campylobacter jejuni*. *J. Antimicrob. Chemother.* **57**, 52–60 (2006).
45. St Maurice, M. et al. Flavodoxin:quinone reductase (FqrB): a redox partner of pyruvate:ferredoxin oxidoreductase that reversibly couples pyruvate oxidation to NADPH production in *Helicobacter pylori* and *Campylobacter jejuni*. *J. Bacteriol.* **189**, 4764–4773 (2007).
46. Weerakoon, D. R. & Olson, J. W. The *Campylobacter jejuni* NADH:ubiquinone oxidoreductase (complex I) utilizes flavodoxin rather than NADH. *J. Bacteriol.* **190**, 915–925 (2008).
47. Gehrig, J. L. et al. Effects of microbiota-directed foods in gnotobiotic animals and undernourished children. *Science* **365**, eaau4732 (2019).
48. Lee, G. et al. Symptomatic and asymptomatic *Campylobacter* infections associated with reduced growth in Peruvian children. *PLoS Negl. Trop. Dis.* **7**, e2036 (2013).
49. Das, A., O’Herlihy, E., Shanahan, F., O’Toole, P. W. & Jeffery, I. B. The fecal mycobiome in patients with Irritable Bowel Syndrome. *Sci. Rep.* **11**, 124 (2021).
50. Seelbinder, B. et al. Antibiotics create a shift from mutualism to competition in human gut communities with a longer-lasting impact on fungi than bacteria. *Microbiome* **8**, 133 (2020).
51. Rhoades, N. S., Cinco, I. R., Hendrickson, S. M., Slifka, M. K. & Messaoudi, I. Taxonomic and functional shifts in the perinatal gut microbiome of rhesus macaques. *Microbiol. Spectr.* **10**, e0081422 (2022).
52. Gough, E. K. et al. Maternal fecal microbiome predicts gestational age, birth weight and neonatal growth in rural Zimbabwe. *EBioMedicine* **68**, 103421 (2021).
53. Nyangahu, D. D. et al. Disruption of maternal gut microbiota during gestation alters offspring microbiota and immunity. *Microbiome* **6**, 124 (2018).
54. Koenig, J. E. et al. Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl Acad. Sci. USA* **108**, 4578–4585 (2011).
55. Nilsen, M. et al. Butyrate levels in the transition from an infant- to an adult-like gut microbiota correlate with bacterial networks associated with *Eubacterium Rectale* and *Ruminococcus Gnavus*. **11**, 1245 (2020).
56. Engevik, M. A. et al. Microbial metabolic capacity for intestinal folate production and modulation of host folate receptors. *Front. Microbiol.* **10**, 2305 (2019).
57. Sugahara, H., Odamaki, T., Hashikura, N., Abe, F. & Xiao, J. Z. Differences in folate production by bifidobacteria of different origins. *Biosci. Microbiota Food Health* **34**, 87–93 (2015).
58. Schirmer, M. et al. Linking the human gut microbiome to inflammatory cytokine production capacity. *Cell* **167**, 1897 (2016).
59. Reese, A. T. et al. Antibiotic-induced changes in the microbiota disrupt redox dynamics in the gut. *Elife* **7**, e35987 (2018).
60. Yatsunenko, T. et al. Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).
61. Laing, S. T. et al. Idiopathic colitis in rhesus macaques is associated with dysbiosis, abundant enterochromaffin cells and altered T-cell cytokine expression. *Vet. Pathol.* **55**, 741–752 (2018).
62. Fan, Q. et al. The impact of age and pathogens type on the gut microbiota in infants with diarrhea in Dalian, China. *Can. J. Infect. Dis. Med. Microbiol.* **2020**, 8837156 (2020).
63. Bharadia, L., Agrawal, N. & Joshi, N. Development and functions of the infant gut microflora: Western vs. Indian infants. *Int. J. Pediatr.* **2020**, 7586264 (2020).
64. De Filippo, C. et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl Acad. Sci. USA* **107**, 14691–14696 (2010).
65. Chen, D. et al. *Campylobacter* colonization, environmental enteric dysfunction, stunting, and associated risk factors among young children in rural Ethiopia: a cross-sectional study from the *Campylobacter* Genomics and Environmental Enteric Dysfunction (CAGED) Project. *Front Public Health* **8**, 615793 (2020).
66. Terefe, Y. 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* **8**, 99 (2020).
67. Guerry, P. *Campylobacter* flagella: not just for motility. *Trends Microbiol.* **15**, 456–461 (2007).
68. Yao, R. et al. Isolation of motile and non-motile insertional mutants of *Campylobacter jejuni*: the role of motility in adherence and invasion of eukaryotic cells. *Mol. Microbiol.* **14**, 883–893 (1994).
69. Johnson, C. C. et al. *Bacteroides gracilis*, an important anaerobic bacterial pathogen. *J. Clin. Microbiol.* **22**, 799–802 (1985).
70. Doan, T. et al. Gut microbiome alteration in MORDOR I: a community-randomized trial of mass azithromycin distribution. *Nat. Med.* **25**, 1370–1376 (2019).
71. Siqueira, J. F. Jr & Rocas, I. N. *Campylobacter gracilis* and *Campylobacter rectus* in primary endodontic infections. *Int. Endod. J.* **36**, 174–180 (2003).
72. Lawson, A. J., On, S. L., Logan, J. M. & Stanley, J. *Campylobacter hominis* sp. nov., from the human gastrointestinal tract. *Int. J. Syst. Evol. Microbiol.* **51**, 651–660 (2001).
73. Richardson, P. T. & Park, S. F. Enterochelin acquisition in *Campylobacter coli*: characterization of components of a binding-protein-dependent transport system. *Microbiology* **141**, 3181–3191 (1995).
74. Guo, Y. et al. Functional identification of HugZ, a heme oxygenase from *Helicobacter pylori*. *BMC Microbiol.* **8**, 226 (2008).
75. Taveirne, M. E., Theriot, C. M., Livny, J. & DiRita, V. J. The complete *Campylobacter jejuni* transcriptome during colonization of a natural host determined by RNAseq. *PLoS ONE* **8**, e73586 (2013).
76. Naikare, H., Palyada, K., Panciera, R., Marlow, D. & Stintzi, A. Major role for FeoB in *Campylobacter jejuni* ferrous iron acquisition, gut colonization, and intracellular survival. *Infect. Immun.* **74**, 5433–5444 (2006).
77. Whitehouse, C. A. et al. *Campylobacter jejuni* cytolethal distending toxin causes a G2-phase cell cycle block. *Infect. Immun.* **66**, 1934–1940 (1998).
78. Vieira, A., Ramesh, A., Seddon, A. M. & Karlyshev, A. V. CmeABC multidrug efflux pump contributes to antibiotic resistance and promotes *Campylobacter jejuni* survival and multiplication in *Acanthamoeba polyphaga*. *Appl. Environ. Microbiol.* **83**, e01600–17 (2017).
79. Ng, L. K., Taylor, D. E. & Stiles, M. E. Characterization of freshly isolated *Campylobacter coli* strains and suitability of selective media for their growth. *J. Clin. Microbiol.* **26**, 518–523 (1988).
80. Guo, B., Lin, J., Reynolds, D. L. & Zhang, Q. Contribution of the multidrug efflux transporter CmeABC to antibiotic resistance in different *Campylobacter* species. *Foodborne Pathog. Dis.* **7**, 77–83 (2010).
81. Lin, J., Sahin, O., Michel, L. O. & Zhang, Q. Critical role of multidrug efflux pump CmeABC in bile resistance and in vivo colonization of *Campylobacter jejuni*. *Infect. Immun.* **71**, 4250–4259 (2003).
82. Bolyen, E. et al. QIIME 2: reproducible, interactive, scalable, and extensible microbiome data science. *Nat. Biotechnol.* **37**, 852–857 (2019).
83. Callahan, B. J. et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
84. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
85. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
86. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).

87. Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J. & Knight, R. UniFrac: an effective distance metric for microbial community comparison. *ISME J.* **5**, 169–172 (2011).
88. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
89. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
90. Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B. & Wu, C. H. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926–932 (2015).
91. Abubucker, S. et al. Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput. Biol.* **8**, e1002358 (2012).
92. Karp, P. D., Riley, M., Paley, S. M. & Pellegrini-Toole, A. The MetaCyc Database. *Nucleic Acids Res.* **30**, 59–61 (2002).
93. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
94. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
95. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
96. Wattam, A. R. et al. Assembly, annotation, and comparative genomics in PATRIC, the all bacterial bioinformatics resource center. *Methods Mol. Biol.* **1704**, 79–101 (2018).
97. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
98. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
99. Galperin, M. Y., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* **43**, D261–D269 (2015).
100. Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G. & Toth, I. K. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* **8**, 12–24 (2016).
101. Shaiber, A. et al. Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. *Genome Biol.* **21**, 292 (2020).
102. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**, 5315–5316 (2022).
103. Parks, D. H. et al. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat. Biotechnol.* **38**, 1079–1086 (2020).
104. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
105. Segata, N. et al. Metagenomic biomarker discovery and explanation. *Genome Biol* **12**, R60 (2011).

### Author contributions

N.S.R., M.K.S., and I.M. conceived and designed the experiments. S.M.H., K.P., A.J.H., and G.E.F. collected samples and provided reagents. N.S.R. and I.R.C. performed the experiments and analyzed the data. N.S.R., I.M. wrote the first draft of the paper. All authors read, edited and approved the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-023-05695-0>.

**Correspondence** and requests for materials should be addressed to Ilhem Messaoudi.

**Peer review information** *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Tobias Goris. A peer review file is available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024