Supplementary Methods

Bacteria strains and culture conditions

The human clinical isolate *C. jejuni* 81–176 and mut*cdtB* were cultured on *Campylobacter* selective medium (Remel) at 37 °C for 48 hours under micro-aerobic conditions using the GasPak system (BD) (1).

Bacterial lysate preparation

Wide type *C. jejuni* 81-176 or mut*cdtB* strain were grown micro-aerobically on *Campylobacter* selective medium (Remel) at 37 °C for 48h. Bacteria were harvested, suspended in sterile PBS, pelleted at 3000rpm and washed twice in sterile PBS. Bacterial suspensions were sonicated (Sonicator 3000, Misonix) on ice for four 30-seconds bursts with 30-seconds intervals in between. After sonication, bacterial lysate was centrifuged at 5000rpm×4 °C for 10min and passed through a 0.22μm sterile syringe filter (Olympus). The concentration of protein content was measured using the Bio-Rad protein assay (Bio-Rad).

RNA extraction and RNA-sequencing

Total RNA was extracted from frozen distal colon tissue snips using bead beater disruption following by the manufacturer's instructions of the mirVana miRNA isolation kit with phenol (ThermoFisher Scientific). Extracted RNA was treated with the Turbo DNA-free Kit (ThermoFisher Scientific) to remove DNA. Quality control, rRNA depletion and cDNA library preparation was performed by the University of Florida's Interdisciplinary Center for Biotechnology Research (ICBR) Gene Expression and Genotyping core using the Agilent 2100 Bioanalyzer (Agilent Genomics), Ribo-Zero Gold rRNA Removal Kit (Epidemiology) (Illumina) and NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB) starting with 550 ng total RNA. Samples were sequenced by the University of Florida ICBR NextGen DNA Sequencing core using one lane of Illumina HiSeq 3000 (2×100 cycles).

16S rDNA sequencing

Mouse stool samples were collected and fecal DNA was extracted using bead beater disruption and phenol:chloroform separation according to the DNeasy Blood & Tissue Kit (Qiagen) manufacturer's recommendations as described before (2). The V1-V3 hypervariable region of the 16S rRNA gene was amplified using primer pair 8F (5'- AGAGTTTGATCCTGGCTCAG -3') and 534R (5'-ATTACCGCGGCTGCTGG-3'). Both the forward and the reverse primers contained universal Illumina paired-end adapter sequences, as well as unique individual 4 to 6 nucleotide barcodes between the PCR primer sequence and the Illumina adapter sequence to allow multiplex sequencing. PCR products were visualized on an agarose gel before and after samples were purified using the Agencourt AMPure XP kit (Beckman Coulter) and quantified by qPCR with the KAPA Library Quantification Kit (KAPA

Biosystems). Equimolar amounts of samples were then pooled and sequenced with an Illumina MiSeq.

Mouse RNA-seq Analysis

Reads were quality filtered at Q20 and trimmed to remove remaining adaptors using Trimmomatic (3) version 0.36. The resulting reads were aligned to Illumina iGenome *Mus musculus* Ensembl GRCm38 reference genome using Tophat (4) version 2.1.1 utilizing Bowtie2 (5) version 2.3.0 following the approach of Gilad and Mizrahi-Man (6). The resulting alignments (averaging 21,051,145 reads per sample aligned uniquely to mouse transcriptome) were processed using Cufflinks (7) version 2.2.1 along with Illumina iGenome *Mus musculus* Ensembl GRCm38 Gene transfer format file, after masking rRNA features. We used cuffquant to perform transcript quantification and exported the raw counts (un-normalized counts) to text files. The Raw counts were then imported to edgeR (8) version 3.16.5 for detecting DE genes. A gene was considered for differential expression test if it was present in at least 3 samples and its edgeR's counts per million (CPM) value is at least 2. We considered a gene DE if its edgeR FDR adjusted P-value < 0.05.

Parallel analysis using featureCounts (9) from the subread package version 1.5.3 for transcript quantification showed similar results (data not shown). Pathway analysis was conducted through GAGE (10) version 2.24 using *Mus musculus* (mmu) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and we considered a pathway significant if its GAGE q-value was less than 0.05. Genes were mapped to KEGG pathways (11) using Pathview (12). Raw sequencing reads have been deposited in the NCBI Sequence Read Archive (SRA) under accession number SRP150780.

Metatranscriptome Analysis

Quality filtered and trimmed reads from above were aligned to iGenome *Mus musculus* Ensembl GRCm38 reference genome using Bowtie2 (5) version 2.3.0 using the "very-sensitive" option and reads with alignments were excluded from further analysis. The remaining reads were then filtered from rRNA and tRNA sequences by aligning to a collection of NCBI rRNA and tRNA sequences and SLIVA database sequences, which were then submitted for *de novo* assembly using Trinity version 2.4.0 (13). The resulting assembly was annotated using Trinotate (14) version 3.0.1 (http://trinotate.github.io) with the following databases: uniprot_uniref90, uniprot_sprot, Pfam (15,16) and Virulence Factor Database (VFDB) (17). The resulting annotations were examined and sequences annotated as non-bacterial were removed. Transcripts abundance was determined using RNA-Seq by Expectation Maximization (RSEM) (18) through Trinity's align_and_estimate_abundance.pl script and resulting in an average of 16,423 bacterial reads per sample. Those counts were imported to edgeR (8) version 3.16.5 for differential expression analysis. A gene was considered for differential expression test if it was present in at least 3 samples and its edgeR's counts per million (CPM) value is at least 2. We considered a transcript DE if its edgeR FDR adjusted P-value

< 0.05. Pathway analysis was conducted through GAGE (10) version 2.24 using Kyoto Encyclopedia of Genes and Genomes (KEGG) reference pathways (11). Genes were mapped to KEGG pathways using Pathview (12). We considered a pathway significant if its q-value < 0.05.

16S rRNA Sequencing Analysis

Reads were preprocessed using Quantitative Insights into Microbial Ecology (QIIME) version 1.9.1 including merging (using QIIME default parameters), trimming and filtering at Q20 (19). The final set of reads was fed to QIIME's pick_closed_reference_otus.py to pick closed-reference OTUs at 97% similarity level using the Greengenes 97% reference dataset (release 13_8). Taxonomic assignment was done using RDP (ribosomal database project) classifier through QIIME's parallel_assign_taxonomy_rdp.py with confidence set to 50% (20). We excluded OTUs that had ≤ 0.005% of the total number of sequences according to Bokulich and colleagues (21). This resulted in an average of 72,588 reads per sample. The resulting OTU counts were normalized and $\log_{10}$ transformed using the following formula (22):

$$log_{10}\left(\frac{RC}{n}\ \mathrm{x}\ \frac{\sum x}{N}+1\right)$$

where RC is the read count for a particular OTU in a particular sample, *n* is the total number of reads in that sample, the sum of *x* is the total number of reads in all samples and N is the total number of samples. The Principle Coordinate Analysis (PCoA) was generated from the Bray-Curtis distance of the normalized and $\log_{10}$ transformed counts using the phyloseq R package (23).

Alpha diversity (Shannon diversity index) was calculated using the phyloseq R package after rarefying the raw counts to an even depth (the minimum count in all samples for each time-point samples: 14 days time-point= 13,449 and 45 days time-point= 31,262).

Pairwise differences between groups (Control, *mutcdtB*, *C. jejuni*) at the PCoA axis and taxa levels were tested using the gls function in the R nlme package, with the REML method to fit a mixed linear model of the form:

variable ~ group + ε

where variable indicates PCoA axis, Alpha diversity index, the $\log_{10}$ normalized abundance of a particular taxa and 1|cage indicates that we used the cage as a random effect. We then ran an ANOVA analysis on the above model to generate p-values. We checked for possible cage effect by comparing the above model and an lme model with the cage (variable ~ group+ 1|cage + ε) using ANOVA. The p-values were adjusted for multiple hypothesis testing in R using the p.adjust function employing the method of Benjamini & Hochberg (24). For all our taxa comparisons, taxa present in less than 25% of the samples were excluded.

16S sequencing reads have been deposited in the NCBI Sequence Read Archive (SRA) under accession number SRP150765.

Nakatsu et al. (25) seqeunces were downloaded from the European Nucleotide Archive (study ID: PRJNA280026) and analyzed as decribed above.

***C. jejuni* colonization in fecal samples**

Stools were homogenized, serially diluted in sterilize PBS and plated on *Campylobacter* selective medium (Remel) at 37 °C for 48 hours under micro-aerobic condition using the GasPak system (BD). Colonies were inspected by eyes and counted. The stools and tissues of SPF colonized mice were negative for *C. jejuni* as assayed by CFU count on *Campylobacter* selective agar plates or FISH signals on tissue sections.

Fluorescence in situ hybridization (FISH)

C. jejuni in colonic tissue sections was visualized using FISH assay as previously described (2). Briefly, Cy3-tagged 5'AGCTAACCACACCTTATACCG3' was used as probe to detect the presence of *C. jejuni*. Deparaffinized 5μm thick tissue sections were incubated in lysozyme solution, washed, hybridized with the FISH probe overnight, washed, stained with DAPI, and imaged using a fluorescent Microscope system.

Immunohistochemistry (IHC)

IHC was performed as described previously (2,26). Briefly, colonic tissue sections were deparaffinized, rehydrated, boiled in citrate buffer for antigen retrieval and blocked. For β-catenin, the mouse anti-β-catenin antibody (1:300, 4 °C overnight; 610153, BD Biosciences) was used. For PCNA, mouse anti-PCNA antibody (M0879, Dako) (1:300, 30 minutes at room temperature) was used. For p-S6, rabbit anti-p-S6 antibody (1:100, 4 °C overnight; 2217S, Cell signaling) was used.

Comet assay

The comet assay (single cell gel electrophoresis) was performed with the Trevigen Comet Assay™ kit (Trevigen) according to the manufacturer's protocol. Briefly, IEC-6 and HT-29 cells were plated at 10^5 /well in 6-well tissue culture plates and serum starved in 2% fetal bovine serum (FBS) reduced medium overnight (16-18h). Cells were incubated with bacterial lysates (5μg/ml) for 24h, washed by PBS, lifted by trypsin and collected in PBS to a concentration of 1 x 10^5 cells/ml, mixed with 37 °C 1% LMAgarose (low-melting agarose) (Trevigen) and loaded on 20-well CometSlides (Trevigen). CometSlides were placed in the pre-cold lysis solution (Trevigen) at 4 °C for 60 min, and then incubated in alkaline electrophoresis solution (300mM NaOH, 1 mM EDTA, pH > 13) at room temperature for 20 min, in the dark. CometSlides were transferred to pre-cold fresh alkaline electrophoresis solution (300mM NaOH, 1mM EDTA, pH > 13) and subjected to electrophoresis using the CometAssay Electrophoresis System II (Trevigen) for 30 min (21 V). Slides were washed twice in dH2O for 5 min and 70% ethanol for 5 min. DNA was stained with 50ul SYBR™ Gold nucleic acid gel stain (1:10,000 dilution in Tris-EDTA solution, S-11494, ThermoFisher

scientific) in a light-protected setting for 30 min and visualized using the Leica DM6000B upright microscope.

Immunofluorescence detection of γH2AX in cell lines

IEC-6 and HT-29 cell lines were plated on 8-well chamber slide system at $5x10^4$/ well and serum starved in 2% FBS reduced medium overnight. Cells were incubated with bacterial lysates (5µg/ml) for 24 h, washed by cold PBS and fixed with 3.7% methanol-free formaldehyde (Thermo Fisher Scientific) for 30 min on ice. Cells were permeabilized using ice-cold methanol for 10 min, washed with PBS to remove methanol, blocked with blocking solution (PBS containing 1% BSA and 5% goat serum) for 1 h on ice, and incubated with the Phospho-Histone H2AX (Ser139) rabbit monoclonal antibody (Cell Signaling) overnight at 4 °C. Cells were washed with PBS and incubated in the secondary Alexa Fluor 488 goat anti-rabbit IgG (H+L) antibody (Life Technologies) for 45 min at room temperature. Cells were washed again with PBS and mounted with VECTASHIELD mounting medium with DAPI (Vector Labs). Images were taken using a Leica DM6000B upright microscope.

γH2AX flow cytometry and cell cycle analysis

For γH2AX flow cytometry, cells were incubated with bacterial lysates for 24h. For cell cycle assay, cells were incubated with bacterial lysates for 48h. After incubation with bacterial lysates, cells were collected in PBS, fixed in ice-cold 1% methanol-free formaldehyde solution (Thermo Fisher Scientific) for 15 min, washed in PBS, and incubated in 70% ethanol overnight at −20 °C. Cells were washed with BSA-T-PBS (PBS containing 1% BSA and 0.2% Triton X-100), and incubated with the Alexa Fluor 647 anti-H2AX-phosphorylated (Ser139) antibody (BioLegend, diluted 1:200 in BSA-T-PBS) overnight at 4 °C. Cells were washed with BSA-T-PBS and incubated in propidium iodide (Life Technologies) with 100µg/ml RNase (Sigma-Aldrich). At least 10,000 cells from each sample were analyzed using the LSR Fortessa flow cytometer (BD Biosciences). Data were further processed and analyzed using FCS Express 5 software (http://www.denovosoftware.com).

REFERENCES

1. Sun X, Threadgill D, Jobin C. Campylobacter jejuni induces colitis through activation of mammalian target of rapamycin signaling. Gastroenterology. 2012;142:86–95.e5.
2. Sun X, Winglee K, Gharaibeh RZ, Gauthier J, He Z, Tripathi P, et al. Microbiota-Derived Metabolic Factors Reduce Campylobacteriosis in Mice. Gastroenterology. 2018;0.
3. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
4. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:R36.
5. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
6. Gilad Y, Mizrahi-Man O. A reanalysis of mouse ENCODE comparative gene expression data. [version 1; referees: 3 approved, 1 approved with reservations]. F1000Res. 2015;4:121.
7. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7:562–78.
8. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
9. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.
10. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics. 2009;10:161.
11. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30.
12. Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics. 2013;29:1830–1.
13. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52.
14. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8:1494–512.
15. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44:D279-85.

16. UniProt Consortium T. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2018;46:2699.
17. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. Nucleic Acids Res. 2016;44:D694-7.
18. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.
19. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7:335–6.
20. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73:5261–7.
21. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods. 2013;10:57–9.
22. McCafferty J, Mühlbauer M, Gharaibeh RZ, Arthur JC, Perez-Chanona E, Sha W, et al. Stochastic changes over time and not founder effects drive cage effects in microbial community assembly in a mouse model. ISME J. 2013;7:2116–25.
23. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8:e61217.
24. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society Series B …. 1995;
25. Nakatsu G, Li X, Zhou H, Sheng J, Wong SH, Wu WKK, et al. Gut mucosal microbiome across stages of colorectal carcinogenesis. Nat Commun. 2015;6:8727.
26. Tomkovich S, Yang Y, Winglee K, Gauthier J, Mühlbauer M, Sun X, et al. Locoregional effects of microbiota in a preclinical model of colon carcinogenesis. Cancer Res. 2017;77:2620–32.