SUPPLEMENTARY FIGURES AND REFERENCES FOR:

Campylobacter jejuni transcriptional and genetic adaptation during acute and persistent human infections

Alexander A. Crofts[a], Frédéric M. Poly[b], Cheryl P. Ewing[b], Janelle M. Kuroiwa[b], Joanna E. Rimmer[bgh], Clayton Harro[e], David Sack[e], Kawsar R. Talaat[e], Chad K. Porter[b], Ramiro L. Gutierrez[b], Barbara DeNearing[e], Jessica Brubaker[e], Renée M. Laird[b], Alexander C. Maue[b], Kayla Jaep[b], Ashley Alcala[b], David R. Tribble[f], Mark. S. Riddle[b,] Amritha Ramakrishnan[i], Andrea J. McCoy[i], Bryan W. Davies[ac], Patricia Guerry[b*] and M. Stephen Trent[d,j*]

Affiliations:

[a]Department of Molecular Biosciences, University of Texas at Austin, Austin, TX, 78712, USA
[b]Enteric Diseases Department, Naval Medical Research Center, Silver Spring, MD,
[c]Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX, 78712, USA
[d]Department of Infectious Diseases, University of Georgia at Athens, College of Veterinary Medicine, Athens, Georgia, 30602, USA
[e]Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
[f]Infectious Disease Clinical Research Program, Uniformed Services University of the Health Sciences, Bethesda, MD, 20814, USA
[g]School of Immunity and Infection, University of Birmingham, Birmingham, UK
[h]Academic Department of Military Medicine, Royal Centre for Defense Medicine, Medical Directorate, Joint Medical Command, Birmingham Research Park, B15 2SQ Birmingham, UK
[i]Naval Medical Research Unit-6, Lima, Peru
[j]Center for Vaccines and Immunology, University of Georgia at Athens, College of Veterinary Medicine, Athens, Georgia, 30602, USA

SUPPLEMENTARY TABLES (EXTERNAL FILE):

Supplementary Table 1: Differential *Campylobacter jejuni* CG8421 Gene Expression Between Infected Human Feces and Laboratory Grown Controls.

Supplementary Table 2: Previously Published Chick Cecum RNA-seq Data with strain 81176 used in this study.

Supplementary Table 3: Variant Frequency per Sample: The frequency of each variant per sample population.

Supplementary Table 4: List of *C. jejuni* CG8421 genome variants selected for during human infection.

Supplementary Table 5: Frequency of genome variants called (listed in Sup. Table 1D) per sample.

Supplementary Table 6: Structural variant analysis detection of the 1416012..1416020 TTAAATTTT 9 nucleotide deletion in the CG8421 homolog of transcription factor CJJ81176_1483

Supplementary Table 7: No Variants Are Associated with Isolates from Rifaximin Prophylactically Treated Patients compared to no treatment patients, via Fisher's Exact Test with one-sided p-value.

Supplementary Table 8: No Variants Are Associated with Isolates Harvested on Days Patients had Severe Disease Symptoms Compared to Isolates from Days of Typical Symptomatic Disease, via Fisher's Exact Test with one-sided p-value. Primary infection isolates only.

Supplementary Table 9: No Variants Are Associated with Late Primary Infeciton Isolates (Last 3 Days of Infection) when Compared to Early Infection (First Three Days of Infection), via Fisher's Exact Test with one-sided p-value. Primary infection isolates only.

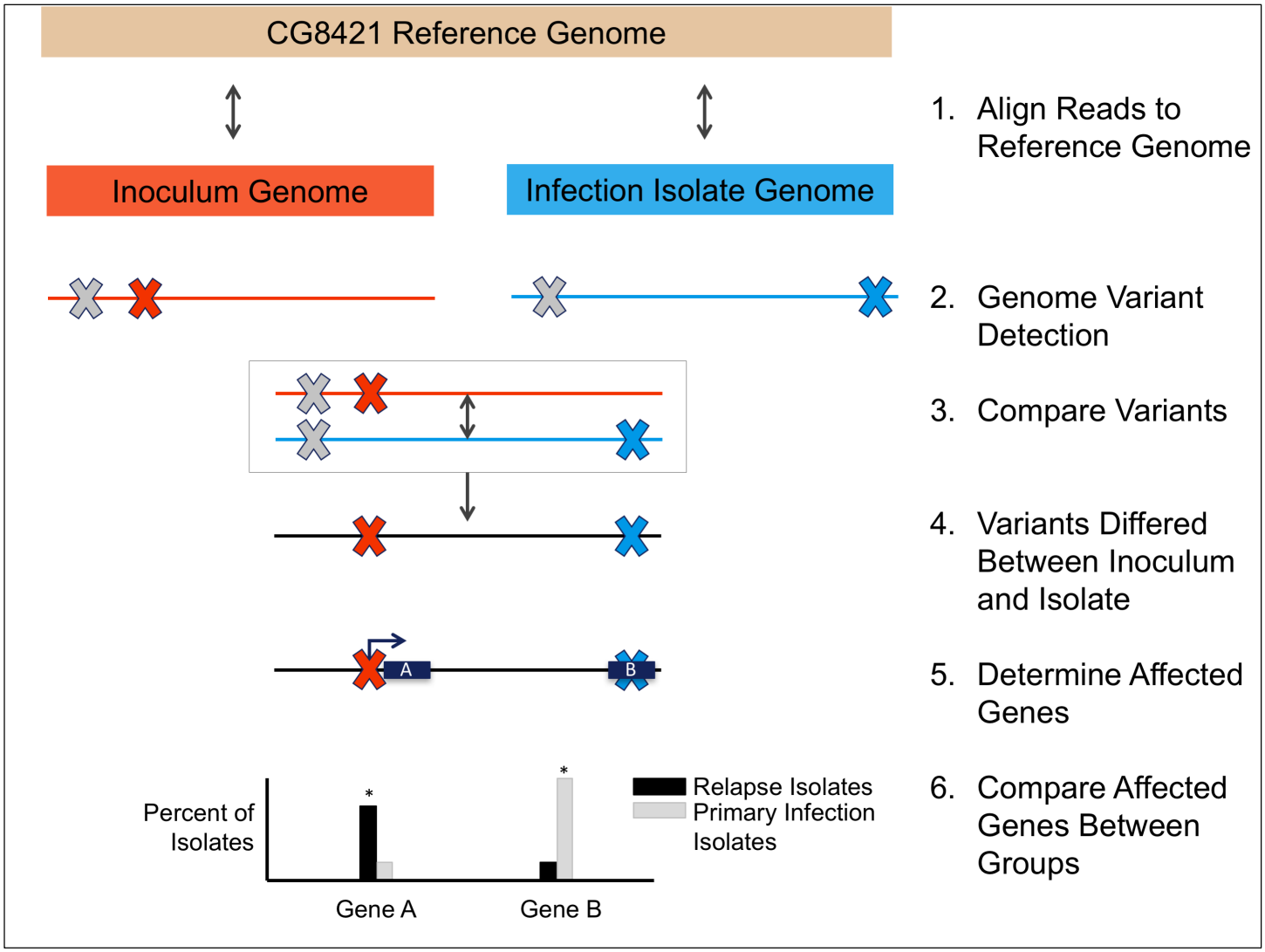
Supplementary Table 10: Fisher Exact Test with FDR corrected one-sided p-value for variable genes association with relapse infection isolates when compared to primary infection isolates.

Supplementary Table 11: Fisher Exact Test with FDR corrected one-sided p-value for variable genes association with primary infection isolates when compared to relapse infection isolates.

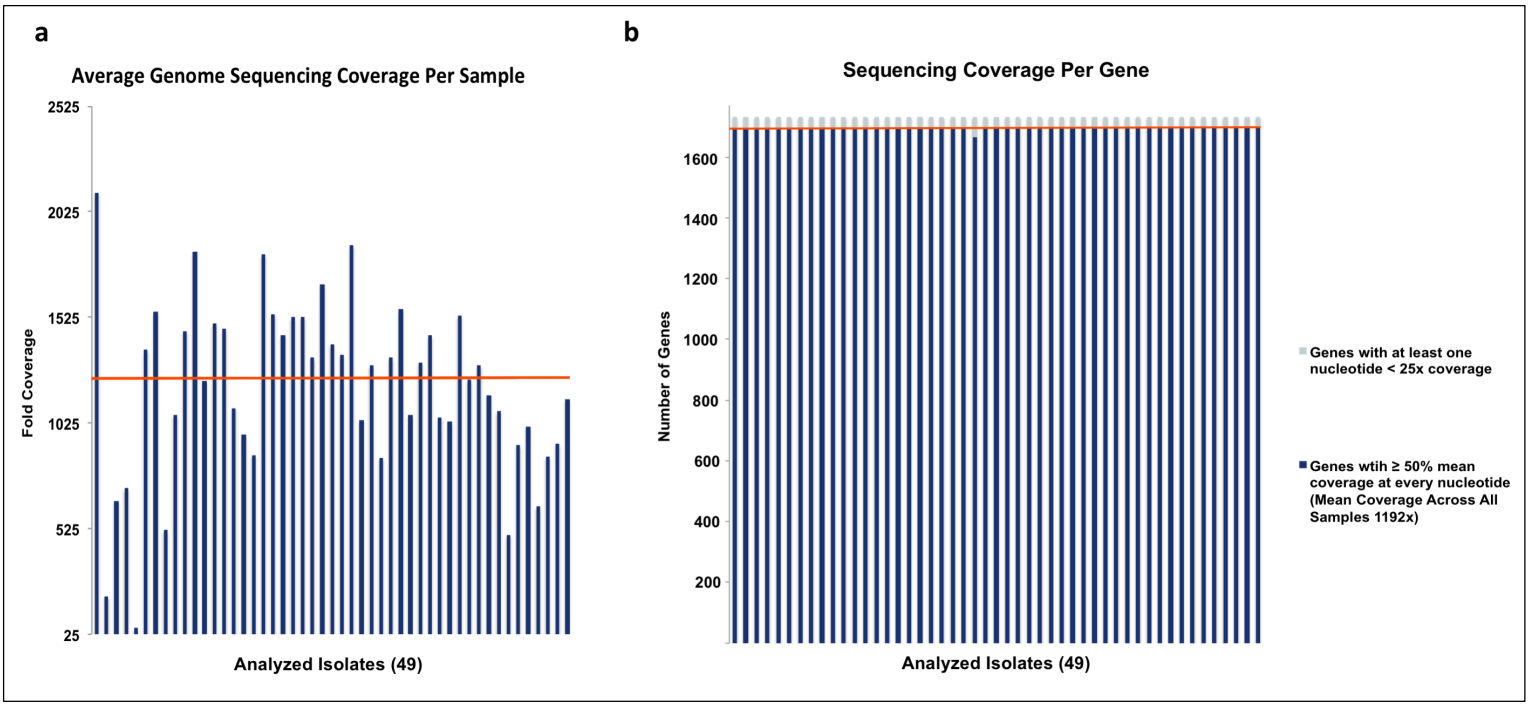
Supplementary Table 12: Genomic Variants Called in the *C. jeuni* CG8421 Inoculum Given to *Aotus* New World Primates.

Supplementary Table 13: Genomic Variants Called in *C. jeuni* CG8421 Infection Samples from Aotus New World Primates.

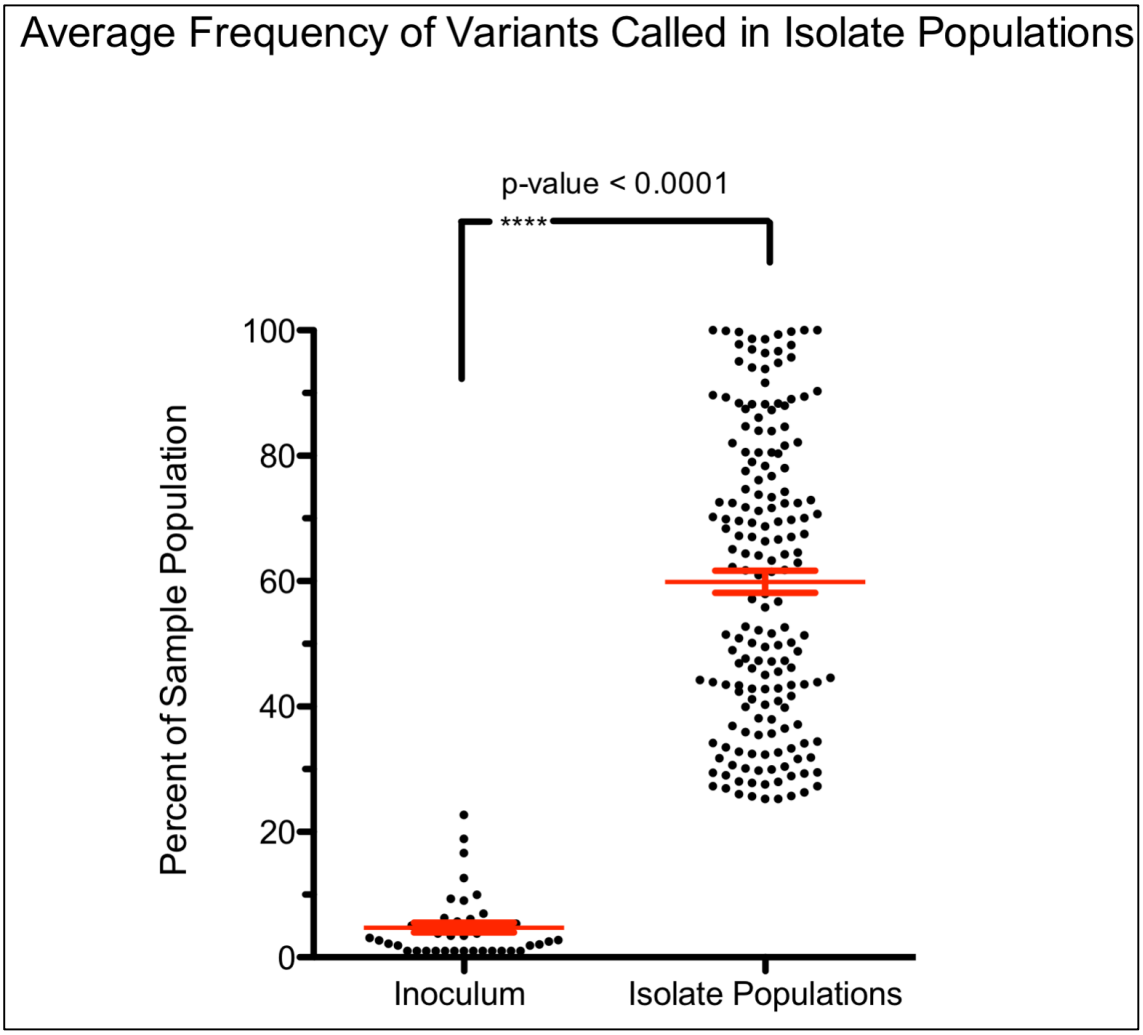
SUPPLEMENTARY FIGURES:



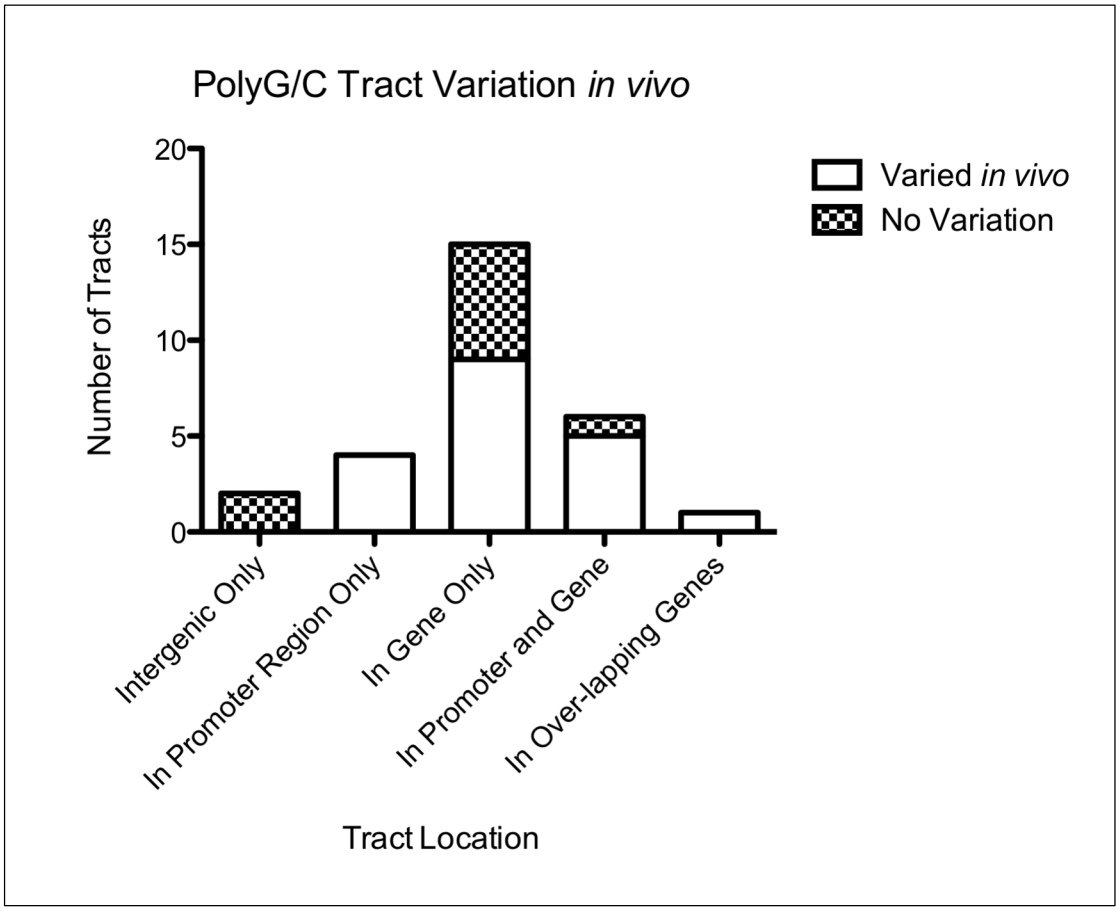
Supplementary Figure 1: Schematic of Genome Variant Analysis
The genomes of the *C. jejuni* CG8421 inoculum and infection isolate population genomes were compared to the published reference genome. Genome variants that occurred in at least 1 percent of any sample population are listed in Sup Table 3. However, to identify influential variants we used a 25 percent frequency cutoff as described in the methods and paper text, and those variants are listed in Sup. Table 4. The variant calls between the inoculum genome and the infection isolate populations where then compared, and variant calls that differed between the inoculum and isolate populations are considered genomic variants selected for *in vivo*. Using the location of each variant in the annotated genome, we determined the genes (or gene promoters) affected by these variants, and then used Fisher's Exact Test with FDR-corrected p-values to determine if particular genes were more likely to genetically vary in *in vivo* between volunteer isolate groups.



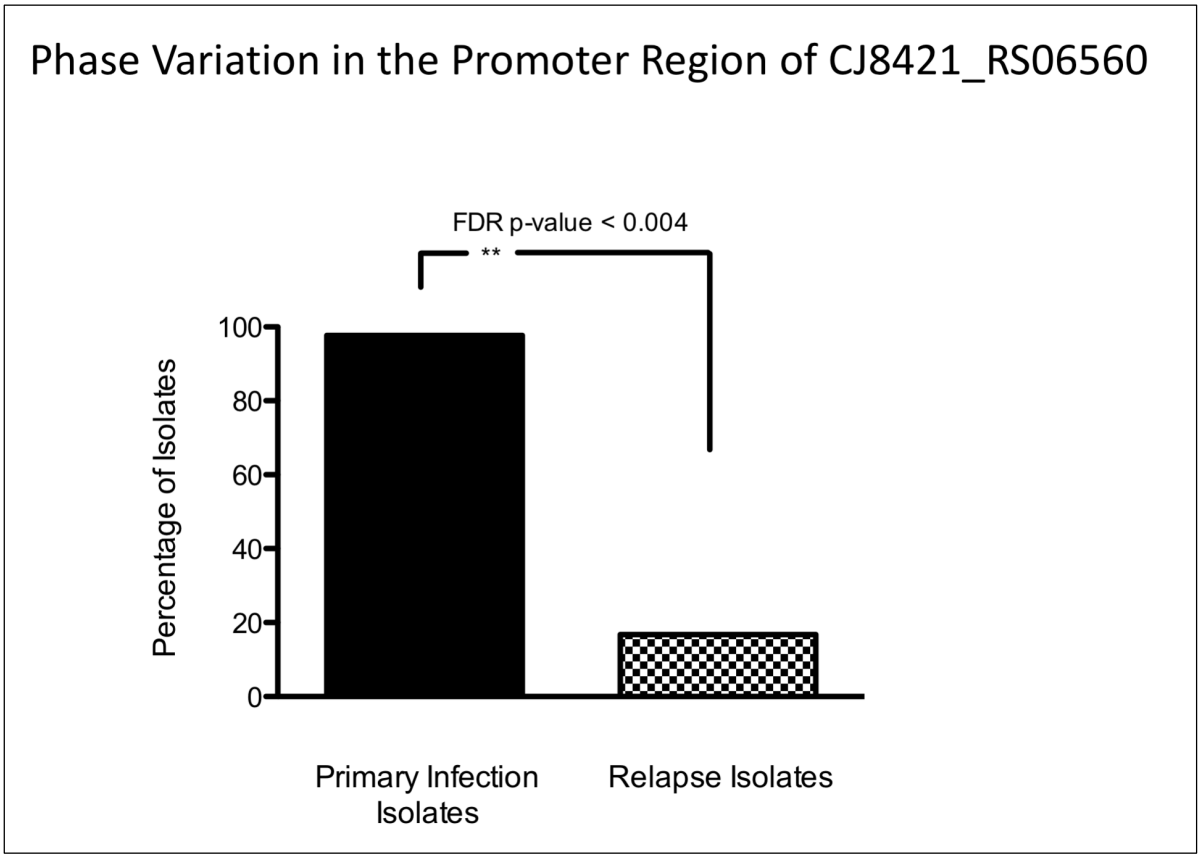
Supplementary Figure 2: Infectious Isolate Populations Genome Sequencing Coverage
a, Sequencing depth across all 49 infection isolate populations. The average fold coverage across samples was 1192x (orange line). **b**, Coverage of open reading frames across samples. On average (orange line) 98 percent of annotated genes had at least 25 fold coverage at every base across samples.



Supplementary Figure 3: Average Frequency of Variants Called in Isolate Populations
Genomic variants were initially called if they occurred in at least 1 percent of a sample population (49 isolate samples and 1 inoculum), resulting in over 600 variants detected across all samples, and that data is found in Sup. Table 3. To focus on variants that had both a large change in frequency during infection and represented a large portion of the population, we used a 25 percent variant calling cutoff. This raised cutoff enriched for variants that had, on average, a large fold change in frequency between the inoculum and an infection isolate populations (~12 fold) while also representing a majority of the infection population (~60%). Detailed information about each variant call, per sample, is provided in Sup.Table 5 and is represented here by a dot plot. An unpaired two-tailed T-test was used to confirm a significant difference in the mean frequency between the pre (single inoculum) and post infection (isolate populations) variant frequency calls. Bars represent the mean variant frequency in the sample(s) and standard error of the mean.

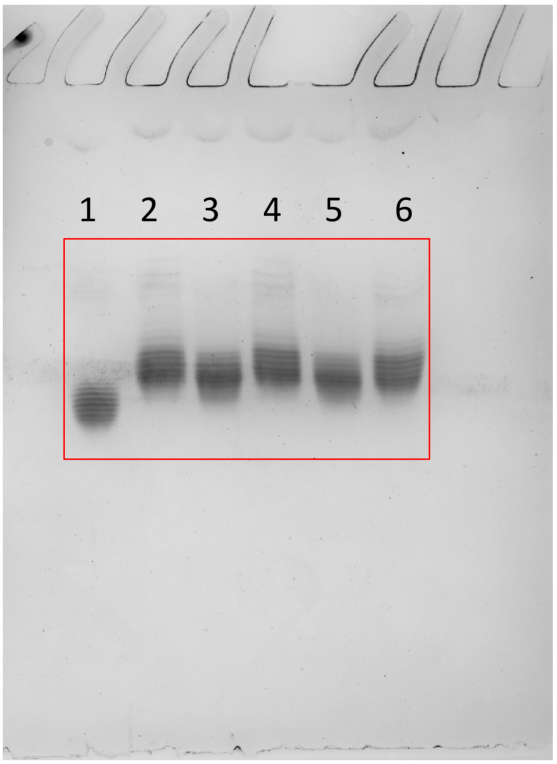


Supplementary Figure 4: Poly G/C Tract Locations and Variations in Vivo
The location of all 28 poly G/C tracts in the CG8421 genome are presented by their relationship to annotated open reading frames in a stacked bar graph. The number of tracts at each location type that varied in at least one sample population are noted.



Supplementary Figure 5: Genome Variant Associated Only With Primary Infection Isolate Populations
A phase variant in the promoter region of the uncharacterized putative transferase gene CJ8421_RS06560, present in the inoculum, was resoundingly selected against during primary infection of humans. The percentage of isolates shown denotes the frequency that the wildtype (reference sequence) variant was called in those samples. Fisher's exact test identified this variant as statistically significantly associated with primary infection isolates (43 samples) compared to relapse infection isolates (6 samples) with a false-discovery rate (FDR) correct p-value (one-sided).

a

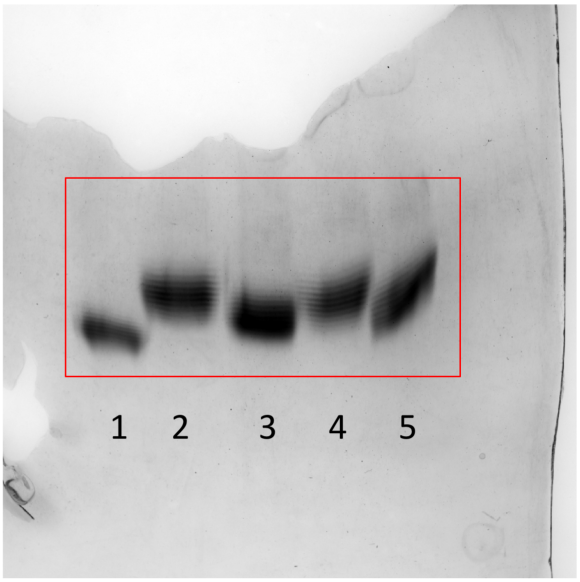


Lane	Purified Flagella of Strain	On/Off Status of CipA in Strain
1	pseA	NA
2	WT	Off
3	cipAR	On
4	WT	Off
5	cipAR	On
6	WT	Off

b

Fisher's exact test ** P-value 0.0045	Disease state associated (either primary or relapse infection association)	No Associations	Total
Flagellin modification genes	6	1	7
All other *in vivo* variable genes	5	18	23
Total	11	19	30

c



Lane	Purified Flagella of Strain	On/Off Status of Cj0617/618 of Strain
1	pseA	NA
2	WT	Off
3	Cj0617/618R	On
4	WT	Off
5	Cj0617/618R	On

Supplementary Figure 6: Genetic Variation in Flagella Modification Genes is Strongly Associated with the Human Disease State

a, A functional *cipA* gene results in modification of the flagella as shown by an isoelectric focusing gel (pH 3-5) of purified flagellins from *C. jejuni* strain 81-176. The *cipA* gene was phase varied "off" in the human inoculum, but was selected to be turned "on" in every human relapse infection isolate population. Therefore, a functional *cipA* gene is strongly associated with relapsing human infections. We therefore aimed to identify the biochemical role of *cipA* in *Campylobacter*. *cipA* contains a domain of unknown function found in putative glycosyl transferase genes as predicted by BLAST search. We therefore sought to identify possible targets of *cipA* activity in *Campylobacter*. We investigated potential flagella modification by preforming an isoelectric focusing gel mobility shift assay on purified flagella from *cipA* "on" and "off" genetic backgrounds. Flagellin modification activity changes the electrical charge of the flagellin and therefore results in a mobility shift in the delicate isoelectric focusing gel based on isolectric point and not molecular weight. Therefore, in lieu of standard ladders, purified flagella from characterized intact and mutant flagella modification strains are used as simultaneous markers and reference controls to interpret modification-induced shifts on IEF gels. Genetic manipulation cannot be performed in strain CG8421 so strain 81-176 was used for these analyses. Lanes contain flagellin from: *pseA*, a mutant lacking the acetamidino form of pseudaminic acid, serving as a control demonstrating a more negatively charged modification state compared to wildtype; WT, wildtype 81-176 flagellin in which *pseA* is intact, however *cipA* is "off" similar to the human inoculum; *cipA*R, a mutant in which

the phase variable homopolymeric G tract in *cipA* was repaired to produce a full-length *cipA* open reading frame in an otherwise WT background. Indeed, a mobility shift of flagella from the *cipA* "on" repaired strain indicates a more negative charge and therefore *cipA* contributes to flagellin modification. Samples were repeated in multiple lanes for clarity of the phenotype and this gel is representative of at least 5 mobility shift experiments. **b,** Contingency table of variable genes and their association with either primary or relapse infections (disease state). A Fisher's exact test shows variation in flagella modification genes (including *O*-link glycosylation locus genes as described in Sup. Table 4) as a group is strongly associated with either primary or relapse infection isolate populations (one-sided p-value). The 8 variable genes related to flagellin glycosylation[1] are Cj11168 homologs Cj0617/618, Cj1295, Cj1306c, Cj1325, Cj1342, Cj1333, Cj1321 (CJ8421_RS06560) (Sup. Table 4), and *cipA* as identified in Sup. Figure 6a. **c,** Isoelectric focusing gel of flagellins from *C. jejuni* strain 81-176 and 81-176 mutants. Cj0617 and Cj0618 are phase variable genes that can be translated as a unified open reading frame (Cj0617/618). Although Cj0617 is the representative gene of the 617 family of *O*-linked glycosylation genes[2], and Cj0618 contributes to motility[3], Cj0617 and Cj0618 are the only genes included in Sup. Fig. 6b, besides *cipA*, that are not located in the *O*-linked glycosylation locus and there are no reports showing the affect of Cj0617/618 on flagellin modification to our knowledge. We therefore sought to verify that Cj0617/618 contributes to flagella modification by preforming an isoelectric focusing gel mobility shift assay on purified flagella. Flagellin modification activity changes the electrical charge of the flagellin and therefore results in a mobility shift in the isoelectric focusing gel based on isolectric point and not molecular weight. Therefore, in lieu of standard ladders, purified flagella from characterized intact and mutant flagella modification strains are used as simultaneous markers and reference controls to interpret modification-induced shifts on IEF gels. Genetic manipulation cannot be performed in strain CG8421 so strain 81-176 was used for these analyses. Lanes contain flagellin from: *pseA*, a mutant lacking the acetamidino form of pseudaminic acid, serving as a control demonstrating a more negatively charged modification state compared to wildtype; WT, wildtype 81-176 flagellin in which *pseA* is intact, however Cj0617/618 is not translated as a continuous open reading frame; Cj0617/618R, a mutant in which the phase variable homopolymeric G tract was repaired to produce a continuous open reading frame of genes Cj0617 and Cj0618 in an otherwise WT background. Indeed, a mobility shift of flagella from the Cj0617/618 repaired strain indicates a more negative charge and therefore Cj0617/618 contributes to a flagellin modification phenotype. This mobility shift is representative of at least 5 mobility shift experiments. Cj0617/618 likely affects pseudaminic acid modifications on flagellin as strain 81-176 lacks the alternative legionaminic acid pathway.

SUPPLEMENTARY REFERENCES

1. Poly, F. *et al.* Characterization of Two Campylobacter jejuni Strains for Use in Volunteer Experimental-Infection Studies. *Infect. Immun.* **76,** 5655–5667 (2008).
2. Parkhill, J. *et al.* The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences. *Nature* **403,** 665–668 (2000).
3. Hendrixson, D. R. & DiRita, V. J. Identification of Campylobacter jejuni genes involved in commensal colonization of the chick gastrointestinal tract. *Mol. Microbiol.* **52,** 471–484 (2004).