SUPPLEMENTARY MATERIALS

Supplementary Materials and Methods

16S rRNA gene sequencing and analyses.

16S rRNA gene sequencing methods were adapted from the methods developed for the NIH-Human Microbiome Project (*63*). Briefly, bacterial genomic DNA was extracted using MO BIO PowerSoil DNA Isolation Kit (MO BIO Laboratories). The 16S rDNA V4 region was amplified by PCR and sequenced in the MiSeq platform (Illumina) using the 2x250 bp paired-end protocol. The 16S rRNA gene pipeline data incorporates phylogenetic and alignment-based approaches to maximize data resolution. The read pairs were demultiplexed based on the unique molecular barcodes, and reads were merged using USEARCH v7.0.1090, allowing zero mismatches and a minimum overlap of 50 bases. Merged reads were trimmed at the first base with Q5. In addition, a quality filter was applied to the resulting merged reads and reads with > 0.05 expected errors were discarded. 16S rRNA gene sequences were clustered into Operational Taxonomic Units (OTUs) at a similarity cutoff value of 97% using the UPARSE algorithm. OTUs were mapped to an optimized version of the SILVA Database containing only the 16S V4 region to determine taxonomies. Abundances were recovered by mapping the demultiplexed reads to the UPARSE OTUs. A rarefied OTU table from the output files generated in the previous two steps was used for downstream analyses of α-diversity, β-diversity, and phylogenetic trends. The IgA-coating index (ICI) was calculated using the relative abundance for each individual taxa as ICI = (log(IgA+) – log(IgA-))/(log(IgA+) + log(IgA-)) as described previously (*23*). A pseudocount of 0.001 or 0.1% relative abundance (the limit of detection for our sequencing) was added to the fractional abundance of each taxa for this analysis.

Bacterial isolation.

Freshly sorted samples were added to PYG media under anaerobic conditions and cultured for 24-48 hours to allow for recovery. Following anaerobic recovery, mixed bacteria cultures were plated by limiting dilution on both selective (MacConkey) and non-selective (CBA) media. MacConkey media plates with isolated colony growth of

50-100 colonies were used to pick *E. coli* library isolates and confirmed by 16S rDNA sequence analysis. Random amplified polymorphic DNA-PCR (RAPD-PCR) was used to identify unique isolates. The presence of virulence-associated genes and phylotype were determined as previously reported (*33*). CUMT8 deletion mutants lacking the *pduC* gene were constructed using the lambda red recombineering system (*33*). The deletion of *pduC* was confirmed by PCR. To complement the *pduC* mutation, the complete *pduC* coding region was amplified by PCR from CUMT8, restriction digested and cloned into the SacI and XbaI sites of plasmid pBAD33. This plasmid construct was electroporated into CUMT8:ΔpduC and selected for resistance to chloramphenicol. Constructs were confirmed by PCR and sequencing using primers flanking the insertion sites.

CD-SpA *E. coli* 2A genome sequence.

Genome sequencing was performed using the Pacific Biosciences RSII sequencing platform. Long reads were assembled *de novo* into 3 contigs (main chromosome and 2 plasmids) using SMRT analysis (*64*). Gene prediction and annotation were carried out using RAST (*65*), incorporating the Glimmer algorithm (*66*). GCView server was used for genome visualization and comparison (*67*).

DSS-induced colitis.

Mice were treated with 2% dextran sulfate sodium (Affymetrix) in drinking water *ad libitum* for 7 days to induce colitis starting at 10 days after colonization. Mice were monitored for weight loss and disease activity and tissues were collected 2 days after DSS removal. For some experiments, functional grade anti-IL23p19 (G23-8) antibody or isotype control rat IgG1 was injected intravenously at 100 μg per mouse on day 1 and day 4 after initiating DSS treatment.

Histology.

Proximal colon sections were fixed in 4% paraformaldehyde, embedded in paraffin, sectioned (5 μm), and then stained with hematoxylin and eosin for histological examination. Tissue slides were evaluated in a blinded fashion using the scoring system established for chemically-induced colitis (*68*).

Fluorescent *In Situ* Hybridization staining.

Methacarn-fixed paraffin-embedded histological sections of terminal ileum were mounted on charged slides and evaluated by FISH using *Escherichia*/*Shigella* (*E. coli*, 16SrRNA) probes as previously described (*30*). Sections were examined with an Olympus BX51 epifluorescence microscope. Images were captured with a DP-70 camera and DP-Controller software with Image files processed using DP-Manager (Olympus America, Center Valley, PA, USA).

Cellular isolation and intracellular cytokine staining.

Lamina propria mononuclear cells were isolated from colonic tissue as previously described (*69*). LIVE/DEAD fixable aqua dead cell stain kit (Molecular Probes) was used to exclude dead cells. For transcription factor analysis, cells were stained with anti-CD3-E780 (eBiosciences 17A2) and anti-CD4-AF700 (eBiosciences GK1.5) before fixation and permeabilization with Cytoperm/Cytofix as per manufacturer's instructions (eBiosciences) for intracellular staining against anti-Foxp3-E450 (eBiosciences FJK-16s) and anti-RORγt-PE (eBiosciences B2D). For cytokine detection, cells were stimulated with phorbol myristate acetate (PMA) and ionomycin with BD GolgiPlug for 4 hours. Following surface-marker staining with anti-CD4-AF700 (eBiosciences GK1.5) and anti-TCRβ-E780 (eBiosciences H57-597), cells were prepared as per manufacturer's instruction with Cytoperm/Cytofix (BD Biosciences) for intracellular cytokine evaluation of IL-17A (eBiosciences 17B7), IL-22 (eBiosciences IL22JOP) and IFNγ (eBiosciences XMG1.2). For the analysis of K/BxN mice, mononuclear cells from spleen, Peyer's Patches and small intestine lamina propria were prepared as described previously (*70*). For intracellular staining, cells were stained with anti-CD19 (BioLegend 6D5), anti-TCRβ (BioLgend H57-597), anti-Fas (BD Biosciences Jo2) and peanut agglutinin (PNA) (Vector Laboratories) before fixation with Cytoperm/Cytofix for intracellular staining against IgG1 (BioLegend RMG1-1) and IgA (BioLegend RMA-1). For cytokine detection, cells were stimulated with PMA and ionomycin with BD GolgiPlug for 4 hours, and stained with anti-CD4 (BioLegend RM4-5), anti-TCRβ, anti-CD45 (BioLegend 30-F11). Cells were then prepared with Cytoperm/Cytofix (BD Biosciences) for intracellular cytokine staining for IL-17A (BioLegend TC11-18H10.1). Data

acquisition was computed with BD LSRFortessa and LSRII flow cytometers and analysis performed with FlowJo software.

qPCR.

Epithelial cells were manually scraped off the luminal surface of the terminal ileum and placed into RLT buffer (Qiagen). Colonic Lamina Propria Mononuclear Cells (LPMC) were isolated and stored in RLT buffer (Qiagen). RNA was extracted and purified using RNeasy Plus Micro Kit (Qiagen) and quantified by Nanodrop prior to reverse transcription with qScript cDNA Supermix kit (Quanta Biosciences). QPCR was performed on an Applied BioSciences Quant Studio 6 using PerfeCTa SYBR Green Fast mix, Low ROX (Quanta Biosciences) according to manufacturer's protocol. Relative levels of *SAA1 (*SAA1-F; 5'-CATTTGTTCACGAGGCTTTCC-3', SAA1-R; 5'-GTTTTTCCAGTTAGCTTCCTTCATGT-3') and *Retnlb* (Retnlb-F; 5'-AGCTCTCAGTCGTCAAGAGCCTAA-3', Retnlb-R; 5'-CACAAGCACATCCAGTGACAACCA-3') in epithelial cells were determined by ΔΔCT using expression of housekeeping gene HPRT. Relative levels of *Il10* (IL10-F; 5'-GAGAGCTGCAGGGCCCTTTGC-3', IL10-R; 5'-CTCCCTGGTTTCTCTTCCCAAGACC-3') were measured in colonic LPMC and ΔΔCT determined using expression of housekeeping gene HPRT.

ELISPOT.

Following ACK lysis, splenic CD4+ T cells were magnetically sorted according to manufacturer's instructions (Miltenyi Biotec). 96-well PVDF multiscreen plates (Millipore) were coated with anti-mouse IL-17 (Affymetrix) as per manufacturer's instructions. A total of 2×10^5 spleen-derived CD4+ T cells were added to the wells and stimulated with T-cell depleted splenocytes loaded with cecal homogenates from monocolonized mice, anti-CD3 and anti-CD28 (positive control), or cecal homogenates from germ-free mice (negative control). Cells were incubated at 37°C for 48 h and plates were stained with biotinylated anti-mouse IL-17 (Affymetrix) followed by streptavidin-horseradish peroxidase (HRP) (Affymetrix). Enumeration of individual colored spots was performed using a dissecting microscope (Leica).

Seroreactivity.

Serum (undiluted, 1:100, 1:333 and 1:1000 dilutions) was incubated with 10^6 bacterial cells (human CD-SpA 2A or *B. vulgatus* isolates) for 30 min on ice. Following 2 washing steps, samples were incubated with PE-labeled anti-human IgG or isotype (eBiosciences) for 30 min. Bacterial cells were then washed to remove excess antibody and fixed in 4% paraformaldehyde. Data was acquired with a BD LSRFortessa flow cytometer (BD Biosciences).

Human serum cytokine staining.

Cytokine concentrations in human serum were determined using a LEGENDplex human Th cytokine 13-plex panel kit (BioLegend) according to the manufacturer's instructions. Data was acquired with a BD LSRFortessa flow cytometer (BD Biosciences) and analyzed using the BioLegend's LEGENDplex Data Analysis Software.

Supplementary Figures

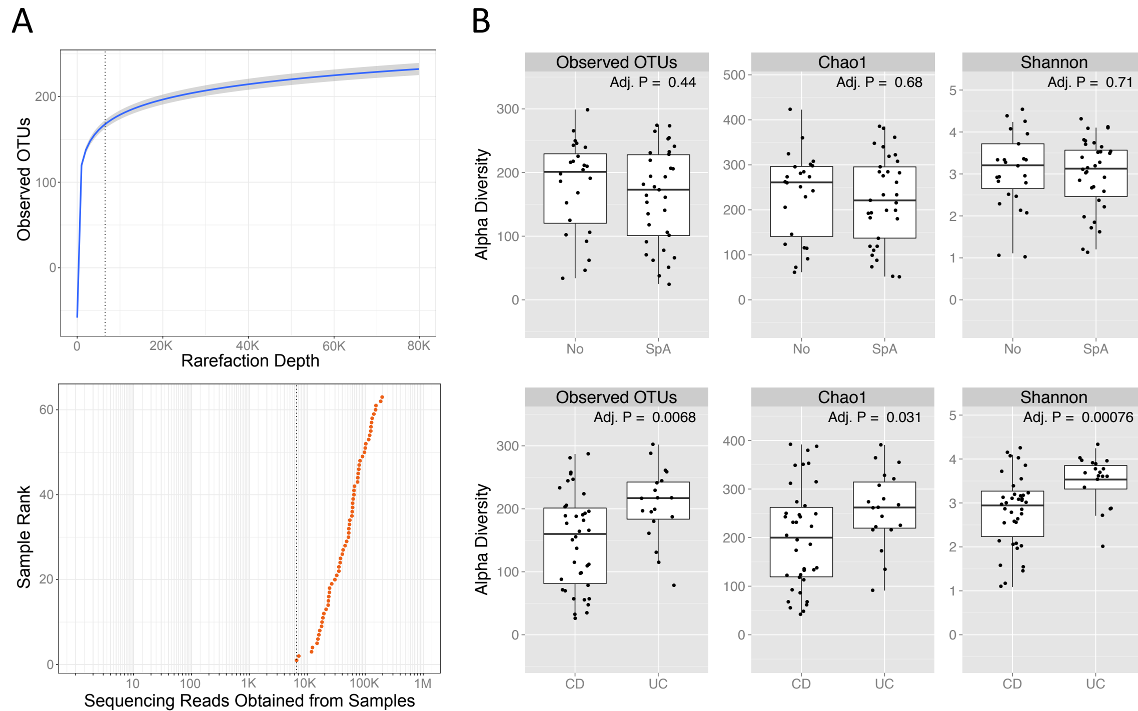


Fig. S1. Alpha and beta diversity metrics of 16S rRNA compositional data. A. Rarefaction plots are shown for 16S rRNA sequencing data. 6500 reads / sample cutoff is indicated. **B.** Alpha diversity for 16S rRNA compositional data grouped by no spondyloarthritis (No) or spondyloarthritis (SpA) in the upper panel and IBD subtype CD (N=40) or UC (N=19) in the bottom panel. Observed OTUs, Chao1, and Shannon index metrics are shown. Adjusted p-values are indicated, Mann-Whitney.

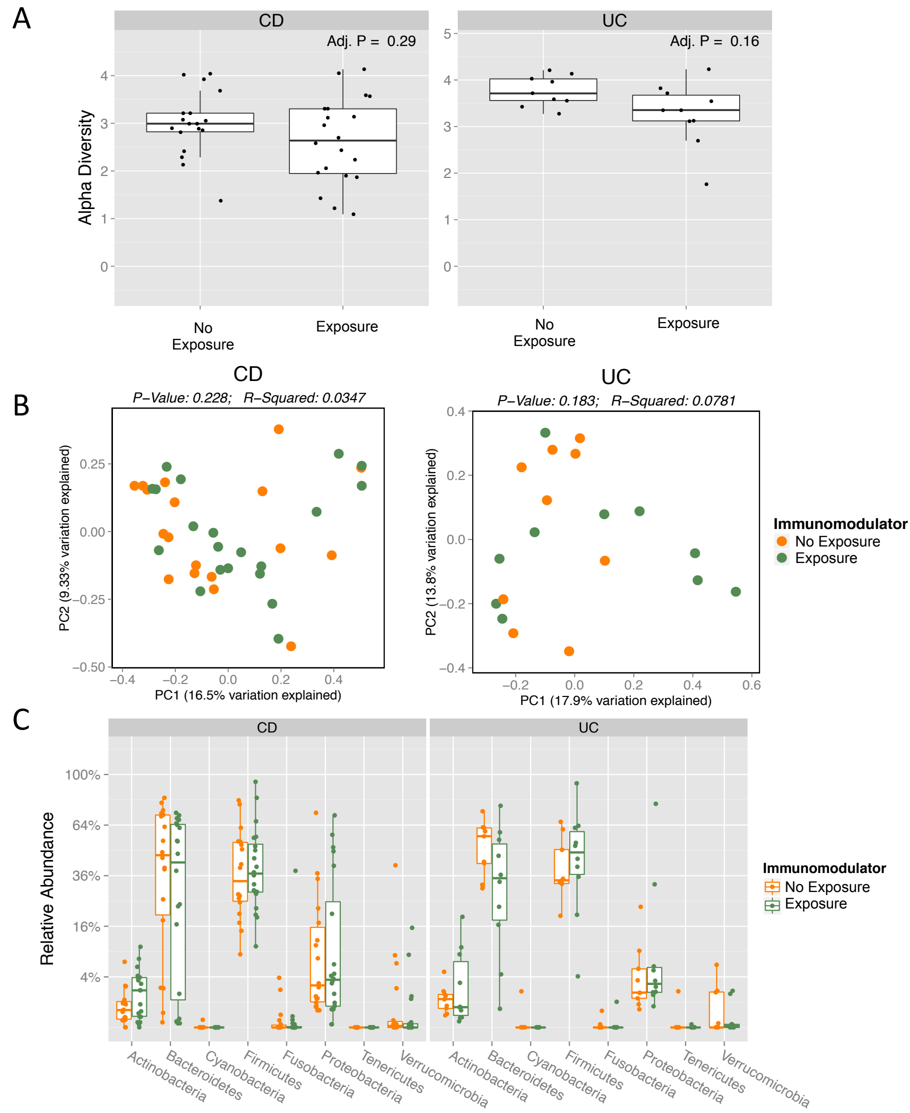


Fig. S2. Alpha and beta diversity metrics stratified by immunomodulator exposure. A. Alpha diversity for 16S rRNA compositional data grouped by immunomodulator medication exposure and faceted by IBD subtype CD or UC. Shannon index metrics are shown. Adjusted p-values are indicated, Mann-Whitney. **B.** Beta diversity Bray Curtis ordination PCoA plot faceted by IBD subtype. No medication exposure: orange. Immunomodulator medication exposure: green. P-value, Monte Carlo permutation test. **C.** Relative abundance of the indicated phyla are shown grouped by medication exposure. Boxplots reflect 25 and 75% confidence interval and horizontal line reflects median. Outliers are indicated by colored dots. For all panels, CD No exposure =19, CD exposure = 21; UC No exposure = 9, UC exposure = 10.

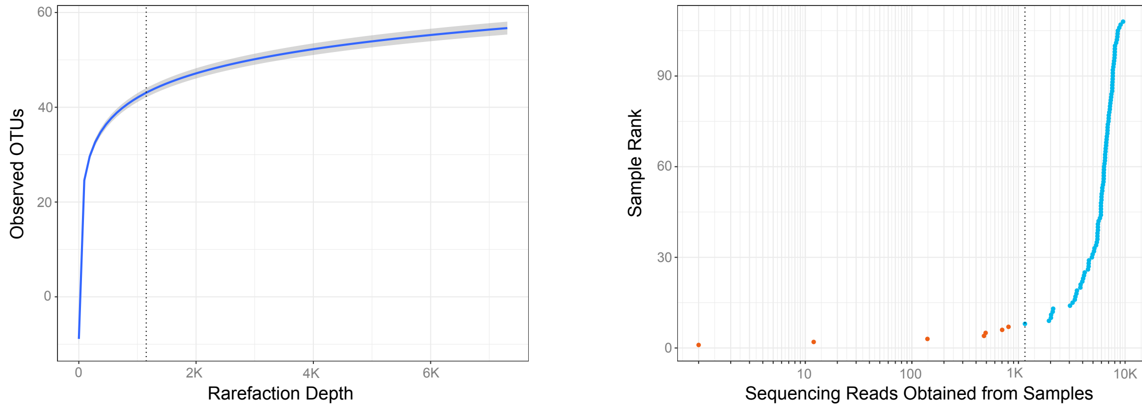


Fig. S3. 16S rRNA IgA-seq sample rarefaction. Rarefaction plots are shown for 16S rRNA sequencing data. 1100 reads/sample cutoff is indicated.

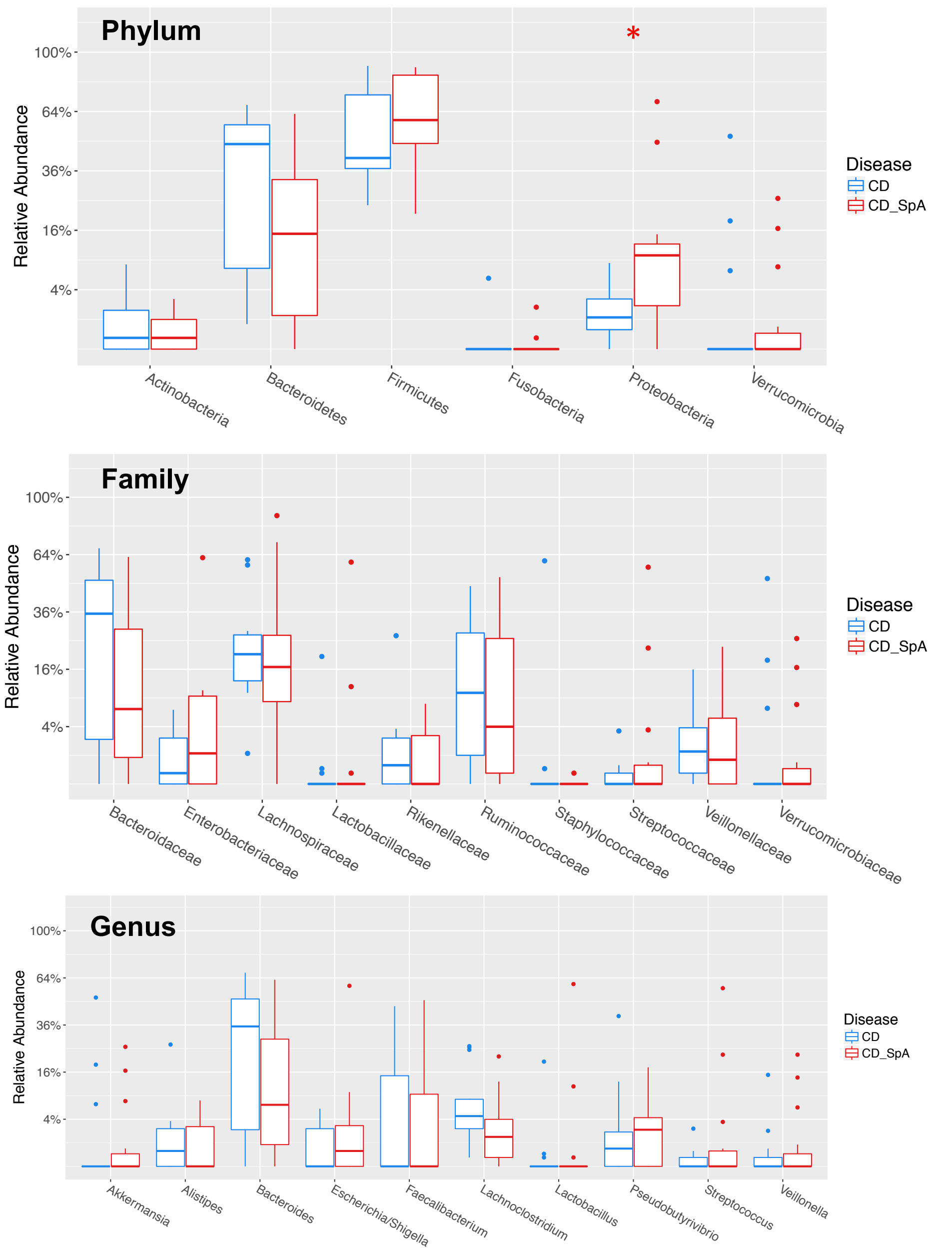
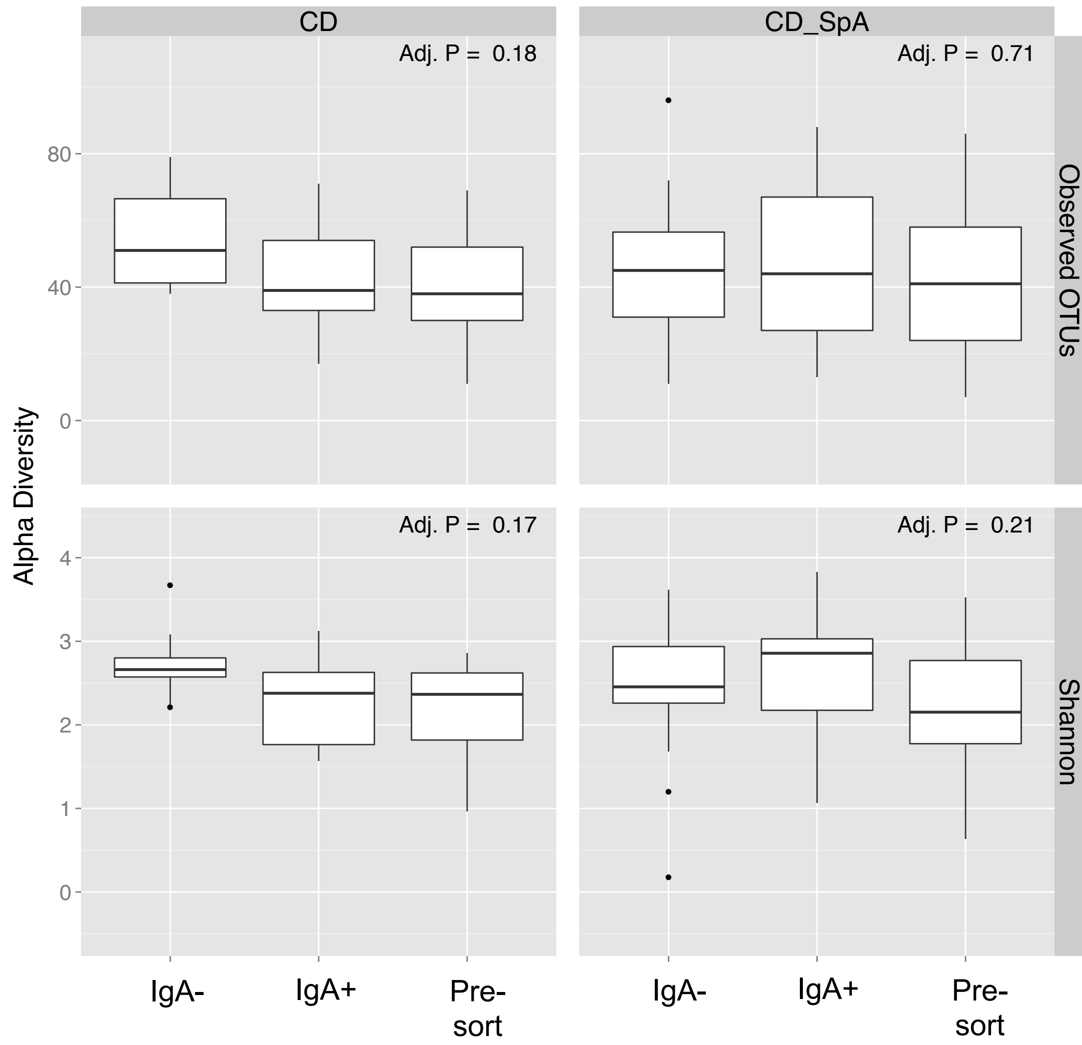


Fig. S4. Relative abundance by phylum, family and genus in the pre-sort fraction. Relative abundance of the indicated phylum (top plot), family (middle plot), or genus (bottom plot) is shown grouped by CD (blue) or CD associated SpA (CD-SpA, red). Boxplots reflect 25 and 75% confidence interval and horizontal line reflects median. Outliers are indicated by colored dots. P-values < 0.05 are indicated by an asterisk, Mann-Whitney. For all panels, CD N=10, CD-SpA N=13.


A
CD
CD_SpA
Adj. P = 0.18
Adj. P = 0.71
Adj. P = 0.17
Adj. P = 0.21
Observed OTUs
Shannon
Alpha Diversity
0
40
80
0
1
2
3
4
IgA-
IgA+
Pre-sort

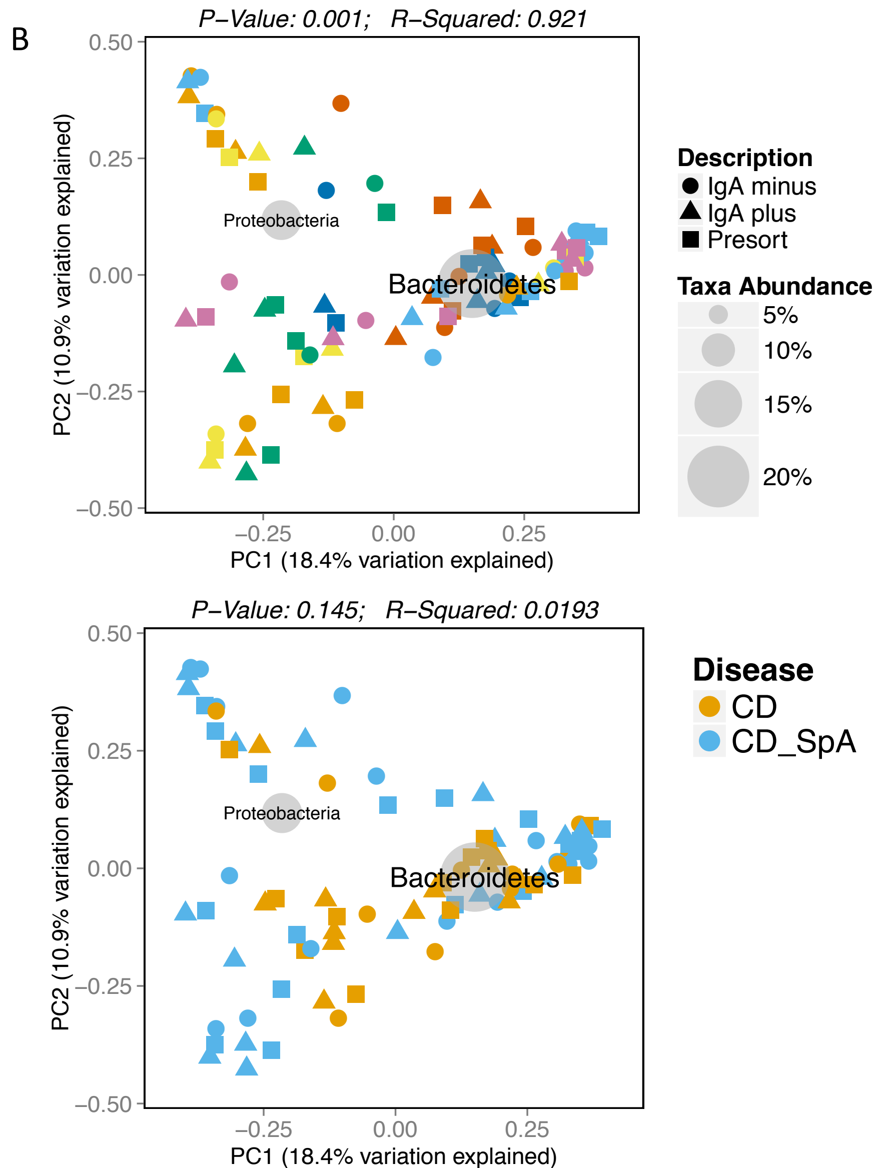


Fig. S5. Alpha and beta diversity of 16S rRNA IgA-seq data. A. Alpha diversity for 16S rRNA IgA-seq data grouped by sort population (IgA+, IgA- or pre-sort) and faceted by CD or CD-associated spondyloarthritis (CD-SpA). Observed OTUs and Shannon index metrics are shown. Boxplots reflect 25 and 75% confidence interval and horizontal line reflects median. Outliers are indicated by black dots. Adjusted p-values are indicated, Mann-Whitney. **B.** Beta diversity Bray Curtis ordination PCoA plot grouped by patient ID represented by a unique color (upper plot) or spondyloarthritis (SpA, lower plot). Taxa abundance is represented by gray circles. P-value, Monte Carlo permutation test. For all panels, CD N=10, CD-SpA N=13.

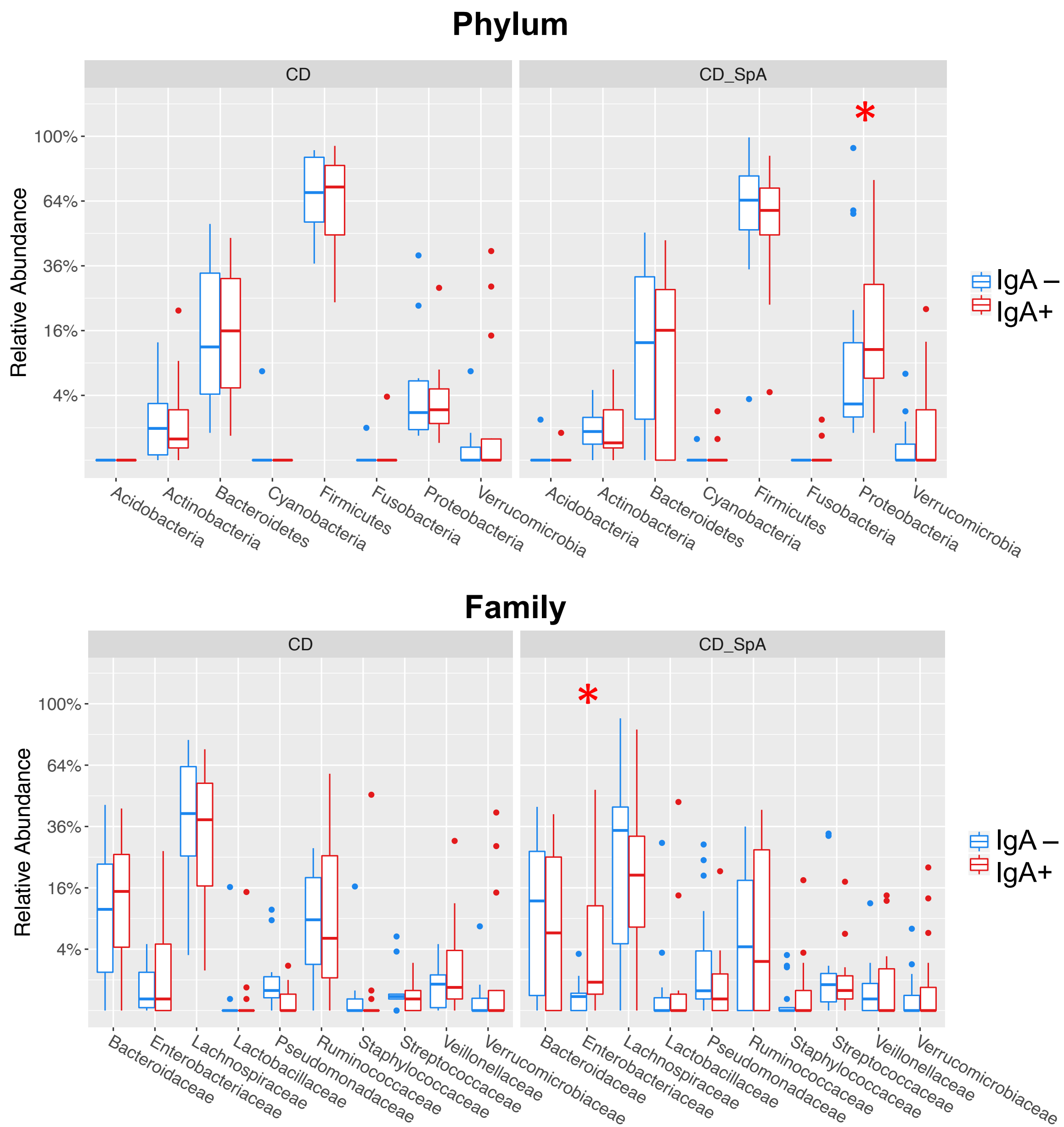


Fig. S6. Relative abundance by phylum and family in the sorted IgA+ and IgA- fractions. Relative abundance of the indicated phylum (top plot) and family (bottom plot) are shown for IgA- (blue) or IgA+ (red) fractions in CD or CD associated SpA (CD-SpA). Boxplots reflect 25 and 75% confidence interval and horizontal line reflects geometric median. Outliers are indicated by colored dots. P-values < 0.05 are indicated, Mann-Whitney. For all panels, CD N=10, CD-SpA N=13.

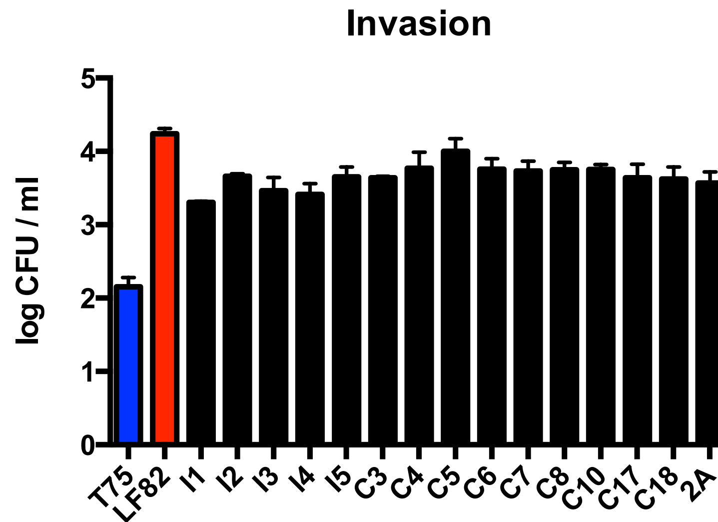


Fig. S7. CD-SpA-derived, IgA-coated *E. coli* isolates are AIEC. Invasion of Caco-2 cells by non-AIEC *E. coli* T75 (blue), AIEC LF82 (red), and CD-SpA *E. coli* isolates (black) was assessed by gentamicin protection assay. The log of mean colony forming units (cfu) / ml is presented. SEM is shown for each group. All samples were run in triplicate.

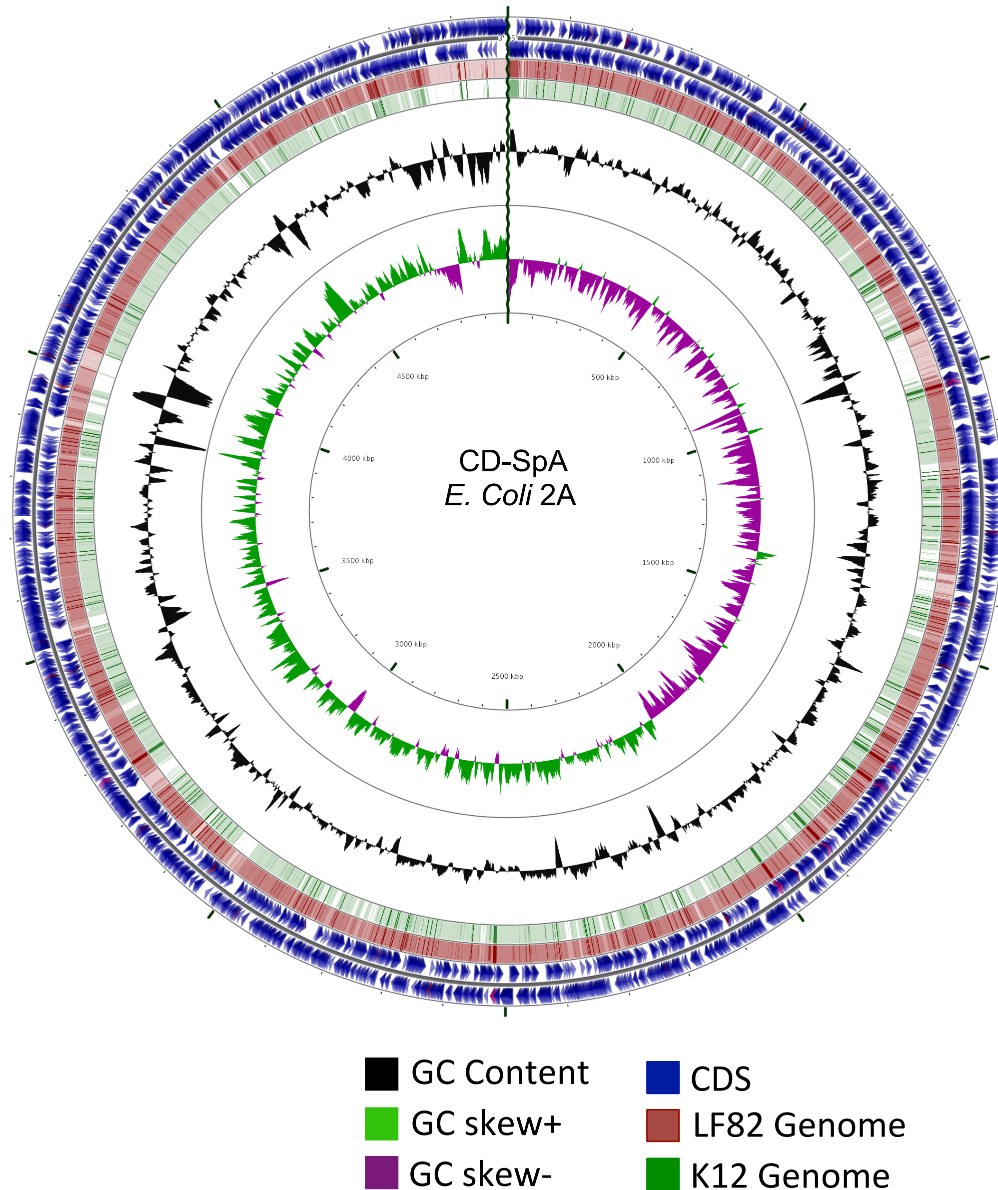


Fig. S8. Circular genome map of CD-SpA-derived *E. coli* isolate 2A. The genome of CD-SpA AIEC isolate 2A was sequenced using PacBio and assembled *de novo* into 3 contigs (main chromosome and 2 plasmids). The complete circular genome was 4,991,385 bp in length with a GC content of 50.6% and encoded 4,845 ORFs. The genome of CD-SpA 2A was annotated using RAST and visualized using the GCView Server. Comparative genomic analyses was performed for CD-SpA 2A (blue, outer circle) using AIEC LF82 (red) and non-AIEC *E. coli* K12 (green) as reference genomes.

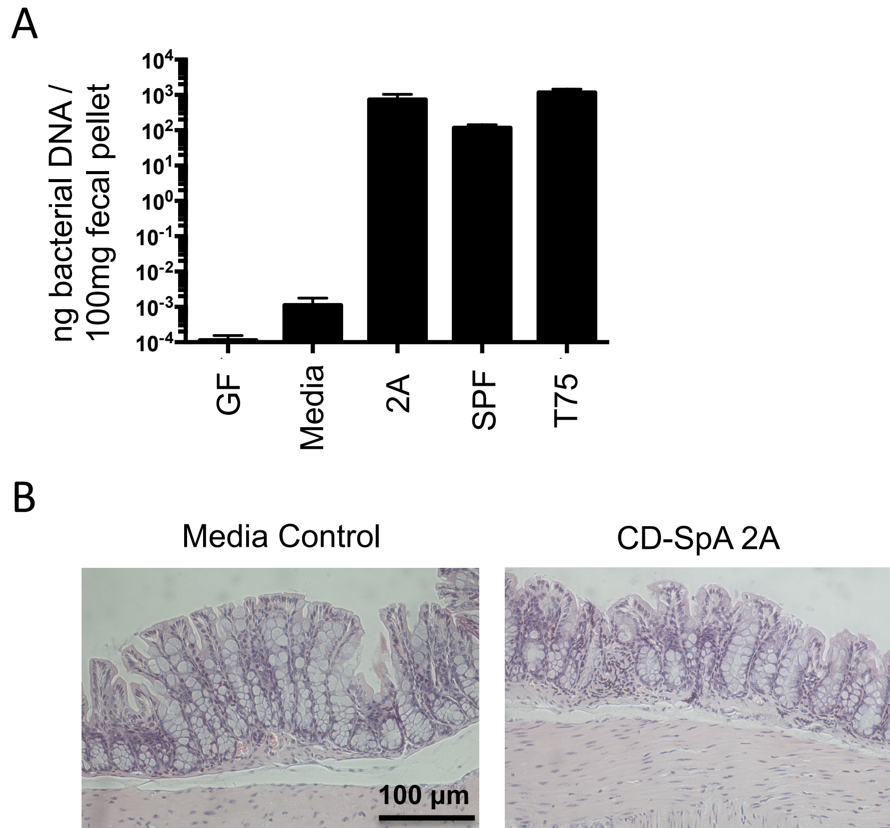


Fig. S9. Colonization of germ-free C57BL/6 mice with CD-SpA *E. coli* 2A. A. Mice are colonized outside the isolator and analyzed at 1-2 weeks post-colonization. Analysis of 16S rDNA (ng) / fecal pellet is shown to assess quality of colonization. Results from a GF control (Ctrl), Media alone gavage (Media), CD-SpA 2A gavage, SPF fecal pellet gavage, and non-AIEC CD-derived *E. coli* T75 are shown. **B**. Colonic histology of control and CD-SpA 2A colonized WT mice 10 days post-colonization.

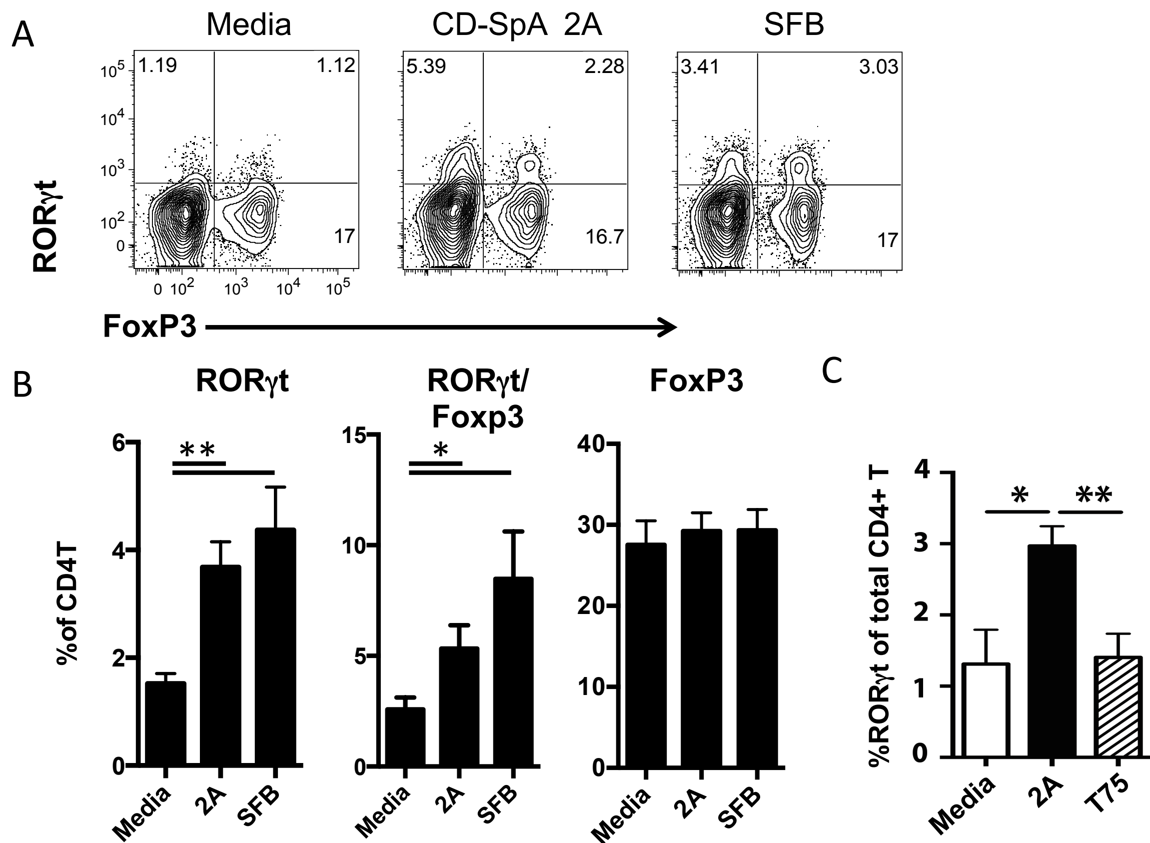


Fig. S10. CD-SpA *E. coli* 2A induces Th17 mucosal immunity. Live, CD3+ CD4+ T cells from the small intestinal lamina propria were identified by flow cytometry. **A.** Intracellular cytokine staining was used to evaluate RORγt and FoxP3 expression two weeks post-colonization. **B.** Percentage of RORγt+ and/or FoxP3+ CD4+ T cells in lamina propria were quantified by flow cytometry. Bar graph represents geometric mean and error bars represent SEM. Data are a composite of 3 independent experiments with 7-10 mice / group. ANOVA, p-values < 0.05 indicated by asterisk. **C.** SFB-free SPF C57BL/6 weanlings were gavaged with media control, AIEC 2A, or T75 and analyzed on day 10 post-gavage. Bar graphs represent geometric mean of 3-5 mice / group. Error bars represent SEM. One of two representative experiments is shown. ANOVA, p-values indicated by asterisk.

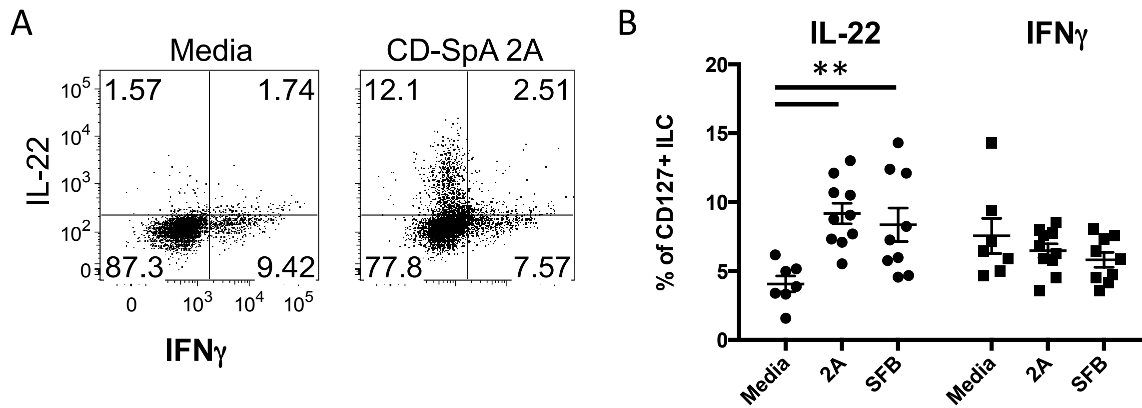


Fig. S11. CD-SpA *E. coli* 2A induces ILC3 mucosal immunity. A. Lineage negative, CD127+ ILC production of IL-22 and IFNγ two weeks after gavage with media alone, $2x10^9$ cfu CD-SpA E. coli 2A or SFB. **B.** Percentage of ILCs expressing IL-22 (circles) or IFNγ (squares) in the lamina propria were quantified by flow cytometry. Line and error bars represent mean and SEM, respectively. Data are a composite of 3 independent experiments with 7-10 mice / group. ANOVA, p-values < 0.01 indicated by asterisks.

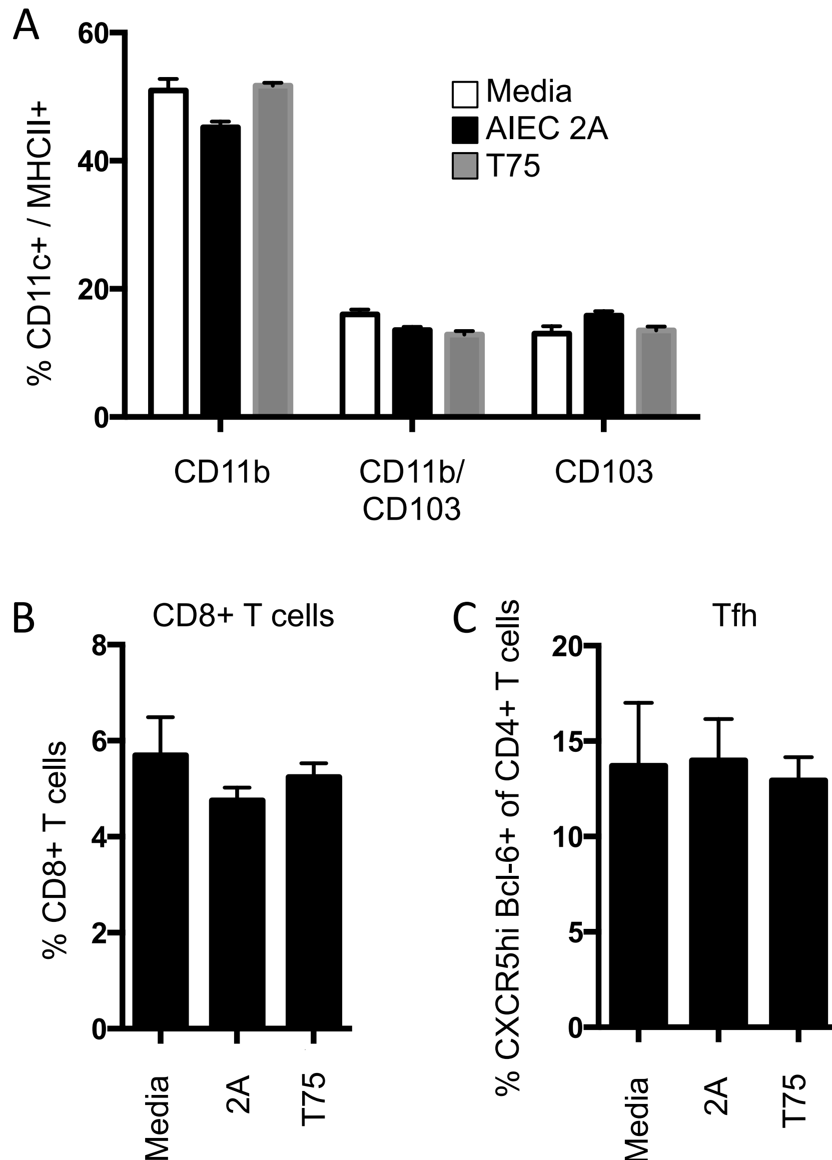


Fig. S12. Characterization of dendritic cells, CD8+ T cells and Tfh. Germ-free C57BL/6 mice were colonized with 2 x 10^9 CFU of CD-SpA-derived *E. coli* isolate 2A or non-adherent *E. coli* T75 and analyzed after 10 days. Flow cytometry gating of live cells was used to assess the percentage of dendritic cells (A) and CD8+ T cells (B) in the colonic lamina propria as well as follicular T helper cells (C) within the CD4+ T cell compartment of the Peyer's patches (C). Bar graphs represent geometric mean for at least 3 mice per experimental group. Data are from 1 of 3 representative experiments. Error bars represent SEM.

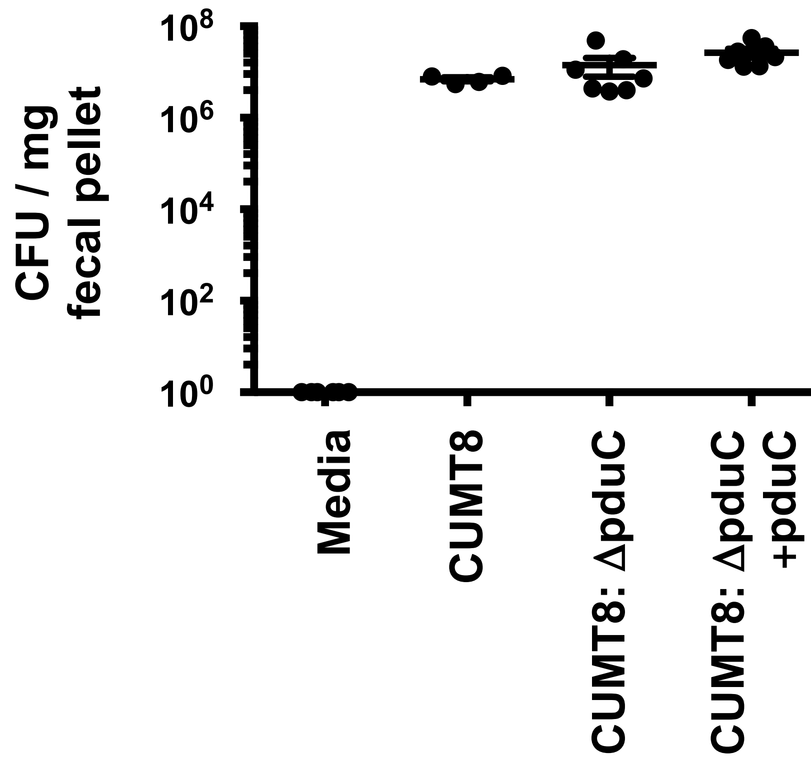


Fig. S13. Assessment of colonization. Mice were colonized outside the isolator in sterile cages and analyzed at 10 days post-colonization. Analysis of CFU / mg fecal pellet is shown. Each dot represents an individual mouse and error bars reflect SEM. One representative of 3 experiments is shown.

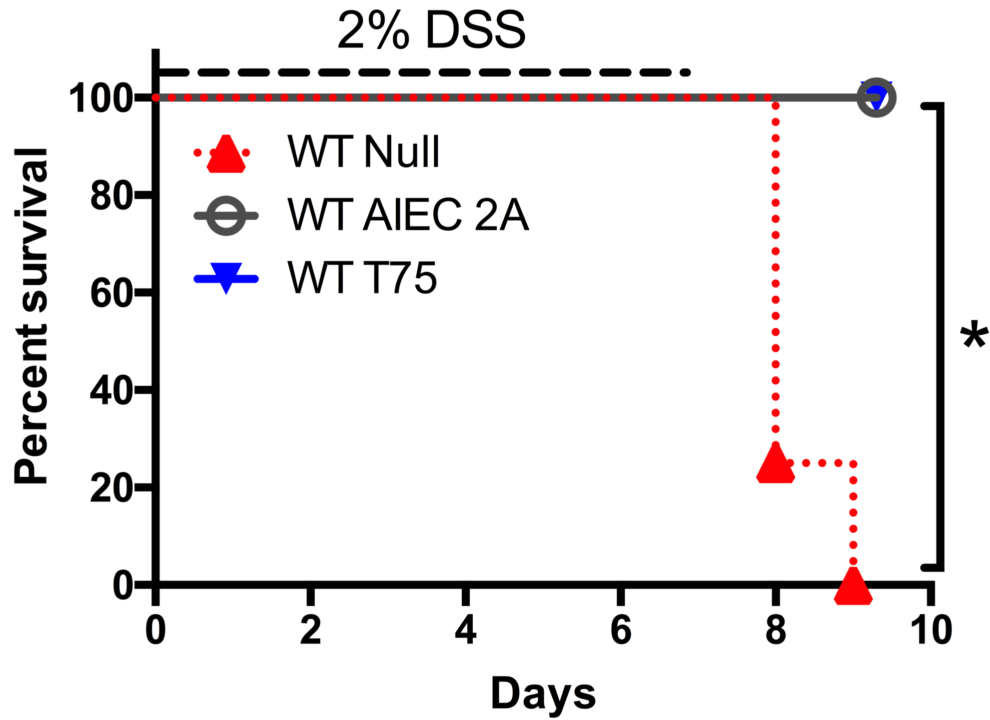


Fig. S14. CD-SpA AIEC 2A protects WT mice against DSS. Survival curves of WT mice colonized with AIEC 2A (N=6), T75 (N=5) or with media control (Null, N=5) and exposed to 2% DSS for 7 days. Data are a composite of 2 independent experiments. Log rank test, $P<0.05$.

	CD (N=15)	CD-SpA (N=25)	UC (N=10)	UC-SpA (N=9)	p-value
Demographics					
Age	39.7	44.6	52.6	45.4	0.26, 0.13
Female, %	71.0	68.0	64.0	67.0	NS
BMI	23.2	24.3	25.0	26.0	0.45, 0.58
Clinical Parameters					
Vitamin D	34.2	29.7	32.5	25.6	0.15, 0.46
ESR	19.1	25.3	11.3	25.0	0.20, 0.22
CRP	2.0	4.1	0.9	1.4	0.19, 0.08
Hemoglobin	12.5	12.0	13.5	13.0	0.25, 0.46
Disease Activity Indexes					
BASDAI	2.5	4.8	2.2	5.1	0.01, 0.01
44-pt joint	1.2	3.3	0.3	4.3	0.006, 0.001
HBI (CD)	6.1	7.8			0.07
UCDAI (UC)			5.0	5.5	0.6
Medication Use					
anti-TNFa, %	53	44	45.5	22.2	
Thp/MTX, %	20	20	18.2	22.2	
anti-integrin, %	7	8	0.0	0.0	

Table S1. Demographics, clinical parameters and disease activity indexes of study subjects. Erythrocyte sedimentation rate (ESR); C-reactive protein (CRP); Bath Ankylosing Spondylitis Disease Activity Index (BASDAI); Harvey-Bradshaw Index (HBI) for disease activity in Crohn's Disease (CD); Ulcerative Colitis Disease Activity Index (UCDAI) for disease activity in UC.

Montreal Classification	CD (N=15)	CD-SpA (N=25)
Age		
A1, %	21	12
A2, %	79	84
Location		
L1, %	22	28
L3, %	78	72
Behavior		
B1, %	20	21
B2, %	60	57
B3, %	20	21
perianal, %	21	12

Table S2. Montreal classification of CD cohort. Age of disease onset, A1<16 years of age, A2 17-40 years of age; Location of disease, L1 ileal, L3 ileocolonic; Behavior of disease, B1 non-stricturing, non-penetrating, B2 stricturing, B3 penetrating.

Reference Strain	ColV	fyuA	lpf141	lpf154	iroN	chuA	pduC	ratA	kpsMII	GSP	traC	ibeA	hcp	malX
NC101	+	+	-	-	-	+	+	+	+	+	-	+	-	+
LF82	+	+	+	-	ND	+	+	+	+	+	-	+	-	+
T75	-	-	-	-	-	-	-	-	-	+	-	-	+	-

CD-SpA	ColV	fyuA	lpf141	lpf154	iroN	chuA	pduC	ratA	kpsMII	GSP	traC	ibeA	hcp	malX
I1	+	-	+	-	-	+	+	-	+	+	-	-	-	+
I2	+	-	+	-	-	+	-	-	+	+	-	-	-	+
I3	+	-	+	-	-	+	-	-	+	+	+	-	-	+
I4	+	-	+	-	-	+	-	-	+	+	-	-	-	+
I5	+	-	+	-	-	+	-	-	+	+	-	-	-	+
2A	+	-	-	-	+	+	+	+	-	+	+	-	-	-
C3	+	+	+	-	+	+	+	+	+	+	-	-	-	-
C10	+	-	-	-	+	+	+	+	+	+	+	-	-	-
C17	+	-	-	-	+	+	+	+	-	+	+	-	-	-
C18	-	-	-	+	+	-	+	+	-	-	-	-	-	-
C4	+	-	-	-	+	+	+	+	-	+	+	-	-	-
C5	+	-	-	-	+	+	+	+	+	+	+	-	-	-
C6	+	-	-	-	+	+	+	+	+	+	+	-	-	-
C7	+	-	-	-	+	+	+	+	+	+	+	-	-	-
C8	+	-	-	-	+	+	+	+	+	+	+	-	-	-

CD isolate	ColV	fyuA	lpf141	lpf154	iroN	chuA	pduC	ratA	kpsMII	GSP	traC	ibeA	hcp	malX
B4	-	-	-	-	-	+	+	+	-	-	+	-	-	-
B5	+	+	-	-	+	+	-	+	+	+	+	+	-	+
B10	-	+	-	+	-	+	+	+	-	-	+	-	-	-
L4	-	-	-	-	-	-	-	-	-	+	+	-	+	-
L10	-	-	-	-	-	-	-	-	-	+	+	-	+	-
LD1	-	+	-	+	-	+	-	+	-	-	+	-	-	-
LD4	-	+	-	+	-	+	-	+	+	-	+	-	-	-
LD6	-	+	-	+	-	+	-	+	-	-	+	-	-	-
LD9	+	+	-	-	+	+	+	+	+	-	+	-	-	+
LD10	-	+	-	+		-	-	+	-	-	+	-	-	-
S7	-	-	+	+	-	-	+	-	-	-	+	-	-	-
S10	+	+	-	-	-	+	-	+	+	-	+	-	-	+
R6	-	-	-	-	-	-	-	-	-	-	+	-	+	-
R8	-	-	+	+	-	-		-	-	-	-	-	+	-
R9	-	-	-	-	-	-	-	-	-	-	+	-	+	-

Table S3. AIEC marker genes in reference strains and clonal isolates from CD and CD-SpA patients. Multiplex PCR was performed for the AIEC-associated virulence genes listed. The presence (yellow, "+") or absence (white, "-") of these factors in CD and CD-SpA isolates as well as reference strains NC101, LF82 and T75 is shown. ND, not determined.