**Supplemental Information**

# Diverse Intestinal Bacteria Contain

# Putative Zwitterionic Capsular Polysaccharides

# with Anti-inflammatory Properties

C. Preston Neff, Matthew E. Rhodes, Kathleen L. Arnolds, Colm B. Collins, Jody Donnelly, Nichole Nusbacher, Paul Jedlicka, Jennifer M. Schneider, Martin D. McCarter, Michael Shaffer, Sarkis K. Mazmanian, Brent E. Palmer, and Catherine A. Lozupone

# Diverse intestinal bacteria with putative

# zwitterionic capsular polysaccharide operons

# can induce IL-10 and T regulatory cells

C. Preston Neff, Matthew E. Rhodes, Kathleen L. Arnolds, Colm B. Collins, Jody Donnelly, Nichole Nusbacher, Paul Jedlicka, Jennifer M. Schneider, Martin D. McCarter, Michael Shaffer, Sarkis K. Mazmanian, Brent E. Palmer, Catherine A. Lozupone
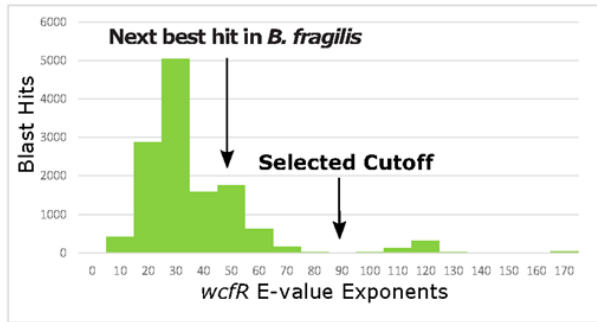
# Supplemental Figures

**Figure S1.** *Related to Experimental Procedures*: **A.** Distribution of best BLAST hits to the *wcfR* gene of *B. fragilis* 9343 across 8024 genomes. The x axis represents the coefficient of the e-value of BLAST hits to *wcfR* (e.g. "10" represents is $1e^{-10}$.) Our threshold for orthology to *wcfR* is indicated at an e-value of $1e^{-90}$ and the location of the second best hit in *B. fragilis* to a *wcfR* homologue is shown ($3e^{-48}$). **B.** Amplification of cDNA from bacterial cultures with *wcfR*-specific primers to verify the expression of the *wcfR* gene in the growth conditions employed.

**Figure S2.** *Related to Figure 1:* Heatmap showing gene family content of different AATGal-PSA operons. The rows represent gene names and the columns different operons. The genes are colored green if present and white if absent in the operon. Rows and columns are both ordered by UPGMA hierarchical clustering such that operons with similar gene membership and genes with similar operon membership are grouped. Only genes families with members found in at least 4 operons are plotted and *upaY* and *upaZ* regulators are not included. The predicted functions of each gene family are indicated in Table S1.
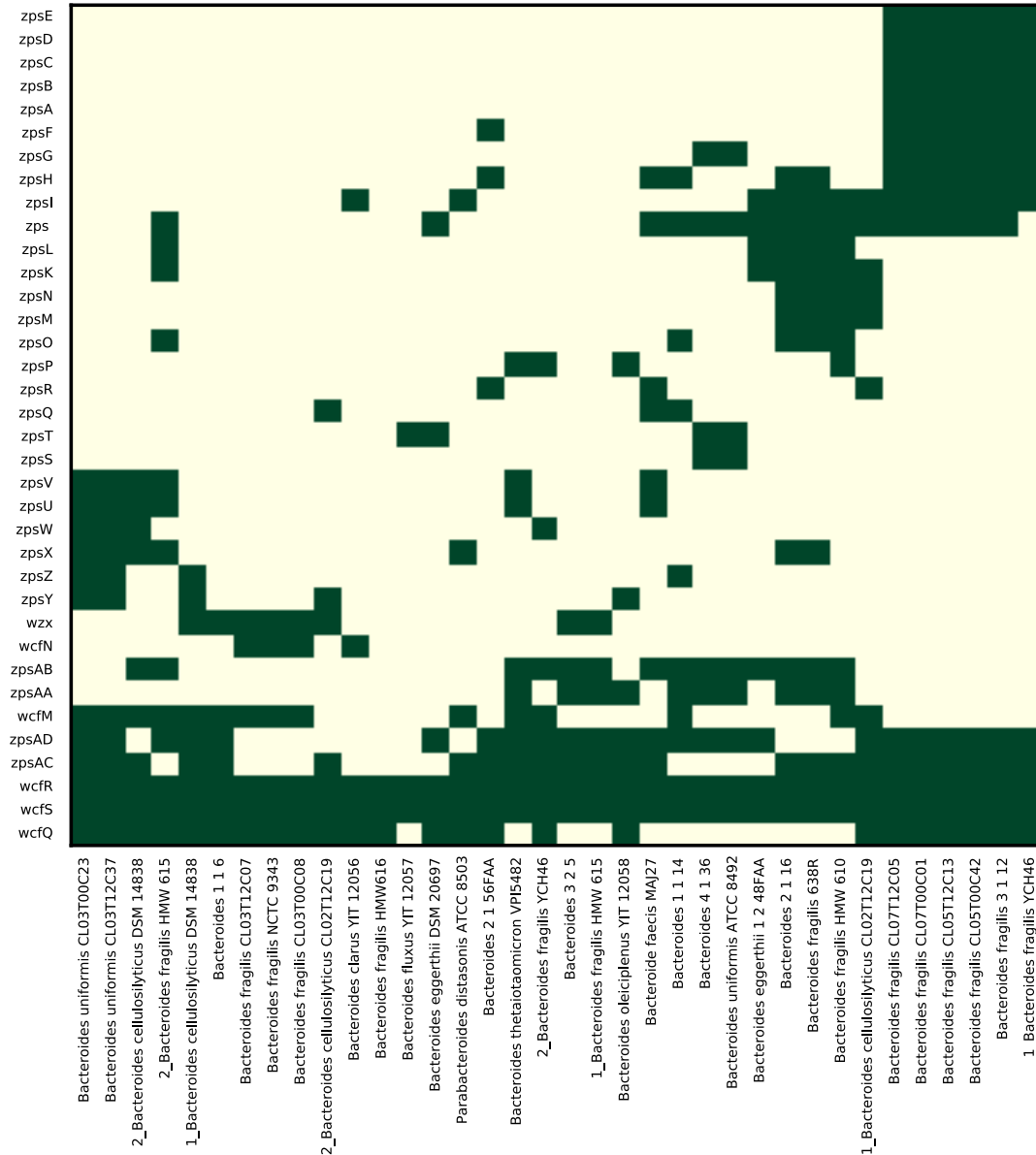
**Figure S3:** *Related to Figure 1*. Neighborjoining tree of operons based on operon gene content, such that genes with identical gene family membership have a branch length of zero. Operons from *B. fragilis* strains are colored red with PSA and PSA2 marked to illustrate the diversity of AATGal-ZPSs within this species alone. The two *B. cellulosilyticus* operons are marked with a green star. The operons contained in the genomes of other assayed bacteria are marked with a yellow star. The chemical structures of PSA and PSA2 are shown to illustrate the types of molecules produced in different parts of the tree.
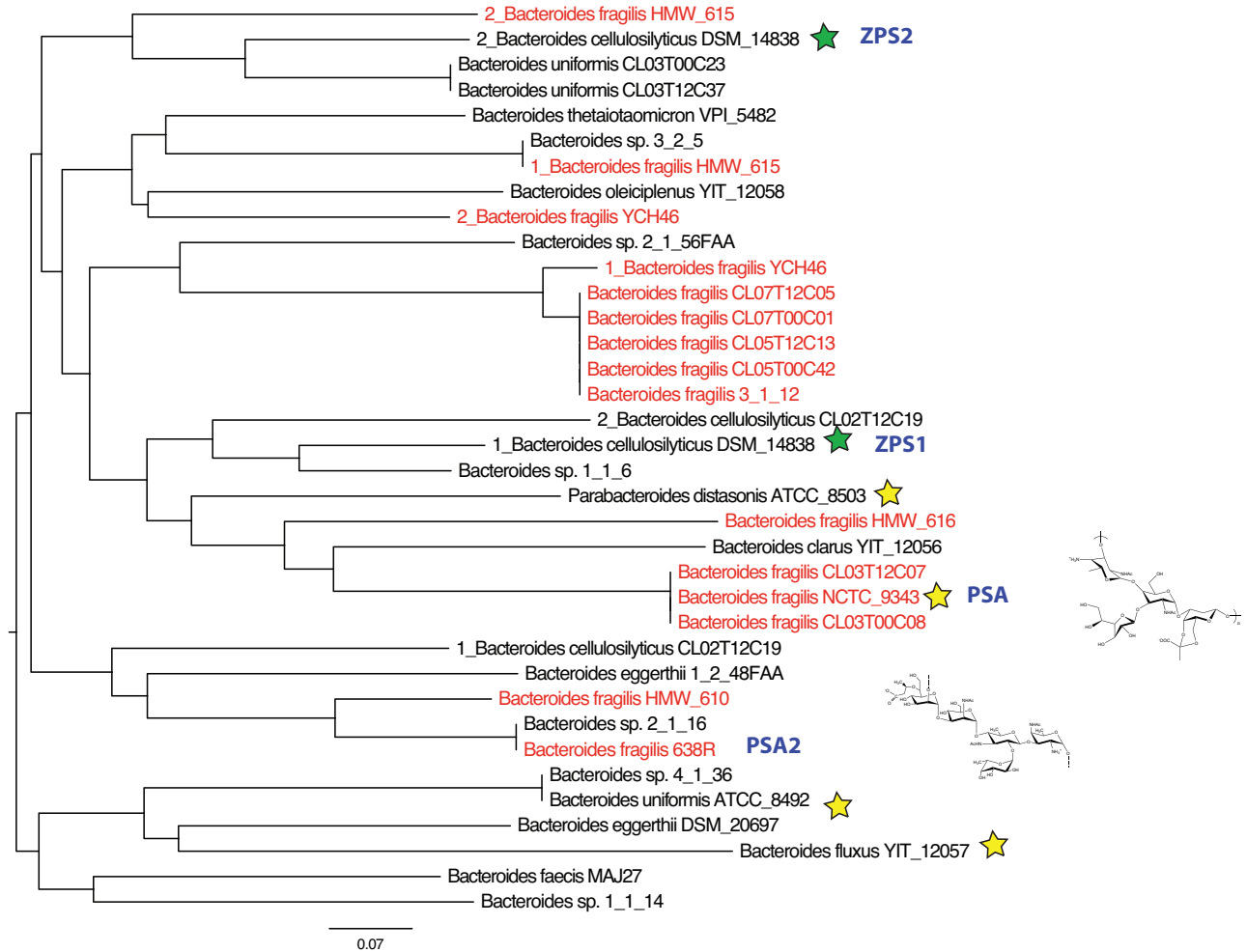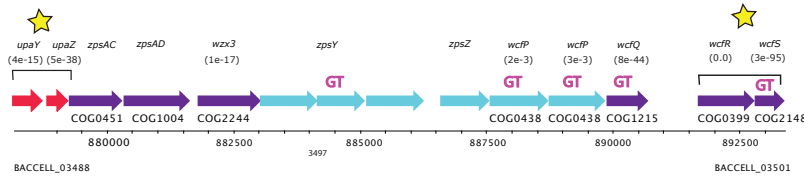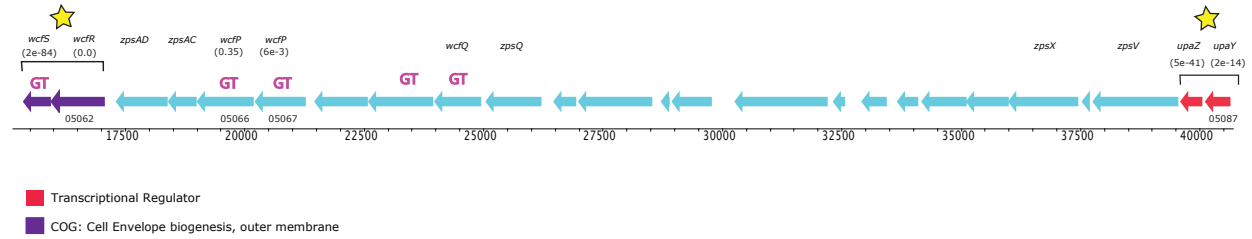
**Figure S4:** *Related to Figure 1*. Predicted AATGal operons of *Bacteroides cellulosilyticus* DSM 14838. Genes labeled GT represent glycosyltransferases. Gene homologues to genes in the PSA operon of *B. fragilis* are noted with their e-value in parentheses. Genes that are conserved across ZPS operons are also named and correspond to the Heatmap in Figure S2 and Table S1. The yellow stars indicate genes that were conserved in the AATGal-ZPS operons of 50 evaluated strains of *B. fragilis* in Coyne *et al*. (Coyne et al. 2001).

*Bacteroides cellulosilyticus* DSM 14838: ZPS1 operon: Scaffold 5



*Bacteroides cellulosilyticus* DSM 14838: ZPS2 operon: Scaffold 9



■ Transcriptional Regulator

■ COG: Cell Envelope biogenesis, outer membrane

**Figure S5.** *Related to Figure 4:* **A.** WT *B. fragilis* induces lower levels of inflammatory cytokines compared to *B. fragilis ΔPSA* in PBMC and LPMC. Human PBMC (circles) and LPMC (triangles) were cultured with bacteria lysate of WT *B. fragilis* or *B. fragilis* ΔPSA for 3 days and levels of IL-6, TNF-α, IL-17 and IL-22 in the supernatant were measured by ELISA. The horizontal line represents the median. Statistical significance was calculated by paired T tests. **B**. Representative staining for CD25+FoxP3+CTLA4+CD127- Tregs in human PBMC and LPMC. Human PBMC was stained for CD3, CD4, CD8, CD25, CD127, FoxP3 and CTLA-4, while human LPMC also included live/dead and CD45 staining.  Representative examples of Tregs for PBMC (top) and LPMC (bottom) are shown.

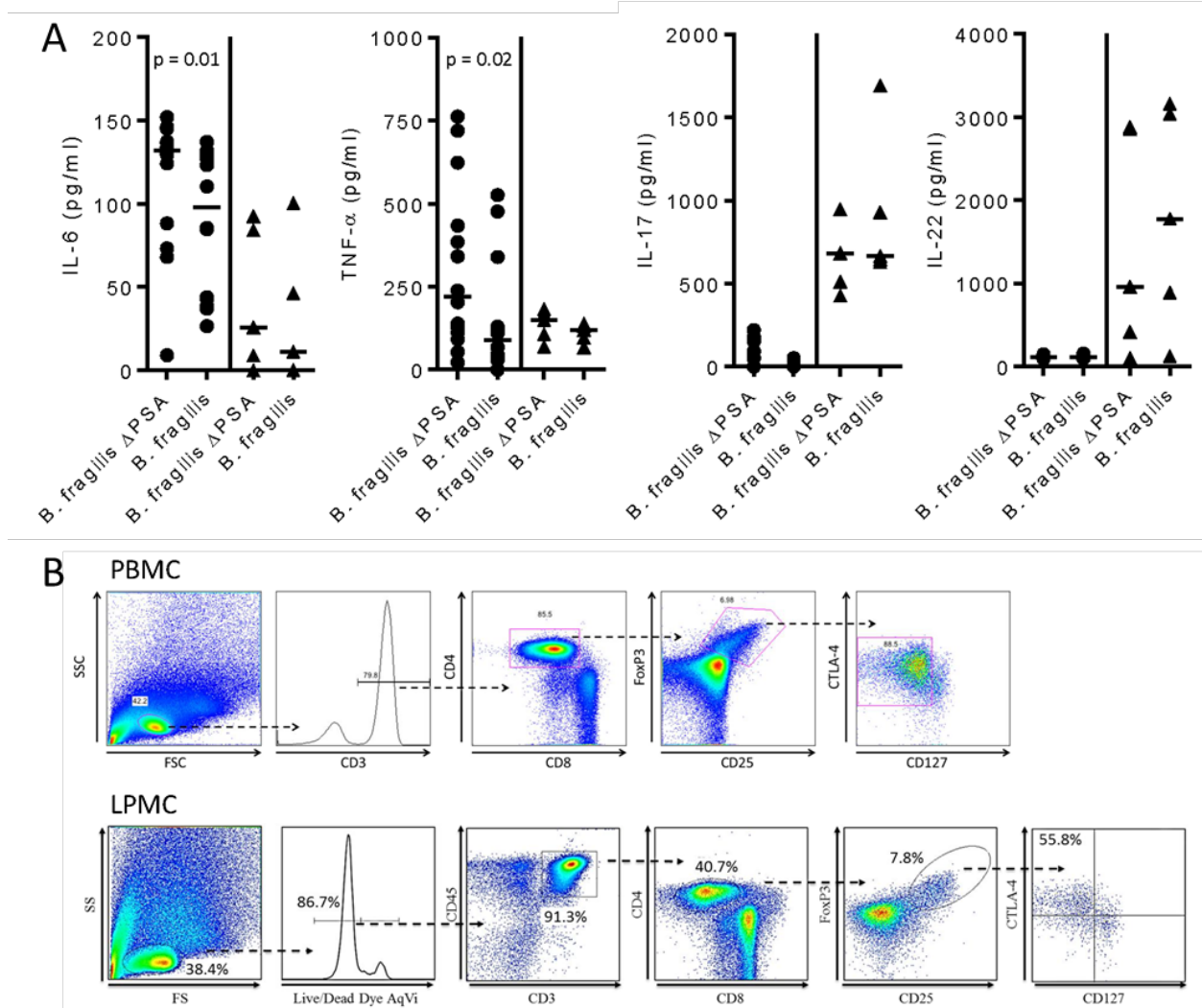**Figure S6.** *Related to Figure 5.* Highly related bacteria, *Bacteroides cellulosilyticus* and *Bacteroides intestinalis,* with opposite PSA designations induce different levels of IL-10 determined by ELISA and proportion CD25+FoxP3+CTLA4+CD127- cells of CD4+ T cells cultured with *B. cellulosyliticus* or *B. intestinalis* for 3 days in PBMC, **A.** and **B.,** and in LPMC, **C.** and **D.**
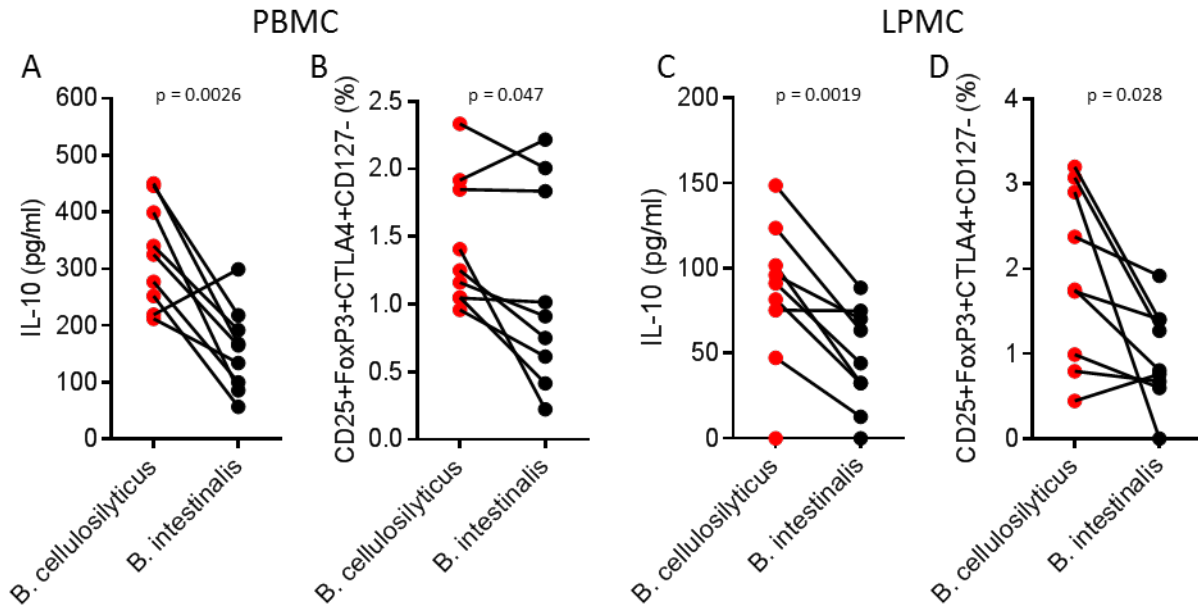
**Figure S7.** *Related to Figure 2.* Electron micrograph of **A.** *Clostridium spiroforme* (reprinted with permission from (Baldassarri et al. 1989)) with red arrows pointing to putative outer membrane vesicles that are morphologically similar to **B.** outer membrane vesicles shown to contain PSA in *B. fragilis* (reprinted with permission from (Shen et al. 2012)).
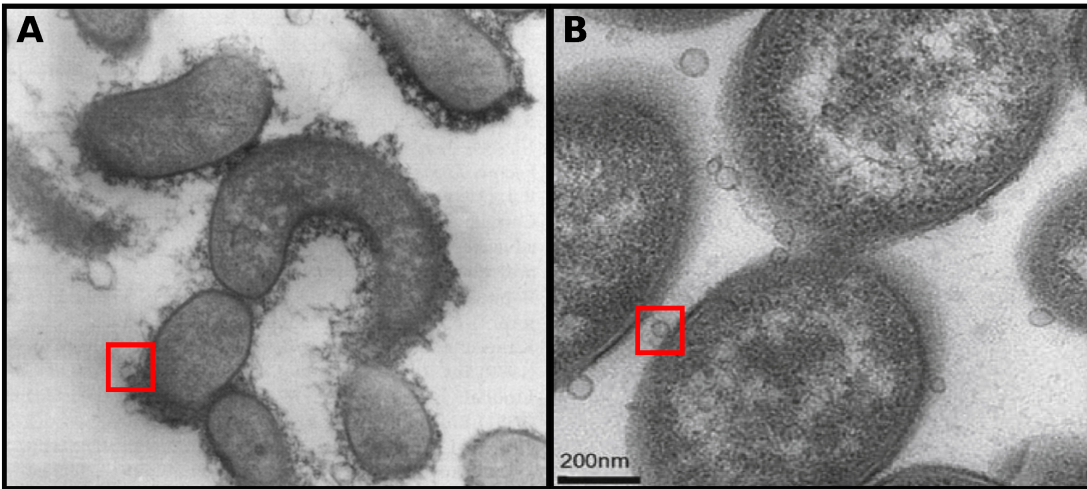
**Table S1: Putative functions for AATGal-ZPS conserved genes in Figure S2**

| Gene | NCBI ACCESSION | Putative Function |
|---|---|---|
| zpsA | YP_098723 | Hypothetical protein |
| zpsB | ZP_17243765 | Capsular polysaccharide polymerase |
| zpsC | YP_098717 | Acyl carrier protein |
| zpsD | ZP_17252722 | AMP-dependent synthetase and ligase |
| zpsE | ZP_17243761.1 | Hypothetical protein |
| zpsF | ZP_07808028 | Hexapeptide transferase |
| zpsG | ZP_17243764.1 | Oligosaccharide repeat unit transporter |
| zpsH | ZP_08589520 | Transferase |
| zpsI | ZP_17243758.1 | Dehydrogenase |
| zpsJ | ZP_09860609.1 | Chloramphenicol acetyltransferase |
| zpsK | ZP_17263668 | Hypothetical protein |
| zpsL | ZP_17258425 | Trans-membrane protein |
| zpsM | WP_014298531 | 3-oxoacyl-ACP synthase |
| zpsN | ZP_06091404 | Acyl carrier protein |
| zpsO | ZP_17263671 | Glucose-1-phosphate thymidylyltransferase |
| zpsP | ZP_18880373 | O-antigen repeat unit transporter |
| zpsQ | ZP_17210313 | Carbamoyl-phosphate synthase large subunit |
| zpsR | ZP_08589530.1 | Chloramphenicol acetyltransferase |
| zpsS | ZP_07939427 | Glycosyl transferase group 1 |
| zpsT | ZP_07939424 | FAD binding domain-containing protein |
| zpsU | ZP_03680696 | Polysaccharide pyruvyl transferase |
| zpsV | ZP_17285045 | Hypothetical protein |
| zpsW | ZP_17263668 | Transcriptional regulator |
| zpsX | ZP_03680708 | Transmembrane protein |
| zpsY | ZP_17285040 | Glycosal transferase |
| zpsZ | ZP_17286368 | Acyl transferase |
| wcfN | YP_211029.1 | LPS biosynthesis related glycosyltransferase |
| wzx | ZP_17230422 | LPS biosynthesis related flippase |
| zpsAA | WP_014298513.1 | UDP-GlcNAc 2-epimerase |
| zpsAB | ZP_06091401 | Conserved putative glycosyl transferase |
| wcfM | ZP_06995100.1 | UDP-galactose-lipooligosaccharide |
| wcfS | WP_005795339.1 | Undecaprenyl-phosphate galactose phosphotransferase |
| wcfR | WP_005786175.1 | Amino sugar synthetase |
| zpsAC | ZP_03680689 | Nucleoside-diphosphate sugar epimerase |
| zpsAD | ZP_07935853 | Nucleotide sugar dehydrogenase |
| wcfQ | ZP_17243768 | Glycosyltransferase |

**Supplemental Experimental Procedures**

*ZPS Operon Identification*

The faa and frn files of 2659 complete and 5406 draft genomes were downloaded from the NCBI FTP server in September of 2013. A BLASTP search of the PSA operon of *B. fragilis* NCTC 9343 against the entire collection of genomes was conducted (Altschul et al. 1997). Genomes with a BLAST hit to the *wcfR* gene with an e-value $< 1e^{-90}$ were selected for further scrutiny. This threshold was based on the observation of a bi-modal distribution of hits to *B. fragilis wcfR*, with the *B. fragilis wcfR* homologue DegT in the lower distribution (Figure S1a). The proximity of *wcfR*, *wcfS*, *upaY*, and *upaZ* gene homologues on genomic contigs were next determined using custom code.

In order to characterize the operon gene membership, genes located between *upaY,/upaZ* regulators and *wcfr/wcfs* gene homologues were considered to be a part of the operon. In the PSA operon, these genes define the operon's beginning and end. In order to determine whether the operon extended beyond the *wcfr/wcfs* gene homologues, the genomic region between the *upaY,/upaZ* regulatory genes through 20 genes past *wcfs/wcfr* were run through the DOOR2 operon prediction program (Mao et al. 2014). Resulting operon gene lists thus sometimes included genes that were located after *wcfs/wcfr* gene homologues based on DOOR2 results.

To group the identified genes into families, pairwise BLAST searches between all genes were conducted. Genes were joined with an edge in a network if the pairwise BLAST search for the pair had a bit score of >50. A k-clique percolation method (Palla et al. 2005)as implemented in Networkx was used to define gene families as network modules. To relate operons based on their gene content, modules with at least 4 elements were retained and used to make a matrix with gene families as rows and with putative AATGal-ZPS operons as columns. The binary distance between operon gene content was calculated based on the Jaccard index using the QIIME software package (Caporaso, Kuczynski, et al. 2010).

*Phylogenetic tree construction*

To make the phylogenetic tree of *wcfR* (Fig. 2), the median length for the genes with BLAST hits to *wcfR* with an e-value less than $10^{-90}$ was determined and all genes whose length fell within 90% of the median were aligned with muscle using default parameters (Edgar 2004). Subsequently a tree was constructed using Phylip Neighbor Joining (Felsenstein 1989).

To make 16S ribosomal RNA (rRNA) phylogenetic trees representing surveyed bacteria in the Bacteroidales and Erysipelotrichales orders, we first assigned taxonomy to all 16S ribosomal RNA (rRNA) genes from the collection of genomes by the uclust algorithm (Edgar 2010) against the greengenes database from May of 2013 (DeSantis et al. 2006) using the QIIME assign_taxonomy.py program (Caporaso, Kuczynski, et al. 2010). The 16S rRNA genes assigned to the orders Bacteroidales and Erysipelotrichales were extracted and all genes shorter than 1,300 base pairs were discarded. Of the remaining genes, the longest 16S rRNA gene from each genome was used for phylogenetic construction. For each order, an outgroup was added (*Halobacteroides halobius* for the Erysipelotrichales and a Flavobacteria for the Bacteroidales). The 16S rRNA sequences were aligned against the QIIME core.set.imputed alignment using PyNAST (Caporaso, Bittinger, et al. 2010). The alignment was filtered against the QIIME lane_mask_in_1's_and_0's, and trimmed so that each sequence began and ended at the same base (Caporaso, Kuczynski, et al. 2010). Trees for both orders were constructed using Phylip Neighbor Joining (Felsenstein 1989).

### Growth of isolates and lysate preparation

Isolates were purchased from the ATCC or DSMZ and grown in rich media depending on their preferences including either 1) Mega Media (Goodman et al. 2011) 2) PYG (DSMZ 104) or 3) BHI (Lee et al. 2013). When robust culture growth was observed (typically after 18-36 hours depending on the bacteria), liquid cultures were centrifuged at 2500 rpm for 10 minutes. Supernatant was removed and the pellet was resuspended in PBS and frozen at -80$^{\circ}$C. Resuspended bacteria were subjected to freeze/thaw or heat killing before being used in immune stimulations. Since immune stimulations with heat killed compared to freeze/thaw lysates resulted in no differences (data not shown), all reported stimulations were done with freeze/thaw lysates in order to avoid protein denaturing.

To determine whether *wcfR* was being expressed under the growth conditions employed, we designed primers that specifically amplified the *wcfR* gene using PrimerProspector software (Walters et al. 2011). Because it was difficult to design primers that amplified all *wcfR* genes within our e$^{-90}$ BLAST threshold and not other homologues of *wcfR*, we developed a primer set that exclusively amplified most *wcfR* genes in the Bacteroides genus, and in this case was used to verify the presence of *wcfR* mRNA in cultures of *B. cellulosilyticus, B. fragilis WT,* and *B. uniformis.* As negative controls, we attempted to amplify from *B. fragilis Δ PSA,* and *B. intestinalis.* The sequences of this primer set are: Forward primer: 5- GGR CGY ATC YTK GTG ATG AA -3, Reverse primer: 5- CCG ACA ATR TCA TAA CGC CA -3. RNA was extracted from bacterial cultures using the PureLink RNA Mini Kit from Life technologies. RNA was reverse transcribed to cDNA using Thermoscript RT-PCR system from Invitrogen. Standard PCR was then performed on cDNA to amplify wcfR in various bacterial strains (Figure S1b).

### Generating wcfR knockouts

The *B. cellulosilyticus*ΔZPS1 strain was developed using the pKNOCK-bla-ermGb vector to disrupt *wcfR* via targeted insertional mutagenesis as previously described (Alexeyev 1999). Briefly, a 207 base pair internal fragment of the target gene, *wcfR*, was first amplified using the following primers; which also encoded a Sal1 or Kpn1 restriction site respectively (Forward Sal-1 ACGCGTCGACTGCGGGTGCAAAGCCTGTTATG, Reverse KPN-1 CGGGGTACCTCCTGTACCAATGAGTCGC). The resulting amplicon was digested and ligated into the pKNOCK-bla-ermGb vector, which was then electroporated into *Escherichia coli* S17-1 lambda pir. Conjugation and plasmid transfer into *B. cellulosilyticus* from *E. coli* S17-1 lambda pir was then initiated by co-culture of the two bacteria on BHI blood agar plates under aerobic conditions for 36 hours, as the E. coli growth becomes dense the conditions become more anaerobic facilitating conjugation and growth of transformed Bacteroides. Disruption of the *wcfR* gene occurs via insertion of the pKNOCK plasmid by homologous recombination. Insertion of the pKNOCK plasmid confers ampicillin and erythromycin resistance, which allowed for selection of *wcfR* knock-outs by growth in media containing 50ug/mL of erythromycin. Successful disruption of *wcfR* was confirmed by attempting to amplify with primers within the *wcfR* gene that flank the insertion site using the following primers: Forward GAG TAA CAG CTG AAG ATG AAG TG Reverse AAA CAG GTT GCG TAT]. Because *B. cellulosilyticus* DSM 14848 has 2 highly related copies of *wcfR*, we selected primers that were

11

specific to one copy and not the other. We also confirmed that the mRNA of only one copy of wcfR was being made using PCR with wcfR targeted primers and Sanger sequencing.

### *Immune Assays*

Human PBMCs were isolated by Ficol gradient centrifugation as previously described (Chain et al. 2013; Kassu et al. 2010; Neff et al. 2015) from the blood of 13 normal individuals. Informed consent was obtained and the study protocol was approved by the Colorado Multiple Institutional Review Board (COMIRB #14-1595). PBMCs were cultured with 10 µg freeze killed bacterial lysate for 3 days at 37°C. Six and 10-day stimulations were also performed and had similar results to those of 3-day stimulations (data not shown). Stimulations were performed in the presence of Streptomycin and Penicillin and in aerobic conditions and no bacterial growth was observed in the cell cultures. Unlike in the assays conducted by Kreisman and Cobb (Kreisman and Cobb 2011) in human PBMC, our stimulations were conducted in absence of exogenous IL-2; the IL-2 receptor CD25, is upregulated in the presence of IL-2, which could compromise our Treg staining.

To quantify cytokine secretion after 3-day stimulations with bacterial lysates, supernatant was collected and subjected to ELISA Ready Set Go! (eBioscience) following manufacture's instructions. Specifically, human IL-6, IL-10, IL-17, IL-22 or TNF-α ELISA Ready-SET-GO! Kits (eBioscience) were used. ELISA plates were coated overnight with the specific capture antibody, then washed and blocked with assay diluent for 30 minutes. 100 µl of supernatant from three day stimulations with bacterial lysate was evaluated for cytokine as per instructions supplied in the kit. To enumerate Tregs, cells were washed with staining buffer containing PBS, 2% BSA, 2 mM EDTA and 0.09% $NaN_3$ and surface staining was performed with BV605-labelled anti-CD3 antibody (BioLegend), PerCP/Cy5.5-labelled anti-CD4 (BioLegend), Alexa405-labelled anti-CD8 antibody (Invitrogen), APC-Cy7-labeled anti-CD25 (BD Biosciences) and FITC-labeled anti-CD127 (BioLegend). Intracellular staining for PE-labeled anti-FoxP3 antibody (eBioscience), APC-labeled anti-CTLA-4 antibody (BD Biosciences), was performed using the FoxP3 staining buffer set (eBioscience).

ICCS was used to determine IL-10+ cells. Briefly, PBMC was cultured with bacteria lysate for 24 hours. Monensin was added 20 hours after the addition of bacterial lysate. Cells were then subjected to surface staining and fixed as above and permeabilized using Fix and Perm (Lifetechnologies) for 30 minutes. Pre-labeled anti-IL-10 was added to cells and stained overnight at 4°C. Cells were washed 2 times and enumerated with a LSR II (BD Biosciences) and data were analyzed using FlowJo software (Treestar).

To rule out if induced IL-10 production and Tregs were due to restimulation of memory T cells, CD14+ monocytes were isolated by magnetic bead selection (Miltenyi Biotec), plated, stimulated with bacterial lysate for 4 hours and washed twice with PBS. Naïve T cells were negatively selected by magnetic beads (Miltenyi Biotec) and were added to the bacterial stimulated APCs. After three days supernatant was collected and subjected to IL-10 ELISA and cells were enumerated for Tregs as above.

### *Isolation of intestinal lamina propria lymphocytes and flow cytometry*

Resected gut tissue of jejunum, duodenum and colon were obtained from patients undergoing elective abdominal surgery. All patients signed a release to allow the unrestricted use of discarded tissues for research purposes and all protected patient information was de-identified to the laboratory investigators. Tissues were opened longitudinally, and trimmed of excess fat,

the mucosal layer and extraneous material, such as stitched portions, damaged tissues or connective tissues. Gut tissue was then washed with PBS supplemented with the antibiotics Streptomycin, Penicillin, and Amphotericin B and incubated at 37°C with PBS containing 1 mM EDTA for 2 cycles of 45 minutes to remove epithelial cells. The lamina propria layers were cut into 1 gram pieces and incubated while shaking with AIMS V media containing antibiotics, 0.5 mg collagenase D and 10 μg DNase I for 3 cycles of 45 minutes at 37 °C. The digested tissues were collected and washed with PBS and resuspended in RPMI1650 containing 10% human serum and overlaid on 15 ml of 60% Percoll in a 50-ml Flacon tube. Percoll gradient separation was performed by centrifugation at 800 g for 20 min. The lamina propria lymphocytes were collected from the interface of the Percoll gradient and suspended in ice-cold PBS. For analysis of Treg cells, isolated lymphocytes were labeled with the LIVE/DEAD fixable dead cell stain kit (Invitrogen) to exclude dead cells from the analysis. The cells were washed with staining buffer, then surface and intracellular staining was performed as above with Pe-Cy7-labeled anti-CD45. Cells were washed 2 times and enumerated with a LSR II (BD Biosciences) analyzed using FlowJo software (Treestar).

### *Murine TNBS Colitis Model*

In two separate trials, female C57/Bl6 mice aged 8 - 12 weeks were gavaged with 5.0 x $10^8$ bacterial cells in a 200 μL volume of PBS (as determined by plate count of fresh culture) or control 200 μL of PBS only on days 0, 7, and 14. The first trial used 5 mice per treatment group and the second used 10, although there was some attrition due to factors such as perforations during the TNBS enema in each case. Mice were cohoused prior to the intervention, and then housed in separate cages once the treatments began to prevent transmission of the introduced bacteria between mice. Mice were anesthetized, shaved and skin painted with 100μl of 1% TNBS in 100% EtOH on day 7. On day 14 mice received a rectal enema of 2.5% TNBS in 40% EtOH (5μl/g body weight). Body weight was recorded daily and mice were euthanized once the first animal reached 85% of initial weight in accordance with institutional guidelines. Colonic tissue was excised, opened along the mesentery by blunt dissection and rinsed of luminal content. Tissue was fixed in 10% formalin for histological assessment of disease or digested as previously described (Collins et al. 2013). Histological inflammatory index scores were assigned by a trained pathologist in a blinded fashion. Cells isolated from the colonic lamina propria were stained for CD4 (BioLegend), CD25 (Biolegend), LIVE/DEAD fixable dye (Invitrogen) and FoxP3 (Biolegend) using the FoxP3 staining kit (eBioscience) according to manufacturer's instructions. Data was acquired on a BD FACSCantoII and analyzed using FlowJo software. Statistical analysis was performed using GraphPad Prism. These experiments were approved by IACUC (B-104413(12)1E).

**Supplemental References**

Alexeyev, M. F. 1999. 'The pKNOCK series of broad-host-range mobilizable suicide vectors for gene knockout and targeted DNA insertion into the chromosome of gram-negative bacteria', *Biotechniques*, 26: 824-6, 28.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic Acids Res*, 25: 3389-402.

Baldassarri, L., A. Pantosti, A. Caprioli, P. Mastrantonio, and G. Donelli. 1989. 'Haemagglutination and surface structures in strains of Clostridium spiroforme', *FEMS Microbiol Lett*, 51: 1-4.

Caporaso, J. G., K. Bittinger, F. D. Bushman, T. Z. DeSantis, G. L. Andersen, and R. Knight. 2010. 'PyNAST: a flexible tool for aligning sequences to a template alignment', *Bioinformatics*, 26: 266-7.

Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Pena, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, and R. Knight. 2010. 'QIIME allows analysis of high-throughput community sequencing data', *Nat Methods*, 7: 335-6.

Chain, J. L., A. K. Martin, D. G. Mack, L. A. Maier, B. E. Palmer, and A. P. Fontenot. 2013. 'Impaired function of CTLA-4 in the lungs of patients with chronic beryllium disease contributes to persistent inflammation', *J Immunol*, 191: 1648-56.

Collins, C. B., C. M. Aherne, S. F. Ehrentraut, M. E. Gerich, E. N. McNamee, M. C. McManus, M. D. Lebsack, P. Jedlicka, T. Azam, E. F. de Zoeten, C. A. Dinarello, and J. Rivera-Nieves. 2013. 'Alpha-1-antitrypsin therapy ameliorates acute colitis and chronic murine ileitis', *Inflamm Bowel Dis*, 19: 1964-73.

Coyne, M. J., A. O. Tzianabos, B. C. Mallory, V. J. Carey, D. L. Kasper, and L. E. Comstock. 2001. 'Polysaccharide biosynthesis locus required for virulence of Bacteroides fragilis', *Infect Immun*, 69: 4342-50.

DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen. 2006. 'Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB', *Appl Environ Microbiol*, 72: 5069-72.

Edgar, R. C. 2004. 'MUSCLE: multiple sequence alignment with high accuracy and high throughput', *Nucleic Acids Res*, 32: 1792-7.

Edgar, R. C. 2010. 'Search and clustering orders of magnitude faster than BLAST', *Bioinformatics*, 26: 2460-1.

Felsenstein, J. . 1989. 'Phylip: phylogeny inference package (version 3.2). ', *Cladistics 5*: 164-66.

Goodman, A. L., G. Kallstrom, J. J. Faith, A. Reyes, A. Moore, G. Dantas, and J. I. Gordon. 2011. 'Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice', *Proc Natl Acad Sci U S A*, 108: 6252-7.

Kassu, A., R. A. Marcus, M. B. D'Souza, E. A. Kelly-McKnight, L. Golden-Mason, R. Akkina, A. P. Fontenot, C. C. Wilson, and B. E. Palmer. 2010. 'Regulation of virus-specific CD4+ T cell function by multiple costimulatory receptors during chronic HIV infection', *J Immunol*, 185: 3007-18.

Kreisman, L. S., and B. A. Cobb. 2011. 'Glycoantigens induce human peripheral Tr1 cell differentiation with gut-homing specialization', *J Biol Chem*, 286: 8810-8.

Lee, S. M., G. P. Donaldson, Z. Mikulski, S. Boyajian, K. Ley, and S. K. Mazmanian. 2013. 'Bacterial colonization factors control specificity and stability of the gut microbiota', *Nature*, 501: 426-9.

Mao, X., Q. Ma, C. Zhou, X. Chen, H. Zhang, J. Yang, F. Mao, W. Lai, and Y. Xu. 2014. 'DOOR 2.0: presenting operons and their functions through dynamic and integrated views', *Nucleic Acids Res*, 42: D654-9.

Neff, C. P., J. L. Chain, S. MaWhinney, A. K. Martin, D. J. Linderman, S. C. Flores, T. B. Campbell, B. E. Palmer, and A. P. Fontenot. 2015. 'Lymphocytic alveolitis is associated with the accumulation of functionally impaired HIV-specific T cells in the lung of antiretroviral therapy-naive subjects', *Am J Respir Crit Care Med*, 191: 464-73.

Palla, G., I. Derenyi, I. Farkas, and T. Vicsek. 2005. 'Uncovering the overlapping community structure of complex networks in nature and society', *Nature*, 435: 814-8.

Shen, Y., M. L. Giardino Torchia, G. W. Lawson, C. L. Karp, J. D. Ashwell, and S. K. Mazmanian. 2012. 'Outer membrane vesicles of a human commensal mediate immune regulation and disease protection', *Cell Host Microbe*, 12: 509-20.

Walters, W. A., J. G. Caporaso, C. L. Lauber, D. Berg-Lyons, N. Fierer, and R. Knight. 2011. 'PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers', *Bioinformatics*, 27: 1159-61.