Supplemental Materials and Methods

Human participants

Participants included in this study were selected from the Studies of the Etiology of Rheumatoid Arthritis (SERA) cohort, described previously*(74)*. Briefly, SERA is a cross-sectional and longitudinal study that follows and studies individuals at-risk for future development of rheumatoid arthritis (RA). At-risk status is defined as individuals with positive concentrations of circulating RA-specific autoantibodies (serum anti-cyclic citrullinated peptide (CCP)3/anti-CCP3.1 positive or ≥2 rheumatoid factor (RF) isotypes), or first-degree relatives (FDRs) of a patient with RA. All at-risk individuals were without a history or findings of inflammatory arthritis at the time of sample acquisition. Early RA was defined as having received a diagnosis of seropositive RA within the past 12 months. Controls were defined as having a negative CCP3, CCP3.1, and RF; not having an FDR with RA or a personal history of autoimmune rheumatic disease; and no inflammatory arthritis by history or on exam at the time of the study.

Participants were recruited at the University of Colorado Anschutz Medical Campus with approval from the Colorado Multiple Institutional Review Board (IRB#: 01-675) or from Benaroya Research Institute (BRI) with approval from the BRI Institutional Review Board (IRB#: IRB07109-139). During study visits, all participants were asked to complete questionnaires that assessed basic demographics, self and family history of disease, and past and current environmental exposures. A 68-joint examination was performed by a rheumatologist or trained study nurse in the at-risk and control individuals to confirm the absence of inflammatory arthritis. Blood for serum draws was drawn into serum separation tubes (Fisher Scientific BD Vacutainer), allowed to clot for 15 minutes at room temperature and then centrifuged at 3000 rpm for 10 minutes. Blood for cellular assays was collected using sodium heparin as an anticoagulant. Peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll underlay, cryopreserved in 7% dimethyl sulfoxide (DMSO), and subsequently thawed to perform T cell assays. A fecal collection kit (BD GasPak EZ Anaerobe Pouch System with Indicator) was provided to the participant for self-collection and overnight shipment on frozen gel packs by FedEx to the study

site. Serum and fecal samples were aliquoted into 2 ml graduated tubes and stored at −80°C until the sample was utilized for further analysis.

Serum Autoantibody Testing

Anti-CCP3 (IgG, Inova Diagnostics) and anti-CCP3.1 (IgG/IgA, Inova Diagnostics) enzyme-linked immunosorbent assays (ELISAs) were performed and analyzed according to the manufacturers' instructions. Results are reported in units/mL. RF IgG and IgA isotypes were measured by ELISA using QUANTA Lite kits (Inova Diagnostics), and results are reported in international units/mL.

Human Plasmablast Monoclonal Antibody Expression

As described previously*(28)*, dual IgA/IgG family plasmablasts (n=94) were isolated from 4 individuals at-risk for RA, defined as serum RF+ only (n=1), anti-CCP+ only (n=1), and RF and anti-CCP+ (n=2), as well as from 2 anti-CCP+ individuals with early RA. Determination of plasmablasts belonging to a clonal family was based on shared, International ImMunoGeneTics (IMGT)-based assignments of heavy chain and light chain V-J-region genes and 60% identity matching. A fraction of these plasmablasts were selected for cloning and immortalization due to their representation of shared IgG/IgA clonal families or their ability to bind citrullinated peptide tetramers. Production of the monoclonal antibodies was performed using an Expi293 Expression System (Thermo Fisher) with Expi293F cells, as previously described*(75-77)*. Fab domains from the plasmablast-derived monoclonal antibodies (PB-mAbs), selected for binding of RA-relevant antigens on a protein microarray, were expressed in a mouse IgG2a scaffold. Antibodies were recombinantly produced in Dr. Robinson's laboratory at Stanford University or by staff at LakePharma. The variable region sequences of these antibodies are shown (data file S2).

RA-Related Autoantigen Array

Titers of IgG antibodies targeting RA-associated autoantigens (n=346) were measured using a previously described Bio-Plex assay*(78)*. Analysis was performed on a Luminex 200 instrument running Bio-Plex

Manager software version 6.1. 94 PB-mAbs isolated from six individuals (n=4 at-risk for RA and n=2 early RA) were analyzed. Subsequently, thirty mice were analyzed across two timepoints utilizing this approach.

Fecal Bacteria Pool

Feces from 5 controls, 8 at-risk (n=3 non-FDR CCP+, n=2 FDR CCP+, n=3 FDR CCP-), and 5 early RA individuals were utilized to create a pool. Inclusion of samples was based on fecal 16S sequencing data from these individuals, with a goal of being broadly representative of bacteria across all study groups. 50 mg of fecal material from each sample was placed onto ceramic bead tubes (MP Biomedicals) and were rehydrated in 1mL sterile phosphate-buffered saline (PBS). Each sample was homogenized by bead beating for 15 minutes at full speed. Samples were then centrifuged at 50 x g to remove debris, and the supernatants from each sample was collected into a combined tube. The combined bacteria were washed 3 times with sterile PBS containing 1% (w/v) bovine serum albumin (BSA, Sigma-Aldrich) and spun at 14,000 x g to pellet bacteria. In order to verify the bacterial composition of this pool, DNA was extracted and 16S rRNA sequencing was performed per the methods below.

Plasmablast-Bound Bacteria Sequencing

100 µL of fecal bacteria pool (400 mg feces/mL), or 5 x 10^6 colony-forming units (CFU) cultured bacteria, were combined with 0.5 µg of PB-mAbs, and was incubated on ice for 30 minutes. The samples were washed 3 times with sterile PBS containing 1% (w/v) BSA. 0.5 µg of phycoerythrin (PE)-conjugated rat anti-mouse IgG2a secondary antibody (Invitrogen) was added to each sample, along with 1:4000 nucleic acid dye Syto9Green (Invitrogen). Negative controls included in-house generated PB-mAbs against the 2010/2011 seasonal trivalent influenza vaccine (H1N1 A/California/7/2009, H3N2 A/Perth/16/2009, and B/Brisbane/60/2008)*(79)* and *Borrelia burgdorferi.* The positive control was polyclonal goat anti-*E. coli* antibody (Invitrogen) with PE mouse anti-goat (SouthernBiotech) secondary

antibody. Each sample was incubated again for 30 minutes on ice and underwent the same washing step. Samples were analyzed by flow cytometry on a BD LSR2 flow cytometer and FlowJo v10 software. A positive mAb-bacteria binding cut-off was established as greater than 2 standard deviations (SD) above the mean fluorescence intensity of the negative controls. Those mAbs with positive binding then underwent subsequent flow cytometric sorting into mAb coated and uncoated fractions utilizing an Astrios EQ 5 laser flow sorter (Beckman Coulter). Sorted bacterial DNA was extracted and 16S rRNA sequenced.

DNA Extraction

50 mg of fecal material from each sample was placed onto ceramic bead tubes (MP Biomedicals) with 1mL sterile PBS. For samples that had undergone flow sorting, the mAb-coated and uncoated fractions were spun at 14,000 x g for 30 minutes, and supernatant was removed to create a total sample volume of 100 µL that was added to ceramic bead tubes. Samples were homogenized by bead beating for 10 minutes at maximum speed and then centrifuged at 50 x g for 15 minutes at 4°C to remove debris. Supernatants were harvested and centrifuged at 14 000 x g for 15 minutes at 4°C to pellet the bacteria. Supernatant was removed, and the bacteria were washed 3 times with sterile PBS containing 1% (w/v) BSA (Sigma-Aldrich) and spun at 14 000 x g to pellet bacteria. Total bacterial genomic DNA extractions were performed using the Qiamp Power Fecal DNA prep kit (Qiagen) according to the manufacturer's instructions.

16S rRNA Gene Sequencing and Analysis

DNA was amplified by polymerase chain reaction (PCR) with broad-range bacterial primers targeting the 16S ribosomal RNA (rRNA) gene hypervariable regions V1 and V2 and pooled amplicons were subjected to Illumina MiSeq sequencing, as previously described*(80-85)*. Assembled sequences were aligned and classified with SINA (1.3.0-r23838)*(86)* using the 418,497 bacterial sequences in Silva 115NR99*(87)* as reference configured to yield the Silva taxonomy. Operational taxonomic units (OTUs) were produced by

clustering sequences with identical taxonomic assignments. OTUs with greater than 0.01% abundance in any sample and observed in greater than 5% of the samples were included in further analyses. Analyses of OTU relative abundance and biodiversity were conducted using Explicet software*(88)*. Microbiome data were analyzed using Explicet and R statistical software, including MicrobiomeAnalyst*(89)*.

Bacterial Isolation

Primary bacterial isolates were established from a fecal sample of an at-risk individual by serially diluting the stool sample 1:10 in sterile oxygen-reduced Mega media*(90)*, adding 1 g of feces to 10mL of media, in a Coy anaerobic chamber. The diluted sample was homogenized for 5 minutes by vortexing, and then solids were allowed to settle for 5 minutes. Serial 1:10 dilutions were carried out to 10^{-9} dilutions in Mega media. For the 10^{-7}, 10^{-8}, and 10^{-9} dilutions, 170 µL of media was plated in each well of a 96 well plate. Each 96 well plate was sealed and allowed to incubate in anaerobic conditions at 37°C for 5 days. The dilution factors that resulted in 30% or fewer wells being turbid at the end of the incubation had less than a 5% chance of having multiple bacterial clones in each well. To ensure strain purity, each established isolate was cultured on a sheep blood agar (SBA) plate under anaerobic conditions for 3 days and then had one isolated colony harvested and grown to larger volume in sterile Mega media under anaerobic conditions. Each isolate was stored at -80°C in 25% sterile oxygen-reduced glycerol.

Taxonomic Identification

Each bacterial isolate was taxonomically identified using bacterial group primers for *Lachnospiraceae* and *Ruminococcaceae*. The primer sequences were 5'→3' Forward: CGGTACCTGACTAAGAAGC and Reverse: AGTTT(C/T)ATTCTTGCGAACG*(91)* as compared to global 16S bacterial primers, which were 5' → 3' rpoB1698 Forward: AACATCGGTTTGATCAAC and rpoB2041 Reverse: CGTTGCATGTTGGTACCCAT*(92)*. The cycling protocol is as follows: 50°C for 2 minutes, 95°C for 10 minutes, and then 40 cycles of 95°C for 15 seconds to 60°C for 1 minute. Isolates whose DNA amplified with the *Lachnospiraceae/Ruminococcaceae* primers were further 16S sequenced for confirmation.

Confirmed isolates were subsequently whole genome sequenced by Novogene, Inc. using a 350 bp insert DNA library and an Illumina Platform PE150. Short reads were cleaned and contigs and scaffolds assembled using Abyss*(29)*. Long reads were verified as derived from order *Clostridiales* through NCBI-BLAST. Assembled long reads were aligned to a reference genome, *Clostridiales* strain MGYG-HGUT-02424 (Genome accession number GCF_902387115.1, located at https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/902/387/115/GCF_902387115.1_UHGG_MGYG-HGUT-02424/) and analyzed using Artemis*(93)*.

Mice

DBA/1j mice were obtained from The Jackson Laboratory and derived germ-free by Taconic by embryonic transfer into germ-free Swiss Webster hosts. Pregnant dams were shipped to the University of Colorado Anschutz Medical Campus Gnotobiotic Core facility, where a colony of germ-free DBA/1 mice was maintained in sterile vinyl isolators. Mice were age and sex matched for all studies. Between 6 and 10 weeks of age (mean=8.2), mice were orally gavaged with 10^7 CFU in 200 µl of primary *Subdoligranulum* species isolates 1 or 7, *Prevotella copri* (DSM 18205), or sterile growth media. This strain of *P. copri* was utilized due to its colonization efficacy as previously determined*(94)*. Mono-colonized mice were monitored weekly for stable bacterial colonization; fecal pellets were collected, DNA was extracted, and each sample was amplified using global 16S primers. Furthermore, cecal contents collected at termination were serially diluted and plated on SBA and incubated for 48 hours at 37°C in anaerobic conditions. Bacterial concentrations were determined by quantifying the number of CFUs at each dilution factor.

Mice that were treated with broad spectrum antibiotics were given an oral mixture of 1mg/mL ampicillin and neomycin, and 0.5mg/mL vancomycin and metronidazole in drinking water for eight days prior to bacterial gavage. The microbiome was confirmed to be depleted by eubacterial quantitative (q)PCR on fecal pellets prior to gavage with *Subdoligranulum* isolate 1, isolate 7, or *P. copri*.

Mice were monitored for arthritis development similar to that performed for the collagen-induced arthritis (CIA) model*(6)*, though mice in these studies were not injected with collagen or other immune stimulus, but solely gavaged with bacteria. Arthritis severity was measured as a mean clinical score for each of the animal's four paws, where 0 = normal, 1 = erythema, 2 = swelling, and 3 = ankylosis. At the specified time points, feces, serum, and tissues were harvested from mice. All animal experiments were approved by the University of Colorado School of Medicine Institutional Animal Care and Use Committee.

Fluorescein isothiocyanate (FITC)-Dextran Flux

Mice were orally gavaged with 0.6 mg/kg body weight 4 kDa dextran labeled with fluorescein (Sigma) and serum was collected 4 hours later. The amount of fluorescence was measured with a fluorimeter (Promega) at 485/530 nm. A standard curve was generated to calculate the amount of dextran that was present in the serum.

Histopathology

Whole colon and ileal tissue was harvested from mice, flushed with PBS, dissected longitudinally, and pinned in wax for fixation in formaldehyde overnight. Tissues were then embedded in paraffin. Sections of 5 μm were cut and stained with hematoxylin and eosin (H&E). Four high-powered fields of well-oriented colon tissue per mouse were analyzed at 40X magnification for quantification of tertiary lymphoid structures, crypts per high-powered field, and crypt height and width.

Mouse paws were removed at mid-limb and fixed in 10% paraformaldehyde. The bones were decalcified in 14% EDTA for 10 weeks and then embedded in paraffin. Sections of 5μm were cut from paraffin embedded tissues and stained with H&E. Pathology was assessed in a blinded manner by a trained pathologist for synovitis, osteomyelitis, vasculitis, subcutaneous/muscle/periosteal inflammation, and fat inflammation, and global incidence rate of pathological findings were reported.

Complement C3 Immunohistochemistry

Paraffin-embedded tissue slides were assessed for C3 complement deposition in the joints. Each slide was exposed to the following: xylene (histology grade, Sigma-Aldrich) for 5 minutes (x2), 100% ethanol for 1 minute (x2), 95% ethanol for 1 minute (x2), tap water for 2 minutes, wash buffer (Dako Cytomation) for 2 minutes, 3% hydrogen peroxide (Millipore-Sigma) for 5 minutes, wash buffer for 2 minutes (x2), serum-free protein block (Dako Cytomation) for 5 minutes, goat anti-mouse complement C3 (1:10,000) (Fisher) overnight at 4°C, wash buffer for 2 minutes (x2), goat probe (BioCare Medical) for 15 minutes, wash Buffer for 2 minutes (x2), goat polymer (BioCare Medical) for 15 minutes, wash buffer for 2 minutes (x2), DAB+ (Dako Cytomation) for 5 minutes, distilled water for 2 minutes (x2), hematoxylin (4-5 dips), distilled water for 1 minute, 95% ethanol for 1 minute, 100% ethanol for 1 minute (x2), xylene for 1 minute (x2), and then were mounted and coverslipped. Each stained section was assessed for complement deposition in the joint and was scored in a blinded fashion from 0 to 3 based on C3 deposition severity: 0=no staining, 1=mild staining, 2=moderate staining, 3=intense staining.

IgG and IgA Immunohistochemistry

Paraffin-embedded tissue slides were assessed for IgA and IgG deposition in the joints. Each slide was exposed to the following: xylene (histology grade, Sigma-Aldrich) for 5 minutes (x2), 100% ethanol for 1 minute (x2), 95% ethanol for 1 minute (x2), tap water for 2 minutes, wash buffer (Dako Cytomation) for 2 minutes, 3% hydrogen peroxide (Millipore-Sigma) for 5 minutes, wash buffer for 2 minutes (x2), serum-free protein block (Dako Cytomation) for 5 minutes, anti-mouse IgA-horseradish peroxidase (HRP) 1:500 or anti-mouse IgG-HRP 1:500, respectively, wash buffer for 2 minutes (x2), 3,3′-Diaminobenzidine (DAB)+ (Dako Cytomation) for 5 minutes, distilled water for 2 minutes (x2), hematoxylin (4 to 5 dips), distilled water for 1 minute, 95% ethanol for 1 minute, 100% ethanol for 1 minute (x2), xylene for 1 minute (x2), and then were mounted and coverslipped. Each stained section was assessed for IgA and IgG deposition in the joint and was scored in a blinded fashion from 0 to 3 based on IgA/IgG deposition severity. 0=no staining, 1=mild staining, 2=moderate staining, 3=intense staining.

Type II Collagen ELISA

Titers of IgG targeting Type II collagen was determined by ELISA. All steps were prepared on ice. Collagen dilution buffer (Chondrex) was prepared and ELISA-grade type II murine collagen (Chondrex) was diluted 1:100 and adsorbed to high protein binding ELISA plates overnight. Plates were washed with well wash solution (PBS + 0.05% Tween 20) three times, and then blocked with blocking buffer (PBS + 2% w/v BSA) for 4 hours at 4°C. Pooled serum from day 35 CIA mice were diluted and used as a standard curve. Serum samples were diluted 1:2000 for analysis. Samples were adsorbed to the plate and were incubated overnight at 4°C. The plate was washing three times with well wash solution. A 1:1000 solution of anti-murine IgG HRP was prepared and bound to the plate. The plate was incubated for four hours at 4°C. The plate was washed three times with well wash solution. Tetramethylbenzidine (TMB) substrate was added to the plate and development was halted with the addition of 1N H_2SO_4. Plates were read at 450nm.

Immunophenotyping Flow Cytometry

Spleen, mesenteric lymph nodes (MLNs), and Peyer's patches (PPs) were harvested from mono-colonized mice at days 14 and 35 following gavage. Tissues were processed by homogenizing in RPMI-1640 media and straining through a 70 micron filter. Cells were pelleted by centrifugation at 500xg for 8 minutes. Splenic cells underwent red blood cell lysis using a commercially available buffer (eBioscience) for 10 minutes and were then centrifuged again. Cells were then resuspended in 1mL FACS buffer consisting of PBS with 5% fetal bovine serum (FBS) for cell counting. 10^6 cells were added to each staining tube. For intracellular stains, cells were permeabilized and fixed using a commercially available buffer (Tonbo Biosciences). Table S8 lists the antibodies and clones used. The panel was validated through antibody titration, full minus one (FMO) controls, and isotype controls. Cytometric analysis was performed on a Cytek spectral cytometer, and downstream analysis was performed using FlowJo v10 software.

Fluorescence in situ hybridization (FISH)

Colons were harvested from mono-colonized mice at 35 days following gavage, and fixed in methacarn (60% methanol, 30% chloroform, 10% glacial acetic acid) overnight. Tissue was then washed in xylene and placed in 100% ethanol. Colons were paraffin embedded and then sectioned for FISH. For FISH, slides were deparaffinizes at 72°C for 30 minutes, then washed twice in xylene for 2 minutes, then washed twice in 100% ethanol for 2 minutes, then washed twice in 95% ethanol for 2 minutes. The slides then underwent ultraviolet C (UVC) crosslinking in PBS. The Eubacterial FISH probe (EUB-338; GCTCCTCCCGTAGGAGT, 5' labeled with Alexa Fluor (AF) 568) was applied in hybridization buffer (20mM Tris-HCl, 0.1% sodium dodecyl sulfate, 0.9M NaCl pH 7.2, 5% formamide). The slides were allowed to hybridize in a hybridization chamber overnight at 50°C. They were then removed and washed in wash buffer (20mM Tris-HCl, 0.9 NaCl pH 7.2) for 30 minutes. Slides were counterstained with wheat germ agglutinin (AF488) and 4'6'-diamidino-2-phenylindole (DAPI), and mounted with ProLong antifade mounting media. To quantify the number of bacteria in the host epithelium, the total number of bacteria in the epithelium across 10 fields of view was quantified by a reviewer blinded as to the treatment group. In order to determine host mucus area in mm^2, mucus area was quantified and averaged across 10 fields of view.

Immunofluorescence on Colonic Sections

Colonic sections fixed in methacarn as described above were heated at 70°C for 15 minutes, then rinsed in xylene for 2 minutes, rinsed in 100% ethanol for 2 minutes, rinsed in 95% ethanol for 2 minutes, and rinsed in 70% ethanol for 2 minutes. The slides were rehydrated in PBS for 3 minutes. Antigen retrieval was performed in sodium citrate buffer (10mM sodium citrate, 0.05% Tween 20, pH 6.0) for 20 minutes between 90 and 100°C. The slides were cooled to room temperature and washed in PBS twice. The slides were blocked at room temperature for 50 minutes (PBS with 5% BSA, 005% Tween 20, 1:100 Fc block (Tonbo), and 1:100 rat serum), then incubated with target antibodies (B220 AF488 (R&D Systems) 1:100 and anti-IgA Cy5 (Bioss) 1:500) for 2 hours at room temperature. The slides were air dried and DAPI was

applied at 1 μg/mL. Slides were incubated at room temperature for 15 minutes and then mounted with Prolong antifade plus mounting medium.

Serum Transfer and Analysis of Arthritis

Serum collected from mono-colonized and control mice was pooled from 10 mice for each treatment group, respectively, and 150 μl was injected intraperitoneally (IP) into germ-free DBA/1 mice, as we know the background autoantibody concentrations for these mice. The mice were then monitored daily for the development of arthritis and scored in a blinded fashion as above. To determine the kinetics of the joint swelling, mice were monitored for 21 days post-serum transfer. The peak of joint swelling was determined to be 7 days post-serum transfer. Then, a second cohort of mice were dosed with serum, and were euthanized 7 days following serum transfer and paws were collected for histology.

Lymphocyte and Granulocyte Depletion Studies

CD4+ T cells, B cells, and neutrophils were depleted from germ-free DBA/1 mice by injection of targeting monoclonal antibodies. For CD4+ T cell depletion, mice were injected IP with 200 μg of antibody GK1.5 (BioXCell) two days prior to colonization, and then with 100 μg of antibody IP every four days thereafter. Mice were depleted of B cells using a CD20 depletion antibody (clone SA271G2, BioLegend). Mice were injected with 250 μg of antibody IP two days before bacterial administration, and then with 250 μg of antibody IP on day 21 of the study. Neutrophils and other myeloid-derived cells were depleted using a Ly6G/Ly6C antibody (clone GR1, BioXCell) injected IP with 200 μg two days before bacterial administration, and then with 200 μg of antibody IP every four days thereafter. Finally, 200 μg of control antibody (anti-Rat IgG2aK, clone RTK4530, BioLegend) was injected IP two days before bacterial administration, and then with 200 μg of antibody IP on day 21 of the study. Cellular depletion was confirmed in the MLNs, PPs, and spleens of mice. The depletions were efficacious at depleting the B cells by two-fold, the CD4+ T cells by ten-fold, and granulocytes by five-fold on average. Mice were

gavaged with 10^7 CFU of *Ruminococcaceae subdoligranulum* isolate 7 on the second day of the study and monitored for the development of arthritis for 35 days. Mice were then euthanized and tissues harvested.

Human T cell studies

Healthy individuals and individuals with RA were recruited with informed consent through the BRI rheumatic disease registry. Sample use was approved and monitored by the BRI Institutional Review Board (IRB number: IRB07109-139). All participants met the 1987 American College of Rheumatology criteria*(95)*. Clinical characteristics of the participants are summarized in tables S4 and S5.

PBMCs were isolated from whole blood by the BRI Clinical Core Laboratory. Briefly, PBMCs were separated from whole blood over Ficoll-Hypaque gradient, cryopreserved in heat inactivated FBS supplemented with 10% DMSO, and stored in liquid nitrogen. PBMCs were thawed in a 37°C water bath, washed in RPMI-1640 supplemented with 10% FBS and 0.001% DNase/Benzonase (Sigma-Aldrich), re-suspended in serum-free RPMI-1640 at a volume of 10^7 cells/mL, and allowed to rest in 37°C 5%-CO_2 incubator for 2 hours. Cells then were centrifuged, re-suspended in RPMI-1640 supplemented with 10% commercial human pooled serum and anti-CD40 antibody (Miltenyi Biotec) at 1µl/million cells, and plated at a volume of 500µl of 5 million PBMCs per well of a 48-well tissue culture plate. PBMCs were stimulated 14 hours in 37°C in the presence of either 0.1% DMSO or 50ng/mL isolate 7 or isolate 1, respectively. For major histocompatibility complex class II blocking, PBMCs were rested in either anti-human leukocyte antigen (HLA)-DR (clone L243) at 20µg/mL, or equal volume PBS, 30 minutes prior to stimulation.

After a 14-hour stimulation, cells were transferred to a 4 mL polypropylene FACS tube for surface stain with PE-CF594 mouse anti-CD154 (clone TRAP1BD Biosciences) and PE-Cy7 mouse anti-CD137 (clone CY1G4, BioLegend) for 15 minutes at room temperature, then enriched with anti-PE microbeads (Miltenyi) following manufacturer's standard protocol with 1% of PBMCs set aside before enrichment to calculate number of CD4+ cells present. Cells were then surface stained for 30 minutes at 4°C with the following antibodies: CD14 FITC, CD19 FITC, CD56 FITC, CD3 peridinin chlorophyll

protein (PerCP)-Cy5.5, CD4 brilliant violet (BV) 510, CD8 allophycocyanin (APC)-Cy7, CD69 AF647, CD45RA BV421, and CCR7 BV711 (table S9). Cells were then stained with Sytox green (Thermo Fisher Scientific) to exclude non-viable cells prior to flow cytometry analysis on a BD FACS Canto II. The gating scheme and panel validation can be found in fig. S5. For comparison of healthy and RA cases to strain 7 cells were surface stained for 30 minutes at 4°C with the following antibodies: CD14 PerCP-Cy5.5, CD16 PerCP-Cy5.5, CD19 PerCP-Cy5.5, CCR7 APC-Cy7, CD45RA AF700, CD69 AF647, CCR4 PE, CXCR5 brilliant blue (BB)515, CCR6 BV786, CD95 BV711, CD278 BV650, CXCR3 BV605, CD4 V500, CD27 BV421, PD-1 brilliant ultraviolet (BUV)737, and CD8 BUV395 (table S10), then stained with BD Viaprobe to exclude non-viable cells prior to flow cytometry analysis on a BD FACS Fortessa. Flow cytometry data were analyzed using FlowJo v10, SAS JMP statistical software V15, and GraphPad Prism 8.0. The frequency of activated CD4+ T cells was calculated as follows: F = (1,000,000 x activated events) / (100 x number of CD4+ T cell events from the pre-enriched fraction).

Murine T cell studies

Bone marrow was isolated from the long bones of healthy specific-pathogen-free DBA/1 mice. The bone marrow was suspended at a concentration of $1x10^7$ cells per well in 4mL of bone marrow dendritic cell culture media (RPMI-1640 + 10% FBS, 100U/mL penicillin/streptomycin, L-glutamine, 100mm sodium pyruvate, 20ng/mL GM-CSF). Cells were cultured for 6 days prior to T cell co-culture. Splenocytes were isolated from mice 35 days after bacterial gavage with either *Subdoligranulum* isolate 1 or isolate 7. CD4+ T cell isolation was performed (EasySep mouse T cell isolation kit, Stemcell), and were suspended at a concentration of $2.5x10^6$ cells/mL in culture media (RPMI-1640 + 10% FBS, 100U/mL penicillin/streptomycin, L-glutamine, 100mm sodium pyruvate). $2.5x10^5$ T cells were cocultured with $2.5x10^5$ BMDCs loaded with $1x10^6$ CFU of heat killed *Subdoligranulum* isolate 1, isolate 7, or without bacterial stimulation for 14 hours at 37°C. Protein transport was blocked with brefeldin A, and CellTrace Far Red was added as a proliferation dye.

After the 14-hour stimulation, cells were transferred to a 4 mL polypropylene FACS tube. Cells were then surface stained for 30 minutes at 4°C. The antibodies and clones used are listed in table S11, and the gating scheme can be found in fig. S14. The panel was validated through antibody titration and FMO controls. Cytometric analysis was performed on a Cytek spectral cytometer, and downstream analysis was performed using FlowJo v10 software.

Subdoligranulum isolate 7 specific qPCR

Regions of genomic DNA present in *Subdoligranulum* isolate 7 but absent in isolate 1 were identified. From these regions, the NCBI primer design tool was utilized to generate 30 sets of potentially isolate 7-specific primers. DNA was extracted from *Subdoligranulum* isolates 7 and 1 as described above. Isolate 1 and 7 DNA were screened utilizing the primer sets and compared to a universal bacterial primer, RpoB (table S12). Three primer sets that had a strong fold change expression over RpoB and isolate 1 were flagged for further optimization. Next, known quantities of isolate 7 were spiked into a human fecal sample that had previously not been identified as containing *Subdoligranulum* to create a standard curve. The curve ranged from $1x10^9$ CFU of isolate 7 per 100mg human feces to 0 CFU of isolate 7 per 100mg of human feces. The three primer sets of interest were utilized against this standard curve,and comparing to RpoB as a control. The best primer set was identified (table S12) and limit of detection of $1x10^5$ CFU of isolate 7 per 100mg feces was established based on qPCR Ct versus CFUs of Isolate 7 (Fig. 7A). A line of best fit was established for the curve (R^2=0.9982), and utilized for regression analysis of human and murine samples. Next, the feces of healthy control individuals (n=12), individuals at-risk for RA (n=12), and individuals with early RA (less than 1 year from diagnosis, n=12) (table S6) were screened for the presence of isolate 7 in 100mg of feces. These samples were compared to the standard curve and through regression analysis the CFUs of isolate 7 per 100mg of fecal weight was established. The feces of 12 specific-pathogen-free mice were screened for the presence of isolate 7 through the same method.

Supplemental Figures:

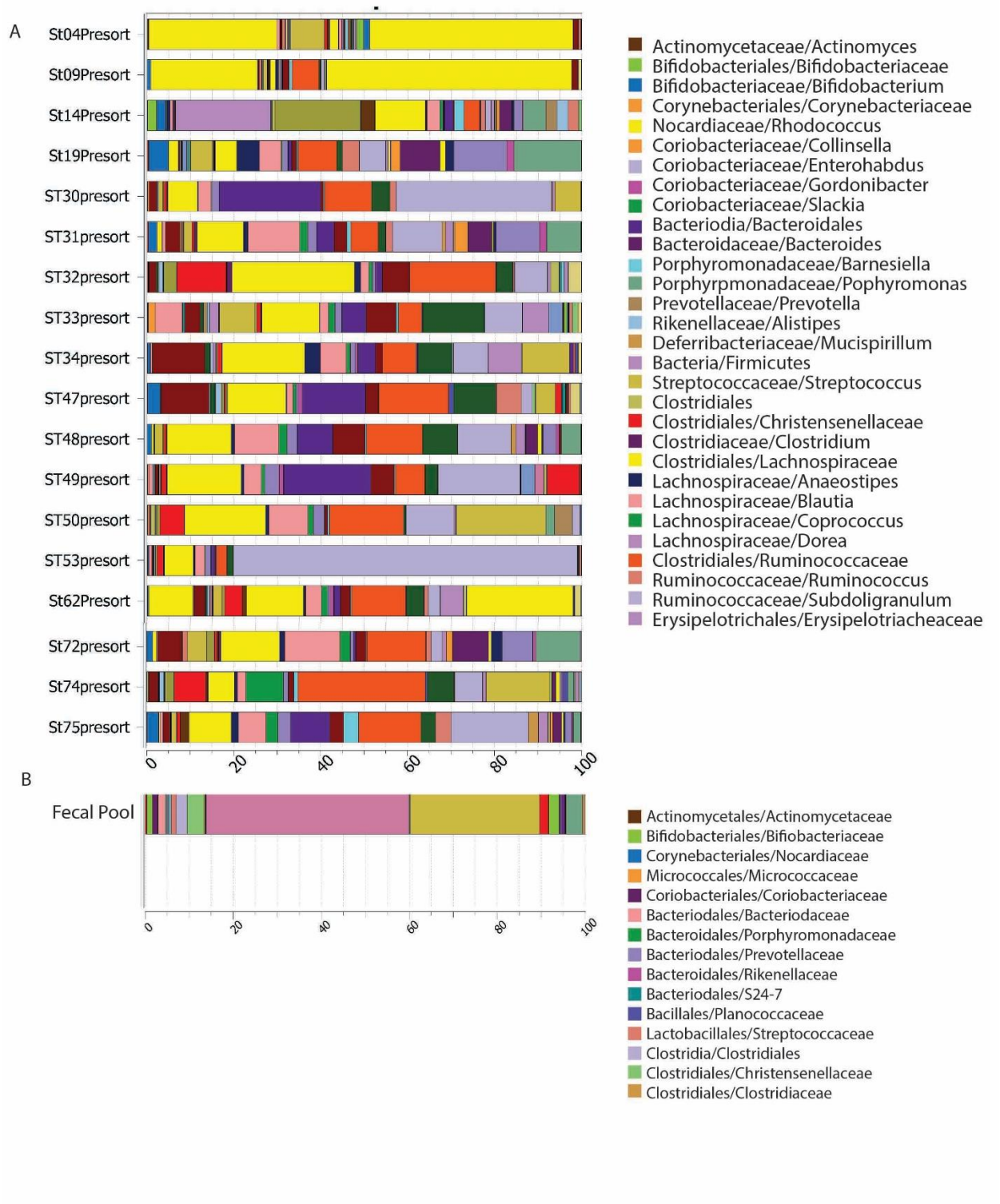


Fig. S1. 16S sequencing of each fecal pool sample.

Fecal samples were collected from healthy control individuals (n=5; ST30, 31, 32, 33, and 34), individuals at-risk for RA (n= 8; ST4, 9, 14, 19, 62, 72, 74, and 75), and individuals with early RA (n=5; ST 47, 48, 49, 50, and 53, less than l year from diagnosis). (**A**) Each sample was individually 16S sequenced and the OTU table graphed in a stacked bar format. The most highly represented taxonomic groups are represented (see figure legend to the right), displaying the percentage of the bacteria sequenced belonging to each taxa (x-axis) for each sample (y-axis). (**B**) A pool of fecal bacteria was created from these human fecal samples. The bacteria pool was l6S rRNA sequenced and its bacterial components are displayed in a bar format. The most highly represented taxonomic groups are presented (see figure legend to the right), displaying the percentage of the bacteria sequenced belonging to each taxa (x-axis) for the combined fecal bacterial pool.

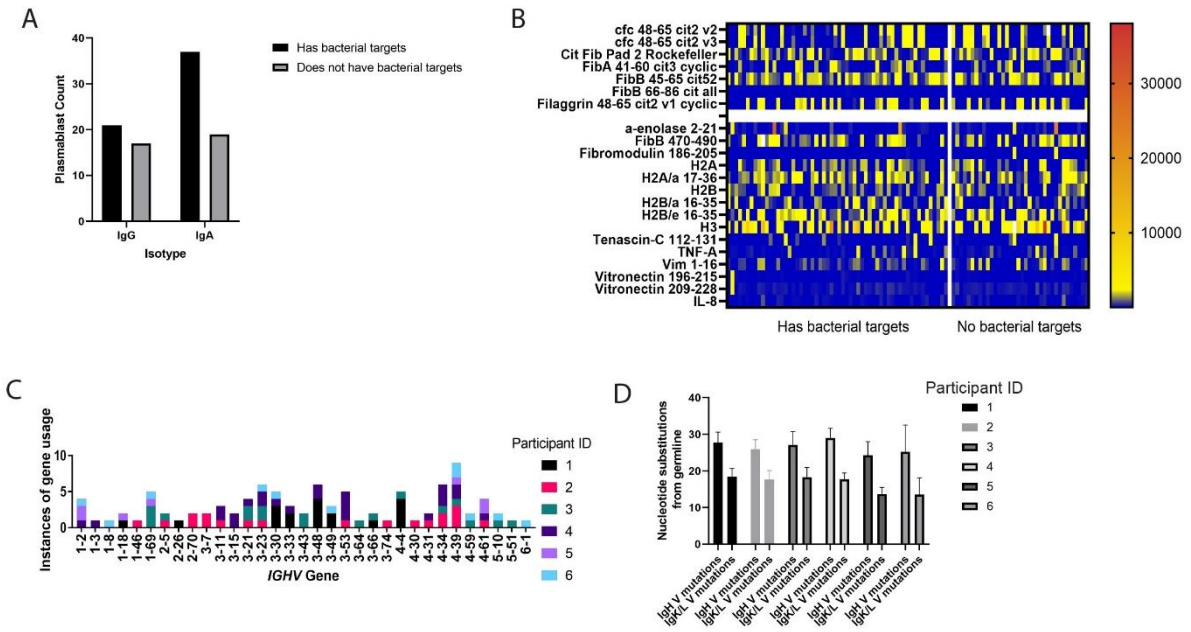
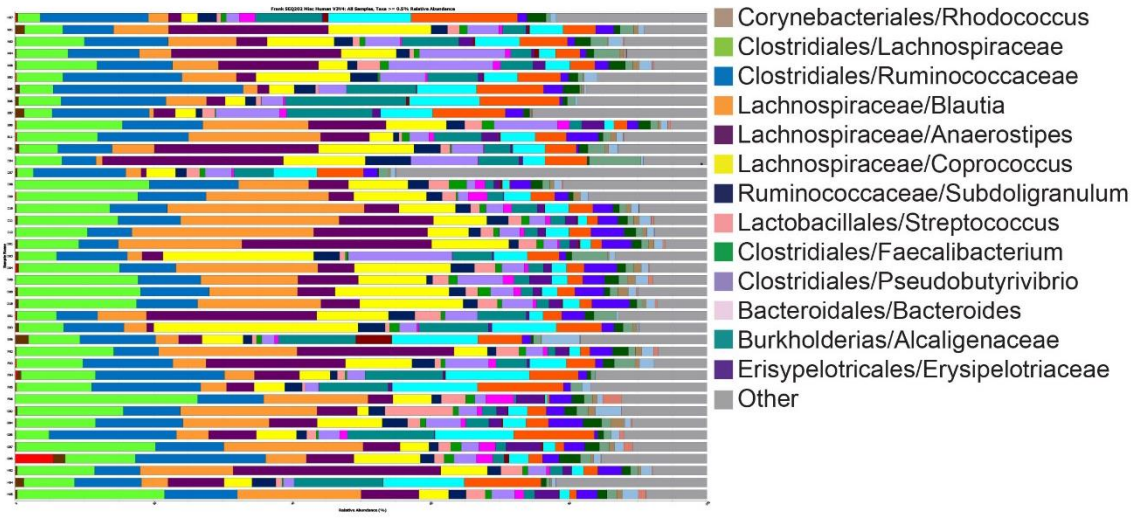


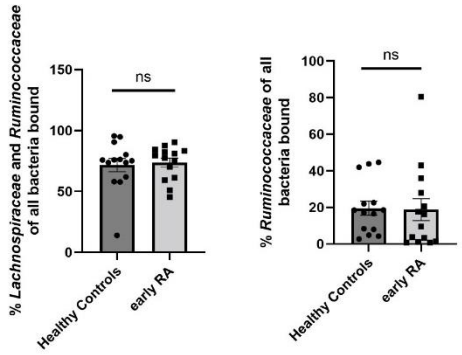
Fig. S2. Plasmablast cohort *IGHV* gene usage, mutations from germline, and bacterial binding characteristics.

(**A**) Displayed is the isotype of the each plasmablast antibody (IgG or IgA, x-axis) segregated by whether or not they target intestinal bacteria, with mAbs targeting bacteria shown in black bars and mAbs without bacterial targets shown in grey bars. The total mAb count for each group is displayed on the y-axis. No significance by Fisher's exact test. (**B**) RA-relevant autoantigenic targets are displayed as in Fig. 1A, but segregated by whether the plasmablasts (listed along the x-axis) have bacterial binding targets or not. 94 PB-mAbs from at-risk (n=4) and early RA (n=2) individuals belonging to shared IgG and IgA clonal families were applied to a planar array containing 346 different citrullinated and native peptide targets. The heatmap demonstrates degree of reactivity between PB-mAbs that have bacterial targets (x-axis, left) and PB-mAbs that don't have bacterial targets (x-axis, right) with specific antigens (y-axis). (**C**) The *IGHV* gene used by each plasmablast is displayed by individual from whom the PB-mAb was derived. The instances of gene usage is displayed on the x-axis, with the total mAb count displayed on the y-axis. (**D**) Instances of nucleotide substitutions from germline (y-axis) are demonstrated for each participant, as shown in the figure legend (x-axis). IgH V amino acid mutations are displayed on the left and IgL/K amino acid mutations are displayed on the right for each participant.

A


Corynebacteriales/Rhodococcus
Clostridiales/Lachnospiraceae
Clostridiales/Ruminococcaceae
Lachnospiraceae/Blautia
Lachnospiraceae/Anaerostipes
Lachnospiraceae/Coprococcus
Ruminococcaceae/Subdoligranulum
Lactobacillales/Streptococcus
Clostridiales/Faecalibacterium
Clostridiales/Pseudobutyrivibrio
Bacteroidales/Bacteroides
Burkholderias/Alcaligenaceae
Erisypelotricales/Erysipelotriaceae
Other

B


% Lachnospiraceae and Ruminococcaceae of all bacteria bound
150
100
50
0
ns
Healthy Controls
early RA
% Ruminococcaceae of all bacteria bound
100
80
60
40
20
0
ns
Healthy Controls
early RA

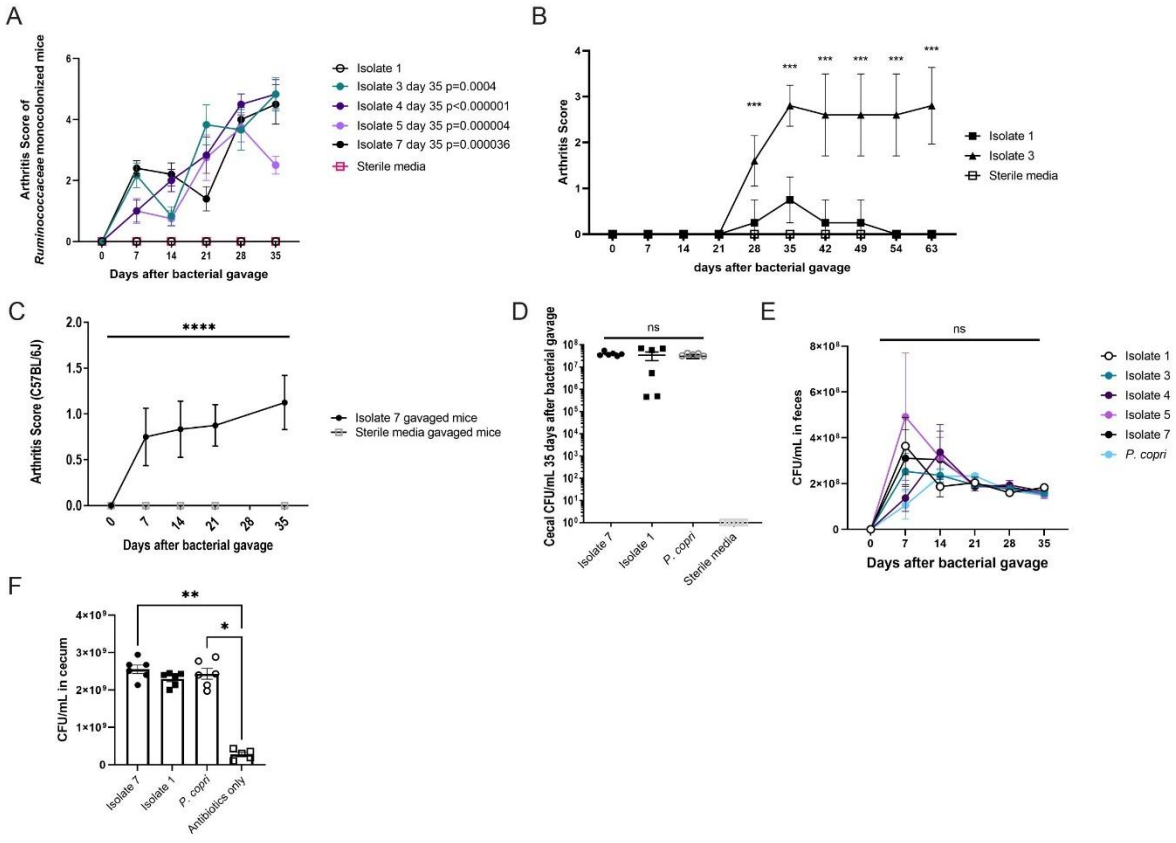
C


ST6


Actinomycetaceae/Actinomyces
Bacteriodia/Bacteroidales
Bacteroidaceae/Bacteroides
Porphyromonadaceae/Barnesiella
Porphyrpmonadaceae/Pophyromonas
Prevotellaceae/Prevotella
Rikenellaceae/Alistipes
Deferribacteriaceae/Mucispirillum
Bacteria/Firmicutes
Streptococcaceae/Streptococcus
Clostridiales
Clostridiales/Christensenellaceae
Clostridiaceae/Clostridium
Clostridiales/Lachnospiraceae
Lachnospiraceae/Anaeostipes
Lachnospiraceae/Blautia
Lachnospiraceae/Coprococcus
Lachnospiraceae/Dorea
Clostridiales/Ruminococcaceae
Ruminococcaceae/Ruminococcus
Ruminococcaceae/Subdoligranulum
Erysipelotrichales/Erysipelotriacheacea

Fig. S3. Plasmablast mAbs bind *Lachnospiraceae* and *Ruminococcaceae* from feces of healthy controls and individuals with RA.

(**A**) The ability of the plasmablasts to bind to the bacteria in the fecal pool was determined by flow cytometry. Each mAb was bound to the fecal pool; the bound bacterial fraction was separated by flow cytometric sorting and each bound fraction was 16S rRNA sequenced. Each sample was individually 16S sequenced and the OTU table graphed in a stacked bar format. The most highly represented taxonomic groups are represented (see figure legend to the right), displaying the percentage of the bacteria sequenced belonging to each taxa (x-axis) for each discrete mAb that bound bacteria (y-axis). (**B**) Four plasmablast-derived mAbs that demonstrated reactivity against *Lachnospiraceae* and *Ruminococcaceae* in a fecal pool (4, 28, 58, and 91) were pooled and screened for reactivity against individual fecal samples from healthy individuals (n=14) and individuals with early RA (n=14). The fecal samples were exposed to each antibody and analyzed by flow cytometry. The mAb bound bacterial fraction underwent 16S rRNA sequencing and taxonomic identification. The percentage of bacteria bound from *Ruminoccoaceae* or *Lachnospiraceae* are shown (y-axis), compared to group (x-axis). ns, not significant, unpaired t-test. Data are presented as mean±SEM. (**C**) The sample from which bacterial strains were isolated was 16S rRNA sequenced and the OTU table graphed in a stacked bar format.


A
Arthritis Score of Ruminococcaceae monocolonized mice
Days after bacterial gavage
Isolate 1
Isolate 3 day 35 p=0.0004
Isolate 4 day 35 p<0.000001
Isolate 5 day 35 p=0.000004
Isolate 7 day 35 p=0.000036
Sterile media
B
Arthritis Score
days after bacterial gavage
Isolate 1
Isolate 3
Sterile media
C
Arthritis Score (C57BL/6J)
****
Days after bacterial gavage
Isolate 7 gavaged mice
Sterile media gavaged mice
D
Cecal CFU/mL 35 days after bacterial gavage
ns
Isolate 7
Isolate 1
P. copri
Sterile media
E
CFU/mL in feces
ns
Days after bacterial gavage
Isolate 1
Isolate 3
Isolate 4
Isolate 5
Isolate 7
P. copri
F
CFU/mL in cecum
**
*
Isolate 7
Isolate 1
P. copri
Antibiotics only

1

Fig. S4. Mono-colonization of germ-free mice with *Ruminococcaceae* isolates results in a spectrum of joint swelling phenotypes.

(**A**) *Ruminococcaceae Subdoligranulum* Isolates 1, 3, 4, 5, and 7 were gavaged separately into germ-free DBA/1 mice. Mice were gavaged with either 5 x 10^6 CFUs of the cultured bacteria or sterile media. N= 6 mice gavaged with isolate 1, n=6 with isolate 3, n= 6 with isolate 4, n=4 with isolate 5, n=5 with isolate 7, and n=6 with sterile media. The mice were then observed weekly for 35 days for the development of joint swelling or ankylosis. The clinical score is shown (y-axis) relative to time post-bacterial gavage (x-axis) for each different isolate. (p-values for each of the isolates relative to sterile media-gavaged control mice at day 35 is shown in the figure legend, unpaired t-test). (**B**) *Ruminococcaceae Subdoligranulum* Isolate 1 and 3 were gavaged separately into germ-free DBA/1 mice. Mice were gavaged with either 5 x 10^6 CFUs of the cultured bacteria or sterile media (n=4 isolate 1-gavaged, n=5 isolate 3-gavaged, and n= 5 sterile media-gavaged mice). The mice were then observed weekly for 63 days for the development and maintenance of joint swelling or ankylosis. The clinical score is shown (y-axis) relative to time (x-axis) for each isolate. ***P<0.001, multiple unpaired t-test. (**C**) *Ruminococcaceae Subdoligranulum* isolate 7 was gavaged into germ-free C57BL/6J mice. Mice were gavaged with either 5 x 10^6 CFUs of the cultured bacteria (n= 8) or sterile media (n=8). The mice were then observed weekly for 35 days for the development of joint swelling or ankylosis. The clinical score is shown (y-axis) relative to time (x-axis) for each treatment group. ****P<0.0001 by repeated measures ANOVA. (**D**) To verify stable colonization of the mono-colonized mice, the ceca of mice were harvested and CFU/mL of bacteria was determined (n=6 mice from each group, respectively). The total CFU/mL at day 35 is displayed (y-axis) relative to each treatment group (x-axis). ns, not significant, unpaired t-test. (**E**) To verify stable colonization for each isolate, feces were collected weekly for 35 days. CFU/mL of bacteria was determined (n= 6 mice from each group, respectively). The total CFU/mL is displayed (y-axis) for each timepoint (x-axis) relative to each treatment group. ns, not significant, unpaired t-test. (**F**) To verify stable colonization in SPF mice gavaged with each isolate, the ceca of mice were harvested and CFU/mL of bacteria was determined (n=6 mice isolate 7-gavaged, n= 6 isolate 1-gavaged, n= 6 *P. copri*-gavaged, n=5 antibiotic treated only). The total CFU/mL at day 35 is displayed (y-axis) relative to each treatment group (x-axis). *P<0.05, **P<0.01, Kruskal-Wallis test with Dunn's post-test. For all plots, data are presented as mean±SEM.

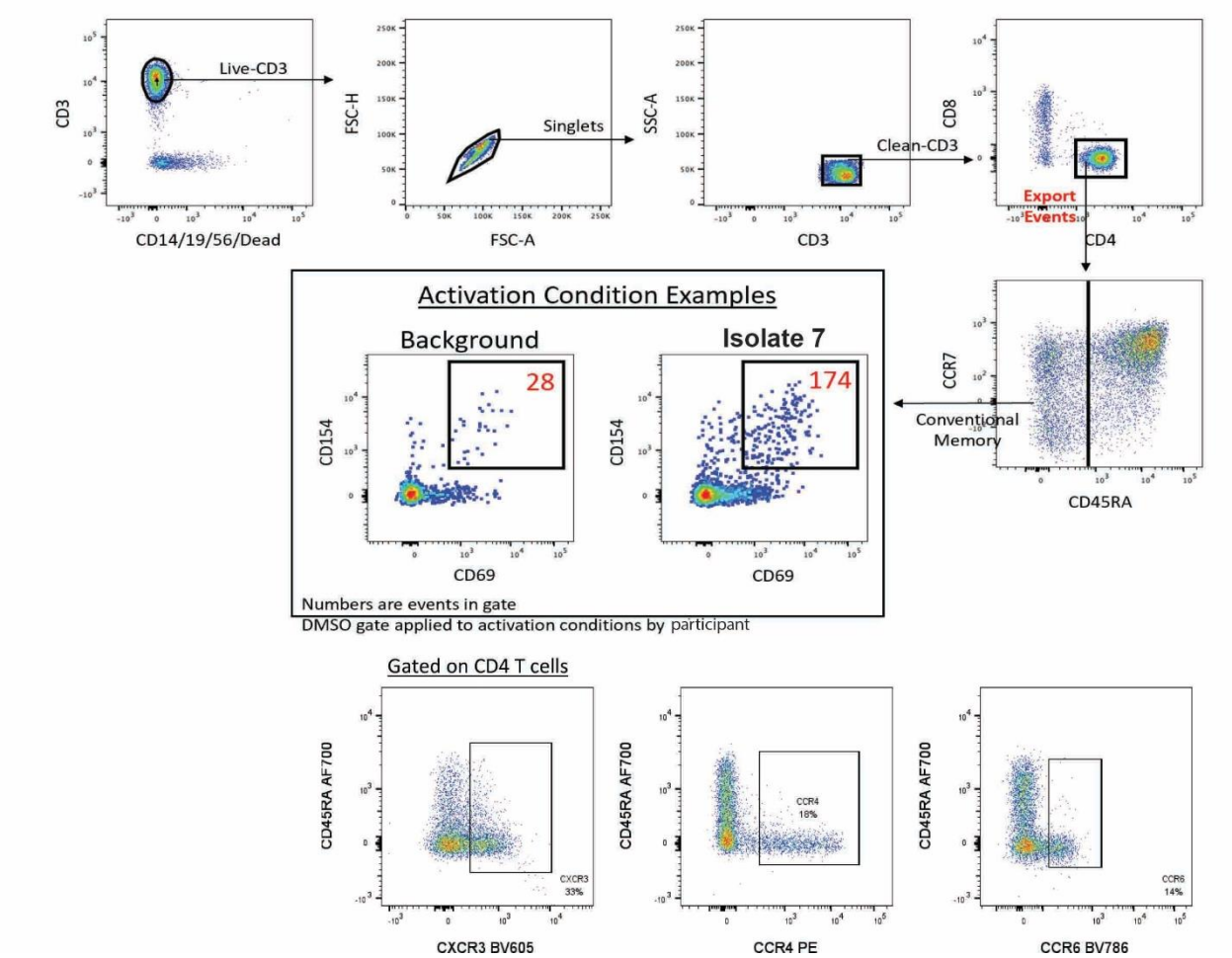


Fig. S5. Gating strategy and validation of stimulated CD4+ T cells.

PBMCs from individuals with RA (n= 11) were surface stained prior to stimulation for 30 minutes at 4 °C with the following antibodies: CD14/CD19/CD56 FITC, CD3 PerCP-Cy5.5, CD4 BV510, CD8 APC-Cy7, CD69 AF647, CD45RA BV421, and CCR7 BV71 l. Cells were then stained with Sytox green to exclude non-viable cells prior to flow cytometry. The gating scheme for this flow cytometric panel is shown, with arrows representing the direction of subsequent gating events. FSC-A, forward scatter area; FSC-H, forward scatter height; SSC-A, side scatter area.

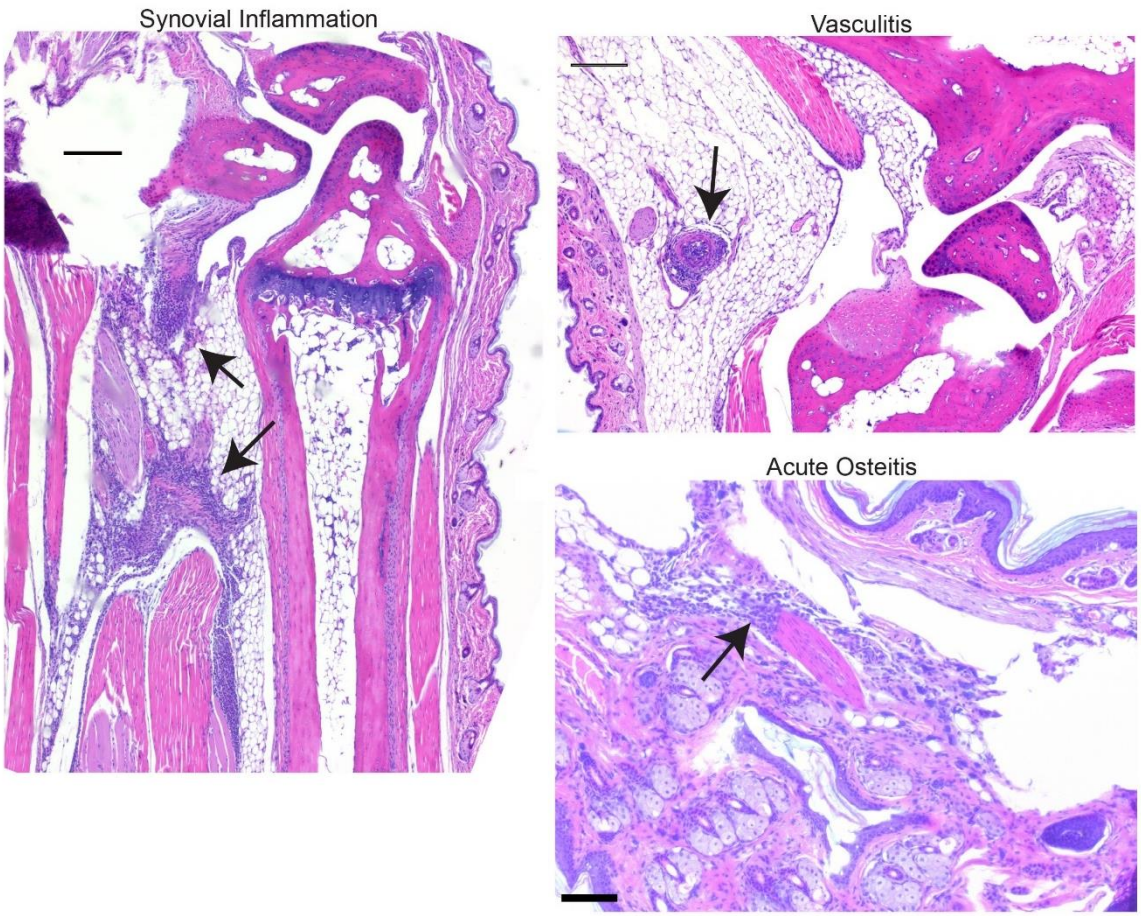


Fig. S6. Mono-colonized mouse paw pathology.

H&E staining was performed on decalcified paw sections from mice gavaged with isolate 7, isolate 1, *P. copri*, or sterile media and paw pathology was assessed by a pathologist in a blinded fashion. Evaluation of pathology included synovitis (top left photo, arrows), vasculitis (top right photo, arrow), and acute osteitis (bottom right photo, arrow) in isolate 7 mono-colonized mice. Scale bar is 100µm.

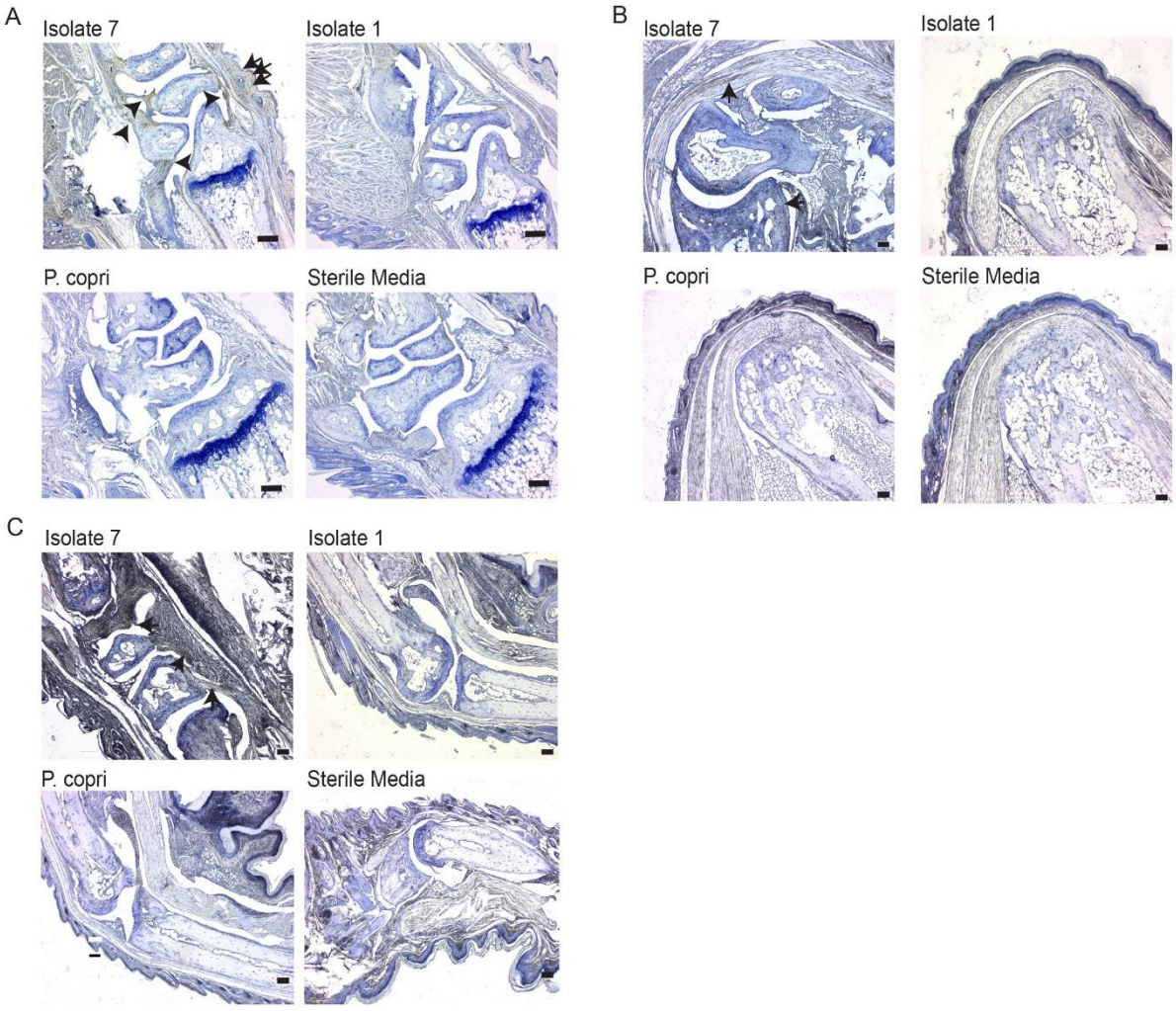


Fig. S7. Complement C3, IgG, and IgA deposition in paws of mono-colonized and sterile media-gavaged mice.

Immunohistochemistry using antibodies targeting the C3 component of the complement cascade, IgG Fc, or IgA Fc was performed on decalcified paw tissue sections. (**A**) C3 deposition, (**B**) IgG deposition, and (**C**) IgA deposition in representative sections is shown, with isolate 7 shown in the top left photo, isolate 1 shown in the top right photo, *P. copri* shown in the bottom left photo, and sterile media shown in the bottom right photo for each condition. Arrows denote regions of deposition, and scale bar is l00µm.

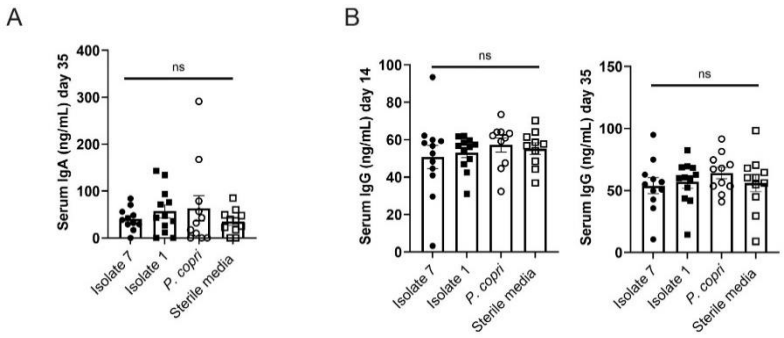


Fig. S8. Serum IgA and IgG of mono-colonized mice.

(**A**) Serum IgA and (**B**) serum IgG concentrations were determined from the mono-colonized mice at 14 days and 35 days after bacterial gavage. At day 14, n= l2 isolate 7-gavaged mice, n=l2 isolate 1-gavaged mice, n= 10 *P. copri*-gavaged mice, and n= 10 sterile media-gavaged mice were tested. At day 35, n= 11 isolate 7-gavaged mice, n= 12 isolate 1-gavaged mice, n= 11 *P. copri*-gavaged mice, and n= 11 sterile media-gavaged mice were tested. The total IgA and IgG are displayed in ng/mL (y-axis), separated by treatment group (x-axis). Data are presented as mean±SEM. ns, not significant as determined by Kruskal-Wallis test.

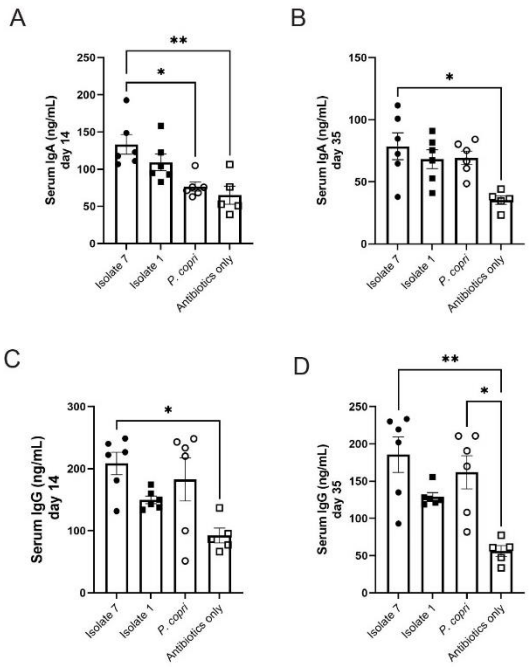


Fig. S9. Serum IgA and IgG of SPF mice gavaged with isolate 7.

Serum from SPF mice gavaged with either isolate 1 (n=6), isolate 7 (n=6), or *P. copri* (n=6), or treated with antibiotics only (n=5) was collected at days 14 and 35 after gavage. (**A and B**) The total serum IgA at 14 (A) and 35 (B) days after gavage was determined by ELISA; serum IgA is displayed (y-axis) against treatment group (x-axis). (**C and D**) The total serum IgG at 14 (C) and 35 (D) days after gavage was determined by ELISA. Symbols represent individual mice while bars are the mean±SEM. *P<0.05 and **P<0.01, Kruskal-Wallis test with Dunn's post-test.

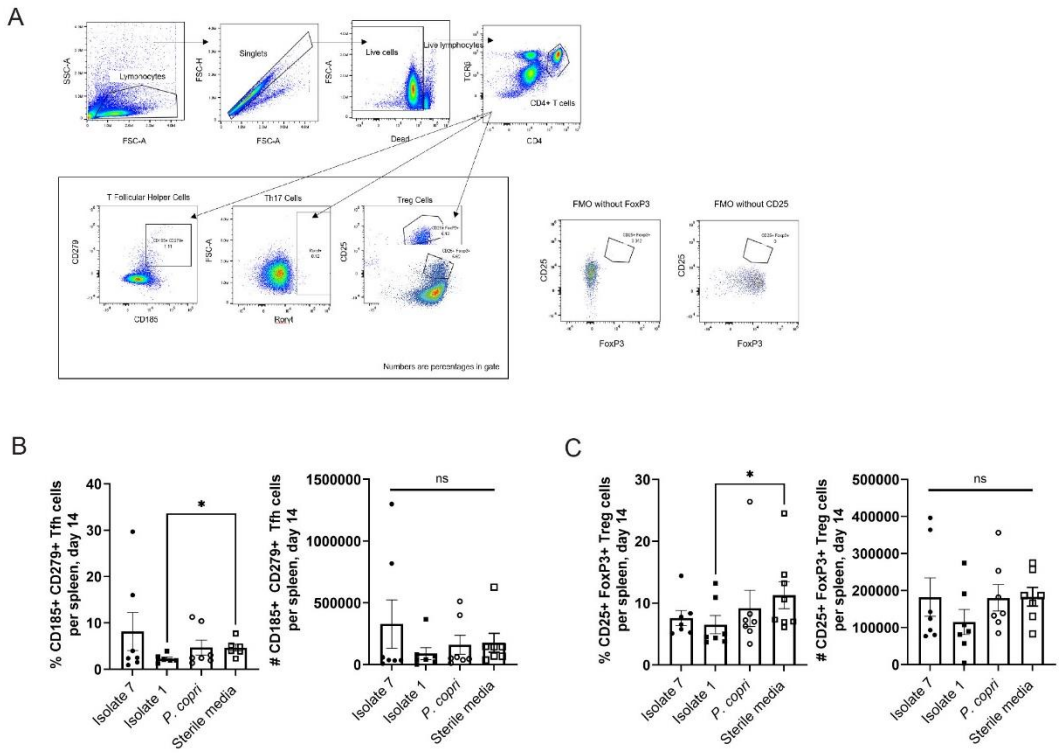


Fig. S10. T cell immunophenotyping panel gating scheme.

(**A**) The gating scheme for the T cell immunophenotyping panel is displayed. Names of gates and percentages of cells in each gate are displayed. The FMOs for FoxP3 and CD25 are shown to rationalize the gating scheme. (**B**) Splenic T cell populations were determined at day 14 after gavage. The percentage and absolute number of T follicular helper cells (Live T cell receptor (TCR) β+ CD4+ CD185+ CD279+ lymphocytes) is displayed (y-axis) compared to treatment groups (x-axis). *P<0.05 and ns, not significant, as determined by Kruskal-Wallis test with Dunn's post-test. (**C**) The percentage and absolute numbers of activated Tregs (Live TCRβ+ CD4+ CD25+ FoxP3+ lymphocytes) are displayed (y-axis) compared to treatment group (x-axis). *P<0.05 and ns, not significant, as determined by Kruskal-Wallis with Dunn's post-test. n=7 isolate 7-gavaged mice, n=7 isolate 1-gavaged mice, n=7 *P. copri*-gavaged mice, and n=8 sterile media-gavaged mice. Data in (B) and (C) are presented as mean±SEM.

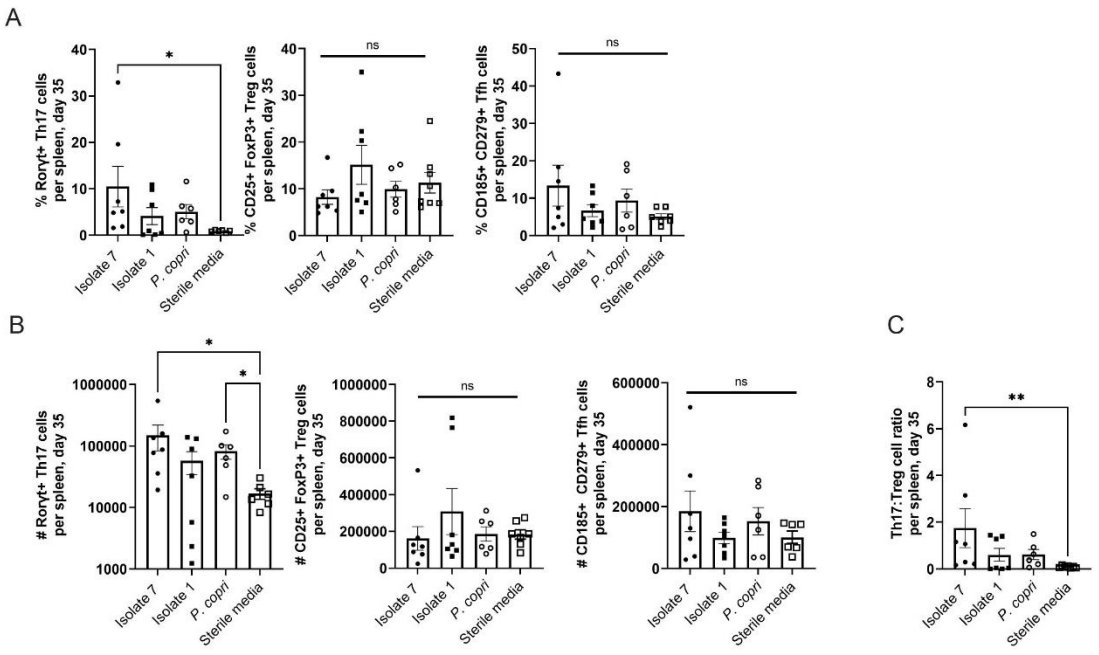


Fig. S11. Splenic immune cell populations of mono-colonized mice 35 days after bacterial gavage.

Splenic T cell populations of treated mice were determined at day 35 after bacterial gavage. (**A**) The percentage of Th17 cells (Live TCRβ+ CD4+ Rorγt+ lymphocytes), Tregs (Live TCRβ+ CD4+ CD25+ FoxP3+ lymphocytes), and T follicular helper cells (Live TCRβ+ CD4+ CD185+ CD279+ lymphocytes) are displayed (y-axis), compared to treatment group (x-axis). (**B**) The absolute numbers of Th17 cells, Tregs, and Tfh cells is displayed (y-axis) compared to treatment group (x-axis). (**C**) The ratio between Th17 cells and Tregs is shown (y-axis) compared to treatment group (x-axis). Data are presented as mean±SEM. *P<0.05, **p<0.01, and ns, not significant, as determined by Kruskal-Wallis with Dunn's post-test. n=7 isolate 7-gavaged mice, n=7 isolate 1-gavaged mice, n=6 *P. copri*-gavaged mice, and n=8 sterile media-gavaged mice.

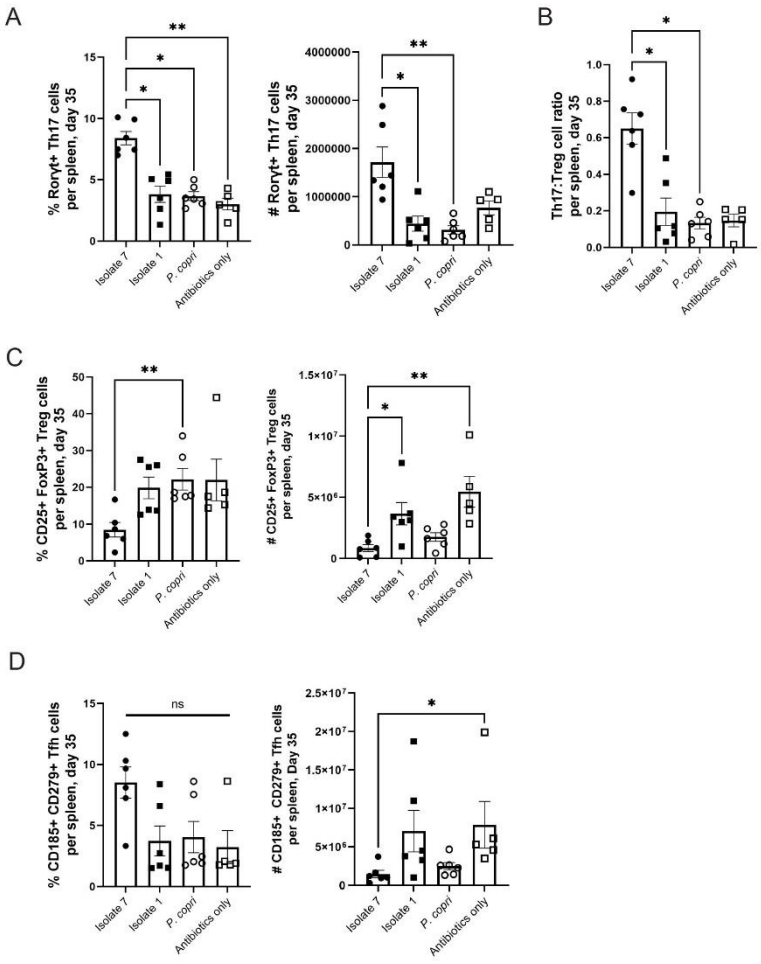


Fig. S12. Splenic immune cell populations of SPF mice 35 days after bacterial gavage.

Splenic T cell populations of SPF mice gavaged with bacteria were determined at day 35 after bacterial gavage. (**A**) The percentage and absolute number of Th17 cells (Live TCRβ+, CD4+ Rorγt+ lymphocytes) are displayed (y-axis), compared to treatment group. (**B**) Additionally, the ratio of Th17 cells to activated Treg cells (Live TCRβ+, CD4+ CD25+, FoxP3+ lymphocytes) is shown. (**C**) The percentage and absolute number of Tregs are displayed (y-axis) compared to treatment group. (**D**) The percentage and absolute number of Tfh cells (Live TCRβ+, CD4+ CD185+ CD279+ lymphocytes) are displayed (y-axis) compared to treatment group. Data are presented as mean±SEM. *P<0.05, **P<0.01, ns, not significant, as determined by Kruskal-Wallis test with Dunn's post-test. n=6 isolate 7-gavaged mice, n=6 isolate 1-gavaged mice, n=6 *P. copri*-gavaged mice, n=5 antibiotic-treated with no bacteria gavage.

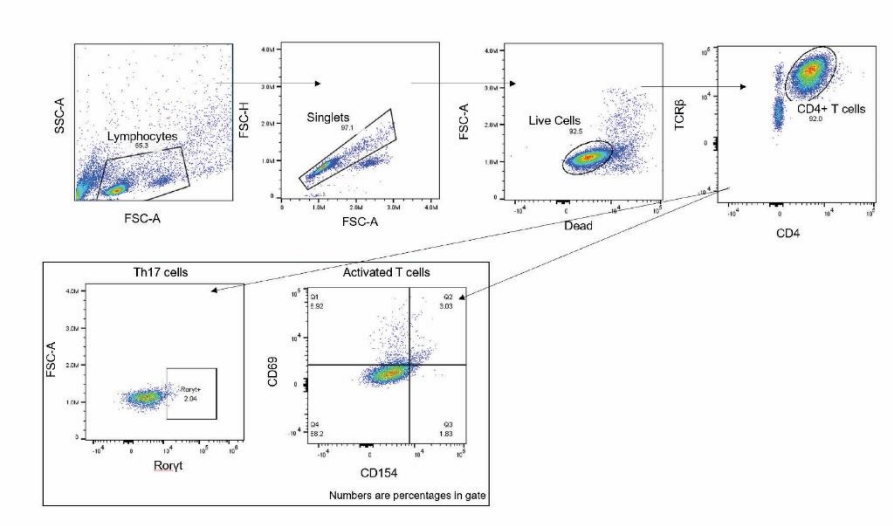


Fig. S13. Murine T cell stimulation panel gating scheme.

The gating scheme for the murine T cell stimulation panel is displayed here. Names of gates and percentages of cells in each gate are displayed.

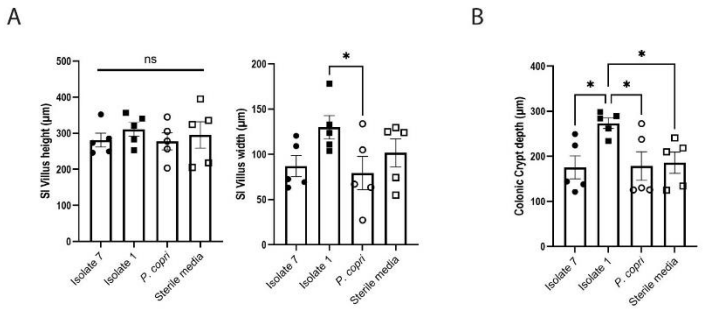


Fig. S14. Small intestinal villus and crypt properties in mono-colonized mice.

(**A**) Small intestine (SI) villus height and width are displayed (y-axis, in µm) and separated by treatment group (x-axis), as determined by intestinal histology on gavaged mice. (**B**) Colon crypt depth (in µm) is displayed. Data are presented as mean±SEM. *P<0.05 and ns, not significant, as determined by one-way ANOVA with Tukey's post-test. n=5 mice per group.

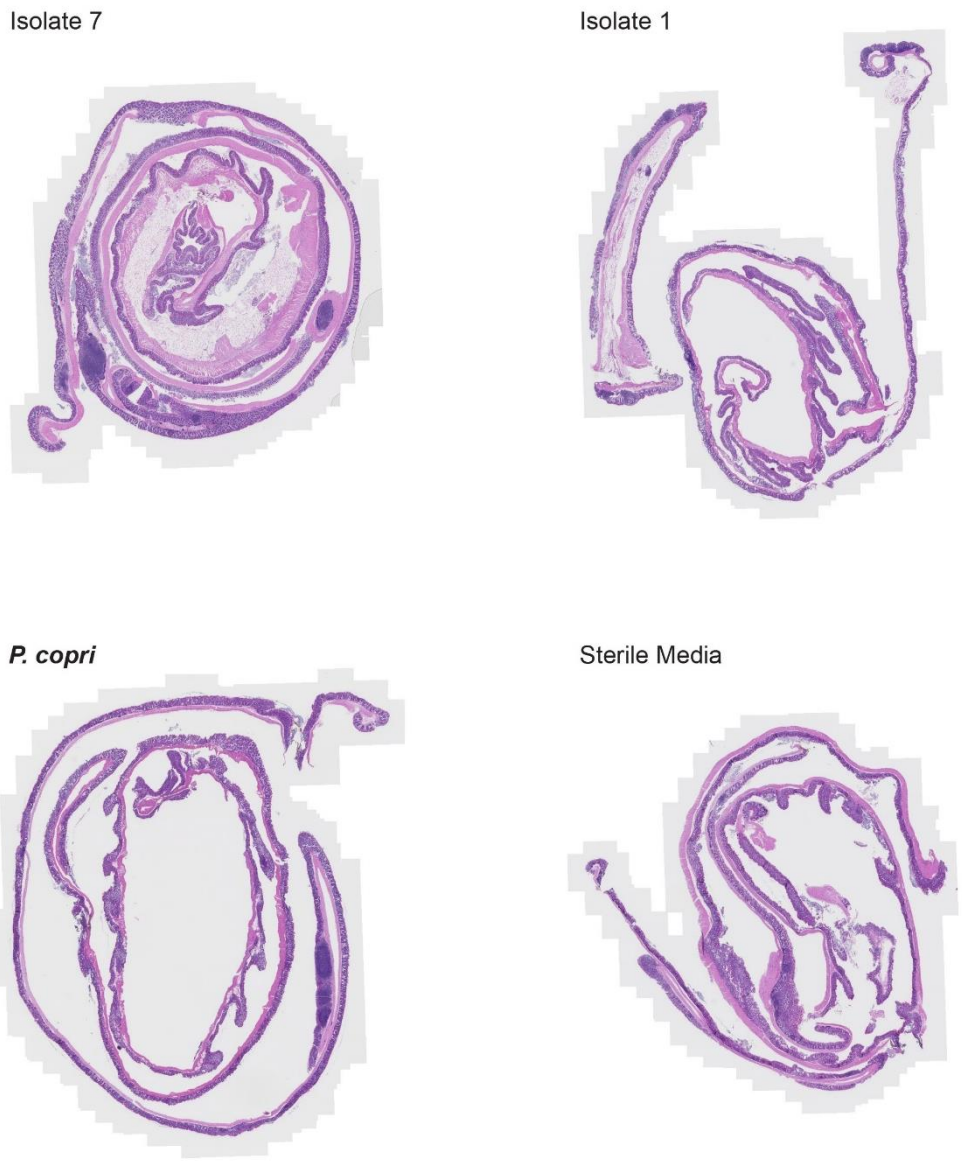


Fig. S15. Swiss roll images for each mono-colonized mouse treatment group.

Representative Swiss roll images of colonic mature isolated lymphoid follicles are shown for each treatment group.

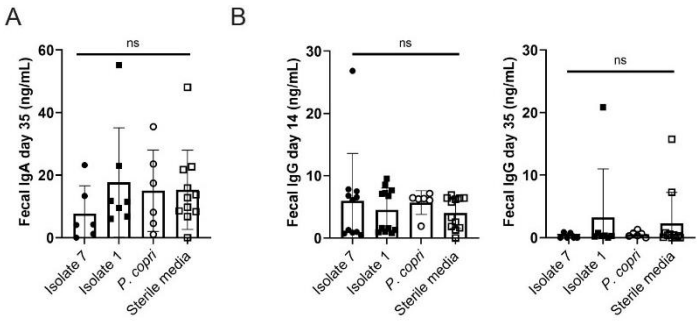


Fig. S16. Fecal IgA and IgG of mono-colonized mice.

(**A**) Fecal IgA and (**B**) IgG concentrations were determined from the mono-colonized mice at 14 days and 35 days after bacterial gavage. The total IgA and IgG are displayed in ng/mL, (y-axis) separated by treatment group (x-axis). For day 14, n=11 isolate 7-gavaged mice, n=l2 isolate 1-gavaged mice, n=6 *P. copri*-gavaged mice, and n=l2 sterile media-gavaged mice were tested. For day 35, n=6 isolate 7-gavaged mice, n=7 isolate 1-gavaged mice, n=6 *P. copri*-gavaged mice, and n=11 sterile media-gavaged mice were tested. Data are presented as mean±SEM. ns, not significant, as determined by Kruskal-Wallis test.

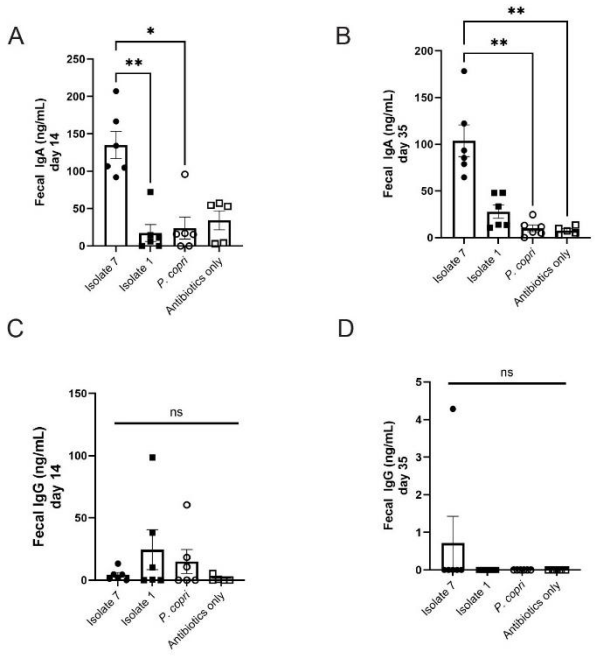


Fig. S17. Fecal IgA and IgG of SPF mice gavaged with isolate 7.

Feces from SPF mice gavaged with either isolate 1 (n=6), isolate 7 (n=6), or *P. copri* (n=6), or treated with antibiotics only (n=5) was collected at days 14 and 35 after gavage. (**A and B**) The total fecal IgA at 14 (A) and 35 (B) days after gavage was determined by ELISA; fecal IgA is displayed (y-axis) against treatment group (x-axis). (**C and D**) The total fecal IgG at 14 (C) and 35 (D) days after gavage was determined by ELISA. Symbols represent individual mice while bars are the mean±SEM. ns, not significant, *P<0.05, and **P<0.01, Kruskal-Wallis test with Dunn's post-test.

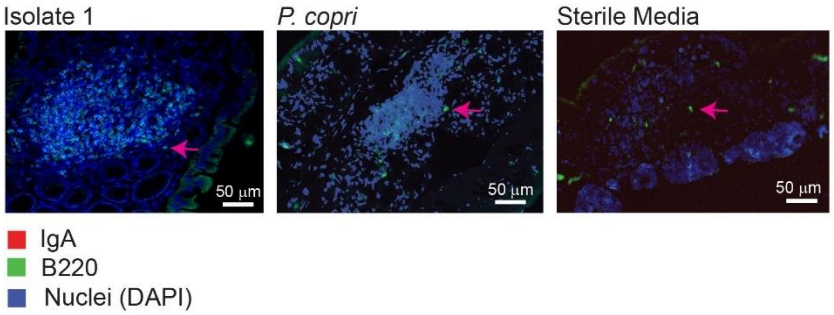


Fig. S18. Immunofluorescence on colonic isolated lymphoid follicles.

Representative colon immunofluorescence containing isolated lymphoid follicles across treatment groups is shown. In red is IgA staining, in green is B220 staining, and in blue is DAPI. Arrows indicate B220+ IgA- B cells.

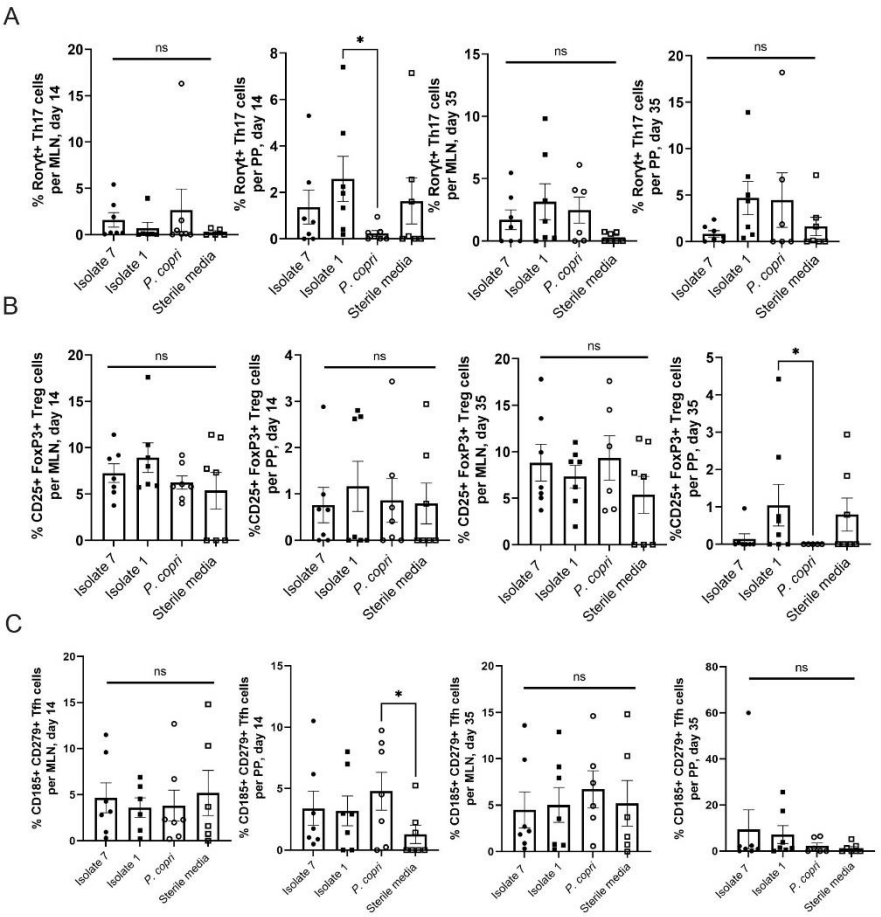


Fig. S19. MLNs and Peyer's patch immune cell populations in mice 14 and 35 days after gavage.

MLNs and Peyer's patch (PP) T cells frequencies and absolute numbers were determined at day 14 and 35 after gavage. (**A**) The percentage and absolute number of Th17 cells (live TCRβ+ CD4+ Rorγt+ lymphocytes) are shown (y-axis) compared to treatment group (x-axis). (**B**) The percentage and absolute number of Tregs (live TCRβ+ CD4+ CD25+ FoxP3+ lymphocytes) are shown. (**C**) The percentage and absolute number of Tfh cells (live TCRβ+ CD4+ CD185+ CD279+ lymphocytes) are shown. Data are presented as mean±SEM. *P<0.05 and ns, not significant as determined by Kruskal-Wallis with Dunn's post-test. n=7 isolate 7-gavaged mice, n=7 isolate l-gavaged mice, n=7 *P. copri*-gavaged mice, and n=8 sterile media-gavaged mice.

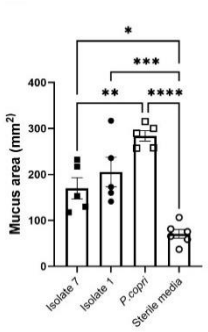


Fig. S20. Colonic mucus area in mice 35 days after bacterial gavage.

Colonic mucus area in square millimeters was determined 35 days after bacterial gavage. The average mucus area across 10 fields of view are shown (y-axis) compared to treatment group (x-axis). Data are presented as mean±SEM. *P<0.05, **P<0.01, ***P<0.001 as determined by ordinary one-way ANOVA with Tukey post-test. n=5 isolate 7-gavaged mice, n=5 isolate 1-gavaged mice, n=5 *P. copri*-gavaged mice, and n=6 sterile media-gavaged mice.

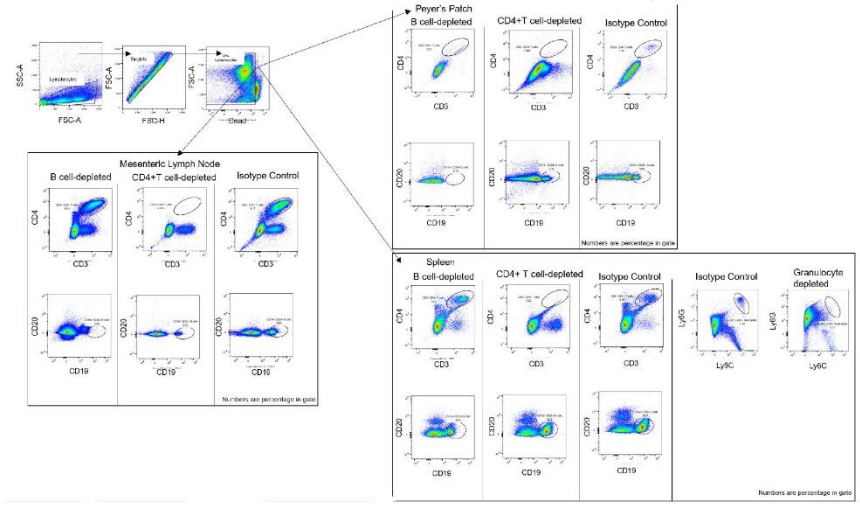


Fig. S21. Gating scheme and validation of depletion antibodies.

The gating scheme and validation of B cell, CD4+ T cell, and granulocyte depletion antibodies is displayed for each tissue of interest, including the MLNs, PP, and spleen. Each gate is labeled and contains the percentage of cells included.

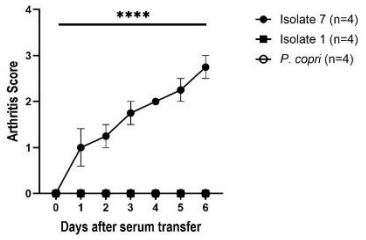


Fig. S22. Arthritis severity in mice up to 6 days post-serum transfer before phenotype resolution occurs.

Serum was collected from mice mono-colonized with isolate 1, isolate 7, or *P. copri* 35 days after bacterial gavage and was IP injected into healthy germ-free DBA/1 mice. These mice were monitored for the development of joint swelling and ankylosis every day for 6 days after serum transfer and then were euthanized for tissue collection. Clinical score is displayed (y-axis) against time since serum transfer (x-axis). Data are presented as mean±SEM. ****P<0.0001, repeated measures ANOVA. n=4 isolate 7 serum transfer, n=4 isolate 1 serum transfer, n=4 *P. copri* serum transfer.

Supplemental Tables:

Table S1. Characteristics of individuals in plasmablast mAb cohort

Abbreviations: NHW = Non-Hispanic White; DMARD = disease modifying anti-rheumatic drug; NSAID = non-steroidal anti-inflammatory drug; CCP = cyclic citrullinated peptide; RF = rheumatoid factor

Characteristic	**At-risk participants**	**Early RA participants**	**All participants**
N	4	2	6
Sex: N (% female)	3 (75.0)	2 (100.0)	5 (83.3)
Age: mean ± SD	58.8 ± 8.7	44.5 ± 16.3	54.0 ± 12.3
Race: n (% NHW)	3 (75.0)	1 (50.0)	4 (66.7)
Ever Smoke: n (% yes)	1 (25.0)	2 (100.0)	3 (50.0)
Current Smoker: n (% yes)	0 (0.0)	1 (50.0)	1 (16.7)
Supplement Use: n (% yes)	1 (25.0)	0 (0.0)	1 (16.7)
Other autoimmune disease: N (disease)			
DMARD: n (% yes)			
NSAID: n (% yes)	1 (25.0)	0 (0.0)	1 (16.7)
Swollen Joint: n (% yes)	0 (0.0)	1 (50.0)	1 (16.7)
Tender Joint: n (% yes)	2 (50.0)	1 (50.0)	3 (50.0)
Tested for Shared Epitope: n (% positive)	4 (100.0)	1 (50.0)	5 (83.3)
Shared Epitope: n (% positive *of those tested*)	2 (50.0)	1 (100.0)	3 (60.0)
CCP and RF Status			
RF- and CCP-: n (% yes)	0 (0.0)	0 (0.0)	0 (0.0)
RF+ only: n (% yes)	1 (25.0)	0 (0.0)	1 (16.7)
CCP+ only: n (% yes)	1 (25.0)	0 (0.0)	1 (16.7)
RF+ and CCP+: n (% yes)	2 (50.0)	2 (100.0)	4 (66.7)
Type of CCP+			
Tested for CCP3: n (% yes)	2 (50.0)	1 (50.0)	3 (50.0)
CCP3+: n (% yes *of those tested*)	2 (100.0)	1 (50.0)	3 (50.0)
Tested for CCP3.1: n (% yes)	4 (100.0)	2 (100.0)	6 (100.0)
CCP3.1+: n (% yes *of those tested*)	3 (75.0)	2 (100.0)	5 (83.3)
Type of RF+			
RF neph.+: n (% yes)			
Tested for RF: n (% yes)	4 (100.0)	2 (100.0)	6 (100.0)
RF IgA+: n (% yes *of those tested*)	0 (0.0)	2 (100.0)	2 (33.3)
RF IgG+: n (% yes *of those tested*)	1 (25.0)	2 (100.0)	3 (50.0)
RF IgM+: n (% yes *of those tested*)	1 (25.0)	2 (100.0)	3 (50.0)

Table S2. Characteristics of plasmablast cohort

	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Participant 6
RA status	At-risk for RA	At-risk for RA	At-risk for RA	At-risk for RA	Early RA	Early RA
Number plasmablasts derived from participant	19	18	19	22	7	5
Plasmablasts from each isolate in circulation						
%IgA+ plasmablasts	61.14	55.19	49.65	68.94	26.40	0
%IgG+ plasmablasts	20.29	29.87	37.83	23.48	45.51	83.12
%IgM+ plasmablasts	4.29	7.14	5.44	2.27	11.8	12.52
Plasmablasts isolated for study (counts)						
Number of IgG plasmablasts	7	2	6	9	7	5
Number of IgA plasmablasts	12	16	13	13	0	0
Plasmablasts targeting bacteria						
%Plasmablasts of cohort with bacterial targets	68.42	83.33	63.16	59.09	28.57	20
%IgG plasmablasts of cohort with bacterial targets	71.4	100	83.33	55.55	28.57	20
%IgA plasmablasts of cohort with bacterial targets	67	81	53.85	61.54	0	0

Table S3. Characteristics of human samples used for fecal pool

Abbreviations: NHW = Non-Hispanic White; DMARD = disease modifying anti-rheumatic drug; NSAID = non-steroidal anti-inflammatory drug; CCP = cyclic citrullinated peptide; RF = rheumatoid factor; neph = nephelometry. *This individual was previously CCP2+, CCP3.1+, and IgG RF+ but had negative titers at the time of draw.

Characteristic	**Healthy Controls**	**At-risk**	**Early RA**
N	5	8	5
Sex: N (% female)	4 (80)	6 (75)	3 (60)
Age: mean ± SD	41.2 ± 13.6	62 ± 11.3	46.2 ± 14.5
Race: N (% NHW)	4 (80)	7 (87.5)	4 (80)
Ever smoke: N (% yes)	1 (20)	1 (12.5)	1 (20)
Current smoker: N (% yes)	0 (0)	0 (0)	0 (0)
Supplement use: N (% yes)	4 (80)	6 (75)	5 (100)
DMARD: N (% yes)	0 (0)	0 (0)	4 (80)
NSAID: N (% yes)	1 (20)	5 (62.5)	2 (40)
Swollen joint: N (% yes)	0 (0)	1 (12.5)	3 (60)
Tender joint: N (% yes)	1 (20)	4 (50)	3 (60)
Shared Epitope: N (% positive)	4 (80)	3 (37.5)	3 (60)
CCP and RF status			
RF- and CCP-: N (% yes)	5 (100)	2 (25)	1 (20*)
RF+ only: N (% yes)	0 (0)	1 (12.5)	1 (20)
CCP+ only: N (% yes)	0 (0)	1 (12.5)	1 (20)
RF+ and CCP+: N (% yes)	0 (0)	4 (50)	2 (40)
Type of CCP			
CCP3+: N (% positive)	0 (0)	3 (37.5)	3 (60)
CCP3.1: N (% positive)	0 (0)	5 (62.5)	3 (60)
Type of RF			
RF neph+: N (% yes, *of those tested*)	0 (0)	2 (28.6)	2 (40)
RF IgA+: N (% positive)	0 (0)	1 (12.5)	2 (40)
RF IgG: N (% positive)	0 (0)	2 (25)	0 (0)
RF IgM: N (% positive)	0 (0)	4 (50)	2 (40)

Table S4. Characteristics of RA cases for CD4+ T cell reactivity assays

Abbreviations: CCP = cyclic citrullinated peptide; RF = rheumatoid factor

Participant ID	Participant Type	Anti-CCP Outcome	RF Outcome	HLA DRB1 (a+b)	Gender	Age at Draw	RA Duration at Draw	Active Autoimmune Medications
RA_A	RA	Positive	Positive	*0401/*0407	Female	60	28.1	Etanercept (Enbrel), Leflunomide (Arava)
RA_B	RA	Positive	Positive	*0401/*0408	Female	68	26.9	Untreated
RA_C	RA	Positive	Negative	*03/*0401	Female	58	2.9	Untreated
RA_D	RA	Positive	Negative	*0401/*13	Male	56	2.0	Hydroxychloro-quine (Plaquenil), Infliximab (Remicade), Leflunomide (Arava)
RA_E	RA	Positive	Positive	*03/*0401	Female	64	18.5	Rituximab
RA_F	RA	Positive	Positive	*01/*0401	Male	72	20.0	Methotrexate (Trexall, Rheumatrex)
RA_G	RA	Positive	Positive	*0401/*0408	Female	61	9.9	Ibuprofen (Advil), Methotrexate (Trexall, Rheumatrex)
RA_H	RA	Positive	Positive	*0401/*0404	Female	50	6.9	Untreated
RA_I	RA	Positive	Negative	*0401/*0408	Male	59	1.1	Meloxicam (Mobic), Etanercept (Enbrel), Leflunomide (Arava)
RA_J	RA	Positive	Positive	*01/*0401	Female	73	0.0	Methotrexate (Trexall, Rheumatrex)
RA_K	RA	Positive	Negative	*0401/*1501	Male	67	7.6	Meloxicam (Mobic)

Table S5. List of RA and Healthy Control cases for CD4+ T cell reactivity assays

Abbreviations: CCP = cyclic citrullinated peptide; RF = rheumatoid factor

Participant ID	Participant Type	Anti CCP Outcome	RF Outcome	HLA DRB1 (a+b)	Gender	Age at Draw	RA Duration At Draw	Active Autoimmune Medications
HC-1	HC			*0401/*0404	male	61		
HC-2	HC			*0401/*0701	male	65		
HC-3	HC			*0401/ Unknown	male	62		
HC-4	HC			*0401/ Unknown	female	25		
HC-5	HC			*0401/*0401	male	37		
HC-6	HC			*01/*0401	male	48		
HC-7	HC			*0401/ Unknown	female	54		
HC-8	HC			*0401/*0301	female	31		
HC-9	HC			*0401/ Unknown	female	49		
HC-10	HC			*0401/*0407	female	46		
HC-11	HC			*0401/ Unknown	male	63		
HC-12	HC			*0401/ Unknown	female	56		
RA-1	RA	Positive	Negative	*0401/ Unknown	male	61	21.3	Azathioprine, Etanercept
RA-2	RA	Positive	Positive	*0401/*0701	male	61	25.9	Analgesic, OTC, Unknown, Hydroxy-chloroquine, Methotrexate
RA-3	RA	Positive	Positive	*0301/*0401	female	40	11.9	Adalimumab, Aspirin (Bayer)
RA-4	RA	Positive	Positive	*0401/*0404	female	25	6.8	Etanercept
RA-5	RA	Positive	Positive	*0401/*1102	male	37	5.5	Sulfasalazine
RA-6	RA	Positive	Negative	*0401/*1501	male	48	8.1	Hydroxy-chloroquine, Leflunomide
RA-7	RA	Positive	Positive	*0401/ Unknown	female	56	20.4	Certolizumab
RA-8	RA	Positive	Positive	*0401/*03	female	30	7.2	Etanercept, Hydroxy-chloroquine, Sulfasalazine

RA-9	RA	Positive	Negative	*0401/*1501	female	51	10.6	Abatacept, Hydroxy-chloroquine
RA-10	RA	Positive	Positive	*01/*0401	female	55	2.5	Etanercept
RA-11	RA	Positive	Positive	*0401/*1501	female	48	8.1	Hydroxy-chloroquine, Methotrexate

1

Table S6. Characteristics of human fecal samples for *Subdoligranulum* isolate-specific qPCR

Abbreviations: NHW = Non-Hispanic White; DMARD = disease modifying anti-rheumatic drug; NSAID = non-steroidal anti-inflammatory drug; CCP = cyclic citrullinated peptide; RF = rheumatoid factor

Characteristic	**Healthy controls**	**At-risk**	**Early RA**	**All**
N	12	12	12	36
Sex: n (% female)	8 (66.7)	10 (83.3)	8 (66.7)	26 (72.2)
Age: mean ± SD	42.5 ± 14.3	57.8 ± 13.8	56.6 ± 13.4	52.3 ± 15.2
Race: n (% NHW)	9 (75.0)	9 (75.0)	10 (83.3)	28 (77.8)
*Ever Smoke: n (% yes)	3 (25.0)	4 (33.3)	6 (54.6)	13 (37.1)
*Current Smoker: n (% yes)	0 (0.0)	3 (25.0)	1 (10.0)	4 (11.8)
Supplement Use: n (% yes)	7 (63.6)	9 (81.8)	10 (100.0)	26 (81.3)
Other autoimmune disease: n (disease)	1 (thyroid and UC)	2 (1 thyroid, 1 UC)	4 (3 thyroid, 1 pemphigus)	7
DMARD: n (% yes)	0 (0.0)	0 (0.0)	4 (33.3)	4 (11.1)
NSAID: n (% yes)	5 (41.7)	8 (66.7)	8 (66.7)	21 (58.3)
Swollen Joint: n (% yes)	0 (0.0)	1 (8.3)	8 (72.7)	9 (25.7)
Tender Joint: n (% yes)	3 (25.0)	4 (33.3)	8 (72.7)	15 (42.9)
Tested for Shared Epitope: n (% tested)	12 (100.0)	12 (100.0)	11 (91.7)	35 (97.2)
Shared Epitope: n (% positive of those tested)	4 (33.3)	8 (66.7)	5 (45.5)	17 (48.6)
CCP and RF status				
RF- and CCP-: n (% yes)	12 (100.0)	0 (0.0)	1 (8.3)	13 (36.1)
RF+ only: n (% yes)	0 (0.0)	0 (0.0)	2 (16.7)	2 (16.7)
CCP+ only: n (% yes)	0 (0.0)	9 (75.0)	6 (50.0)	15 (41.7)
RF+ and CCP+: n (% yes)	0 (0.0)	3 (25.0)	3 (25.0)	6 (16.7)
Type of CCP+				
Tested for CCP3: n (% yes)	12 (100.0)	12 (100.0)	12 (100.0)	36 (100.0)
CCP3+: n (% yes of those tested)	0 (0.00	11 (91.7)	9 (75.0)	20 (55.6)
Tested for CCP3.1: n (% yes)	12 (100.0)	12 (100.0)	12 (100.0)	36 (100.0)
CCP3.1+ (% yes of those tested)	0 (0.0)	11 (91.7)	9 (75.0)	20 (55.6)
Type of RF+				
RF neph.+: n (% yes)	0 (0.0)	1 (8.3)	2 (20.0)	3 (13.0)
Tested for RF isotypes: n (% yes)	12 (100.0)	12 (100.0)	12 (100.0)	36 (100.0)
RF IgA+: n (% yes of those tested)	0 (0.0)	1 (8.3)	2 (16.7)	3 (8.3)
RF IgG+: n (% yes of those tested)	0 (0.0)	3 (25.0)	1 (8.3)	4 (11.1)
RF IgM+: n (% yes of those tested)	0 (0.0)	2 (16.7)	3 (25.0)	5 (13.9)
*missing data summary 1 subject missing ever smoke 2 subjects missing current smoker 4 missing supplement use 1 missing joint exam data 1 missing SE data 13 subjects missing RF neph. testing				

Table S7. Estimated percentage of Isolate 7 out of total bacterial load in human fecal samples.

The estimated percentage of total bacteria per 100mg feces that is isolate 7 was estimated for each positive sample, based off of a previously published estimate of total CFUs of bacteria per fecal weight (Stephen, A.M., *J Med Microbial*, 1980). The mean percentage of bacteria that are isolate 7 is displayed ± SEM for all positive samples (n=4), for positive samples from individuals at-risk for RA (n=2) and for positive samples for individuals with early RA (n=2).

Sample Subset	**Estimated percentage of isolate 7 out of total bacteria**
All positive samples	0.04 ± 0.07
At-risk participants	0.004 ± 0.003
Early RA participants	0.08 ± 0.11

Table S8. Fluorochromes used for murine immunophenotyping

Antibody Target	Fluorophore	Clone	Company
CD3	PerCP Cy5.5	145-2C11	Tonbo
CD4	PE	GK1.5	Tonbo
CD25	FITC	PC61.5	Tonbo
CD185	APC-Cy7	L138D7	BioLegend
CD279	BV421	29F.1A12	BioLegend
FoxP3	PE-Cy7	3G3	Tonbo
Rorγt	BV786	Q31-378	BD Horizon
Tbet	PE Vio615	REA102	Miltenyi
TCRβ	APC	H57-597	Tonbo
viability	Ghost 510	n/a	Tonbo

Table S9. Human T cell immunophenotyping panel information

Target	Fluorochrome	Clone	Manufacturer
CD3	PerCP-Cy5.5	UCHT1	BioLegend
CD4	BV510	RPA-T4	BioLegend
CD8	APC-Cy7	RPA-T8	BioLegend
CD69	AF-657	FN50	BioLegend
CD137	PE-Cy7	CY1G4	BioLegend
CD45RA	BV421	HI100	BioLegend
CCR7	BV711	G043H7	BioLegend
CD14	FITC	61D3	BioLegend
CD56	FITC	NCAM16.2	BD Biosciences
CD19	FITC	HIB19	BioLegend
CD154	PE-CF594	TRAP1	BD Biosciences
Sytox Dead	FITC	n/a	Thermo Fisher

Table S10. T Human T cell stimulation antibody information

Flow Cytometry Panel RA versus Control (Figure 2)				
Target	Fluorochrome	Manufacturer	Clone	Volume
CD8	BUV395	BD Biosciences	RPA-T8	1:50
PD-1	BUV 737	BD Biosciences	EH12.1	1:33
CD27	BV421	BioLegend	M-T271	1:66
CD4	V500	BD Biosciences	RPA-T4	1:40
CXCR3	BV605	BioLegend	G025H7	1:50
ICOS	BV650	BD Biosciences	DX29	1:50
CD95	BV711	BD Biosciences	DX2	1:50
CCR6	BV786	BioLegend	G034E3	1:50
CXCR5	BB515	BD Biosciences	RF8B2	1:50
CCR4	PE	BioLegend	L291H4	1:66
CD154	PE-Texas Red	BD Biosciences	TRAP1	1:20
CD137	PE-Cy7	BioLegend	4B4-1	1:20
CD69	APC	BioLegend	FN50	1:50
CD45RA	AF700	BD Biosciences	HI100	1:30
CCR7	APC-Cy7	BioLegend	G043H7	1:40
CD14	PerCP-Cy5.5	BioLegend	HCD14	1:50
CD16	PerCP-Cy5.5	BioLegend	3G8	1:50
CD19	PerCP-Cy5.5	BioLegend	HIB19	1:50
Viability	PerCP-Cy5.5	BD Biosciences	n/a	1:7

Table S11. Murine T cell stimulation antibody information

Target	Fluorochrome	Clone	Manufacturer
CD69	APC-eF780	H1.2F3	eBioscience
CD4	BUV395	GK1.5	BD Bioscience
IL-17	BV605	TC11-18H10.1	BioLegend
Rorγt	BV786	Q31-378	BD Bioscience
TCRβ	FITC	H57-597	Tonbo
CD154	PE	MR1	Tonbo
CD3	eF450	17A2	Invitrogen
Proliferation	Cell Trace Far Red	N/A	Thermo Fisher
Viability	Aqua Stain	N/A	Thermo Fisher

Table S12. Bacterial primers for qPCR

Target	Forward sequence 5' → 3'	Reverse sequence 5' → 3'
Universal 16S (RpoB)	AACATCGGTTTGATCAAC	CGTTGCATGTTGGTACCCAT
Isolate 7	ACCGAAAACACACACATAAACACT	GGTTACTGGCTTTCAGGCGA

Data File S1. Plasmablast cohort RA-relevant antigen binding

Data File S2. Variable region sequences for expressed plasmablast antibodies

Data File S3. Raw, individual level data for experiments where n<20.