Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Bhatt AS, Freeman SS, Herrera AF, et al. Sequence-based discovery of *Bradyrhizobium enterica* in cord colitis syndrome. N Engl J Med 2013;369:517-28. DOI: 10.1056/NEJMoa1211115

Supplementary appendix

Supplement to: *Sequence-based discovery of Bradyrhizobium enterica in cord colitis syndrome*

Table of contents

Ami S. Bhatt, MD, PhD, Samuel S. Freeman, B.S.E., Alex F. Herrera, MD, Chandra Sekhar Pedamallu, PhD, Dirk Gevers, PhD, Fujiko Duke, B.S., Joonil Jung, PhD, Monia Michaud, M.Sc., Bruce J. Walker, B.S., Sarah Young, PhD, Ashlee Earl, PhD, Aleksander D. Kostic, Akinyemi I. Ojesina, MD, PhD, Robert Hasserjian, MD, Karen K. Ballen, MD, Yi-bin Chen, MD, Gabriela Soriano, MD, Joseph H. Antin, MD, Robert J. Soiffer, MD, Lindsey R. Baden, MD, Wendy S. Garrett, MD, PhD, Jason L. Hornick, MD, PhD, Francisco M. Marty, MD, and Matthew Meyerson, MD, PhD

I. PathSeq results and genome assembly methods

PathSeq classification of sequencing reads from colon-biopsy samples 5b, 5c, 11b and 11d revealed a large number of unclassifiable sequences (Table S1).

Sequencing reads from short fragment sequencing libraries (insert size 150 – 400bp) were pooled from temporally separated biopsies from each separate patient (5b + 5c and 11b + 11d) as well as all four patients (5b + 5c + 11b + 11d). All paired-end sequences were treated as single-end reads and were run through the PathSeq algorithm for computational subtraction of human reads after quality filtering. All non-human reads from these samples and pair-mates of these non-human reads were also included in the assembly, regardless of the quality score of the pair-mate. Two separate computational assembly methods, Velvet[1] and ALLPATHS[2,3], were employed, as previously described. ALLPATHS was developed as a tool for genome assembly using dual inputs of short fragment sequencing libraries and large fragment (jumping) libraries. In order to use ALLPATHS for assembly, reads were first assembled into a temporary genome. All paired-end reads were aligned using the Burrows-Wheeler alignment algorithm to this temporary genome and insert size was inferred based on alignment of reads pairs.[4,5] Paired-end reads were then split into "shorter" and "longer" fragment pools and were taken forward for formal ALLPATHS assembly. Both assembly methods assembled a total contig length (of contigs > 2.5kb) of greater than 7.5Mb when applied to the pooled set of reads from all four sequenced samples. The ALLPATHS assembly generated a longer set of contigs for sequences obtained from a single patient (patient 11) and was thus taken forward for further analysis. The results of the final assembly using nonhuman sequences from colon-biopsy samples 11b and 11d are presented in Table S2. Thus, the ALLPATHS assembly of sequences from patient 11, which consisted of a set of 99 contigs was taken forward as the draft genome.

Each contig of greater than 2.5kb was analyzed for percent GC content and read coverage. Contigs were analyzed by BLASTN[6] against the NCBI nt database and were defined by the top hit (that with the lowest E value; Figure S1).

Each contig was plotted as a function of % GC and read coverage (Figure S2)

The BLASTN results of each individual contig were evaluated by our genome annotation team (ASB, SSF, CSP, SY, DG, AE, BW). The contig corresponding to the SEN virus was determined to be unlikely inserted into the novel organism's genome and was removed from the draft genome. The vast majority of the remaining contigs mapped to members of the family Bradyrhizobaceae and all other contigs mapping to other bacterial families were maintained in the draft genome due to similar coverage and GC content. As there are gaps in the draft genome, there remains the possibility that a small subset of these contigs is not a part of the true *B. enterica* genome. Future efforts to isolate, culture and complete the genome of this organism will be revealing in this regard, and will also illuminate the question of whether this organism has a circular or linear genome and whether it has a single chromosome or multiple chromosomes.

Contigs were taken forward for further assembly and from the 99 contigs, 90 scaffolds or supercontigs were generated (by end joining of contigs). One of these supercontigs (3,621bp) corresponded to the SEN virus and was excluded from further analysis.

========== Scaffold Stats ==========
Scaffolds 90
Max Scaffold 533,022
Mean Scaffold 84,997
Scaffold N50 155,300
Total Scaffold Length 7,649,768

SEN virus supercontig length 3,621
Total Scaffold number (minus SEN virus supercontig) 89
Total Scaffold Length (minus SEN virus supercontig) 7,646,147
==========

As the *B. enterica* genome was assembled from a complex human tissue sample, the genome has been submitted as a "multispecies" sample to the NCBI, as it was not derived from a isolated, purified culture or a true metagenomic sample. The strain has been designated DFCI-1 (Dana-Farber Cancer Institute-1) for the institution and location of care of cord colitis syndrome-affected patients.

II. Comparative genomic analysis and Circos plot construction

In order to perform comparative genome analysis of *B. enterica*, genome annotation was carried out by PRODIGAL (as previously described and cited in the main manuscript).[7] Gene annotations are available on NCBI.

The most closely related species in a phylogenetic analysis reported in the main text was *Bradyrhizobium japonicum* (strain USDA 110). In order to determine the homology between genes in *B. enterica* and *B. japonicum*, each PRODIGAL-predicted gene was compared to the *B. japonicum* amino acid sequence by peptide BLAST.[6] The full sequence of the top hit was extracted and the full-length genes were then aligned using the Needleman-Wunsch global alignment algorithm.[8] The percentage identity was then calculated for each gene. This value was plotted at the location of the gene on the circular genome plot in the main manuscript.[9] A histogram of global sequence identity by individual gene is provided (Figure S3).

B. enterica genes for which no homologous *B. japonicum* gene was found or for which the global amino acid sequence identity was less than 5% were identified and are plotted in the circular genome plot in the main manuscript. A list of the genes that are specific to *B. enterica* compared to *B. japonicum* is provided (Table S3). Note that the PRODIGAL algorithm is a highly specific method that conservatively assigns gene annotations, resulting in a significant number of hypothetical gene "calls".

III. Contamination analysis

Several limitations are introduced by the execution of a single center study that may increase the likelihood of contamination including (1) common paraffin baths used for the generation of FFPE samples, (2) a common nosocomial microbiome, (3) FFPE block handling by a single laboratory, (4) preparation of libraries using very limited DNA in a single laboratory location.

The experimental method employed in this single-center study was designed to minimize the likelihood that the results obtained were due to a contaminant as follows: (1) FFPE colon biopsy samples from normal controls and post-stem cell transplantation GVHD controls processed at the same institution were included and did not demonstrate appreciable *B. enterica* by PCR. (2) Additional frozen colon cancer controls were also included in this analysis and did not demonstrate appreciable *B. enterica* by PCR. (3) DNA extraction for the samples that were sequenced was started on the same day but was completed on successive days. (4) Two different type of barcodes generated at different facilities were used to generate sequencing libraries. (5) Samples 5b+5c and 11b+11d were sequenced at two different sequencing facilities. (6) Buffers and ultrapure water used in the extraction of DNA and generation of the libraries were subjected to targeted PCR to investigate for *B. enterica* in the stock solutions used (Figure S4). (7) DNA extraction and sequencing library construction was carried out in a dedicated "clean area" away from lab areas where organisms are cultured. (8) As samples were very limited, the reserved "top scrolls" from two of the samples (samples 9d and 9e) were subjected to DNA extraction several months after the original extraction and *B. enterica* was present in both scrolls that were studied (Figure S4). (9) Single nucleotide polymorphism analysis was limited by the reported intrinsic low polymorphism rate of organisms such as *Bradyrhizobium japonicum* USDA 110[10,11] and relatively low coverage of *B. enterica* for samples 5b+5c. Despite this, it appeared that there were at least five to 11 SNPs at an allelic fraction of at least 40% between *B. enterica* reads from patient 5 vs. patient 11. Additional intrinsic difficulties in evaluation for SNPs include the lack of a completed genome and the high GC content of the organism, which can lead to more frequent sequencing errors.

(10) All FFPE samples prepared for sequencing in our laboratory within four months of the cord colitis syndrome samples were analyzed by PathSeq for the presence of *B.* enterica.Samples sequenced at the Broad sequencing facility with the Broad barcode set were found to have the following total number of *B. enterica* reads. Note that the bone marrow and spleen samples were sequenced on the same flow-cell as samples 11b and 11d. There is a known small amount of barcode contamination between the 96 barcodes used at the Broad sequencing facility and the number of reads that correspond to *B. enterica* in the bone marrow and spleen samples are therefore felt to represent cross-contamination from the cord colitis syndrome libraries. Note that bone marrow and spleen RNA samples also had *B. enterica* reads. Mapping of these reads onto the draft *B. enterica* genome was visualized in IGV 2.0.[12] May reads were found to map to intergenic regions of *B. enterica* suggesting that these reads correspond to contamination from the

cord colitis syndrome libraries and not RNA sequence present in the bone marrow and spleen samples, themselves.

Sample descriptions and names are followed by the number of total reads from the samples that map to *B. enterica* by Burrows Wheeler alignment of raw reads against the *B. enterica* draft genome.[4,5] Samples included in the analysis below have at least 20 million high quality sequencing reads per sample. No samples studied in the four months before or after cord colitis syndrome sample sequencing had more reads mapping to *B. enterica* than the aforementioned bone marrow and spleen samples (raw *B.* enterica read count noted below). Note that this method has much lower specificity for the identification of *B. enterica* reads than the PathSeq method, which competitively classifies reads against a large database of known microbes.

Below are the file names of the samples sequenced at the Broad on the same lane as samples 5b and 5c followed by the number of raw reads that map to *B. enterica*:

Bone marrow aspirate (frozen DNA)
1512
Bone marrow aspirate (frozen RNA)
1904
Spleen (frozen DNA)
3141
Spleen (frozen RNA)
2118

Below are the de-identified file names of human FFPE tissue specimens sequenced at the same facility as samples 11b and 11d followed by the number of raw reads that map to *B. enterica*:

CC-M-066_unmapped.sam
4
CC-M-067_unmapped.sam
8
CC-M-068_unmapped.sam
52
CC-M-069_unmapped.sam
13
CC-M-070_unmapped.sam
3
CC-M-071_unmapped.sam
48
CC-M-072_unmapped.sam
4
CC-M-073_unmapped.sam
2
CC-M-074_unmapped.sam

14
CC-M-075_unmapped.sam
3
CC-M-076_unmapped.sam
7
CC-M-077_unmapped.sam
3

IV. Viral reads in sequenced cord colitis syndrome samples

Samples 5b, 5c, 11b and 11d were carried through PathSeq analysis, as described in the main text of the manuscript. Manually reviewed hits are presented in Table S4. A detailed list of all viral hits is also presented (Figure S5).

V. PCR conditions

PCR was performed using 10 mM forward and reverse primers, 0.2 ng of input DNA and the AccuPrime Taq DNA polymerase system (Invitrogen, Grand Island, NY, USA) per manufacturer's directions in a total volume of 10 ml with the following cycle protocol: 95°C for 2 minutes, followed by 35 cycles of: 95°C for 30 seconds, 62.1°C for 30 seconds, 68°C for 40 seconds, and finally an extension at 68° C for 5 minutes. PCR was carried out on an Eppendorf AG Mastercycler Pro (Hauppauge, NY, USA).

VI. Fluorescence *in situ* hybridization methods

FISH was carried out according to Swidsinski's method as follows.[13,14] Following sectioning and mounting of 5um FFPE tissue samples to glass slides, the paraffin embedded tissue was fixed to the slides by incubation at 50°C for 30 minutes. The slides were then deparaffinized by successive incubation in xylene baths (four incubations of five minutes each) followed by four washes in 100% ethanol for five minutes each. The slides were then incubated at 50°C for 25 minutes. A cold PAP pen was used to encircle the samples followed by incubation with 1mg/mL lysozyme (diluted in water) for 15 minutes. Oligonucleotide FISH probes were selected using probeBase (http://www.microbial-ecology.net/probebase/) and were custom synthesized by IDT. Details on oligonucleotide probes are available at probeBase.[15] Two probes were ordered, specific to a eubacterial 16S RNA sequence (EUB338, 5'-/Cy5/-GCT GCC TCC CGT AGG AGT-3') and specific to a Bradyrhizobium 16S RNA sequence (Brady, 5'-/Cy3/-CTG CCG CTG ACA TAT TGC TA-3'). The Brady target sequence was present in the *B. enterica* genomic sequence. Probes were diluted in water to a stock concentration of 50ng/mL and 0.5uL of probe was added to 50uL of hybridization buffer containing 1% formamide (0.88M NaCl, 20mM Tris HCl, pH 7.4, 0.05% SDS). The sample was incubated with the probe solution for 45 minutes at 50°C in a humid, pre-warmed chamber and in the dark. The slides were washed in a prewarmed wash buffer (900mM NaCl, 100mM Tris HCl, pH 7.4, 0.03% SDS) for five minutes. The slides were then washed with ddH_2O and dried in a 50°C oven for five minutes. The slides were stained with 1ug/mL DAPI (450mM NaCl, 20mM Tris HCl, pH 7.4) and incubated for 5minutes in the dark. The samples were washed with ddH_2O and dried at 50°C for five minutes.

A competition assay as a negative control to test for specificity of the Brady probe was performed as follows. Prior to hybridization of the labeled probes, samples were incubated with a 100-fold molar excess of unlabeled Brady probe in hybridization buffer at 50°C for 45 minutes. This buffer was removed and fresh hybridization buffer containing both labeled probes as well as the unlabeled Brady probe at a 100-fold molar excess was applied to the sample. The sample was incubated at 50°C for 45 minutes. Thereafter, the same wash and DAPI staining steps were followed as noted above. Results of the competition assay, demonstrating high specificity of the Brady probe, are displayed in Figure S6.

Prior to visualization, AquaMount aqueous mounting medium (Cardinal Health, Dublin, Ohio) was applied to the sample and the sample was covered with a cover slip.

FISH and hematoxylin and eosin stained tissues were visualized and images captured using a Nikon Eclipse Ni microscope, DS-Qi1 monochrome camera for fluorescence or a DS-Fi2 color camera, and the NIS Elements BR imaging and software station. Exposure times were fixed and consistent for each individual fluor across all imaged samples. Images were pseudocolored and overlayed using the NIS Elements BR software station and cropped for figure generation using Adobe Photoshop CS5.

VII. Quantitation of *Bradyrhizobium enterica* in sequenced samples before and after antibiotics therapy

The relative abundance of *B. enterica* before and after antibiotic therapy was calculated in the four samples that were sequenced. The number of *B. enterica* reads was corrected for the total number of human reads. This "relative abundance" was compared in the pre- vs. post-antibiotic therapy samples (5b and 11b were pre-antibiotics, 5c and 11d were post-antibiotics). A 6.4-fold reduction in *B. enterica* relative abundance was seen in patient 5 with the initiation of antibiotics. A 2.5-fold reduction in *B. enterica* relative abundance was seen in patient 11 with the initiation of antibiotics (Table S5).

VIII. Detection of *Bradyrhizobium enterica* in upper GI biopsies of cord colitis syndrome-affected patients and in post-HSCT colitis samples from Massachusetts General Hospital (MGH)

In the original report of cord colitis syndrome, Herrera et al noted the presence of granulomas in selected upper GI biopsies of cord colitis syndrome-affected patients. Given this finding, we tested the hypothesis that *Bradyrhizobium enterica* may be the inciting organism of this host response by obtaining upper GI biopsies from patients of the original cord colitis syndrome-cohort. Four upper GI biopsies (stomach and duodenum) were obtained from three affected patients from the original cohort (two samples from patient 5, one sample from patient 6 and one sample from patient 11). FFPE blocks were retrieved for each of these samples, the first 20um was shaved off and discarded and the subsequent 40um section was subjected to DNA extraction as described in the main text. DNA was quantified using the Q-bit system. 1uL of DNA from each sample was subjected to PCR using the *B. enterica* primer set and PCR conditions indicated in section V of the supplementary appendix. A nontemplate control and positive control (0.2ng of the bar-coded sequencing library for patient 5b) were also subjected to PCR simultaneously.

Results of this experiment are presented in Figure S7 and demonstrate the presence of *B. enterica* in all of the upper GI biopsy samples tested. This suggests that *B. enterica* colonizes not only the lower gastrointestinal tract but also the upper gastrointestinal tract in affected patients.

In order to determine if *B. enterica* was an institution-specific potential pathogen, samples from MGH were obtained for sequencing-based analysis. The following samples were obtained: endoscopic colon, duodenum and stomach biopsies were obtained from four post-HSCT colitis patients and an ileum sample from a control patient with Crohn's disease. All samples were subjected to DNA extraction and PathSeq analysis, as previously described in this manuscript. Pathological review of all patients who had undergone UCB-SCT and had colonic biopsies for evaluation of diarrhea revealed six patients with biopsies that could be consistent with cord colitis syndrome. Of those six patients, three were felt to be lower probability. Upon clinical review, one of the three higher probability cases was felt to have some features that were suggestive of a diagnosis of cord colitis syndrome, although there were some atypical features.

The single potential cord colitis syndrome patient was a 35 year-old female who underwent double umbilical cord transplantation at MGH for FLT3-positive AML Her initial transplant course was complicated by engraftment syndrome requiring glucocorticoid treatment, febrile neutropenia treated with cefepime, vancomycin, and micafungin, *Streptococcus mitis* bacteremia treated with ceftriaxone, HHV-6 viremia treated with foscarnet, and elevated liver function tests. On day 33, the patient was readmitted to MGH with worsening liver function tests and a week of watery diarrhea. Flexible sigmoidoscopy and a liver biopsy were performed. Pathology showed normal colonic mucosa in the sigmoid colon and extensive hepatic iron deposition with ceroid-

laden macrophages and pericentral sinusoidal dilation consistent with recent injury. There was a non-caseating granuloma seen on the liver biopsy. The patient received one dose of vancomycin, cefepime, and metronidazole. The diarrhea was self-limited and the elevated liver function tests resolved with discontinuation of fluconazole. On day 48, the patient presented to a local hospital with recurrent watery diarrhea associated with abdominal cramping and a 4.5 kg weight loss. The patient underwent esophagogastroduodenoscopy (EGD) and colonoscopy, which showed ulcerated mucosa and on pathology demonstrated findings that were not typical for cord colitis syndrome (Subtle increased crypt epithelial cell apoptosis and ulceration with mixed inflammation). A colon biopsy from this time point was available for molecular investigation using PCR-based methods. The patient was treated with ciprofloxacin and metronidazole for 25 days with improvement in her symptoms. On day 92, the patient was started on sorafenib as part of a clinical research trial. On day 98, the patient again developed voluminous diarrhea, nausea, 5kg weight loss, and low-grade fevers, and was re-admitted to the hospital on day 103. She was started on empiric glucocorticoids and cefepime. EGD and colonoscopy were performed; duodenal, gastric and colonic biopsies showed patchy granulomatous inflammation, but no significant increase in gland or crypt epithelial cell apoptosis, and minimal active inflammation. The colonic biopsy showed no architectural distortion to suggest chronic mucosal injury. The presence of granulomas would be consistent with cord colitis syndrome, but the absence of increased apoptosis, architectural distortion, and neutrophilic cryptitis would argue against this diagnosis. Biopsies from the stomach and duodenum were available from this time point for molecular investigation; the colonic specimen was not available for molecular investigation.

AFB, GMS, PAS/d, Giemsa, and Brown-Hopps stains were negative for any organisms. Additionally, immunohistochemical stains for CMV, adenovirus, HSV, and VZV were negative. Glucocorticoids were tapered and the patient was switched to ciprofloxacin and metronidazole to complete a 14-day course with improvement in her diarrhea. Shortly after discontinuing ciprofloxacin and metronidazole, the patient again developed watery, large volume diarrhea, nausea, and poor appetite on day 129 for which she was re-treated with ciprofloxacin and metronidazole. Again, she responded well with decreased diarrhea and weight gain. However, the patient developed recurrent diarrhea and nausea while being treated with antibacterial agents. On day 160, the metronidazole was stopped with the goal of decreasing nausea. On day 176, she developed a fever and because of her persistent upper GI symptoms and diarrhea was restarted on both ciprofloxacin and metronidazole. On day 183, she again underwent EGD and colonoscopy that showed a subtle increase in crypt epithelial cell apoptosis, patchy chronic inflammation of a random colon biopsy, including granulomas that were not associated with crypt rupture, focal Paneth cells, without evidence of neutrophilic inflammation. Again, the absence of active (neutrophilic) colitis would be unusual for cord colitis syndrome.

She completed a course of ciprofloxacin and metronidazole by day 203 with improvement in her symptoms. Her subsequent course was complicated by an intensive-care unit admission for *Citrobacter spp*. Bacteremia and *Clostridium difficile* infection. Ultimately, her diarrhea returned and the patient has remained on chronic suppressive ciprofloxacin and her symptoms have been relatively well controlled.

Clinically, the patient has a relapsing, antibiotic-responsive, chronic diarrheal syndrome that could be consistent with the cord colitis syndrome. However, the onset of symptoms as early as day 33 or 48 is early relative to the original cohort of cord colitis syndrome patients (median onset day 131, range 88 to 314). Later in the patient's course, she seems to have developed worsening diarrhea while taking ciprofloxacin and metronidazole, which has not been seen in prior cases of cord colitis syndrome. The hepatic granulomas would be unusual for cord colitis syndrome and may suggest an alternate inflammatory or infectious etiology. Similarly, the absence of increased crypt epithelial cell apoptosis and neutrophilic cryptitis argue against cord colitis syndrome. The introduction of sorafenib with subsequent upper GI symptoms and diarrhea is also a confounding factor.

DNA was extracted from all available tissue blocks from this "variant cord colitis syndrome" case, as described in a previous section of the supplementary appendix. All samples were subjected to PCR as was described in section V of this Appendix. *B. enterica.* PCR was negative for all samples studied. As deep sequencing is an inherently more sensitive method for investigation of the presence of components of the microbiome, a subset of samples for which adequate amounts of DNA were available (stomach and duodenum from the day 98 EGD) were taken forward for sequencing-based analysis. Sequencing demonstrated the presence of *B. enterica* reads, although in low abundance, in both of these samples (Table S6, Figure S8).

These findings confirm the existence of *B. enterica* outside of the original institution where it was discovered. Given the relatively small number of samples that were studied and the lack of a very clearly defined case, it is not possible to ascertain the existence of cord colitis syndrome outside of the original institution where it was described. Based on the data presented, it is not possible to establish the generalizability of the association between *B. enterica* presence and cord colitis syndrome.

IX. Immunoglobulin levels in cord colitis syndrome patients

Hypogammaglobulinemia post-HSCT was common in this population with four of the original 11 patients receiving intravenous immunoglobulin (IVIG) post-transplantation. Immunoglobulin levels were available for the two months before or after presumed onset of cord colitis syndrome for 11 of the 11 patients. Immunoglobulin levels were available for the one month before or after presumed onset of cord colitis syndrome for nine of the 11 patients. Of the five patients who had lower GI biopsies that were included as a part of this study, four of the five patients were hypogammaglobulinemic at the time of cord colitis syndrome. Two of the patients required treatment with IVIG. Only one patient, patient 4, was not hypogammaglobulinemic at a time point of two weeks after clinical onset of cord colitis syndrome. Given the relative hypogammaglobulinemia of patients within this cohort and the well-described defects of adaptive immunity in post-HSCT patients, the likelihood of serological evidence of an adaptive immune response against *B. enterica* would be unlikely. Details of clinical status of patients and immunoglobulin status are presented (Table S7).

X. Figures

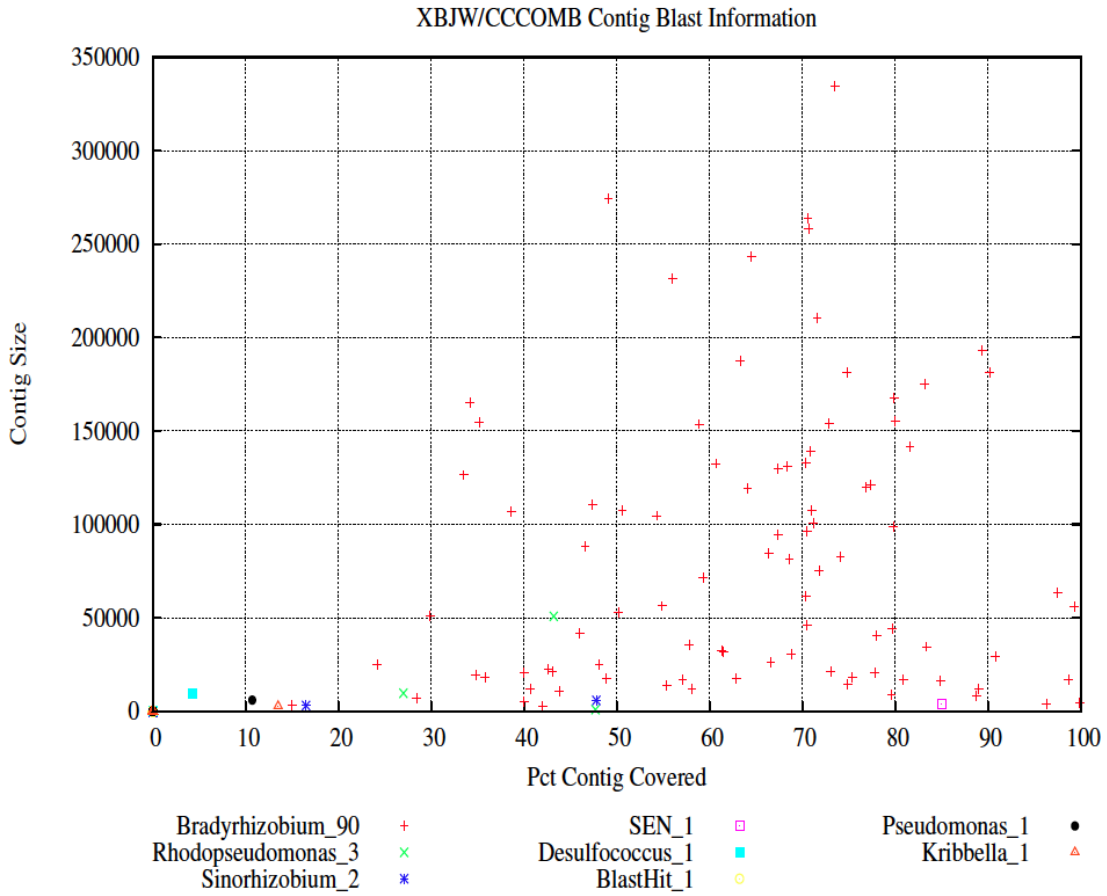

Figure S1. BLASTN of contigs >2.5kb generated by the ALLPATHS assembly of nonhuman reads of Samples 5b and 5c. Each contig was subjected to nucleotide BLAST against the NCBI nt database. The top hit was taken for each contig and the organism corresponding to the top hit is indicated on the scatter plot as described in the legend. The *x*-axis indicates the percentage of the contig that was contained in the top hit and the *y*-axis indicates the contig size.

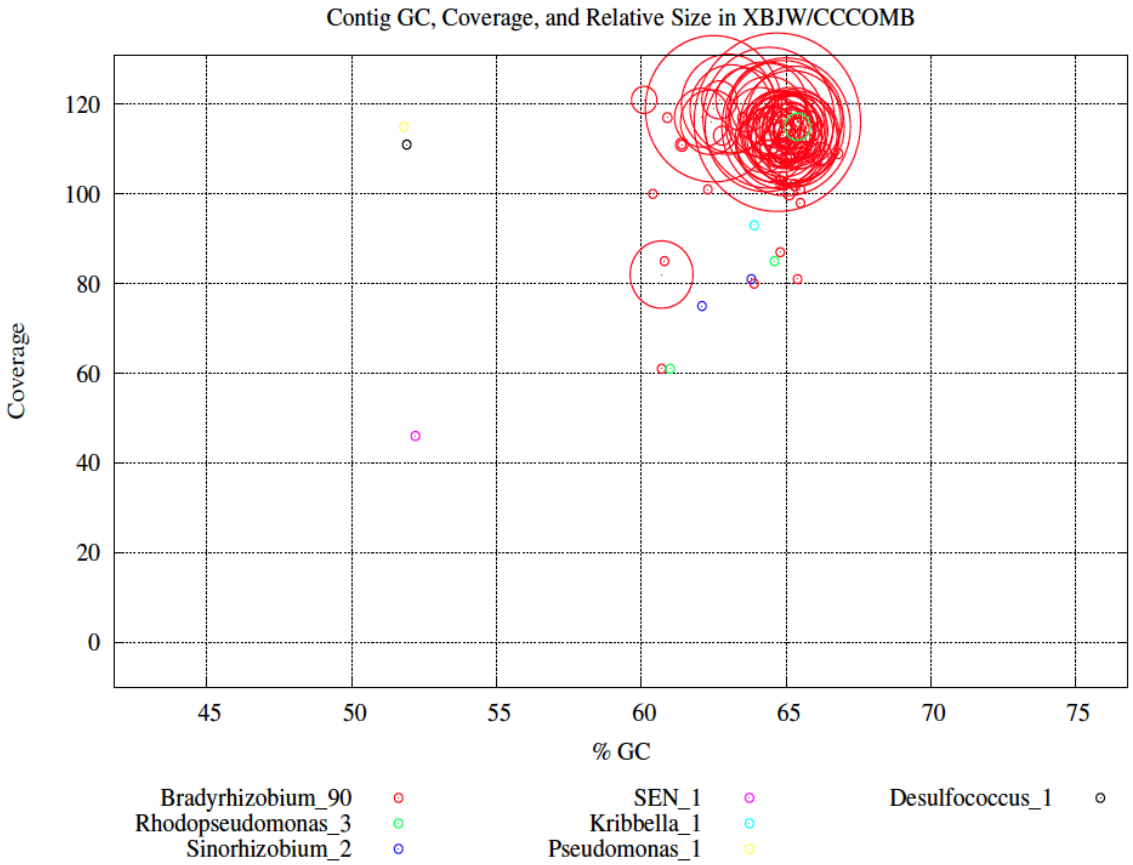

Figure S2. GC content, size and read coverage for contigs generated by the ALLPATHS assembly of samples 5b and 5c. Each contig is indicated as a colored circle (the color corresponds to the organism encoded by the top nucleotide BLAST hit as described in Figure 1). The size of the circle correlates with the relative size of each contig. Percent GC content is indicated on the *x*-axis and read coverage is indicated on the *y*-axis.

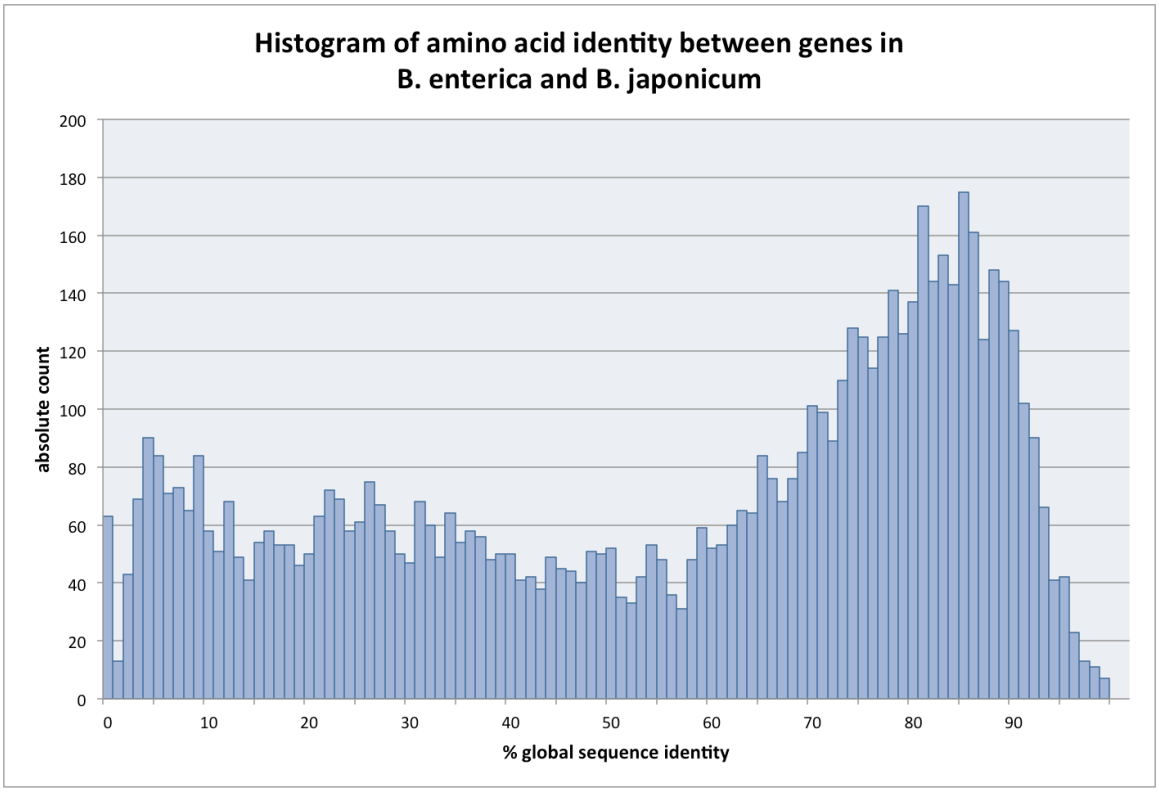

Figure S3. Histogram indicating the number of predicted *B. enterica genes* based on percentage global amino acid sequence identity to the closes *B. japonicum* homologue.

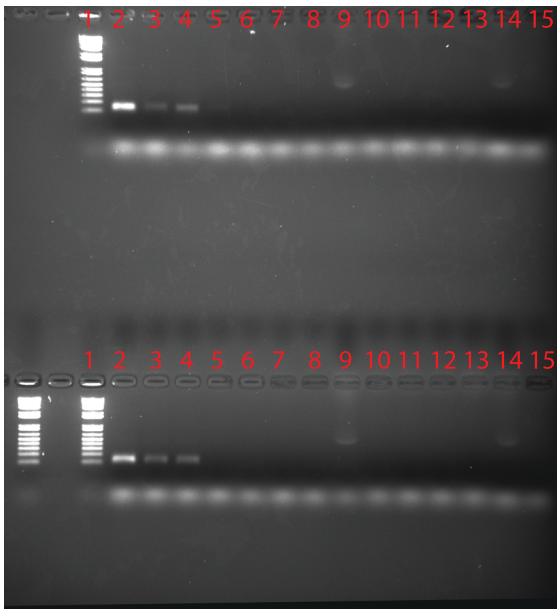

Figure S4. PCR (in duplicate) to detect *B. enterica*. PCR was performed using the conditions indicated in the main text with the exception that 40 cycles of PCR were carried out. Lanes are indicated with red text and correspond to the following:

1. 100bp MW marker
2. CC006 (positive control) - middle scroll
3. CC011 - top scroll
4. CC010 - top scroll
5. Non template control
6. Hemo-D
7. Wash 2/3 (bottle 1)
8. Wash 2/3 (bottle 2)
9. Digestion buffer
10. Wash 1 (bottle 1)
11. Wash 1 (bottle 2)
12. Wash 1 (bottle 3)
13. Isolation additive
14. Digestion buffer
15. Nuclease free water

NAME OF VIRUS	NUMBER OF READS
5b	
ALL VIRUSES	35
HUMAN HERPESVIRUS 6	18
ENTEROBACTERIA PHAGE	7
PROPIONIBACTERIUM PHAGE	4
ACINETOBACTER PHAGE	3
PSEUDOMONAS PHAGE	1
PHAGE GIFSY-1	1
INFLUENZA A	1

NAME OF VIRUS	NUMBER OF READS
5c	
ALL VIRUSES	94
HUMAN HERPESVIRUS 6	46
HUMAN HERPESVIRUS 7	39
TORQUE TENO VIRUS	9

11b	
ALL VIRUSES	300
HUMAN HERPESVIRUS 6	241
ENTEROBACTERIA PHAGE	26
PROPIONIBACTERIUM PHAGE	8
HUMAN CYTOMEGALOVIRUS	5
BACTERIOPHAGE FELIX	4
TORQUE TENO VIRUS	4
STREPTOCOCCUS PHAGE	2
PSEUDOMONAS PHAGE	2
HUMAN ADENOVIRUS	2
BACTERIOPHAGE EJ-1	2
TTV-LIKE VIRUS	1
CAFETERIA ROENBERGENSIS	1
BURKHOLDERIA CEPACIA PHAGE	1
BACTERIOPHAGE F1	1

11c	
ALL VIRUSES	478
TORQUE TENO VIRUS	231
ENTEROBACTERIA PHAGE	88
BACTERIOPHAGE FELIX	57
HUMAN HERPESVIRUS 6	42
SEN VIRUS 41	41
PROPIONIBACTERIUM PHAGE	9
HUMAN CYTOMEGALOVIRUS	7
TTV-LIKE VIRUS	1
MICRO TORQUE VIRUS	1
HUMAN ADENOVIRUS	1

Figure S5. PathSeq quantification of viral reads in sequences from cord colitis syndrome samples.

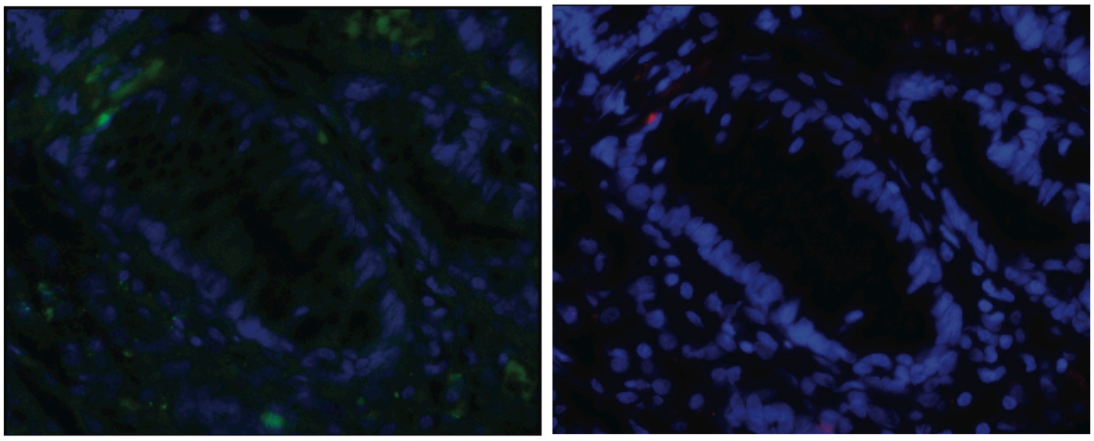

Figure S6. Competition assay with unlabeled Bradyrhizobium 16S RNA probe. The panel on the left demonstrates diffuse staining with the Bradyrhizobium 16S RNA labeled probe, panel on the right demonstrates that staining with the Brady probe is abrogated in the presence of unlabeled probe).

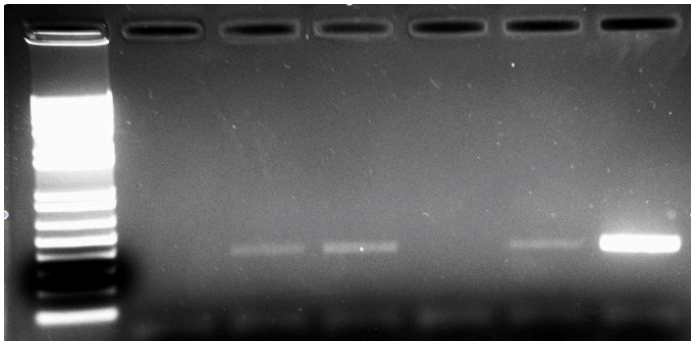

Figure S7. PCR based identification of *B. enterica* in upper GI biopsies from cord colitis syndrome-affected patients. (Left to R, 1kb plus MW marker, non-template control, patient 5 sample 1, patient 5 sample 2, patient 6 sample 1, patient 11 sample 1, positive control). PCR of all samples except for the non-template control also amplified human actin, to varying degrees (data not shown).

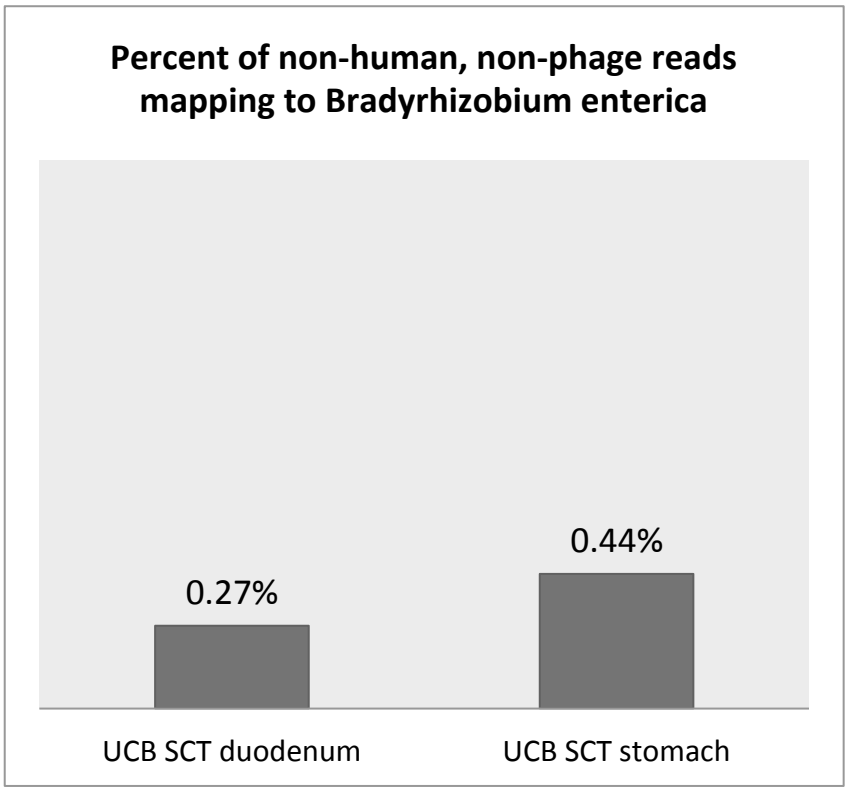

Figure S8. Sequencing-based identification of *Bradyrhizobium enterica* in post-HSCT colitis patients from a second institution. Shotgun DNA sequencing and PathSeq-based taxonomic classification of reads from gastrointestinal biopsies of patients with post-HSCT colitis syndromes suggests that *B. enterica* exists outside of the original institution in which it was discovered.

XI. Tables

Sample number	***5b***	***5c***	***11b***	***11d***
Read length	101	101	76	76
Total number of reads	110,856,860	134,251,634	31,045,710	41,992,012
Low quality reads (removed)	28,610,052	35,389,363	11,015,365	5,689,381
Duplicate/repeat reads	1,625,164	2,492,830	1,982,166	1,351,105
Human reads	79,945,073	96,204,998	16,282,123	33,546,179
Known bacterial reads	274,698	65,792	573,551	455,742
Known other microbial reads (Archaea, Fungi, etc)	77	151	417	162
Known viral reads	35	94	300	478
Unmapped reads	401,761	98,406	1,191,788	955,165

Table S1. Classification of reads from whole genome shotgun sequencing of formalin-fixed, paraffin embedded colon biopsy samples from patients with cord colitis. Computational analysis of massively parallel DNA sequencing from human tissue samples was performed using PathSeq software. Human reads were computationally subtracted, followed by taxonomic classification with BLASTN to microbial and viral databases. A large proportion of non-human reads were "unmappable" to available reference genomes.

Samples used for assembly	**11b + 11d**
Number of input reads for assembly*	4,619,184
Number of contigs (>2.5kb)	99
Maximum contig length (bp)	334,780
Mean contig length (bp)	77,268
Contig N50	141,525
Total Contig length (bp)	7,649,492
Assembly GC content (% of total bp)	64.4

Table S2. Results of contig generation from unmapped read assembly. The ALLPATHS software program was used to assemble unmapped reads from pooled samples (11b and 11d) into longer, contiguous sequences.

Gene identification number	Prodigal-predicted gene name	amino acid length	Is gene absent in B. japonicum or <5% identity to a predicted B. japonicum homologue?
C207_02513	Bradyrhizobium enterica 2-dehydro-3-deoxyphosphogalactonate aldolase	174	<5% identity
C207_05358	Bradyrhizobium enterica 2-haloacid dehalogenase	50	<5% identity
C207_00881	Bradyrhizobium enterica 3-oxoacid CoA-transferase subunit B	33	absent
C207_06559	Bradyrhizobium enterica 3-oxoacyl-[acyl-carrier-protein] synthase III	70	<5% identity
C207_01707	Bradyrhizobium enterica 4-hydroxyacetophenone monooxygenase	129	<5% identity
C207_02017	Bradyrhizobium enterica 4-hydroxyphenylacetate-3-monooxygenase large chain	526	<5% identity
C207_00016	Bradyrhizobium enterica 6-aminohexanoate-cyclic-dimer hydrolase	61	<5% identity
C207_06840	Bradyrhizobium enterica acetoacetyl-CoA synthetase	48	<5% identity
C207_04517	Bradyrhizobium enterica alanyl-tRNA synthetase	48	<5% identity
C207_04847	Bradyrhizobium enterica alkanesulfonate monooxygenase	103	<5% identity
C207_04970	Bradyrhizobium enterica antibiotic transport system ATP-binding protein	71	<5% identity
C207_02911	Bradyrhizobium enterica ApaG protein	49	<5% identity
C207_03323	Bradyrhizobium enterica aspartate ammonia-lyase	174	<5% identity
C207_01988	Bradyrhizobium enterica aspartyl-tRNA (Asn)/glutamyl-tRNA (Gln) amidotransferase subunit C	64	<5% identity
C207_04254	Bradyrhizobium enterica ATP-dependent Clp protease ATP-binding subunit ClpB	154	<5% identity
C207_06703	Bradyrhizobium enterica ATP-dependent Clp protease subunit	61	<5% identity
C207_02710	Bradyrhizobium enterica biopolymer transporter ExbD	103	<5% identity
C207_02204	Bradyrhizobium enterica branched-chain amino acid transport system ATP-binding protein	112	<5% identity
C207_00214	Bradyrhizobium enterica branched-chain amino acid transport system ATP-binding protein	59	<5% identity
C207_01742	Bradyrhizobium enterica branched-chain amino acid transport system permease	123	<5% identity
C207_02874	Bradyrhizobium enterica branched-chain amino acid transport system substrate-binding protein	338	<5% identity
C207_00321	Bradyrhizobium enterica branched-chain amino acid transport system substrate-binding protein	257	<5% identity
C207_01678	Bradyrhizobium enterica carbamate kinase	320	<5% identity
C207_01177	Bradyrhizobium enterica CDF family cation efflux system protein	155	<5% identity
C207_03088	Bradyrhizobium enterica cell division protein FtsI (penicillin-binding protein 3)	116	<5% identity
C207_01915	Bradyrhizobium enterica cobalt transporter subunit CbtB (proposed)	64	<5% identity

C207_00003	Bradyrhizobium enterica cobalt-precorrin 5A hydrolase	138	<5% identity
C207_05190	Bradyrhizobium enterica cytochrome d ubiquinol oxidase subunit II	633	<5% identity
C207_06585	Bradyrhizobium enterica D-threo-aldose 1-dehydrogenase	84	<5% identity
C207_01857	Bradyrhizobium enterica DNA repair protein RecN (Recombination protein N)	66	<5% identity
C207_02942	Bradyrhizobium enterica DNA-3-methyladenine glycosylase II	63	<5% identity
C207_00234	Bradyrhizobium enterica DOPA 4,5-dioxygenase	137	<5% identity
C207_01169	Bradyrhizobium enterica dTDP-4-dehydrorhamnose 3,5-epimerase	111	<5% identity
C207_00459	Bradyrhizobium enterica FdhD protein	140	<5% identity
C207_06358	Bradyrhizobium enterica Fe-S cluster assembly protein SufD	67	<5% identity
C207_03698	Bradyrhizobium enterica filamentous hemagglutinin family domain-containing protein	4428	<5% identity
C207_01723	Bradyrhizobium enterica filamentous hemagglutinin family domain-containing protein	4282	<5% identity
C207_01969	Bradyrhizobium enterica filamentous hemagglutinin family domain-containing protein	4010	<5% identity
C207_04905	Bradyrhizobium enterica filamentous hemagglutinin family domain-containing protein	3769	<5% identity
C207_02878	Bradyrhizobium enterica flagellum-specific ATP synthase	226	<5% identity
C207_04305	Bradyrhizobium enterica formyl-CoA transferase	127	<5% identity
C207_05832	Bradyrhizobium enterica galactarate dehydratase	82	<5% identity
C207_02559	Bradyrhizobium enterica general secretion pathway protein D	104	<5% identity
C207_07133	Bradyrhizobium enterica glutathione transport system permease	148	<5% identity
C207_00007	Bradyrhizobium enterica glycerol-3-phosphate dehydrogenase	895	<5% identity
C207_06841	Bradyrhizobium enterica haloacetate dehalogenase	46	<5% identity
C207_07098	Bradyrhizobium enterica hypothetical protein	2910	<5% identity
C207_06833	Bradyrhizobium enterica hypothetical protein	1855	<5% identity
C207_06429	Bradyrhizobium enterica hypothetical protein	816	<5% identity
C207_01070	Bradyrhizobium enterica hypothetical protein	599	<5% identity
C207_02136	Bradyrhizobium enterica hypothetical protein	587	<5% identity
C207_03202	Bradyrhizobium enterica hypothetical protein	545	absent
C207_01463	Bradyrhizobium enterica hypothetical protein	543	<5% identity
C207_03999	Bradyrhizobium enterica hypothetical protein	463	<5% identity
C207_02931	Bradyrhizobium enterica hypothetical protein	437	absent
C207_01798	Bradyrhizobium enterica hypothetical protein	431	<5% identity
C207_05230	Bradyrhizobium enterica hypothetical protein	430	<5% identity
C207_06785	Bradyrhizobium enterica hypothetical protein	415	<5% identity
C207_01242	Bradyrhizobium enterica hypothetical protein	382	<5% identity
C207_02843	Bradyrhizobium enterica hypothetical protein	366	absent

C207_02120	Bradyrhizobium enterica hypothetical protein	334	<5% identity
C207_04081	Bradyrhizobium enterica hypothetical protein	334	<5% identity
C207_03341	Bradyrhizobium enterica hypothetical protein	327	<5% identity
C207_07094	Bradyrhizobium enterica hypothetical protein	294	<5% identity
C207_05219	Bradyrhizobium enterica hypothetical protein	288	absent
C207_03599	Bradyrhizobium enterica hypothetical protein	283	<5% identity
C207_01150	Bradyrhizobium enterica hypothetical protein	259	<5% identity
C207_05854	Bradyrhizobium enterica hypothetical protein	255	<5% identity
C207_00970	Bradyrhizobium enterica hypothetical protein	233	<5% identity
C207_01966	Bradyrhizobium enterica hypothetical protein	225	<5% identity
C207_06967	Bradyrhizobium enterica hypothetical protein	225	<5% identity
C207_05333	Bradyrhizobium enterica hypothetical protein	206	<5% identity
C207_06540	Bradyrhizobium enterica hypothetical protein	197	<5% identity
C207_05378	Bradyrhizobium enterica hypothetical protein	196	<5% identity
C207_01191	Bradyrhizobium enterica hypothetical protein	196	absent
C207_06786	Bradyrhizobium enterica hypothetical protein	186	absent
C207_00400	Bradyrhizobium enterica hypothetical protein	183	<5% identity
C207_03995	Bradyrhizobium enterica hypothetical protein	176	<5% identity
C207_02696	Bradyrhizobium enterica hypothetical protein	174	<5% identity
C207_01068	Bradyrhizobium enterica hypothetical protein	160	<5% identity
C207_01535	Bradyrhizobium enterica hypothetical protein	157	<5% identity
C207_03228	Bradyrhizobium enterica hypothetical protein	152	<5% identity
C207_04620	Bradyrhizobium enterica hypothetical protein	151	<5% identity
C207_07089	Bradyrhizobium enterica hypothetical protein	146	<5% identity
C207_01330	Bradyrhizobium enterica hypothetical protein	145	<5% identity
C207_00316	Bradyrhizobium enterica hypothetical protein	144	<5% identity
C207_06454	Bradyrhizobium enterica hypothetical protein	143	<5% identity
C207_06934	Bradyrhizobium enterica hypothetical protein	140	<5% identity
C207_06065	Bradyrhizobium enterica hypothetical protein	137	<5% identity
C207_06412	Bradyrhizobium enterica hypothetical protein	130	absent
C207_04600	Bradyrhizobium enterica hypothetical protein	128	<5% identity
C207_02656	Bradyrhizobium enterica hypothetical protein	126	<5% identity
C207_06022	Bradyrhizobium enterica hypothetical protein	125	<5% identity
C207_00934	Bradyrhizobium enterica hypothetical protein	121	<5% identity
C207_05116	Bradyrhizobium enterica hypothetical protein	121	<5% identity
C207_05441	Bradyrhizobium enterica hypothetical protein	120	<5% identity
C207_04449	Bradyrhizobium enterica hypothetical protein	119	<5% identity
C207_06118	Bradyrhizobium enterica hypothetical protein	118	<5% identity
C207_05934	Bradyrhizobium enterica hypothetical protein	115	<5% identity
C207_01403	Bradyrhizobium enterica hypothetical protein	113	<5% identity
C207_03963	Bradyrhizobium enterica hypothetical protein	111	<5% identity

C207_05797	Bradyrhizobium enterica hypothetical protein	108	<5% identity
C207_00550	Bradyrhizobium enterica hypothetical protein	106	<5% identity
C207_04611	Bradyrhizobium enterica hypothetical protein	106	<5% identity
C207_01406	Bradyrhizobium enterica hypothetical protein	105	<5% identity
C207_01734	Bradyrhizobium enterica hypothetical protein	105	<5% identity
C207_02902	Bradyrhizobium enterica hypothetical protein	105	<5% identity
C207_04614	Bradyrhizobium enterica hypothetical protein	105	<5% identity
C207_05167	Bradyrhizobium enterica hypothetical protein	104	<5% identity
C207_01791	Bradyrhizobium enterica hypothetical protein	103	<5% identity
C207_05570	Bradyrhizobium enterica hypothetical protein	103	<5% identity
C207_01794	Bradyrhizobium enterica hypothetical protein	102	<5% identity
C207_03993	Bradyrhizobium enterica hypothetical protein	100	<5% identity
C207_04236	Bradyrhizobium enterica hypothetical protein	100	<5% identity
C207_06932	Bradyrhizobium enterica hypothetical protein	100	<5% identity
C207_06922	Bradyrhizobium enterica hypothetical protein	99	<5% identity
C207_06562	Bradyrhizobium enterica hypothetical protein	98	<5% identity
C207_05824	Bradyrhizobium enterica hypothetical protein	97	<5% identity
C207_06950	Bradyrhizobium enterica hypothetical protein	96	<5% identity
C207_04264	Bradyrhizobium enterica hypothetical protein	94	<5% identity
C207_06139	Bradyrhizobium enterica hypothetical protein	94	<5% identity
C207_01751	Bradyrhizobium enterica hypothetical protein	93	<5% identity
C207_03614	Bradyrhizobium enterica hypothetical protein	90	<5% identity
C207_04833	Bradyrhizobium enterica hypothetical protein	90	<5% identity
C207_06299	Bradyrhizobium enterica hypothetical protein	89	<5% identity
C207_01183	Bradyrhizobium enterica hypothetical protein	88	<5% identity
C207_01430	Bradyrhizobium enterica hypothetical protein	88	<5% identity
C207_02833	Bradyrhizobium enterica hypothetical protein	88	<5% identity
C207_04597	Bradyrhizobium enterica hypothetical protein	88	<5% identity
C207_03399	Bradyrhizobium enterica hypothetical protein	88	absent
C207_04481	Bradyrhizobium enterica hypothetical protein	87	<5% identity
C207_06845	Bradyrhizobium enterica hypothetical protein	87	<5% identity
C207_01212	Bradyrhizobium enterica hypothetical protein	86	<5% identity
C207_01529	Bradyrhizobium enterica hypothetical protein	86	<5% identity
C207_07126	Bradyrhizobium enterica hypothetical protein	86	<5% identity
C207_01077	Bradyrhizobium enterica hypothetical protein	85	<5% identity
C207_01552	Bradyrhizobium enterica hypothetical protein	85	<5% identity
C207_02712	Bradyrhizobium enterica hypothetical protein	85	<5% identity
C207_03892	Bradyrhizobium enterica hypothetical protein	85	<5% identity
C207_04468	Bradyrhizobium enterica hypothetical protein	85	<5% identity
C207_03949	Bradyrhizobium enterica hypothetical protein	84	<5% identity
C207_04454	Bradyrhizobium enterica hypothetical protein	83	<5% identity

C207_05444	Bradyrhizobium enterica hypothetical protein	82	<5% identity
C207_00280	Bradyrhizobium enterica hypothetical protein	81	<5% identity
C207_05913	Bradyrhizobium enterica hypothetical protein	80	<5% identity
C207_04376	Bradyrhizobium enterica hypothetical protein	79	<5% identity
C207_01418	Bradyrhizobium enterica hypothetical protein	78	<5% identity
C207_02008	Bradyrhizobium enterica hypothetical protein	78	<5% identity
C207_06615	Bradyrhizobium enterica hypothetical protein	78	<5% identity
C207_06707	Bradyrhizobium enterica hypothetical protein	78	<5% identity
C207_00504	Bradyrhizobium enterica hypothetical protein	75	<5% identity
C207_04314	Bradyrhizobium enterica hypothetical protein	75	<5% identity
C207_05933	Bradyrhizobium enterica hypothetical protein	74	<5% identity
C207_06935	Bradyrhizobium enterica hypothetical protein	74	<5% identity
C207_07049	Bradyrhizobium enterica hypothetical protein	73	absent
C207_00413	Bradyrhizobium enterica hypothetical protein	72	<5% identity
C207_05375	Bradyrhizobium enterica hypothetical protein	72	<5% identity
C207_05382	Bradyrhizobium enterica hypothetical protein	72	<5% identity
C207_06111	Bradyrhizobium enterica hypothetical protein	72	<5% identity
C207_00150	Bradyrhizobium enterica hypothetical protein	71	<5% identity
C207_00595	Bradyrhizobium enterica hypothetical protein	70	<5% identity
C207_01078	Bradyrhizobium enterica hypothetical protein	69	<5% identity
C207_04557	Bradyrhizobium enterica hypothetical protein	69	<5% identity
C207_06134	Bradyrhizobium enterica hypothetical protein	69	<5% identity
C207_02575	Bradyrhizobium enterica hypothetical protein	67	<5% identity
C207_03144	Bradyrhizobium enterica hypothetical protein	67	<5% identity
C207_05053	Bradyrhizobium enterica hypothetical protein	67	<5% identity
C207_02003	Bradyrhizobium enterica hypothetical protein	67	absent
C207_01786	Bradyrhizobium enterica hypothetical protein	66	<5% identity
C207_03465	Bradyrhizobium enterica hypothetical protein	65	<5% identity
C207_04529	Bradyrhizobium enterica hypothetical protein	65	absent
C207_04394	Bradyrhizobium enterica hypothetical protein	64	<5% identity
C207_05648	Bradyrhizobium enterica hypothetical protein	64	<5% identity
C207_05858	Bradyrhizobium enterica hypothetical protein	64	<5% identity
C207_01792	Bradyrhizobium enterica hypothetical protein	63	<5% identity
C207_04266	Bradyrhizobium enterica hypothetical protein	63	<5% identity
C207_04616	Bradyrhizobium enterica hypothetical protein	63	<5% identity
C207_06462	Bradyrhizobium enterica hypothetical protein	63	<5% identity
C207_03940	Bradyrhizobium enterica hypothetical protein	63	absent
C207_00554	Bradyrhizobium enterica hypothetical protein	62	<5% identity
C207_01735	Bradyrhizobium enterica hypothetical protein	61	<5% identity
C207_07090	Bradyrhizobium enterica hypothetical protein	61	<5% identity
C207_03964	Bradyrhizobium enterica hypothetical protein	60	<5% identity

C207_00944	Bradyrhizobium enterica hypothetical protein	59	<5% identity
C207_02529	Bradyrhizobium enterica hypothetical protein	59	<5% identity
C207_05937	Bradyrhizobium enterica hypothetical protein	59	<5% identity
C207_01419	Bradyrhizobium enterica hypothetical protein	58	<5% identity
C207_05381	Bradyrhizobium enterica hypothetical protein	58	<5% identity
C207_07127	Bradyrhizobium enterica hypothetical protein	58	<5% identity
C207_03618	Bradyrhizobium enterica hypothetical protein	56	<5% identity
C207_04383	Bradyrhizobium enterica hypothetical protein	56	<5% identity
C207_04477	Bradyrhizobium enterica hypothetical protein	56	<5% identity
C207_06492	Bradyrhizobium enterica hypothetical protein	56	<5% identity
C207_01534	Bradyrhizobium enterica hypothetical protein	55	<5% identity
C207_02125	Bradyrhizobium enterica hypothetical protein	55	<5% identity
C207_02301	Bradyrhizobium enterica hypothetical protein	55	<5% identity
C207_03151	Bradyrhizobium enterica hypothetical protein	55	<5% identity
C207_05406	Bradyrhizobium enterica hypothetical protein	55	<5% identity
C207_02339	Bradyrhizobium enterica hypothetical protein	55	absent
C207_02516	Bradyrhizobium enterica hypothetical protein	54	<5% identity
C207_01148	Bradyrhizobium enterica hypothetical protein	54	absent
C207_01785	Bradyrhizobium enterica hypothetical protein	53	<5% identity
C207_05735	Bradyrhizobium enterica hypothetical protein	53	<5% identity
C207_01140	Bradyrhizobium enterica hypothetical protein	52	<5% identity
C207_04958	Bradyrhizobium enterica hypothetical protein	52	<5% identity
C207_00534	Bradyrhizobium enterica hypothetical protein	52	absent
C207_00542	Bradyrhizobium enterica hypothetical protein	51	<5% identity
C207_02937	Bradyrhizobium enterica hypothetical protein	51	<5% identity
C207_01699	Bradyrhizobium enterica hypothetical protein	50	<5% identity
C207_05214	Bradyrhizobium enterica hypothetical protein	50	<5% identity
C207_01531	Bradyrhizobium enterica hypothetical protein	50	absent
C207_06225	Bradyrhizobium enterica hypothetical protein	50	absent
C207_02152	Bradyrhizobium enterica hypothetical protein	49	<5% identity
C207_05285	Bradyrhizobium enterica hypothetical protein	48	<5% identity
C207_00588	Bradyrhizobium enterica hypothetical protein	48	absent
C207_00722	Bradyrhizobium enterica hypothetical protein	48	absent
C207_05494	Bradyrhizobium enterica hypothetical protein	48	absent
C207_05703	Bradyrhizobium enterica hypothetical protein	47	<5% identity
C207_01554	Bradyrhizobium enterica hypothetical protein	47	absent
C207_01097	Bradyrhizobium enterica hypothetical protein	46	<5% identity
C207_00969	Bradyrhizobium enterica hypothetical protein	45	<5% identity
C207_01800	Bradyrhizobium enterica hypothetical protein	45	<5% identity
C207_00634	Bradyrhizobium enterica hypothetical protein	45	absent
C207_04590	Bradyrhizobium enterica hypothetical protein	45	absent

C207_05503	Bradyrhizobium enterica hypothetical protein	45	absent
C207_04915	Bradyrhizobium enterica hypothetical protein	43	<5% identity
C207_05938	Bradyrhizobium enterica hypothetical protein	43	<5% identity
C207_06288	Bradyrhizobium enterica hypothetical protein	43	<5% identity
C207_06994	Bradyrhizobium enterica hypothetical protein	43	<5% identity
C207_01562	Bradyrhizobium enterica hypothetical protein	43	absent
C207_02714	Bradyrhizobium enterica hypothetical protein	42	<5% identity
C207_05112	Bradyrhizobium enterica hypothetical protein	42	<5% identity
C207_01514	Bradyrhizobium enterica hypothetical protein	40	<5% identity
C207_00810	Bradyrhizobium enterica hypothetical protein	40	absent
C207_01141	Bradyrhizobium enterica hypothetical protein	40	absent
C207_06362	Bradyrhizobium enterica hypothetical protein	38	<5% identity
C207_04129	Bradyrhizobium enterica hypothetical protein	38	absent
C207_01374	Bradyrhizobium enterica hypothetical protein	37	absent
C207_04916	Bradyrhizobium enterica hypothetical protein	37	absent
C207_05693	Bradyrhizobium enterica hypothetical protein	37	absent
C207_01743	Bradyrhizobium enterica hypothetical protein	36	<5% identity
C207_04602	Bradyrhizobium enterica hypothetical protein	36	<5% identity
C207_01872	Bradyrhizobium enterica hypothetical protein	36	absent
C207_02486	Bradyrhizobium enterica hypothetical protein	36	absent
C207_04189	Bradyrhizobium enterica hypothetical protein	36	absent
C207_04202	Bradyrhizobium enterica hypothetical protein	36	absent
C207_03601	Bradyrhizobium enterica hypothetical protein	35	absent
C207_00591	Bradyrhizobium enterica hypothetical protein	34	<5% identity
C207_04350	Bradyrhizobium enterica hypothetical protein	34	absent
C207_06503	Bradyrhizobium enterica hypothetical protein	34	absent
C207_06891	Bradyrhizobium enterica hypothetical protein	34	absent
C207_00911	Bradyrhizobium enterica hypothetical protein	33	<5% identity
C207_01059	Bradyrhizobium enterica hypothetical protein	33	absent
C207_04209	Bradyrhizobium enterica hypothetical protein	33	absent
C207_05684	Bradyrhizobium enterica hypothetical protein	33	absent
C207_06783	Bradyrhizobium enterica hypothetical protein	32	<5% identity
C207_02248	Bradyrhizobium enterica hypothetical protein	32	absent
C207_03468	Bradyrhizobium enterica hypothetical protein	32	absent
C207_06482	Bradyrhizobium enterica hypothetical protein	32	absent
C207_07159	Bradyrhizobium enterica hypothetical protein	32	absent
C207_03294	Bradyrhizobium enterica hypothetical protein	31	absent
C207_00015	Bradyrhizobium enterica hypothetical protein	30	absent
C207_00039	Bradyrhizobium enterica hypothetical protein	30	absent
C207_01058	Bradyrhizobium enterica hypothetical protein	30	absent
C207_01698	Bradyrhizobium enterica hypothetical protein	30	absent

C207_01804	Bradyrhizobium enterica hypothetical protein	29	absent
C207_03018	Bradyrhizobium enterica hypothetical protein	23	absent
C207_06871	Bradyrhizobium enterica hypothetical protein	20	absent
C207_03997	Bradyrhizobium enterica indolepyruvate ferredoxin oxidoreductase	166	<5% identity
C207_02900	Bradyrhizobium enterica light-harvesting protein B-880 alpha chain	64	<5% identity
C207_02901	Bradyrhizobium enterica light-harvesting protein B-880 beta chain	73	<5% identity
C207_06138	Bradyrhizobium enterica lipid A biosynthesis lauroyl acyltransferase	256	<5% identity
C207_04704	Bradyrhizobium enterica long-chain acyl-CoA synthetase	71	<5% identity
C207_01846	Bradyrhizobium enterica magnesium transporter	51	<5% identity
C207_00945	Bradyrhizobium enterica malate dehydrogenase (oxaloacetate-decarboxylating)	81	<5% identity
C207_01039	Bradyrhizobium enterica membrane protein	179	<5% identity
C207_04947	Bradyrhizobium enterica membrane-bound serine protease (ClpP class)	174	<5% identity
C207_02895	Bradyrhizobium enterica MFS transporter, BCD family, chlorophyll transporter	452	<5% identity
C207_03996	Bradyrhizobium enterica MFS transporter, BCD family, chlorophyll transporter	168	<5% identity
C207_03721	Bradyrhizobium enterica MFS transporter, BCD family, chlorophyll transporter	158	<5% identity
C207_02016	Bradyrhizobium enterica muconolactone delta-isomerase	97	<5% identity
C207_01524	Bradyrhizobium enterica multidrug efflux transporter MdtA	69	absent
C207_06828	Bradyrhizobium enterica multiple sugar transport system substrate-binding protein	174	<5% identity
C207_03140	Bradyrhizobium enterica NAD(P) transhydrogenase subunit beta	186	<5% identity
C207_07161	Bradyrhizobium enterica nitrite reductase (NAD(P)H) large subunit	64	absent
C207_00317	Bradyrhizobium enterica NitT/TauT family transport system ATP-binding protein	86	<5% identity
C207_03397	Bradyrhizobium enterica oxidoreductase	155	<5% identity
C207_04149	Bradyrhizobium enterica penicillin-binding protein 1A	343	<5% identity
C207_03586	Bradyrhizobium enterica peptide/nickel transport system permease	61	absent
C207_04636	Bradyrhizobium enterica periplasmic protein TonB	55	<5% identity
C207_04848	Bradyrhizobium enterica permease	104	<5% identity
C207_04512	Bradyrhizobium enterica phosphinothricin acetyltransferase	222	<5% identity
C207_00961	Bradyrhizobium enterica phosphoglycolate phosphatase	72	<5% identity
C207_05411	Bradyrhizobium enterica phytoene synthase	65	<5% identity
C207_01174	Bradyrhizobium enterica protease	492	<5% identity
C207_02899	Bradyrhizobium enterica reaction center protein L	279	absent

	chain		
C207_02763	Bradyrhizobium enterica RelE/StbE family addiction module toxin	99	<5% identity
C207_05676	Bradyrhizobium enterica ribose 5-phosphate isomerase A	52	absent
C207_03424	Bradyrhizobium enterica simple sugar transport system ATP-binding protein	62	<5% identity
C207_05162	Bradyrhizobium enterica small GTP-binding protein	171	<5% identity
C207_03318	Bradyrhizobium enterica starch synthase	64	<5% identity
C207_05284	Bradyrhizobium enterica starvation-inducible DNA-binding protein	438	absent
C207_00516	Bradyrhizobium enterica sulfonate transport system substrate-binding protein	670	<5% identity
C207_06059	Bradyrhizobium enterica tat (twin-arginine translocation) pathway signal sequence	101	<5% identity
C207_05879	Bradyrhizobium enterica threonine synthase	238	<5% identity
C207_02700	Bradyrhizobium enterica TonB family domain-containing protein	237	<5% identity
C207_05118	Bradyrhizobium enterica transcriptional regulator	73	<5% identity
C207_03226	Bradyrhizobium enterica transmembrane sensor	211	<5% identity
C207_05463	Bradyrhizobium enterica two-component system, chemotaxis family, sensor kinase CheA	42	<5% identity
C207_06398	Bradyrhizobium enterica two-component system, chemotaxis family, sensor kinase CheA	31	<5% identity
C207_06941	Bradyrhizobium enterica two-component system, chemotaxis family, sensor kinase CheA	31	<5% identity
C207_07005	Bradyrhizobium enterica two-component system, chemotaxis family, sensor kinase CheA	31	<5% identity
C207_07110	Bradyrhizobium enterica two-component system, OmpR family, phosphate regulon response regulator OmpR	137	<5% identity
C207_04538	Bradyrhizobium enterica type IV secretion system protein VirB2	99	<5% identity
C207_00251	Bradyrhizobium enterica UDPglucose 6-dehydrogenase	99	<5% identity
C207_03021	Bradyrhizobium enterica urease accessory protein ureE	204	<5% identity
C207_06301	Bradyrhizobium enterica uroporphyrinogen-III synthase	92	<5% identity
C207_04870	Bradyrhizobium enterica YD repeat (two copies)	63	<5% identity

Table S3. A list of genes present in *B. enterica* that are absent in *B. japonicum* or have homologues with less than 5% identity to *B. japonicum*.

Virus	**5b**	**5c**	**11b**	**11d**
	number of reads			
TTV	0	9	3	185
HHV6b	14	46	19	42
CMV	0	0	224	7
EBV	0	0	0	1
KSHV	0	0	0	1
HHV7	2	39	0	0

Table S4. The abundance (number of reads) of a subset of known human viruses is also presented. All viral read assignments reported in this table have been manually confirmed by BLASTN analysis against the complete nt database.

Sample number		5b	5c	11b	11d
Human reads		79,951,010	96,212,072	14,612,284	30,119,587
B. enterica reads		631,733	119,186	1,670,372	1,361,453
Relative abundance (RA) of B. enterica reads compared to human reads	(# reads Bent corrected for genome size/# human reads corrected for diploid genome size)	6.65	1.04	96.26	38.06
Fold change in B. enterica RA in Pre v. post treatment	(pre treatment RA B. enterica/post RA B. enterica)	6.38		2.53	

Table S5. Relative abundance of *B. enterica* reads. The relative abundance of *B. enterica* compared to human reads is represented in this table, as is the fold-reduction in *B. enterica* abundance with antibiotic-initiation. Note that the post-antibiotic biopsies were not taken at convalesence; rather, they were taken in the clinical setting of continued diarrhea.

MGH UCB SCT duodenum		**MGH UCB SCT stomach**	
Non-human Organism present in sample	**Number of reads**	**Non-human Organism present in sample**	**Number of reads**
Escherichia coli	300	*Escherichia coli*	532
Ralstonia pickettii	115	*Ralstonia pickettii*	234
Propionibacterium acnes	109	*Propionibacterium acnes*	182
Ralstonia solanacearum	66	*Ralstonia solanacearum*	126
Acidovorax ebreus	33	*Acidovorax ebreus*	89
Acidovorax sp.	18	Human Herpesvirus-7	62
Polyomavirus HPYV6	17	Acidovorax sp.	55
Burkholderia sp.	12	***Bradyrhizobium enterica***	32
Acinetobacter baumannii	11	*Pseudomonas putida*	31
Bradyrhizobium enterica	10	*Acinetobacter baumannii*	29
Pseudomonas putida	9	*Acinetobacter calcoaceticus*	28
Stenotrophomonas maltophilia	8	*Enterobacter cloacae*	27
Staphylococcus epidermidis	8	*Stenotrophomonas maltophilia*	25
Pseudomonas entomophila	8	*Pseudomonas mendocina*	24
Agrobacterium sp.	8	*Alicycliphilus denitrificans*	24
Pseudomonas fluorescens	7	*Pseudomonas entomophila*	20
Acidovorax avenae	7	*Pseudomonas stutzeri*	17
Kluyveromyces thermotolerans	6	*Comamonas testosteroni*	17
Pseudomonas mendocina	5	*Pseudomonas aeruginosa*	16
Delftia sp.	5	*Dechloromonas aromatic*	16

Table S6. PathSeq-based quantification of reads assigned to non-human organisms (viruses, bacteria, fungi) in sequenced samples from MGH. Biopsies from the duodenum and stomach of a patient with a diarrheal syndrome, with some features suggestive of cord colitis syndrome, were subjected to shotgun whole genome sequencing and subsequent PathSeq analysis. Non-human, non-phage reads are assigned as demonstrated (the top twenty non-human, non-phage organisms to which reads mapped are presented).

Patient ID	Days post-SCT of CCS diagnosis	Current diarrhea	Current antibiotics	Alive at last follow up	Days post SCT of TP1 Ig level	TP1 IgA	TP1 IgM	TP1 IgG	Days post SCT of TP2 Ig level	TP2 IgA	TP2 IgM	TP2 IgG	Date of IVIG	GD
1	153	N	None	Y	122	14	11	536	153	11	10	442		Y
2	117	N	TMP/SMX for PCP prophylaxis	Y	79	42	89	306	120	NA	NA	643	278	Y
3	101	N	TMP/SMX for nocardiosis	N	-101	104	91	922	136	28	74	415		Y
4	111	N	None	N		NA	NA	NA	122	NA	NA	925		Y
5	187	Y*	Ciprofloxacin & metronidazole	Y	173	NA	NA	701	187	NA	NA	607		Y
6	197	N	None	N	127	NA	NA	648	197	NA	NA	613		Y
7	146	N	None	N		NA	NA	NA	205	NA	NA	397		Y
8	135	N	Valacyclovir for VZV prophylaxis	Y	109	NA	NA	734	187	NA	NA	430		Y
9	335	N	None	N	298	NA	NA	368	348	NA	NA	88	201	Y
10	271	N	TMP/SMX for PCP prophylaxis	Y	212	NA	NA	349	244	NA	NA	365	452	Y
11	325	N	Atovaquone for PCP prophylaxis	Y	312	NA	NA	477	323	NA	NA	1120	321	Y

Table S7. Clinical status and Immunoglobulin levels of cord colitis syndrome patients. Abbreviations are as follows: *Lost to follow up in May 2011; TMP/SMX: Trimethoprim, Sulfamethoxazole; PCP: Pneumocystis (carinii) jiroveci; Ig: Immunoglobulin; IVIG: Intravenous immunoglobulin; TP: Time point; GD: Gut decontamination

XII. References

1. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 2008;18:821-9.
2. Gnerre S, Maccallum I, Przybylski D, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A 2011;108:1513-8.
3. Butler J, MacCallum I, Kleber M, et al. ALLPATHS: de novo assembly of whole-genome shotgun microreads. Genome Res 2008;18:810-20.
4. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.
5. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-9.
6. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403-10.
7. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:119.
8. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 1970;48:443-53.
9. Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics. Genome Res 2009;19:1639-45.
10. Lieberman TD, Michel JB, Aingaran M, et al. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. Nat Genet 2011;43:1275-80.
11. Van Rossum D, Schuurmans FP, Gillis M, et al. Genetic and phenetic analyses of Bradyrhizobium strains nodulating peanut (Arachis hypogaea L.) roots. Appl Environ Microbiol 1995;61:1599-609.
12. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 2012.
13. Swidsinski A. Standards for bacterial identification by fluorescence In situ hybridization within eukaryotic tissue using ribosomal rRNA-based probes. Inflamm Bowel Dis 2006;12:824-6; author reply 6-7.
14. Swidsinski A, Loening-Baucke V, Lochs H, Hale LP. Spatial organization of bacterial flora in normal and inflamed intestine: a fluorescence in situ hybridization study in mice. World J Gastroenterol 2005;11:1131-40.
15. Loy A, Maixner F, Wagner M, Horn M. probeBase--an online resource for rRNA-targeted oligonucleotide probes: new features 2007. Nucleic Acids Res 2007;35:D800-4.