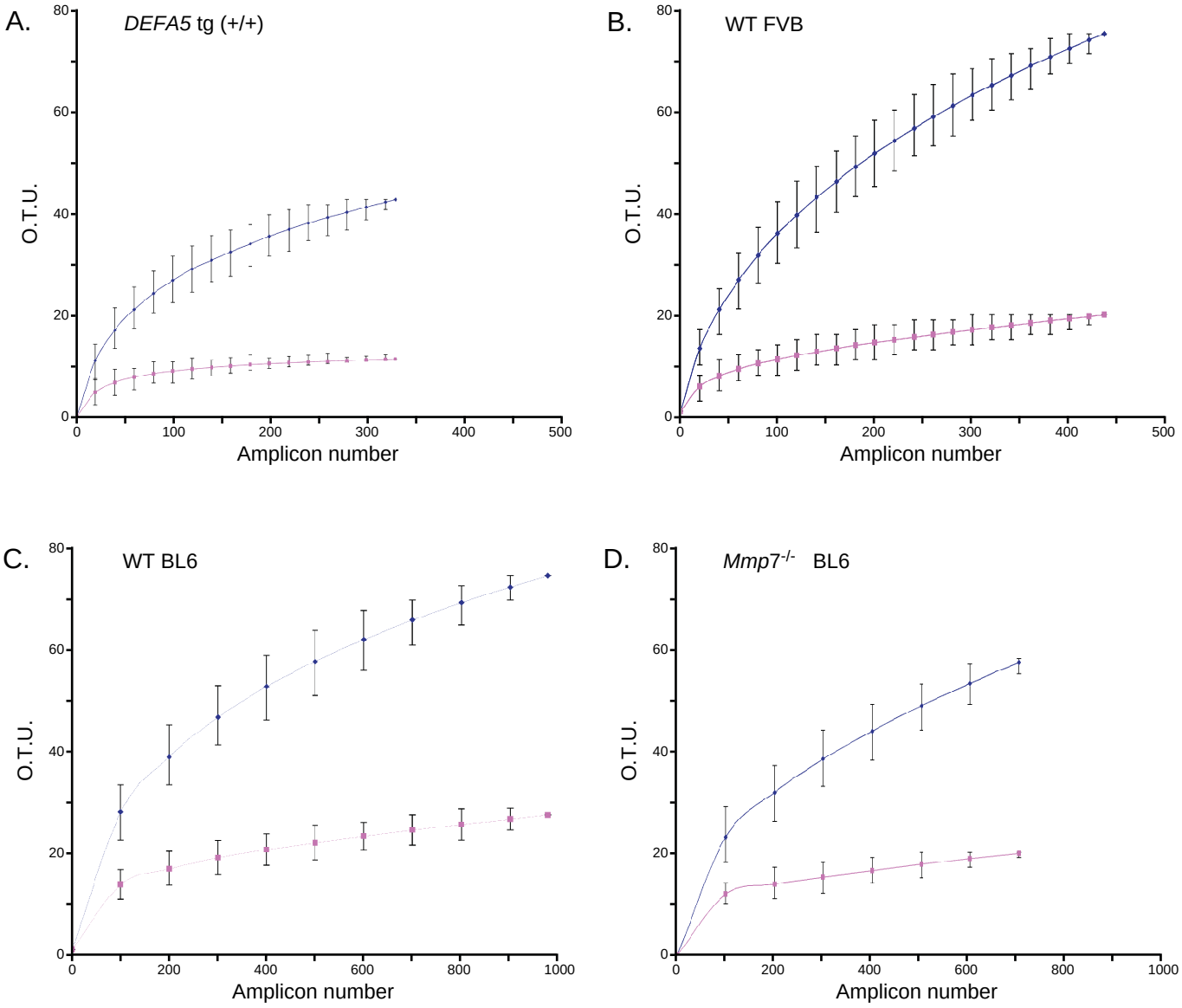
Enteric defensins are essential regulators of intestinal microbial ecology.

Nita H. Salzman*, Kuiechun Hung, Dipica Haribhai, Hiutung Chu, Jenny Karlsson-Sjöberg, Elad Amir, Paul Teggatz, Melissa Barman, Michael Hayward, Daniel Eastwood, Maaike Stoel, Yanjiao Zhou, Erica Sodergren, George M. Weinstock, Charles L. Bevins, Calvin B. Williams[2,3], and Nicolaas A. Bos.

*Corresponding Author

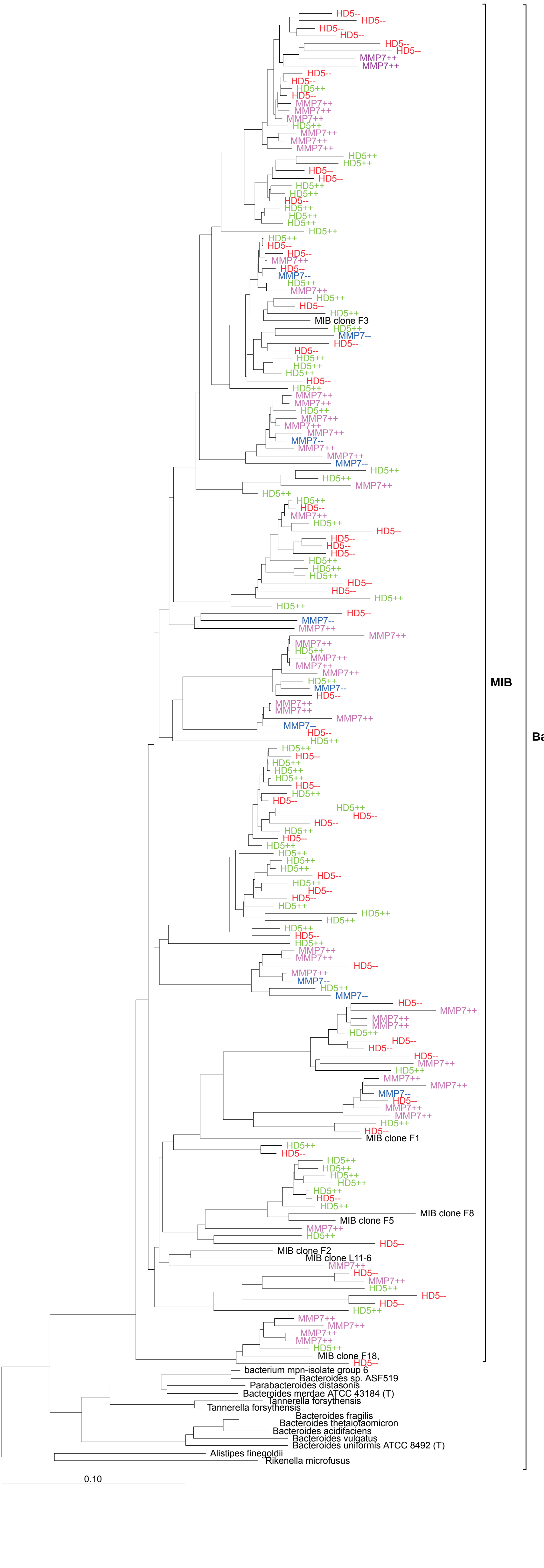
Supplementary Fig. 1



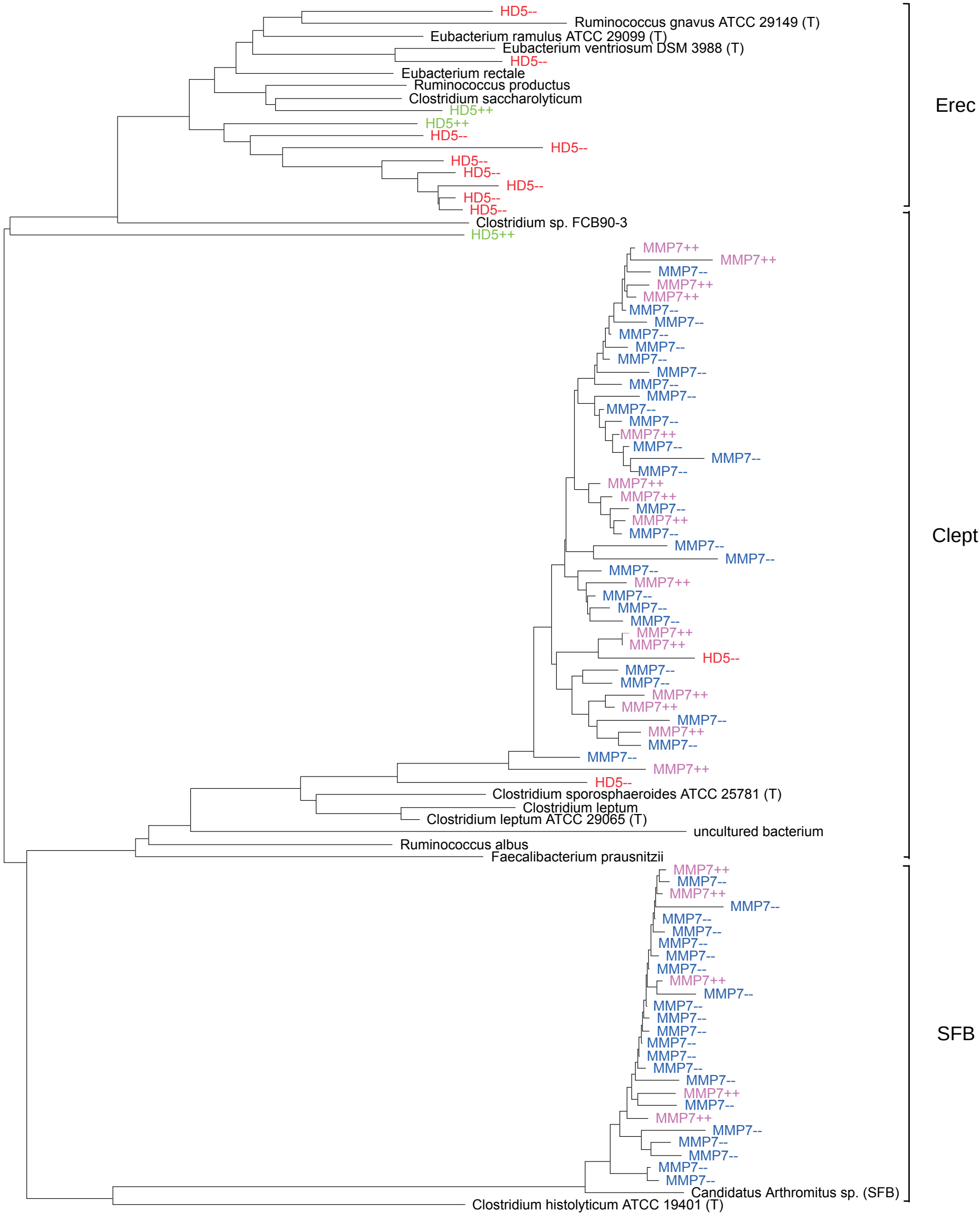
Supplementary Fig. 1. Rarefaction curves demonstrating community coverage at the phylum and class level. Rarefaction curves were generated for sequences identified from the distal small intestines of **(a)** *DEFA5* tg (+/+) FVB, **(b)** *DEFA5* tg (-/-) WT FVB, **(c)** *Mmp7*$^{+/+}$(WT B6), and **(d)** *Mmp7*$^{-/-}$ (B6) mice, using DOTUR [1], both at the phylum and class level. The rarefaction curve depicts the number of operational taxonomic units (OTU) observed at different sampling depth, where the X axis is the number of sequenced amplicons and the Y axis is the number of OTU observed. Flattening of the curve suggests that sampling is of adequate depth and that additional sampling will yield few additional OTU. Pink (squares) represent analysis by phylum, while blue (diamonds) represent analysis by class.

Supplementary Fig. 2a

Supplementary Fig. 2b



Supplementary Fig. 2c



Supplementary Fig. 2. Phylogenetic analysis of bacterial composition of ileum. DNA isolated from the ileum of mice was pooled by genotype using litters from both *DEFA5* tg and *Mmp7* knockout lines. 16S rDNA sequences were obtained by PCR amplification, cloning, and sequencing. **(a)** Phylogenetic sequence analyses of 309 clones. ARB software [2] was used for analysis and the sequences were aligned to the closest related 16S sequences of known, described bacteria within the ARB-Silva database (http://www.arb-silva.de). The tree was calculated using the neighbor-joining method with Jukes-Cantor 2-correction parameter. The distance matrix used in the neighbor-joining method included stretches of sequence corresponding to *E.coli* positions 27 to 1480. HD5++, *DEFA5* tg (+/+); HD5--, *DEFA5* tg (-/-) (WT FVB); MMP7++, $Mmp7^{+/+}$ (WT B6); MMP7--, $Mmp7^{-/-}$. **(b)** Bacteroidetes sub-tree. Sequences were analyzed using the ARB software to determine whether they contained the target sequence of the primer sets used for qPCR, as noted by the brackets on the right. *DEFA5* tg (+/+) sequences are in green, *DEFA5* tg (-/-) (WT FVB) sequences are in red, $Mmp7^{+/+}$ (WT B6) sequences are in violet, and $Mmp7^{-/-}$ sequences are blue. HD5++ (green), *DEFA5* tg (+/+); HD5-- (red), *DEFA5* tg (-/-) (WT FVB); MMP7++ (purple), $Mmp7^{+/+}$ (WT B6); MMP7-- (blue), $Mmp7^{-/-}$. **(c)** Phylogenetic tree for Clostridiae was completed with the closest known relatives. Sequences were analyzed using the ARB software to determine whether they contained the target sequence of the primer sets used for qPCR, as noted by the brackets on the right. HD5++ (green), *DEFA5* tg (+/+); HD5-- (red), *DEFA5* tg (-/-) (WT FVB); MMP7++ (purple), $Mmp7^{+/+}$ (WT B6); MMP7-- (blue), $Mmp7^{-/-}$.

a.

Bacteria (phylum - class)	***DEFA5* tg (+/+)**				***DEFA5* tg (+/+) combined**	***DEFA5* tg (-/- , WT)**							***DEFA5* tg (-/-, WT) combined**
Firmicutes	**11**	**22**	**21**	**40**	**94**	**50**	**30**	**34**	**85**	**24**	**55**	**83**	**361**
-Bacilli	10	1	10	17	38	49	9	20	27	12	4	7	128
-Clostridia	1	0	1	9	11	1	2	14	15	6	51	3	92
- Erysipelotrichi	0	20	10	14	44	0	19	0	43	6	0	73	141
Bacteroidetes	**74**	**63**	**68**	**49**	**254**	**33**	**59**	**52**	**2**	**59**	**14**	**0**	**219**
-Bacteroidetes	74	63	68	49	254	33	59	52	2	59	14	0	219
Actinobacteria	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**1**	**0**	**2**
Proteobacteria	**5**	**6**	**4**	**3**	**18**	**5**	**3**	**2**	**0**	**5**	**0**	**0**	**15**
Deferribacteres	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**18**	**0**	**19**
Total	90	91	93	92	366	89	92	89	87	88	88	83	616

b.

Bacteria (phylum - class)	***Mmp7*$^{-/-}$**								***Mmp7*$^{-/-}$ combined**	***Mmp7*$^{+/+}$ (WT)**									***Mmp7*$^{+/+}$ combined**
Firmicutes	**50**	**37**	**47**	**77**	**53**	**51**	**134**	**41**	**490**	**32**	**18**	**53**	**16**	**35**	**25**	**108**	**40**	**47**	**374**
-Bacilli	11	0	3	9	26	17	10	6	82	6	4	4	1	2	4	54	12	7	94
-Clostridia	38	37	44	67	27	34	124	35	406	26	14	49	15	33	21	54	28	40	280
-Erysipelotrichi	1	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
Tenericutes	**5**	**15**	**7**	**16**	**1**	**23**	**26**	**2**	**95**	**3**	**0**	**10**	**0**	**16**	**10**	**9**	**5**	**9**	**62**
Bacteroidetes	**18**	**32**	**14**	**0**	**28**	**12**	**19**	**5**	**128**	**45**	**52**	**21**	**52**	**27**	**39**	**42**	**34**	**24**	**336**
-Bacteroidetes	18	32	14	0	28	12	19	5	128	45	52	21	52	27	39	42	34	24	336
Actinobacteria	**5**	**1**	**1**	**0**	**2**	**2**	**4**	**1**	**16**	**0**	**1**	**2**	**1**	**1**	**0**	**4**	**1**	**0**	**10**
Proteobacteria	**3**	**4**	**10**	**2**	**4**	**0**	**6**	**4**	**33**	**9**	**5**	**3**	**20**	**7**	**15**	**12**	**7**	**2**	**80**
Deferribacteres	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**2**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**1**
Total	82	89	79	95	89	88	189	53	764	89	76	89	89	86	89	176	87	82	863

Supplementary Table 1. 16S subclone categorization by bacterial phylum. Each full-length 16S rRNA sequence was analyzed using the Ribosomal Database Project II (RDP) classifier and categorized by phylum. Columns represent the sequence analysis of the distal small intestine of individual mice. Mice were **(a)** *DEFA5* tg (+/+) and *DEFA5* tg (-/-, WT FVB) littermates, or **(b)** *Mmp7*$^{-/-}$ and *Mmp7*$^{+/+}$ (WT B6) littermates. Bolded text values indicate phylum and unbolded text values indicate class within the plylum.

Group	Reference strain	Primers	Product length	Reference
Eubacteria	*Ruminococcus productus* (ATCC 27340D)	UniF334: ACTCCTACGGGAGGCAGCAGT UniR514: ATTACCGCGGCTGCTGGC	172 bp	[3]
Bacteroides (Bact)	*Bacteroides fragilis* (ATCC 25285D)	BactF: GGTTCTGAGAGGAGGTCCC UniR349: CTGCCTCCCGTAGGAGT	65	[4]
Mouse Intestinal Bacteria (MIB)	Plasmid DNA (CT11-6)	UniF516-CCAGCAGCCGCGGTAATA MIBR-CGCATTCCGCATACTTCTC	162	[5]
Lactobacillus/Enterococcus group (Lact)	*Lactobaccillus acidophilus* (ATCC 4357D)	LABF: AGCAGTAGGGAATCTTCCA LABR: CACCGCTACACATGGAG	341	[6]
E. rectale/C. coccoides group (Erec)	*Ruminococcus productus* (ATCC 27340D)	UniF338: ACTCCTACGGGAGGCAGC CcocR: GCTTCTTAGTCAGGTACCGTCAT	139	[7]
C. leptum (Clept)	Plasmid DNA (Mmp7+/+-3)	ClepF1123: GTTGACAAAACGGAGGAAGG ClepR1367:GACGGGCGGTGTGTACAA	244	[8]
Segmented Filamentous Bacteria (SFB)	Plasmid DNA (CT25-6)	SFBF –GACGCTGAGGCATGAGAGCAT SFBR -GACGGCACGGATTGTTATTCA	109	[9]

Supplementary Table 2. 16S rDNA group-specific and kingdom-specific oligonucleotide primers for qPCR. Primer sequences for qPCR were designed based on published probes specific for the different bacterial groups in combination with a universal primer. Groups were defined in the ARB software [2] based on the published probe in combination with a suitable universal primer and compatibility of primer combinations for qPCR was analyzed by Primer Designer 5 (Sci Ed Programs). For the *Bacteroides* group the Bac303 probe [4] for the *E. rectale/ C. coccoides* group the Erec482 probe [7], for the *Mouse Intestinal Bacteroides* the MIB661 probe [5] and for the Eubacteria group the Bact338 probe [3] were used for definition of the specified groups. For *Segmented Filamentous Bacterium* a new combination of primers was designed based on the published 16S DNA sequence [9]. For the *Lactobacillus/Enterococcus* group and the *Clostridium leptum* group, published PCR primer sets were used [6, 8].

Supplementary Table 3. Oligonucleotide primers for qPCR analysis of Paneth cell products.

Gene[1]	Sense	Antisense
Gapdh	5' TCATCAACGGGAAGCCCATCAC 3'	5' AGACTCCACGACATACTCAGCACCG 3'
Lyz1	5' GCCAAGGTCTACAATCGTTGTGAGTTG 3'	5' CAGTCAGCCAGCTTGACACCACG 3'
Llyz2	5' GGCTGGCTACTATGGAGTCAGCCTG 3'	5' GCATTCACAGCTCTTGGGGTTTTG 3'
Pla2g2	5' AGGATTCCCCCAAGGATGCCAC 3'	5' CAGCCGTTTCTGACAGGAGTTCTGG 3'
Defa-rs1c	5' CACCACCCAAGCTCCAAATACACAG 3'	5' ATCGTGAGGACCAAAAGCAAATGG 3'
Defcr1	5' TCAAGAGGCTGCAAAGGAAGAGAAC 3'	5' TGGTCTCCATGTTCAGCGACAGC 3'
Defcr4(FVB)[2]	5' GCTGTGTCTATCTCCTTTGGAGGC 3'	5' CGTATTCCACAAGTCCCACGAAC 3'
Defcr4(B6)[2]	5' CCAGGGGAAGATGACCAGGCTG 3'	5' TGCAGCGACGATTTCTACAAAGGC 3'

[1] – The common names for these genes are: *Lys1*, P-lysozyme; *Lyz2*, M-lysozyme, *Pla2g2*, sPLA2, *Defa-rs1c*, CRS-1c; *Defcr1*, cryptdin-1; *Defcr4*, cryptdin-4.

[2] – There is strain-specific differences in defcr4. The primers target FVB and B6 sequences as indicated.

Supplementary Methods.

Amplification, subcloning, sequencing, and sequence alignment. Amplification of 16S rDNA was carried out using primers 27F (5'-AAGAGTTTGATCMTGGCTC-3') and 1492R (5'-TACGGYTACCTTGTTACGACTT-3'). PCR fragments were subcloned into TOP TA pCR4.0, followed by transformation into *Escherichia coli* TOP10 (Invitrogen). Individual sub-clones were sequenced on 3730 XL Applied Biosystems capillary sequencer to a depth of 96 amplicons. The M13F (5'-TGTAAAACGACGGCCAGT-3'), M13R (5'-CAGGAAACAGCTATGACC-3') and 907R (5'-CCGTCAATTCCTTTRAGTTT-3') primers were used for sequencing. The three sequences were assembled using the PHRED and PHRAP software package. Bases with a PHRAP quality score of <10 were trimmed prior to assembly. Short sequences (<1250 nts) and poor quality sequences were removed prior to alignment. Chimeric sequences were also removed prior to alignment using Kr0yer version 0.0.1, a tool developed by the Genome Center (Washington University, St. Louis). Sequence alignments were done using the online NAST (Nearest Alignment Space Termination) algorithm from greengenes http://greengenes.lbl.gov/cgi-bin/nph-NAST_align.cgi. Both conserved regions and hypervariable regions were compared in NAST.

Construction of phylogenetic trees. For construction of phylogenetic trees, equal quantities of genomic DNA were pooled from mice of each genotype. The 16S sequences from the pooled DNA were amplified and subcloned as described above. Plasmids were sequenced at the University of Chicago Cancer Research Center DNA sequencing facility (Chicago, IL). Sequences were edited using the SeqMan program in the DNAstar software package. Phylogenetic analyses of the sequences were done using ARB software [2]. Sequences were aligned against the closest related 16S sequences of known, described bacteria within the ARB-Silva database (http://www.arb-silva.de). The phylogenetic tree was completed with the closest known relatives. Sequences were tested using the ARB software to determine whether they contained the target sequence of the primer sets used for qPCR.

Quantitative PCR for microbiota analysis. The abundance of specific intestinal bacterial groups was measured by qPCR using the MyiQ Single-Color Real-Time PCR Detection System (BioRad) using group specific 16S rDNA primers (Operon Technologies) (**Supplementary Table 2**) as previously described [10]. A short segment of the 16S rRNA gene (174 bp) was specifically amplified by real time PCR using the conserved 16S rRNA specific primer pair UniF340 and UniR514 to determine the total amount of bacteria in each intestinal segment. The real time PCR program started with an initial step at 95°C for 3 minutes, followed by 40 cycles of 10s at 95°C and 45s at 63°C, using an IQ SYBR Green Supermix (Biorad). Data was acquired in the final step at 63°C. Using genomic DNA from each sample, real time PCR reactions were completed using group specific primers to determine the amount of bacteria in each of the following major groups: *Eubacterium rectal-/Clostridium coccoides* (Erec), *Clostridium leptum* (Clept), *Lactobacillus sp*. (Lact), *Bacteroides sp*. (Bact), Mouse Intestinal Bacteria (MIB), and Segmented Filamentous Bacteria (SFB). Bacterial numbers were determined using standard curves constructed with reference bacterial DNA specific for each bacterial group analyzed (**Supplementary Table 2**). qPCR measures 16S gene copies per sample, not actual bacterial numbers or colony forming units. Nevertheless, these values are directly related and correlate well.

Paneth cell effector expression. Tissue samples from the distal small intestine were isolated and homogenized in a guanidine thiocyanate buffer. RNA was isolated by cesium chloride ultracentrifugation as described [11]. Isolated total RNA was quantified by ultraviolet absorbance at 260 nm using a spectrophotometer (Nanodrop ND-1000), and 0.5 μg-1.0μg of this RNA was reverse transcribed with reverse transcriptase using an oligo-(dT) 12-18 primer according to the supplier's protocol (Superscript II , Invitrogen). The reaction product was then treated with RNase-H (2units), and the resulting tissue specific cDNA was purified using column adsorption chromatography (Qiagen). The eluate was diluted to an equivalent of 10ng/ul in 10mM Tris-HCl (pH 8.5) based on the initial input concentration of total RNA. Gene specific real-time PCR primers (**Supplementary Table 3**) were selected using MacVector software (MacVector, Inc.), and purchased from Invitrogen. Real-

time PCR was performed on the tissue-specific cDNA as a template with specific oligonucleotide primer pairs as described previously [11]. Each 10μl PCR reaction contained 4mM $MgCl_2$, 0.5μg of each oligonucleotide primer and 1x LightCycler-Fast Start DNA SYBR Green I Mastermix (Roche Diagnostics). The qRT-PCR was performed using the LightCycler 2.0 system (Roche Diagnostics). A no-template reaction was included as a negative control for each qRT-PCR experiment, and for absolute quantification gene-specific plasmid standards were included within every set of reactions. The PCR cycling times and temperatures, and methods of data analysis were described previously [11]. All samples were analyzed in duplicate, and variation between duplicates was < 10% for every reported value.

Supplementary References

1. Schloss, P.D. & Handelsman, J. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied & Environmental Microbiology* **71**, 1501-1506 (2005).
2. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Research* **32**, 1363-1371 (2004).
3. Amann, R.I. *et al.* Combination of 16S rRNA-Targeted Oligonucleotide Probes with Flow Cytometry for Analyzing Mixed Microbial Populations. *Appl Environ Microbiol* **56**, 1919-1925 (1990).
4. Dore, J., Sghir, A., Hannequart-Gramet, G., Corthier, G. & Pochart, P. Design and evaluation of a 16S rRNA-targeted oligonucleotide probe for specific detection and quantitation of human faecal Bacteroides populations. *Syst Appl Microbiol* **21**, 65-71 (1998).
5. Salzman, N.H. *et al.* Analysis of 16S libraries of mouse gastrointestinal microflora reveals a large new group of mouse intestinal bacteria. *Microbiology* **148**, 3651-3660 (2002).
6. Rinttila, T., Kassinen, A., Malinen, E., Krogius, L. & Palva, A. Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. *Journal of Applied Microbiology* **97**, 1166-1177 (2004).
7. Franks, A.H. *et al.* Variations of Bacterial Populations in Human Feces Measured by Fluorescent In Situ Hybridization with Group-Specific 16S rRNA-Targeted Oligonucleotide Probes. *Appl Environ Microbiol* **64**, 3336-3345 (1998).
8. Matsuki, T., Watanabe, K., Fujimoto, J., Takada, T. & Tanaka, R. Use of 16S rRNA gene-targeted group-specific primers for real-time PCR analysis of predominant bacteria in human feces. *Applied & Environmental Microbiology* **70**, 7220-7228 (2004).
9. Snel, J. *et al.* Comparison of 16S rRNA Sequences of Segmented Filamentous Bacteria Isolated from Mice, Rats, and Chickens and Proposal of "*Candidatus* Arthromitus". *Int J Syst Bacteriol* **45**, 780-782 (1995).
10. Barman, M. *et al.* Enteric Salmonellosis disrupts the microbial ecology of the murine gasrointestinal tract. *Infect Immun* **76**, 907-915 (2008).
11. Wehkamp, J. *et al.* Paneth cell antimicrobial peptides: topographical distribution and quantification in human gastrointestinal tissues. *FEBS Letters* **580**, 5344-5350 (2006).