SUPPLEMENTAL MATERIALS

Methods

Mouse genotyping

Genotyping for the Nod2 knockout allele was performed on mouse DNA extracted from tail clippings. Analysis of the Nod2 knockout allele was performed on genomic DNA extracted with REDExtract-N-Amp extraction buffer (Sigma-Aldrich, St. Louis, MO), while genomic DNA extracted with 50 mM sodium hydroxide was used to determine the presence of the WT allele. PCR analysis identified *Nod2*[-/-] mice based on the increased size of exon 3 in the disrupted *Nod2* gene (Suppl. Fig. 2). Specific primer sequences used were as follows: 1) WT allele: forward, 5′-ACAGAGATGCCGACACCATACTG-3′; reverse, 5′-TGGAGAAGGTTGAAGAGCAGAGTC-3′; and 2) Nod2 knockout allele: forward, 5′-TGACTGTGGCTAATGTCCTTTGTG-3′; reverse, 5′-TTCTATCGCCTTCTTGACGAGTTC-3′. Confirmation of the knockout allele was accomplished by sequencing the PCR genotyping products, which showed that the larger amplicon includes a thymidine kinase gene that is part of the engineered construct that was originally used to disrupt the Nod2 gene (Suppl. Figs. 2-3).[1]

Quantitative reverse-transcriptase polymerase chain reaction

Total RNA was extracted from ileal tissue using an RNeasy Mini Kit (Qiagen, Valencia, CA) per manufacturer's instructions. Complementary DNA was generated using SuperScript II reverse transcriptase (Invitrogen, Carlsbad,

CA). Quantitative RT-PCR was performed using TaqMan Gene Expression

Master Mix (Applied Biosystems, Foster City, CA) per manufacturer's

instructions. Specific primer/probe sets were obtained from Applied Biosystems

as follows: *Actb* (Mm00607939_s1), *Lyz* (Mm00727183_s1), *Ang4*

(Mm03647554_g1), *Reg3γ* (Mm00441127_m1), *Defa3* (Mm04205962_gH),

Defa5 (Mm00651548_g1), *Defa4* (Mm00651736_g1), *Defa20* (Mm00842045g1),

and *Defa-rs10* (Mm00833275_g1). The following primer sets were utilized with

SYBR Green PCR Master Mix (Applied Biosystems), per manufacturer's

protocols: 1) Global α-defensins: forward, 5′-

GGTGATCATCAGACCCCAGCATCAGT-3′; reverse, 5′-

AAGAGACTAAAACTGAGGA GCAGC-3′; 2) CRS1C: forward, 5′-

TGCTCTTCAAGATGTAGCCCAACG-3′; reverse, 5′-

TGGAGCTTGGGTGGTGATTGCA-3′; and 3) CRS4C: forward, 5′-

GCATGGAATCTGGGTCAAGATAAC-3′; reverse, 5′-

AGAAGGAAGAGCAATCAAGGCTAAG-3′.

Acid urea polyacrylamide gel electrophoresis

Samples were subject to 2 extraction steps using 60% acetonitrile, 1%

trifluoroacetic acid (TFA), incubated at 4°C with rotation and clarification by

centrifugation. Resulting supernatants were lyophilized, re-suspended and

dialyzed in 5% acetic acid. Dialysates were then lyophilized, and 300 µg protein

aliquots were solubilized in 30 µl of AU-PAGE loading solution (3 M urea, 5%

acetic acid). These were next electrophoresed on a 12.5% AU-PAGE gel for 1 h

at 150 V and 4 h at 400 V, alongside a sample of recombinant α-defensin 4 provided by Dr. André Ouellette (University of Southern California). Resolved proteins were then visualized by staining with 0.05% Coomassie Brilliant Blue in 30% methanol and 15% formalin, followed by destaining in 25% methanol and 1% formalin.

Individual Coomassie-stained AU-PAGE bands were excised and de-stained twice in 50% acetonitrile, 25 mM ammonium bicarbonate at room temperature for 10 min. Lyophilized gel slices were extracted for protein with 0.5% TFA, 50% acetonitrile in two steps at room temperature for 10 min. Extracted protein was subsequently lyophilized and re-suspended in 0.1% TFA for analysis. Samples were submitted to the Proteomics Core Facility of the University of North Carolina at Chapel Hill School of Medicine for mass spectrometric analysis via matrix assisted laser desorption ionization-time of flight tandem mass spectrometry in the linear mode using α-cyano-4-hydrocinnamic acid for the matrix.

Immunohistochemistry

2 cm segments of terminal ileum were fixed in 10% phosphate-buffered formalin for 12 hr. The tissue segments were longitudinally embedded in paraffin and cut in 5 μm sections for histological analysis. Sections were de-parrafinized in separate containers of fresh xylene for a total of 8 min. Rehydration was accomplished in a series of ethanol dilutions. Sections were then treated with 3% hydrogen peroxide to inhibit the action of endogenous peroxidase, blocked

with 3% bovine serum albumin (BSA) in phosphate buffered saline (PBS) for 1 h at 25°C, and incubated with 1:1500 rabbit polyclonal antibody to human lysozyme (Diagnostic BioSystems, Pleasanton, CA) in 3% BSA overnight at 4°C. Sections were subsequently incubated with 1:200 biotinylated anti-rabbit IgG (Vector Laboratories, Inc., Burlingame, CA) in 1% BSA for 30 min at 25°C, and a 1:50 dilution of Avidin DH and biotinylated enzyme (Vector Laboratories, Inc.) in PBS for 30 min at 25°C. The sections were visualized with DAB chromogen reagent (Dako, Carpinteria, CA).

Bactericidal gel overlay assay

Bacterial strains *E. coli* NC101[2], *Listeria monocytogenes* 10403S, and *Salmonella enterica* serovar Typhimurium were grown to mid-log phase in trypticase soy broth (TSB) media, washed with 10 mM sodium citrate-phosphate buffer, re-suspended in warm 0.03% TSB, 1% low-melt agarose, 10 mM sodium citrate-phosphate buffer and 0.02% Tween 20 at 4×10^5 CFU/mL, and plated onto a Petri dish as a 1 mm deep undergel. Ileal protein samples (100 µg) were prepared by electrophoresis on a small-scale 12.5% AU-PAGE gel for 1 h at 150V. Resolved gels were washed with ice-cold 10 mM sodium phosphate buffer for 15 min, placed atop the bacteria-laden solid agarose layer, and incubated at 37°C for 3 h. Subsequently, the gel was removed and replaced with a layer of warm 6% TSB, 0.8% low-melt agarose to form a nutrient-rich overgel. Gel overlay plates were incubated overnight at 20°C and then imaged for band-associated zones of bacterial clearance.

Bacterial composition analyses

     1) <u>Bacterial DNA extraction</u> – 100 mg of frozen tissue or feces was re-suspended in sterile lysis buffer (200 mM NaCl, 100 mM Tris [pH 8.0], 20 mM EDTA, 20 mg/ml lysozyme [Sigma-Aldrich, St. Louis, MO]) for 30 min at 37°C. This was then supplemented with 40 μl of proteinase K (20 mg/ml) and 85 μl of 10% SDS and incubated for 30 min at 65°C. Homogenization was accomplished by adding 300 mg of 0.1 mm zirconium beads (BioSpec Products, Bartlesville, OK) and bead beating for 2 min (BioSpec). Supernatants were then isolated and DNA extracted using phenol/chloroform/iso-amyl alcohol (25:24:1), followed by precipitation with absolute ethanol for 1 hr at -20°C. Finally, the precipitated DNA was cleaned up using a DNeasy Blood and Tissue extraction kit (Qiagen) per manufacturer's instructions.

     2) <u>16S rRNA gene sequencing</u> – Forward primers were tagged with 10 bp unique barcode labels at the 5' end along with the adaptor sequence to allow multiple samples to be included in a single 454 GS FLX Titanium sequencing plate.[3, 4] 16S rRNA PCR products were quantified, pooled, and purified for the sequencing reaction. 454 GS FLX Titanium sequencing was performed on a 454 Life Sciences Genome Sequencer FLX machine (Roche, Florence, SC) at the microbiome core at UNC-Chapel Hill.

     3) <u>Analysis of 16S rRNA sequences</u> – Taxonomic and phylogenetic analyses of 16S rRNA sequence data generated by the 454 GS FLX Titanium sequencer were processed as follows. Sequences were removed if (i) they were less than 300 base-pairs, (ii) there was not an exact match to a 5' primer, or (iii)

there were any N's in the sequences. 135,982 sequences met all QC criteria.

These sequences were clustered with the program AbundantOTU[5] which

produced 257 OTUs representing 126,526 (93%) of the sequences. Consensus

sequences were checked for chimeras with UCHIME[6] using the gold database

(http://greengenes.lbl.gov/Download/Sequence_Data/Fasta_data_files/gold_strai

ns_gg16S_aligned.fasta.gz) as reference. UCHIME detected 5 chimeras in the

consensus sequences, for which the corresponding OTUs were removed from

downstream analysis. To check for non-microbial contamination, consensus

sequences were mapped with blastn to version 108 of the Silva 16S rRNA

database (http://www.arb-silva.de/). All consensus sequences had a match of >=

295 basepairs with a percent identity >= 91%. We conclude that non-microbial

contamination was not a significant problem in our dataset. OTUs were assigned

to a taxonomy using the java-based stand-alone version of the RDP classifier (v.

2.4). Consensus sequences representing each OTU are given in supplemental

File 1.

   To correct for variation in total sequence counts between samples, the

abundance of each OTU in a given sample was standardized by calculating the

logged sequence abundance using the following formula:

   LOG_{10} [(Frequency/ # sequences in sample)*Average # of

   sequences per sample +1]

The logarithm was used to lessen the influence of more dominant OTUs. In

order to take the log of zero, a pseudo-count (one sequence) is added to each

sample for each taxa. In order to minimize the effect of this pseudo-count, all

samples are normalized to the average number of sequences per sample before the addition of the pseudo-count.

PCoAs were performed in the package Mothur 1.24.1 (using the program "pcoa") based on Bray-Curtis dissimilarity defined as:

$$\frac{\sum_{k=1}^{k=n}|y_{ik}-y_{jk}|}{\sum_{k=1}^{k=n}(y_{ik}+y_{jk})}$$

where y_{ik} and y_{ik} are the log-transformed and normalized values for taxa k in samples i and j respectively and n is the number of taxa in all samples.

Supplemental Figure Legends

Supplemental Figure 1. *Breeding strategy used to ensure WT and Nod2$^{-/-}$ littermates were reliably generated and utilized throughout the course of the study.* Littermates were generated from Nod2$^{+/-}$ (purple) mice on a pure C57BL/6 background. WT (blue) or Nod2$^{-/-}$ (red) mice were sacrificed, while Nod2$^{+/-}$ (purple) mice were re-bred to produce additional littermates. The dashed line highlights the strategy of previous studies, in which littermates were used to start independently housed homozygous WT and Nod2$^{-/-}$ lines.

Supplemental Figure 2. *WT, Nod2$^{+/-}$, and Nod2$^{-/-}$ mice can be distinguished by PCR.* Agarose gel electrophoresis of PCR reactions of genomic DNA from eight mice: three WT (lanes 2-4), three *Nod2$^{+/-}$* (lanes 5-7), and two *Nod2$^{-/-}$* mice (lanes 8-9). The PCR reaction analyzed in lane 10 contained no template and therefore serves as a negative control.

Supplemental Figure 3. *Confirmation of WT and Nod2$^{-/-}$ PCR products.* Above: DNA sequence alignment of the 348 bp WT allele amplicon (Suppl. Fig. 2, top row) with the corresponding sequence of exon 3 of the wild-type Nod2 gene from NCBI's assembly of the C57BL/6 mouse genome. The alignment was generated by the T-COFFEE DNA sequence alignment algorithm.[7] Alignment gaps are indicated by hyphens, indeterminate bases are denoted with "N", and non-identical aligned bases are identified with red textual coloring. Below: Similar DNA sequence alignment of the 776 bp of the 5' end of the 945 bp Nod2 knockout allele amplicon (Suppl. Fig. 2, bottom row). The presence of two canonical sequences for *loxP* sites are indicated by black outlined boxes in the Nod2 knockout allele amplicon sequence. The 3' end and orientation of a thymidine kinase (tk) gene is indicated by a black, bent arrow.

Supplemental Figure 4. *Variations in α-defensin mRNA and protein expression based on mouse background strain.* Transcript expression of total *CRS4C* (A, top) and the CRS4C isoform *Defa-rs10* (A, bottom) as well as *Defa4* (B) in WT 129, WT B6 and *Nod2$^{-/-}$* B6 ileal tissue (n=6-7 mice/group). Copy number is normalized to β-actin and expressed as a fold Δ relative to the WT 129 group. Data are shown as means with SEM. N.D. - not detected. (C) AU-PAGE of ileal protein from WT 129, WT B6 and *Nod2$^{-/-}$* B6 mice. The Paneth cell α-defensin region is indicated by the black outlined box. First lane is recombinant Defa4 control; each additional lane represents a pooled sample from three mice.

Supplemental File 1. Consensus sequences representing each OTU in fasta

format.

References

1   Kobayashi KS, Chamaillard M, Ogura Y, et al. Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. Science 2005;**307**:731-4.

2   Kim SC, Tonkonogy SL, Albright CA, et al. Variable phenotypes of enterocolitis in interleukin 10-deficient mice monoassociated with two different commensal bacteria. Gastroenterology 2005;**128**:891-906.

3   Hamady M, Walker JJ, Harris JK, et al. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Methods 2008;**5**:235-7.

4   Fierer N, Hamady M, Lauber CL, et al. The influence of sex, handedness, and washing on the diversity of hand surface bacteria. Proc Natl Acad Sci U S A 2008;**105**:17994-9.

5   Ye Y. Identification and Quantification of Abundant Species from Pyrosequences of 16S rRNA by Consensus Alignment. Proceedings (IEEE Int Conf Bioinformatics Biomed) 2011;**2010**:153-7.

6   Edgar RC, Haas BJ, Clemente JC, et al. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 2011;**27**:2194-200.

7   Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 2000;**302**:205-17.