Supplemental Experimental Procedures

Bacterial Strains and Culture Conditions

To isolate commensal bacteria from C57Bl/6 mice, fecal contents were streaked on Lactobacillus MRS agar (Neogen Accumedia) plates or Schaedler agar (BD Biosciences) supplemented with 5% defibrinated sheep's blood (Hemostat Laboratories), followed by aerobic or anaerobic culture at 37°C for 24-72h. Individual colonies were picked, re-streaked twice, and frozen at -80C in 10% glycerol. Individual isolates were characterized by 16S rDNA sequencing. Briefly, colonies were resuspended in 100ul PBS and boiled for 10min at 100°C. The 16S gene was amplified by PCR using 27F (AGAGTTTGATCCTGGCTCAG) and 1429R (GGTTACCTTGTTACGACTT) primers using DreamTaq PCR reagents (Thermo/Fisher). PCR products were sequenced using the 27F and 1429R primers and reads were aligned using MacVector. 16S sequences were classified using the Michigan State University Ribosomal Database Project classifier function and NCBI BLAST program.

For calculation of colony forming units (cfu) in the mLN, organs were harvested aseptically into PBS and dissociated using a Polytron PT2100 homogenizer (Kinematica). Dilutions were plated onto Schaedler agar (BD Biosciences) supplemented with 5% defibrinated sheep's blood (Hemostat Laboratories). Plates were grown under aerobic or anaerobic culture conditions at 37C for 24-72 hours before enumeration of colonies.

For bacterial flow cytometry of *Salmonella enterica* serovar Typhimurium SL1344 and *Citrobacter rodentium,* fresh overnight cultures were obtained in LB broth and processed as described for the rest of bacterial and fecal samples.

DSS colitis

Mice were cohoused 1-7 weeks before administration of 2% dextran sodium sulfate (40,000 – 50,000 Da; Affymetrix) in the drinking water. Mice were treated with DSS for 7 days. Disease severity was measured by monitoring weight every other day. Serum samples were taken on days 6/7, 10 and 14 while fecal pellets were collected on d14 post DSS-administration.

Generation of Hybridomas

Sera from 21 week old C57BL/6J mice was screened for binding activity to bacterial isolates cultured from our colony. Mice with sera reactive to isolates were immunized with $10x10^6$ total bacteria of *Lactobacillus sp.* and *Enterobacter sp.* to boost antigen specific B cells. The pooled mixture was administered subcutaneous in both flanks, intra-peritoneal, and intra-venous. Spleens were harvested three days post immunization and B cells were enriched using CD19 MicroBeads (Miltenyi Biotec). Splenocytes were fused to murine myeloma cell line Ag8.653 using PEG 1500 (Roche) at a ratio of 1:2, respectively, and plated at a concentration of $2.5x10^6$ cells/mL in complete media (IMDM plus 4mM L-glutamine, 4,5000 mg/L of L-glucose, 10% (v/v) FCS, 50 units/L penicillin, 50µg/mL streptomycin, 1mM sodium pyruvate, 50uM b-mercaptoethanol) supplemented with 1X HFCS (Sigma) and 2X HAT (Sigma). Wells containing IgG3-secreting cells were confirmed by ELISA, and supernatants from IgG3-positive wells were screened for reactivity to *Lactobacillus sp.* and *Enterobacter sp.* isolates by flow cytometry. Limiting dilution was used to clone hybridomas producing commensal-specific antibodies.

Enzyme-linked immunosorbent assay (ELISA), ELISpot and CBA assay

Nunc Hi Affinity ELISA plates were coated with isotype-specific antibodies to IgG2b, IgG3 or IgM (500ng/mL; Jackson Immunoresearch) and blocked with PBS with 1% BSA (w/v) and 2% goat serum (Gibco; v/v). Secondary peroxidase conjugated goat antibodies to IgG2b, IgG3 or IgM (Jackson Immunoresearch) were used at 1:5000 in PBS. Purified IgG2b, IgG3, and IgM standards were from eBioscience. For NP-specific responses, mice were bled at d0 and d7 post immunization with 20µg NP-Ficoll (Biosearch Technologies) and NP-specific antibody titers were determined using NP_{30}-BSA (Biosearch Technologies) coated plates (5µg/mL) and IgG3 detection reagents. Plates were developed with 1mg/mL OPD (Sigma) in Citrate Buffer (PBS with 0.05M NaH_2PO_4 and 0.02M Citric acid) with 3M HCl acid stop. Absorbance at 490nm was measured on a SpectraMax M2. For measurements of proteins in the lumen, intestinal contents were resuspended at 100mg/mL in PBS, homogenized by vortexing and centrifuged at 13000xg for 10min. Supernantants were used to measure lipocalin-2 according to the manufacturer's instructions (R&D systems). Immunoglobulin titers were measured as described above, except for IgA, which was measured according to the manufacturer's instructions (Ebioscience).

For ELISpot analysis, multiscreen plates (Millipore) were coated with isotype-specific antibodies to IgG2b, IgG3 or IgM (500ng/mL; Jackson Immunoresearch) in PBS and blocked with PBS with 1% BSA (w/v) and 2% goat serum (v/v). Plates were washed with PBS and cells from various tissues were serially diluted in complete RPMI and incubated at 37°C overnight. Following several washes in PBS, secondary peroxidase conjugated goat

antibodies to IgG2b, IgG3 (Jackson Immunoresearch) and IgA (Southern Biotech) were used at 1:5000 in PBS to detect antibody-secreting cells. Plates were developed with AEC developing reagent (Vector Laboratories) according to manufacturer's instructions. Plates were read on an ImmunoSpot C.T.L. Analyzer (v3.2) and quantitated using ImmunoSpot 5.1 Pro.

Serum cytokines were measured using a Cytometric Bead Array assay according to the manufacturer's instructions (BD Biosciences).

Immunostaining

Feces-filled ileum and distal colon tissues were fixed in methacarn (60% dry methanol, 30% chloroform, 10% glacial acetic acid), processed into paraffin blocks, and sectioned at 5µm similar to as previously described (Johansson and Hansson, 2012). For immunostaining, slides were de-waxed with two incubations in pre-warmed 60ºC Histo-Clear (National Diagnostics), hydrated in stepwise washes of decreasing concentrations (100, 95, 70, 50, 35%) of ethanol:water, and washed twice with PBS. Slides were blocked for 30 minutes in PBS, 1% serum, 0.2% Triton X-100, rinsed with PBS and stained with 1:50 dilution of serum for 2 hours. Control (- serum) sections were incubated in PBS alone. Tissues were rinsed twice with PBS and stained with secondary FITC conjugated anti-mouse IgG2b, IgG3, IgG1, IgA antibodies (BD Biosciences) for 2 hours in the dark. Slides were counterstained with 10mg/mL DAPI for 5 min, and mounted with Vectashield Hardset (Vector Laboratories). Images were acquired on a Zeiss LSM 710 confocal microscope with the ZEN 2010 software (Zeiss). Samples were imaged with a 20x air objective and acquired at a frame size of 2048x2048 with 16-bit depth. Images were processed using FIJI (Schindelin et al., 2012).

16S rRNA Gene Sequencing and Microbial Community Analysis

Sequence quality filtering

16S rRNA sequencing was performed by the Alkek Center for Metagenomics and Microbiome Research at Baylor College of Medicine. A total of 1,087,113 paired end reads, each 250 bp in length were pre-filtered to good quality giving a total sequencing amount of 543,556,500 bp across all 70 samples. The target 16s v4 amplicon size was 254 bp, so we required all mate pairs to overlap. Any sequencing pair that did not overlap was thus discarded. We used SeqPrep (John, 2011) to merge the reads and remove Illumina adapters as follows:

- Limit Phred quality scores to a maximum of 60 (default)
- Allow up to 2% of bases in the overlapping reads to be of high quality mismatches (default)
- Require a minimum of 15 bp overlap when merging reads (default)
- Require at least 90% of aligned region to have matches (default)

Then we ran PrinSeq (Schmieder and Edwards, 2011) on the merged sequences, requiring the following:

- length greater than 30 nt (-min_len 30)
- GC between 10-90% (-min_gc 10 -max_gc 90)
- A mean Phred quality greater than or equal to 25 (-min_qual_mean 25)
- A maximum of 5 unknown nucleotides (N) in the sequence (-ns_max_n 5)
- Trim sequences from both edges until a Phred score of 22 is reached (-trim_qual_left 22 -trim_qual_right 22 -trim_qual_type min -trim_qual_rule lt -trim_qual_window 1 -trim_qual_step 1)
- Sequence must have a low-complexity entropy value greater than 5 (-lc_method entropy -lc_threshold 5)

Next we de-barcoded the sequences, requiring at most one single insertion, deletion, or mismatch. This yielded 767,758 sequences totaling 193,996,082 bp. See Table S1 for further details.

OTU clustering

We used an expectation maximization estimation of OTU abundances (McMurdie and Holmes, 2014; Paulson et al., 2013). For the complete clustering and classification pipeline, please see our GitHub code repository (https://github.com/tdseher/16s-em-clustering). The following is a high-level summary of the steps:

1. Put the merged FASTQ reads from all samples through Mothur (Schloss et al., 2009):
    a. Convert FASTQ sequences to FASTA by removing quality scores
    b. Remove sequences with homopolymer repeats longer than 10 bp
    c. Remove sequences less than 200 bp in length

   d. Align sequences to SILVA bacterial 16s database (Pruesse et al., 2007; Quast et al., 2013; Yilmaz et al., 2014) (Release 102 downloaded from http://www.mothur.org/w/images/9/98/Silva.bacteria.zip)
   e. Require bases to align between positions 13862 and 23444 on the Silva alignment
   f. Remove extraneous characters from the gapped alignments
2. Remove gaps (hyphens), effectively de-aligning the sequences
3. Dereplicate the sequences
4. Remove any sequences that appear a single time
5. Cluster sequences with USEARCH (Edgar, 2010, 2013; Edgar et al., 2011) at 97% identity to get the initial set of OTUs, also performing an initial *de novo* chimera filtering
6. Align the reads to the OTUs exhaustively (each read can map to multiple OTUs), and keep only the OTUs with 10 or more uniquely-mapped reads
7. Align the reads to the OTUs exhaustively, and use eXpress (Roberts and Pachter, 2013) to estimate the true sequence counts
8. Perform another round of *de novo* chimera filtering
9. Perform a reference-based chimera filtering using the FastTree (Price et al., 2009, 2010) "two-study filter, 3%" (downloaded from http://morgannprice.org/16S/403K/sel03n.masked.gz)
10. Align the reads to the OTUs exhaustively, and use eXpress to estimate the true sequence counts, and keep only OTUs with 10 or more reads

This yielded a total of 527 OTUs across all samples.

OTU classification

We use Qiime 1.8.0 (Caporaso et al., 2010) to assign each OTU to the Greengenes (DeSantis et al., 2006; McDonald et al., 2012; Werner et al., 2012) release 13.8 97% identity taxonomy tree using the BLASTn (Altschul et al., 1997; Altschul et al., 1990; Gish and States, 1993; Karlin and Altschul, 1990, 1993) algorithm.

Contaminant removal

Following OTU classification, we removed additional OTUs that were likely to be contaminants:

1. For each sample, we counted the number of times an OTU appears in N samples. This revealed that samples IgG3+IgA+.D4, IgG3+IgA-.D1, and IgG3+IgA-.B3 were the most likely to have spurious OTUs. Thus we removed any OTU whose family appears in less than 2 samples.
2. We remove any OTU with high abundance in all "sort" samples but not in "input" samples (i.e. Pseudomonadaceae).
3. We removed any OTU assigned to fungus or archaea or given no assignment.
4. We specifically looked at the set of "common contaminant" sequences for 16s experiments (Salter et al., 2014), and removed OTUs assigned to these taxonomies if they also had low sequence counts.

This brought the total OTU count from 527 down to 423.

Finally, to be extra sure to minimize the false-positive OTUs, we required the mean relative abundance across all samples for each OTU to be greater than 0.01%, which reduced the OTU count to 255.

Calculating UniFrac distances

After clustering and classifying 97% identity OTUs, we used Qiime to calculate unweighted UniFrac distances (Lozupone and Knight, 2005; Lozupone et al., 2011; Lozupone et al., 2007) using the Greengenes 97% identity taxonomy tree. We also used Qiime for principal coordinate analysis (PCoA) and the calculation of bootstrapped alpha diversity indices (e.g. Shannon diversity index), resampling 10 times.

Sequence availability

Merged sequences can be downloaded from the SRA study accession SRP056248. Additional information can be found at the BioProject accession PRJNA278658.

Table S1: Barcodes used for 16S Sequencing

Filtered sequences	SRA run	Barcode	Sort	Enclosure	IgG3	IgA	Sex	Mouse	Gneotype
11086	SRR1917257	ATTCTGTGAGCG	A	3	+	-	male	1	B6.SJL
12584	SRR1917260	GCACTACCGAAT	A	3	+	-	male	2	B6.SJL
9380	SRR1917263	GGTAACAGCTCG	A	3	+	-	male	3	B6.SJL
10784	SRR1917266	CAGCATGTGTTG	A	3	+	-	male	4	B6.SJL
8620	SRR1917269	GATAGCTGTCTT	B	4	+	-	female	5	B6.SJL
11380	SRR1917275	TGACCGGTCAAT	B	4	+	-	female	6	B6.SJL
7773	SRR1917279	TGTGATGGAGAA	B	4	+	-	female	7	B6.SJL
10465	SRR1917299	CCTCTACCTACG	B	4	+	-	female	8	B6.SJL
15453	SRR1917304	GTCGTCCTAAAT	B	4	+	-	female	9	B6.SJL
6213	SRR1917675	ATTATACCTCGG	C	5	+	-	male	10	B6
13867	SRR1917680	GCTAATTACGCT	C	5	+	-	male	11	B6
9833	SRR1917684	GTCATTCACGAG	C	5	+	-	male	12	B6
7257	SRR1917691	TAGAACTCACCT	C	6	+	-	female	13	B6.SJL
16626	SRR1917731	TGTGAGCACGGT	D	7	+	-	female	14	B6
9168	SRR1917738	ACACAAAGGGAG	D	7	+	-	female	15	B6
7052	SRR1917751	TTAGGTGCAGCT	D	8	+	-	female	16	B6
7563	SRR1917758	CATCGGTCAAGG	D	8	+	-	female	17	B6
14658	SRR1917258	ACGTGAGAGAAT	A	3	+	+	male	1	B6.SJL
7479	SRR1917261	CGTTTAGAGTCG	A	3	+	+	male	2	B6.SJL
9858	SRR1917264	CTTGCTGAAGAC	A	3	+	+	male	3	B6.SJL
10970	SRR1917267	GGAGGTTATCCG	A	3	+	+	male	4	B6.SJL
10620	SRR1917270	CCTTGGCTATCC	B	4	+	+	female	5	B6.SJL
9221	SRR1917276	CGTAACCAACCA	B	4	+	+	female	6	B6.SJL
7120	SRR1917287	CATTTGGACGAC	B	4	+	+	female	7	B6.SJL
14073	SRR1917300	TGGAGCACGTTG	B	4	+	+	female	8	B6.SJL
8928	SRR1917305	GCGACTTGTGTA	B	4	+	+	female	9	B6.SJL
9430	SRR1917677	AGATAGGACAGG	C	5	+	+	male	10	B6
17288	SRR1917681	GCGTGTTAAACC	C	5	+	+	male	11	B6
10432	SRR1917685	CTCCTACTGTCT	C	5	+	+	male	12	B6
6262	SRR1917733	TGAGGATGATAG	D	7	+	+	female	14	B6
4447	SRR1917739	ACCGGCTAGAGT	D	7	+	+	female	15	B6
4573	SRR1917753	TAAGGCCTATCG	D	8	+	+	female	16	B6
15470	SRR1917941	CACGAGGTCATT	D	8	+	+	female	17	B6
10595	SRR1917256	TAAAGTCACCCT	A	3	-	-	male	1	B6.SJL
10623	SRR1917259	TCGCATGAAGTC	A	3	-	-	male	2	B6.SJL
5499	SRR1917262	TTGCAGACAGGC	A	3	-	-	male	3	B6.SJL
16089	SRR1917265	GAGGCCATCAGT	A	3	-	-	male	4	B6.SJL
9771	SRR1917268	ATTCATGGACGC	B	4	-	-	female	5	B6.SJL
14156	SRR1917272	GTTACGCATTAC	B	4	-	-	female	6	B6.SJL
10691	SRR1917278	ATGTACGGCGAC	B	4	-	-	female	7	B6.SJL
10332	SRR1917298	TAATCGGATTCC	B	4	-	-	female	8	B6.SJL
9279	SRR1917303	AACCATCGGGTG	B	4	-	-	female	9	B6.SJL
11691	SRR1917674	GGTAGGAACAAT	C	5	-	-	male	10	B6
6458	SRR1917679	CAAGTTTACGGC	C	5	-	-	male	11	B6
6463	SRR1917683	CTGGAAATCTGC	C	5	-	-	male	12	B6
6983	SRR1917689	GACTGATCATCT	C	6	-	-	female	13	B6.SJL
8225	SRR1917729	AAGTGGACTCTC	D	7	-	-	female	14	B6
10357	SRR1917736	TACGTCCCGTTC	D	7	-	-	female	15	B6
7065	SRR1917750	AAGCCTACACGT	D	8	-	-	female	16	B6

6648	SRR1917756	CTGTAGGAGACC	D	8	-	-	female	17	B6
7510	SRR1917690	GTAAGTCGTGGC	C	6	-	+	female	13	B6.SJL
1911	SRR1917730	CACATGCCTAAG	D	7	-	+	female	14	B6
3439	SRR1917737	TGCTATATCTGG	D	7	-	+	female	15	B6
11612	SRR1917252	CGCAAATTCGAC	A	3	a	a	male	1	B6.SJL
13325	SRR1917253	TAACTCTGATGC	A	3	a	a	male	2	B6.SJL
12886	SRR1917254	TACGATGACCAC	A	3	a	a	male	3	B6.SJL
13014	SRR1917255	TTCGCCCTTCAG	A	3	a	a	male	4	B6.SJL
15836	SRR1917271	TGCGGCATCGAA	B	4	a	a	female	5	B6.SJL
14713	SRR1917277	ACACGAGCCACA	B	4	a	a	female	6	B6.SJL
21161	SRR1917297	CTCCTGAAAGTT	B	4	a	a	female	7	B6.SJL
19236	SRR1917301	TTATCACGTGCA	B	4	a	a	female	8	B6.SJL
14827	SRR1917306	TGGACACCGAAC	B	4	a	a	female	9	B6.SJL
10080	SRR1917678	TACCTCTCAGAA	C	5	a	a	male	10	B6
10664	SRR1917682	TTCCGTAGGGAT	C	5	a	a	male	11	B6
11753	SRR1917686	CCTAATGGAACC	C	5	a	a	male	12	B6
18258	SRR1917692	TTCTGGGAACAC	C	6	a	a	female	13	B6.SJL
18046	SRR1917734	AGAACACGTCTC	D	7	a	a	female	14	B6
16337	SRR1917741	GTGCCGGTGATA	D	7	a	a	female	15	B6
18779	SRR1917755	ATCTCTGGCATA	D	8	a	a	female	16	B6
17513	SRR1917942	TGATTTGGACCT	D	8	a	a	female	17	B6

Table S2: Details of Antibodies Used in This Study

Antibody	Clone	Vendor
IgM	Il/41	BD Biosciences
IgG3	R40-82	BD Biosciences
IgG2b	R12-3	BD Biosciences
IgG1	A85-1	BD Biosciences
IgD	11-26	eBioscience
IgA	11-44-2	Southern Biotech
IgG2c	goat anti-mouse; polyclonal	Jackson Immunoresearch
IgG(total)	goat anti-mouse; polyclonal	Jackson Immunoresearch
CD19	1D3	eBioscience
B220	RA3-6B2	eBioscience
CD44	IM7	eBioscience
CD69	H1.2F3	eBioscience
CD11b	M1/70	eBioscience
CD8a	53-6.7	eBioscience
CD4	GK1.5	eBioscience
GL-7	GL7	eBioscience
Ly6G	1A8	eBioscience
CD11c	N418	eBioscience
gdTCR	GL3	eBioscience
PD-1	J43	eBioscience
CXCR5	2G8	eBioscience
PNA		Vector Laboratories
SA-PECy7		eBioscience
Foxp3	FJK-16s	eBioscience
T-bet	4B10	eBioscience
RORgt	AJKJS-9	eBioscience

Supplemental References

Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res *25*, 3389 - 3402.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of Molecular Biology *215*, 403-410.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I.*, et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. Nat Meth *7*, 335-336.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. (2006). Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. Applied and Environmental Microbiology *72*, 5069-5072.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460-2461.

Edgar, R.C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Meth *10*, 996-998.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. Bioinformatics *27*, 2194-2200.

Gish, W., and States, D.J. (1993). Identification of protein coding regions by database similarity search. Nat Genet *3*, 266-272.

John, J.S. (2011). SeqPrep: merging overlapping paired-end reads.

Karlin, S., and Altschul, S.F. (1990). Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proceedings of the National Academy of Sciences *87*, 2264-2268.

Karlin, S., and Altschul, S.F. (1993). Applications and statistics for multiple high-scoring segments in molecular sequences. Proceedings of the National Academy of Sciences *90*, 5873-5877.

Lozupone, C., and Knight, R. (2005). UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. Applied and Environmental Microbiology *71*, 8228-8235.

Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J., and Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. ISME J *5*, 169-172.

Lozupone, C.A., Hamady, M., Kelley, S.T., and Knight, R. (2007). Quantitative and Qualitative β Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities. Applied and Environmental Microbiology *73*, 1576-1585.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., and Hugenholtz, P. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J *6*, 610-618.

McMurdie, P.J., and Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLoS Comput Biol *10*, e1003531.

Paulson, J.N., Stine, O.C., Bravo, H.C., and Pop, M. (2013). Differential abundance analysis for microbial marker-gene surveys. Nat Meth *10*, 1200-1202.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. Molecular Biology and Evolution *26*, 1641-1650.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE *5*, e9490.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research *35*, 7188-7196.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research *41*, D590-D596.

Roberts, A., and Pachter, L. (2013). Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Meth *10*, 71-73.

Salter, S., Cox, M., Turek, E., Calus, S., Cookson, W., Moffatt, M., Turner, P., Parkhill, J., Loman, N., and Walker, A. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology *12*, 87.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J.*, et al.* (2009). Introducing mothur: Open-Source, Platform-Independent, Community-

Supported Software for Describing and Comparing Microbial Communities. Applied and Environmental Microbiology *75*, 7537-7541.
Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics *27*, 863-864.
Werner, J.J., Koren, O., Hugenholtz, P., DeSantis, T.Z., Walters, W.A., Caporaso, J.G., Angenent, L.T., Knight, R., and Ley, R.E. (2012). Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. ISME J *6*, 94-103.
Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., and Glöckner, F.O. (2014). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Research *42*, D643-D648.