Table S4. Analysis of 16S rDNA Sequences, Related to Extended Experimental Procedures

Over 1.1 million raw pyrosequencing reads were obtained, filtered, and quality-trimmed as previously described (Dethlefsen and Relman, 2011) with mothur (Schloss et al., 2009). In brief, reads with more than one ambiguous character were removed, the proximal primer and barcode sequences were trimmed off, and reads shorter than 390 bp and longer than 520 bp were discarded. These steps resulted in a total of 808,810 reads that could be assigned unambiguously to a sample. The numbers of sequences are summarized in the table below.

Sample (numbers)	MMb inocula (2)	MMb mice (29)	HMb inocula (2)	HMb mice (28)
Number of reads (range)	26,764 (10.536,16.228)	328,417 (6.619- 7.813)	37,198 (9.833, 27.365)	288,793 (6.300-18.307)
Number of OTUs (range)	855 (488,569)	1168 (153-465)	574 (226,374)	610 (115-230)

Phylogenetic assignments were obtained by clustering reads with Uclust (http://www.drive5.com/) (Edgar, 2010) at a maximum distance of 5% with the Needleman–Wunsch algorithm based on a high-quality seed library. The seed library consisted of 47,768 high-quality sequences derived from a pre-clustered SILVA SSU rRNA reference database release 100 (http://www.arb-silva.de/). Omitted from further analysis were 127,176 reads that did not cluster with a seed within a genetic distance of 5%. A total of 681,634 reads clustered with a reference sequence, of which 462 reads were singletons and therefore were discarded. The remaining 681,172 reads were assigned to a total of 1,866 OTUs. There were 546,855 reads with ≥99% similarity to their respective seed (refOTU); these reads were assigned to 1,113 OTUs. The 112,333

reads that were 98.9–98% similar to a seed were assigned to 371 OTUs, the 14,862 reads that were 97.9–97% similar to a seed were assigned to 189 OTUs, the 6,579 reads that were 96.9–96% similar to a seed were assigned to 129 OTUs, and the 543 reads that were 95.9–95% similar to a seed were assigned to 64 OTUs. In this study, the term *operational taxonomic unit* (OTU) is used synonymously with the term *bacterial species*. As described, OTUs were defined on the basis of sequence clustering of V1-V2-V3 16S rDNA pyrosequence reads with a high-quality reference database. To evaluate differences between microbial communities, unweighted UniFrac distances—i.e., a phylogenetic beta diversity metric (Lozupone and Knight, 2005)—were calculated from normalized read numbers per sample between all pairs of samples. The relatedness of community membership in each sample was assessed through principal coordinates analysis with the QIIME software package (Quantitative Insights Into Microbial Ecology, http://qiime.sourceforge.net/) (Caporaso et al., 2010). The sequence data used for these analyses are available at http://sites.google.com/site/davidrelmanlab/databases/supplements.