Supplementary Materials:
Microbiome composition in both wild-type and disease model mice is heavily influenced by mouse facility

Kristopher D. Parker[1], Shannon E. Albeke[2], Jason P. Gigley[1], Allan M. Goldstein[3], Naomi L. Ward[1*]

***Correspondence:**
Dr. Naomi L. Ward
nlward@uwyo.edu

SUPPLEMENTARY METHODS

Mouse genotyping

Mouse genomic DNA was extracted from ear tissue with either Red Extract-N-Amp Tissue PCR kit (Sigma-Aldrich, St. Louis, MO, USA) or Viagen DirectPCR Ear Lysis Reagent (Viagen Biotech, Los Angeles, CA, USA), using manufacturer protocols. For PCR, the following primer set was used: KO-forward: oIMR7770F ATAGATTCGCCCTTGTGTCC, WT-forward oIMR6449F GATGAACCTGCTCAGTGCAA, common reverse: oIMR6448R CATGGTCTTGTTCCTGATGC. Amplifications were performed in 25μL reactions with 1μL of each 10μM primer, 0.5μL 10mM dNTPs, 2μL 25mM $MgCl_2$, 0.125μL DNA polymerase, and 4μL template. Reactions were performed under the following thermal profile: 94°C for 3 min, then 35 cycles of 94°C for 30 sec, 60°C for 1 min, 72°C for 1 min, followed by one cycle of 72°C for 2 min and a 4°C hold. Amplification products were visualized using standard agarose gel electrophoresis. $Ednrb^{-/-}$ (KO) mice exhibit a fragment of 244bp. $Ednrb^{+/+}$ (WT) mice exhibit a fragment of 444bp. $Ednrb^{+/-}$ (heterozygous) mice exhibit fragments of 444bp and 244bp.

DNA extraction

Colon and fecal samples were thawed and DNA extraction was performed using the QIAmp DNA stool MiniKit according to manufacturer's instructions with the following modifications: fecal samples were vortexed at max speed for approximately one minute or until thoroughly homogenized in existing buffer. To colon samples, 600μL of sterile PBS (pH 7.4) was added, followed by vortexing at max speed for one minute. For both sample types, 600μL of supernatant was added to a 2mL tube prefilled with sterile 0.1mm diameter zirconium beads (Spectrum Scientifics, Philadelphia, PA). To each tube, 1mL buffer ASL was added. Samples were processed on a TissueLyzer II (QIAGEN, Valencia, CA, USA) for 5 minutes at 30Hz. Supernatants were removed following centrifugation at 13,000 x g for 1 minute. One tablet of InhibitEX (provided in the kit) was added to each to tube to absorb PCR inhibitors, followed by 5 minutes at 15Hz on the TissueLyzer II. Hereafter, samples were processed according to manufacturer instructions. Samples were eluted in 50μL buffer AE (pre-warmed to 50°C) provided in the kit.

16S rRNA gene sequencing – Roche 454

Samples from the HSCR studies were sequenced at different times: 2012 for HSCR-Boston and 2015 for HSCR-Laramie. Amplification and sequencing was performed by RTL Genomics, Lubbock, TX, USA. Samples were amplified for pyrosequencing using a forward and reverse

fusion primer. The forward primer was constructed with the Roche A linker: CCATCTCATCCCTGCGTGTCTCCGACTCAG, an 8bp barcode, and the 28F primer: GAGTTTGATCNTGGCTCAG. The reverse fusion primer was constructed with a biotin molecule, the Roche B linker: CCTATCCCCTGTGTGCCTTGGCAGTCTCAG, and the 519R primer: GTNTTACNGCGGCKGCTG. This primer pair corresponds to the V1-V3 hypervariable region of the 16S rRNA gene. Amplifications were performed in 25μL reactions with Qiagen HotStarTaq master mix (QIAGEN, Valencia, CA, USA), 1μL of each 5μM primer, and 1μL of template. Reactions were performed on ABI Veriti thermocyclers (Applied Biosytems, Carlsbad, CA, USA) under the following thermal profile: 95°C for 5 min, then 35 cycles of 94°C for 30 sec, 54°C for 40 sec, 72°C for 1 min, followed by one cycle of 72°C for 10 min and 4°C hold. Amplification products were visualized with eGels (Life Technologies, Grand Island, NY, USA). Products were then pooled equimolar and each pool was cleaned and size selected using Agencourt AMPure XP (BeckmanCoulter, Indianapolis, IN, USA) following Roche 454 protocols (454 Life Sciences, Branford, CT, USA). Size-selected pools were then quantified, diluted, and used in emPCR reactions, which were performed and subsequently enriched. Sequencing followed established manufacturer protocols (454 Life Sciences).

16S rRNA gene sequencing – Illumina MiSeq

Samples from the C57BL/6J study were sequenced at the same time, in the same lane. Resamples from P24 Laramie mice were sequenced at the same time, in the same lane. Amplification and sequencing was performed by RTL Genomics, Lubbock, TX, USA. Samples were amplified for sequencing in a two-step process. The forward primer was constructed with the Illumina i5 sequencing primer: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG, and the 28F primer: GAGTTTGATCNTGGCTCAG. The reverse primer was constructed with the Illumina i7 sequencing primer: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG, and the 388R primer: TGCTGCCTCCCGTAGGAGT. This primer pair corresponds to the V1-V2 hypervariable region of 16S rRNA. Amplifications were performed in 25μL reactions with Qiagen HotStarTaq master mix (QIAGEN, Valencia, CA, USA), 1μL of each 5μM primer, and 1μL of template. Reactions were performed on ABI Veriti thermocyclers (Applied Biosytems, Carlsbad, CA, USA) under the following thermal profile: 95°C for 5 min, then 25 cycles of 94°C for 30 sec, 54°C for 40 sec, 72°C for 1 min, followed by one cycle of 72°C for 10 min and 4°C hold. Products from the first-stage amplification were added to a second PCR based upon qualitatively determined concentrations. Primers for the second PCR were designed based on the Illumina Nextera PCR primers as follows: forward: AATGATACGGCGACCACCGAGATCTACAC-i5index-TCGTCGGCAGCGTC and reverse: CAAGCAGAAGACGGCATACGAGAT-i7index-GTCTCGTGGGCTCGG. The second-stage amplification was run under the same conditions as the first except for only 10 cycles. Amplification products were visualized with eGels (Life Technologies, Grand Island, NY, USA). Products were then pooled equimolar and each pool was size selected in two rounds using Agencourt AMPure XP (BeckmanCoulter, Indianapolis, IN, USA) in a 0.7 ratio for both rounds. Size-selected pools were then quantified using the Quibit 2.0 fluorometer (Life Technologies, Grand Island, NY, USA) and loaded on an Illumina MiSeq (Illumina, Inc. San Diego, CA, USA) 2x300 flow cell at 10pM.

Quality trimming, chimera checking, and denoising of raw datasets

These pre-processing steps improve the overall accuracy of sequences thus reducing the chance of falsely classified OTUs in downstream analyses. RTL performed pre-processing as follows: for

Illumina only, the forward and reverse reads were taken in FASTQ format and merged using the PEAR Illumina paired-end read merger (Zhang et al. 2014). For both Illumina and 454, the FASTQ-formatted files were converted into FASTA-formatted sequence and quality files. Reads were run through an internally developed quality-trimming algorithm. During this stage, a running average for each read was taken across the sequence and was trimmed back at the last base where the total average was greater than 25. Sequence reads were sorted by length from longest to shortest. Prefix dereplication was performed using the USEARCH algorithm (Edgar 2010). Briefly, prefix dereplication grouped reads into clusters such that each sequence of equal or shorter length to the centroid sequence must be a 100% match to the centroid sequence for the length of the sequence. Each cluster was marked with the total number of member sequences. Sequences less than 100bp in length were not written to the output file; however, no minimum cluster size restriction was applied, allowing singleton clusters to exist in the output. Following dereplication, clustering at 4% divergence was performed using the USEARCH clustering algorithm. The result of this stage was the consensus sequence from each new cluster, tagged to show its total number of member sequences (dereplicated + clustered). Clusters containing fewer than two members were not added to the output file, thus removing singletons from the dataset. Selection of Operational Taxonomic Units was performed using the UPARSE OTU selection algorithm to classify the large number of clusters into OTUs (Edgar 2013). Following OTU selection, chimera checking was performed using the UCHIME chimera detection software executed in *de novo* mode (Edgar et al. 2011). Each clustered centroid from USEARCH clustering was then mapped to its corresponding OTUs and marked as either chimeric or non-chimeric. All chimeric sequences were then removed. Each quality-trimmed read was then mapped to its corresponding non-chimeric cluster using the USEARCH global alignment algorithm. Using the consensus sequence for each centroid as a guide, each sequence in a cluster was then aligned to the consensus sequence and each base was corrected using the following rules where C was the consensus sequence and S was the aligned sequence: (1) if the current base pair in S was marked to be deleted, then the base was removed from the sequence if the quality score for that base was less than 30, (2) if the current position in S was marked to have a base from C inserted, then the base was inserted into the sequence if the mean quality score from all sequences that mark the base as existing was greater than 30, (3) if the current position in S was marked as a match to C, but the bases were different, then the base in S was changed if the quality score for that base was less than 30, (4) if a base was inserted or changed, the quality score for that position was updated, (5) if the base was deleted, the quality score for that position was removed, (6) otherwise, the base in S was left alone and correction moved to the next position. Finally, all of the corrected sequences were written to the output file in FASTA format.

OTU table fates and rarefaction depths

The unrarefied master OTU table (master-HSCR) was used for facility- or genotype-based comparisons of relative abundance of taxa. Prior to rarefaction, master-HSCR was split into separate OTU tables as follows: colon samples from both facilities (cBL-HSCR), fecal samples from both facilities (fBL-HSCR), all Boston colon samples (cBos-HSCR), all Boston fecal samples (fBos-HSCR), all Laramie colon samples (cLar-HSCR), all Laramie fecal samples (fLar-HSCR). Filtered OTU tables were individually rarefied to the depths listed in Table S8. Rarefaction creates a subsampled OTU table by randomly selecting equal numbers of sequences from each sample. This subsampling reduces the variability in sequencing depth between sequencing runs, while also reducing the disparity of depth between sequencing platforms. When

rarefying OTU tables, we recognize that valuable data may be lost from samples with higher coverage, especially when merging sequences from 454 and Illumina platforms. However, to apply statistical approaches and thus make meaningful conclusions, we needed the added samples from the HSCR-Laramie Illumina resample. For these reasons, the 454 and Illumina sequence files from the HSCR-Laramie study were merged prior to OTU picking. Colon or fecal specific rarefied OTU tables (cBL-HSCR-rare and fBL-HSCR-rare) were used for inter-facility comparisons of alpha and beta, and core microbiome computation. We did not make inter-facility comparisons using OTU tables of different rarefaction depths. For intra-facility comparisons, individually rarefied OTU tables (cBos-HSCR-rare, fBos-HSCR-rare, cLar-HSCR-rare, and fLar-HSCR-rare) were used to assess genotype-based differences in alpha and beta diversity. For the C57BL/6J dataset, the unrarefied master OTU table (master-C57) was used for facility-based comparisons of relative abundance of taxa. The master-C57 OTU table was rarefied to a depth of 12,790 and used for all other inter-facility comparisons.

REFERENCES FOR SUPPLEMENTARY METHODS

Edgar, Robert C. 2010. "Search and Clustering Orders of Magnitude Faster than BLAST." *Bioinformatics (Oxford, England)* 26 (19): 2460–61. https://doi.org/10.1093/bioinformatics/btq461.

———. 2013. "UPARSE: Highly Accurate OTU Sequences from Microbial Amplicon Reads." *Nature Methods* 10 (10): 996–98. https://doi.org/10.1038/nmeth.2604.

Edgar, Robert C., Brian J. Haas, Jose C. Clemente, Christopher Quince, and Rob Knight. 2011. "UCHIME Improves Sensitivity and Speed of Chimera Detection." *Bioinformatics* 27 (16): 2194–2200. https://doi.org/10.1093/bioinformatics/btr381.

Zhang, Jiajie, Kassian Kobert, Tomáš Flouri, and Alexandros Stamatakis. 2014. "PEAR: A Fast and Accurate Illumina Paired-End ReAd MergeR." *Bioinformatics (Oxford, England)* 30 (5): 614–20. https://doi.org/10.1093/bioinformatics/btt593.

SUPPLEMENTARY FIGURES

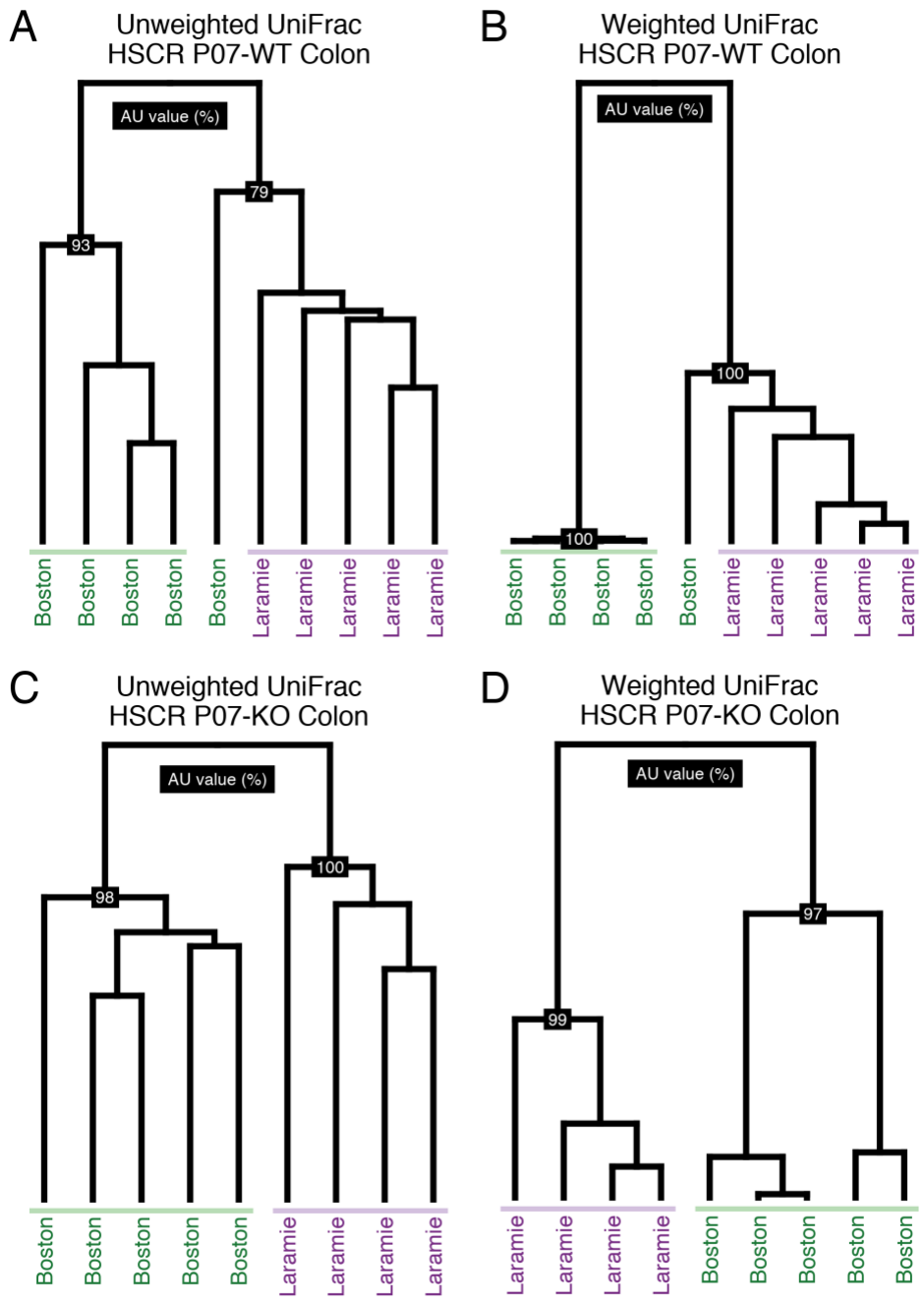

Figure S1 | Colon samples from P07 HSCR-mice cluster by facility of origin. Hierarchical clustering of unweighted (A, C) and weighted (B, D) UniFrac distances for sequences obtained from P07-WT and P07-KO colon samples. Shading emphasizes each facility's cluster.

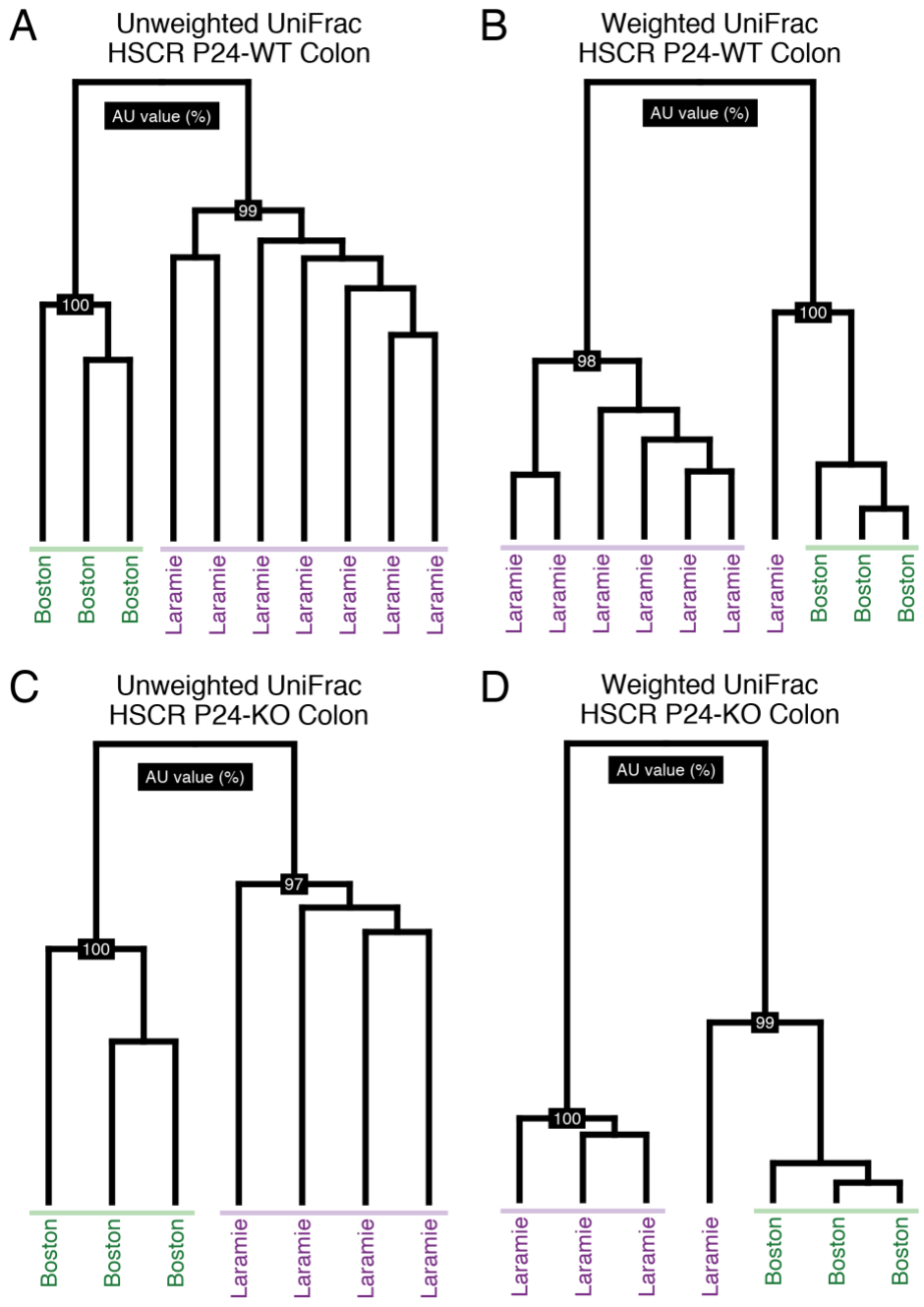

Figure S2 | Colon samples from P24 HSCR-mice cluster by facility of origin. Hierarchical clustering of unweighted (A, C) and weighted (B, D) UniFrac distances for sequences obtained from P24-WT and P24-KO colon samples. Shading emphasizes each facility's cluster.

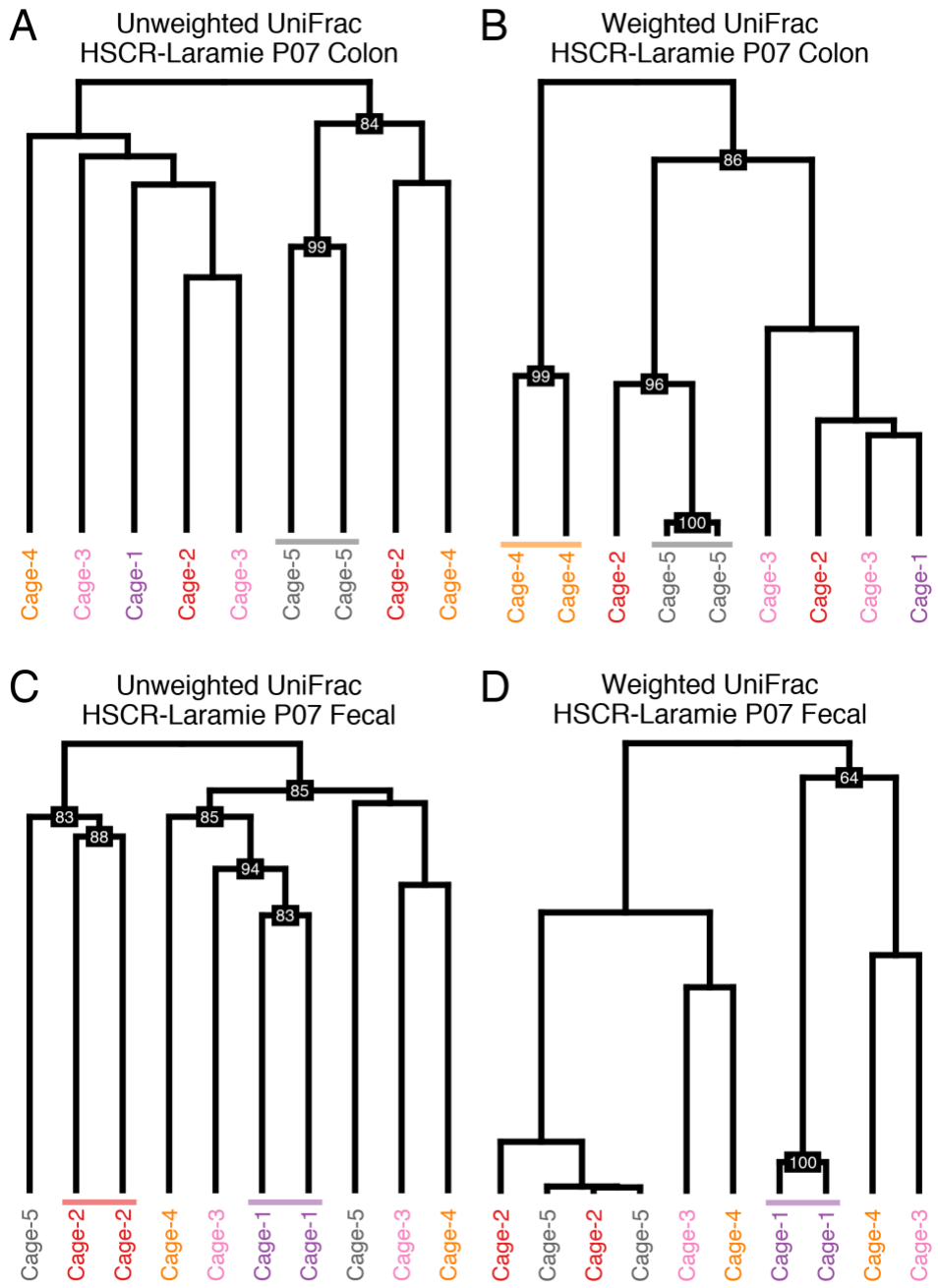

Figure S3 | Colon and fecal samples from Laramie P07 HSCR-mice exhibit some clustering by cage. Hierarchical clustering of unweighted (A, C) and weighted (B, D) UniFrac distances for sequences obtained from P07-WT and P07-KO colon and fecal samples from Laramie. Labels and colors denote separate cages. Shading represents instances where clusters contained mice from the same cage.

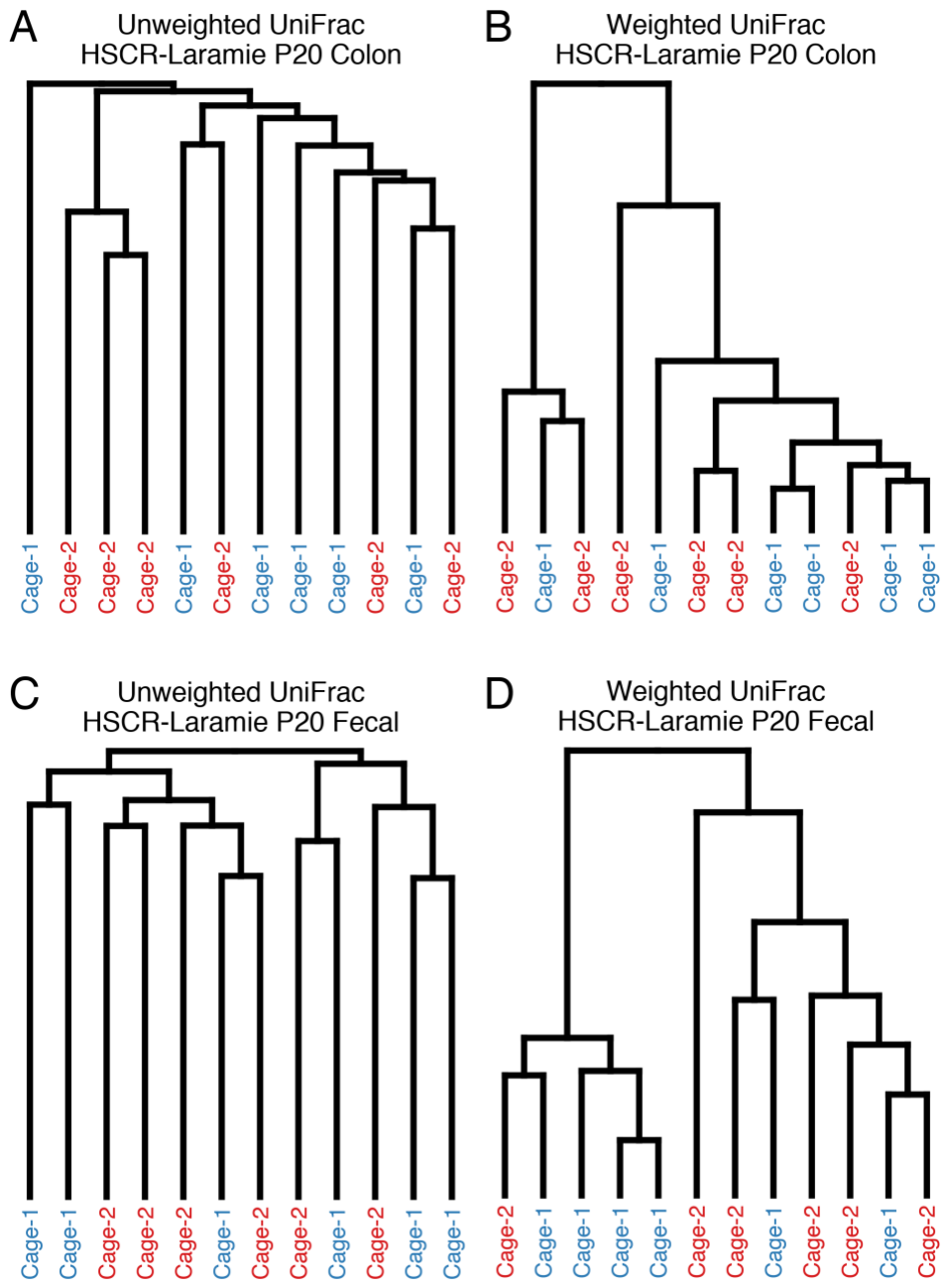

Figure S4 | Colon and fecal samples from Laramie P20 HSCR-mice exhibit no clear clustering by cage. Hierarchical clustering of unweighted (A, C) and weighted (B, D) UniFrac distances for sequences obtained from P20-WT and P20-KO colon and fecal samples from Laramie. Labels and colors denote separate cages.

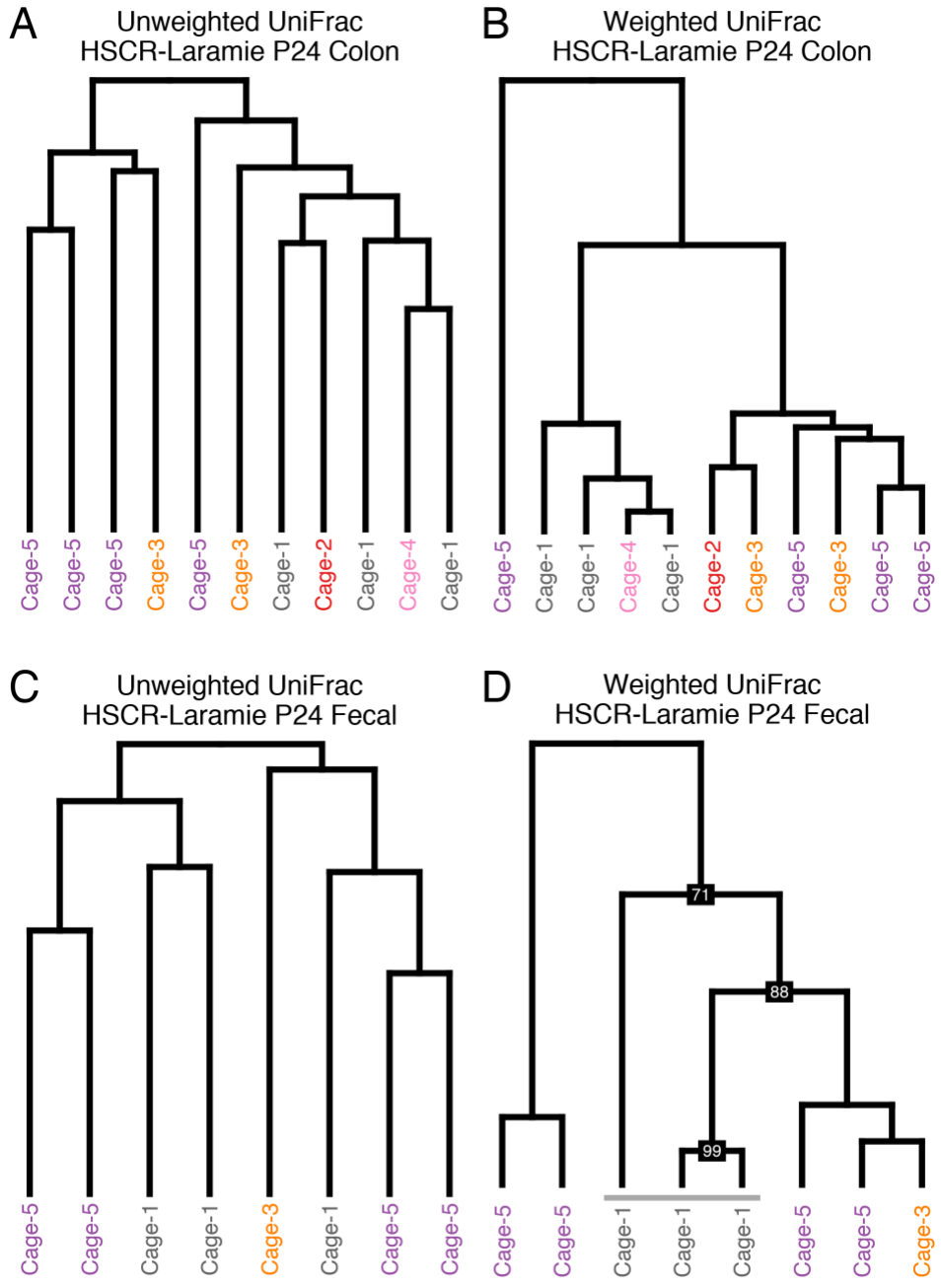

Figure S5 | Colon and fecal samples from Laramie P24 HSCR-mice exhibit very minimal clustering by cage. Hierarchical clustering of unweighted (A, C) and weighted (B, D) UniFrac distances for sequences obtained from P24-WT and P24-KO colon and fecal samples from Laramie. Labels and colors denote separate cages.

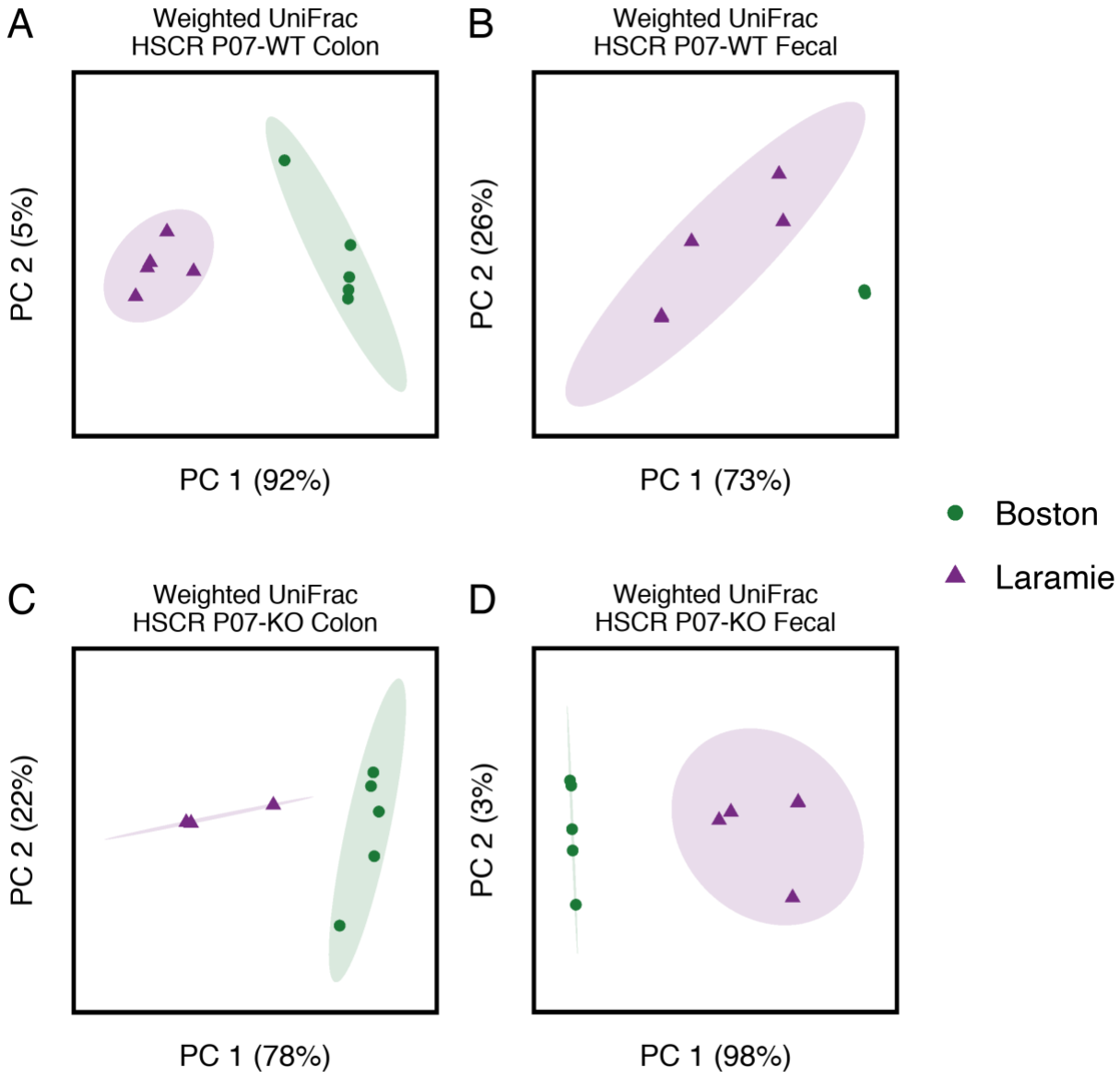

Figure S6 | Core microbiomes of P07 HSCR-mice separate by facility of origin, with no overlap. Principal Coordinates Analysis of weighted UniFrac distances for core microbiome OTUs at the 50% threshold obtained from P07-WT and P07-KO colon and fecal samples. Percentage values along each axis indicate the amount of variability in the data explained by each of the first two principal coordinates. Ellipses indicate 95% confidence intervals.

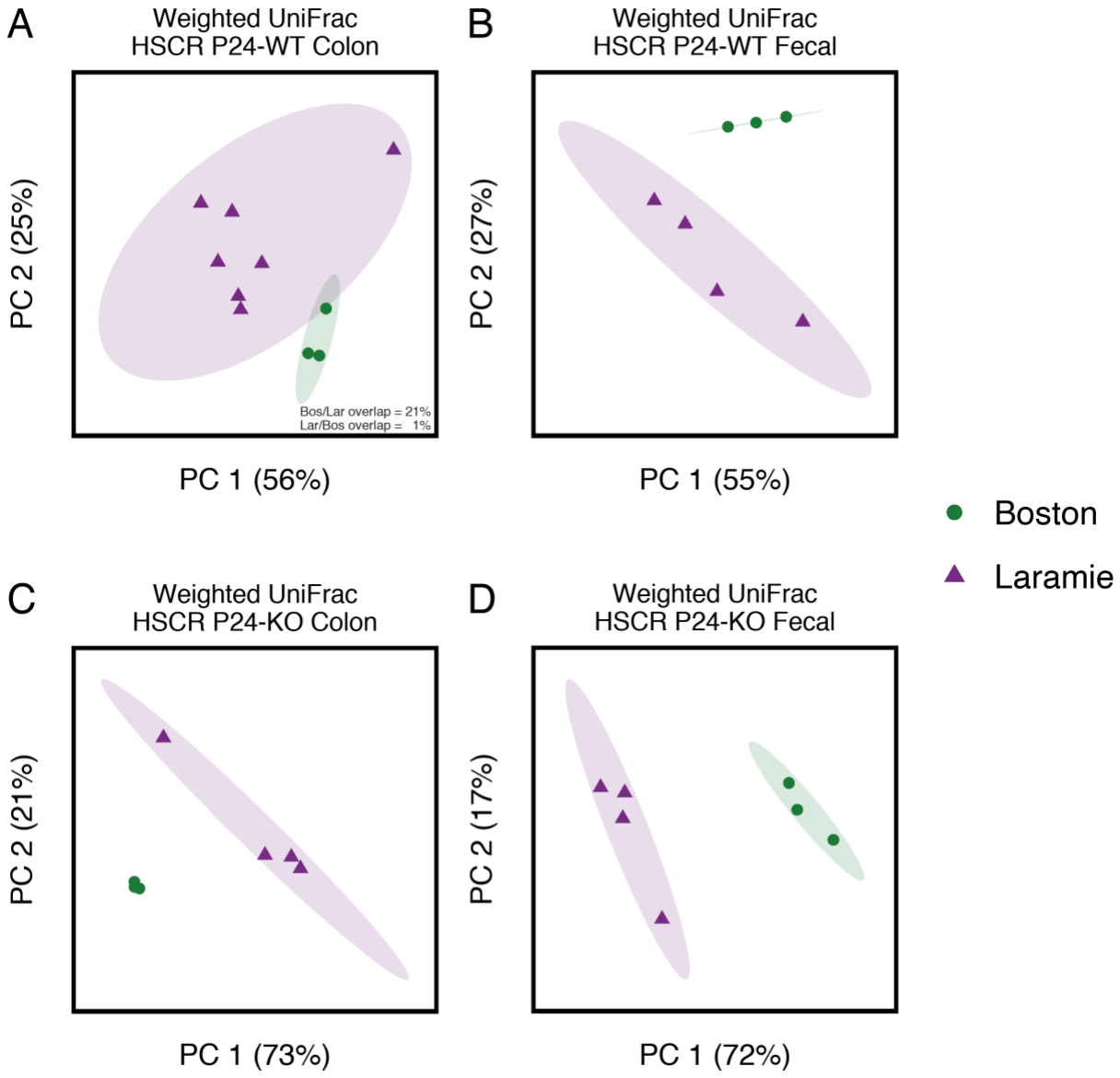

Figure S7 | Core microbiomes of P24 HSCR-mice separate by facility of origin, with very minimal overlap. Principal Coordinates Analysis of weighted UniFrac distances for core microbiome OTUs at the 50% threshold obtained from P24-WT and P24-KO colon and fecal samples. Percentage values along each axis indicate the amount of variability in the data explained by each of the first two principal coordinates. Ellipses indicate 95% confidence intervals.

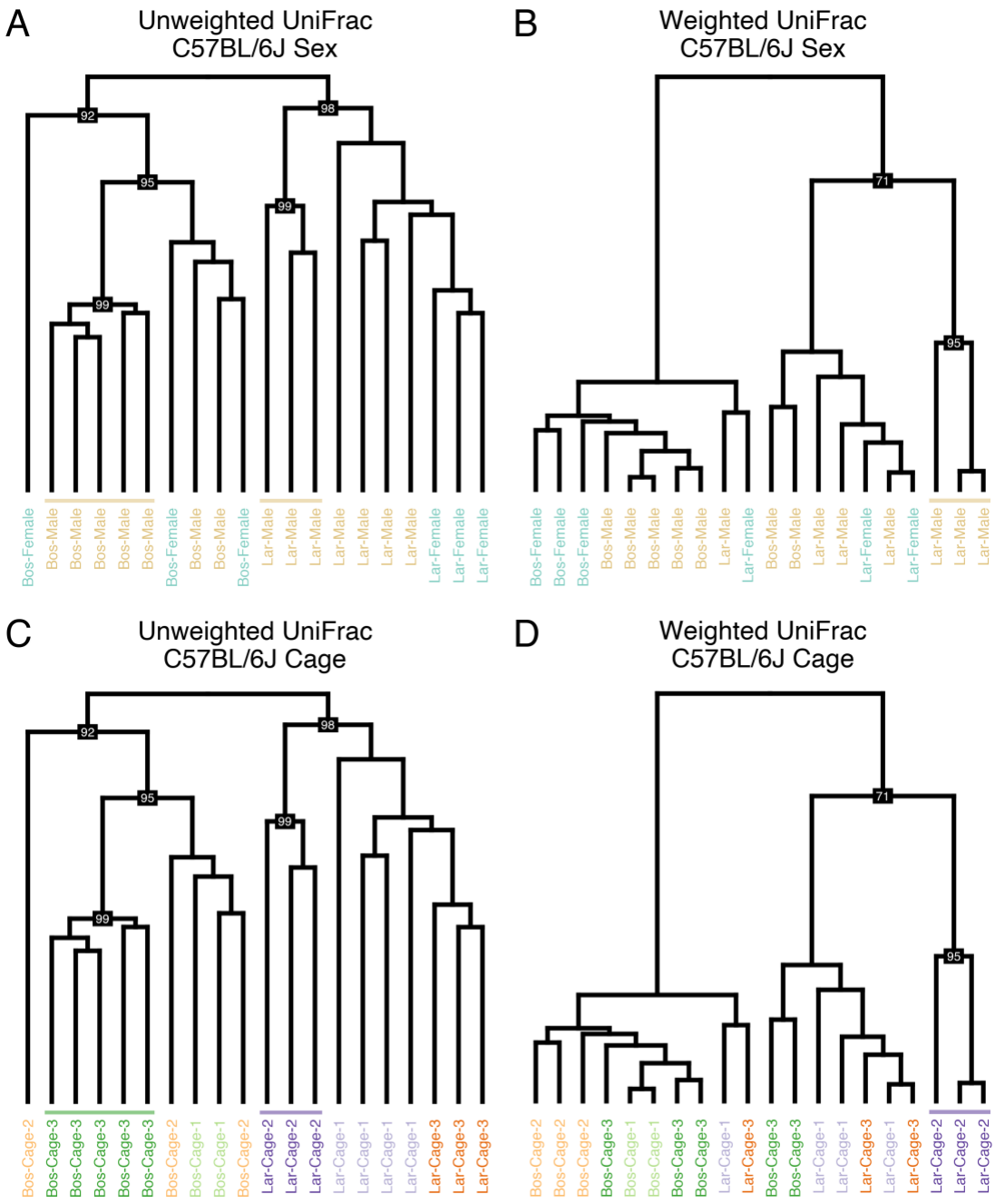

Figure S8 | Fecal samples from C57BL/6J mice show minimal clustering by sex or cage. (A-D) Hierarchical clustering of unweighted and weighted UniFrac distances for sequences obtained from C57BL/6J fecal samples. (A-B) Shading represents instances where individual clusters contained only mice of the same sex. (C-D) Each color denotes a separate cage. Shading represents instances where clusters contained only mice from the same cage.

SUPPLEMENTARY TABLES

Table S1 Inter-Facility PERMANOVA statistics

HSCR dataset *Boston vs. Laramie*	Total Beta Diversity								Core 50% Beta Diversity			
	ADONIS				ANOSIM				ADONIS		ANOSIM	
	Unweighted UniFrac		Weighted UniFrac		Unweighted UniFrac		Weighted UniFrac		Weighted UniFrac		Weighted UniFrac	
Colon Samples	R2	*P*	R2	*P*	R	*P*	R	*P*	R2	*P*	R	*P*
P07-WT vs. P07-WT	31.4%	0.008	58.5%	0.008	0.884	0.009	0.880	0.008	86.0%	0.008	1.000	0.008
P20-WT vs. P20-WT	38.6%	0.002	78.3%	0.002	1.000	0.002	0.992	0.002	91.4%	0.002	1.000	0.002
P24-WT vs. P24-WT	29.1%	0.009	30.3%	0.034	0.968	0.008	0.484	0.034	28.7%	0.042	0.464	0.034
P07-KO vs. P07-KO	29.8%	0.009	51.6%	0.008	0.962	0.007	0.831	0.008	70.0%	0.008	0.913	0.008
P20-KO vs. P20-KO	44.2%	0.002	66.9%	0.002	1.000	0.002	1.000	0.002	71.3%	0.002	1.000	0.002
P24-KO vs. P24-KO	37.2%	0.029	55.7%	***0.057***	1.000	0.028	0.815	***0.057***	56.0%	***0.057***	0.537	***0.058***
Fecal Samples												
P07-WT vs. P07-WT	50.3%	0.008	61.8%	0.008	1.000	0.008	0.940	0.008	62.3%	0.008	0.780	0.008
P20-WT vs. P20-WT	34.0%	0.002	82.3%	0.002	1.000	0.002	1.000	0.002	84.6%	0.002	1.000	0.002
P24-WT vs. P24-WT	36.6%	0.029	17.9%	***0.400***	0.926	0.029	0.130	***0.315***	30.9%	***0.114***	0.333	***0.086***
P07-KO vs. P07-KO	29.9%	0.008	59.6%	0.008	0.708	0.008	0.936	0.008	92.0%	0.008	1.000	0.008
P20-KO vs. P20-KO	38.4%	0.002	74.4%	0.002	1.000	0.002	1.000	0.002	57.8%	0.002	0.813	0.002
P24-KO vs. P24-KO	42.6%	0.029	54.6%	***0.057***	1.000	0.029	0.759	***0.058***	68.9%	0.029	1.000	0.029
C57BL/6J dataset												
Boston vs. Laramie	15.9%	0.00002	28.0%	0.001	0.583	0.00003	0.409	0.0004	30.9%	0.0012	0.402	0.0006

Bold-Italic text indicates *P*-values > 0.05

Table S2 Core OTUs

HSCR dataset Colon Samples		Total # Observed OTUs	Core 50% # of core OTUs	Core 50% % Observed OTUs = core	Core 75% # of core OTUs	Core 75% % Observed OTUs = core	Core 100% # of core OTUs	Core 100% % Observed OTUs = core
P07-WT	Boston Core	22	7	31%	4	18%	3	13%
	Laramie Core	105	39	37%	19	18%	12	11%
	Conserved Core	64	16	25%	0	0%	0	0%
P07-KO	Boston Core	27	12	45%	6	23%	5	19%
	Laramie Core	67	37	55%	20	30%	12	18%
	Conserved Core	47	8	17%	1	2%	0	0%
P20-WT	Boston Core	110	50	46%	25	23%	7	6%
	Laramie Core	102	53	52%	18	18%	11	11%
	Conserved Core	106	19	18%	0	0%	0	0%
P20-KO	Boston Core	85	48	56%	31	36%	15	18%
	Laramie Core	95	53	56%	22	23%	17	18%
	Conserved Core	90	32	36%	3	3%	1	1%
P24-WT	Boston Core	150	82	55%	34	23%	34	23%
	Laramie Core	122	43	35%	13	11%	4	3%
	Conserved Core	136	42	31%	2	1%	1	1%
P24-KO	Boston Core	137	81	59%	30	22%	30	22%
	Laramie Core	91	57	62%	25	27%	8	9%
	Conserved Core	114	23	20%	2	2%	1	1%
C57BL/6J dataset								
Fecal Samples	Boston Core	513	357	70%	164	32%	76	15%
	Laramie Core	534	335	63%	142	27%	65	12%
	Conserved Core	524	275	53%	129	25%	39	7%

Conserved Core is calculated as the average of Boston and Laramie Cores. % Observed OTUs = core is calculated as the # of core OTUs divided by the Total # of Observed OTUs.

Table S3 Alpha Diversity statistics

HSCR Dataset *Boston vs. Laramie*	Chao1		Observed OTUs	
	Kruskal-Wallis	Pairwise Wilcoxon	Kruskal-Wallis	Pairwise Wilcoxon
Colon Samples	*P*	*P*	*P*	*P*
P07-WT vs. P07-WT	0.118	0.067	0.029**	0.067
P20-WT vs. P20-WT	0.573	0.254	0.754	0.881
P24-WT vs. P24-WT	0.421	0.323	0.453	0.184
P07-KO vs. P07-KO	0.308	0.255	0.458	0.091
P20-KO vs. P20-KO	0.198	0.087	0.757	0.537
P24-KO vs. P24-KO	0.110	0.107	0.206	0.486
Fecal Samples				
P07-WT vs. P07-WT	0.107	0.089	0.058	0.080
P20-WT vs. P20-WT	0.727	0.522	0.810	0.701
P24-WT vs. P24-WT	0.852	1.000	0.658	0.961
P07-KO vs. P07-KO	0.484	0.132	0.544	0.118
P20-KO vs. P20-KO	0.445	0.234	0.816	0.438
P24-KO vs. P24-KO	0.121	0.118	0.308	0.127
Boston vs. Boston Colon Samples				
P07-WT vs. P07-KO	0.385	0.118	0.573	0.154
P20-WT vs. P20-KO	0.385	0.061	0.401	0.046**
P24-WT vs. P24-KO	1.000	1.000	0.669	0.190
Fecal Samples				
P07-WT vs. P07-KO	0.525	0.055	0.394	0.036**
P20-WT vs. P20-KO	0.200	0.055	0.418	0.462
P24-WT vs. P24-KO	0.759	0.663	0.394	0.104
Laramie vs. Laramie Colon Samples				
P07-WT vs. P07-KO	0.752	0.905	0.369	0.486
P20-WT vs. P20-KO	0.752	0.905	0.678	0.588
P24-WT vs. P24-KO	0.752	0.905	0.369	0.504
Fecal Samples				
P07-WT vs. P07-KO	0.697	0.784	0.497	0.638
P20-WT vs. P20-KO	0.850	0.784	0.597	0.638
P24-WT vs. P24-KO	0.215	0.586	0.497	0.971
C57BL/6J dataset				
Boston vs. Laramie	0.174	0.186	0.762	0.791

** indicates *P*-values < 0.05

Table S4 Inter-Facility taxonomy statistics

HSCR Dataset Colon Samples						
Association		Taxonomic Classification		Mean Relative Abundance	Kruskal-Wallis	Pairwise Wilcoxon
Facility	Group	Rank	Taxon		*P*	*P*
Laramie	P07-WT	Phylum	Actinobacteria	19%	0.011	NS
			Proteobacteria	46%	0.019	NS
		Family	Enterobacteriaceae	35%	0.006	NS
		Genus	*Propionibacterium*	11%	0.002	0.038
Boston	P07-WT	Phylum	Firmicutes	89%	0.013	NS
Laramie	P20-WT	Order	Bacteroidales	12%	0.003	0.023
		Genus	*Bacteroides*	25%	0.005	0.023
			Parabacteroides	9%	0.006	0.023
Boston	P20-WT	Phylum	Firmicutes	73%	0.048	NS
		Family	Ruminococcaceae	6%	0.048	NS
		Genus	*Oscillospira*	8%	0.011	0.046
Laramie	P24-WT	Genus	*Bacteroides*	20%	0.025	0.041
Laramie	P07-KO	Phylum	Actinobacteria	7%	0.020	NS
			Proteobacteria	58%	0.016	NS
		Family	Enterobacteriaceae	54%	0.005	NS
Boston	P07-KO	Phylum	Firmicutes	99%	0.009	NS
Laramie	P20-KO	Phylum	Bacteroidetes	77%	NS	0.045
		Order	Bacteroidales	14%	0.003	0.023
			Bacteroidales.1	8%	0.003	0.023
		Genus	*Bacteroides*	34%	0.005	0.023
			Parabacteroides	7%	0.006	0.023
Boston	P20-KO	Family	S24-7	42%	0.042	NS
Laramie	P24-KO	Genus	*Bacteroides*	37%	0.009	NS
			Parabacteroides	12%	0.016	NS
C57BL6/J Dataset						
Laramie		Phylum	Firmicutes	39%	0.010	0.011
Boston		Phylum	Bacteroidetes	76%	0.028	0.031

Only those taxa with mean relative abundances above 6% and that exhibited significant differences between age- and genotype-matched Boston and Laramie mice are shown.

Table S5 Intra-Facility PERMANOVA statistics

HSCR dataset *Boston vs. Boston*	Total Beta Diversity							
	ADONIS				ANOSIM			
	Unweighted UniFrac		Weighted UniFrac		Unweighted UniFrac		Weighted UniFrac	
Colon Samples	R2	*P*	R2	*P*	R	*P*	R	*P*
P07-WT vs. P07-KO	17.0%	0.064	34.4%	0.071	0.356	0.024**	0.284	0.087
P20-WT vs. P20-KO	31.0%	0.008**	67.5%	0.007**	0.604	0.008**	1.000	0.008**
P24-WT vs. P24-KO	29.2%	0.100	57.6%	0.100	0.889	0.099	1.000	0.100
Fecal Samples								
P07-WT vs. P07-KO	42.6%	0.008**	18.1%	0.133	1	0.007**	0.244	0.016**
P20-WT vs. P20-KO	26.2%	0.008**	79.6%	0.008**	0.588	0.007**	1	0.008**
P24-WT vs. P24-KO	29.8%	0.100	33.3%	0.200	0.889	0.1006	0.037**	0.499
Laramie vs. Colon Samples								
P07-WT vs. P07-KO	14.9%	0.142	19.1%	0.166	0.156	0.128	0.225	0.128
P20-WT vs. P20-KO	9.3%	0.354	5.9%	0.620	0.002	0.458	0.070	0.731
P24-WT vs. P24-KO	13.6%	0.046**	14.3%	0.165	0.286	0.050**	0.085	0.305
Fecal Samples								
P07-WT vs. P07-KO	13.4%	0.118	12.0%	0.388	0.104	0.2289	0.016	0.42
P20-WT vs. P20-KO	9.3%	0.413	8.0%	0.499	0.004	0.4431	0.011	0.362
P24-WT vs. P24-KO	24.1%	0.029**	25.7%	0.171	0.615	0.028**	0.271	0.142

** indicates *P*-values < 0.05

Table S6 Intra-Facility taxonomy statistics

HSCR Dataset Colon Samples Association		Taxonomic Classification		Mean Relative Abundance	Kruskal-Wallis	Pairwise Wilcoxon
Facility	Group	Rank	Taxon		*P*	*P*
Boston	P20-KO	Family	Enterobacteriaceae	7%	0.026	NS
Fecal Samples						
	P07-WT	Phylum	Firmicutes	99.9%	NS	0.037
Boston	P20-KO	Phylum	Bacteroidetes	60%	NS	0.040
	P20-KO	Family	S24-7	60%	NS	0.040
	P20-WT	Family	S24-7	20%	NS	0.042
	P20-KO	Family	Enterobacteriaceae	6%	NS	0.038
	P24-WT	Family	S24-7	26%	NS	0.042
Laramie	P24-KO	Phylum	Proteobacteria	11%	NS	0.031
	P24-KO	Family	Enterobacteriaceae	11%	NS	0.050
	P24-KO	Genus	*Bacteroides*	39%	NS	0.046
	P24-KO	Genus	*Parabacteroides*	22%	0.034	NS

Only those taxa with mean relative abundances above 6% and that exhibited significant differences between age- and genotype-matched Boston and Laramie mice are shown.

Table S7 Differentially abundant non-candidate OTUs and taxa conserved between facilities

OTU-ID	Taxonomic Identity	Statistical Test	Sample Type	Association	Facility	Log_2 FC
549756^	G: *Lactobacillus*	G-Test	Fecal	P07-WT	Boston	-0.29
				P07-KO	Laramie	+2.4
n/a	G: *Streptococcus*	Kruskal-Wallis	Fecal	P07-KO	Boston	+10
				P07-WT	Laramie	-2.5
n/a	P: Actinobacteria	Kruskal-Wallis	Colon	P20-KO	Boston	+10
				P20-WT	Laramie	-6.7
		Nonparametric T		P20-KO	Boston	+10
				P20-WT	Laramie	-6.7
n/a	F: S24-7^	Kruskal-Wallis	Fecal	P20-KO	Boston	+1.4
				P20-WT	Laramie	-0.94
		Nonparametric T		P20-KO	Boston	+1.4
				P20-WT	Laramie	-0.94
n/a	F: Ruminococcaceae	Kruskal-Wallis	Fecal	P24-KO	Boston	+2.7
				P24-WT	Laramie	-4.8

^ Indicates OTUs or taxa observed at multiple ages. P: represents Phylum. F: represents Family. G: represents Genus. Log_2 Fold Change (FC) was calculated KO/WT, therefore (+) indicates association with KO mice and (-) indicates association with WT mice.

Table S8 Rarefaction depths

OTU Table Name	Depth
master_HSCR	374
cBL_HSCR	374
fBL_HSCR	396
cBos_HSCR	695
fBos_HSCR	1791
cLar_HSCR	374
fLar_HSCR	396
master_C57	12790

DATA SHEET 2 FILE KEY

-QIIME_Code_1 (.html) Command-line code for demultiplexing, quality filtering, and OTU picking using QIIME 1.
-QIIME_Code_2 (.html) Command-line code for OTU table filtering, rarefaction, and all analyses using QIIME 1.
-Pre_R_Code.html (.html) Command-line code for moving QIIME 1 outputs prior to using R.
-HSCR_metadata (.txt) Tab-delimited text file of metadata containing information about all samples from the HSCR studies.
-HSCR_OTU_table (.biom) BIOM formatted file of unrarefied OTU counts of observations and assigned taxonomies for all samples from the HSCR studies.
-HSCR_rep_set (.tre) FastTree formatted file containing phylogenetic tree of representative sequences for all samples from the HSCR studies.
-HSCR_diffabund_map (.txt) Tab-delimited text file of metadata formatted for differential abundance testing (using QIIME 1) of samples from the HSCR studies.
-C57_metadata (.txt) Tab-delimited text file of metadata containing information about all samples from the C57BL/6J study.
-C57_OTU_table (.biom) BIOM formatted file of unrarefied OTU counts of observations and assigned taxonomies for all samples from the C57BL/6J study.
-C57_rep_set (.tre) FastTree formatted file containing phylogenetic tree of representative sequences for all samples from the C57BL/6J study.
-hscr_R_map (.txt) Tab-delimited text file (R map) for the HSCR studies, needed when working in R.
-hscr_R_format_file (.txt) Tab-delimited text file (R format file) for the HSCR studies, needed when working in R.
-c57_R_map (.txt) Tab-delimited text file (R map) for the C57BL/6J study, needed when working in R.
-c57_R_format_file (.txt) Tab-delimited text file (R format file) for the C57BL/6J study, needed when working in R.
-R_Code_1 (.R) R code for dendrogram construction and visualization to generate Figure 1, Supplementary Figures 1-5, and for principal coordinates analysis and visualization to generate Figure 2, and Supplementary Figures 6 and 7.
-R_Code_1_WS (.Rdata) The final R workspace generated for Figure 1-2, and Supplementary Figures 1-7.
-R_Code_2 (.R) R code for generating Figure 3.
-R_Code_2_WS (.Rdata) The final R workspace generated for Figure 3.
-R_Code_3 (.R) R code for generating Figure 4 and Supplementary Figure 8.
-R_Code_3_WS (.Rdata) The final R workspace generated for Figure 4 and Supplementary Figure 8.
-R_Code_4 (.R) R code for generating values for Table 1, and Supplementary Tables 2, 3, 4, 6, and 7.