Supporting Information

Benson et al. 10.1073/pnas.1007028107

SI Methods

Pyrosequencing. The V1-V2 region of the 16S rRNA gene was amplified using bar-coded fusion primers with the Roche-454 A or B titanium sequencing adapters (in italics), followed by a unique 8-base barcode sequence (B) and finally the 5′ ends of primer A-8FM (5′-*CCATCTCATCCCTGCGTGTCTCCGACTCAG*BBBBBBBBAGAGTTTGATCMTGGCTCAG) and of primer B-357R (5′-*CCTATCCCCTGTGTGCCTT-GGCAGTCTCAG*BBBBBBBBCTGCTGCCTYCCGTA-3′). All PCR reactions were quality- controlled for amplicon saturation by gel electrophoresis; band intensity was quantified against standards using GeneTools software (Syngene). For each region of a two-region picotiter plate, amplicons from 48 reactions were pooled in equal amounts and gel-purified. The resulting products were quantified using PicoGreen (Invitrogen) and a Qubit fluorometer (Invitrogen) before sequencing using Roche-454 GS FLX titanium chemistry.

Data Processing Pipeline. The raw data from the 454 pyrosequencing machine were first processed through a quality filter that removed unqualified sequence reads that did not meet the following criteria:

1. A complete forward primer and barcode
2. ≤ 2 "N" in a sequence read, where N is equivalent to an interrupted and resumed signals from sequential flows
3. 200 nt $\leq$ sequence length $\leq$ 500 nt
4. Average quality score ≥ 20.

After filtering, each read was trimmed to remove 3′ adapter and primer sequences and was parsed by barcode. The corresponding .QUAL file also was updated to remove quality scores from reads not passing quality filters. The files are associated with sample information in a hierarchical manner in MySQL tables. The processed data and the MySQL database tables are stored on a database server and available to the public at http://cage.unl.edu.

Given the massive size of the pyrosequencing data set and the need to normalize the taxonomy across the entire data set in a hierarchical fashion, a limited number of current algorithms could be modified and implemented. The CLASSIFIER algorithm assigns taxonomic status to each sequence read based on a covariance model developed from a training set (1). This algorithm is capable of processing very large data sets and was recently shown to provide adequate taxonomic assignments to pyrosequencing data (2). We implemented a parallelized version of the CLASSIFIER (kindly provided by the Center for Microbial Ecology, Ribosomal Database Project at Michigan State University), using the standard threshold of 0.8, with reads classified down to the lowest level until the score <0.8, at which point reads are classified as "unclassified" at the next-higher taxonomic rank.

The hierarchical output data from the from CLASSIFIER were further processed by computing the absolute proportion of each sequence, calculated as

$$absolute\ proportion = \frac{\#reads\ of\ a\ taxon}{total\ number\ of\ reads\ in\ a\ sample}$$

The absolute proportion is referred to as the Prop value. The multi-CLASSIFIER algorithm, proportion calculation, and assembly of the Prop table for the entire data set were performed sequentially on a Linux cluster of computer nodes, with the jobs controlled by the PBS portable batch system. The data were partitioned into a number of smaller groups, and the calculations were computed independently in a cluster node for each group, with the final results compiled when all were complete. At a threshold of 0.8, the data from all 645 animals in our data set included 420 different genera, 143 families, 53 orders, 24 classes, and 16 phyla that contained at least one assigned sequence. Of the 420 observed genera, 47 genera accounted for >99% of the sequences, and 19 accounted for >90% of the sequences.

To test the robustness of the CLASSIFIER algorithm, we compared the CLASSIFIER-based taxonomic assignments to the RDP database using SEQMATCH. Samples of 40,000 sequences assigned to one of several representative taxa were chosen and compared with the RDP database using the SEQMATCH program. Results for the top hits were compiled and are reported in Table S1.

Details of the QTL Analysis. QTL analyses generated *P* values for the original population and the GRAIP-permuted populations (n = 50,000); these were performed on log-transformed traits using the multiple-imputation method (3) within R/qtl (4). Statistical models included parent-of-origin type [i.e., whether a G_4 individual was descended from a progenitor (F_0) cross HR♀ X B6♂ or B6♀ X HR♂, coded as 1 or 0, respectively] and parity (i.e., order of litters from individual F_3 dams). The X chromosome was treated as an autosome, because R/qtl assumes a F_2 population and requires the identity of the cross direction. The output from R/qtl was then used to calculate locus-specific *P* values as described previously (5). Locus-specific *P* values were calculated for each marker of the original data set, using the value of that specific marker in each of the permuted maps at each locus as a null distribution. The null distribution for each marker was compared with the value for the original G_4 mapping data set to generate locus-specific *P* values at marker positions. These *P* values were interpolated onto the genome based on known physical positions of markers and placed on a scaffold at regular physical intervals. Finally, genome-wide, adjusted *P* values were computed by creating an ordered list of the minimum possible *P* values (or highest −log *P*, LOD) from each GRAIP-permuted map. Because we used 50,000 permutations, the minimum possible *P* value was 0.00002 (1/50,000) and the maximum −log *P* was 4.7. The 95th percentile ($P = 0.05$; LOD $\geq$ 3.9) and 90th percentile ($P = 0.1$; LOD ≥ 3.5) defined significant and suggestive loci, respectively. Confidence intervals were approximated by 1 LOD drop support intervals (relative to the GRAIP-permuted LOD score). Standard linear regression was used to estimate the percent variation by fitting the imputed QTL marker genotypes; the additive and dominance QTL effects were calculated using R/qtl.

1. Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73:5261–5267.
2. Liu Z, DeSantis TZ, Andersen GL, Knight R (2008) Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res* 36:e120.
3. Sen S, Churchill GA (2001) A statistical framework for quantitative trait mapping. *Genetics* 159:371–387.
4. Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19:889–890.
5. Peirce JL, et al. (2008) Genome Reshuffling for Advanced Intercross Permutation (GRAIP): Simulation and permutation for advanced intercross population analysis. *PLoS One* 3:e1977.
6. Walter J, et al. (2001) Detection of *Lactobacillus*, *Pediococcus*, *Leuconostoc*, and *Weissella* species in human feces by using group-specific PCR primers and denaturing gradient gel electrophoresis. *Appl Environ Microbiol* 67:2578–2585.

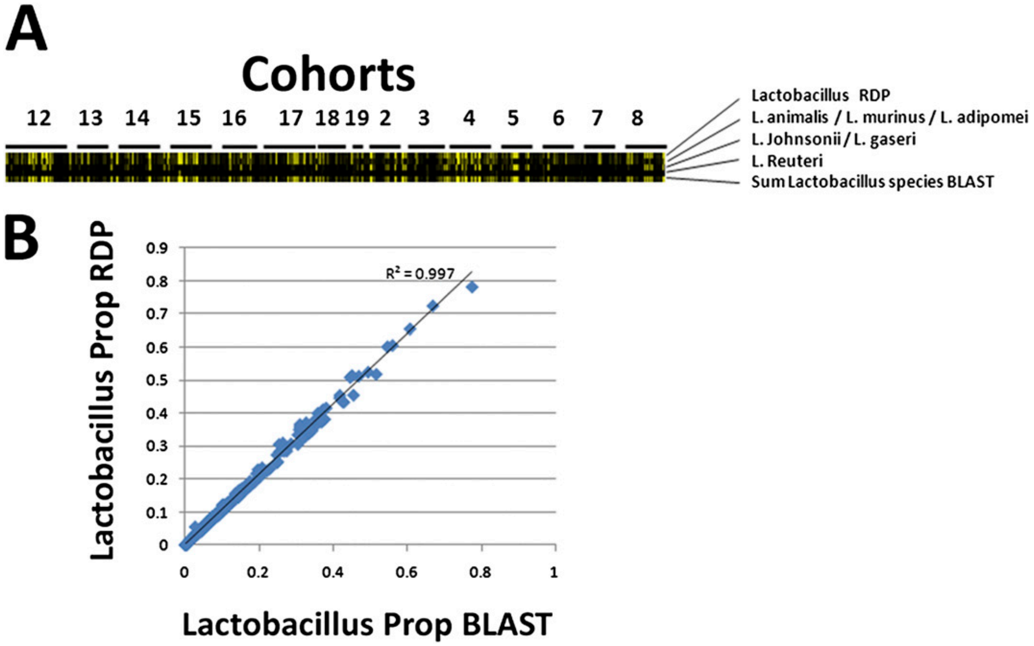

Fig. S1. BLAST Analysis of *Lactobacillus* species. To analyze the *Lactobacillus* at the species level, Bioedit v7.0.9 was used to perform a local nucleotide BLAST (blastn) search using the murine *Lactobacillus* type strain sequences: *L. animalis (*AB326350.1), *L. apodemi* (AJ871178.1), *L. murinus* (AB326349.1), *L. reuteri* (CP000705.1), *L. gasseri* (CP000413.1), and *L. johnsonii* (ACGR01000047.1). These sequences were trimmed to ~340 nucleotides to match the length of the V1-V2 amplicons and used as queries against entire sets of read sequences from each sample with a 97% identity threshold for species assignment. The number of each *Lactobacillus* species hits for each sample was then divided by the total number of reads and used as the Prop value for the sample. (*A*) A heat map depicting the relative abundance of BLAST hit distribution for the species groups of *L. animalis/murinus/, L. johnsonii/gasseri*, and *L. reuteri*. The top row depicts the relative abundance of the genus *Lactobacillus* from the CLASSIFIER algorithm, and the bottom row shows the pooled cumulative Prop for BLAST hits of all three species groups. (*B*) A scatterplot of relative Prop of *Lactobacillus* from the RDP CLASSIFIER versus the cumulative Prop of the *Lactobacillus* species groups from the BLAST analysis.

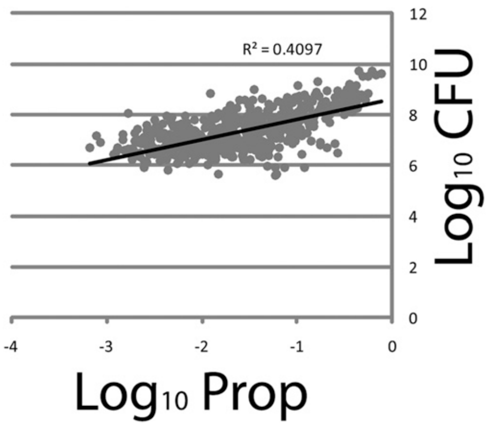

Fig. S2. Correlation between pyrosequencing and qPCR estimates for *Lactobacillus* in the G_4 AIL population. To quantify organisms in the Lactobacilli group, real-time qPCR was performed using a Mastercycler ep *realplex* (Eppendorf) and the group-specific primers Lac1 and Lac2 described previously (6). The primers target the 16S rDNA of *Lactobacillus* spp., *Pediococcus* spp., *Leuconostoc* spp., and *Weissella* spp., and result in a product length of 340 bp. The reaction mixture (25 μL) consisted of 1× QuantiFast SYBR Green PCR Master Mix (Qiagen), 25 pmol of each primer, template DNA, and RNase-free water. The amplification program was an initial denaturation at 94 °C for 2 min, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 61 °C for 1 min, and extension at 68 °C for 1 min. A melting curve analysis was performed after each run. Standard curves were generated from 10-fold serial dilutions of DNA extracted from pure cultures of *L. reuteri* (DSM 20016^T) and *L. gasseri* (ATCC 33323^T). A plot of the threshold cycles (C_t) vs. bacterial counts (CFU/mL) resulted in a linear relationship with a correlation coefficient (r) of −0.989 ($R^2 = 0.98$). The total number of bacteria (CFU/g) for each stool sample was determined by interpolation of the standard curve. Both standards and samples were run in duplicate, and the counts were averaged. To measure the linear relationship between pyrosequencing and qPCR, a correlation analysis was performed on the amount of bacteria quantified by each method. Specifically, the bacterial counts (in $\log_{10}$ CFU/mL) obtained by qPCR was plotted against the $\log_{10}$ proportion of *Lactobacillus, Leuconostoc, Pediococcus*, and *Weissella* reads over the total reads for each sample. A significant correlation ($P < 0.0001$) was obtained, with $r = 0.625$.

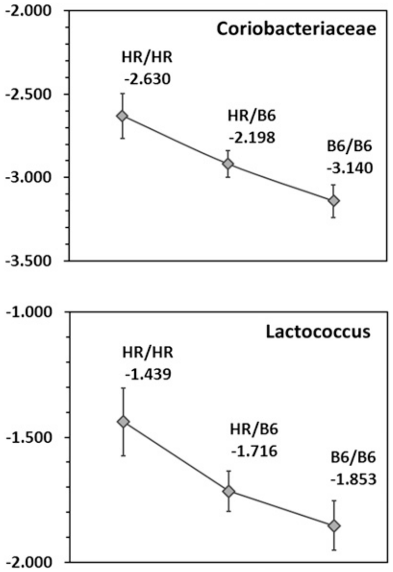

Fig. S3. Association of alleles at the JAX0030095 marker on MMU10 with the relative abundance of Coriobacteriaceae and *Lactococcus*. The log10-transformed Prop values for the family Coriobacteriaceae and the genus *Lactococcus* were averaged for each combination of JAX0030095 alleles. Alleles and the average log10-transformed Prop values are indicated above the relevant data points. Error bars indicate 2 SDs.

Table S1. SEQMATCH best hits of selected taxonomic bins from CLASSIFIER output

Taxonomic rank	Taxa[*]	Top organisms[†]	Taxa represented[‡]	Counts[§]	Prop total[¶]	Prop top hits[**]	Average S_ab score[††]
Genus	*Variovorax*	Variovorax paradoxus; Iso1; AY127900	Variovorax	15,303	0.382575	0.4481244	0.9253395
	40k[‡‡]	Variovorax sp. TUT1027; AB098595	Variovorax	10,450	0.26125	0.3060119	0.9306991
		Uncultured eubacterium WD2115; AJ292627	Variovorax	4,424	0.1106	0.1295499	0.9046336
		Variovorax paradoxus S110; CP001635	Variovorax	3,972	0.0993	0.1163138	0.9478978
				34,149	**0.853725**		
Genus	*Helicobacter*	*Helicobacter* ganmani; ES-5; AY561831	*Helicobacter*	38,664	0.840522	0.8623043	0.8908472
	46k[‡‡]	*Helicobacter* hepaticus; AJ007931	*Helicobacter*	4,944	0.107478	0.1102636	0.8031028
		Uncultured bacterium; L-123; EU622666	*Helicobacter*	646	0.014043	0.0144074	0.8069954
		Uncultured bacterium; MD2_aap36e09; EU508632	*Helicobacter*	584	0.012696	0.0130247	0.928738
				44,838	**0.974739**		
Genus	*Bacteroides*	*Bacteroides* acidifaciens; AB021157	*Bacteroides*	6,672	0.128308	0.3118048	0.8855073
	52k[‡‡]	Uncultured bacterium; SWPT13_aaa01g04; EF096855	*Bacteroides*	6,455	0.124135	0.3016637	0.8953075
		Uncultured bacterium; HY1_h06_1; EU458381	Odoribacter	4,220	0.081154	0.1972147	0.7443513
		Uncultured bacterium; K80N2_04b08; EU454172	*Bacteroides*	4,051	0.077904	0.1893168	0.906758
				21,398	**0.4115**		
Genus	*Parabacteroides*	Uncultured bacterium; lean2_aaa01f09; EF096000	Parabacteroides	17,825	0.445625	0.6227509	0.8851896
	40k[‡‡]	Uncultured bacterium; SJTU_A2_04_88; EF403654	Parabacteroides	4,034	0.10085	0.1409356	0.9071552
		Uncultured bacterium; RL246_aai73h07; DQ793582	Parabacteroides	3,733	0.093325	0.1304196	0.9290656
		Uncultured bacterium; WF16S_154; EU939416	Parabacteroides	3,031	0.075775	0.1058939	0.9160894
				28,623	**0.715575**		
Genus	*Marinilabilia*	Uncultured bacterium; HD5++50; EU791010	Barnesiella	10,475	0.261875	0.5164423	0.8619934
	40k[‡‡]	Uncultured bacterium; nbt15e03; FJ893065	Barnesiella	6,548	0.1637	0.3228319	0.8312596
		Uncultured bacterium; mcbc135; AM932661	Odoribacter	1,842	0.04605	0.090815	0.7321471
		Uncultured bacterium; C20_j04; AY991881	Odoribacter	1,418	0.03545	0.0699108	0.8218131
				20,283	**0.507075**		
Genus	*Alistipes*	Uncultured bacterium; WD3_aak03b12; EU510226	Alistipes	8,234	0.20585	0.2924006	0.8541191
	40k[‡‡]	Uncultured bacterium; cc_74; GQ175415	Alistipes	7,759	0.193975	0.2755327	0.8011231
		Uncultured bacterium; WD4_aal37e01; EU510373	Alistipes	7,640	0.191	0.2713068	0.8777465
		Uncultured bacterium; 16saw34-1g01.w2k; EF603689	Alistipes	4,527	0.113175	0.1607599	0.8833928
				28,160	**0.704**		
Genus	*Rikenella*	Uncultured bacterium; WD3_aak01e03; EU510108	Unclassified Bacteroidales	30,563	0.72769	0.8172799	0.8550692
	42k[‡‡]	Uncultured bacterium; C21_e10; AY993107	Unclassified Bacteroidales	2,858	0.068048	0.0764253	0.8142789
		Uncultured bacterium; cc_96; GQ175429	Rikenella	2,277	0.054214	0.0608889	0.6285806
		Uncultured bacterium; 2.16F; EU655924	Unclassified Bacteroidales	1,698	0.040429	0.0454059	0.7522668
				37,396	**0.890381**		
Family	Peptostreptococcaceae	Uncultured bacterium; R-9612; FJ880565	Sporacetigenium	14,429	0.327932	0.5388982	0.8899369
	44k[‡‡]	Uncultured bacterium; MD23_2aaa04g05; EU507538	Sporacetigenium	5,811	0.132068	0.2170308	0.9109019
		Uncultured bacterium; MD18_aaa01c10; EU506158	Sporacetigenium	4,063	0.092341	0.151746	0.9240386
		Uncultured bacterium; MD19_aaa01c03; EU506401	Sporacetigenium	2,472	0.056182	0.0923249	0.897591
				26,775	**0.608523**		
Genus	*Lactococcus*	Lactococcus lactis subsp. cremoris; YIT 2007 (ATCC 19257); AB008214	Lactococcus	10,278	0.25695	0.3364982	0.8874481
	40k[‡‡]	Lactococcus lactis subsp. cremoris SK11; CP000425	Lactococcus	8,971	0.224275	0.2937074	0.8944487
		Uncultured bacterium; 1–5D; EU289440	Lactococcus	6,533	0.163325	0.2138882	0.896064

Table S1. Cont.

Taxonomic rank	Taxa[*]	Top organisms[†]	Taxa represented[‡]	Counts[§]	Prop total[¶]	Prop top hits[**]	Average S_ab score[††]
		Lactococcus lactis subsp. lactis; RO6; AF515224	Lactococcus	4,762	0.11905	0.1559062	0.9152703
				30,544	**0.7636**		
Genus	*Roseburia*	Uncultured bacterium; RL184_aao65g01; DQ809900	Roseburia	9,117	0.227925	0.4072635	0.8149241
	40k[‡‡]	Uncultured bacterium; CRWD2_aaa03d03; EU503700	Roseburia	8,557	0.213925	0.3822478	0.8707943
		Uncultured bacterium; CRWD5_aaa04f02; EU504227	Unclassified Lachnospiraceae	2,445	0.061125	0.10922	0.9149681
		Uncultured bacterium; K74N1_19e08; EU455153	Unclassified Clostridiales	2,267	0.056675	0.1012687	0.9076903
				22,386	**0.55965**		
Genus	Turicibacter	Uncultured bacterium; control_7 d-F2; EF406422	Turicibacter	18,682	0.44481	0.4727705	0.8986114
	42k[‡‡]	Uncultured bacterium; infected_7 d-E1; EF406660	Turicibacter	17,816	0.42419	0.4508553	0.8995282
		Uncultured bacterium; R-6524; FJ880085	Turicibacter	1,621	0.038595	0.0410214	0.9036231
		Uncultured bacterium; R-9107; FJ881096	Turicibacter	1,397	0.033262	0.0353528	0.8997015
				39,516	**0.940857**		
Order	Coriobacteriales	Uncultured bacterium; C18_f09_1; EF614565	Unclassified Coriobacteriaceae	2,951	0.210786	0.4407767	0.8278333
	14k[‡‡]	Uncultured bacterium; MD2_aap35a10; EU508535	Asaccharobacter	1,520	0.108571	0.2270351	0.9184368
		Coriobacteriaceae bacterium B7; DQ789120	Unclassified Coriobacteriaceae	1,201	0.085786	0.1793876	0.9162223
		Uncultured bacterium; SWPT20_aaa03a06; EF097741	Unclassified Coriobacteriaceae	1,023	0.073071	0.1528006	0.9142815
				6,695	**0.478214**		
Family	Coriobacteriaceae	Uncultured bacterium; C18_f09_1; EF614565	Unclassified Coriobacteriaceae	2,951	0.210786	0.4407767	0.8278333
	14k[‡‡]	Uncultured bacterium; MD2_aap35a10; EU508535	Asaccharobacter	1,520	0.108571	0.2270351	0.9184368
		Coriobacteriaceae bacterium B7; DQ789120	Unclassified Coriobacteriaceae	1,201	0.085786	0.1793876	0.9162223
		Uncultured bacterium; SWPT20_aaa03a06; EF097741	Unclassified Coriobacteriaceae	1,023	0.073071	0.1528006	0.9142815
				6,695	**0.478214**		

---

[*] All sequences from respective taxa assigned by CLASSIFIER were extracted. At least 40,000 random sequences were selected from each taxon and analyzed by RDP SEQMATCH. Taxa with fewer than 40,000 sequences were analyzed to completion.

[†] Top four bacteria with the most sequence matches to the RDP SeqMatch database for the given taxa.

[‡] The lowest taxonomic rank assigned by RDP SeqMatch for the given top organism.

[§] The number of matches to the database for the given top organism.

[¶] The proportion of sequences matching the given top organism divided by the total number of sequences pooled for analysis of the given taxa.

[**] The proportion of the sequences matching the given top organism divided by the compiled amount of sequences making up all four top organisms for the given taxa.

[††] The average of RDP SeqMatch score (S_ab). The S_ab score is the number of (unique) 7-base oligomers shared between the query sequence and a given RDP sequence divided by the lowest number of unique oligos in either of the two sequences.

[‡‡] The total number of sequences pooled for RDP SeqMatch analysis for the given taxa.

Table S2. Descriptive statistics for CMM traits measured in the G_4 population

		Average*	SD	Min	Max
Phylum	Actinobacteria	−2.67739	0.590533	−4.39794	−0.39324
Class	Actinobacteria	−2.67739	0.590533	−4.39794	−0.39324
Subclass	Coriobacteridae	−2.94055	0.584882	−4.39794	−0.54654
Order	Coriobacteriales	−2.94055	0.584882	−4.39794	−0.54654
Suborder	Coriobacterineae	−2.94055	0.584882	−4.39794	−0.54654
Family	Coriobacteriaceae	−2.94055	0.584882	−4.39794	−0.54654
Subclass	Actinobacteridae	−3.28907	0.685807	−4.65758	−0.39482
Phylum	Bacteroidetes	−0.64014	0.322578	−2.21247	−0.07597
Class	Flavobacteria	−3.24884	0.768164	−4.63827	−1.20137
Order	Flavobacteriales	−3.24884	0.768164	−4.63827	−1.20137
Family	Flavobacteriaceae	−3.24962	0.768473	−4.63827	−1.20137
Class	Bacteroidetes	−0.86399	0.340153	−2.55486	−0.30995
Order	Bacteroidales	−0.86399	0.340153	−2.55486	−0.30995
Family	Rikenellaceae	−1.45665	0.361402	−2.86902	−0.78168
Genus	*Odoribacter*	−2.69635	0.658767	−4.55284	−1.61057
Genus	*Alistipes*	−1.82236	0.403583	−3.16052	−0.96128
Genus	*Rikenella*	−3.0305	0.75621	−4.55284	−1.60478
Family	Bacteroidaceae	−1.81256	0.524582	−4.25181	−0.56101
Genus	*Bacteroides*	−1.8127	0.524608	−4.25181	−0.56101
Family	Porphyromonadaceae	−1.83477	0.483651	−4.25181	−0.69437
Genus	*Parabacteroides*	−1.83713	0.483785	−4.25181	−0.69497
Phylum	Proteobacteria	−1.29749	0.441019	−2.74642	−0.19835
Class	Epsilonproteobacteria	−2.12995	0.931753	−4.65758	−0.50796
Order	Campylobacterales	−2.12999	0.931747	−4.65758	−0.50796
Family	Helicobacteraceae	−2.14262	0.94777	−4.65758	−0.50813
Genus	*Helicobacter*	−2.15126	0.950455	−4.65758	−0.51233
Class	Deltaproteobacteria	−2.18371	0.669213	−4.21467	−0.81547
Class	Alphaproteobacteria	−2.79909	0.689063	−4.60206	−1.0188
Order	Rhizobiales	−3.16046	0.793923	−4.65758	−1.25349
Class	Gammaproteobacteria	−2.38611	0.639983	−4.23657	−0.21968
Order	Pseudomonadales	−2.82208	0.638174	−4.45593	−0.21992
Order	Enterobacteriales	−2.83163	0.598018	−4.38722	−1.44615
Family	Enterobacteriaceae	−2.83163	0.598018	−4.38722	−1.44615
Class	Betaproteobacteria	−2.2471	0.633134	−4.05552	−0.41858
Order	Burkholderiales	−2.38423	0.667937	−4.05552	−0.42322
Family	Comamonadaceae	−2.4195	0.676075	−4.05552	−0.43046
Genus	*Variovorax*	−2.66522	0.751411	−4.25181	−0.43876
Phylum	Firmicutes	−0.27565	0.143062	−1.06802	−0.0228
Class	Bacilli	−1.20876	0.500503	−2.47638	−0.10353
Order	Lactobacillales	−1.23337	0.502172	−2.5085	−0.10555
Family	Lactobacillaceae	−1.73651	0.687982	−4.08619	−0.10924
Genus	*Lactobacillus*	−1.74217	0.687414	−4.08619	−0.11181
Family	Leuconostocaceae	−2.67244	0.558744	−4.45593	−1.14704
Genus	*Weissella*	−2.80507	0.626531	−4.65758	−1.21328
Family	Streptococcaceae	−1.73707	0.5654	−3.28651	−0.28054
Genus	*Lactococcus*	−1.75409	0.572448	−3.28651	−0.28122
Order	Bacillales	−2.96039	0.641711	−4.36653	−0.69596
Class	Erysipelotrichi	−2.41441	0.808503	−4.27572	−0.42666
Order	Erysipelotrichales	−2.41441	0.808503	−4.27572	−0.42666
Family	Erysipelotrichaceae	−2.41441	0.808503	−4.27572	−0.42666
Genus	*Turicibacter*	−2.69515	0.992398	−4.55284	−0.42707
Class	Clostridia	−0.42739	0.192067	−1.52896	−0.07883
Order	Clostridiales	−0.43079	0.192643	−1.53304	−0.08154
Family	Lachnospiraceae	−0.70714	0.232744	−2.04648	−0.26048
Genus	*Lachnobacterium*	−3.35505	0.842755	−4.5376	−0.92087
Genus	*Dorea*	−2.38523	0.446171	−4.18709	−1.11065
Genus	*Lachnospiraceae Incertae Sedis*	−2.55034	0.383489	−4.25964	−1.68721
Genus	*Roseburia*	−2.89953	0.590418	−4.65758	−0.5072
Family	Peptostreptococcaceae	−2.84529	0.996391	−4.58503	−0.27802
Genus	*Peptostreptococcaceae Incertae Sedis*	−2.85821	0.998041	−4.58503	−0.28497
Family	Ruminococcaceae	−1.5107	0.244235	−2.61386	−0.62938
Family	Clostridiaceae	−3.44843	0.821475	−4.72125	−0.70714
Subfamily	Clostridiaceae 1	−3.4492	0.821532	−4.72125	−0.70714
Genus	*Clostridium*	−3.55616	0.770174	−4.72125	−0.86786

*Prop values of 0 were replaced with 0.5/total reads. and all Prop values were log10-transformed for descriptive statistics.

Table S3. Mixed-model analysis of CMM traits with an across-taxa FDR < 0.05

Rank	Taxon	Source of variation[*]	*P* value[†]	FDR[‡]
Phylum	Proteobacteria	Parent of origin	0.0022	0.017925
Class	Deltaproteobacteria	Parent of origin	<0.0001	0.000453
Class	Epsilonproteobacteria	Parent of origin	<0.0001	0.000453
Order	Campylobacterales	Parent of origin	<0.0001	0.000453
Family	Clostridiaceae	Parent of origin	0.0112	0.049683
Family	Helicobacteraceae	Parent of origin	<0.0001	0.000453
Family	Peptostreptococcaceae	Parent of origin	0.0115	0.049683
Family	Ruminococcaceae	Parent of origin	0.0006	0.005208
Subfamily	Clostridiaceae 1	Parent of origin	0.0111	0.049683
Genus	*Dorea*	Parent of origin	0.0025	0.018041
Genus	*Helicobacter*	Parent of origin	<0.0001	0.000453
Genus	*Lachnobacterium*	Parent of origin	0.006	0.035161
Genus	*Lachnospiraceae Incertae sedis*	Parent of origin	0.0005	0.005208
Genus	*Peptostreptococcaceae Incertae sedis*	Parent of origin	0.0116	0.049683
Genus	*Rikenella*	Parent of origin	0.0049	0.031611
Phylum	Actinobacteria	Sex	0.0024	0.016972
Class	Actinobacteria	Sex	0.0024	0.016972
Class	Epsilonproteobacteria	Sex	0.0073	0.033396
Class	Erysipelotrichi	Sex	0.0068	0.033396
Subclass	Coriobacteridae	Sex	0.0006	0.006676
Order	Bacillales	Sex	0.0108	0.04052
Order	Campylobacterales	Sex	0.0073	0.033396
Order	Coriobacteriales	Sex	0.0006	0.006676
Order	Erysipelotrichales	Sex	0.0068	0.033396
Suborder	Coriobacterineae	Sex	0.0006	0.006676
Family	Coriobacteriaceae	Sex	0.0006	0.006676
Family	Erysipelotrichaceae	Sex	0.0068	0.033396
Family	Helicobacteraceae	Sex	0.0092	0.036682
Family	Peptostreptococcaceae	Sex	<0.0001	0.002721
Genus	*Helicobacter*	Sex	0.0082	0.035158
Genus	Peptostreptococcaceae Incertae sedis	Sex	<0.0001	0.002721
Genus	Turicibacter	Sex	0.0015	0.013717
Phylum	Actinobacteria	Cohort	<0.0001	0.000025
Phylum	Bacteroidetes	Cohort	<0.0001	0
Phylum	Firmicutes	Cohort	<0.0001	0.000025
Phylum	Proteobacteria	Cohort	<0.0001	0.000121
Subclass	Actinobacteridae	Cohort	0.0013	0.002488
Subclass	Coriobacteridae	Cohort	<0.0001	0.00002
Order	Enterobacteriales	Cohort	<0.0001	0.000001
Order	Flavobacteriales	Cohort	0.0002	0.000369
Order	Lactobacillales	Cohort	0.0007	0.001321
Order	Pseudomonadales	Cohort	<0.0001	0.000063
Order	Rhizobiales	Cohort	<0.0001	0.000197
Suborder	Coriobacterineae	Cohort	<0.0001	0.00002
Class	Bacteroidetes	Dam	0.0004	0.011485
Order	Bacteroidales	Dam	0.0004	0.011485
Phylum	Actinobacteria	Litter	0.0009	0.004262
Phylum	Proteobacteria	Litter	0.0017	0.006141
Class	Actinobacteria	Litter	0.0009	0.004262
Class	Deltaproteobacteria	Litter	0.0104	0.025511
Class	Epsilonproteobacteria	Litter	<0.0001	0.001275
Class	Erysipelotrichi	Litter	0.0001	0.001275
Class	Flavobacteria	Litter	0.0004	0.002262
Subclass	Actinobacteridae	Litter	0.0017	0.006141
Subclass	Coriobacteridae	Litter	0.0019	0.006141
Order	Bacillales	Litter	0.011	0.026008
Order	Burkholderiales	Litter	0.0064	0.016302
Order	Campylobacterales	Litter	<0.0001	0.001275
Order	Coriobacteriales	Litter	0.0019	0.006141
Order	Erysipelotrichales	Litter	0.0001	0.001275
Order	Flavobacteriales	Litter	0.0004	0.002262
Order	Pseudomonadales	Litter	0.0155	0.03545
Order	Rhizobiales	Litter	0.0007	0.003648

Table S3. Cont.

Rank	Taxon	Source of variation*	*P* value[†]	FDR[‡]
Suborder	Coriobacterineae	Litter	0.0019	0.006141
Family	Clostridiaceae	Litter	0.0056	0.01482
Family	Comamonadaceae	Litter	0.0164	0.036246
Family	Coriobacteriaceae	Litter	0.0019	0.006141
Family	Erysipelotrichaceae	Litter	0.0001	0.001275
Family	Flavobacteriaceae	Litter	0.0004	0.002262
Family	Helicobacteraceae	Litter	0.0001	0.001275
Subfamily	Clostridiaceae 1	Litter	0.0055	0.01482
Genus	*Clostridium*	Litter	0.0031	0.009582
Genus	*Dorea*	Litter	0.0198	0.042126
Genus	*Helicobacter*	Litter	0.0002	0.001275
Genus	*Lachnobacterium*	Litter	0.0055	0.01482
Genus	*Turicibacter*	Litter	0.0002	0.001275
Genus	*Weissella*	Litter	0.0204	0.042126

*Abbreviated notation for sources of variation: cohort for cohort(parent of origin), dam for dam(parent of origin), and litter for litter (parent of origin*cohort*dam).

[†] The *P* value is the probability of obtaining a larger F value in the individual taxon analysis.

[‡] FDR is the across-taxa false discovery rate adjusted *P* value calculated separately for each source of variation.

Table S4. REML estimated variance components of CMM traits

Rank	Taxon	Proportion of total variance*				Variance[†]			
		Cohort	Family	Litter	Residual	Cohort	Family	Litter	Residual
Phylum	Actinobacteria	0.39788	0	0.08413	0.51798	0.13479	0	0.0285	0.1755
Phylum	Bacteroidetes	0.27611	0.06455	0	0.65935	0.03189	0.00745	0	0.07615
Phylum	Firmicutes	0.32759	0.06585	0	0.60656	0.00601	0.00121	0	0.01114
Phylum	Proteobacteria	0.41389	0.04939	0.08617	0.45055	0.05838	0.00697	0.01215	0.06355
Class	Actinobacteria	0.39788	0	0.08413	0.51798	0.13479	0	0.0285	0.1755
Class	Alphaproteobacteria	0.38916	0.00749	0.04531	0.55804	0.16557	0.00319	0.01928	0.2374
Class	Bacilli	0.24173	0.04747	0.01526	0.69553	0.06647	0.01305	0.0042	0.1913
Class	Bacteroidetes	0.26202	0.07093	0	0.66705	0.03386	0.00917	0	0.08619
Class	Betaproteobacteria	0.25345	0.01973	0.03592	0.6909	0.07182	0.00559	0.01018	0.1958
Class	Clostridia	0.05253	0.04713	0	0.90035	0.00187	0.00167	0	0.03199
Class	Deltaproteobacteria	0.18166	0	0.1135	0.70485	0.0465	0	0.02905	0.1804
Class	Epsilonproteobacteria	0.17242	0.16151	0.16859	0.49747	0.11118	0.10414	0.10871	0.3208
Class	Erysipelotrichi	0.1029	0.04101	0.12422	0.73187	0.07042	0.02807	0.08501	0.5009
Class	Flavobacteria	0.36772	0	0.08908	0.54319	0.17613	0	0.04267	0.2602
Class	Gammaproteobacteria	0.46878	0.01188	0.05323	0.46612	0.19338	0.0049	0.02196	0.1923
Order	Bacillales	0.21766	0.03951	0.08784	0.65499	0.08841	0.01605	0.03568	0.266
Order	Bacteroidales	0.26202	0.07093	0	0.66705	0.03386	0.00917	0	0.08619
Order	Burkholderiales	0.27349	0	0.0543	0.67221	0.08385	0	0.01665	0.2061
Order	Clostridiales	0.05309	0.04697	0	0.89994	0.0019	0.00168	0	0.03214
Order	Coriobacteriales	0.42787	0	0.08143	0.49069	0.15041	0	0.02863	0.1725
Order	Enterobacteriales	0.41728	0.00428	0.05511	0.52333	0.14701	0.00151	0.01941	0.1844
Order	Erysipelotrichales	0.1029	0.04101	0.12422	0.73187	0.07042	0.02807	0.08501	0.5009
Order	Flavobacteriales	0.36772	0	0.08908	0.54319	0.17613	0	0.04267	0.2602
Order	Lactobacillales	0.23713	0.04833	0.01495	0.6996	0.06562	0.01337	0.00414	0.1936
Order	Pseudomonadales	0.35992	0.00417	0.06987	0.56604	0.1392	0.00161	0.02702	0.2189
Order	Rhizobiales	0.42862	0	0.07539	0.496	0.22995	0	0.04044	0.2661
Suborder	Coriobacterineae	0.42787	0	0.08143	0.49069	0.15041	0	0.02863	0.1725
Family	Bacteroidaceae	0.30932	0.00968	0.06251	0.6185	0.08616	0.0027	0.01741	0.1723
Family	Clostridiaceae	0.21729	0.02276	0.08559	0.67436	0.144	0.01509	0.05672	0.4469
Family	Comamonadaceae	0.27078	0	0.04536	0.68385	0.08496	0	0.01423	0.2146
Family	Coriobacteriaceae	0.42787	0	0.08143	0.49069	0.15041	0	0.02863	0.1725
Family	Enterobacteriaceae	0.41728	0.00428	0.05511	0.52333	0.14701	0.00151	0.01941	0.1844
Family	Erysipelotrichaceae	0.1029	0.04101	0.12422	0.73187	0.07042	0.02807	0.08501	0.5009
Family	Flavobacteriaceae	0.36838	0	0.08826	0.54336	0.17655	0	0.0423	0.2604
Family	Helicobacteraceae	0.14769	0.17841	0.16148	0.51241	0.0982	0.11863	0.10737	0.3407
Family	Lachnospiraceae	0.06745	0.03029	0.01847	0.88379	0.00366	0.00164	0.001	0.04789
Family	Lactobacillaceae	0.08002	0.07373	0	0.84625	0.0404	0.03722	0	0.4273
Family	Leuconostocaceae	0.47525	0.00714	0.06711	0.45051	0.15215	0.00229	0.02148	0.1442
Family	Peptostreptococcaceae	0.16267	0.01352	0.05315	0.77067	0.15868	0.01318	0.05184	0.7518
Family	Porphyromonadaceae	0.19482	0.09877	0.01168	0.69472	0.05108	0.0259	0.00306	0.1822
Family	Rikenellaceae	0.26116	0.05894	0	0.67991	0.03695	0.00834	0	0.09619
Family	Ruminococcaceae	0.08045	0.05935	0	0.8602	0.00437	0.00322	0	0.04669
Family	Streptococcaceae	0.46151	0.02891	0.05591	0.45368	0.14802	0.00927	0.01793	0.1455
Subfamily	Clostridiaceae 1	0.21721	0.02305	0.08595	0.67379	0.14392	0.01527	0.05695	0.4465
Genus	*Alistipes*	0.30995	0	0.05017	0.63988	0.05545	0	0.00898	0.1145
Genus	*Bacteroides*	0.30922	0.00943	0.06259	0.61876	0.08614	0.00263	0.01744	0.1724
Genus	*Clostridium*	0.22068	0.01223	0.09866	0.66843	0.13105	0.00726	0.05859	0.3969
Genus	*Dorea*	0.18903	0.07916	0.07951	0.6523	0.03365	0.01409	0.01415	0.1161
Genus	*Helicobacter*	0.15294	0.17651	0.16352	0.50703	0.1031	0.11899	0.11023	0.3418
Genus	*Lachnobacterium*	0.11745	0.11622	0.08846	0.67787	0.0778	0.07698	0.05859	0.449
Genus	*Lachnospiraceae Incertae sedis*	0.08521	0.10581	0	0.80897	0.01086	0.01349	0	0.1031
Genus	*Lactobacillus*	0.07871	0.07461	0	0.84668	0.03968	0.03762	0	0.4269
Genus	*Lactococcus*	0.46108	0.02927	0.0544	0.45525	0.15159	0.00962	0.01788	0.1497
Genus	*Odoribacter*	0.13062	0.02303	0.035	0.81134	0.06023	0.01062	0.01614	0.3741
Genus	*Parabacteroides*	0.19493	0.09813	0.0125	0.69444	0.0511	0.02572	0.00328	0.182
Genus	*Peptostreptococcaceae Incertae sedis*	0.15946	0.01438	0.05114	0.77502	0.1563	0.0141	0.05013	0.7597
Genus	*Rikenella*	0.1678	0.02807	0.05314	0.75099	0.08676	0.01452	0.02748	0.3883
Genus	*Roseburia*	0.0487	0.12158	0.08587	0.74385	0.01454	0.0363	0.02564	0.2221
Genus	*Turicibacter*	0.09938	0.04869	0.12768	0.72425	0.09663	0.04734	0.12415	0.7042
Genus	*Variovorax*	0.33673	0.00279	0.04756	0.61292	0.13608	0.00113	0.01922	0.2477
Genus	*Weissella*	0.52388	0.00223	0.0754	0.3985	0.1991	0.00085	0.02865	0.1514

*Proportion of total variance is the variance divided by the sum of the cohort, family, litter, and residual variances.

[†] Variances were estimated using REML with a linear mixed model that included fixed effects for parent of origin and sex and random effects for cohort(parent of origin), family(parent of origin), and litter(parent of origin*cohort*family).

Table S5. QTL detected and respective statistics for Prop1 traits

Trait		Nearest marker	Chromosome	Peak position, Mb	Naive LOD	GRAIP LOD*	95% CI, Mb[†]	% Var[‡]	Additive ± SE[§]	Dominance ± SE[§]
Phylum	Actinobacteria									
Subclass	Coriobacteridae	JAX00300375	10	119	7.2	3.9**	104–123	5.7	0.20 ± 0.03[¶]	–0.03 ± 0.05
Order	Coriobacteriales	JAX00300375	10	119	7.1	4.0**	105–122	5.7	0.20 ± 0.03[¶]	–0.03 ± 0.05
Suborder	Coriobacterineae	JAX00300375	10	119	7.0	3.9**	104–123	5.7	0.20 ± 0.03[¶]	–0.03 ± 0.05
Family	Coriobacteriaceae	JAX00300375	10	119	7.3	4.2**	106–122	5.7	0.20 ± 0.03[¶]	–0.03 ± 0.05
Phylum	Proteobacteria	JAX00139228	6	28	8.6	4.1**	–40	1.5	–0.05 ± 0.02	0.08 ± 0.03
		JAX00666793	8	43	8.6	4.1**	33–63	3.2	–0.08 ± 0.02	0.12 ± 0.03
Class	Epsilonproteobacteria	JAX00603343	6	13	9.2	4.7**	–39	1.7	–0.03 ± 0.05	0.24 ± 0.07
Order	Campylobacterales	JAX00603343	6	13	9.2	4.7**	–39	1.7	–0.03 ± 0.05	0.24 ± 0.07
Family	Helicobacteraceae	JAX00603343	6	13	8.7	4.4**	–39	1.7	–0.03 ± 0.05	0.24 ± 0.07
Genus	*Helicobacter*	JAX00603343	6	13	8.8	4.4**	–39	1.6	–0.02 ± 0.05	0.24 ± 0.08
Class	Deltaproteobacteria	JAX00480903	19	56	5.1	3.9**	54-	2.5	–0.10 ± 0.04	0.17 ± 0.05
Class	Gammaproteobacteria	JAX00707462	9	119	6.2	3.6	117-	4.0	–0.14 ± 0.04	0.19 ± 0.05
Order	Pseudomonadales	JAX00707462	9	119	6.8	3.8	117-	4.4	–0.14 ± 0.04	0.21 ± 0.05
Class	Betaproteobacteria	JAX00633165	7	19	8.6	4.7**	15–29	6.3	–0.22 ± 0.03[¶]	–0.10 ± 0.05
Order	Burkholderiales	JAX00633165	7	19	10.7	4.7**	14–33	7.9	–0.26 ± 0.04[¶]	–0.11 ± 0.05
Family	Comamonadaceae	JAX00633165	7	19	10.8	4.7**	13–34	7.9	–0.26 ± 0.04[¶]	–0.12 ± 0.05
Genus	*Variovorax*	JAX00633165	7	19	9.7	4.7**	14–28	7.2	–0.27 ± 0.04[¶]	–0.16 ± 0.06
Phylum	Firmicutes									
Species	*L.johnsonii/L.gasseri* 97%	JAX00641805	7	66	6.8	4.7**	47–71	4.7	–0.27 ± 0.05[¶]	–0.11 ± 0.07
		JAX00387018	14	93	5.8	4.7**	86–103	3.9	–0.23 ± 0.05[¶]	–0.16 ± 0.07
Family	Streptococcaceae	JAX00022058	10	107	8.0	4.7**	101–111	7.0	0.21 ± 0.03[¶]	–0.05 ± 0.04
Genus	*Lactococcus*	JAX00022058	10	107	8.0	4.7**	100–111	7.0	0.21 ± 0.03[¶]	–0.05 ± 0.05
Class	Erysipelotrichi	JAX00643377	7	73	6.4	4.0**	65–88	5.0	–0.24 ± 0.04[¶]	0.03 ± 0.06
Order	Erysipelotrichales	JAX00643377	7	73	6.5	4.2**	67–87	5.0	–0.24 ± 0.04[¶]	0.03 ± 0.06
Family	Erysipelotrichaceae	JAX00643377	7	73	6.5	4.0**	66–88	5.0	–0.24 ± 0.04[¶]	0.03 ± 0.06
Genus	*Turicibacter*	JAX00643377	7	73	7.1	4.6**	71–88	5.3	–0.30 ± 0.05[¶]	0.09 ± 0.08
Family	Peptostreptococcaceae	JAX00010715	1	148	5.8	3.8	143–150	4.4	–0.25 ± 0.05[¶]	0.16 ± 0.08
Genus	*Peptostreptococcaceae IS*	JAX00010715	1	148	5.7	3.7	143–150	4.3	–0.25 ± 0.05[¶]	0.17 ± 0.08
Family	Ruminococcaceae	JAX00327082	12	17	5.5	4.4**	–26	3.4	0.06 ± 0.01[¶]	0.04 ± 0.02
Phylum	Bacteriodetes									
Genus	*Barnesiella*	JAX00005735	1	80	10.7	4.7**	63–139	9.0	–0.23 ± 0.03[¶]	0.14 ± 0.05
		JAX00173791	9	87	4.6	3.5	72–104	3.4	–0.14 ± 0.04	0.14 ± 0.05

*LOD exceeding the 95% ($P = 0.05$, LOD ≥ 3.9) permutation threshold are denoted by **; other QTL exceeded the 90% ($P = 0.1$, LOD ≥ 3.5) threshold.

[†] Confidence intervals for QTL positions were obtained using a 1.0 LOD drop in Mb (relative to the GRAIP-permuted LOD score).

[‡] Percentage of phenotypic variance accounted for by the QTL effect.

[§] For additive and dominance effects: positive values indicate increasing effect of the HR allele or increasing effect of the heterozygote, respectively.

[¶] Indicates that additive and/or dominance effects were statistically significant at $P < 0.0$.

Table S6. **Genotype (C57BL/6J = BB; HR = AA) frequencies (% of total calls) at a given SNP location**

	Trait	SNP	MMU	% of BB	% of BA	% of AA
Subclass	Coriobacteridae	JAX00300375	10	30.3	51.6	18.1
Order	Coriobacteriales	JAX00300375	10	30.3	51.6	18.1
Suborder	Coriobacterineae	JAX00300375	10	30.3	51.6	18.1
Family	Coriobacteriaceae	JAX00300375	10	30.3	51.6	18.1
Genus	*Odoribacter*	JAX00005735	1	23.0	44.7	32.3
		JAX00173791	9	22.1	53.6	24.3
Phylum	Proteobacteria	JAX00139228	6	29.2	45.6	25.2
		JAX00666793	8	27.9	47.0	25.1
Class	Epsilonproteobacteria	JAX00603343	6	32.3	45.4	22.2
Order	Campylobacterales	JAX00603343	6	32.3	45.4	22.2
Family	Helicobacteraceae	JAX00603343	6	32.3	45.4	22.2
Genus	*Helicobacter*	JAX00603343	6	32.3	45.4	22.2
Class	Deltaproteobacteria	JAX00480903	19	23.0	55.3	21.7
Class	Gammaproteobacteria	JAX00707462	9	29.5	47.6	22.8
Order	Pseudomonadales	JAX00707462	9	29.5	47.6	22.8
Class	Betaproteobacteria	JAX00633165	7	21.5	50.6	27.9
Order	Burkholderiales	JAX00633165	7	21.5	50.6	27.9
Family	Comamonadaceae	JAX00633165	7	21.5	50.6	27.9
Genus	*Variovorax*	JAX00633165	7	21.5	50.6	27.9
Family	Streptococcaceae	JAX00022058	10	37.4	43.2	19.4
Genus	*Lactococcus*	JAX00022058	10	37.4	43.2	19.4
Species	*L.johnsonii/L.gasseri* 97%	JAX00641805	7	24.7	48.1	27.3
		JAX00387018	14	23.9	54.0	22.1
Class	Erysipelotrichi	JAX00643377	7	26.9	45.2	27.9
Order	Erysipelotrichales	JAX00643377	7	26.9	45.2	27.9
Family	Erysipelotrichaceae	JAX00643377	7	26.9	45.2	27.9
Genus	*Turicibacter*	JAX00643377	7	26.9	45.2	27.9
Family	Peptostreptococcaceae	JAX00010715	1	29.8	41.3	28.9
Genus	*Peptostreptococcaceae IS*	JAX00010715	1	29.8	41.3	28.9
Family	Ruminococcaceae	JAX00327082	12	22.8	48.2	29.0