Supplementary Information

Supplementary Figures

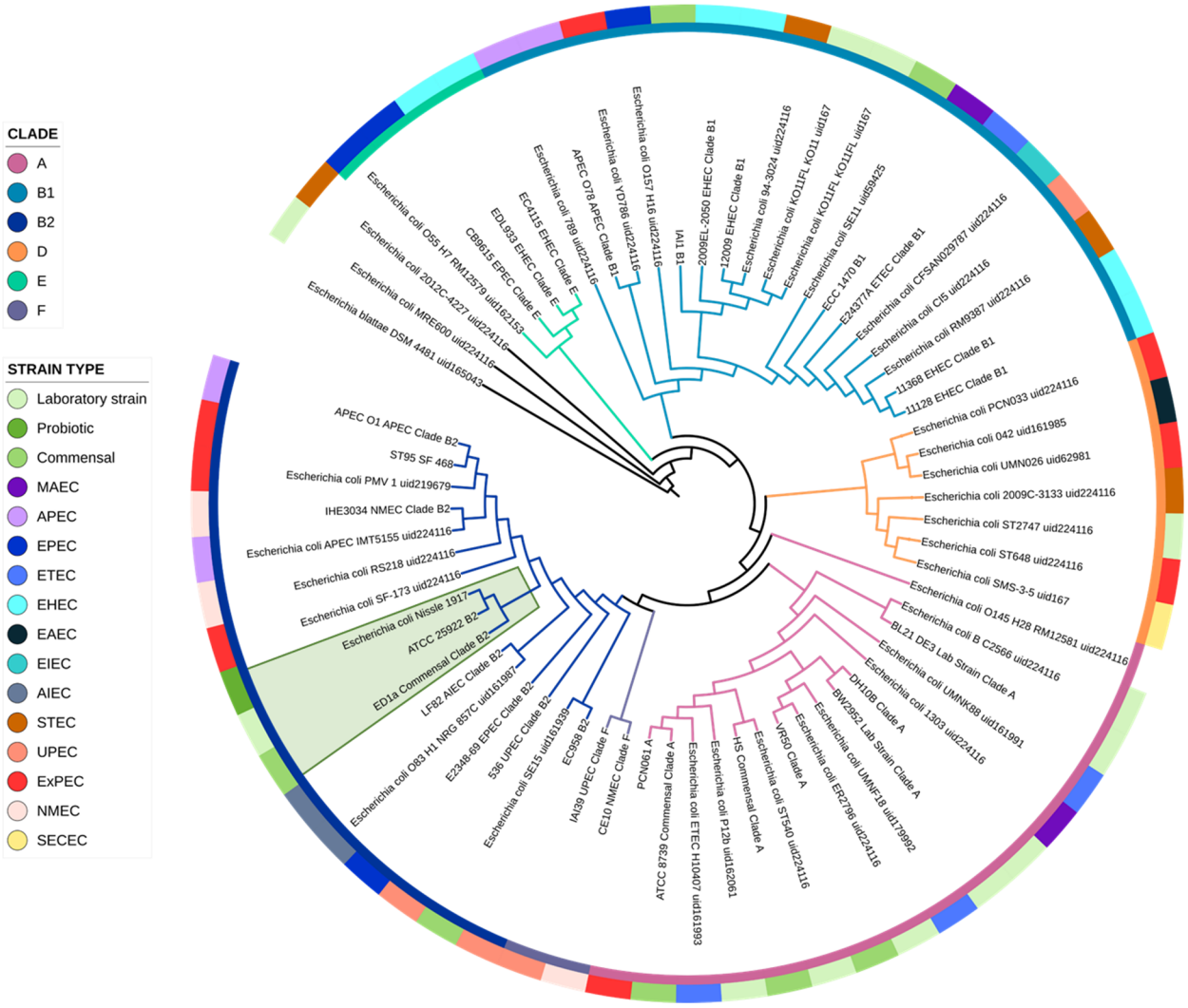


Figure S1. Phylogeny of *Escherichia* strains. Related to Figure 1. Maximum likelihood phylogeny of representative *Escherichia* strains annotated by clade (inner track, branches) and by strain type (outer track).

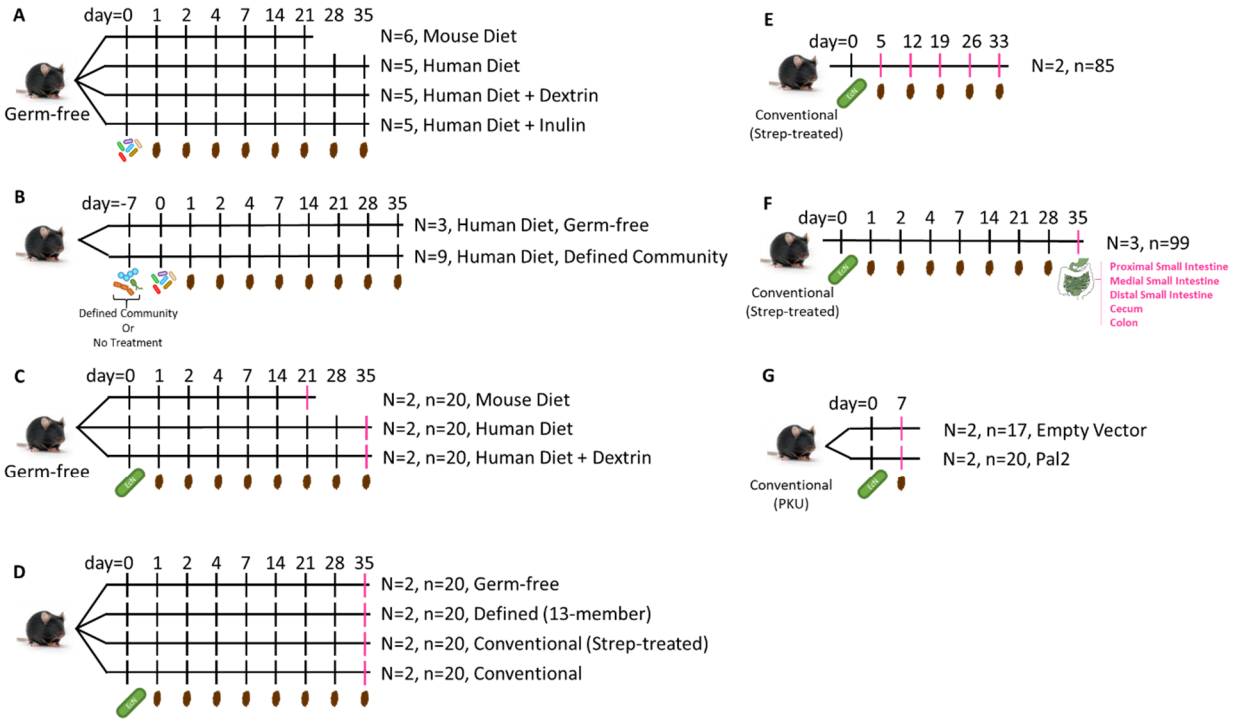
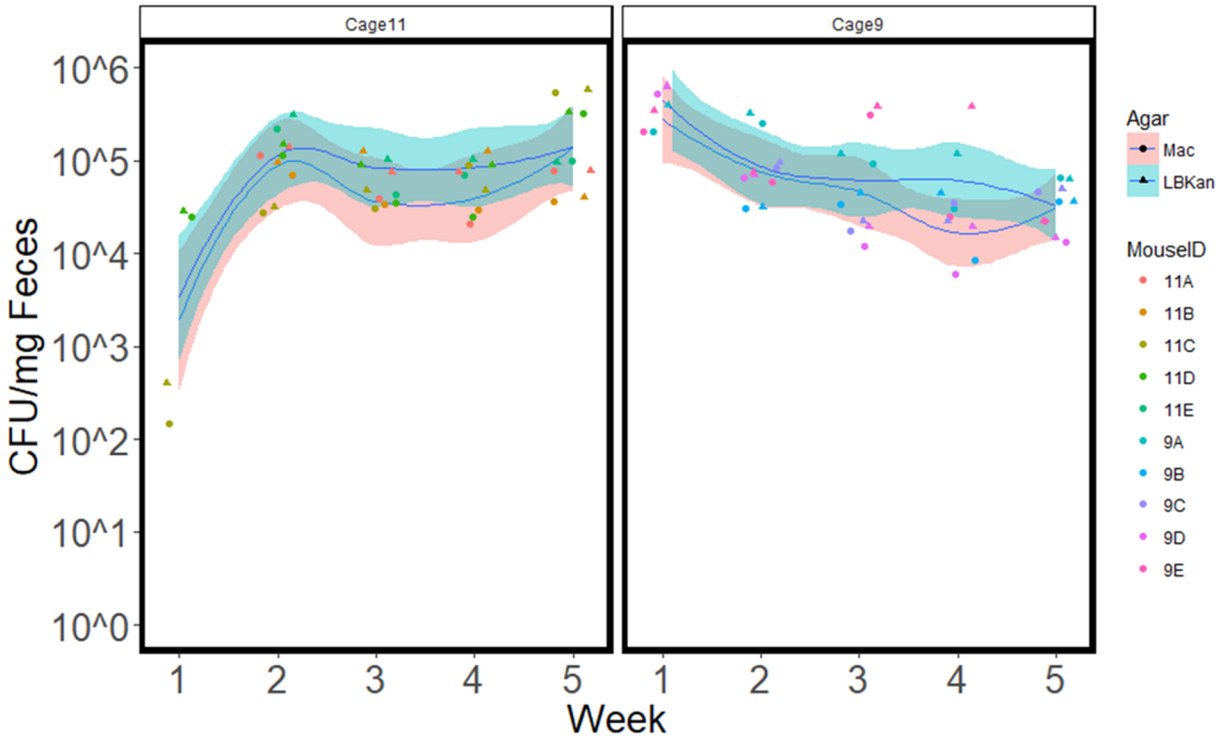


Figure S2. Schematic of *in vivo* functional metagenomic selections and adaptation experiments. Related to Table 1, Figures 1-3, and Figures 5-6. Experimental design and sampling timelines for A) functional metagenomic selections in germ-free mice B) functional metagenomic selections in mice pre-colonized with a defined community, C) genomic adaptation experiments in mice fed different diets, D) genomic adaptation experiments in mice with different gut microbiota complexities, E) longitudinal analysis of genomic adaptation in conventional, streptomycin treated mice, F) analysis of adaptations specific to gut sites, and G) genomic adaptation experiments in Pahenu2 mutant mice (PKU model) that were treated with EcN harboring either empty vector or the Pal2 gene. Mice were gavaged with EcN on day 0. N: number of replicate mice, n: total number of sequenced isolates. Pink bars: sampling time point from which EcN isolates were harvested. Mice with a defined community were pre-colonized 7 days before EcN Gavage.



21

22 **Figure S3. Assessment of *in vivo* plasmid loss. Related to Figure 1.** CFUs/mg of EcN recovered
23 from mouse fecal samples over 5 weeks when plated either on MacConkey agar (selects for all
24 EcN) or LB agar with Kanamycin (selects for EcN containing pZE21). Shaded regions represent
25 95% confidence intervals of LOESS regressions on the data. 10 mice from two different cages
26 were sampled.

27

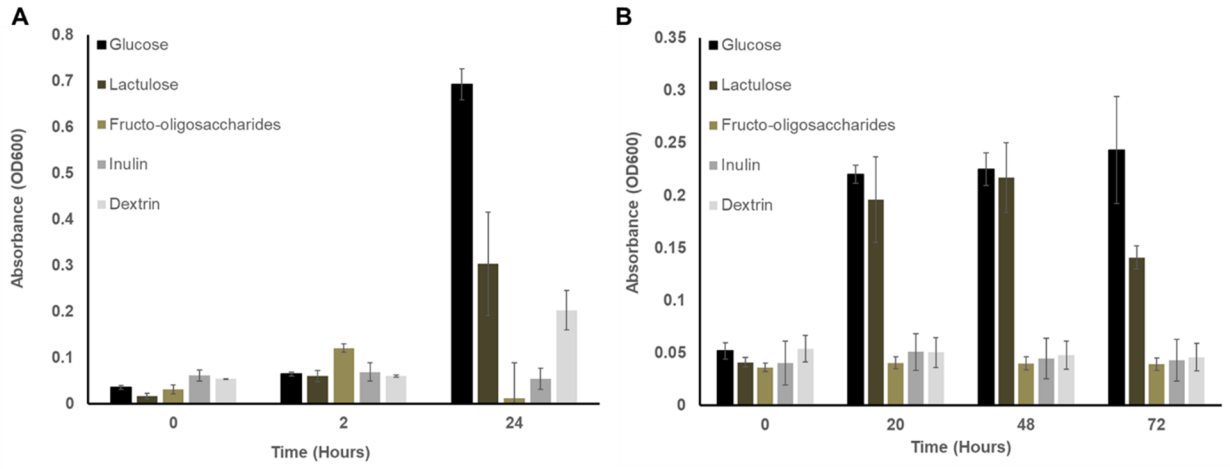


Figure S4. Growth of wild type EcN on mono- and polysaccharides. Related to Figures 1-3.

In vitro growth of EcN on minimal media supplemented with the monosaccharides glucose or lactulose, or with various polysaccharides under aerobic **(A)** and anaerobic **(B)** conditions. Shown are means of 3 replicates. Error bars are one standard deviation above and below the mean.

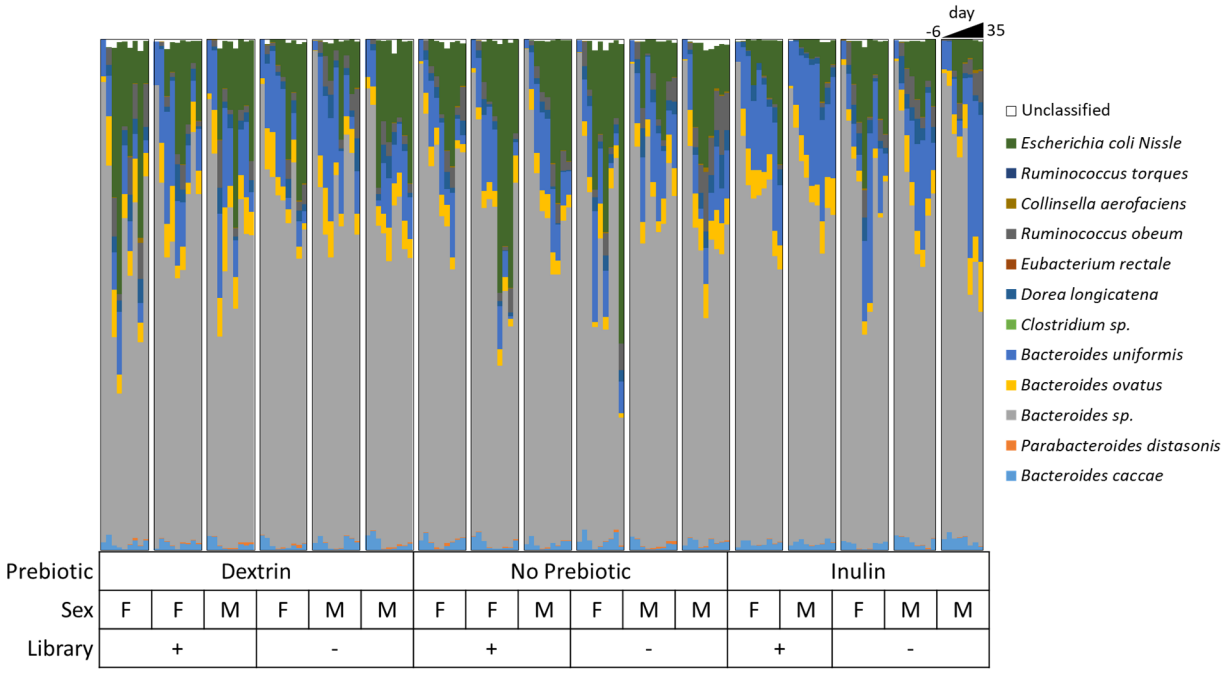


Figure S5. Relative taxonomic abundances from gnotobiotic mice humanized with a defined 13-member community. Related to Figure 3. gDNA from longitudinal fecal samples was analyzed by 16S rRNA sequencing. Mice were fed a Human Diet, and in some arms were supplemented with prebiotic in the drinking water. One week after humanization with the 13-member community, mice were gavaged with EcN containing either empty pZE21 vector (- Library) or the functional metagenomic library (+ Library).

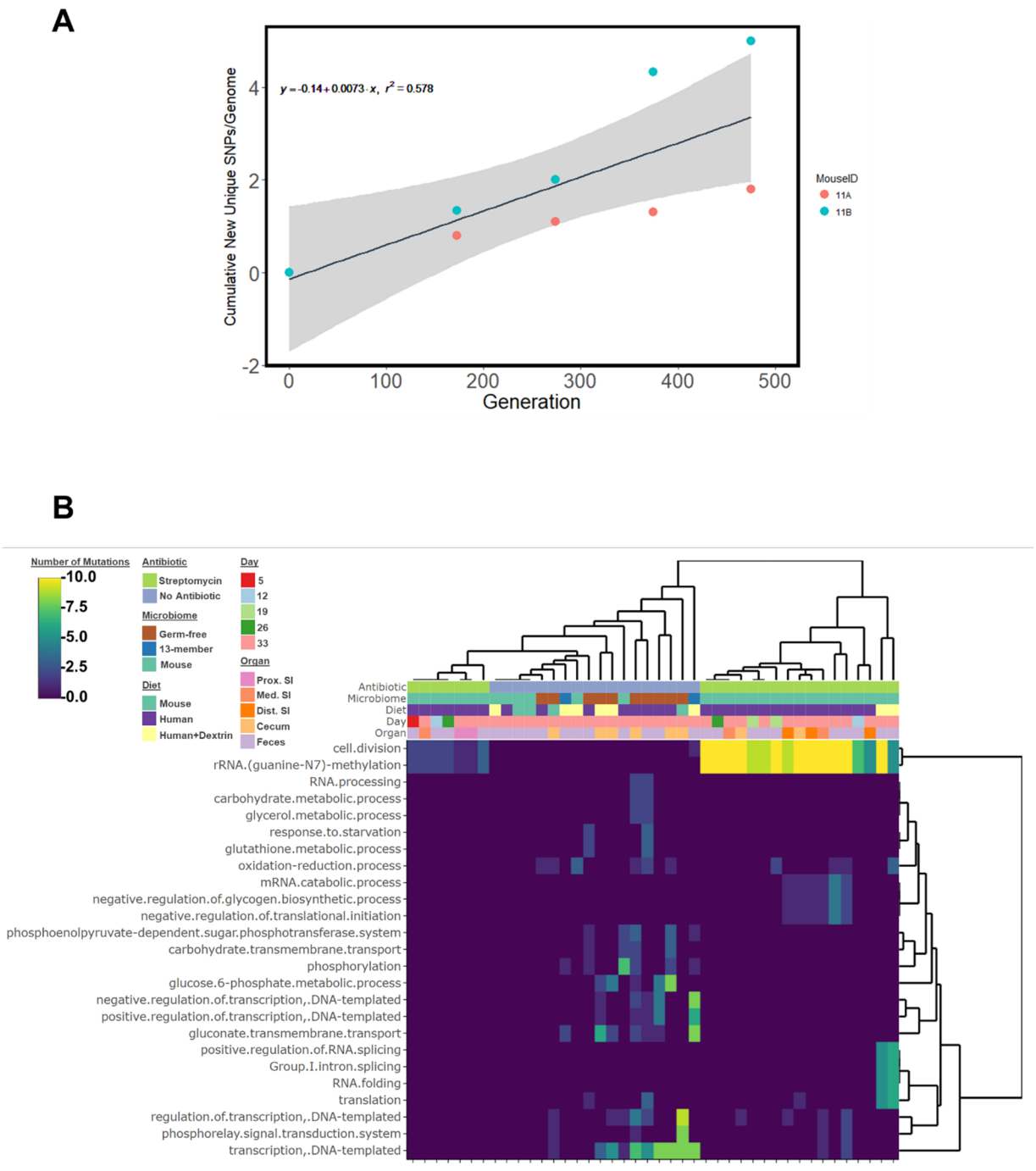


Figure S6. Estimation of *in vivo* mutation rate and comparison of functions mutated in EcN after passage through mice with a range of gut microbiome complexities and fed different diets. Related to Figure 4. A) Regression of SNP accumulation on number of bacterial generations. New unique SNPs detected in EcN isolates from fecal samples collected weeks 2-5

after gavage from two streptomycin treated mice are plotted against number of generations, where the *in vivo* generation time is assumed to be 100 minutes (Rang et al., 1999), and the number of SNPs at time 0 is assumed to be 0. 10 isolates per mouse per timepoint were whole-genome sequenced for SNP detection. The shaded area represents the 95% confidence interval of a linear regression of new unique SNPs/genome on generation. The estimated mutation rate of 0.007 (± 0.002) mutations per genome per generation is on the same order of magnitude as the previously published rate of 0.001 new SNPs per genome per generation (Lee et al., 2012). **B)** Heatmap of the cumulative number of nonsynonymous SNPs detected in genes grouped by their GO terms. Inclusion criteria for the heatmap was any function with at least two cumulative mutations identified across isolates from at least two mice. Dendrograms represent the clustering applied to group similar rows/columns of the heatmap. With 168 mutations, the GO pathway for Cell Division was the most commonly-mutated, but it should be noted that this GO pathway was almost entirely (165/168 mutations) composed of the ribosomal genes mutated during streptomycin treatment. The next most abundant pathways were transcription (57 mutations), regulation of transcription (23 genes), and gluconate transmembrane transport (22 mutations).

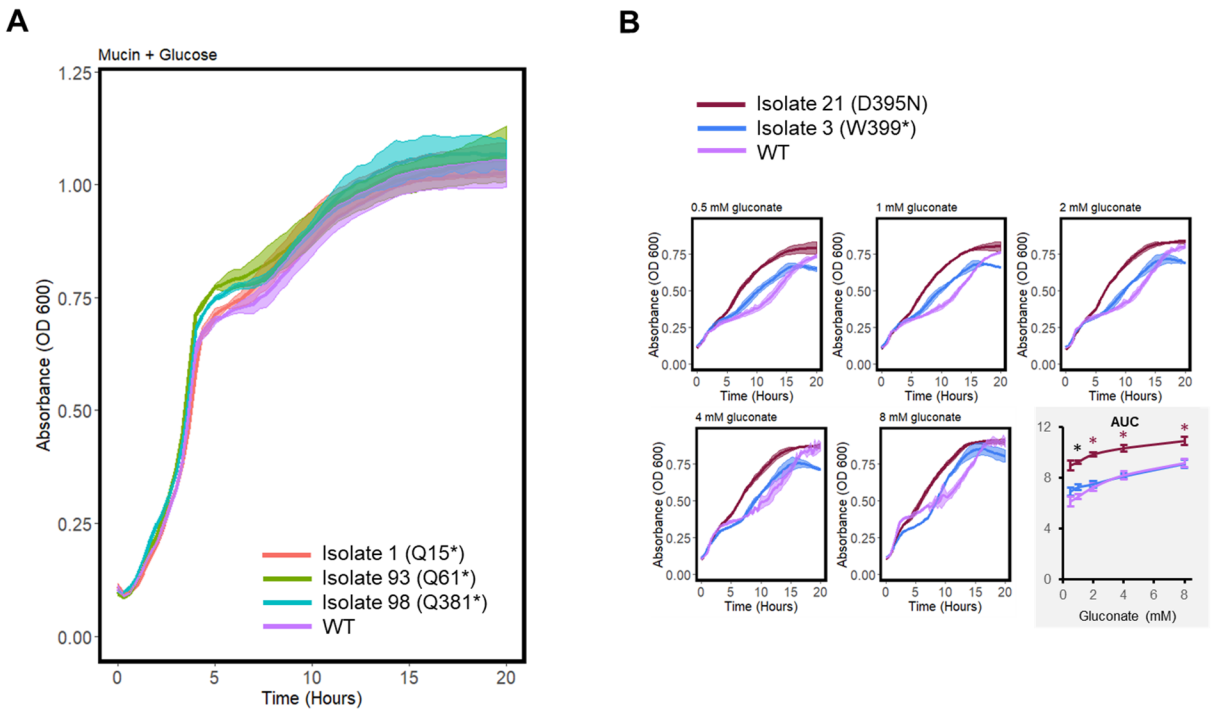


Figure S7. Growth of *NagC* and *gntT* mutants on mucin-containing minimal medias. Related to Figure 5. A) Diauxic growth curves of *NagC* mutants on glucose and porcine gastric mucin minimal media. **B)** Growth curves of gntT mutants on porcine gastric mucin minimal media supplemented with either 0.5, 1, 2, 4, or 8 mM gluconate. The gray panel summarizes the growth via the areas under the curves (AUC) for each gluconate concentration. N = 3, * $P < 0.05$, Welch's t-tests, Benjamini-Hochberg correction for multiple comparisons. Black stars indicate that both Isolate 3 and Isolate 21 had a significantly greater AUC at a given gluconate concentration; red – only Isolate 21. Areas under the curves were calculated using the growthcurver R package.

Supplementary Tables

Table S1. Comparison of nutrient compositions of the standard Mouse Chow diet (Laboratory Rodent Diet 5001*, LabDiet) and the High-fat ('Human') Diet (TD.88137, Teklad). Related to Table 1.

	Mouse Chow		High-fat ('Human') Diet	
	% By Weight	% kcal from	% By Weight	% kcal from
Protein	24.1	28.7	17.3	15.2
Fat	11.4	13.4	21.2	42
Carbohydrates	48.7	57.9	48.5	42.7
kcal/gm	2.89		4.50	

Table S2. Comparison of formulaic compositions of the standard Mouse Chow diet (Laboratory Rodent Diet 5001*, LabDiet) and the High-fat ('Human') Diet (TD.88137, Teklad). Related to Table 1. Ingredients are listed in order of most to least abundant by weight.

Mouse Chow	**High-fat ('Human') Diet**
Ground Corn	Sucrose
Dehulled Soybean Meal	Anhydrous Milk Fat
Dried Plain Beet Pulp	Casein
Fish Meal	Corn Starch
Ground Oats	Cellulose
Dehydrated Alfalfa Meal	Mineral Mix
Brewers Dried Yeast	Vitamin Mix
Cane Molasses	Calcium Carbonate
Wheat Germ	DL-Methionine
Dried Whey	Cholesterol
Porcine Animal Fat	Ethoxyquin
Porcine Meat and Bone Meal	
Wheat Middlings	
Salt	

Calcium Carbonate	
Vitamin Mix	
Mineral Mix	

80

81 **Table S3. Summary of *in vivo* adapted EcN isolates. Related to Figure 4.**

Microbiome	pZE21 Plasmid Insert	Diet	Organ	Day	# Isolates	# Mice
None	Metagenomic	Human	Cecum	35	20	2
None	Metagenomic	Human	Cecum	35	20	2
None	Metagenomic	Human+Dextrin	Cecum	35	20	2
None	Metagenomic	Mouse	Cecum	21	20	2
None	No insert	Human	Feces	35	20	2
13-member	No insert	Human+Dextrin	Feces	35	20	2
Mouse+strep	No insert	Human	Prox. SI	35	2	2
Mouse+strep	No insert	Human	Med. SI	35	22	3
Mouse+strep	No insert	Human	Dist. SI	35	25	3
Mouse+strep	No insert	Human	Cecum	35	20	2
Mouse+strep	No insert	Human	Colon	35	30	3
Mouse+strep	No insert	Human	Feces	5	5	1
Mouse+strep	No insert	Human	Feces	12	20	2
Mouse+strep	No insert	Human	Feces	19	20	2
Mouse+strep	No insert	Human	Feces	26	20	2
Mouse+strep	No insert	Human	Feces	33	20	2
Mouse+strep	No insert	Human+Dextrin	Feces	35	20	2
Mouse	No insert	Human	Feces	35	20	2
Mouse	No insert	Human+Dextrin	Feces	35	20	2
PKU Mouse	No insert	Mouse	Feces	35	17	2
PKU Mouse	PAL2	Mouse	Feces	35	20	2

82

83 **Table S4. Taxonomic composition of the 13-member defined bacterial community (Goodman**

84 **et al., 2011). Related to Table 1.**

Species	ATCC/DSMZ
Bacteroides caccae	ATCC 43185T

Parabacteroides distasonis	ATCC 8503
Bacteroides novel WH2	A. Salyers strain collection
Bacteroides ovatus	ATCC 8483T
Bacteroides uniformis	ATCC 8492
Bacteroides vulgatus	ATCC 8482
Clostridium scindens	ATCC 35704
Clostridium symbiosum	ATCC 14940
Dorea longicatena	DSM 13814
Eubacterium rectale	ATCC 33656
Ruminococcus obeum	ATCC 29174
Collinsella aerofaciens	ATCC 25986
Ruminococcus torques	ATCC 27756

Table S5: Putative phage regions in the EcN genome. Related to Figure 4.

REGION_LENGTH	SPECIFIC_KEYWORD	REGION_POSITION
18.8Kb	transposase,terminase	276890-295730
45.5Kb	transposase,lysis,head,capsid,tail	1325775-1371307
59Kb	tail,head,portal,terminase,lysin,integrase	2011601-2070656
12.5Kb	integrase,capsid,head	2535374-2547895
38.8Kb	integrase,transposase,tail	3318001-3356837

Table S6: Primers. Related to Figures 1-3 and Figure 4.

Primer #	Primer Name	Primer Sequence
1	pZE21_79_102For	CCGAATTCATTAAAGAGGAGAAAG
2	pZE21_80_103For	CGAATTCATTAAAGAGGAGAAAGG
3	pZE21_81_106For	GAATTCATTAAAGAGGAGAAAGGTAC
4	pZE21_126_149rc Rev	GATATCAAGCTTATCGATACCGTC
5	pZE21_128_150rc Rev	CGATATCAAGCTTATCGATACCG
6	pZE21_129_151rc Rev	TCGATATCAAGCTTATCGATACC

7	Barcode1	/5Phos/ANNNNNNNGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
8	Barcode2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNN*T
9	IlluminaPEAmpF	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
10	IlluminaPEAmpR	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT

89

90 **Table S7: Mutations in pZE21. Related to Figure 4**

Position	Type of mutation	Mutation	# isolates	# mice	# cages
74	intergenic (in ori)	G→T	1	1	1
295	intergenic (between kanR and ori)	G→T	1	1	1
313	intergenic (between kanR and ori)	G→T	3	2	1
386	intergenic (between kanR and ori)	G→T	1	1	1
480	intergenic (between kanR and ori)	C→A	1	1	1
532	intragenic (in kanR)	P249P (CCC→CCA)	7	3	2
648	intragenic (in kanR)	R211R (CGG→AGG)	1	1	1
742	intragenic (in kanR)	P179P (CCC→CCA)	3	1	1
802	intragenic (in kanR)	D159E (GAC→GAA)	1	1	1
805	intragenic (in kanR)	L158L (CTG→CTT)	1	1	1
890	intragenic (in kanR)	T130N (ACC→AAC)	3	2	1
963	intragenic (in kanR)	H106N (CAC→AAC)	1	1	1
1,019	intragenic (in kanR)	A87E (GCG→GAG)	1	1	1
1,085	intragenic (in kanR)	A65V (GCG→GTG)	1	1	1
1,372	intergenic (upstream of promoter)	G→T	1	1	1
1,440	intergenic (upstream of promoter)	T→C	1	1	1
1,482	intergenic (upstream of promoter)	G→T	1	1	1
1,558	intergenic (in promoter)	G→T	1	1	1
1,681	intergenic (in terminator)	G→T	1	1	1
1,717	intergenic (in terminator)	G→T	1	1	1

91 The kanR mutations could have been accumulated during growth in selective media prior to
92 delivery. As all kanamycin was removed from the gavage bolus by 3 washes in PBS, no antibiotic
93 selection was present during growth *in vivo*. Further, these isolates were from mice never treated
94 with streptomycin, eliminating cross-resistance as a possible explanation. No inactivating
95 mutations in kanR were observed, potentially limited by the fact that kanamycin was used to
96 recover EcN from fecal and intestinal contents.