Title: Adaptive strategies of the candidate probiotic *E. coli* Nissle in the mammalian gut

Authors: Nathan Crook[1,2], Aura Ferreiro[1,3,4], Andrew J. Gasparrini[3], Mitchell Pesesky[3], Molly K. Gibson[3], Bin Wang[3], Xiaoqing Sun[3], Zevin Condiotte[3], Stephen Dobrowolski[5], Daniel Peterson[6], Gautam Dantas[3,4,7-9]

[1]These authors contributed equally to this work

[2]Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, North Carolina, USA.

[3]The Edison Family Center for Genome Sciences & Systems Biology, Washington University School of Medicine, St. Louis, Missouri 63110, USA.

[4]Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, USA.

[5]Department of Pathology, Children's Hospital of Pittsburgh, 4401 Penn Avenue, Pittsburgh, Pennsylvania 15224, USA.

[6]Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.

[7]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110, USA.

[8]Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, Missouri 63110, USA.

[9]To whom correspondence should be addressed: Gautam Dantas, Ph.D. (dantas@wustl.edu)

Correspondence during review: Nathan Crook, Ph.D. (nccrook@ncsu.edu)

Summary

Probiotics are living microorganisms that are increasingly used as gastrointestinal therapeutics by virtue of their innate or engineered genetic function. Unlike abiotic therapeutics, probiotics can replicate in their intended site, subjecting their genomes and therapeutic properties to natural selection. By exposing the candidate probiotic *E. coli* Nissle (EcN) to the mouse gastrointestinal tract over several weeks, we uncovered the consequences of gut transit, inter-species competition, antibiotic pressure, and engineered genetic function on the processes under selective pressure during both within-genome and horizontal evolutionary modes. We then show the utility of EcN as a chassis for engineered function by achieving the highest reported reduction in serum phenylalanine levels in a mouse model of phenylketonuria using an engineered probiotic. Collectively, we demonstrate a generalizable pipeline which can be applied to other probiotic strains to better understand their safety and engineering potential.

Keywords

Probiotic; *E. coli*; Microbiome; Evolution; Phenylketonuria; Engineering

Introduction

The human gut microbiota is the collection of microbes which have evolved to live in the human gastrointestinal tract (Huttenhower et al., 2012). The microbiota's metagenome complements the human genome by assisting in, among other functions, metabolism of dietary fiber (El Kaoutari et al., 2013), synthesis of vitamins (Backhed et al., 2005), and exclusion of pathogens (Theriot et al., 2014). Chemical and/or microbial perturbations to this community can result in dysbiotic community compositions, defined by their negative impact on human health (Sampson et al., 2016; Smith et al., 2013; Tamboli et al., 2004; Turnbaugh et al., 2006). Correction of specific dysbioses through administration of probiotics has shown therapeutic promise (Sartor, 2005). Probiotics are live microorganisms that, when consumed in sufficient quantities, provide health benefits to the host, including enhancement of epithelial barrier function, blocking of pathogen binding, and vitamin synthesis (Hill et al., 2014). To improve efficacy and expand therapeutic scope, probiotics can be genetically engineered, with recent work demonstrating preclinical efficacy against infectious and metabolic diseases (Durrer et al., 2017; Hwang et al., 2017; Mansour and Abdelaziz, 2016; Palmer et al., 2018). Unlike conventional small-molecule (abiotic) therapeutics, wild type or engineered probiotics replicate in the gut and are therefore subject to genomic mutation and natural selection, potentially to the detriment of their intended therapeutic effect and safety profile. Indeed, both commensal and pathogenic microbes have been shown to adapt in a host-specific manner during gut passage (Lieberman et al., 2014; Zhao et al., 2018). Clinical use of probiotics therefore demands a thorough understanding of their *in vivo* evolutionary trajectories. In this study, we focus on the *in vivo* evolutionary trajectories of the candidate probiotic *Escherichia coli* Nissle 1917 (EcN), an *E. coli* clade B2 probiotic [Figure S2]. EcN has a long history of probiotic usage without evolution of pathogenic phenotypes (Wassenaar et al., 2015; Westendorf et al., 2005), has

demonstrated efficacy against inflammatory bowel disease (Scaldaferri et al., 2016), and has been gaining increased attention as a “chassis” for engineered therapeutic function (Hwang et al., 2017; Palmer et al., 2018).

One example of a disorder which could be treated with engineered probiotics is phenylalanine-hydroxylase-deficient phenylketonuria (PKU), an inborn error of host metabolism. Individuals with phenylketonuria lack some or all function of the phenylalanine hydroxylase enzyme (PAH), leading to accumulation of neurotoxic phenylalanine (Phe) levels which, if left untreated, impair neurocognitive development and lead to severe intellectual disability (Nardecchia et al., 2015; Regier and Greene, 2000). The primary therapy for PKU is dietary restriction through use of low Phe and low protein medical foods to achieve serum Phe concentrations in the range of 120-360 µmol/L (Vockley et al., 2014). However, the efficacy of dietary restriction is substantially hindered by poor compliance (MacDonald et al., 2010; MacLeod and Ney, 2010). The lone FDA approved therapeutic alternative is sapropterin, but this is effective only for a subset of PKU patients (Vockley et al., 2014). While an enzyme replacement therapy (Pegvaliase) (Eavri and Lorberboum-Galski, 2007) has undergone promising Phase III clinical trials (Charlesworth, 2016), this option is expensive and requires continual administration. Recent work establishing efficacy of an engineered *Lactobacillus* strain in a murine model of PKU demonstrates that engineered probiotics could innovatively address shortcomings in current therapies (Durrer et al., 2017). However, the level of reduction of murine serum Phe concentrations achieved in this *Lactobacillus* study was substantially lower than those reported with the injected enzyme therapy (Sarkissian et al., 2008), motivating further optimization of engineered probiotics for PKU. We hypothesized that EcN, with established genetic tools enabling high-output enzyme production, would provide

an alternative and potentially superior host for therapeutic degradation of gut Phe through regulated expression of a heterologous phenylalanine-ammonia lyase (PAL). PAL converts Phe into ammonia and trans-cinnamic acid, both of which are nontoxic and readily excreted (Sarkissian and Gamez, 2005). Additionally, we hypothesized that, due to their differing serum Phe levels, the optimal probiotic PAL expression level would differ between male and female PKU mice.

Prior work has uncovered the unique *in vivo* evolutionary trajectories (relative to *in vitro*) of commensal, laboratory, and pathogenic strains of *E. coli* in the mammalian gut under specific dietary conditions and gut microbiome architectures (Barroso-Batista et al., 2014; Ghalayini et al., 2018; Gumpert et al., 2017; Karami et al., 2007; Lescat et al., 2017; Lourenco et al., 2016). Nutrient availability and inter-species metabolic dependencies are major selective pressures for the gut microbiome (Hoek and Merks, 2017; Rakoff-Nahoum et al., 2016; Zhao et al., 2018). A study of a large human cohort (N > 1000) demonstrated that host diet dominates host genetics as a driver of the composition and function of the gut microbiota (Rothschild et al., 2018). It is also well-known that engineered metabolism can dramatically compromise microbial growth (Fletcher et al., 2016; Lechner et al., 2016), and in general may alter the primary selective pressures relative to wild type. We therefore set out to investigate the effects of host diet, microbiome complexity, and engineered metabolism on the *in vivo* evolutionary trajectories of wild type and engineered EcN, using mice as a model of the mammalian gut (Reyes et al., 2013; Ridaura et al., 2013) and heterologous expression of PAL as a model of engineered probiotic metabolism.

Probiotics face a variety of selective pressures in the mammalian gut that may lead to *in vivo* evolution. First, EcN cannot natively degrade many of the complex polysaccharides present in the

intestine (Fabich et al., 2008; Hoskins et al., 1985), and is dependent on other species for production of metabolizable di- and mono-saccharides (Conway and Cohen, 2015), which may lead to adaptations in carbohydrate utilization. Second, the high prevalence of antagonistic interspecies interactions in the gut (Wexler et al., 2016) may drive adaptation in EcN's encoded colicins (Patzer et al., 2003). Third, since EcN normally resides in sessile multi-species biofilms on the outer mucin layer (Adediran et al., 2014; Conway and Cohen, 2015), introduction to the murine gut may drive adaptation in adhesins, increasing their specificity to mouse mucins. Finally, expression of a Phe-degrading enzyme may impose a metabolic burden. For these reasons, we hypothesized that the primary genomic adaptations (mutations) which EcN would undergo in the mouse gut could involve i) carbohydrate utilization, ii) inhibitory interactions with other microbes, iii) motility/adhesion, or iv) inactivation of engineered function.

For gut commensals, genetic adaptation can occur through genomic SNPs or indels, as well as horizontal gene transfer (evolutionary scales 1 and 2, respectively, as described in (Ferreiro et al., 2018)). Horizontal gene transfer is a major mechanism driving bacterial evolution (Ferreiro et al., 2018; Rodriguez-Beltran et al., 2018; Soucy et al., 2015), occurring more rarely than single nucleotide polymorphisms (SNPs) and insert/deletion events but accounting for greater net genetic distance on timescales shorter than bacterial speciation, and frequently associated with key functional determinants such as antibiotic resistance (Smillie et al., 2011). Bacterial strains differ in their capacity to engage in HGT, due to origin of replication incompatibilities and lack of natural competence (Thomas and Nielsen, 2005). In this study, we describe the fitness and adaptation of EcN in the murine gut using both whole-genome sequencing of *in vivo* adapted isolates (to recover changes in the EcN genome) and high-throughput functional metagenomic selections (to simulate

horizontal gene transfer). In each case, replicate mice were fed diverse diets including a standard mouse diet, a high-fat Western diet [Tables S1-S2], and diets supplemented with prebiotics. Further, we included mice with a spectrum of background microbiota complexities, ranging from germ-free mice mono-colonized with EcN, germ-free mice colonized with synthetic biota of intermediate complexity, and conventional mice with antibiotic-perturbed or unperturbed complex microbiota [Figure S1]. Using this reductionist approach, we elucidated the impact of host diet and microbial competition on the *in vivo* adaptation of EcN. In total we analyzed 225 longitudinal samples from functional metagenomic selections in 33 mice [Table 1], and 401 EcN isolates from 20 mice [Table S3], comprising the most comprehensive study of *in vivo* probiotic evolution to date. Finally, using engineered probiotic EcN, we achieve the highest reported reduction in murine serum Phe levels from an engineered probiotic, and describe genomic changes occurring *in vivo* during course of treatment.

Results

Interrogation of HGT-mobilizable functions

To identify functions that may be acquired for improved EcN fitness in the mammalian gut, we carried out functional metagenomic selections where metagenomic insert libraries heterologously expressed in EcN were passaged through mouse gastrointestinal tracts for up to five weeks (Gibson et al., 2014). By creating a functional metagenomic library, we were able to bypass the normal limitations and timescale of HGT and ask: if EcN could receive any genetically encoded function from any member of a healthy gut microbiome independent of mode of HGT, what functions would provide the greatest fitness advantage? By competing millions of recombinant EcN strains against each other *in vivo*, we expected to observe significant enrichment of a subpopulation of

strains with DNA inserts conferring EcN a fitness advantage. These functional metagenomic selections served as an upper bound proxy for the potential impact of horizontal gene transfer on the fitness of EcN in the gut. The source of our metagenomic DNA was stool from two healthy human infants and six healthy human mothers, to approximate the diversity of microbial taxa present in the gut through different life stages (Moore et al., 2015). Approximately 10^6 unique 2-5kb fragments of metagenomic DNA were cloned into EcN, generating a library $>10^9$ bp in size, which is equivalent to hundreds of bacterial genomes.

To more easily parse the selective pressures placed on this functional metagenomic library, we performed *in vivo* selections using a reductionist approach. That is, we began with the *in vivo* setting with the fewest selective factors: the germ-free gut of mice fed only standard mouse chow and water. Subsequently, we carried out functional selections in mice pre-colonized with a defined community of 13 representative human gut bacterial species (Goodman et al., 2011), conventional mice that had had their microbiota depleted through streptomycin treatment, and untreated conventional mice. We also carried out selections in mice fed either standard mouse chow, a high-fat Western diet, or a high-fat diet supplemented with either dextrin or inulin, two prebiotic polysaccharides. We specifically hypothesized that in EcN library mono-colonized mice, we would select for enzymes capable for degrading complex polysaccharides, a function that *E. coli* typically depends on gut anaerobes to carry out. Conversely, we hypothesized that co-colonization with a defined bacterial community would relieve this selective pressure, which would be reflected in a greater diversity of enriched DNA inserts.

Selection for carbohydrate utilization in the germ-free gut

Metagenomic libraries were delivered to six germ-free C57BL/6 mice via oral gavage. These mice consumed a standard mouse diet (MD) composed mainly of plant-based carbohydrates and protein [Tables S1-S2]. Over 5 weeks of colonization, we observed a striking reduction in the diversity of metagenomic inserts in mouse feces. In all 6 mice, the initial ~10^6 metagenomic fragments were reduced to the same 5 metagenomic inserts. [Figure 1b]. Supporting our hypothesis that in the absence of microbial taxa capable of degrading complex polysaccharides, the main selective pressure on EcN in the gut is polysaccharide degradation, four out of the five inserts contained a predicted glycosyl hydrolase family 32 (GH32) enzyme. GH32 is a large protein family involved in the hydrolysis of glycosidic bonds within polysaccharides. Several of these inserts also contained predicted transporters and other enzymes involved in central carbon metabolism [Figure 1g]. To experimentally validate our hypothesis that these *in vivo*-enriched inserts directly enhance the metabolic activity of EcN, we cloned these metagenomic fragments into wild type EcN and tested additional carbon source utilization conferred by these fragments using *in vitro* phenotypic microarrays [Supplementary File 1].

GH32 is reported to be involved in sucrose utilization. Commensal *E. coli*, including EcN, are unable to consume sucrose (Conway and Cohen, 2015), and each GH32-containing fragment enabled growth on this substrate [Figure 1f]. Similarly, several GH32-containing fragments enabled growth on longer-chain oligosaccharides containing sucrose as a terminal moiety, including raffinose and stachyose [Figure 1f]. The growth rate of EcN strains containing 2 different GH32-containing fragments (MD02 and MD05) was higher on these longer-chain oligosaccharides than sucrose, indicating likely specialization of these enzymes toward these molecules. The one fragment we isolated from this experiment which did not contain GH32

(fragment MD04) instead contained a transcription factor which was identical (100% nucleotide identity) to the transcriptional regulator gadX in *E. coli* Nissle. This protein plays a major role in regulating acid tolerance by regulating the expression of the *gad* operon, which consumes intracellular H+ during glutamate metabolism (Tramonti et al., 2002; Tucker et al., 2003). It has been previously shown that cytosolic pH regulates glycolytic flux as mediated by vacuolar ATPases - proton pumps that sense intracellular pH and cease activation of the cAMP-dependent protein kinase A under acidic conditions (Dechant et al., 2010). Supporting this hypothesized role, *in vitro* recombinant expression of this fragment in EcN significantly reduced lag phase when switching from LB to galactose-containing minimal media [Figure 1f].

Our observations regarding the similarity of recovered fragment functions were confirmed after assigning Gene Ontology (GO) annotations to these 5 fragments. In agreement with our observations, >87% of functionally selected fragments were annotated as belonging to the GO pathway "Carbohydrate Metabolic Process." The next most abundant GO pathways were "Phosphorylation" followed by "Carbohydrate Transmembrane Transport" and "Transcription, DNA-templated". Together, these results indicated that the dominant selective pressure on EcN in the germ-free mouse gut is for enzymatic degradation and utilization of long-chain polysaccharides.

Diet-dependent selection

We hypothesized that because selective pressure for long-chain polysaccharide degradation was so strong in the EcN mono-colonized mouse gut, we would observe polysaccharide-specific enrichment of DNA inserts in response to the presence to prebiotics. To test this hypothesis, we

gavaged EcN containing the same metagenomic library to 3 groups of 5 germ-free C57BL/6 mice. All 15 mice were maintained on a high-fat, high-carbohydrate diet which mimics the diet of Western humans (HD) (Plump et al., 1992). One group of mice was maintained on the Western diet alone, while the other two were supplemented with the prebiotics inulin or dextrin in the drinking water. Inulin is a fructose polysaccharide with chain-terminating glucosyl moieties and is not degradable by wild type EcN, while dextrin is a glucose polysaccharide that can support limited EcN growth, e.g. 20-30% of the cell density that glucose can support under aerobic conditions (Fabich et al., 2008), [Figure S3]. Over 5 weeks, we again observed strong convergence toward ≤6 metagenomic fragments in each dietary condition [Figure 1c-1e]. Supporting our diet-dependency hypothesis, the collection of enriched fragments in each condition was different. However, one "generalist" metagenomic fragment was enriched in all three groups of mice (fragment HD04). This fragment was unique among the enriched fragments in that it contained a GH32 enzyme, again validating our hypothesis that a requirement for growth during mono-colonization is the ability to break down polysaccharides that wild type EcN alone cannot degrade. Indeed, *in vitro* minimal media growth assays with a freshly transformed copy of this fragment enabled EcN consumption of sucrose [Figure 1f]. We further experimentally verified that of the fragments recovered in this experiment, this GH32-containing fragment uniquely enabled increased growth in media composed of the water-soluble extract of the Western diet, likely due to the high amount of sucrose (341 g/kg) in this formulation [Figure 1f].

To gain insight into the broad functional classes providing a selective advantage in each dietary condition, ORFs in each enriched fragment were classified into GO Pathways [Figure 2a]. In agreement with our qualitative observation of different fragments enriched in each condition, we

also observed strong clustering of fragment function based on prebiotic supplementation. In all conditions, reflecting the "generalist" fragment that we recovered, "Carbohydrate Metabolic Process" was abundant (>23% relative abundance in all samples). In the presence of dextrin, we observed a unique additional signal for "Pyrimidine Nucleotide Biosynthetic Process", reflecting a predicted glutamate synthase that reached high abundance. It has been previously demonstrated that *E. coli* has two primary pathways for glutamate synthesis, the glutamate synthase pathway and the glutamate dehydrogenase pathway, the latter of which is preferred during glucose-limited growth (Helling, 1998). As the dextrin monomer glucose is expected to be abundant in the dextrin condition, presumably due to the activity of the generalist GH32 enzyme, it was unsurprising that the glutamate synthase, and not the glutamate dehydrogenase pathway was enriched. In the presence of inulin, the major fragment (reaching 32-67% relative abundance in all mice at the end of the experiment) contained an ABC transporter that was classified as an "ATPase-coupled sulfate transmembrane transporter". In the absence of prebiotic, the dominant fragment was the generalist glycosyl-hydrolase-containing fragment. Alone, no "specialist" fragment enabled increased *in vitro* growth on dextrin or inulin, or on Western diet extract broth. We conclude that in these EcN mono-colonized mice the dominant selective pressure remained the ability to enzymatically degrade long-chain polysaccharides, but that once this necessary function was fulfilled by a subpopulation of GH32-carrying EcN, as it would be by other taxa in a conventional gut community, secondary functions had room for enrichment in a host diet-specific manner.

Advantageous functions in a simplified microbial community.

We then tested our hypotheses that i) in the presence of other taxa capable of degrading long-chain polysaccharides, this particular selective pressure on EcN would be relieved, and ii) with

increasing complexity of the microbial gut community, EcN would be subject to selective pressures arising from direct competition and interactions with other microbes. Accordingly, 8 germ-free C57BL/6 mice were first colonized with a 13-member community of human gut-derived bacteria which broadly mimic the phylogenetic makeup of the human microbiome (Goodman et al., 2011), [Table S4]. After allowing this community to stabilize, the EcN metagenomic library was delivered via oral gavage and allowed to colonize for 5 weeks. Sequencing the metagenomic DNA contained in EcN after 5 weeks in the mouse gut revealed a highly diverse pool of fragments, relative to the mono-colonization experiments, supporting our hypothesis that in the presence of taxa expressing hydrolases that degrade complex polysaccharides, thus liberating di- and mono-saccharides for utilization by EcN, the selective pressure for acquisition of this family of hydrolases is relieved.

While no fragment reached a high abundance (>20% relative abundance) in all mice in the same experimental treatment, we did observe some functional convergence. The GO term "Carbohydrate Metabolic Process" was still consistently recovered across the majority of experimental replicates. One of the strongest signals we observed for condition-specific enrichment of a particular pathway was "Gluconate Metabolic Process", with a transcriptional regulator, kinase, and transporter related to gluconate achieving >60% abundance in mice containing the 13-member defined community [Figure 2a]. We hypothesized that the presence of this pathway would directly alter the ability of EcN to metabolize gluconate, and tested this in *vivo*. When these fragments were cloned on a plasmid into wild type EcN, the recombinant strain was no longer able to grow on minimal media containing gluconate as the sole carbon source [Figure 2b], indicating constitutive activity of the gntR repressor present in these fragments.

Gluconate utilization has been previously shown to be necessary for *E. coli* to colonize the mouse gut (Chang et al., 2004; Sweeney et al., 1996), so the consistent recovery of fragments reducing or abolishing gluconate utilization was puzzling. These past studies used conventional or streptomycin-treated mice hosting diverse mouse flora. In this study, it is possible that a different species in the 13-member human commensal consortium preferentially utilizes gluconate as a carbon source, making expression of the EcN gluconate utilization pathway metabolically wasteful, and providing a selective pressure for EcN to effectively downregulate the gluconate pathway by upregulating its repressor.

Together, these data indicate that the primary selective pressures imposed on EcN in the mammalian gut are related to carbohydrate utilization. Regardless of diet, there was a strong enrichment over the course of the functional metagenomic selections for glycosyl hydrolases that can degrade complex polysaccharides present in the diet, a function that in conventional gut microbiomes is provided by other taxa in the community. Once these taxa were reintroduced this selective pressure was relieved, but even in this case the most enriched functions were involved in carbohydrate utilization.

Within-genome evolutionary change

To more thoroughly test our hypothesis that the main drivers of EcN adaptation in the mouse gut include factors involved in i) carbohydrate utilization ii) direct inhibitory interactions with other microbes, or iii) motility or adhesion in response to the mouse-specific mucins, we performed whole-genome sequencing on 401 isolates collected from mice that were either mono-colonized with EcN, colonized with a defined community, streptomycin treated, or conventional, and that in

addition were either fed standard mouse chow, a Western diet, or a Western diet supplemented with dextrin [Tables S1-S2]. We also sequenced a reference EcN strain using both short and long read sequencing platforms, generating a high quality and manually curated closed assembly to which experimental isolates were compared. Among sequenced isolates, we observed a diverse array of genomic changes, with 456 total (336 unique) genomic changes distributed across 171 EcN genes [Figure 3]. The most highly-mutated strain had 5 SNPs, while the average strain had 0.86 SNPs/genome.

Adaptations involving nutrient utilization.

Competition for nutrients plays a critical role in microbial community dynamics and colonization resistance (Pereira and Berry, 2017). Therefore, mutations that influence nutrient utilization in EcN are likely to give it a strong selective advantage, relative to wild type. In support of our hypothesis, we found that several genes involved in nutrient utilization were extensively mutated. The gene *gntT* was found to be mutated in 6 different ways across 21 strains isolated from four different mice, including 2 isolates with >10kb deletions containing this gene [Figure 4a]. The 4 other kinds of mutations included 3 unique early stop codons, and a D395N mutation. *GntT* encodes a high-affinity gluconate transporter (Porco et al., 1997). We hypothesized that *gntT* mutants isolated from these *in vivo* experiments would have an impaired ability to utilize gluconate as a carbon source, and so we tested the ability of representative isolates harboring these mutations to grow *in vitro* in gluconate minimal media. Isolates with the early stop codons were unable to grow, and the isolate with the D395N mutation exhibited an increased lag phase compared to wild type as indicated by the significantly lower optical density of its culture at 8 hours [Figure 4b]. When comparing all isolates from mice spanning the axis of microbiome complexity (mono-

colonized to conventional), *gntT* mutants were significantly enriched in mice that had been pre-colonized with a 13-member defined community (p = 1.06E-6, hypergeometric test), although *gntT* mutants were also observed in EcN-monocolonized mice. Strikingly, this was the same phenotype recovered in the functional metagenomic selection in mice colonized with the defined community, where the strongest signal for functional enrichment was that of a gluconate utilization pathway whose phenotypic consequence was strongly inhibited growth on gluconate compared to wild type. This may be similarly due to preferential gluconate utilization by another member of the defined community balanced against the metabolic cost for EcN to maintain this pathway.

Additionally, we found 10 different mutations in *nagC* in 19 isolates [Figure 4c]. These isolates were from 4 different mice used in the functional metagenomic selection for dietary conditions, and exclusively housed one of two metagenomic inserts, containing either GH32 or *gadX* transcriptional regulator, which may simply reflect the fact that those were the most abundant strains in the EcN population after the functional selection. All mutations resulted in either a premature stop codon, elimination of the start codon, or a frameshift. NagC is a repressor of N-acetylglucosamine and chitobiose import and utilization and is an activator of N-acetylglucosamine synthesis in its absence (Konopka, 2012; Plumbridge and Pellegrini, 2004). When mutated in this way, the pathways for N-acetylglucosamine and chitobiose utilization are likely permanently activated. We hypothesized that this may enhance the ability of EcN to consume host mucins, which have high levels of N-acetylglucosamine (Tailford et al., 2015). To test this hypothesis, we carried out an *in vitro* diauxic growth experiment and found that *nagC* mutants were able to switch more rapidly to growth on mucin under glucose limiting conditions than wild type EcN [Figure 4d]. We also found that *nagC* mutants were able to grow better on N-

acetylglucosamine minimal media, and switch more rapidly to growth on N-acetylglucosamine under glucose limiting conditions [Figure 4d]. *NagC* mutants were exclusively isolated from germ-free mice that had been mono-colonized with EcN; such mutations enabling enhanced growth on a carbon source that is typically less preferred may be explained by the absence of taxa upon which EcN depends for polysaccharide degradation products. It has also been found that *nagC* is mutated during adaptation to osmotic stress (Winkler et al., 2014), which may provide EcN a selective advantage during transit through the GI tract.

Adaptations involved in stress response

Probiotics are exposed to stressful conditions as integral parts of their life cycle. They must survive transit through the stomach and then take up residence in the nutrient-limiting and highly competitive environment of the GI tract. Our probiotics were subjected to selective pressure to remain viable in feces to gain repeated access to the GI tract through coprophagy. This is supported by the fact that stress response pathways were a strong target for natural selection in our experiment.

In particular, the gene *sgrR* was mutated 4 different ways in 8 isolates from 3 mice [Figure 4g]. These mutations were all nonsynonymous and did not result in stop codon formation. SgrR is a transcription factor which activates the expression of *sgrS*. SgrS is a small RNA which in turn inhibits the glucose-specific PTS transporter (Vanderpool and Gottesman, 2004). Under conditions of excess glucose-phosphate, SgrR is necessary to protect the cell from stress (Vanderpool and Gottesman, 2007). When comparing against all isolates from mice spanning the axis of biota complexity, *sgrR* mutations were significantly enriched in mice that had been pre-colonized with

the 13-member defined community (p = 0.00732, hypergeometric test); however, because 6 of the isolates originated from one mouse that was in the defined community treatment group, this enrichment may be independent of the complexity of the microbiome. None of the mutations appear to be inactivating, and while a dN/dS analysis determined that there was not sufficient evidence to support positive selection, repeated nonsynonymous mutations in isolates from multiple, separately-caged mice suggests this stress response function is important for EcN fitness in the murine gut. Further, the *rssB* gene was mutated 4 different times in 9 different strains. Strikingly, all mutations to this gene were frameshift mutations. RssB regulates *rpoS* expression by directing it to the ClpX protease. RssB mutants display elevated RpoS levels (Battesti et al., 2011). RpoS is a master regulator of the stress response in *E. coli*, and *rpoS* mutants show enhanced resistance to stress (Fontaine et al., 2008).

Adaptation in membrane-associated functions

The gene *kfiB* was mutated in 24 strains from 7 mice, each with the same loss of an early stop codon at amino acid 366, resulting in a full-length protein of 562 amino acids, indicating that in our wild type strain of EcN *kfiB* is either inactivated or truncated [Figure 4e]. When comparing isolates from mice spanning the axis of microbiome complexity, this potential gain-of-function variant was significantly enriched in mice that were pre-colonized with a defined community (p=3.01E-15, hypergeometric test). KfiB is a member of the K5 capsule polysaccharide (heparosan) biosynthesis cluster, which has been implicated in mediating the interaction between EcN and epithelial cells (Hafez et al., 2010; Leroux and Priem, 2016; Rigg et al., 1998). The role of *kfiB* in this gene cluster is not fully elucidated, but it has been shown to stabilize the *kfiAC* glycosyltransferase complex and be involved in heparosan export (Leroux and Priem, 2016). We

hypothesized that our *kfiB* variants would exhibit altered adhesion to human epithelial cells compared to our wild type strain that harbors the early stop codon. To experimentally test this hypothesis, we carried out an *in vitro* adhesion assay with representative *kfiB* mutants and Caco-2 cell monolayers. As all *kfiB* variants had the same stop-loss mutation, 3 isolates were chosen as biological replicates. The results of this assay revealed that the loss of truncation in the kfiB mutants provided them significantly enhanced adhesion to Caco-2 cells [Figure 4f]. This loss of truncation may reflect re-adaptation of our reference EcN strain, which likely has been passaged often under laboratory conditions, to *in vivo* conditions and interactions with epithelial cells.

Antibiotic resistance mutations

In colonization experiments in streptomycin-treated mice, we observed that the majority of isolates (167/204) contained mutations in *rsmG*, a 16S rRNA methyltransferase. *rsmG* is responsible for methylating G527 in the 530 loop of the 16S rRNA. The 530 loop is a target for streptomycin binding, and it has been shown that mutation of key residues in *rsmG* to alanine cause low-level resistance to streptomycin (Benitez-Paez et al., 2012). In particular, the three positions we identified with mutations (53, 71, and 79), were also mutated in prior work to give low-level streptomycin resistance and are predicted to be involved in catalysis, s-adenosyl-L-methionine (the methyl group donor) binding, and SAM binding, respectively. Unsurprisingly, when comparing isolates from mice that were treated with streptomycin to all other isolates, mutants were significantly enriched in mice that had been given streptomycin (p=1E-87, hypergeometric test). Analysis of mutations to this gene revealed strong evidence for extensive strain sharing among mice in the same cage. In particular, *rsmG* mutations clustered based on the cage from which the isolate was recovered. Isolates with the same mutation at position 53 were exclusively recovered

from one cage (J11), while isolates with mutations at positions 71 and 79 were recovered from the other sampled cage (J1) [Figure 4h]. While the precise mechanism of microbiota transfer among humans is not known, this result indicates that if such a route is present, probiotics which evolved in one individual can be readily transferred to others. These results could indicate residual streptomycin present in the mouse gut, 1 day after delivery of the drug, and consequently that antibiotic treatment shortly before probiotic delivery may select for the presence of antibiotic-resistant probiotics.

We also observed 2 mutations to *rpsL* in 11 strains. All mutations occurred at residue 86 and converted it either to a cysteine or serine. These mutations only occurred in streptomycin-treated mice, and 10/11 times occurred in conjunction with a mutation in *rsmG*. Residue 86 is adjacent to residues that have been mutated previously to provide streptomycin resistance to EcN (Timms et al., 1992; Toivonen et al., 1999). From longitudinal sampling of 2 mice fed a Western diet and treated with streptomycin prior to EcN gavage, we calculate the mutation rate of EcN in the mouse gut to be 0.002 SNPs/genome/generation in the context of antibiotic treatment [Figure S4]. This rate is double that recently obtained for *E. coli* grown *in vitro* in the absence of antibiotic pressure (Lee et al., 2012). We expect that the combination of *in vivo* residence and antibiotic selection drove increased mutation rates in our experiment. Taken together, these findings are highly relevant considering the common practice of administering probiotics and fecal microbiota transplants after antibiotic administration (Hempel et al., 2012; Koenigsknecht and Young, 2013).

No in vivo *plasmid acquisition*

We searched through our genome assemblies for contigs that did not map to the EcN genome, as this would indicate horizontal gene transfer of novel genetic material. Interestingly, we did not

observe acquisition of new sequences over a 5-week timescale. The cause for this may be that EcN contains two endogenous plasmids (pMut1 and pMut2) (Grozdanov et al., 2004) in addition to the cloning vector that was used for this study. It is possible that these plasmids restrict incoming plasmids due to plasmid incompatibility groups or other defense mechanisms. In agreement with these results, *E. coli* Nissle has been shown to be a poor recipient of horizontally-transferred DNA, with a rate of DNA acceptance <10% the rate of *E. coli* K-12 for a variety of plasmid incompatibility groups (Sonnenborn and Schulze, 2009).

In vivo *excision of large genomic sections*

In total, we observed 8 deletions larger than 1kb in size in 7 isolates. Two of these large deletions (24.3kb and 15.5kb in size) encompassed *gntT*, as further support that inactivation of this gene is beneficial for colonization in the context of a defined community. While EcN contains 5 putative phage regions (Arndt et al., 2016) [Table S5], only one (11.5kb) was observed to be excised *in vivo*. A 5.7kb deletion included a fimbriae locus. 3 deletions (23.3kb, 25kb, and 93kb in size) contained inverted repeat sequences, transposase, or integrase sequences, providing a potential mechanism for their excision. For example, the 25kb deletion, encompassing *glpR* (a ribokinase involved in ribose metabolism), *ydcZ* (an inner membrane exporter), *php* (a phosphotriesterase), along with various hypothetical proteins, was flanked by an insertion sequence 66 C-terminal element on one end and multiple transposases on the other. Interestingly, the 23.3kb and 15.5kb deletions were shared by the same strain. Copy number variation (CNV) analysis indicated that multiple isolates from the same mouse also exhibited copy number increases in two hypothetical proteins; a BLAST search revealed that these hypothetical proteins align significantly to an integrase and an inovirus Gp2 family protein. Together, these results indicate that large deletions

or duplications occur in the EcN genome at low but significant rates during gut colonization, partially mediated by phage and transposon activity.

Carbohydrate utilization, adhesion, and stress drive adaptations in EcN

Beyond the genes just described, we also performed GO annotation of all mutated genes. With 168 mutations, the GO pathway for Cell Division was the most commonly-mutated, but it should be noted that this GO pathway included 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 5]. Collectively, this analysis revealed a diverse array of genomic changes undergone by probiotic EcN as it adapts to the mammalian gut. Our hypothesis that a main driver of probiotic adaptation in the gut is carbohydrate availability and utilization was supported; specifically, we have demonstrated that in the absence of other bacterial taxa the greatest selective pressure on EcN is glycosyl hydrolase activity. Ubiquitous mutations in kfiB provide support for our hypothesis that EcN will undergo adaptation specific to motility or adhesion in the mouse gut. Interestingly, other than one isolate with one mutation in its colicin receptor, we found no evidence of mutations involved in direct inhibitory interactions with other taxa such as toxins or secretion systems, but instead multiple signals for adaptation in stress response regulators. These data paint EcN as not an opportunistically offensive player, but instead a scavenger and survivalist.

PKU treatment by engineered EcN

Engineered probiotics are gaining increased attention as low-cost delivery vehicles of drugs, microbial modulators, and metabolic functions (Hwang et al., 2017; Mansour and Abdelaziz, 2016;

Palmer et al., 2018). Here we demonstrate the efficacy of EcN as a host for synthetic metabolism in the gut microbiota and determine whether mutations to engineered EcN accrue during treatment. To accomplish this, we turned to PKU, which is gaining increased attention as a metabolic disorder which may be treated by an engineered probiotic (Al-Hafid et al., 2015; Durrer et al., 2017; Falb et al., 2017).

PAH-deficient PKU creates a blockage in the tyrosine synthetic pathway leading to Phe accumulation in tissues. It has been shown that expression of a phenylalanine-metabolizing enzyme by probiotic *Lactobacillus* can lower the levels of circulating Phe by ~33% in a mouse model of the disease (Durrer et al., 2017). We utilized PKU as a model to test the Phe catabolic properties of EcN. To this end, we expressed phenylalanine-ammonia lyase from *Arabidopsis thaliana* (McKenna and Nielsen, 2011) under the control of a weak, medium, or strong synthetic promoter (Anderson, 2006). The rationale for using different strength promoters is the known sexual dimorphism exhibited in the Pah^{enu2} mouse model, where females show moderately higher hyperphenylalaninemia than males (Sarkissian et al., 2008). We anticipate different levels of Phe reduction may be required to balance metabolic burden with equivalent therapeutic effect between the sexes. These expression cassettes were placed on a high copy plasmid and transformed into EcN. We observed strong rates of Phe consumption *in vitro* correlated to promoter strength, with the low strength promoter driving slower Phe consumption than the medium or high strength promoters [Figure 6a].

To test whether the differences in promoter activity observed *in vitro* translate to activity differences *in vivo*, the strains with the high and low strength synthetic promoters were delivered to PKU mice via oral gavage (10^9 cfu). Serum Phe was measured before and 24 hours after

treatment. After 24 hours, male mice treated with EcN expressing PAL2 under a high strength promoter, and female mice treated with strains harboring either the high and low strength promoters, saw significant reductions in serum Phe concentrations [Figure 6b-c]. In male mice, serum Phe levels were reduced to half of their baseline levels 24 hours after probiotic delivery by the strongly-expressing probiotic [Figure 6b], indicating the potential of this strain for treatment of metabolic diseases. As expected, we observed dimorphism in the efficacy of this probiotic among males and female mice [Figure 6c]. We further observed *in vivo* differences in Phe reduction in keeping with the expected difference in activity of the low and high strength promoters, indicating the feasibility of tuning the performance of engineered probiotics *in situ*. We were able to recover EcN isolates containing the PAL gene one week after delivery, indicating that the recombinant plasmid was maintained by EcN in the absence of selection. Sequencing of recovered isolates failed to recover any mutations to the PAL2 gene over one week of treatment, indicating that this gene does not impose a substantial metabolic burden over this time course. We also only saw 1/16 strains accrue genomic changes. The two SNPs we observed in this strain were 1) a nonsynonymous mutation in the endoribonuclease ybeY and 2) a nonsense mutation in 4-alpha-glucanotransferase malQ. No other isolate sequenced during this study had mutations in these genes. Given the low frequency of the recovered mutations, we could not confirm our hypothesis that PAL2 expression confers a significant selective pressure to EcN in the PKU gut. Given the substantial amount of strain engineering that can readily be performed to further improve the efficacy of this treatment, we concluded that EcN represents a promising platform with which to treat PKU or deliver additional metabolic functions.

Discussion

Unlike conventional abiotic therapeutics, the performance characteristics of probiotics can vary during treatment in response to natural selection. This variation can be relatively benign (i.e. increased residence time or loss of engineered function) or actively harmful (i.e. acquisition of virulence or antibiotic resistance), and therefore motivates detailed assessment. In this work, we described the evolutionary trajectories traversed by the candidate probiotic *E. coli* Nissle upon delivery to the GI tract.

We first simulated an upper bound for evolutionary rate through horizontal gene transfer via forced expression of a metagenomic DNA library derived from the healthy human fecal microbiota. By competing the resulting strains in the mouse gut, we found that the most advantageous fragments encoded nutrient utilization functions, rather than virulence factors, and that the specific nutrient utilization function was highly dependent upon host diet and microbiome composition. As probiotics are often administered to individuals with gastrointestinal or metabolic disorders, it would be interesting to determine whether these results extend to fragments sourced from dysbiotic microbiotas. While we recovered the highly conserved ABC transporters (Wilkens, 2015) in our experiments, our selections may have under-recovered other functions that require membrane translocation and display or secretion, due to unrecognizable signal sequences and incompatible translocation machinery. Alternate library construction strategies, such as massively parallelized fusion of robust translocation modules with metagenomic material, library expression enhancement with heterologous sigma functions, and expansion of heterologous host species, would be required to further explore this highly interesting functional class (Craig et al., 2010; Fleetwood et al., 2014; Gaida et al., 2015).

We next performed the most extensive analysis of *in vivo* probiotic genome evolution to date through whole-genome sequencing of 401 gut-adapted EcN isolates. We again observed strong selective signatures in nutrient utilization pathways, and even saw convergence between horizontal and vertical evolution in the inactivation of gluconate metabolism. In response to the harsh conditions of the gastrointestinal tract, we found that pathways involved in stress response were also significantly mutated. Surprisingly, we did not observe plasmid acquisition in any sequenced EcN isolate during gut passage, suggesting that the extent of HGT is low for this strain and validating the use of functional metagenomics to characterize in detail the heterologous functions conferring a selective advantage to EcN *in vivo*. In contrast to our expectation, functions involved in direct antagonistic interactions with other bacteria were not significantly mutated, collectively indicating that EcN defines its gut niche through metabolic competition. Finally, we did not observe any mutations in pathways associated with virulence, with the caveat that antibiotic exposure prior to probiotic delivery resulted in selection for antibiotic-resistant strains. Because probiotics and other commensals are often administered after antibiotic treatment to restore gut microbial diversity (Ciorba, 2012), refining the timescale over which antibiotics continue to exert selective pressure in the gut after administration is of critical importance.

Given the promising evolutionary characteristics of *E. coli* Nissle, we next demonstrated its potential as a delivery vehicle for added metabolic function, using phenylketonuria as a model disease. We found that expression of PAL was effective at reducing plasma Phe levels in male PKU mice by over 50% at 24h after administration, exceeding the reported efficacy of an engineered *Lactobacillus* (Durrer et al., 2017), and supporting engineered probiotics as a promising treatment route for PKU. This result also indicates that the powerful genetic tools

available for *E. coli* Nissle manipulation position it well as a chassis for further engineering efforts, potentially enabling the generation of "personalized" probiotic therapies whose Pal expression is tuned to an individual's unique Phe levels. Finally, we were unable to observe significant changes to the genome or plasmids of our engineered probiotic over one week of treatment, indicating that some types of engineered metabolism may not impose a substantial growth defect in the gut over the measured time scales.

Probiotics are an emerging class of therapeutics with increasing interest in diverse areas. Their use extends beyond the human large intestine and into habitats such as the mouth (Burton et al., 2006), skin (Wang et al., 2014), and urogenital tract (Stapleton et al., 2011). Microbes (natural and engineered) are also being deployed in agricultural (Berg, 2009), built environment (Liu et al., 2013), and natural (Pieper and Reineke, 2000) settings. Due to their wide use and propensity for genomic change we have just described, we argue that probiotics demand additional safety considerations beyond what is normally undertaken for traditional, nonliving entities. Coupled with continued development of biocontainment approaches (Chan et al., 2016), we envision the experiments described herein as a general pipeline to be applied to other probiotic strains in development to better understand their safety and engineering potential. It is possible to use *in vivo* functional metagenomic selections or genome evolution studies to develop engineered probiotics with increased colonization rates and longer residence times. However, the major selective pressures operating on any engineered probiotic may be substantially different than their wild type counterparts and would warrant further study of their *in vivo* evolutionary trajectories. As such, this work presents a generalizable framework for developing and regulating living therapeutics.

Acknowledgments

We would like to acknowledge Chyi Hsieh and members of his lab for shared setup and maintenance of our germ-free mouse facility, Jessica Hoisington-Lopez, Eric Martin, and Brian Koebbe for next-generation sequencing and high-throughput computing support at the Edison Family Center for Genome Sciences and Systems Biology at Washington University in St Louis School of Medicine, and Thaddeus Stappenbeck, Tae Seok Moon, and members of the Dantas lab for helpful discussions relating to work described herein, This work is supported in part by the NIH Director's New Innovator Award (http://commonfund.nih.gov/newinnovator/), the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK: http://www.niddk.nih.gov/), the National Institute of General Medical Sciences (NIGMS: http://www.nigms.nih.gov/), the National Institute of Allergy and Infectious Diseases (NIAID: https://www.niaid.nih.gov/), and the Eunice Kennedy Shriver National Institute of Child Health & Human Development (https://www.nichd.nih.gov/) of the National Institutes of Health (NIH) under award numbers DP2DK098089, R01GM099538, R01AI123394, and R01HD092414 to G.D. This work was also supported in part by the Kenneth Rainin Foundation Innovation and Breakthrough Awards (13H5) to G.D. N.C. received support from the NIDDK Pediatric Gastroenterology Research Training Program of the NIH, under award number T32 DK077653 (Phillip I. Tarr, Principal Investigator). A.F. received support from the Chancellor's Graduate Research Fellowship Program at Washington University in St. Louis. A.J.G. received support from a NIGMS training grant through award number T32 GM007067 (Jim Skeath, Principal Investigator) and the NIDDK Pediatric Gastroenterology Research Training Program through award number T32 DK077653 (Phillip I. Tarr, Principal Investigator). M.W.P. received support from a NIGMS training grant through award number T32 GM007067 (Jim Skeath, Principal Investigator). M.K.G. received support as a

Mr. and Mrs. Spencer T. Olin Fellow at Washington University in St Louis and from the National Science Foundation (NSF) as a graduate research fellow under award number DGE-1143954. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Author Contributions

N.C., A.F., and A.J.G. designed and performed *in vivo* functional metagenomic selections in germ-free and conventional mice and carried out high-throughput sequencing of functional metagenomic hits. N.C. and A.F. also carried out high-throughput sequencing of adapted isolates, designed and performed *in vitro* microbiological experiments, designed and performed *in vivo* PKU experiments with Pah^{enu2} mice, analyzed data, and wrote the manuscript. M.P. and M.G. designed and performed *in vivo* functional metagenomic selections in germ-free mice. B.W. transformed metagenomic libraries into EcN and engineered a GFP-tagged EcN reference strain. S.S. carried out *in vivo* functional metagenomic selections in germ-free and conventional mice. Z.C. generated the experimental EcN:PAL2 strains for the PKU experiments. S.D. created and shared the Pah^{enu2} mouse model and provided guidance on its establishment and use. D.P carried out *in vivo* functional metagenomic selections in germ-free mice. G.D. designed experiments, analyzed data, and wrote the manuscript.

Declaration of Interests

The authors declare no competing interests

Tables

Table 1. Summary of Functional Metagenomic Selection Experiments. See also Table S1, Table S2, and Table S4 for descriptions of the diet compositions and the 13-member community.

Microbiome	Diet	# Mice
Germ-free	Mouse	6
Germ-free	Human	5
Germ-free	Human+Dextrin	5
Germ-free	Human+Inulin	5
Germ-free	Human	3
13-member	Human	3
13-member	Human+Dextrin	3
13-member	Human+Inulin	3

Methods

Media and Strains

E. coli Nissle was obtained as a kind gift from Dr. Phillip I. Tarr (Washington University in St. Louis School of Medicine). Unless otherwise indicated, EcN was grown in LB media under aerobic conditions at 37 °C. Agar was added at a concentration of 15 g/L for growth on solid media. For maintenance of the pZE21 plasmid (Lutz and Bujard, 1997) and its derivatives, kanamycin was added to a final concentration of 50 µg/mL after autoclaving.

Thirteen strains that were selected to mimic a healthy human microbiota composition were obtained as a kind gift from Dr. Andrew Goodman (Yale University) (Supplementary Table 4). Strains were cultured in TYG medium supplemented with D-(+)-Cellobiose (0.1% w/v; Sigma), D-(+)-Maltose (0.1% w/v; Sigma), D-(-)-Fructose (0.1% w/v; Sigma), Tween 80 (0.05% v/v; Sigma), Meat Extract (0.5% w/v; Sigma), ATCC Trace Mineral Supplement (1% v/v), ATCC Vitamin Supplement (1% v/v), N-butyric acid (4mM), propionic acid (8mM), isovaleric acid

(1mM), and acetic acid (30mM) as in Goodman *et. al.* (Goodman et al., 2009). Culturing occurred under anaerobic conditions in a soft-sided plastic anaerobic chamber (Coy Laboratory Products). Stocks of strains were stored in E-Z crimp top vials (Wheaton) at -80˚C in Mega Medium with 20% glycerol. Stocks were titered by plating on BHI blood agar plates. Strains were pooled immediately prior to gavage such that each member was equally represented, and washed 3x in PBS. *Eubacterium rectale* and *Ruminococcus torques* displayed poor recovery rates from freezer stocks, so overnight cultures at early stationary phase were included in the final pool *in lieu* of stocks. Mice were gavaged with a total dose of $1x10^8$ CFU in a volume of 200 µL.

Metagenomic library construction

Metagenomic libraries were sourced from a prior study examining the antibiotic resistome of the healthy maternal and infant microbiomes (Moore et al., 2015). Frozen glycerol stocks containing these libraries were thawed on ice, and plasmid DNA was extracted using a Qiagen Spin Miniprep kit. 100 ng of DNA was transformed into electrocompetent *E. coli* Nissle, and allowed to recover in SOC media for 1h at 37 °C. After 1h, 1 µL of SOC media was plated on LB+kan, to count transformants, and the remainder was placed in 50mL LB media containing kanamycin at room temperature overnight. After overnight growth, the library was centrifuged at 4000 rpm for 7 m, resuspended in 10mL LB+15% w/v glycerol, and frozen in 1mL aliquots at -80 °C. After freezing, one aliquot was thawed on ice, and serial dilutions were plated on LB+kan media to count viable cells. This procedure yielded a library containing 32.77 Gb of metagenomic DNA.

Plasmids used in this study

All plasmids used in this study were constructed using a Golden Gate Assembly Mastermix from NEB, according to the manufacturer's instructions. Plasmid pSPAL2At containing the PAL2 enzyme was obtained from Addgene (cat #78286), as a kind gift from David Nielsen (McKenna and Nielsen, 2011) Genbank files of all plasmids are provided in [Supplementary File 2].

Mouse experiments

All mouse experiments were approved by the Washington University in Saint Louis School of Medicine Division of Comparative Medicine.

E. coli Nissle monocolonization experiments were performed in germ-free C57BL/6 mice (University of Michigan). Upon arrival, mice were provided food and water *ad libitum* for one week prior to colonization. For the mouse chow experiment, mice were provided autoclaved feed (Purina Conventional Mouse Diet (JL Rat/Mouse 6F Auto) #5K67) and autoclaved water. Western diet consisted of Envigo TD.88137, irradiated and vacuum-sealed. Inulin or Dextrin (Sigma) was provided in the drinking water at a concentration of 20 g/L, and autoclaved. 6 (for the mouse chow experiment) or 5 (for the Western Diet experiment) co-housed male mice were exposed to 10^8 colony forming units of *E. coli* Nissle via oral gavage of 100 µL cell suspension in phosphate buffered saline on day 1 of the experiment. Feces were collected at weekly intervals and immediately frozen at -80 °C. At the end of the experiment, mice were sacrificed through carbon dioxide asphyxiation and cecal contents were collected.

Experiments involving the 13-member synthetic microbiota were also performed in germ-free C57BL/6 mice (University of Pennsylvania Gnotobiotic Mouse Facility). Upon arrival, 3 germ-free mice (1-2 male, 1-2 female) per condition were housed separately in germ-free isolators. Mice were acclimated to their diet for one week prior to colonization with the synthetic microbiota. The

synthetic microbiota was allowed to stabilize for one week prior to delivery of 10^8 colony forming units of *E. coli* Nissle via oral gavage of 100 µL cell suspension in phosphate buffered saline on day 1 of the experiment. Fecal samples were subsequently collected weekly for 5 weeks, and immediately frozen at -80 °C. At the end of the experiment, mice were sacrificed through carbon dioxide asphyxiation and cecal contents were collected.

Experiments involving the conventional mouse microbiota were performed in a specific-pathogen-free facility. 2 cages of 5 mice (Jackson Labs C57BL/6J) for each condition were maintained on the Western Diet with or without prebiotic for one week after arrival. After one week, mice were deprived of food and water for 4h, and given either 20mg streptomycin in 100 µL water, or 100 µL water alone via oral gavage. After this treatment, food and water was immediately returned. Daily over the next three days, 10^8 CFU of EcN was delivered in 100 µL PBS via oral gavage. Fecal samples and intestinal contents were collected as above.

Experiments involving PKU mice were performed in a specific-pathogen-free mouse facility. For each condition (High PAL expression, Low PAL expression, or No PAL expression), 3 Male and 3 Female PKU mice (10-12 weeks of age) were co-housed and maintained on Phe-containing mouse chow. 100 µL blood was sampled from the femoral artery one week before the experiment to serve as a baseline for each mouse. Then, 10^9 CFU of the appropriate EcN strain was delivered via oral gavage. One day after gavage, 100 µL blood was again sampled from the femoral artery. Fecal samples were collected 1 week after gavage. After collection, all blood samples were allowed to coagulate then centrifuged in BD Microtainer ® SST ™ tubes for 2 minutes at 10,000g to collect serum. Serum samples were then passed through 10 kDA molecular weight cut-off filters by

centrifugation. Phe levels of the filtrates were determined using a Phenylalanine Assay Kit (MAK005, Sigma-Aldrich).

Functional metagenomic library sequencing

DNA was extracted from fecal samples and intestinal contents using the PowerSoil DNA extraction kit (MoBio/Qiagen). 10 ng of this DNA was used as a template for a multiplex PCR with primers 1-6 [Supplementary Table 6] at equimolar concentrations using Taq Reddymix (Fisher Scientific). PCR recipe per sample was as follows: 12.5 µL Taq Reddymix, 3 µL primer mix (10 µM total concentration), 10 ng template, nuclease free water to 25 µL. PCR protocol was as follows: 94 °C for 10 m, 94 °C for 45 s, 55 °C for 45 s, 72 °C for 5.5 m, go to step 2 24 times, 72 °C for 10 m, 4 °C forever. PCR products were purified using a Qiagen Spin PCR Purification Kit. 500 ng PCR products in 200 µL elution buffer (from purification kit) were then placed in a AB1900 half-skirted plate (Fisher Scientific) and sonicated using a Covaris E220 sonicator. The settings on the sonicator were as follows: Peak Incident Power: 140, Duty Cycle: 10%, Cycles per Burst: 200, Treatment Time: 600 s, Temp: 7 °C. Sonicated products were cleaned using a Qiagen MinElute PCR purification kit according to the manufacturer's instructions, and eluted in 22 µL elution buffer. 20 µL sonicated products were end-repaired through addition of 2.5 µL DNA ligase buffer with 10 mM ATP (NEB), 1 µL 1 mM dNTP, 0.5 µL T4 Polymerase (NEB), 0.5 µL T4 PNK (NEB), and 0.5 µL Taq Polymerase (NEB), and incubating for 30 m at 25 °C, followed by 20 m at 75 °C. Then, 0.8 µL T4 DNA ligase (NEB) and 5 µL pre-annealed sequencing barcodes were added to the full end-repair reaction, followed by 40 m at 16 °C, and 10 m at 65 °C.

Sequencing barcodes were designed according to the template described in Primers 7-8 (Supplementary Table 6), and synthesized by IDT in 2 96-well plates to a concentration of 500 µM. Each barcode was designed to be at least 2 mismatches away from every other barcode. 2 µL

of each primer pair were added to 96 µL of TES buffer (10mM Tris, 1mM EDTA, 50mM NaCl, pH 8.0), and this mixture was diluted 1:10 in TES buffer, to a final concentration of 1 µM. Adapters were annealed by incubating at 95 °C for 1 m, followed by slowly cooling (0.1 °C per second) to 4 °C. Annealed oligos were kept cold and transferred to a -20 °C freezer until use.

Next, ligation products were size-selected to 300-400bp in size through gel electrophoresis. Briefly, the total ligation volume was loaded on a 2% (w/v) agarose gel in 0.5x TBE buffer using thin combs. This gel was run at 120V for 2h, or until the loading dye front reached ~2/3 of the way to the end of the gel. DNA fragments between 300 and 400bp were excised, and purified using the Qiagen MinElute Gel Extraction Kit, eluting in 15 µL elution buffer.

Sequencing adaptors were then added to the size-selected ligation products through PCR. PCR recipe was as follows: 12.5 µL Phusion HF Mastermix (Finnzymes), 9.5 µL nuclease-free water, 1 µL 10 µM primer mix (Primers 9-10) and 2 µL gel-purified DNA. PCR conditions were as follows: 98 °C for 30 s, 98 °C for 10 s, 65 °C for 30 s, 72 °C for 30 s, return to step 2 17 times, 72 °C for 5 min, 4 °C forever. PCR products were then size-selected on a 2% agarose, 0.5x TBE gel as above. Sequencing libraries were quantified on a Qubit fluorimeter, and pooled at equimolar concentrations for sequencing. 100,000 2x150 paired-end sequencing reads were obtained per sample.

Functional metagenomic sequencing analysis

Functional metagenomic contigs were assembled from sequencing reads and annotated using PARFuMS (Boolchandani et al., 2017; Forsberg et al., 2012). For each experiment, contigs were clustered at 95% nucleotide identity using CD-HIT (Fu et al., 2012). Then, vector-trimmed sequencing reads (from PARFuMS) were uniquely mapped to each contig using bowtie2

(Langmead and Salzberg, 2012), and mapping reads were counted using SAMtools (Li, 2011). Contig abundance was estimated as the fraction of reads mapping to each contig. To perform Gene Ontology mapping, open reading frames (from PARFuMS) were mapped against all nonredundant protein sequences using BLASTp (Boratyn et al., 2012), and annotated using InterProScan (Jones et al., 2014). Outputs of these two operations were then used as input to Blast2GO, according to manufacturer's instructions. Each contig was associated with a GO annotation if at least one of its constituent open reading frames was. The abundance of each GO annotation was then estimated as the sum of contig abundances associated with that GO annotation. Heatmaps of GO annotations were plotted using the heatmaply package in R.

Isolate sequencing

To isolate *in vivo*-adapted EcN isolates, intestinal contents or fecal samples were vortexed in phosphate-buffered saline and streaked on LB plates containing kanamycin. Single colonies were placed in liquid media and grown at 37 °C overnight. Un-adapted EcN strains were also grown overnight. Total DNA was extracted from each culture using the DNeasy UltraClean 96 Microbial kit (10196-4, Qiagen), and prepared for whole-genome sequencing using Nextera Tagmentation (Baym et al., 2015). Briefly, gDNA was brought to a concentration of 0.5 ng/µL in 1 µL volume. To each sample, 1.25 µL TD buffer, 0.125 µL TDE1 enzyme, and 0.125 µL nuclease-free water was added, and incubated at 55 °C for 15 m. Then, 11.2 µL KAPA HiFi master mix was added to each tagmented sample. Indexed sequencing adaptors (5 µM each) were thawed, and 8.8 µL was added to each sample. Then, PCR was performed using the following protocol: 72 °C for 3 m, 98 °C for 5 m, 98 °C for 10 s, 63 °C for 30 s, 72 °C for 30s, return to step 2 13 times, 72 °C for 5 m, 4 °C forever. PCR reactions were purified using AMPure XP beads according to the

manufacturer's protocol, and eluted in 60 µL resuspension buffer. Sequencing libraries were quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Fisher Scientific), and pooled at equimolar concentrations for sequencing. 2 million 2x150bp sequencing reads were obtained per sample.

MinION sequencing

Genomic DNA from 30 mL of overnight culture of control EcN was extracted using a Genomic DNA Buffer Set (19060, Qiagen) with the Qiagen protocol for Gram-negative bacteria (Qiagen Genomic DNA Handbook 06/2015) and the Genomic-tip 500/G (10262, Qiagen). Purified genomic DNA was sheared to a target fragment size of 10 kilobases using the Covaris g-TUBE™. All following steps were carried out in Eppendorf DNA LoBind tubes. Sheared DNA was repaired using the NEBNext FFPE Repair Mix (M6630), and repaired DNA purified from the reaction using Agencourt AMPure XP beads. DNA was then end-repaired and dA-tailed using the NEBNext End repair / dA-tailing Module (E7546) and again purified with AMPure XP beads. DNA was then prepared for MinION sequencing using the MinION Ligation Sequencing Kit 1D (SQK-LSK108). Briefly, adapters were ligated to the DNA using the MinION Adapter Mix (AMX1D) and NEB Blunt/TA Ligase Master Mix (M0367). Then the DNA product was purified using AMPure XP beads and the Minion Adapter Bead Binding Buffer (ABB). 350 ng of DNA product was then sequenced on a MinION R9.4 flow cell for 48 hours using the Running Buffer with Fuel Mix (RBF) and the Library Loading Bead Kit (EXP-LLB001). Average read lengths were 7 kilobases and the genome was sequenced with an average coverage of 50 reads.

Isolate sequencing data analysis

Illumina sequencing reads were quality-filtered and adaptor-trimmed using Trimmomatic (Bolger et al., 2014). Isolate genomes were assembled with Illumina data using SPAdes (Bankevich et al., 2012). The control EcN genome (which had both Illumina and MinION data) was assembled using SPAdes in hybrid mode. Illumina reads were mapped to the assembled genome using bowtie2 and MinION reads were mapped using LAST (Frith and Noe, 2014). Mapping results were converted to indexed, sorted bam files using SAMtools and visualized using IGV (Thorvaldsdottir et al., 2013). The control EcN genome was then corrected manually by inspecting read alignments for dips in coverage, and correcting these regions using MinION data as well as homology to a published EcN genome (Reister et al., 2014) using an iterative process. After each iteration, Illumina and MinION reads were re-mapped to the corrected genome, and read alignments were inspected to verify an improved assembly. Coding sequences were annotated using Prokka (Seemann, 2014). One region of spurious assembly was due to imperfect tandem repeats present in the UpaH gene; this region was resolved by PCR amplification from the genome followed by Sanger sequencing. The result of this finishing process became the control genome to which all *in vivo*-adapted isolates were compared.

Mutation Calling

Single-nucleotide polymorphisms and indels were annotated using two complementary approaches. The first method was Breseq, using default parameters (Deatherage and Barrick, 2014). The second method was VarScan2, where SNPs and indels were called if supported by a minimum read coverage of 30 and 80% abundance in those reads (Koboldt et al., 2012). VarScan2 variant call format output was then annotated with ANNOVAR (Wang et al., 2010). Mutations called by each method were manually verified through visualization of Illumina read alignments

using IGV. Anecdotally, both methods did recover some false positive mutation calls, primarily due to the presence of a homologous metagenomic DNA insert causing spurious read alignments to the EcN genome. Phage sequences were identified using the Phaster web server (Arndt et al., 2016). To look for potential plasmid sequences, assembled contigs for each isolate were aligned to 1) all nonredundant nucleotide sequences, and 2) the control EcN genome using BLASTn. Our criteria for plasmid sequences was the following: >500bp in length, <90% of bases mapping to the EcN genome, no homology to the functional metagenomic insert present in the strain (if any), and no obvious contamination from other species. No contig satisfied these criteria, and therefore we concluded that no acquired plasmid sequences were obviously present in our *in vivo* adapted strains.

Annotation

To perform Gene Ontology mapping, mutated open reading frameswere mapped against all nonredundant protein sequences using BLASTp, and annotated using InterProScan. Outputs of these two operations were then used as input to Blast2GO, according to manufacturer's instructions. The abundance of each GO annotation was then estimated as the number of open mutated open reading frames with that annotation among the strains within a certain condition. Heatmaps of GO annotations were plotted using the heatmaply package in R.

Copy Number Variation

Copy number variation was assessed using the CNVkit pipeline (Talevich et al., 2016). Briefly, short reads for each isolate were mapped to an indexed reference assembly using bowtie2 with the --very-senstitive-local option to generate SAM files. SAM files were then converted to sorted

BAM files using Samtools. Coverage was calculated in bins using the CNVkit coverage function and compared to the coverage of the reference strain using the CNVkit reference function with the --no-edge option. Biases were adjusted using the CNVkit fix and segment functions. CNV hits were called if the bias-adjusted log2 copy ratio was greater than 1.3 or less than -1.3.

Growth assays on functional metagenomic hits

Three biological replicates of EcN containing the appropriate plasmid were grown overnight in LB media containing kanamycin at 37 °C under anaerobic conditions. The cell densities of these overnight cultures were measured, and EcN was transferred to M9 MOPS minimal media containing the carbon source of interest at a starting optical density of 0.01 in a 96-well plate. BIOLOG plates were prepared with M9-MOPS minimal media per the manufacturer's recommendations. A Breathe-Easy Sealing Membrane (Sigma) was placed over the plate. This plate was placed in a plate reader set to maintain 37 °C and read optical density (absorbance at 600 nm) every minute for 72h.

Growth assays on mutant isolates

GntT mutants were grown on 0.4% gluconate minimal media. Three representative GntT mutant isolates were grown overnight in LB media at 37 °C aerobically, along with wild type EcN. The cell densities of these overnight cultures were measured, and the strains were transferred to M9 minimal media supplemented with 0.4% gluconate at a starting optical density of 0.05 in a 96-well plate as before. This plate was set in a plate shaker at 37 °C. Optical density (absorbance at 600nm) was measured at 0, 4, 8, and 24 hours.

NagC mutants were grown in diauxic growth experiments. As before, overnight cultures were transferred to M9 supplemented with 0.2% glucose 0.75% porcine gastric mucin (Sigma). Alternatively, overnight cultures were transferred to M9 supplemented with either 0.4% N-acetylglucosamine, or 0.2% glucose and 0.2% N-acetylglucosamine. These plates were set in a plate reader at 37 °C. Optical density (absorbance at 600 nm) was measured for 18 hours at an interval of 20 minutes.

Epithelial cell binding assay

Caco-2 cells were seeded overnight in a 24-well plate at 50,000 cells/cm^2 in DMEM-HEPES media supplemented with Glutamax (10564-011, ThermoFisher) and 10% Fetal Bovine Serum. They were incubated at 37 °C, 5.0% CO_2 and 90% relative humidity. Overnight cultures of representative *kfiB* mutants or wild type EcN were grown in LB at 37 °C. The next day, the wells with Caco-2 cells were inoculated with mutant or wild type EcN that had been washed and resuspended in PBS. The multiplicity of infection was 1:100. The human and bacterial cells were co-incubated 37 °C, 5.0% CO_2 and 90% relative humidity for 3 hours. Then, media was aspirated from each well. Three replicate wells for each strain were gently washed 3x with PBS, while three replicate wells for each strain were left untreated. All wells were then treated with 100 µL of 1% Triton-X for 10 minutes at room temperature to lyse the human cells. Then, 900 µL LB was added to each well and the cells resuspended by pipetting. These cultures were diluted and plated on LB agar plates and grown overnight at 37 °C. The next day CFUs were enumerated, and the percentage of adherent cells was calculated as the mean ratio of CFUs/mL from the washed and unwashed wells for each strain.

In vitro *phenylalanine degradation assay*

EcN strains expressing PAL2 in the pZE21 plasmid under promoters of low (p08), medium (p01), or high (p19) strength, or harboring only empty vector, were grown overnight in LB at 37 °C. The next day the optical densities (absorbance at 600nm) of these cultures were measured and the strains were transferred to M9 minimal media supplemented with 0.4% glucose and 20 µM L-phenylalanine with a starting optical density of 1.0. Fractions of these cultures were removed at 0, 30, 60, and 90 minutes and immediately centrifuged. Supernatants were passaged through 10 kDa molecular weight cut-off filters and the phenylalanine in the filtrates was measured using a Phenylalanine Assay Kit (MAK005, Sigma-Aldrich). Phenylalanine levels were reported as average absolute nanomoles of phenylalanine per well. Three technical replicates were carried out for each strain.

Gene phylogeny construction and enrichment testing

We searched for genes with variants significantly enriched in specific diet conditions or in specific microbiome complexities, using the hypergeometric test for enrichment and Bonferroni correction for multiple hypothesis testing. The significant genes included *gntT*, *nagC*, *sgrR*, *rsmG*, and *kfiB*. Gene phylogenies were created for these genes by aligning the gene sequence for all isolates involved in the comparison in which the gene variants were found enriched in a condition (that is, all isolates along the axis of different diets, or the axis of microbiome complexity). Alignments were generated using Clustal Omega. Maximum parsimony phylogenies for each gene were generated using the Dnapars program included in PHYLIP 3.695 (Felsenstein, 2005), and visualized using iTOL (Letunic and Bork, 2016).

E. coli phylogeny construction

Representative E. coli genomes were downloaded from NCBI. PROKKA was used to generate general feature format (gff) files for each *E. coli* genome, including our assembled reference EcN genome (Seemann, 2014). Roary was used to generate a core gene alignment, using these gff files and the corresponding genome sequences (Page et al., 2015). The strain phylogeny was inferred with RaxML using the GTRGAMMA model and 1000 bootstraps (Stamatakis, 2014). The best tree was visualized using iTOL.

Analysis of selection

SgrR was analyzed for evidence of positive or negative selection. The previously generated gene alignment and phylogeny were provided as input to the aBSREL, MEME, and FUBAR modules included in the HyPhy application hosted by the Datamonkey webserver (Delport, 2010). Neither of the three modules reported evidence of selection.

References

Adediran, J., Leatham-Jensen, M.P., Mokszycki, M.E., Frimodt-Moller, J., Krogfelt, K.A., Kazmierczak, K., Kenney, L.J., Conway, T., and Cohen, P.S. (2014). An Escherichia coli Nissle 1917 missense mutant colonizes the streptomycin-treated mouse intestine better than the wild type but is not a better probiotic. Infect Immun *82*, 670-682.

Al-Hafid, N., Christodoulou, J., and Tong, X.Z. (2015). Genetically-modified probiotic for treatment of phenylketonuria (Google Patents).

Anderson, C. (2006). Anderson promoter collection. iGEM.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D.S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res *44*, W16-21.

Backhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A., and Gordon, J.I. (2005). Host-bacterial mutualism in the human intestine. Science *307*, 1915-1920.

978 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I.,
979 Pham, S., Prjibelski, A.D.*, et al.* (2012). SPAdes: a new genome assembly algorithm and its applications to
980 single-cell sequencing. J Comput Biol *19*, 455-477.

981 Barroso-Batista, J., Sousa, A., Lourenco, M., Bergman, M.L., Sobral, D., Demengeot, J., Xavier, K.B., and
982 Gordo, I. (2014). The first steps of adaptation of Escherichia coli to the gut are dominated by soft sweeps.
983 PLoS Genet *10*, e1004182.

984 Battesti, A., Majdalani, N., and Gottesman, S. (2011). The RpoS-mediated general stress response in
985 Escherichia coli. Annu Rev Microbiol *65*, 189-213.

986 Baym, M., Kryazhimskiy, S., Lieberman, T.D., Chung, H., Desai, M.M., and Kishony, R. (2015). Inexpensive
987 multiplexed library preparation for megabase-sized genomes. PLoS One *10*, e0128036.

988 Benitez-Paez, A., Villarroya, M., and Armengod, M.E. (2012). Regulation of expression and catalytic activity
989 of Escherichia coli RsmG methyltransferase. Rna *18*, 795-806.

990 Berg, G. (2009). Plant-microbe interactions promoting plant growth and health: perspectives for
991 controlled use of microorganisms in agriculture. Appl Microbiol Biotechnol *84*, 11-18.

992 Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence
993 data. Bioinformatics *30*, 2114-2120.

994 Boolchandani, M., Patel, S., and Dantas, G. (2017). Functional Metagenomics to Study Antibiotic
995 Resistance. Methods Mol Biol *1520*, 307-329.

996 Boratyn, G.M., Schaffer, A.A., Agarwala, R., Altschul, S.F., Lipman, D.J., and Madden, T.L. (2012). Domain
997 enhanced lookup time accelerated BLAST. Biol Direct *7*, 12.

998 Burton, J.P., Chilcott, C.N., Moore, C.J., Speiser, G., and Tagg, J.R. (2006). A preliminary study of the effect
999 of probiotic Streptococcus salivarius K12 on oral malodour parameters. J Appl Microbiol *100*, 754-764.

1000 Chan, C.T., Lee, J.W., Cameron, D.E., Bashor, C.J., and Collins, J.J. (2016). 'Deadman' and 'Passcode'
1001 microbial kill switches for bacterial containment. Nat Chem Biol *12*, 82-86.

1002 Chang, D.E., Smalley, D.J., Tucker, D.L., Leatham, M.P., Norris, W.E., Stevenson, S.J., Anderson, A.B.,
1003 Grissom, J.E., Laux, D.C., Cohen, P.S.*, et al.* (2004). Carbon nutrition of Escherichia coli in the mouse
1004 intestine. Proc Natl Acad Sci U S A *101*, 7427-7432.

1005 Charlesworth, D. (2016). BioMarin Phase 3 Study of Pegvaliase for Phenylketonuria (PKU) Meets Primary
1006 Endpoint of Blood Phenylalanine (Phe) Reduction (p<0.0001).

1007 Ciorba, M.A. (2012). A gastroenterologist's guide to probiotics. Clin Gastroenterol Hepatol *10*, 960-968.

1008 Conway, T., and Cohen, P.S. (2015). Commensal and Pathogenic Escherichia coli Metabolism in the Gut.
1009 Microbiol Spectr *3*.

1010 Craig, J.W., Chang, F.-Y., Kim, J.H., Obiajulu, S.C., and Brady, S.F. (2010). Expanding Small-Molecule
1011 Functional Metagenomics through Parallel Screening of Broad-Host-Range Cosmid Environmental DNA
1012 Libraries in Diverse Proteobacteria. Appl Environ Microbiol *76*, 1633-1641.

1013 Deatherage, D.E., and Barrick, J.E. (2014). Identification of mutations in laboratory-evolved microbes from
1014 next-generation sequencing data using breseq. Methods Mol Biol *1151*, 165-188.

1015 Dechant, R., Binda, M., Lee, S.S., Pelet, S., Winderickx, J., and Peter, M. (2010). Cytosolic pH is a second
1016 messenger for glucose and regulates the PKA pathway through V-ATPase. Embo j *29*, 2515-2526.

1017 Delport, W.P., A.F., Frost, S.D.W., Pond, S.L.K. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools
1018 for evolutionary biology Bioinformatics *26*, 2455-2457.

1019 Durrer, K.E., Allen, M.S., and Hunt von Herbing, I. (2017). Genetically engineered probiotic for the
1020 treatment of phenylketonuria (PKU); assessment of a novel treatment in vitro and in the PAHenu2 mouse
1021 model of PKU. PLoS One *12*, e0176286.

1022 Eavri, R., and Lorberboum-Galski, H. (2007). A novel approach for enzyme replacement therapy. The use
1023 of phenylalanine hydroxylase-based fusion proteins for the treatment of phenylketonuria. J Biol Chem
1024 *282*, 23402-23409.

1025 El Kaoutari, A., Armougom, F., Gordon, J.I., Raoult, D., and Henrissat, B. (2013). The abundance and variety
1026 of carbohydrate-active enzymes in the human gut microbiota. Nat Rev Microbiol *11*, 497-504.

1027 Fabich, A.J., Jones, S.A., Chowdhury, F.Z., Cernosek, A., Anderson, A., Smalley, D., McHargue, J.W.,
1028 Hightower, G.A., Smith, J.T., Autieri, S.M.*, et al.* (2008). Comparison of carbon nutrition for pathogenic
1029 and commensal Escherichia coli strains in the mouse intestine. Infect Immun *76*, 1143-1152.

1030 Falb, D., Isabella, V.M., Kotula, J.W., Miller, P.F., Millet, Y., and Sarah, R. (2017). Bacteria engineered to
1031 reduce hyperphenylalaninemia (Google Patents).

1032 Felsenstein, J. (2005). PHYLIP (Phylogeny Inference Package) (Distributed by the author).

1033 Ferreiro, A., Crook, N., Gasparrini, A.J., and Dantas, G. (2018). Multiscale Evolutionary Dynamics of Host-
1034 Associated Microbiomes. Cell *172*, 1216-1227.

1035 Fleetwood, F., Andersson, K.G., Ståhl, S., and Löfblom, J. (2014). An engineered autotransporter-based
1036 surface expression vector enables efficient display of Affibody molecules on OmpT-negative E. coli as well
1037 as protease-mediated secretion in OmpT-positive strains. Microbial Cell Factories *13*, 179.

1038 Fletcher, E., Krivoruchko, A., and Nielsen, J. (2016). Industrial systems biology and its impact on synthetic
1039 biology of yeast cell factories. Biotechnol Bioeng *113*, 1164-1170.

1040 Fontaine, F., Stewart, E.J., Lindner, A.B., and Taddei, F. (2008). Mutations in two global regulators lower
1041 individual mortality in Escherichia coli. Mol Microbiol *67*, 2-14.

1042 Forsberg, K.J., Reyes, A., Wang, B., Selleck, E.M., Sommer, M.O., and Dantas, G. (2012). The shared
1043 antibiotic resistome of soil bacteria and human pathogens. Science *337*, 1107-1111.

1044 Frith, M.C., and Noe, L. (2014). Improved search heuristics find 20,000 new alignments between human
1045 and mouse genomes. Nucleic Acids Res *42*, e59.

1046 Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation
1047 sequencing data. Bioinformatics *28*, 3150-3152.

1048 Gaida, S.M., Sandoval, N.R., Nicolaou, S.A., Chen, Y., Venkataramanan, K.P., and Papoutsakis, E.T. (2015).
1049 Expression of heterologous sigma factors enables functional screening of metagenomic and heterologous
1050 genomic libraries. In Nat Commun, pp. 7045.

1051 Ghalayini, M., Launay, A., Bridier-Nahmias, A., Clermont, O., Denamur, E., Lescat, M., and Tenaillon, O.
1052 (2018). "Evolution of a dominant natural isolate of Escherichia coli in the human gut over a year suggests
1053 a neutral evolution with reduced effective population size". Appl Environ Microbiol.

1054 Gibson, M.K., Pesesky, M.W., and Dantas, G. (2014). The yin and yang of bacterial resilience in the human
1055 gut microbiota. J Mol Biol *426*, 3866-3876.

1056 Goodman, A.L., Kallstrom, G., Faith, J.J., Reyes, A., Moore, A., Dantas, G., and Gordon, J.I. (2011). Extensive
1057 personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice.
1058 Proc Natl Acad Sci U S A *108*, 6252-6257.

1059 Goodman, A.L., McNulty, N.P., Zhao, Y., Leip, D., Mitra, R.D., Lozupone, C.A., Knight, R., and Gordon, J.I.
1060 (2009). Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell
1061 Host Microbe *6*, 279-289.

1062 Grozdanov, L., Raasch, C., Schulze, J., Sonnenborn, U., Gottschalk, G., Hacker, J., and Dobrindt, U. (2004).
1063 Analysis of the genome structure of the nonpathogenic probiotic Escherichia coli strain Nissle 1917. J
1064 Bacteriol *186*, 5432-5441.

1065 Gumpert, H., Kubicek-Sutherland, J.Z., Porse, A., Karami, N., Munck, C., Linkevicius, M., Adlerberth, I.,
1066 Wold, A.E., Andersson, D.I., and Sommer, M.O.A. (2017). Transfer and Persistence of a Multi-Drug
1067 Resistance Plasmid in situ of the Infant Gut Microbiota in the Absence of Antibiotic Treatment. Front
1068 Microbiol *8*, 1852.

1069 Hafez, M., Hayes, K., Goldrick, M., Grencis, R.K., and Roberts, I.S. (2010). The K5 capsule of Escherichia coli
1070 strain Nissle 1917 is important in stimulating expression of Toll-like receptor 5, CD14, MyD88, and TRIF
1071 together with the induction of interleukin-8 expression via the mitogen-activated protein kinase pathway
1072 in epithelial cells. Infect Immun *78*, 2153-2162.

1073 Helling, R.B. (1998). Pathway choice in glutamate synthesis in Escherichia coli. J Bacteriol *180*, 4571-4575.

1074 Hempel, S., Newberry, S.J., Maher, A.R., and et al. (2012). Probiotics for the prevention and treatment of
1075 antibiotic-associated diarrhea: A systematic review and meta-analysis. JAMA *307*, 1959-1969.

1076 Hill, C., Guarner, F., Reid, G., Gibson, G.R., Merenstein, D.J., Pot, B., Morelli, L., Canani, R.B., Flint, H.J.,
1077 Salminen, S.*, et al.* (2014). Expert consensus document. The International Scientific Association for
1078 Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic.
1079 Nat Rev Gastroenterol Hepatol *11*, 506-514.

1080 Hoek, M.J.A.v., and Merks, R.M.H. (2017). Emergence of microbial diversity due to cross-feeding
1081 interactions in a spatial model of gut microbial metabolism. BMC Systems Biology *11*, 56.

1082 Hoskins, L.C., Agustines, M., McKee, W.B., Boulding, E.T., Kriaris, M., and Niedermeyer, G. (1985). Mucin
1083 degradation in human colon ecosystems. Isolation and properties of fecal strains that degrade ABH blood
1084 group antigens and oligosaccharides from mucin glycoproteins. J Clin Invest *75*, 944-953.

1085 Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J.H., Chinwalla, A.T., Creasy, H.H., Earl,
1086 A.M., FitzGerald, M.G., and Fulton, R.S. (2012). Structure, function and diversity of the healthy human
1087 microbiome. Nature *486*, 207.

1088 Hwang, I.Y., Koh, E., Wong, A., March, J.C., Bentley, W.E., Lee, Y.S., and Chang, M.W. (2017). Engineered
1089 probiotic Escherichia coli can eliminate and prevent Pseudomonas aeruginosa gut infection in animal
1090 models. Nat Commun *8*, 15028.

1091 Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A.,
1092 Nuka, G.*, et al.* (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*,
1093 1236-1240.

1094 Karami, N., Martner, A., Enne, V.I., Swerkersson, S., Adlerberth, I., and Wold, A.E. (2007). Transfer of an
1095 ampicillin resistance gene between two Escherichia coli strains in the bowel microbiota of an infant
1096 treated with antibiotics. J Antimicrob Chemother *60*, 1142-1145.

1097 Kau, A.L., Ahern, P.P., Griffin, N.W., Goodman, A.L., and Gordon, J.I. (2011). Human nutrition, the gut
1098 microbiome and the immune system. Nature *474*, 327-336.

1099 Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L.,
1100 and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by
1101 exome sequencing. Genome Res *22*, 568-576.

1102 Koenigsknecht, M.J., and Young, V.B. (2013). Fecal microbial transplantation for the treatment of
1103 Clostridium difficile infection: current promise and future needs. Curr Opin Gastroenterol *29*, 628-632.

1104 Konopka, J.B. (2012). N-acetylglucosamine (GlcNAc) functions in cell signaling. Scientifica (Cairo) *2012*.

1105 Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-
1106 359.

1107 Lechner, A., Brunk, E., and Keasling, J.D. (2016). The Need for Integrated Approaches in Metabolic
1108 Engineering. Cold Spring Harb Perspect Biol *8*.

1109 Lee, H., Popodi, E., Tang, H., and Foster, P.L. (2012). Rate and molecular spectrum of spontaneous
1110 mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. Proc Natl Acad
1111 Sci U S A *109*, E2774-E2783.

1112 Leroux, M., and Priem, B. (2016). Chaperone-assisted expression of KfiC glucuronyltransferase from
1113 Escherichia coli K5 leads to heparosan production in Escherichia coli BL21 in absence of the stabilisator
1114 KfiB. Appl Microbiol Biotechnol *100*, 10355-10361.

1115 Lescat, M., Launay, A., Ghalayini, M., Magnan, M., Glodt, J., Pintard, C., Dion, S., Denamur, E., and
1116 Tenaillon, O. (2017). Using long-term experimental evolution to uncover the patterns and determinants
1117 of molecular evolution of an Escherichia coli natural isolate in the streptomycin-treated mouse gut. Mol
1118 Ecol *26*, 1802-1817.

1119 Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and
1120 annotation of phylogenetic and other trees. Nucleic Acids Res *44*, W242-245.

1121 Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and
1122 population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987-2993.

1123 Lieberman, T.D., Flett, K.B., Yelin, I., Martin, T.R., McAdam, A.J., Priebe, G.P., and Kishony, R. (2014).
1124 Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of
1125 selective pressures. Nat Genet *46*, 82-87.

1126 Liu, Z., Ruan, Z., Xiao, Y., Yi, Y., Tang, Y.J., Liao, W., and Liu, Y. (2013). Integration of sewage sludge digestion
1127 with advanced biofuel synthesis. Bioresour Technol *132*, 166-170.

1128 Lourenco, M., Ramiro, R.S., Guleresi, D., Barroso-Batista, J., Xavier, K.B., Gordo, I., and Sousa, A. (2016). A
1129 Mutational Hotspot and Strong Selection Contribute to the Order of Mutations Selected for during
1130 Escherichia coli Adaptation to the Gut. PLoS Genet *12*, e1006420.

1131 Lutz, R., and Bujard, H. (1997). Independent and tight regulation of transcriptional units in Escherichia coli
1132 via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Res *25*, 1203-1210.

1133 MacDonald, A., Gokmen-Ozel, H., van Rijn, M., and Burgard, P. (2010). The reality of dietary compliance
1134 in the management of phenylketonuria. Journal of inherited metabolic disease *33*, 665-670.

1135 MacLeod, E.L., and Ney, D.M. (2010). Nutritional Management of Phenylketonuria. Annales Nestlé *68*, 58-
1136 69.

1137 Mansour, N.M., and Abdelaziz, S.A. (2016). Oral immunization of mice with engineered Lactobacillus
1138 gasseri NM713 strain expressing Streptococcus pyogenes M6 antigen. Microbiol Immunol *60*, 527-532.

1139 McKenna, R., and Nielsen, D.R. (2011). Styrene biosynthesis from glucose by engineered E. coli. Metab
1140 Eng *13*, 544-554.

1141 Moore, A.M., Ahmadi, S., Patel, S., Gibson, M.K., Wang, B., Ndao, M.I., Deych, E., Shannon, W., Tarr, P.I.,
1142 Warner, B.B*., et al.* (2015). Gut resistome development in healthy twin pairs in the first year of life.
1143 Microbiome *3*, 27.

1144 Nardecchia, F., Manti, F., Chiarotti, F., Carducci, C., Carducci, C., and Leuzzi, V. (2015). Neurocognitive and
1145 neuroimaging outcome of early treated young adult PKU patients: A longitudinal study. Mol Genet Metab
1146 *115*, 84-90.

1147 Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T., Fookes, M., Falush, D., Keane,
1148 J.A., and Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics *31*,
1149 3691-3693.

1150 Palmer, J.D., Piattelli, E., McCormick, B.A., Silby, M.W., Brigham, C.J., and Bucci, V. (2018). Engineered
1151 Probiotic for the Inhibition of Salmonella via Tetrathionate-Induced Production of Microcin H47. ACS
1152 Infect Dis *4*, 39-45.

1153 Patzer, S.I., Baquero, M.R., Bravo, D., Moreno, F., and Hantke, K. (2003). The colicin G, H and X
1154 determinants encode microcins M and H47, which might utilize the catecholate siderophore receptors
1155 FepA, Cir, Fiu and IroN. Microbiology *149*, 2557-2570.

1156 Pereira, F.C., and Berry, D. (2017). Microbial nutrient niches in the gut. Environ Microbiol *19*, 1366-1378.

1157 Pieper, D.H., and Reineke, W. (2000). Engineering bacteria for bioremediation. Curr Opin Biotechnol *11*,
1158 262-270.

1159 Plumbridge, J., and Pellegrini, O. (2004). Expression of the chitobiose operon of Escherichia coli is
1160 regulated by three transcription factors: NagC, ChbR and CAP. Mol Microbiol *52*, 437-449.

1161 Plump, A.S., Smith, J.D., Hayek, T., Aalto-Setälä, K., Walsh, A., Verstuyft, J.G., Rubin, E.M., and Breslow, J.L.
1162 (1992). Severe hypercholesterolemia and atherosclerosis in apolipoprotein E-deficient mice created by
1163 homologous recombination in ES cells. Cell *71*, 343-353.

1164 Porco, A., Peekhaus, N., Bausch, C., Tong, S., Isturiz, T., and Conway, T. (1997). Molecular genetic
1165 characterization of the Escherichia coli gntT gene of GntI, the main system for gluconate metabolism. J
1166 Bacteriol *179*, 1584-1590.

1167 Porse, A., Schonning, K., Munck, C., and Sommer, M.O. (2016). Survival and Evolution of a Large Multidrug
1168 Resistance Plasmid in New Clinical Bacterial Hosts. Mol Biol Evol *33*, 2860-2873.

1169 Rang, C.U., Licht, T.R., Midtvedt, T., Conway, P.L., Chao, L., Krogfelt, K.A., Cohen, P.S., and Molin, S. (1999).
1170 Estimation of growth rates of Escherichia coli BJ4 in streptomycin-treated and previously germfree mice
1171 by in situ rRNA hybridization. Clin Diagn Lab Immunol *6*, 434-436.

1172 Rakoff-Nahoum, S., Foster, K.R., and Comstock, L.E. (2016). The evolution of cooperation within the gut
1173 microbiota. Nature *533*, 255.

1174 Regier, D.S., and Greene, C.L. (2000). Phenylalanine Hydroxylase Deficiency. In Gene Reviews, M.P. Adam,
1175 H.H. Ardinger, and R.A. Pagon, eds. (Seattle (WA): University of Washington, Seattle).

1176 Reister, M., Hoffmeier, K., Krezdorn, N., Rotter, B., Liang, C., Rund, S., Dandekar, T., Sonnenborn, U., and
1177 Oelschlaeger, T.A. (2014). Complete genome sequence of the gram-negative probiotic Escherichia coli
1178 strain Nissle 1917. J Biotechnol *187*, 106-107.

1179 Reyes, A., Wu, M., McNulty, N.P., Rohwer, F.L., and Gordon, J.I. (2013). Gnotobiotic mouse model of
1180 phage-bacterial host dynamics in the human gut. Proc Natl Acad Sci U S A *110*, 20236-20241.

1181 Ridaura, V.K., Faith, J.J., Rey, F.E., Cheng, J., Duncan, A.E., Kau, A.L., Griffin, N.W., Lombard, V., Henrissat,
1182 B., Bain, J.R.*, et al.* (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice.
1183 Science *341*, 1241214.

1184 Rigg, G.P., Barrett, B., and Roberts, I.s. (1998). The localization of KpsC, S and T, and KfiA, C and D proteins
1185 involved in the biosynthesis of the Escherichia coli K5 capsular polysaccharide: evidence for a membrane-
1186 bound complex. Microbiology *144*, 2905-2914.

1187 Rodriguez-Beltran, J., Hernandez-Beltran, J.C.R., DelaFuente, J., Escudero, J.A., Fuentes-Hernandez, A.,
1188 MacLean, R.C., Peña-Miller, R., and San Millan, A. (2018). Multicopy plasmids allow bacteria to escape
1189 from fitness trade-offs during evolutionary innovation. Nature Ecology & Evolution.

1190 Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P.I., Godneva, A.,
1191 Kalka, I.N., Bar, N.*, et al.* (2018). Environment dominates over host genetics in shaping human gut
1192 microbiota. Nature *555*, 210-215.

1193 Sampson, T.R., Debelius, J.W., Thron, T., Janssen, S., Shastri, G.G., Ilhan, Z.E., Challis, C., Schretter, C.E.,
1194 Rocha, S., Gradinaru, V.*, et al.* (2016). Gut Microbiota Regulate Motor Deficits and Neuroinflammation in
1195 a Model of Parkinson's Disease. Cell *167*, 1469-1480.e1412.

1196 Sarkissian, C.N., and Gamez, A. (2005). Phenylalanine ammonia lyase, enzyme substitution therapy for
1197 phenylketonuria, where are we now? Mol Genet Metab *86 Suppl 1*, S22-26.

1198 Sarkissian, C.N., Gamez, A., Wang, L., Charbonneau, M., Fitzpatrick, P., Lemontt, J.F., Zhao, B., Vellard, M.,
1199 Bell, S.M., Henschell, C.*, et al.* (2008). Preclinical evaluation of multiple species of PEGylated recombinant
1200 phenylalanine ammonia lyase for the treatment of phenylketonuria. Proc Natl Acad Sci U S A *105*, 20894-
1201 20899.

1202 Sartor, R.B. (2005). Probiotic therapy of intestinal inflammation and infections. Curr Opin Gastroenterol
1203 *21*, 44-50.

1204 Scaldaferri, F., Gerardi, V., Mangiola, F., Lopetuso, L.R., Pizzoferrato, M., Petito, V., Papa, A., Stojanovic, J.,
1205 Poscia, A., Cammarota, G.*, et al.* (2016). Role and mechanisms of action of Escherichia coli Nissle 1917 in
1206 the maintenance of remission in ulcerative colitis patients: An update. World J Gastroenterol *22*, 5505-
1207 5511.

1208 Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *30*, 2068-2069.

1209 Smillie, C.S., Smith, M.B., Friedman, J., Cordero, O.X., David, L.A., and Alm, E.J. (2011). Ecology drives a
1210 global network of gene exchange connecting the human microbiome. Nature *480*, 241-244.

1211 Smith, M.I., Yatsunenko, T., Manary, M.J., Trehan, I., Mkakosya, R., Cheng, J., Kau, A.L., Rich, S.S.,
1212 Concannon, P., Mychaleckyj, J.C.*, et al.* (2013). Gut microbiomes of Malawian twin pairs discordant for
1213 kwashiorkor. Science *339*, 548-554.

1214 Sonnenborn, U., and Schulze, J. (2009). The non-pathogenic Escherichia coli strain Nissle 1917 – features
1215 of a versatile probiotic. Microbial Ecology in Health and Disease *21*, 122-158.

1216 Soucy, S.M., Huang, J., and Gogarten, J.P. (2015). Horizontal gene transfer: building the web of life. Nat
1217 Rev Genet *16*, 472-482.

1218 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
1219 phylogenies. Bioinformatics *30*, 1312-1313.

1220 Stapleton, A.E., Au-Yeung, M., Hooton, T.M., Fredricks, D.N., Roberts, P.L., Czaja, C.A., Yarova-Yarovaya,
1221 Y., Fiedler, T., Cox, M., and Stamm, W.E. (2011). Randomized, placebo-controlled phase 2 trial of a
1222 Lactobacillus crispatus probiotic given intravaginally for prevention of recurrent urinary tract infection.
1223 Clin Infect Dis *52*, 1212-1217.

1224 Sweeney, N.J., Klemm, P., McCormick, B.A., Moller-Nielsen, E., Utley, M., Schembri, M.A., Laux, D.C., and
1225 Cohen, P.S. (1996). The Escherichia coli K-12 gntP gene allows E. coli F-18 to occupy a distinct nutritional
1226 niche in the streptomycin-treated mouse large intestine. Infect Immun *64*, 3497-3503.

1227 Tailford, L.E., Crost, E.H., Kavanaugh, D., and Juge, N. (2015). Mucin glycan foraging in the human gut
1228 microbiome. Front Genet *6*, 81.

1229 Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: Genome-Wide Copy Number
1230 Detection and Visualization from Targeted DNA Sequencing. PLOS Computational Biology *12*, e1004873.

1231 Tamboli, C.P., Neut, C., Desreumaux, P., and Colombel, J.F. (2004). Dysbiosis in inflammatory bowel
1232 disease. Gut *53*, 1-4.

1233 Theriot, C.M., Koenigsknecht, M.J., Carlson, P.E., Jr., Hatton, G.E., Nelson, A.M., Li, B., Huffnagle, G.B., J,
1234 Z.L., and Young, V.B. (2014). Antibiotic-induced shifts in the mouse gut microbiome and metabolome
1235 increase susceptibility to Clostridium difficile infection. Nat Commun *5*, 3114.

1236 Thomas, C.M., and Nielsen, K.M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between
1237 bacteria. Nat Rev Microbiol *3*, 711-721.

1238 Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-
1239 performance genomics data visualization and exploration. Brief Bioinform *14*, 178-192.

1240 Timms, A.R., Steingrimsdottir, H., Lehmann, A.R., and Bridges, B.A. (1992). Mutant sequences in the rpsL
1241 gene of Escherichia coli B/r: mechanistic implications for spontaneous and ultraviolet light mutagenesis.
1242 Mol Gen Genet *232*, 89-96.

1243 Toivonen, J.M., Boocock, M.R., and Jacobs, H.T. (1999). Modelling in Escherichia coli of mutations in
1244 mitoribosomal protein S12: novel mutant phenotypes of rpsL. Mol Microbiol *31*, 1735-1746.

1245 Tramonti, A., Visca, P., De Canio, M., Falconi, M., and De Biase, D. (2002). Functional characterization and
1246 regulation of gadX, a gene encoding an AraC/XylS-like transcriptional activator of the Escherichia coli
1247 glutamic acid decarboxylase system. J Bacteriol *184*, 2603-2613.

1248 Tucker, D.L., Tucker, N., Ma, Z., Foster, J.W., Miranda, R.L., Cohen, P.S., and Conway, T. (2003). Genes of
1249 the GadX-GadW regulon in Escherichia coli. J Bacteriol *185*, 3190-3201.

1250 Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006). An obesity-
1251 associated gut microbiome with increased capacity for energy harvest. Nature *444*, 1027-1031.

1252 Vanderpool, C.K., and Gottesman, S. (2004). Involvement of a novel transcriptional activator and small
1253 RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system.
1254 Mol Microbiol *54*, 1076-1089.

1255 Vanderpool, C.K., and Gottesman, S. (2007). The novel transcription factor SgrR coordinates the response
1256 to glucose-phosphate stress. J Bacteriol *189*, 2238-2248.

1257 Vockley, J., Andersson, H.C., Antshel, K.M., Braverman, N.E., Burton, B.K., Frazier, D.M., Mitchell, J., Smith,
1258 W.E., Thompson, B.H., and Berry, S.A. (2014). Phenylalanine hydroxylase deficiency: diagnosis and
1259 management guideline. Genet Med *16*, 188-200.

1260 Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from
1261 high-throughput sequencing data. Nucleic Acids Res *38*, e164-e164.

1262 Wang, Y., Kuo, S., Shu, M., Yu, J., Huang, S., Dai, A., Two, A., Gallo, R.L., and Huang, C.M. (2014).
1263 Staphylococcus epidermidis in the human skin microbiome mediates fermentation to inhibit the growth
1264 of Propionibacterium acnes: implications of probiotics in acne vulgaris. Appl Microbiol Biotechnol *98*, 411-
1265 424.

1266 Wassenaar, T.M., Zschuttig, A., Beimfohr, C., Geske, T., Auerbach, C., Cook, H., Zimmermann, K., and
1267 Gunzer, F. (2015). Virulence genes in a probiotic E. coli product with a recorded long history of safe use.
1268 Eur J Microbiol Immunol (Bp) *5*, 81-93.

1269 Westendorf, A.M., Gunzer, F., Deppenmeier, S., Tapadar, D., Hunger, J.K., Schmidt, M.A., Buer, J., and
1270 Bruder, D. (2005). Intestinal immunity of Escherichia coli NISSLE 1917: a safe carrier for therapeutic
1271 molecules. FEMS Immunol Med Microbiol *43*, 373-384.

1272 Wexler, A.G., Bao, Y., Whitney, J.C., Bobay, L.-M., Xavier, J.B., Schofield, W.B., Barry, N.A., Russell, A.B.,
1273 Tran, B.Q., Goo, Y.A.*, et al.* (2016). Human symbionts inject and neutralize antibacterial toxins to persist
1274 in the gut. Proc Natl Acad Sci U S A *113*, 3639-3644.

1275 Wilkens, S. (2015). Structure and mechanism of ABC transporters. F1000Prime Rep *7*, 14.

1276 Winkler, J.D., Garcia, C., Olson, M., Callaway, E., and Kao, K.C. (2014). Evolved osmotolerant Escherichia
1277 coli mutants frequently exhibit defective N-acetylglucosamine catabolism and point mutations in cell
1278 shape-regulating protein MreB. Appl Environ Microbiol *80*, 3729-3740.

1279 Zhao, S., Lieberman, T.D., Poyet, M., Groussin, M., Gibbons, S.M., Xavier, R.J., and Alm, E.J. (2018).
1280 Adaptive evolution within the gut microbiome of individual people. bioRxiv.

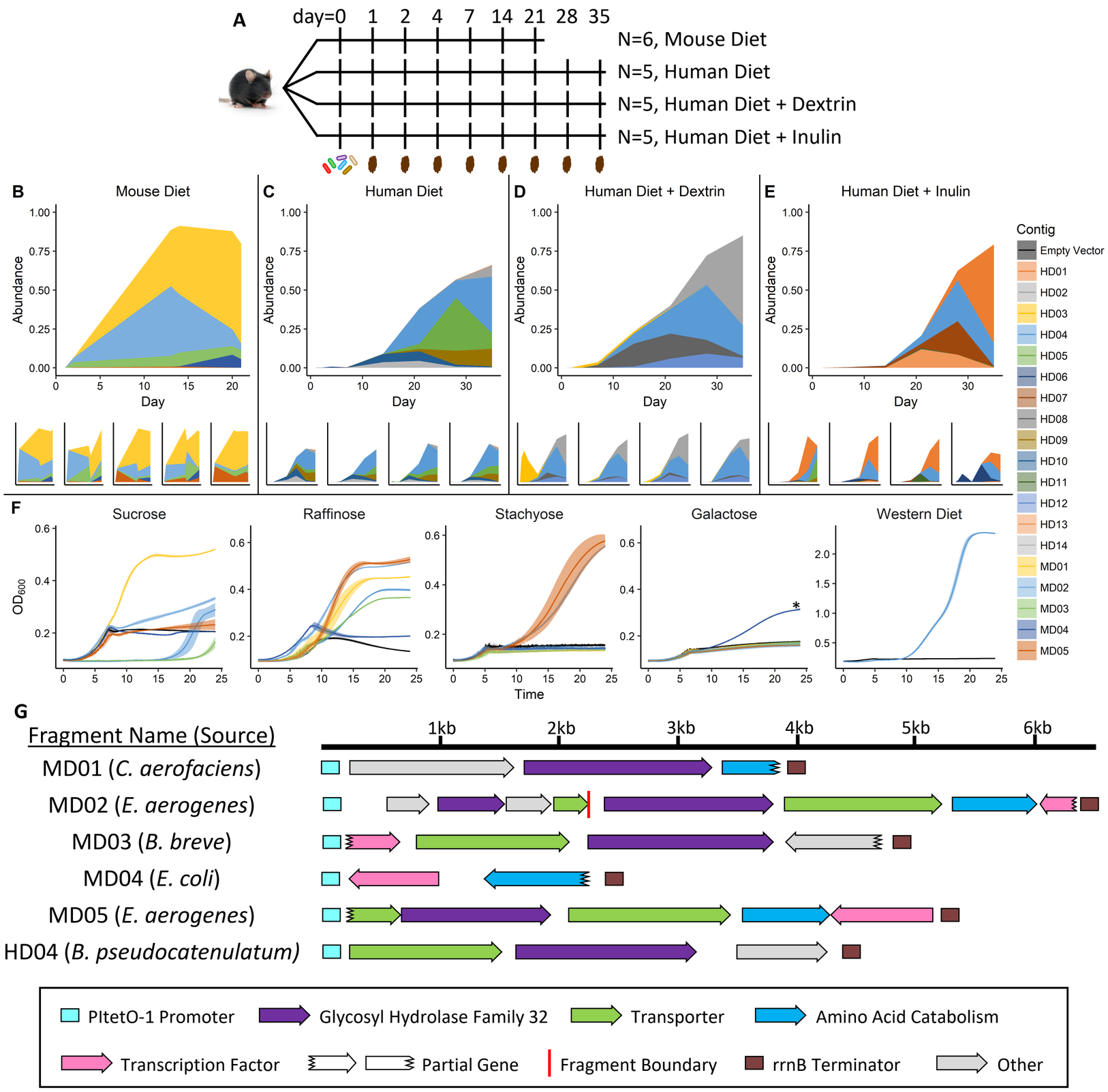

Figure 1. Functional metagenomic selections in EcN mono-colonized mice. A) Experimental design and sampling timeline. Mice were gavaged on day 0. N: number of replicate mice. **B-E)** Relative abundance of metagenomically sourced fragments in the EcN population in replicate mice fed standard mouse diet and water (B), a high-fat human diet and water (C), a high-fat human diet and dextrin (D), or a high-fat human diet and inulin (E). **F)** Growth curves of select metagenomic fragments enriched in the functional selections depicted in B-E on minimal media supplemented with different carbon sources. * P < 6x10-3, Student's t-test comparing time taken to reach OD_{600}=0.25 between WT and MD04. **G)** Annotations of the metagenomic fragments enriched in the functional selection depicted in B, and the generalist fragment enriched in the functional selections depicted in C-E. MD: fragments enriched in mice fed a Mouse Diet ; HD: fragments enriched in mice fed a Human Diet. See also Figures S1-S3, and Table S6.

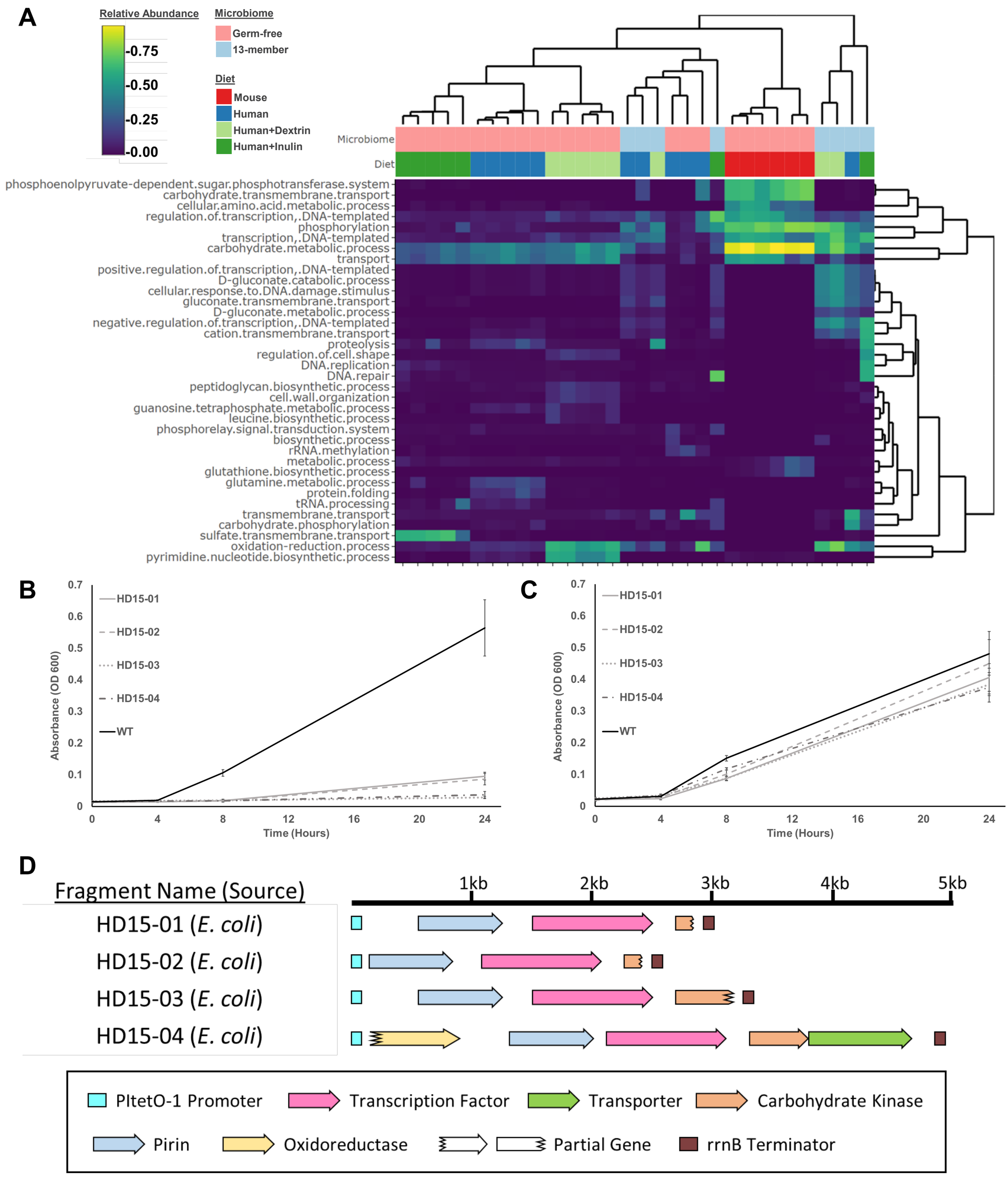

Figure 2. Comparison of functions enriched in functional selections in mice with different gut microbiome complexities. A) Heatmap of the relative abundances of metagenomic functions (GO classifications) enriched in EcN mono-colonized mice and mice pre-colonized with a 13-member defined community at the conclusion of the selection period. **B)** Growth curves of EcN strains harboring metagenomic inserts containing genes classified as a gluconate transmembrane transport or gluconate metabolic process (as depicted in A) on gluconate minimal media, or **(C)** glucose. **D)** Annotations of the metagenomic fragments characterized in B-C. Error bars represent the standard deviation of 6 replicates. See also Figure S1 and S3.

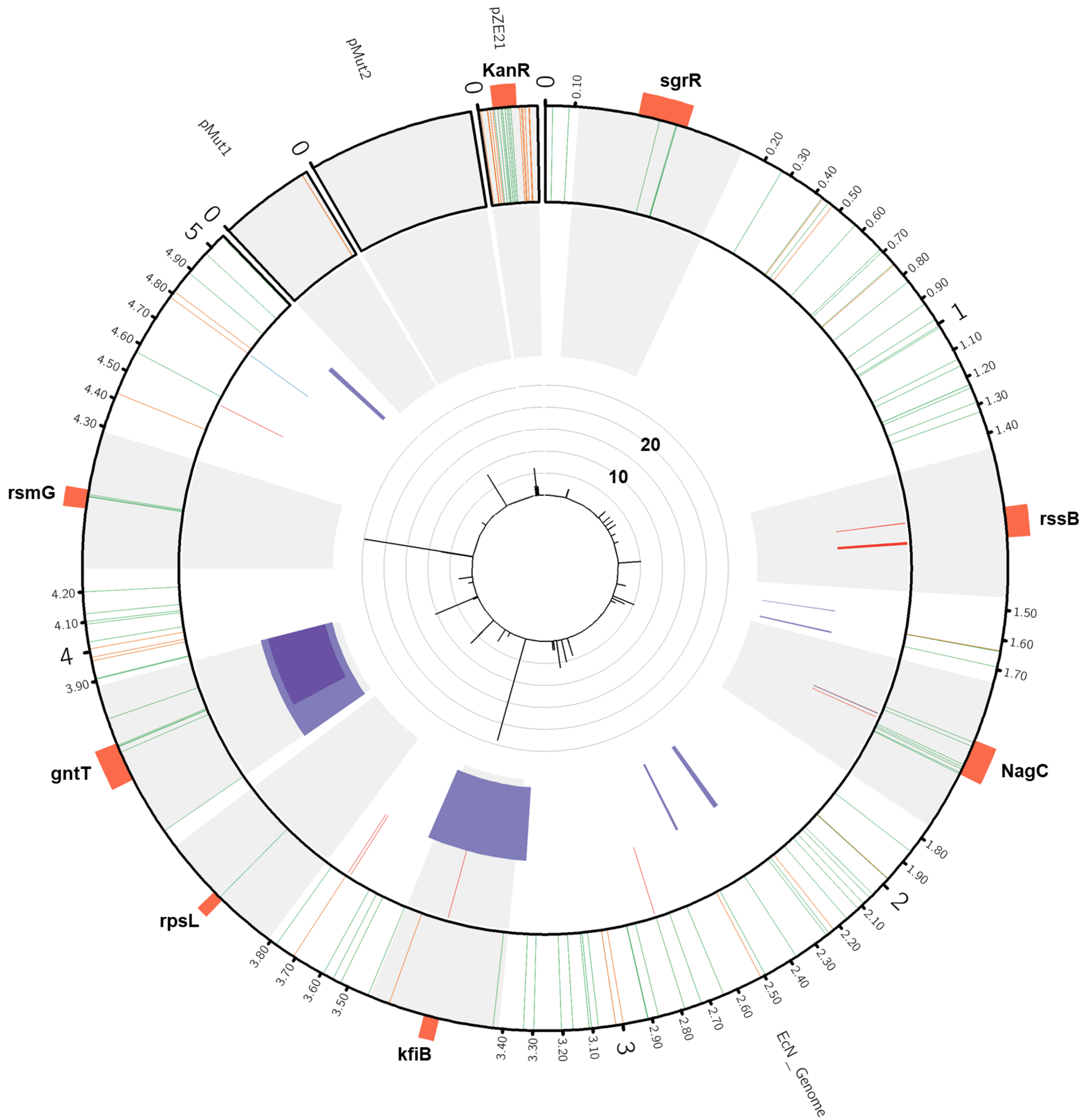

Figure 3. Summary of mutations detected in *in vivo* adapted EcN isolates. The EcN chromosome, its 2 native plasmids pMut1 and pMut2, and the expression vector pZE21 are depicted in a Circos plot. Light gray regions are on a scale 100x that of the rest of the chromosome for visibility. Genes of interest (orange) are labelled around the Circos plot. The outer track depicts all detected intergenic (orange) and nonsynonymous (green) SNPs. The next track depicts all small insertions (blue) and deletions (red). The third tack depicts all large deletions > 1kb (purple). Overlapping deletions are in dark purple. The inner track depicts the number of isolates that contained each mutation. Mutations found in only one isolate are excluded from this bar plot for visibility. Isolate counts for each mutation are capped at 25; this only affects the count for *rsmG* mutants, (167). See also Figure S4, Table S3, and Tables S5-S7.

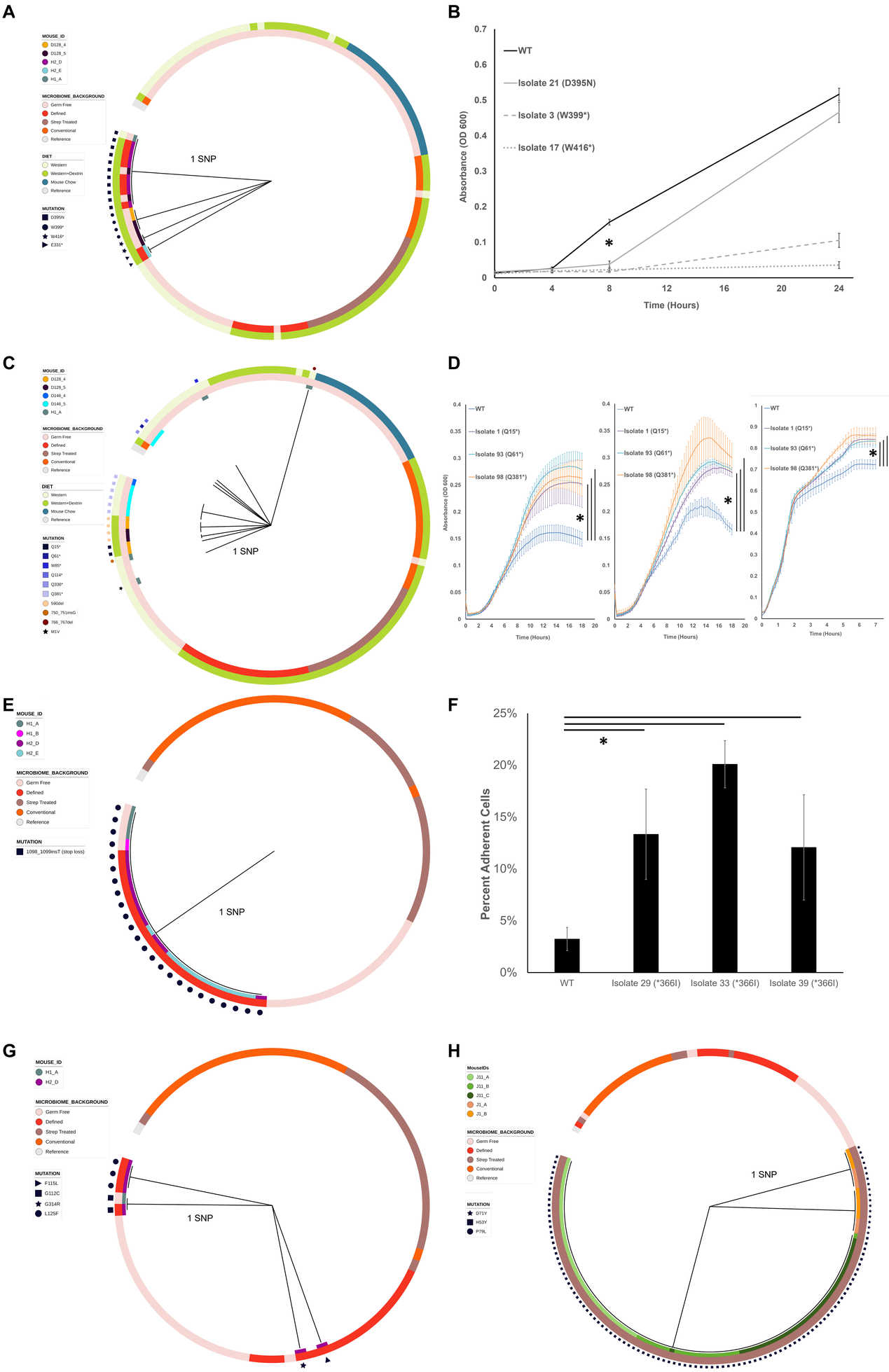

Figure 4. Phylogenies and phenotypic effects of mutant genes. Maximum parsimony phylogenies for adapted isolates based on their nucleotide sequences for **A)** *gntT*, **C)** *nagC*, **E)** *kfiB*, **G)** *sgrR*, or **H)** *rsmG*. Specific mutations are indicated. Color tracks depict the treatment conditions and the Mouse IDs from which isolates were isolated. The prefix of each Mouse ID indicates the cage in which the mouse was housed, e.g. Mouse H1_A and H1_B were housed together, but separately from Mouse H2_E. **B)** Growth curves of gntT mutants on gluconate minimal media. * P < 1x10-9, Student's t-test comparing WT against Isolate 21 at 8 hours. **D)** Growth curves of *nagC* mutants on N-acetylglucosamine minimal media (left), N-acetylglucosamine and glucose minimal media (middle), and porcine gastric mucin and glucose minimal media (right). * P < 0.05, Welch's t-tests comparing the area under the growth curves of each isolate and wild type, with Benjamini-Hochberg correction for multiple comparisons. Areas were determined by computing the integral of logistic models fit to the data for N-acetylglucosamine, and N-acetylglucosamine with glucose growth assays, or by numerically summing the area under the empirical growth curves for the gastric porcine mucin growth assay (growthcurvers R package). **F)** Percentages of *kfiB* mutant cells adherent to Caco-2 monolayers. Shown are the means of ratios of adherent CFU counts to total CFU counts (n = 3). Error bars are one standard deviation above and below the mean. * P < 0.05, Welch's t-test with Benjamini-Hochberg correction for multiple comparisons.

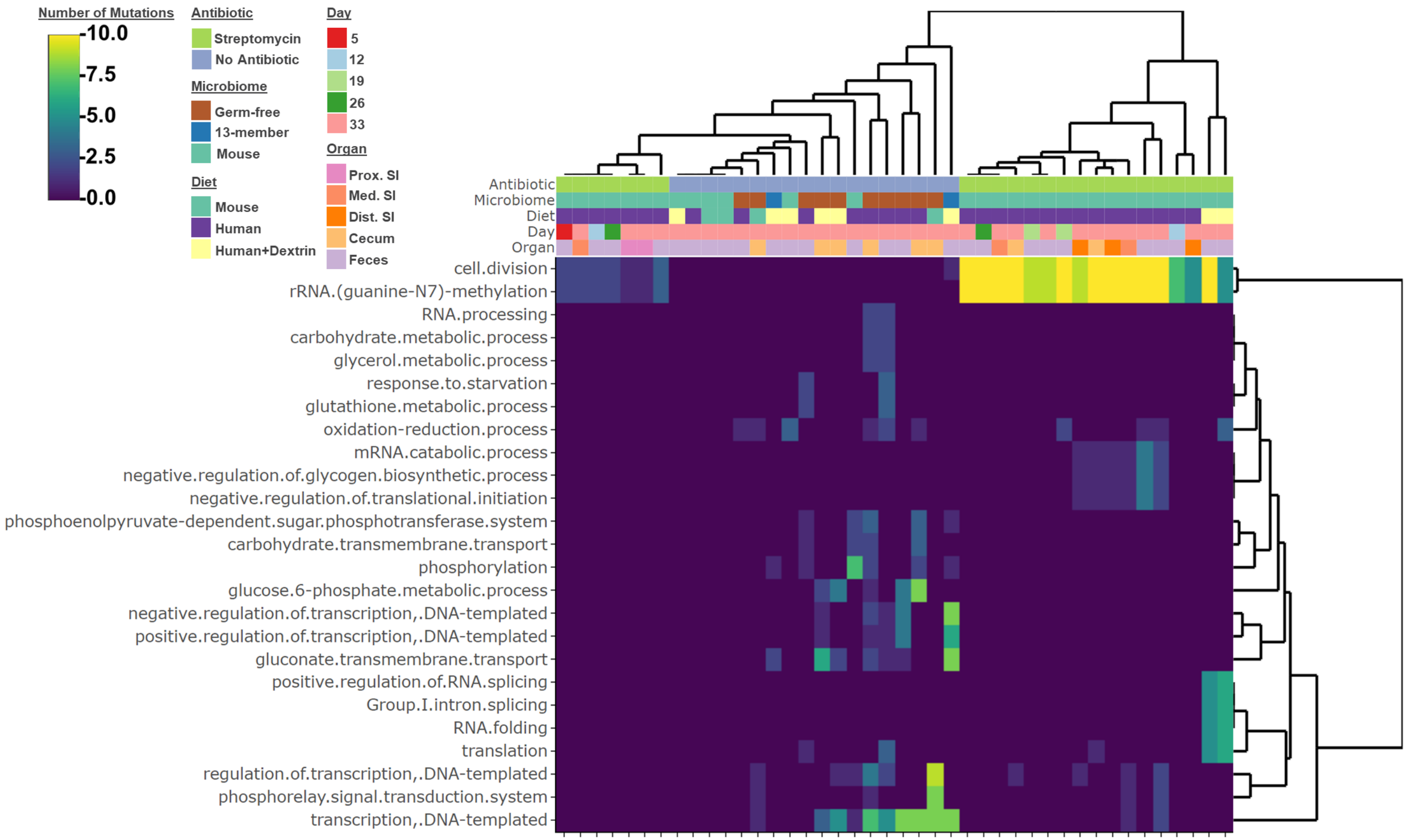

Figure 5. Comparison of functions mutated in EcN after passage through mice with a range of gut microbiome complexities and fed different diets. Heatmap of the cumulative number of nonsynonymous SNPs detected in genes grouped by their GO terms. Inclusion criteria for the heatmap was any function with at least two cumulative mutations identified across isolates from at least two mice. See also Figure S1.

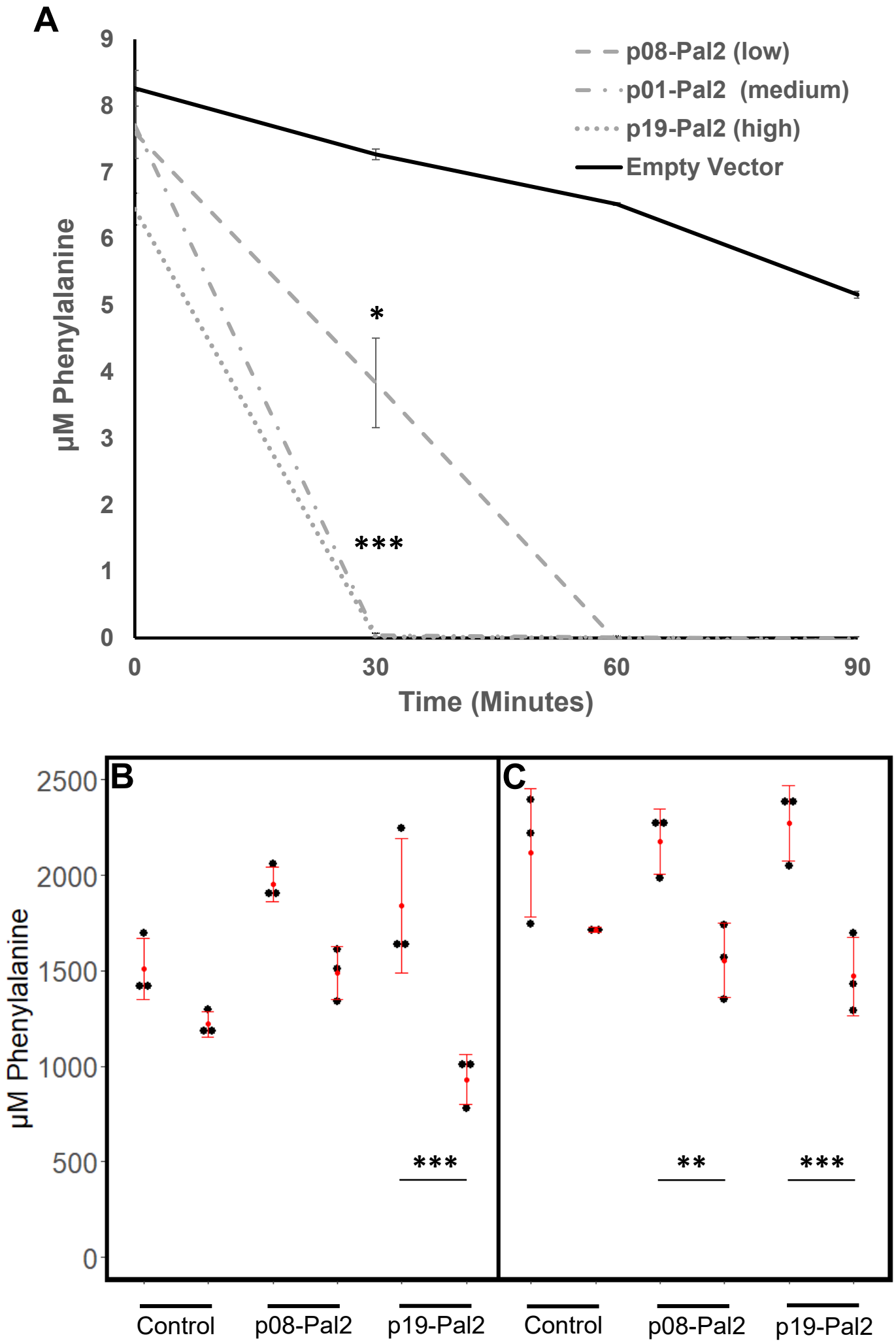

Figure 6. *In vitro* and *in vivo* degradation of phenylalanine by EcN:PAL2. A) *In vitro* phenylalanine concentration over time in minimal media supplemented with glucose and 10 µM Phe. Wells were inoculated with EcN expressing PAL2 under promoters of low, medium, or high strength. Shown are the mean of three technical replicates. Error bars represent one standard deviation above and below the mean. * $P < 0.01$, Welch's t-test with Bonferroni correction for multiple comparisons. **B-C)** Paired serum phe concentrations for male (B) and female (C) homozygous mutant Pah^{enu2} mice before and 24 hours after gavage with wild type EcN or EcN expressing PAL2 under a high (p19) or low strength promoter (p08). Shown are means and standard deviations for three replicate mice. *** $P < 0.0001$, ** $P < 0.01$; P-values were calculated using the generalized linear hypothesis test with Bonferroni correction (glht function from the multcomp R package) on the coefficient estimates from a linear mixed-effects model generated by regressing Phe measurements on an interaction factor comprised of mouse gender, measurement timepoint (before or after treatment), and treatment (lme function from the nlme R package). Random intercepts by mouse were specified in the model.