Title: Randomly barcoded transposon mutant libraries for gut commensals II: applying libraries for functional genetics

Authors: Carlos Geert Pieter Voogdt[1,3], Surya Tripathi[2], Stefan Oliver Bassler[1,4], Saria A McKeithen-Mead[6,7], Emma R Guiberson[7], Alexandra Koumoutsi[1], Afonso Martins Bravo[5], Cullen Buie[8], Michael Zimmermann[3], Justin Sonnenburg[7,9], Athanasios Typas[1,3,†], Adam Deutschbauer[2,10,†], Anthony L. Shiver[6,†], Kerwyn Casey Huang[6,7,9,*,†]

[1]Genome Biology Unit, EMBL Heidelberg, Meyerhofstraße 1, 69117 Heidelberg, Germany
[2]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA
[3]Structural and Computational Biology Unit, EMBL, Heidelberg, Germany
[4]Faculty of Biosciences, Heidelberg University, Grabengasse 1, 69117 Heidelberg, Germany
[5]Department of Fundamental Microbiology, University of Lausanne, CH–1015, Lausanne, Switzerland
[6]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA
[7]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305, USA
[8]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139 USA
[9]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA
[10]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

[†]Correspondence: typas@embl.de, amdeutschbauer@lbl.gov, ashiver@stanford.edu, kchuang@stanford.edu

[*]Lead contact: kchuang@stanford.edu

Keywords: transposon mutagenesis; chemical genomics; gut microbiota; genotype-phenotype mapping; host-microbe interactions

Abstract

The critical role of the intestinal microbiota in human health and disease is well recognized. Nevertheless, there are still large gaps in our understanding of the functions and mechanisms encoded in the genomes of most members of the gut microbiota. Genome-scale libraries of transposon mutants are a powerful tool to help us address this gap. Recent advances in barcoded transposon mutagenesis have dramatically lowered the cost of mutant fitness determination in hundreds of *in vitro* and *in vivo* experimental conditions. In an accompanying review, we discuss recent advances and caveats for the construction of pooled and arrayed barcoded transposon mutant libraries in human gut commensals. In this review, we discuss how these libraries can be applied across a wide range of applications, the technical aspects involved, and expectations for such screens.

Introduction

The human intestinal tract harbors an enormous diversity of microbial genes, many of whose functions remain completely unexplored[1-7]. Harnessing the full potential of the gut microbiome to improve health and counter disease will heavily depend on better understanding the function of (de-orphanizing) these microbial genes and how they link to other genes in the cellular network[8,9]. Exploration of the vast space of microbial genes demands the application of high-throughput systems biology approaches to a wide variety of organisms. For organisms that are genetically tractable, random transposon mutagenesis offers a robust and relatively inexpensive approach to uncovering genotype-phenotype relationships at genomic scale[10-16]. Transposon mutant libraries have proven powerful tools for rapidly screening genetic perturbations for phenotypes under various environmental conditions[11,17], and thus often form the basis for mechanistic discovery of gene functions. Up to this point, such libraries have typically been constructed in model organisms and pathogens[10,14,18-22], but the expansion and maturation of the underlying genetic tools to phylogenetically diverse microbes enables their broader application to gut microbes.

Transposon mutagenesis is typically used to create dense libraries with insertions in most (if not all) non-essential genes in the target organism. In a pooled format, these mutant libraries are coupled with sequencing methods to quantify the relative fitness of each mutant in the pool across a set of environments[10,11,17,23]. Such pooled libraries can also be used to construct non-redundant arrayed libraries[10,24,25], which requires considerable effort but provides an invaluable tool for easily accessing mutant strains. Such access facilitates validation of results from pooled screens and further molecular investigation of the underlying mechanism. It also enables studying single-cell and/or fitness-independent phenotypes, such as cell morphology, biofilm formation, extracellular metabolism, adhesion, metabolite secretion, and many others[26,27]. In random barcode transposon-site sequencing (RB-TnSeq[12]), each transposon carries a random DNA barcode linked to its genomic insertion position. Once the barcode-insertion linkages are mapped via sequencing of the initial library, mutant fitness can be estimated simply from relative barcode abundance as determined by sequencing of the barcodes (Bar-Seq)[12]. As a result, barcoding greatly reduces the time, effort, and cost of pooled library sequencing[12]. Moreover, the location of barcoded transposon mutants in an arrayed library can be determined more easily and with greater accuracy through barcode amplification[28].

Pooled or arrayed mutant libraries enable rapid genome-scale *in vitro* screening across a wide variety of conditions, providing phenotype profiles for each gene in a strategy known as “forward genetics”[11,12,14,29,30]. Such profiles can link genes to phenotypes, and grouping genes with similar phenotypic profiles reveals gene-gene links and higher order genetic networks. Since *in vitro* screening of mutant libraries is reasonably scalable, a large number of conditions can be tested to broadly probe the phenotypic landscape. For example, two decades of work with the *Escherichia coli*

KEIO collection[18], an arrayed library of deletion mutants that covers the non-essential genome, has provided the first phenotypes for hundreds of genes[29,31-35] and served as a basis for dissecting gene function. Similar success has been achieved in the budding yeast *Saccharomyces cerevisiae*[7]. Yet, even for well-studied model organisms such as *E. coli* and *Bacillus subtilis*, a substantial fraction of genes (>25%) have no clear functions or phenotypes[36,37], suggesting the need to expand the space of screening conditions and explore more unorthodox perturbations.

Ultimately, the goal of genotype-phenotype mapping is to understand the role of each gene in the natural habitat and context in which the organism lives. Transposon mutant libraries of pathogens have been used to identify key genes important for virulence in animal infection models[38-40]. Other *in vivo* studies have identified genes in gut commensals involved in animal host colonization and nutrient utilization, within a community of other microbiome members or when colonizing germ-free mice alone[10,14,23]. However, the inherent complexity of the gut ecosystem, involving a large number of species and interspecies interactions as well as contributions of host factors (e.g., diet, immune status, spatial localization within the gut), means that more conditions must be studied to reveal the phenotypes and functions of genes in gut commensals. Further *in vitro* and *in vivo* experimentation, with transposon mutant libraries that represent a broader range of gut commensals, will be needed to understand key representative gene functionalities in the gut environment.

In an accompanying review[41], we discuss important considerations and strategies for constructing and arraying barcoded transposon mutant libraries, as well as prioritization of organisms for future library construction. In this review, we discuss the design and structuring of *in vitro* and *in vivo* functional genetic experiments with pooled and arrayed barcoded transposon mutant libraries. We propose testing conditions that will broaden and maximize the power of chemical genomics. We then focus on the translation from *in vitro* data to *in vivo* phenotypes in popular model animal hosts. Finally, we discuss how to facilitate global knowledge dissemination of phenotypic screening data, to accelerate gene function discovery in the gut microbiome.

Technical aspects of *in vitro* screening of mutant libraries

In vitro screening of mutant libraries is reasonably high throughput and cost effective, hence screens typically involve measurements of phenotypes across hundreds of conditions[11,14,29]. These fitness measurements can directly link the phenotype of a given gene disruption to a condition (e.g., a glycan transporter mutant does not grow in media with only that glycan as a carbon source), and indirectly link genes that share the same phenotypes across conditions (a strong indication of genes operating as part of the same functional unit). While screening libraries *in vitro* (Figure 1) is almost always faster and cheaper than *in vivo* screens, enabling the exploration of more diverse conditions, the relevance and translatability of phenotypes observed *in vitro* to *in vivo* systems should be considered.

Libraries can be screened in two modes. Positive selection employs conditions such as phage predation, toxins, and chemicals that have large negative impact on the wild-type strain[42], enabling outgrowth of a small number of mutants with higher fitness in these conditions (e.g., resistance). Negative selection employs conditions such as growth on carbon and nitrogen sources, pH, osmolarity, and sublethal concentrations of toxins and antibiotics that deplete a small subset of mutants due to their lower fitness compared to the large fraction of unaffected mutants. Positive-selection screens require fewer reads than negative-selection screens (depending on the condition, sequencing depth can be at least four times lower), as the pool of remaining mutants is less complex and thus easier to sequence. Yet, in positive-selection screens only the strongest phenotypes are selected and identified. Negative selection requires deep sequencing to accurately quantify mutant abundance and thus is more costly, but the higher resolution reveals mutants with more subtle fitness defects.

Growth medium

The selection of (species-specific) growth media can strongly influence phenotypes. For example, screening of a *Bacteroides thetaiotaomicron* mutant library revealed genes whose disruption provides a fitness advantage during treatment with the antibiotic vancomycin when grown in the complex, undefined Brain Heart Infusion (BHI) medium but not in a minimal, defined medium[14]. Although many of the known microbial phyla do not have a single cultured representative[43], most members of the gut microbiota can be cultured in complex, undefined media such as GMM, Mega medium, BHI, and GAM[44]. These media can be used as a robust base for screening a library against numerous conditions (e.g., drugs, pH, osmolarity), but their complexity is prohibitive for screening metabolism-related functions. The use of rich, undefined media can also obscure phenotypes due to transcriptional feedback on physiological systems such as carbon catabolite repression[35,45].

If the target organism can be grown in a defined medium such as Varel Bryant minimal media designed for *Bacteroides* species[46] or other recipes (see e.g., refs. [44,47]) that support robust growth, then medium compositional changes can be used to identify pathways relevant to carbon/nitrogen utilization, breakdown of complex substrates,

and biosynthesis of amino acids, nucleotides, vitamins, and co-factors. Typically, phenotypes are stronger under nutrient-limited conditions such as minimal media, especially for metabolism, import, and core physiology[48]. It may be preferential to screen mutant libraries of distantly related organisms in the same medium to shed light on their potential functions in a community. While it may be difficult to establish a "one-size-fits-all" medium applicable across phyla, recent efforts have identified a few minimal media that can be used to culture most gut commensals[44].

Inoculum size and culture volume

The experimental setup for library screening can influence the robustness and resolution limit of fitness measurements. It is critical to inoculate cultures with a sufficiently large population to avoid bottlenecks and sampling artifacts that may bias library growth; for a library with 10^5 mutants, the inoculum should ideally be about 10^7 cells (which corresponds to a standard cuvette optical density of ~0.01 for bacteria with similar size as *E. coli*). For the same starting number of cells, larger culture volumes support more generations of growth than smaller volumes, making subtle fitness defects more apparent, but running the risk that mutants with moderate or stronger fitness effects will decrease in abundance below the limit of detection. For example, the fraction M of any mutant in the total pool P of mutants at a given time t is $M(t)/(M(t)+P(t)) = 2^{fg}/(2^{fg}+2^{g})$, where f is the fitness of the mutant relative to that of the pool of mutants (which will typically respond like wild type and thus have a relative fitness of 1) and g is the number of generations. Over the course of 7 generations (128-fold expansion of the inoculum), the fraction of a mutant with a growth rate 10% lower than the rest of the pool (relative fitness = 0.9), would be 38% of that of an unaffected mutant, while over 10 generations (1024-fold expansion) the mutant fraction would decrease to 33%. In contrast, a mutant with a 50% growth defect (relative fitness = 0.5) will decrease to 8% compared to an unaffected mutant after 7 generations and to 3% after 10 generations, and thus would be much harder to detect.

Ultimately, the choice of inoculum size and growth format/volume (e.g., 96-well versus 24-well plate) for screening will depend on the diversity of the library and on the conditions tested, and should be balanced against other limitations such as scaling factors (e.g., larger culture volumes require more compound, which increases costs). As a general guideline, typical *in vitro* screening for growth of a library on specific nutrients or in the presence of chemicals can be performed in deep 96-well plates with 2 mL of media for about 6-8 generations (64- to 256-fold expansion)[11,12,14].

Timing of library experiments and DNA extraction

Differences in the growth lag time among mutants may skew library dynamics and the final results. These effects can be minimized by growing the library under standard conditions (e.g., the condition under which the library was generated) to early log phase before exposing the library to screening conditions. A sample of the library should be collected before exposure to the screening conditions to enable comparison of mutant abundances after exposure. For convenience, libraries are typically grown

to saturation (stationary phase) in the presence of the perturbation before cells are collected for DNA extraction. However, large fitness differences during stationary phase or survival differences may obscure subtle fitness effects during log phase specific to the perturbation of interest, which may be the primary focus. To avoid such cases, it may be preferable to first determine the growth rate of the wild-type strain in the conditions of interest and based on this information, to grow the library for a fixed number of generations below saturation and harvest DNA before stationary phase is reached (although such a practice would substantially increase the effort of screening). Knowing the growth rate in each condition can also be used to cluster conditions with similar growth rates on the same assay plate for better timing and more straightforward handling. After growth of the library for the desired amount of time, cells are pelleted and DNA extraction can be simplified by the use of commercial kits designed for microbial communities in multi-well plates. However, these kits are expensive (several dollars per well), in part justified by their ability to evenly lyse a broad phylogenetic range of species[49]. For screening a single strain, simpler commercial kits or custom methods exploiting liquid handling robotics can lower the costs.

Sequencing depth and cost

The appropriate number of reads per condition will depend on the number of barcoded strains in the library (diversity) as well as the fraction of unique barcodes. If sequencing results in too few reads, estimates of the relative abundance of mutants will be noisy and hence unreliable. Conversely, very high read counts are unnecessary and the sequencing depth could instead be spread over more conditions. Rough estimates of the appropriate sequencing depth can be calculated using a naive power analysis, as was recently described for CRISPRi-based sequencing of guide RNAs[50], yet to determine the optimal sequencing depth that balances these considerations, intrinsic biases in the distribution of mutants across genes should be taken into account, as these biases may necessitate increased sequencing depth to accurately measure the fitness of less abundant barcodes. Moreover, fluctuations in depth per condition for a given sequencing flow cell are an inevitable bias that either requires resequencing of the conditions that had below average reads by chance or a reduction in the number of conditions per flow cell.

As a rough estimate, an Illumina NextSeq500 Mid-Output flow cell generates ~120 million reads, which is typically sufficient for ~40 samples screening a library of 300,000 mutants representing ~5,000 ORFs. Such a library will on average have ~36 insertions in the central 20-80% of each ORF (ignoring intergenic insertions). With 3 million reads per sample mapping to unique barcodes, each insertion would be represented by ~10 reads on average. Ten reads are not sufficient to analyze fitness for each mutant individually, hence reads mapping to insertions in the same gene must be summed to accurately quantify gene-level fitness. In this case, with ~360 reads per gene on average, the effects of a 40% growth defect (relative fitness = 0.6) accumulated over 7 generations will lower the abundance of this gene to ~5% and its cumulative read count to ~18. Such estimations can be useful for designing sample

multiplexing with the required sequencing depth, but it should be noted that library-specific biases (that must be determined empirically) can considerably skew the actual numbers. With increased reads per sample (greater sequencing depth or fewer samples multiplexed), it may not be necessary to sum counts of all mutants on a per-gene basis and instead quantify fitness on a mutant-by-mutant basis. Such a strategy could provide more statistical power since each mutant is treated independently and thus aberrant mutants (e.g., that acquired a secondary mutation driving the phenotype rather than the transposon insertion) can be identified and removed from calculation of the median fitness across mutants in a gene).

Starting with libraries with more controlled population size (e.g. re-pooled, non-redundant arrayed libraries) can increase dramatically the throughput. The total cost of screening consists of plastic ware, media, chemicals/nutrients for testing, DNA extraction, library preparation, and sequencing, which all scale linearly with the number of technical replicates. Of these factors, plastic ware and media are relatively inexpensive when screens are performed in a 96-well format. As sequencing costs continue to decrease, a large fraction of the expense will typically be represented by certain conditions (e.g., antineoplastic drugs or host-relevant molecules, such as mucin) and library preparation. Thus, users should increasingly avoid economizing on sequencing and instead aim for more reads per sample than necessary as compared to maximizing multiplexing of samples.

In vitro screening conditions relevant to gut bacterial physiology

Intestinal bacteria are exposed to a wide variety of conditions and stresses *in vivo*, some of which can be mimicked *in vitro*. While the relevance of some conditions is more specific to certain target organisms, it is likely worthwhile to screen a broad range of common conditions/perturbations to maximize the chances of discovering phenotypes for genes of unknown function.

Nutrients and metabolism

To identify genes involved in catabolic or anabolic pathways, the library should be grown in a defined medium in which molecules of interest are left out or added in excess. Basic molecules to screen include amino acids**,** nucleotides**,** short chain fatty acids, and trace elements like metals (which cannot be synthesized) and vitamins. For some organisms, certain nutrient classes are natural candidates for screening; for example, many *Bacteroides* species forage on host mucus or degrade complex carbohydrates[47], thus screening *Bacteroides* libraries on a diverse panel of glycans can identify genes involved in their complex carbohydrate catabolic capacities. Identifying microbiome genes involved in prebiotic carbohydrate utilization has strong relevance to health[51-53]. Carbohydrates of interest include human and animal milk oligosaccharides**,** complex polysaccharides from diverse plant sources such as inulin from chicory[54] and glucosinolates from broccoli[55], and beta-glucans from fungal cell walls[56].

Environmental/abiotic factors

The gastrointestinal tract poses a range of physical challenges that force microbes to adapt. Variation in pH along the intestines motivates screening of growth at starting pH ranging from 4-10[57,58], with and without a buffer to counteract the ability of some organisms to modify the environmental pH[59,60]. Sensitivity of growth to high osmolarity may explain the effects of osmotic diarrhea on the gut microbiota[61], motivating library screening across concentrations of non-metabolizable osmolytes; while salt has often been used as an osmolyte, its indirect electrostatic effects are not optimal[62] and thus sugar alcohols such as sorbitol may be more appropriate[63]. The gut lumen can exhibit variable viscosity, for instance due to mucus release, which may impact microbial growth, localization, and transit[64]. The effects of viscosity on growth *in vitro* can be studied by adding various concentrations of polyethylene glycol or glycerol. Finally, an obvious environmental feature of the gut lumen is anoxia; screening libraries in various oxygen concentrations may reveal genes involved in oxygen sensitivity[14].

Host factors

Intestinal bacteria interact intimately with the host, and identifying phenotypes related to host-derived signals can provide insight into host-microbe relationships. Bacteria can acquire nutrients from the host, for example by foraging on mucus. Mucus is a complex mixture of glycosylated proteins that are secreted by goblet cells and can be tethered to the epithelial membrane[65]. Porcine gastric mucus is often used for *in vitro* microbiota studies as it is relatively inexpensive, although it should be noted that mucus composition and properties vary along the gastrointestinal tract and gastric mucus may be a poor model of mucus from the small or large intestine[66]. Enterocytes shed from intestinal villi may also provide a highly complex source of nutrients to intestinal microbes[67] that can be mimicked *in vitro*. Host immune factors such as antimicrobial peptides[10], bile salts[68], immunoglobulins (in particular secretory IgA)[69], and hormones play a key role in shaping the microbiota, and the genes that allow gut bacteria to sense and cope with these factors largely remain to be elucidated.

Other microbiota members

Bacteria in the gut are exposed to many other microbial species, with some of which they interact directly; these interactions may be particularly important in the presence of species whose niche overlaps highly with the target organism[70]. Screening a library for phenotypes at various levels of complexity of the surrounding microbial community, including phages, protists, and fungi as well as other bacterial species, can elucidate the genetic basis for key questions such as microbiome stability, colonization, and use of and defense against antagonisms. To probe these questions, a library can be grown in either co-culture or in the spent supernatant of other microbes to identify phenotypes involved in cross-feeding or sensitivity to released molecules. One disadvantage of screening libraries in a pooled format is that mutants in cell-autonomous phenotypes (e.g., secretion of an autocrine signal) can be complemented by other mutants or strains in the pool and thus their fitness will not be compromised. As with host-derived nutrients, microbe-derived nutrients are usually complex mixtures that may result in

multiple phenotypes, making interpretation challenging. One way to address this issue can be through fractionation of such complex mixtures coupled to metabolomics data to enable linking phenotypes to specific metabolites. Even in the absence of clearly interpretable phenotypes, the ability to quantify fitness across many conditions provides the power to link genes together. The impacts of a natural gut microbiota can be studied by co-culturing libraries with highly diverse synthetic[10,71] or stool-derived communities[72], which have been shown to recapitulate many aspects of the gut microbiota *in vivo*. Such assays have the potential to reveal phenotypes involved in direct cell-cell interactions or ones that emerge from a given community context. The increased complexity of community assays requires consideration of several potential issues: other species may create a severe bottleneck for the focal library species, decreasing its growth rate and yield. In this case, the initial inoculum of the focal species and the assay time would need to be adjusted.

Xenobiotics

Gut bacteria are exposed to diverse xenobiotics (compounds foreign to the body) ingested by the host or released in bile. Such compounds can impact microbial fitness and community composition, and may drive the evolution of resistance mechanisms[73,74]. Xenobiotics relevant for the gut microbiome include antibiotics, human-targeting drugs, food additives, toxins, and excipients (support substances that serve as the vehicle for a drug). Screening of libraries can enable associations between xenobiotics and other conditions based on a common phenotypic profile across mutants. Many companies now sell standard or custom-arrayed compound libraries that can be used for library screening. Prescreening of these compounds on the wild-type strain is important, as the target organism will not be affected by some (potentially most) compounds, and those compounds are less interesting for further screening. A target concentration for screening is the IC_{50}, the concentration at which wild-type growth is reduced by 50%, which enables the identification of both more sensitive and more resistant mutants. Alternatively, the library can be grown across a range of compound concentrations to determine the concentration at which library growth is partially hampered and hence is appropriate for fitness measurements, avoiding the time required for prescreening, at the cost of re-arraying the appropriate samples for sequencing preparation. To obtain a more targeted selection of xenobiotics for screening, recent studies on the impact of medication on the microbiota[73] can be mined to identify compounds that affect the abundance of the target organism.

Applications of arrayed libraries

In pooled form, mutant libraries can be used to mostly screen for conditions that affect fitness, in which change in relative abundance measures the impact of the condition on each mutant's genotype. While such screens have generated powerful insights into bacterial physiology, certain conditions and behaviors are difficult if not impossible to probe in a pooled library format due to trans-complementation of cell-autonomous deleterious phenotypes by other mutants in the pool. Such conditions and behaviors

include: cross-feeding, in which degradation of nutrients and/or the release of waste products by conspecific or heterospecific cells alleviates genetic defects; degradation of drugs (including antibiotics), which lowers the effective concentration and hence modifies the interpretation of mutant sensitivity; and secretion of toxins, enzymes, vesicles, or signaling molecules that end up as community property. Droplet TnSeq (dTnSeq[75,76]), which uses microfluidics-based encapsulation of single cells, has been developed to expand measurable phenotypes beyond fitness using pooled libraries. Moreover, arrayed libraries enable the study of non-growth-related phenotypes and/or single-cell phenotypes such as changes in cell morphology, biofilm formation, and the intracellular and extracellular metabolome and proteome. These and other key aspects of bacterial physiology can justify the effort required to array the pool into a non-redundant collection of mutants.

In the accompanying review, we discuss recent advances that have dramatically lowered the barriers to arrayed library construction. With an arrayed library, single-cell readouts, such as shape defects[27,77,78] or protein stability and abundance[79], or non-growth-related phenotypes, such as biofilm formation[80] or survival in stationary phase[81], have been probed. In addition, an arrayed library can be used to create smaller sub-libraries that focus on mutants related to a specific process, such as metabolism or stress responses. Moreover, due to their lower complexity and/or more balanced coverage of the genome, sub-libraries enable higher throughput and avoid population size bottlenecks, and hence allow for testing of more conditions. Importantly, for many bacterial species (particularly gut commensals), genetic tools are still lacking. Transposon vectors do not require maintenance in the cell (and thus knowledge about maintenance systems is not necessary) so for many organisms transposon vectors are currently the only accessible starting tools for genetic manipulation[82,83]. Thus, arrayed transposon mutant libraries in otherwise genetically intractable microbes provide a highly valuable resource for mechanistic investigation of genotype-phenotype relationships.

Screening of a standard panel of conditions may elucidate strain/species-specific versus conserved genotype-phenotype relations. Yet, even for just the classes of perturbations and conditions mentioned above, the chemical space is enormous and trade-offs between relevance, coverage, and feasibility/costs will be necessary. Moreover, our list is by no means exhaustive for probing gene functions in gut bacteria using mutant libraries. Depending on the target organism, specific screening conditions relevant to microbe-specific lifestyles should be considered.

Expected results from *in vitro* screening

In vitro screening of pooled or arrayed libraries will typically enable identification of phenotypes for many (although not all) genes. General expected behaviors would be lower relative abundance for required/beneficial genes and higher relative abundance for detrimental/toxic genes under the probed condition. For example, when only one of several transporter mutants has a fitness defect in minimal media with glucose as

the carbon source, that transporter can be linked to primary glucose metabolism. However, interpretation of phenotypes may be more difficult if there is redundancy involving two transporters; in this case, more complex conditions may be required to dissect function. For an antibiotic, the exporter or efflux pump transposon mutant would show lower relative abundance due to increased cellular drug concentration[14] and the transposon mutant of the porin used for drug uptake would show higher relative abundance due to decreased cellular drug concentration.**Error! Bookmark not defined.**

A major benefit of high-throughput library screening is the ability to reveal mutants in different genes that exhibit the same behavior across conditions[11]. Results from *in vitro* screening may lead to reannotation of certain genes[14]. Even genes without any known function can be implicated in a genetic network due to their association with other genes of known function with similar fitness in the same conditions ("co-fitness"). The strength of the gene co-fitness metric depends on both the number and orthogonality of conditions tested. Care should be taken to avoid undue bias in the distribution of conditions (e.g., due to screening a large number of antibiotics with the same target), as such bias can emphasize certain conditions in phenotypic correlations and thus discount signal from other conditions. A metric that systematically clusters and normalizes results from diverse conditions to enable high confidence correlation estimates would substantially improve detection of gene-gene linkages.

Screening libraries for phenotypes from a "gene-centric" viewpoint can be an overwhelming endeavor since it is unclear whether useful estimates of practical screening scale (i.e. the number of conditions) can be derived from features such as behavior (e.g., growth capacity across different media), genome size (gene count), genetic network complexity, or predicted enzymatic capacity. Yet, automated microbiology platforms powered by artificial intelligence show promise for easing this challenge[84]. Alternatively, library screening can also be approached from a "condition-centric" perspective, in which specific conditions of interest are probed for any mutant phenotypes. This design inherently constrains the number of conditions, making library screening more feasible, at the cost of ignoring the unknown unknowns. Ideally, at some point a consistent framework for future *in vitro* screening will emerge (be it gene-centric or condition-centric) following the analysis of screening many phylogenetically diverse organisms. Ultimately, for gut microbes a major consideration is how to position *in vitro* screening results to aid interpretation of *in vivo* experiments.

In vivo screening of mutant libraries

Although *in vitro* experiments with barcoded transposon mutant libraries can be carried out at a throughput, relatively low cost, and scale that are typically inaccessible using *in vivo* models, when studying gut bacteria it can be difficult to sufficiently model the intestinal environment *in vitro*. Thus, *in vivo* experimentation is a critical complement to fully understand the physiological role of bacterial genes, including nutrient acquisition through diet and competition, direct agonism/antagonism with other

organisms and phages, and the impact of the host (e.g., inflammation) and xenobiotics on bacterial fitness (Figure 2). Hosts that can be made germ-free enable library screening in a wide range of controlled colonization conditions, in particular to ensure that the library can colonize at high density and not be outcompeted by other species.

Pre-colonization

For a typical *in vivo* experiment, the library is grown to saturation prior to colonization of the host animal[14]. Care should be taken not to grow the library for too long after saturation, since the degree of starvation can affect *in vivo* fitness[85]. Alternatively, to prevent loss of mutants with severe fitness defects, the library can be grown for a limited number of generations into early logarithmic phase and then be immediately introduced into the host. It is generally advisable to colonize with as large a population as possible to prevent bottlenecks, and at least to ensure that the inoculum is substantially larger than the diversity of the library. The inoculum should be sequenced for use as the reference to which animal (fecal) samples are compared. Less complex and more balanced libraries are less prone to bottleneck effects.

Mono-colonization

Mono-colonization of germ-free animals with a mutant library enables analysis of bacterial phenotypes driven by the host environment rather than confounding interspecies interactions with residential microbiota members. The host diet is a major environmental factor that can influence bacterial fitness[14,23,86], and dietary variations are relevant for host health and straightforward to implement. In addition, different host genotypes can be used to interrogate host-microbe genetic interactions[87]. For example, colonization of *Rag1*$^{-/-}$ or *Myd88*$^{-/-}$ mice can highlight bacterial genes that are influenced by the host adaptive or innate immune system, respectively[10]. Other models such as TGR5 (bile acid receptor) knockout mice may provide the ability to probe the host-microbe-metabolism axis. Animal models for colorectal cancer, inflammation (induced by chemicals such as DSS[88]), diarrhea[61], and viral infections can provide insight into bacterial genes specifically required for survival in a diseased host.

Colonization with other microbes

While mono-colonization provides a focused view on specific host-microbe interactions, other microbiota members play an intrinsic role in the life cycle of the library organism *in vivo* by influencing its fitness[10,23]. Germ-free animals form powerful model systems that enable careful design of the host-microbial ecosystem in which to probe the target organism. For example, germ-free animals can be colonized with a synthetic community of microbes that either lacks or has an excess of members of the same species/genus/family to investigate the impact of competition or support on the target organism phenotypes[10].

Use of well characterized, standard synthetic communities that contain a breadth of functionalities can expedite comparisons across labs. One prime example of such a community is the oligo mouse microbiota (OMM12), a widely used 12-member

community of mouse gut commensals[89]. More recently, synthetic communities of human gut species have been used to stably colonize mice, and a highly diverse (>100 member) community of human gut commensals was shown to reproducibly colonize germ-freemice[71]. Yet, whether such rich communities pose bottlenecks for testing colonization of pooled libraries remains to be tested; prior examples of pathogen colonization in mice[90] suggest that bottlenecking may be an issue. Of note, as most synthetic communities are composed of strains that share no evolutionary history (i.e., not isolated from the same host in which the strains co-evolved) phenotypes of the target organism related to important interspecies interactions may be missed. As an interesting possibility, strains that have co-evolved can be introduced in germ-free hosts as a stool-derived *in vitro* community[72], a synthetic community of isolates from a single individual, or a human fecal sample ("humanized"). In these cases, one would construct a library in a strain from a particular individual, and then test the fitness of this library in a host animal that is colonized with the community or fecal sample, also called a bacterial xenograft. It is probably best to first colonize the host with the community without the target organism (if possible) to allow the host and the community to adapt and stabilize, and afterwards introduce the library organism.

Novel behaviors may emerge as community complexity is increased, including nutrient competition, beneficial cross-feeding, and non-nutrient competition-based interactions (e.g., for spatial niches and direct antagonisms). When such interactions have previously been identified, transposon mutant library screens can identify the genetic basis of the interaction as long as the interaction of interest can be separated from other interactions.

Ultimately, it is unclear how much the phenotypic landscape will be affected by the presence of other commensals. If resource competition is the major driver of community composition[91], then the phenotypic impact of the community may be subtle (e.g., altering the relative strength of a phenotype). However, nutrient competition could also lead to metabolic reprogramming such that a gene for processing a certain carbon source becomes dispensable in the community context. Thus, investigating the library organism in hosts colonized with different communities can provide the opportunity to uncover emergent and general principles behind adaptation, colonization, and colonization resistance.

From a technical standpoint, the abundance of the library organism (which is influenced by competition and cooperation with the other community members and host factors such as immune system pressure) will influence barcode diversity and thus the capacity to quantify gene fitness. Some species, such as those in the *Bacteroides* genus, are typically at high enough abundance to avoid these issues (especially in the absence of competition with closely related species), but in other cases choosing a community and/or host environment that increases the abundance of the library organism (e.g., MAC-deficient diets promote *Akkermansia muciniphila*[92]) may be necessary to retain enough barcode diversity. An arrayed library may provide

a remedy for bottlenecking via the construction of a re-pooled library with lower diversity and thus higher numbers of each mutant for a given population size. However, re-pooling has the trade-off of fewer mutants in each gene, and with the limitation of very few mutants per gene it may be difficult to discern when *de novo* mutations are the driver of a high-fitness strain rather than the transposon insertion. One way to counter this issue is to sequence more replicates of the same library across all conditions, which enables more consistent determination of mutant behavior.

Potential animal hosts for *in vivo* experimentation

In vivo experimentation with mutant libraries can be performed in a variety of model and non-model host organisms, each with pros and cons. The choice of host organism may depend on biologically relevant considerations such as the aspect(s) of bacterial physiology of interest and the colonization capacity of the host, as well as practical factors such as availability, cost, and ethical considerations.

Mice

Due to powerful genetics, relatively easy husbandry (including germ-free), and extensive development of disease models, mice have served as the predominant host for *in vivo* experimentation with transposon mutant libraries of human gut bacteria[10,14,23,55]. In addition to the contexts of disease and diet switches, the knowledgebase regarding inbred laboratory mice also forms an excellent baseline to study host-microbe co-evolution using outbred mouse lines, wild mice, or other species in the *Mus* genus. Mice have a microbiome that is largely distinct from humans[93] and many mouse gut microbes are commercially available, including the OMM12 synthetic community. This convenience, in combination with the ability to humanize germ-free mice through colonization with human stool samples[71], enables detailed studies about host-microbiome interactions with transposon mutant libraries that can identify the conserved and unique factors influencing human gut commensal fitness during host colonization.

Other animal hosts

While other mammals such as germ-free miniature pigs are better suited as models for humans compared with mice in terms of natural diet, diurnal activity, and disease translatability, the costs of raising and maintaining germ-free pigs are considerably higher than for rodent model animals[94]. By contrast, gnotobiotic chickens provide certain benefits: they are easy to work with (egg shell sterilization prevents colonization of the chick) and inexpensive, and chickens have a relatively similar gut microbiota composition to mice[95], although their physiology is markedly different (e.g., the body temperature of birds is substantially higher than mammals). It is unknown to what extent bacterial fitness landscapes vary across hosts, although comparisons between colonization of germ-free mice and chicks with a *Bifidobacterium breve* transposon mutant library revealed surprisingly similar phenotypic landscapes given similar diets[96]; in such cases, outlier phenotypes provide insight into host-related

differences. To study host colonization and microbial evolution, several other animals across the vertebrate subphylum could provide complementary insights. Screening libraries in a model fish (e.g., zebrafish[97]), amphibian, or reptilian in addition to a bird and mammal might enable charting the impact of 500 million years of host divergence on bacterial adaptation. Of course, such investigations would only be possible for target bacteria that are able to colonize many hosts and adapt to widely varying host diets. *E. coli* is the best-studied intestinal bacterium and has a very wide host range[98], hence it may be appropriate for linking bacterial genes to host evolutionary divergence. Besides mice, zebrafish form a powerful model system due to their extensive genetics, ease in obtaining large population sizes, and the option to generate germ-free fish.

From the standpoint of experimental ease, population numbers, and host genetics, the nematode *Caenorhabditis elegans* and the fruit fly *Drosophila melanogaster* presents intriguing models for studying basic principles of host-microbe interactions[99,100]. *C. elegans* can be used in combination with bacterial transposon mutant libraries to rapidly identify virulence factors and genes involved in host responses to therapeutics[101-103] and for *Drosophila*, recent studies have also developed the capacity for *in situ* time-lapse imaging of the fly gut with single-cell resolution[104]. However, it is not clear whether human gut commensals can properly colonize *C. elegans* or *Drosophila*, as the physiology of these animals is very different from that of mammals in aspects such as body temperature, intestinal oxygen levels, diet, and more. Moreover, since the number of bacterial cells in the nematode and fly gut is much lower than in larger animals (~10^6 cells in the *Drosophila* gut[105]), bottleneck effects must be circumvented by using lower diversity libraries; when the library contains very few mutants representing each gene, it may be difficult to distinguish whether a fitness advantage is due to a particular transposon insertion or due to a *de novo* mutation that arose in that mutant without comparing fitness results across a large number of hosts (replicates).

Insights that can be obtained from *in vivo* colonization

The *in vivo* environment represents many environmental parameters such as chemical diversity and physical features that are not captured (and may be inaccessible) by typical *in vitro* experiments. As a result, *in vivo* colonization has the potential to reveal phenotypes for genes that would otherwise not exhibit any phenotypes *in vitro*, highlighting genes that may have evolved specifically for life in a host. For example, during mono-colonization of germ-free mice with *B. breve*, dozens of genes related to carbohydrate metabolism exhibited phenotypes *in vivo* but not in the hundreds of *in vitro* conditions screened; genes predicted to be responsible for nutrient uptake also exhibited specific phenotypes only *in vivo*[96].

Changes in the host environment can also affect fitness dramatically. Diet is a major determinant of the fitness of most gut commensals[86], and hence represents a natural knob for tuning the host-microbe interface. Exploring a broad spectrum of diets (e.g., polysaccharide replete versus deficient, high fat/sugar, fasting, caloric restriction, time-

restricted eating) enables the generation of broad hypotheses about commensal metabolism, while careful tuning of diet (e.g., a particular polysaccharide such as inulin[106]) enables testing of mechanistic hypotheses. In this context, it could be interesting to quantify the fitness of each mutant *in vitro* after growth on ground-up mouse chow[107] as a way to uncouple diet and host-diet interactions. Experiments with barcoded but otherwise genotypically identical lineages demonstrated that little to no contamination occurs between cages of gnotobiotic mice in the same isolator[108]. Co-housing mice colonized with the same library and fed a cage-specific diet in the same isolator simplifies the process of exploring a broad range of diets. Moreover, the high rates of transmission in mono-colonized mice within a cage[108] suggest that using >3 mice per cage may have diminishing returns. Single housing maintains each animal as a distinct, independent biological unit, although it remains unclear whether the microbial phenotypic landscape is affected by host behavioral changes that may emerge upon isolating a social animal. Like diet, xenobiotics such as excipients and drugs can have large effects on the microbiota in terms of both composition[109] and the fitness of individual strains[73]. Exposing mice colonized with a transposon mutant library to water-soluble drugs dissolved in drinking water is a straightforward way to determine the bacterial fitness determinants during treatment in hosts, and has the potential to reveal how host metabolism of the drug ends up protecting or sensitizing bacteria.

Links between knowledge obtained from *in vitro* and *in vivo* screens

An ultimate goal of gut commensal mutant library screening is to utilize information from *in vitro* fitness assays with carefully controlled conditions to shed light on the role of genes in the *in vivo* environment[110,111]. In general, nutrients, phage predation, survival in communities, and pressure from invasive (pathogenic) bacteria may be most robustly screened in high-throughput *in vitro* and the results are expected to be largely translatable to an *in vivo* setting. Certain conditions such as an antibiotic challenge translate straightforwardly in experimental design between the *in vitro* and *in vivo* contexts[110], but the interpretation of *in vivo* fitness data will be more complex given the likely off-target impacts on community composition, which may then indirectly select for certain mutants in the library organism (especially since the impact of drugs can be very different across microbiotas[112]).

Other conditions such as the influence of host immune responses and host behavior/physiology will be much harder to mimic *in vitro.* Probing these types of complex perturbations comes with a trade-off between throughput and translatability. For example, methods using immortalized (intestinal) cell lines to represent host factors can be easily scaled but such cells often poorly recapitulate healthy host functions. A more appropriate model would be intestinal organoids, which can closely mimic human physiology, but the costs to upscale organoid production and organoid heterogeneity may be prohibitive for testing highly diverse mutant libraries[113]. Other (larger) *in vitro* systems such as the Simulator of Human Intestinal Microbial

Ecosystem (SHIME)[114] may better capture host environmental parameters and are likely better suited for use with complex libraries.

A less complex application that may be well suited for translation from *in vitro* to *in vivo* is the high-throughput screening of mutant libraries with specific perturbations to discover "biosensor" mutants. For instance, mutants of oxidative stress pathways identified *in vitro* may be used *in vivo* to indicate where/if in the host such stress occurs. Similarly, certain mutants may be used to sense early stages of pathogen invasion or dysbiosis. In the case of *B. thetaiotaomicron*, comparison of *in vitro* screens to *in vivo* fitness measurements revealed that diet affects ammonium levels[115,116] and hence the utilization of ammonium-dependent alternative pathways[14], indicating that mutants in such pathways can indeed act as biosensors for host diet-mediated environmental changes. Host colonization may also provide the ability to distinguish genes/pathways that appear redundant *in vitro*.

Centralized collection of a broad data set of phenotypes from many diverse *in vitro* conditions (e.g., antibiotics, carbon sources, environmental perturbations, in the presence of other microbes) would enable comparisons across organisms, and hopefully the same will be true for *in vivo* fitness measurements (e.g., libraries in hosts with different diets, microbiotas, host genotypes). It remains to be seen whether genetic architecture and regulation will lead to general principles or species-specific solutions. For example, comparisons across libraries could shed light on the impact of host diet; in *B. thetaiotaomicron*, but not *B. breve*, lysine biosynthesis was critical for colonization on a standard diet[96]. These different requirements could be due to amino acids becoming limiting for *B. thetaiotaomicron* due to their higher abundance relative to *B. breve*. In that case, if another factor limits *B. thetaiotaomicron* abundance (for instance, the presence of other microbiota members), amino acid synthesis pathways could become non-essential even without any changes in amino acid concentrations within the gut. As a corollary, it is important to note (both *in vitro* and *in vivo*) that phenotypes may be concentration dependent. Such scenarios underscore the potential for the abundance of the library organism to impact mutant fitness[96].

The above points constitute a non-comprehensive set of links between *in vitro* and *in vivo* knowledge. Many outstanding questions related to *in vivo* experimentation remain. For example, should a "gold standard" set of *in vivo* conditions be established (e.g., diet, disease, xenobiotics) to screen and compare mutant libraries of diverse species? To what extent (and for what conditions) can *in vitro* screening, supplemented by field paradigms and literature, be used to formulate hypotheses about the *in vivo* function(s) of genes in commensal bacteria? And at what stage should efforts be made to delve deeper into phenotypes identified from *in vivo* experiments? Gaining a better understanding of the potential and limitations of *in vitro* data to inform *in vivo* experimentation will improve the design of future *in vivo* experiments with bacterial mutant libraries.

Considerations for the design of *in vivo* experiments

It is important to note that during the course of any colonization experiment, selection will take place on the entire genome, not just specific transposon mutants. Several studies have demonstrated rapid selection of *de novo* mutations when germ-free mice are colonized with a single species[86,108]. Within approximately one week (or even less) of colonization with barcoded transposon libraries of *Bacteroides* species, single mutants start to take over[14,117], signifying a large fitness increase due to a *de novo* mutation rather than the transposon insertion itself (since none of the other insertion mutants in that gene expand significantly). Thus, there is only a short interval over which the fitness of the transposon library should be measured. The cumulative bias in a pooled library due to *de novo* mutations may be exacerbated for hyper-mutators such as mutants in genes involved in DNA damage repair. Such genes may show larger variation in mutant fitness level due to the larger number of advantageous or disadvantageous *de novo* mutations.

To extend the time scale of the experiment while avoiding the impact of *de novo* mutations on *in vivo* screening, it may be useful to first allow the target organism to genetically adapt to the host and then isolate an evolved clone from which the barcoded insertion library is constructed. The adapted population is likely to be heterogeneous and multiple host-adapted colonies should be sequenced to determine the mutational diversity. Transposon libraries may then be constructed in several of these adapted backgrounds; methods for accelerating pooled library construction (see our accompanying review) could facilitate this process and allow for faster exploration of genetic interactions. However, it is important to note that *de novo* mutations with fitness benefits will likely continue to accumulate in these adapted-background libraries as well, and further adaptation will occur upon community or dietary changes.

Indirect effects and/or selection may also lead to changes in the relative abundance of the library organism, which could alter all phenotypes. One potential general issue during colonization is bottlenecks associated with host physiology (e.g., the acidic environment of the stomach) that stochastically affect the initial pool, thereby making fitness quantification challenging or impossible. In such cases, it may be necessary to inject a library directly into the gut. Bottlenecks will likely not be apparent from quantification of CFUs in stool since rapid growth in mice can restore the population to maximal levels within a single day[108]. During mono-colonization of mice, bottlenecks are not an issue for *B. thetaiotaomicron*[14] and *B. breve*[96]; in the latter case, most (if not all) barcodes colonize across mice even though *Bifidobacterium* species are sensitive to oxygen and cold (conditions they experience prior to inoculation and during migration between host individuals).

Analysis and sharing of fitness data

In most cases, pooled library experiments will generate large amounts of sequencing data, particularly since transposon barcoding enables screening of hundreds of conditions at reasonable cost (tens of dollars per condition). To analyze these data,

barcodes are typically quantified based on the number of assigned reads, grouped by gene (or other genomic feature) to average fitness variation across insertions, and the difference in cumulative barcode counts between the test and control conditions is used to infer the fitness effect of each gene in each condition[11,12]. Variability in phenotypic profiles across mutants in the same gene and calculation of the confidence in fitness estimates must be considered. Through this analysis, genes or groups of genes with fitness deviations under a particular set of conditions can be identified, enabling mapping of genetic networks and prioritization of interesting cases for follow-up mechanistic studies.

To maximally capitalize on these rich datasets, they must be made available to the scientific community at large. Adopting similar guidelines as the MIAME (Minimum Information About a Microarray Experiment)[118] and MINSEQE (Minimum Information About a Next-generation Sequencing Experiment)[119] for transposon library analysis, alongside a community-wide standardized pipeline to process raw data (e.g., how to handle barcode read quality, mismatches, alignment tolerance, etc.) will streamline communication and increase reproducibility, thereby enabling examination from diverse user perspectives and to empowering gene-specific investigations. Raw data can be deposited in online repositories[120] (https://journals.plos.org/ploscompbiol/s/recommended-repositories) that allow experienced users to use the data in custom manners, and processed data (e.g., calculated fitness values) can be made available through software or web-based platforms. Web-based platforms have the advantage of accommodating a wider network of researchers and thus increasing data distribution and application, as well as ease of updating. Usage of such tools can be stimulated by ensuring that data is easily browsable through a user-friendly interface supported by examples and tutorials and that the search input accepts cross-platform, stable feature identifiers (such as commonly used gene locus tags). The Fitness Browser (https://fit.genomics.lbl.gov/)[11] focuses on a single data type (gene-level fitness scores from pooled library screens) to enable fast comparison of gene fitness across diverse conditions and species, rapid incorporation of new datasets, and relatively straightforward maintenance. Users can browse fitness data by organism, gene, sequence, or condition, and the data base currently contains pre-computed fitness values from barcode sequencing of tens of thousands of *in vitro* and *in vivo* experiments involving dozens of bacterial mutant libraries. Moreover, linking to tools such as PaperBLAST, which mines the text of published papers for information about homologs, readily connects fitness data to other phenotypes[121].

Toward a general genotype-phenotype platform for bacteria

With the construction of more libraries in diverse bacteria and their distribution among research groups, the resulting genome-wide phenotypic measurements using pooled or arrayed libraries will likely encompass data of a wide range of types beyond sequencing including microscopy images, mass spectra, colony features (size, morphology, color), and more. Ideally, data from these experiments will be centrally

deposited and made accessible online, and other phenotype browsers may be worth developing using the Fitness Browser as a working model. The development of multiple phenotype browsers by different labs would ensure maximum flexibility and innovation with regards to data storage, analysis, and visualization. If such browsers provide the option to use standardized feature identifiers as queries, users can easily switch among phenotype browsers and platforms such as Uniprot and BLAST (preferentially through built-in linkages) with their feature of interest to assemble genotype-phenotype data and generate hypotheses. As researchers move on to new projects, there is always concern about the maintenance of modular browsers. A simple way to avoid issues with browser disappearance is to enable re-creation of the site by others from backups stored in a data repository (as is continually implemented for the Fitness Browser).

Interconnected phenotype browsers linked by standardized feature nomenclature would provide synergism that may help fund their maintenance. Once a sufficiently large and diverse collection of data is accumulated for an organism of interest (which should be facilitated by strategies that increase the number of labs performing experiments on mutant libraries), it may be reasonable to construct an organism-specific database similar to EcoCyc[122] in which multiple types of phenotypic data are integrated and linked to genomic data. The development of additional phenotype browsers and/or species/genus-specific integrated genotype-phenotype databases focused on non-model organisms would be an exciting step forward in modern microbiology and promote the discovery of bacterial functions through systems biology. With this goal in mind, the scientific community focused on transposon libraries could learn from other large-scale, multi-omics data collection initiatives such as the Human Cell Atlas[123] to collect, organize, share, and integrate data. To pave the way for these future advances, the research community should prioritize publication of raw data and the development of accessible, easy-to-use analytical pipelines to empower reproducibility and the development of novel analytical methods, reporting of technically correct but “negative” data to prevent unnecessary replication, and standardization of the implementation and frequent updating of genomic annotations to maintain correct inter-database communication.

Concluding remarks

Recent methodological advances such as RB-TnSeq have empowered mapping of phenotypic landscapes at massive scale, providing deeper insight into certain aspects of bacterial physiology. However, our comprehension of the host environment in which gut commensals coexist and evolve is still limited. Insight into host factors that impact microbiota physiology provides an opportunity to design conditions for *in vitro* screening that elucidate molecular mechanisms. For example, in the gut, host diet strongly affects gut microbiota composition and behavior, and the choice of medium can strongly affect *in vitro* phenotypes[14]. Future efforts that calibrate laboratory media to more closely mimic the nutrient environment in the mouse or human gut[124,125], perhaps guided by comparisons between community composition *in vitro* and *in vivo*[72] or analyses of gut metabolomes[126], could provide a sensible starting point for RB-TnSeq *in vitro* screens.

Simulation of other key aspects of the host environment can also be improved. Porcine gastric mucin is commercially available and hence is usually used as a proxy for intestinal mucus, but the structure of mucus varies throughout the gastrointestinal tract[127] and hence methods to produce mucins more relevant to the small and large intestines could enhance our ability to mimic both the nutrient and the spatial/adhesive roles of mucus in the gut. We also lack precise measurements of host environmental parameters such as pH, salinity, viscosity, oxygen, and temperature (e.g., during fever and exercise[128]), which can vary (perhaps in correlated manners) across the gut[129]. Screening of transposon mutant libraries in animal hosts may be the most direct route toward inferring the most relevant variables *in vivo*. Environmental variables most strongly correlated with changes in community composition in the gut also provide clear focal points for *in vitro* screening.

Another technical consideration is the stage of growth at which the pooled library is collected for *in vitro* fitness measurements. It is probably most expedient to continue collecting cultures in stationary phase, to avoid the undesired noise from differences in lag time across a plate or across experiments due to small environmental fluctuations. However, it is not clear which of log-phase growth, stationary phase, or deep starvation is the most relevant fitness determinant in a given condition, or whether compensatory effects mask fitness changes in a given mutant[130]. *E. coli* cells can adapt to long-term starvation[131], and the phenotype of certain *V. cholerae* mutants in a rabbit infection model depends on whether cells are in log or stationary phase at the time of inoculation[85]. Finally, an outstanding question is the desired properties of the background community for screening in the presence of interspecies interactions, both *in vitro* and *in vivo*. Whether the community should be diverse, contain closely related species, and be composed of mouse versus human commensals may be question dependent; hopefully studies of a few focal species will help to establish general guidelines and principles, such as the extent to which a background community changes the phenotypic landscape of the focal species.

Ultimately, achieving mechanistic understanding of gut microbiota function will require major leaps forward. Through many years of effort, functions or phenotypes have been uncovered for >75% of the genes in *E. coli*, establishing it as a preeminent model organism. By contrast, for virtually all gut commensals, only a minor fraction of their genes has known functions, and the low abundance of *E. coli* in healthy gut microbiotas[132] suggests that it is a poor model for many gut functions. Efforts to build phenotypic landscapes of similar breadth in even a few representatives of the major gut phyla would likely represent a transformative advance given the probable synergy due to overlap in genetic content among closely related gut species. To achieve this goal, it will be desirable for as many screening conditions as possible to provide phenotypic information, which can be accomplished by acquiring more knowledge of the gut environment or through unbiased screening of a broad range of conditions such as media, toxins , human drugs[73], or even non-Western chemicals such as herbal remedies. Fortunately, transposon barcoding facilitates high-throughput screening, and the increasingly low cost of sequencing means that it should be routine to screen thousands of conditions for each species. With robotics and rapid sequencing turnaround, it may even be feasible to construct a closed-loop system in which new conditions are selected and evaluated based on automated analysis of existing data[84]. We look forward to a near future in which data sets of hundreds of thousands of species/condition combinations can be systematically compared using user-friendly computational tools available to the global research community.

Author Contributions

C.G.P.V., S.T., A.T., A.D., A.L.S., and K.C.H. conceptualized the review. The review was written mainly by C.G.P.V., S.T., and K.C.H. with contributions from all authors. All authors edited the review.

Acknowledgements

We thank members of the Huang and Typas labs for helpful discussions. This work was funded by NSF Award EF-2125383 (to K.C.H.), NIH Award RM1 GM135102 (to K.C.H. and A.D.), ERC grant uCARE ID 819454 (to A.T.), the Liliane Bettencourt Prize for Life Sciences (to A.T.), and ERC grant GutTransForm ID 101078353 (to M.Z.). C.G.P.V was supported by a fellowship from the EMBL Interdisciplinary Postdoc (EIPOD4) program under the Marie Curie Skłodowska-Curie Actions COFUND (grant number: 847543) and by an Add-On Fellowship for Interdisciplinary Life Science from the Joachim Herz Foundation. S.O.B. was supported by an Add-On Fellowship for Interdisciplinary Life Science from the Joachim Herz Foundation. K.C.H. and J.L.S. are Chan Zuckerberg Biohub Investigators. A.T. and M.Z. are ERC investigators. Figures were created in part using Biorender.com.

Declaration of interests

The authors declare no competing interests.

Figures

Figure 1: *In vitro* library screening process and output.
Barcoded transposon mutant libraries can be screened in high throughput using a vial of the pooled library diluted to an appropriate starting OD (dependent on the testing conditions and library diversity). The library is introduced in multi-well plates to various conditions such as diverse drugs. The library should be grown in the conditions for a limited number of generations, after which DNA is extracted from the pellet. A single PCR is performed (Bar-Seq) to amplify the barcode in each transposon and barcode amplicons are sequenced en masse and quantified. The barcodes are then linked to genes through the library-specific gene-barcode map, and barcode abundances can be used to quantify mutant fitness, to calculate co-fitness, and to reconstruct gene networks.

Figure 2: Trade-offs and synergies between experimental platforms. Standard *in vitro* systems, particularly multi-well plates, provide the scalability for high-throughput screening of hundreds or thousands of conditions. The translatability of *in vitro* screening results may be unclear but can be established with other systems with higher complexity at the cost of lower scalability. Ultimately, *in vitro* screening results (lower left) can provide synergistic information to interpret *in vivo* experiments; for example, if the phenotypes of two semi-redundant enzymes differ between low and high ammonium conditions *in vitro*, their phenotypes *in vivo* suggest that the host environment is low in ammonium[14]. Blue and yellow indicate genes with low or high fitness, respectively, in a given condition.

References

1. Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z.J., Pollard, K.S., Sakharova, E., Parks, D.H., Hugenholtz, P., et al. (2021). A unified catalog of 204,938 reference genomes from the human gut microbiome. Nature Biotechnology *39*, 105-114. 10.1038/s41587-020-0603-3.
2. Almeida, A., Mitchell, A.L., Boland, M., Forster, S.C., Gloor, G.B., Tarkowska, A., Lawley, T.D., and Finn, R.D. (2019). A new genomic blueprint of the human gut microbiota. Nature *568*, 499-504.
3. Forster, S.C., Kumar, N., Anonye, B.O., Almeida, A., Viciani, E., Stares, M.D., Dunn, M., Mkandawire, T.T., Zhu, A., Shao, Y., et al. (2019). A human gut bacterial genome and culture collection for improved metagenomic analyses. Nat Biotechnol *37*, 186-192. 10.1038/s41587-018-0009-7.
4. Zou, Y., Xue, W., Luo, G., Deng, Z., Qin, P., Guo, R., Sun, H., Xia, Y., Liang, S., and Dai, Y. (2019). 1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses. Nature biotechnology *37*, 179-185.
5. Kim, C.Y., Lee, M., Yang, S., Kim, K., Yong, D., Kim, H.R., and Lee, I. (2021). Human reference gut microbiome catalog including newly assembled genomes from under-represented Asian metagenomes. Genome Medicine *13*, 1-20.
6. Heintz-Buschart, A., and Wilmes, P. (2018). Human gut microbiome: function matters. Trends in microbiology *26*, 563-574.
7. Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., and Ghensi, P. (2019). Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. Cell *176*, 649-662. e620.
8. Fan, Y., and Pedersen, O. (2021). Gut microbiota in human metabolic health and disease. Nature Reviews Microbiology *19*, 55-71.
9. Lynch, S.V., and Pedersen, O. (2016). The human intestinal microbiome in health and disease. New England Journal of Medicine *375*, 2369-2379.
10. Goodman, A.L., McNulty, N.P., Zhao, Y., Leip, D., Mitra, R.D., Lozupone, C.A., Knight, R., and Gordon, J.I. (2009). Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell host & microbe *6*, 279-289.
11. Price, M.N., Wetmore, K.M., Waters, R.J., Callaghan, M., Ray, J., Liu, H., Kuehl, J.V., Melnyk, R.A., Lamson, J.S., and Suh, Y. (2018). Mutant phenotypes for thousands of bacterial genes of unknown function. Nature *557*, 503.
12. Wetmore, K.M., Price, M.N., Waters, R.J., Lamson, J.S., He, J., Hoover, C.A., Blow, M.J., Bristow, J., Butland, G., Arkin, A.P., and Deutschbauer, A. (2015). Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. mBio *6*, e00306-00315. 10.1128/mBio.00306-15.
13. Rosconi, F., Rudmann, E., Li, J., Surujon, D., Anthony, J., Frank, M., Jones, D.S., Rock, C., Rosch, J.W., Johnston, C.D., and van Opijnen, T. (2022). A bacterial pan-genome makes gene essentiality strain-dependent and evolvable. Nat Microbiol *7*, 1580-1592. 10.1038/s41564-022-01208-7.

14. Liu, H., Shiver, A.L., Price, M.N., Carlson, H.K., Trotter, V.V., Chen, Y., Escalante, V., Ray, J., Hern, K.E., Petzold, C.J., et al. (2021). Functional genetics of human gut commensal Bacteroides thetaiotaomicron reveals metabolic requirements for growth across environments. Cell Rep *34*, 108789. 10.1016/j.celrep.2021.108789.
15. Goodman, A.L., Wu, M., and Gordon, J.I. (2011). Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. Nature protocols *6*, 1969.
16. Cullen, T., Schofield, W., Barry, N., Putnam, E., Rundell, E., Trent, M., Degnan, P., Booth, C., Yu, H., and Goodman, A. (2015). Antimicrobial peptide resistance mediates resilience of prominent gut commensals during inflammation. Science *347*, 170-175.
17. Cain, A.K., Barquist, L., Goodman, A.L., Paulsen, I.T., Parkhill, J., and van Opijnen, T. (2020). A decade of advances in transposon-insertion sequencing. Nat Rev Genet *21*, 526-540. 10.1038/s41576-020-0244-x.
18. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol *2*, 2006 0008. 10.1038/msb4100050.
19. Cameron, D.E., Urbach, J.M., and Mekalanos, J.J. (2008). A defined transposon mutant library and its use in identifying motility genes in Vibrio cholerae. Proceedings of the National Academy of Sciences *105*, 8736-8741.
20. Koo, B.-M., Kritikos, G., Farelli, J.D., Todor, H., Tong, K., Kimsey, H., Wapinski, I., Galardini, M., Cabal, A., and Peters, J.M. (2017). Construction and analysis of two genome-scale deletion libraries for Bacillus subtilis. Cell systems *4*, 291-305. e297.
21. Van Opijnen, T., Bodi, K.L., and Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nature methods *6*, 767-772.
22. Langridge, G.C., Phan, M.D., Turner, D.J., Perkins, T.T., Parts, L., Haase, J., Charles, I., Maskell, D.J., Peters, S.E., Dougan, G., et al. (2009). Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. Genome Res *19*, 2308-2316. 10.1101/gr.097097.109.
23. Wu, M., McNulty, N.P., Rodionov, D.A., Khoroshkin, M.S., Griffin, N.W., Cheng, J., Latreille, P., Kerstetter, R.A., Terrapon, N., Henrissat, B., et al. (2015). Genetic determinants of in vivo fitness and diet responsiveness in multiple human gut Bacteroides. Science *350*, aac5992. 10.1126/science.aac5992.
24. Arjes, H.A., Sun, J., Liu, H., Nguyen, T.H., Culver, R.N., Celis, A.I., Walton, S.J., Vasquez, K.S., Yu, F.B., Xue, K.S., et al. (2022). Construction and characterization of a genome-scale ordered mutant collection of Bacteroides thetaiotaomicron. BMC Biol *20*, 285. 10.1186/s12915-022-01481-2.
25. Baym, M., Shaket, L., Anzai, I.A., Adesina, O., and Barstow, B. (2016). Rapid construction of a whole-genome transposon insertion collection for Shewanella oneidensis by Knockout Sudoku. Nature communications *7*, 13270.
26. Fuhrer, T., Zampieri, M., Sévin, D.C., Sauer, U., and Zamboni, N. (2017). Genomewide landscape of gene–metabolome associations in Escherichia coli. Molecular Systems Biology *13*, 907.

27. Campos, M., Govers, S.K., Irnov, I., Dobihal, G.S., Cornet, F., and Jacobs-Wagner, C. (2018). Genomewide phenotypic analysis of growth, cell morphogenesis, and cell cycle events in Escherichia coli. Molecular systems biology *14*, e7573.
28. Shiver, A.L., Culver, R., Deutschbauer, A.M., and Huang, K.C. (2021). Rapid ordering of barcoded transposon insertion libraries of anaerobic bacteria. Nat Protoc *16*, 3049-3071. 10.1038/s41596-021-00531-3.
29. Nichols, R.J., Sen, S., Choo, Y.J., Beltrao, P., Zietek, M., Chaba, R., Lee, S., Kazmierczak, K.M., Lee, K.J., Wong, A., et al. (2011). Phenotypic landscape of a bacterial cell. Cell *144*, 143-156. 10.1016/j.cell.2010.11.052.
30. French, S., Mangat, C., Bharat, A., Côté, J.-P., Mori, H., and Brown, E.D. (2016). A robust platform for chemical genomics in bacterial systems. Molecular biology of the cell *27*, 1015-1025.
31. Chalabaev, S., Chauhan, A., Novikov, A., Iyer, P., Szczesny, M., Beloin, C., Caroff, M., and Ghigo, J.-M. (2014). Biofilms formed by gram-negative bacteria undergo increased lipid a palmitoylation, enhancing in vivo survival. MBio *5*, e01116-01114.
32. Orman, M.A., and Brynildsen, M.P. (2016). Persister formation in Escherichia coli can be inhibited by treatment with nitric oxide. Free Radical Biology and Medicine *93*, 145-154.
33. Gray, A.N., Egan, A.J., Van't Veer, I.L., Verheul, J., Colavin, A., Koumoutsi, A., Biboy, J., Altelaar, A.F., Damen, M.J., Huang, K.C., et al. (2015). Coordination of peptidoglycan synthesis and outer membrane constriction during Escherichia coli cell division. Elife *4*. 10.7554/eLife.07118.
34. Typas, A., Banzhaf, M., van den Berg van Saparoea, B., Verheul, J., Biboy, J., Nichols, R.J., Zietek, M., Beilharz, K., Kannenberg, K., von Rechenberg, M., et al. (2010). Regulation of peptidoglycan synthesis by outer-membrane proteins. Cell *143*, 1097-1109. 10.1016/j.cell.2010.11.038.
35. Shiver, A.L., Osadnik, H., Kritikos, G., Li, B., Krogan, N., Typas, A., and Gross, C.A. (2016). A chemical-genomic screen of neglected antibiotics reveals illicit transport of kasugamycin and blasticidin S. PLoS genetics *12*, e1006124.
36. Ghatak, S., King, Z.A., Sastry, A., and Palsson, B.O. (2019). The y-ome defines the 35% of Escherichia coli genes that lack experimental evidence of function. Nucleic acids research *47*, 2446-2454.
37. Wicke, D., Meißner, J., Warneke, R., Elfmann, C., and Stülke, J. (2023). Understudied proteins and understudied functions in the model bacterium Bacillus subtilis—A major challenge in current research. Molecular Microbiology *120*, 8-19. https://doi.org/10.1111/mmi.15053.
38. Le Breton, Y., Mistry, P., Valdes, K.M., Quigley, J., Kumar, N., Tettelin, H., and McIver, K.S. (2013). Genome-wide identification of genes required for fitness of group A Streptococcus in human blood. Infection and immunity *81*, 862-875.
39. Hubbard, T.P., Chao, M.C., Abel, S., Blondel, C.J., Abel zur Wiesch, P., Zhou, X., Davis, B.M., and Waldor, M.K. (2016). Genetic analysis of Vibrio parahaemolyticus intestinal colonization. Proceedings of the National Academy of Sciences *113*, 6283-6288.
40. Weerdenburg, E.M., Abdallah, A.M., Rangkuti, F., Abd El Ghany, M., Otto, T.D., Adroub, S.A., Molenaar, D., Ummels, R., Ter Veen, K., and van Stempvoort, G. (2015). Genome-wide transposon mutagenesis indicates that

Mycobacterium marinum customizes its virulence mechanisms for survival and replication in different hosts. Infection and immunity *83*, 1778-1788.
41. Tripathi, S., Voogdt, C.G.P., Bassler, S.O., Anderson, M., Huang, P.-H., Sakenova, N., Capraz, T., Jain, S., Koumoutsi, A., Bravo, A.M., et al. (2023). Randomly barcoded transposon mutant libraries for gut commensals I: strategies for efficient library construction. Cell Reports.
42. Park, S.Y., Rao, C., Coyte, K.Z., Kuziel, G.A., Zhang, Y., Huang, W., Franzosa, E.A., Weng, J.K., Huttenhower, C., and Rakoff-Nahoum, S. (2022). Strain-level fitness in the gut microbiome is an emergent property of glycans and a single metabolite. Cell *185*, 513-529 e521. 10.1016/j.cell.2022.01.002.
43. Baker, B.J., and Dick, G.J. (2013). Omic approaches in microbial ecology: charting the unknown. Microbe *8*, 353-359.
44. Tramontano, M., Andrejev, S., Pruteanu, M., Klünemann, M., Kuhn, M., Galardini, M., Jouhten, P., Zelezniak, A., Zeller, G., and Bork, P. (2018). Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. Nature microbiology *3*, 514-522.
45. Christensen, D.G., Orr, J.S., Rao, C.V., and Wolfe, A.J. (2017). Increasing Growth Yield and Decreasing Acetylation in Escherichia coli by Optimizing the Carbon-to-Magnesium Ratio in Peptide-Based Media. Appl Environ Microbiol *83*. 10.1128/AEM.03034-16.
46. Varel, V.H., and Bryant, M.P. (1974). Nutritional features of Bacteroides fragilis subsp. fragilis. Applied microbiology *28*, 251-257.
47. Pudlo, N.A., Urs, K., Crawford, R., Pirani, A., Atherly, T., Jimenez, R., Terrapon, N., Henrissat, B., Peterson, D., and Ziemer, C. (2022). Phenotypic and Genomic Diversification in Complex Carbohydrate-Degrading Human Gut Bacteria. Msystems *7*, e00947-00921.
48. Tong, M., French, S., El Zahed, S.S., Ong, W.k., Karp, P.D., and Brown, E.D. (2020). Gene dispensability in Escherichia coli grown in thirty different carbon environments. MBio *11*, e02259-02220.
49. Celis, A.I., Aranda-Diaz, A., Culver, R., Xue, K., Relman, D., Shi, H., and Huang, K.C. (2022). Optimization of the 16S rRNA sequencing analysis pipeline for studying in vitro communities of gut commensals. iScience *25*, 103907. 10.1016/j.isci.2022.103907.
50. de Bakker, V., Liu, X., Bravo, A.M., and Veening, J.-W. (2022). CRISPRi-seq for genome-wide fitness quantification in bacteria. Nature Protocols *17*, 252-281.
51. Sanders, M.E., Merenstein, D.J., Reid, G., Gibson, G.R., and Rastall, R.A. (2019). Probiotics and prebiotics in intestinal health and disease: from biology to the clinic. Nature reviews Gastroenterology & hepatology *16*, 605-616.
52. Bedu-Ferrari, C., Biscarrat, P., Langella, P., and Cherbuy, C. (2022). Prebiotics and the human gut microbiota: From breakdown mechanisms to the impact on metabolic health. Nutrients *14*, 2096.
53. Shepherd, E.S., DeLoache, W.C., Pruss, K.M., Whitaker, W.R., and Sonnenburg, J.L. (2018). An exclusive metabolic niche enables strain engraftment in the gut microbiota. Nature *557*, 434-438.
54. Le Bastard, Q., Chapelet, G., Javaudin, F., Lepelletier, D., Batard, E., and Montassier, E. (2020). The effects of inulin on gut microbial composition: a systematic review of evidence from human studies. European Journal of Clinical Microbiology & Infectious Diseases *39*, 403-413.

55. Liou, C.S., Sirk, S.J., Diaz, C.A., Klein, A.P., Fischer, C.R., Higginbottom, S.K., Erez, A., Donia, M.S., Sonnenburg, J.L., and Sattely, E.S. (2020). A metabolic pathway for activation of dietary glucosinolates by a human gut symbiont. Cell *180*, 717-728. e719.
56. Jayachandran, M., Chen, J., Chung, S.S.M., and Xu, B. (2018). A critical review on the impacts of β-glucans on gut microbiota and human health. The Journal of nutritional biochemistry *61*, 101-110.
57. Evans, D., Pye, G., Bramley, R., Clark, A., Dyson, T., and Hardcastle, J. (1988). Measurement of gastrointestinal pH profiles in normal ambulant human subjects. Gut *29*, 1035-1041.
58. Firrman, J., Liu, L., Mahalak, K., Tanes, C., Bittinger, K., Tu, V., Bobokalonov, J., Mattei, L., Zhang, H., and Van den Abbeele, P. (2022). The impact of environmental pH on the gut microbiota community structure and short chain fatty acid production. FEMS Microbiology Ecology *98*, fiac038.
59. Ratzke, C., and Gore, J. (2018). Modifying and reacting to the environmental pH can drive bacterial interactions. PLoS biology *16*, e2004248.
60. Cremer, J., Arnoldini, M., and Hwa, T. (2017). Effect of water flow and chemical environment on microbiota growth and composition in the human colon. Proceedings of the National Academy of Sciences *114*, 6438-6443.
61. Tropini, C., Moss, E.L., Merrill, B.D., Ng, K.M., Higginbottom, S.K., Casavant, E.P., Gonzalez, C.G., Fremin, B., Bouley, D.M., Elias, J.E., et al. (2018). Transient Osmotic Perturbation Causes Long-Term Alteration to the Gut Microbiota. Cell *173*, 1742-1754 e1717. 10.1016/j.cell.2018.05.008.
62. Marquis, R.E. (1968). Salt-induced contraction of bacterial cell walls. Journal of bacteriology *95*, 775-781.
63. Rojas, E., Theriot, J.A., and Huang, K.C. (2014). Response of Escherichia coli growth rate to osmotic shock. Proc Natl Acad Sci U S A *111*, 7807-7812. 10.1073/pnas.1402591111.
64. Tamargo, A., Cueva, C., Álvarez, M.D., Herranz, B., Bartolomé, B., Moreno-Arribas, M.V., and Laguna, L. (2018). Influence of viscosity on the growth of human gut microbiota. Food Hydrocolloids *77*, 163-167.
65. Pelaseyed, T., Bergström, J.H., Gustafsson, J.K., Ermund, A., Birchenough, G.M., Schütte, A., van der Post, S., Svensson, F., Rodríguez-Piñeiro, A.M., and Nyström, E.E. (2014). The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. Immunological reviews *260*, 8-20.
66. Ermund, A., Schütte, A., Johansson, M.E., Gustafsson, J.K., and Hansson, G.C. (2013). Studies of mucus in mouse stomach, small intestine, and colon. I. Gastrointestinal mucus layers have different properties depending on location as well as over the Peyer's patches. American Journal of Physiology-Gastrointestinal and Liver Physiology *305*, G341-G347.
67. Desai, M.S., Seekatz, A.M., Koropatkin, N.M., Kamada, N., Hickey, C.A., Wolter, M., Pudlo, N.A., Kitamoto, S., Terrapon, N., and Muller, A. (2016). A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. Cell *167*, 1339-1353. e1321.
68. Collins, S.L., Stine, J.G., Bisanz, J.E., Okafor, C.D., and Patterson, A.D. (2023). Bile acids and the gut microbiota: metabolic interactions and impacts on disease. Nature Reviews Microbiology *21*, 236-247.
69. Nakajima, A., Vogelzang, A., Maruya, M., Miyajima, M., Murata, M., Son, A., Kuwahara, T., Tsuruyama, T., Yamada, S., and Matsuura, M. (2018). IgA

regulates the composition and metabolic function of gut microbiota by promoting symbiosis between bacteria. Journal of Experimental Medicine *215*, 2019-2034.

70. Foster, K.R., and Bell, T. (2012). Competition, not cooperation, dominates interactions among culturable microbial species. Curr Biol *22*, 1845-1850. 10.1016/j.cub.2012.08.005.
71. Cheng, A.G., Ho, P.-Y., Aranda-Díaz, A., Jain, S., Feiqiao, B.Y., Meng, X., Wang, M., Iakiviak, M., Nagashima, K., and Zhao, A. (2022). Design, construction, and in vivo augmentation of a complex gut microbiome. Cell *185*, 3617-3636.
72. Aranda-Diaz, A., Ng, K.M., Thomsen, T., Real-Ramirez, I., Dahan, D., Dittmar, S., Gonzalez, C.G., Chavez, T., Vasquez, K.S., Nguyen, T.H., et al. (2022). Establishment and characterization of stable, diverse, fecal-derived in vitro microbial communities that model the intestinal microbiota. Cell Host Microbe *30*, 260-272 e265. 10.1016/j.chom.2021.12.008.
73. Maier, L., Pruteanu, M., Kuhn, M., Zeller, G., Telzerow, A., Anderson, E.E., Brochado, A.R., Fernandez, K.C., Dose, H., Mori, H., et al. (2018). Extensive impact of non-antibiotic drugs on human gut bacteria. Nature *555*, 623-628. 10.1038/nature25979.
74. Maier, L., Goemans, C.V., Wirbel, J., Kuhn, M., Eberl, C., Pruteanu, M., Muller, P., Garcia-Santamarina, S., Cacace, E., Zhang, B., et al. (2021). Unravelling the collateral damage of antibiotics on gut bacteria. Nature *599*, 120-124. 10.1038/s41586-021-03986-2.
75. Price, S.L., Thibault, D., Garrison, T.M., Brady, A., Guo, H., Kehl-Fie, T.E., Garneau-Tsodikova, S., Perry, R.D., van Opijnen, T., and Lawrenz, M.B. (2023). Droplet Tn-Seq identifies the primary secretion mechanism for yersiniabactin in Yersinia pestis. EMBO reports, e57369.
76. Thibault, D., Jensen, P.A., Wood, S., Qabar, C., Clark, S., Shainheit, M.G., Isberg, R.R., and van Opijnen, T. (2019). Droplet Tn-Seq combines microfluidics with Tn-Seq for identifying complex single-cell phenotypes. Nature Communications *10*, 5729.
77. Peters, J.M., Colavin, A., Shi, H., Czarny, T.L., Larson, M.H., Wong, S., Hawkins, J.S., Lu, C.H., Koo, B.-M., and Marta, E. (2016). A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. Cell *165*, 1493-1506.
78. Shi, H., Colavin, A., Lee, T.K., and Huang, K.C. (2017). Strain Library Imaging Protocol for high-throughput, automated single-cell microscopy of large bacterial collections arrayed on multiwell plates. Nat Protoc *12*, 429-438. 10.1038/nprot.2016.181.
79. Mateus, A., Hevler, J., Bobonis, J., Kurzawa, N., Shah, M., Mitosch, K., Goemans, C.V., Helm, D., Stein, F., and Typas, A. (2020). The functional proteome landscape of Escherichia coli. Nature *588*, 473-478.
80. Kritikos, G., Banzhaf, M., Herrera-Dominguez, L., Koumoutsi, A., Wartel, M., Zietek, M., and Typas, A. (2017). A tool named Iris for versatile high-throughput phenotyping in microorganisms. Nature microbiology *2*, 17014.
81. Silvis, M.R., Rajendram, M., Shi, H., Osadnik, H., Gray, A.N., Cesar, S., Peters, J.M., Hearne, C.C., Kumar, P., Todor, H., et al. (2021). Morphological and Transcriptional Responses to CRISPRi Knockdown of Essential Genes in Escherichia coli. mBio *12*, e0256121. 10.1128/mBio.02561-21.

82. Ronda, C., Chen, S.P., Cabral, V., Yaung, S.J., and Wang, H.H. (2019). Metagenomic engineering of the mammalian gut microbiome in situ. Nat Methods *16*, 167-170. 10.1038/s41592-018-0301-y.
83. Davey, L.E., Malkus, P.N., Villa, M., Dolat, L., Holmes, Z.C., Letourneau, J., Ansaldo, E., David, L.A., Barton, G.M., and Valdivia, R.H. (2023). A genetic system for Akkermansia muciniphila reveals a role for mucin foraging in gut colonization and host sterol biosynthesis gene expression. Nat Microbiol. 10.1038/s41564-023-01407-w.
84. Dama, A.C., Kim, K.S., Leyva, D.M., Lunkes, A.P., Schmid, N.S., Jijakli, K., and Jensen, P.A. (2023). BacterAI maps microbial metabolism without prior knowledge. Nature Microbiology, 1-8.
85. Dörr, T., Möll, A., Chao, M.C., Cava, F., Lam, H., Davis, B.M., and Waldor, M.K. (2014). Differential requirement for PBP1a and PBP1b in in vivo and in vitro fitness of Vibrio cholerae. Infection and immunity *82*, 2115-2124.
86. Dapa, T., Ramiro, R.S., Pedro, M.F., Gordo, I., and Xavier, K.B. (2022). Diet leaves a genetic signature in a keystone member of the gut microbiota. Cell Host Microbe *30*, 183-199 e110. 10.1016/j.chom.2022.01.002.
87. Frazão, N., Sousa, A., Lässig, M., and Gordo, I. (2019). Horizontal gene transfer overrides mutation in Escherichia coli colonizing the mammalian gut. Proceedings of the National Academy of Sciences *116*, 17906-17915.
88. Benoit, C., Jesse, D., Madhu, M., and Matam, V. (2014). Dextran sulfate sodium (DSS)-induced colitis in mice. Curr Protoc Immunol *25*, 1-15.
89. Eberl, C., Ring, D., Münch, P.C., Beutler, M., Basic, M., Slack, E.C., Schwarzer, M., Srutkova, D., Lange, A., and Frick, J.S. (2020). Reproducible colonization of germ-free mice with the oligo-mouse-microbiota in different animal facilities. Frontiers in microbiology *10*, 2999.
90. Maier, L., Diard, M., Sellin, M.E., Chouffane, E.-S., Trautwein-Weidner, K., Periaswamy, B., Slack, E., Dolowschiak, T., Stecher, B., and Loverdo, C. (2014). Granulocytes impose a tight bottleneck upon the gut luminal pathogen population during Salmonella typhimurium colitis. PLoS pathogens *10*, e1004557.
91. Ho, P.-Y., Nguyen, T.H., Sanchez, J.M., DeFelice, B.C., and Huang, K.C. (2022). Resource competition predicts assembly of in vitro gut bacterial communities. bioRxiv.
92. Earle, K.A., Billings, G., Sigal, M., Lichtman, J.S., Hansson, G.C., Elias, J.E., Amieva, M.R., Huang, K.C., and Sonnenburg, J.L. (2015). Quantitative Imaging of Gut Microbiota Spatial Organization. Cell Host Microbe *18*, 478-488. 10.1016/j.chom.2015.09.002.
93. Kieser, S., Zdobnov, E.M., and Trajkovski, M. (2022). Comprehensive mouse microbiota genome catalog reveals major difference to its human counterpart. PLoS Comput Biol *18*, e1009947. 10.1371/journal.pcbi.1009947.
94. Rose, E.C., Blikslager, A.T., and Ziegler, A.L. (2022). Porcine Models of the Intestinal Microbiota: The Translational Key to Understanding How Gut Commensals Contribute to Gastrointestinal Disease. Front Vet Sci *9*, 834598. 10.3389/fvets.2022.834598.
95. Stanley, D., Hughes, R.J., and Moore, R.J. (2014). Microbiota of the chicken gastrointestinal tract: influence on health, productivity and disease. Appl Microbiol Biotechnol *98*, 4301-4310. 10.1007/s00253-014-5646-2.
96. Shiver, A.L., Violette, A., Sun, J., Culver, R., Drescher, S.P.M., Sekhon, P.K., Feissler, L., Carlson, H.K., Wong, D., Wynter, C., et al. (2023). An atlas of

gene function in *Bifidobacterium breve* reveals molecular determinants of colonization and host-microbe interactions. bioRxiv.
97. Roeselers, G., Mittge, E.K., Stephens, W.Z., Parichy, D.M., Cavanaugh, C.M., Guillemin, K., and Rawls, J.F. (2011). Evidence for a core gut microbiota in the zebrafish. ISME J *5*, 1595-1608. 10.1038/ismej.2011.38.
98. Murphy, R., Palm, M., Mustonen, V., Warringer, J., Farewell, A., Parts, L., and Moradigaravand, D. (2021). Genomic epidemiology and evolution of Escherichia coli in wild animals in Mexico. Msphere *6*, e00738-00720.
99. Zhang, J., Holdorf, A.D., and Walhout, A.J. (2017). C. elegans and its bacterial diet as a model for systems-level understanding of host–microbiota interactions. Current opinion in biotechnology *46*, 74-80.
100. Ludington, W.B., and Ja, W.W. (2020). Drosophila as a model for the gut microbiome. PLoS Pathogens *16*, e1008398.
101. Garvis, S., Munder, A., Ball, G., de Bentzmann, S., Wiehlmann, L., Ewbank, J.J., Tümmler, B., and Filloux, A. (2009). Caenorhabditis elegans semi-automated liquid screen reveals a specialized role for the chemotaxis gene cheB2 in Pseudomonas aeruginosa virulence. PLoS pathogens *5*, e1000540.
102. Lewenza, S., Charron-Mazenod, L., Giroux, L., and Zamponi, A.D. (2014). Feeding behaviour of Caenorhabditis elegans is an indicator of Pseudomonas aeruginosa PAO1 virulence. PeerJ *2*, e521.
103. García-González, A.P., Ritter, A.D., Shrestha, S., Andersen, E.C., Yilmaz, L.S., and Walhout, A.J. (2017). Bacterial metabolism affects the C. elegans response to cancer chemotherapeutics. Cell *169*, 431-441. e438.
104. Koyama, L.A.J., Aranda-Diaz, A., Su, Y.H., Balachandra, S., Martin, J.L., Ludington, W.B., Huang, K.C., and O'Brien, L.E. (2020). Bellymount enables longitudinal, intravital imaging of abdominal organs and the gut microbiota in adult Drosophila. PLoS Biol *18*, e3000567. 10.1371/journal.pbio.3000567.
105. Pais, I.S., Valente, R.S., Sporniak, M., and Teixeira, L. (2018). Drosophila melanogaster establishes a species-specific mutualistic interaction with stable gut-colonizing bacteria. PLoS Biol *16*, e2005710. 10.1371/journal.pbio.2005710.
106. Hryckowian, A.J., Van Treuren, W., Smits, S.A., Davis, N.M., Gardner, J.O., Bouley, D.M., and Sonnenburg, J.L. (2018). Microbiota-accessible carbohydrates suppress Clostridium difficile infection in a murine model. Nat Microbiol *3*, 662-669. 10.1038/s41564-018-0150-6.
107. Maritan, E., Gallo, M., Srutkova, D., Jelinkova, A., Benada, O., Kofronova, O., Silva-Soares, N.F., Hudcovic, T., Gifford, I., and Barrick, J.E. (2022). Gut microbe Lactiplantiballus plantarum undergoes different evolutionary trajectories between insects and mammals. bioRxiv.
108. Vasquez, K.S., Willis, L., Cira, N.J., Ng, K.M., Pedro, M.F., Aranda-Diaz, A., Rajendram, M., Yu, F.B., Higginbottom, S.K., Neff, N., et al. (2021). Quantifying rapid bacterial evolution and transmission within the mouse intestine. Cell Host Microbe *29*, 1454-1468 e1454. 10.1016/j.chom.2021.08.003.
109. Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., Kurilshikov, A., Bonder, M.J., Valles-Colomer, M., Vandeputte, D., et al. (2016). Population-level analysis of gut microbiome variation. Science *352*, 560-564. 10.1126/science.aad3503.
110. Leshchiner, D., Rosconi, F., Sundaresh, B., Rudmann, E., Ramirez, L.M.N., Nishimoto, A.T., Wood, S.J., Jana, B., Buján, N., and Li, K. (2022). A genome-

wide atlas of antibiotic susceptibility targets and pathways to tolerance. Nature Communications *13*, 3165.
111. van Opijnen, T., and Camilli, A. (2012). A fine scale phenotype–genotype virulence map of a bacterial pathogen. Genome research *22*, 2541-2551.
112. Ng, K.M., Aranda-Diaz, A., Tropini, C., Frankel, M.R., Van Treuren, W., O'Loughlin, C.T., Merrill, B.D., Yu, F.B., Pruss, K.M., Oliveira, R.A., et al. (2019). Recovery of the Gut Microbiota after Antibiotics Depends on Host Diet, Community Context, and Environmental Reservoirs. Cell Host Microbe *26*, 650-665 e654. 10.1016/j.chom.2019.10.011.
113. Kim, J., Koo, B.-K., and Knoblich, J.A. (2020). Human organoids: model systems for human biology and medicine. Nature Reviews Molecular Cell Biology *21*, 571-584.
114. Van de Wiele, T., Van den Abbeele, P., Ossieur, W., Possemiers, S., and Marzorati, M. (2015). The Simulator of the Human Intestinal Microbial Ecosystem (SHIME((R))). In The Impact of Food Bioactives on Health: in vitro and ex vivo models, K. Verhoeckx, P. Cotter, I. Lopez-Exposito, C. Kleiveland, T. Lea, A. Mackie, T. Requena, D. Swiatecka, and H. Wichers, eds. pp. 305-317. 10.1007/978-3-319-16104-4_27.
115. Birkett, A., Muir, J., Phillips, J., Jones, G., and O'Dea, K. (1996). Resistant starch lowers fecal concentrations of ammonia and phenols in humans. The American Journal of Clinical Nutrition *63*, 766-772.
116. Mortensen, P.B. (1992). The effect of oral-administered lactulose on colonic nitrogen metabolism and excretion. Hepatology *16*, 1350-1356.
117. Wong, D.P., and Good, B.H. (2022). Quantifying the adaptive landscape of commensal gut bacteria using high-resolution lineage tracking. bioRxiv.
118. Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., and Causton, H.C. (2001). Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. Nature genetics *29*, 365-371.
119. Rustici, G., Williams, E., Barzine, M., Brazma, A., Bumgarner, R., Chierici, M., Furlanello, C., Greger, L., Jurman, G., and Miller, M. (2021). Transcriptomics data availability and reusability in the transition from microarray to next-generation sequencing. bioRxiv, 2020.2012. 2031.425022.
120. Goodman, A., Pepe, A., Blocker, A.W., Borgman, C.L., Cranmer, K., Crosas, M., Di Stefano, R., Gil, Y., Groth, P., and Hedstrom, M. (2014). Ten simple rules for the care and feeding of scientific data. PLoS computational biology *10*, e1003542.
121. Price, M.N., and Arkin, A.P. (2017). PaperBLAST: text mining papers for information about homologs. MSystems *2*, e00039-00017.
122. Karp, P.D., Ong, W.K., Paley, S., Billington, R., Caspi, R., Fulcher, C., Kothari, A., Krummenacker, M., Latendresse, M., Midford, P.E., et al. (2018). The EcoCyc Database. EcoSal Plus *8*. 10.1128/ecosalplus.ESP-0006-2018.
123. Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., and Clatworthy, M. (2017). The human cell atlas. elife *6*, e27041.
124. Javdan, B., Lopez, J.G., Chankhamjon, P., Lee, Y.J., Hull, R., Wu, Q., Wang, X., Chatterjee, S., and Donia, M.S. (2020). Personalized Mapping of Drug Metabolism by the Human Gut Microbiome. Cell *181*, 1661-1679 e1622. 10.1016/j.cell.2020.05.001.

125. Ostrem Loss, E., Thompson, J., Cheung, P.L.K., Qian, Y., and Venturelli, O.S. (2023). Carbohydrate complexity limits microbial growth and reduces the sensitivity of human gut communities to perturbations. Nature Ecology & Evolution, 1-16.
126. Han, S., Van Treuren, W., Fischer, C.R., Merrill, B.D., DeFelice, B.C., Sanchez, J.M., Higginbottom, S.K., Guthrie, L., Fall, L.A., and Dodd, D. (2021). A metabolomics pipeline for the mechanistic interrogation of the gut microbiome. Nature *595*, 415-420.
127. Johansson, M.E., Sjövall, H., and Hansson, G.C. (2013). The gastrointestinal mucus system in health and disease. Nature reviews Gastroenterology & hepatology *10*, 352-361.
128. Huus, K.E., and Ley, R.E. (2021). Blowing hot and cold: body temperature and the microbiome. Msystems *6*, e00707-00721.
129. Shalon, D., Culver, R.N., Grembi, J.A., Folz, J., Treit, P.V., Shi, H., Rosenberger, F.A., Dethlefsen, L., Meng, X., and Yaffe, E. (2023). Profiling the human intestinal environment under physiological conditions. Nature, 1-11.
130. Li, Y., Petrov, D.A., and Sherlock, G. (2019). Single nucleotide mapping of trait space reveals Pareto fronts that constrain adaptation. Nature ecology & evolution *3*, 1539-1551.
131. Finkel, S.E. (2006). Long-term survival during stationary phase: evolution and the GASP phenotype. Nature Reviews Microbiology *4*, 113-120.
132. Gevers, D., Knight, R., Petrosino, J.F., Huang, K., McGuire, A.L., Birren, B.W., Nelson, K.E., White, O., Methe, B.A., and Huttenhower, C. (2012). The Human Microbiome Project: a community resource for the healthy human microbiome. PLoS Biol *10*, e1001377. 10.1371/journal.pbio.1001377.