Cell Reports

CellPress
OPEN ACCESS

Review

Randomly barcoded transposon mutant libraries for gut commensals I: Strategies for efficient library construction

Surya Tripathi,[1,12] Carlos Geert Pieter Voogdt,[2,3,12] Stefan Oliver Bassler,[2,4] Mary Anderson,[5] Po-Hsun Huang,[6] Nazgul Sakenova,[2] Tümay Capraz,[2,6] Sunit Jain,[7] Alexandra Koumoutsi,[2] Afonso Martins Bravo,[8] Valentine Trotter,[9] Michael Zimmerman,[3] Justin L. Sonnenburg,[7,10] Cullen Buie,[6] Athanasios Typas,[2,3,*] Adam M. Deutschbauer,[1,9,*] Anthony L. Shiver,[11,*] and Kerwyn Casey Huang[7,10,11,*]

[1]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA 94720, USA
[2]Genome Biology Unit, EMBL Heidelberg, Meyerhofstraße 1, 69117 Heidelberg, Germany
[3]Structural and Computational Biology Unit, EMBL Meyerhofstraße 1, 69117 Heidelberg, Germany
[4]Faculty of Biosciences, Heidelberg University, Grabengasse 1, 69117 Heidelberg, Germany
[5]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[6]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[7]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA
[8]Department of Fundamental Microbiology, University of Lausanne, 1015 Lausanne, Switzerland
[9]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[10]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305, USA
[11]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA
[12]These authors contributed equally
*Correspondence: typas@embl.de (A.T.), amdeutschbauer@lbl.gov (A.M.D.), ashiver@stanford.edu (A.L.S.), kchuang@stanford.edu (K.C.H.)
https://doi.org/10.1016/j.celrep.2023.113517

SUMMARY

Randomly barcoded transposon mutant libraries are powerful tools for studying gene function and organization, assessing gene essentiality and pathways, discovering potential therapeutic targets, and understanding the physiology of gut bacteria and their interactions with the host. However, construction of high-quality libraries with uniform representation can be challenging. In this review, we survey various strategies for barcoded library construction, including transposition systems, methods of transposon delivery, optimal library size, and transconjugant selection schemes. We discuss the advantages and limitations of each approach, as well as factors to consider when selecting a strategy. In addition, we highlight experimental and computational advances in arraying condensed libraries from mutant pools. We focus on examples of successful library construction in gut bacteria and their application to gene function studies and drug discovery. Given the need for understanding gene function and organization in gut bacteria, we provide a comprehensive guide for researchers to construct randomly barcoded transposon mutant libraries.

INTRODUCTION

With the recent explosion of interest in microbiomes from various hosts and across the planet, it is becoming increasingly clear that the relatively few currently established model organisms do a poor job representing major functional segments of the vast microbial diversity on our planet. The human gut microbiome is a salient example, in which the microbial metagenome represents hundreds of millions of unique genes, but only 60% have annotated functions[1] and only a small fraction have been experimentally verified in a handful of model organisms. Interest in particular bacteria has typically been driven by association with a host phenotype or disease state, and the causality of such connections is usually tested via either colonization into complex microbiotas that lack the focal species (top-down approaches) or using assembled synthetic communities (bottom-up approaches).[2–6] However, we lack fundamental understanding of the physiology of most bacterial species, particularly insight into how their genetic information contributes to specific bacterial phenotypes.[2,7] Genome-scale mutant libraries in which genes are deleted or perturbed (e.g., by transposon mutagenesis, targeted deletions, or CRISPRi) have proven a powerful tool for systematic and unbiased discovery of genotype-phenotype relationships. Such libraries have been constructed in several microbial species,[8–23] often with a focus on pathogens or established model organisms, and have contributed substantially to the efficient identification of gene function and organization into networks.[11,24]

Using the forward genetics approach of measuring gene fitness of loss-of-function or gain-of function mutant libraries across various conditions,[25,26] genetic associations can be uncovered based on shared phenotypes, and genes can be

annotated and mapped to genetic modules or pathways. One way to screen libraries is as pools in which phenotypes of all mutants under a common environmental perturbation are interrogated simultaneously. Pooled libraries enable high-throughput identification of the genes important for growth in many conditions, which has many utilities, including highlighting potential inhibitory targets for antibiotics. However, it is challenging or impossible to use pooled screens to measure phenotypes unrelated to fitness (e.g., biofilm formation) or fitness-related phenotypes related to extracellular action, such as molecule/toxin secretion or product degradation, since these are masked by other mutants in the pool. Similarly, determining which mutants are important in the context of a host disease state is difficult in a pooled format if the underlying mechanism is not directly tied to growth. While microencapsulation methods such as droplet transposon sequencing (dTn-Seq)[27,28] can be applied to pooled libraries to discover genes involved in such growth-independent phenotypes,[29] arrayed libraries enable the study of each mutant in isolation as well as direct evaluation of candidate mutants.[12,13,18,20–22,24]

Transposon-insertion mutant libraries have risen in popularity since the development of methods for high-throughput fitness quantification empowered by massively parallel sequencing, including transposon insertion sequencing (Tn-Seq), insertion sequencing (INSeq), transposon directed insertion sequencing (TraDIS), and high-throughput insertion tracking by deep sequencing (HITS).[19,30–32] In these methods, a genome-scale library of mutants with disruptions in non-essential genes is constructed by introducing a transposon that integrates randomly into the genome. Sequencing the pooled library after exposure to a perturbation enables quantification of the relative abundance of each mutant, which is a proxy for fitness in the tested condition. Several transposon systems, such as Tn*5* and Himar1/Mariner, can integrate at high frequency relatively uniformly across the genome[33] and thus can be used to create dense libraries with a high fraction of mutants that carry the transposon inserted in the central region of a gene to maximize gene disruption. In this review, we will focus on the application of transposons for gene disruption, but it is important to note that many variations based on transposon mutagenesis, such as gene overexpression, have been developed to address other questions (reviewed in Cain et al.[29]).

Inspired by barcode analysis by sequencing (Bar-Seq[34]), the introduction of random barcodes into transposons that are linked to the genomic integration site through an initial round of insertion sequencing has dramatically increased the throughput and decreased the cost of experiments.[35] This strategy, termed randomly barcoded Tn-Seq (RB-Tn-Seq), transforms all screens from the costly (in time and money) protocols developed earlier for Tn-Seq into a single-step PCR to amplify the unique transposon barcode. Other than the effort required initially to introduce barcodes with high diversity, there are no obvious drawbacks to the use of barcoded libraries; hence, in this review we will focus on transposon insertion mutants that are barcoded.[29] Many of these improved genome-scale genetics methodologies have accelerated the discovery of microbial gene functions[11] and the study of non-model organisms. Several bacteria from diverse phyla present in the human gut have already been targeted with this approach, including *Bacteroides* species,[19,36–39] *Lactobacillus casei*,[40] *Enterococcus faecalis*,[41] *Clostridium* species,[42,43] and *Akkermansia muciniphila*,[44] providing insights into their physiology and genetic architecture. These advances have uncovered just the tip of the iceberg, and to start unraveling bacterial gene function in the human gut and other microbiomes necessitates targeting many more non-model organisms using similar approaches.

In this review, we focus on the generation of genome-scale randomly barcoded mutant libraries via transposon mutagenesis for diverse bacterial species with a focus on members of the human gut microbiome. We discuss methodologies for vector design and transformation, as well as technical challenges and pitfalls often encountered in the generation of saturated, pooled transposon mutant libraries. We present strategies for efficient arraying of pooled libraries and identification of the location of mutants within the arrayed library and argue for improved planning for library distribution. Finally, we present a roadmap for prioritization of non-model bacterial species from the human gut microbiome as candidates for future library generation. Our goal is to provide a general guide and outlook that motivates and enables the efficient construction of mutant libraries in relevant gut commensals. In the accompanying review by Voogdt et al. in this issue of *Cell Reports*,[45] we focus on applications of such mutant libraries to gene function discovery in gut bacteria via large-scale genotype-phenotype mapping.

DESIGN OF A POOLED TRANSPOSON MUTANT LIBRARY

Many factors have an impact on the overall utility of a pooled transposon mutant library, including target library size, transposition system, transposon design, and potential issues with library diversity. The topics discussed below provide a general foundation for library design, but species-specific modifications and considerations may be important for successful library construction in certain bacterial species.

Target library size

A library for RB-Tn-Seq should ideally contain at least ~10 mutants in the central region (20%–80%) of each gene (where insertion of the transposon is most likely to disrupt gene function) so that insertions in the same gene can be compared for consistency to verify that phenotypes are in fact due to inactivation of the gene. Due to insertion biases, some genes will end up represented by more than 10 mutants. The minimum target library size can be estimated using the Newman-Shepp generalization of the coupon counting problem,[46] such that a library with at least 10 mutants in the central 20%–80% of each gene of a typical bacterial species with an ~5-Mb genome and ~5,000 genes requires a library of at least ~268,000 mutants. This estimate increases further if a significant fraction of the genome is intergenic or non-coding and may also require correcting due to uneven gene size distribution and insertion bias. Constructing dense libraries with hundreds of thousands of mutants increases the odds of disrupting small (<50 residues) proteins and non-coding RNAs, which can play important physiological roles.[47] Taking these factors into account, ideal libraries may require up to 500,000 mutants. The ability of forward genetics to discover

gene functions and map genetic networks generally improves as more conditions are tested[25] and with the density of the mutant library, motivating larger library sizes. However, accurate mapping of insertions and barcode matching becomes challenging in hyper-diverse libraries, and the potential for bottlenecks increases, as discussed in detail below. In addition, for downstream screens involving barcode sequencing, the distribution of a fixed number of reads across all barcodes could lead to an undesirably low number of reads per barcode in hyper-diverse libraries.

Transposition systems

Achieving tens of insertions in each gene is assisted by using a transposition system that functions independent of host factors and has minimal specific sequence requirements for insertion. Two of the most popular systems for transposon mutagenesis in bacteria include the Tn*5*[48] and mariner/Tc1[49] transposase families, neither of which requires host factors to function. Whereas Tn*5* has little sequence insertion bias, mariner/Tc1 transposases, such as the commonly used Himar1, insert preferentially at TA dinucleotide sites. Even with little sequence insertion bias, transposons typically insert near the chromosomal origin of replication with higher frequency than in other regions of the genome, simply due to the greater amount of origin-proximal DNA, a phenomenon that is exacerbated in conditions in which bacteria grow rapidly and undergo multi-fork replication.[50] The baseline expectation, therefore, is an approximately log-linear decrease in insertion density away from the origin.[50] Tn*5* and mariner transposons have been used successfully for transposon mutagenesis in a wide range of bacteria, including (among others) Proteobacteria (Pseudomonadota),[12,14,20,22] Bacteroidetes (Bacteroidota),[11,19,38] Clostridia,[43,51] *Lactobacillus*,[40] and *Treponema*[52] species. Thus, it is likely that these transposases will have broad utility for library construction in gut commensals.

Transposon design

In addition to the transposase, other elements such as the promoter (driving the transposase and the selection marker), the ribosomal binding site (RBS), and the selection marker (typically a gene that confers resistance to an antibiotic) require careful design and optimization to maximize efficiency and minimize biases in the library. The magic pools methodology is a combinatorial strategy for constructing a catalog of thousands of uniquely assembled transposon delivery vectors from a collection of parts (transposases, promoters, RBSs, selection markers)[53] (Figure 1). This method streamlines the process of identifying vectors that have the optimal combination of parts by simultaneously testing the efficacy of all possible vectors in the catalog to generate mutants in a new species of interest. To achieve high expression of the selection marker, sets of strong promoters and RBSs are selected from conserved genes and from the literature, including phage-derived sequences.[54,55] The candidates for each of the four parts (transposase, promoter, RBS, and selection marker) along with a barcode that uniquely identifies each combination of these four elements are assembled into a catalog of vectors (the magic pool), and long-read sequencing (e.g., Nanopore or PacBio) is used to map barcodes to vectors. The strain of interest is mutagenized with the magic pool, and transposon insertion sequencing is used to identify which combination of parts was most prevalent and least biased and hence most successful at mutant generation (Figure 1). It is important to generate enough mutants for the insertion sequencing analysis to determine whether the most successful part combination introduces any gene or strand insertion biases. Ultimately, a single vector is then assembled with the optimal combination of parts for construction of the final mutant library of interest that minimizes biases.

While the magic pool approach can in principle be applied to any species, several factors can have an impact on efficiency and hence should be considered in the design of a magic pool. If the delivery vector contains sequences that are known targets for degradation by the restriction modification systems of the recipient, these vector sequences should be altered. Magic pool tests to date have shown that, while transposase promoters of different strengths all successfully produce mutants (as the transposase needs only to function transiently), the promoters and RBSs that drive the selection marker are critical.[35] If the promoter driving the selection marker is too weak, then mutants with transposon insertions in the same transcriptional orientation as the mutated gene will predominate, because mutant selection will be aided (at least in part) by the native promoter. In addition, weak promoters will lead to biases in genes that are strongly expressed in the condition in which the library is constructed, since expression of the selection marker will be driven mainly by the promoter of the strongly expressed gene. On the other hand, if the promoter is too strong, insertion of the transposon may cause artificial overexpression of downstream genes or silencing of antisense genes (although insertions in only one strand can represent a specific phenotype[56]). This scenario can be prevented by including a terminator sequence in the transposon, thereby isolating expression of the selection marker. However, such terminators will introduce polar effects for operons, which will be particularly apparent in cases in operons containing essential genes.[32,57] With or without a transcriptional terminator, the transposon may have an impact on natural gene expression; thus, phenotypes found using transposon libraries should be verified with complementary methods such as clean deletions.

In addition to promoter and RBS optimization, codon optimization of the selection marker for the target organism may improve expression.[58] For understudied organisms, resistance markers from evolutionarily distant species may function suboptimally or not all. In these cases, (meta)genomic data can be mined to identify potential selection markers beyond the most commonly used antibiotic resistance genes that are likely to function in the organism of interest.[59]

In addition to enhanced identification of a construct that is optimal for a given species, magic pool design can also potentiate broad utility for related organisms by including more general parts. Certain constructs that exhibited high efficiency in *Bacteroides thetaiotaomicron* can also be effective with other *Bacteroides* species.[60,61] One strategy for realizing this generalizability is to employ promoters with a broad host range among the higher-level taxon (e.g., family) of interest. Selection of these promoters could be facilitated by *in silico* identification of promoters of essential genes (which are more likely to be constitutive) and

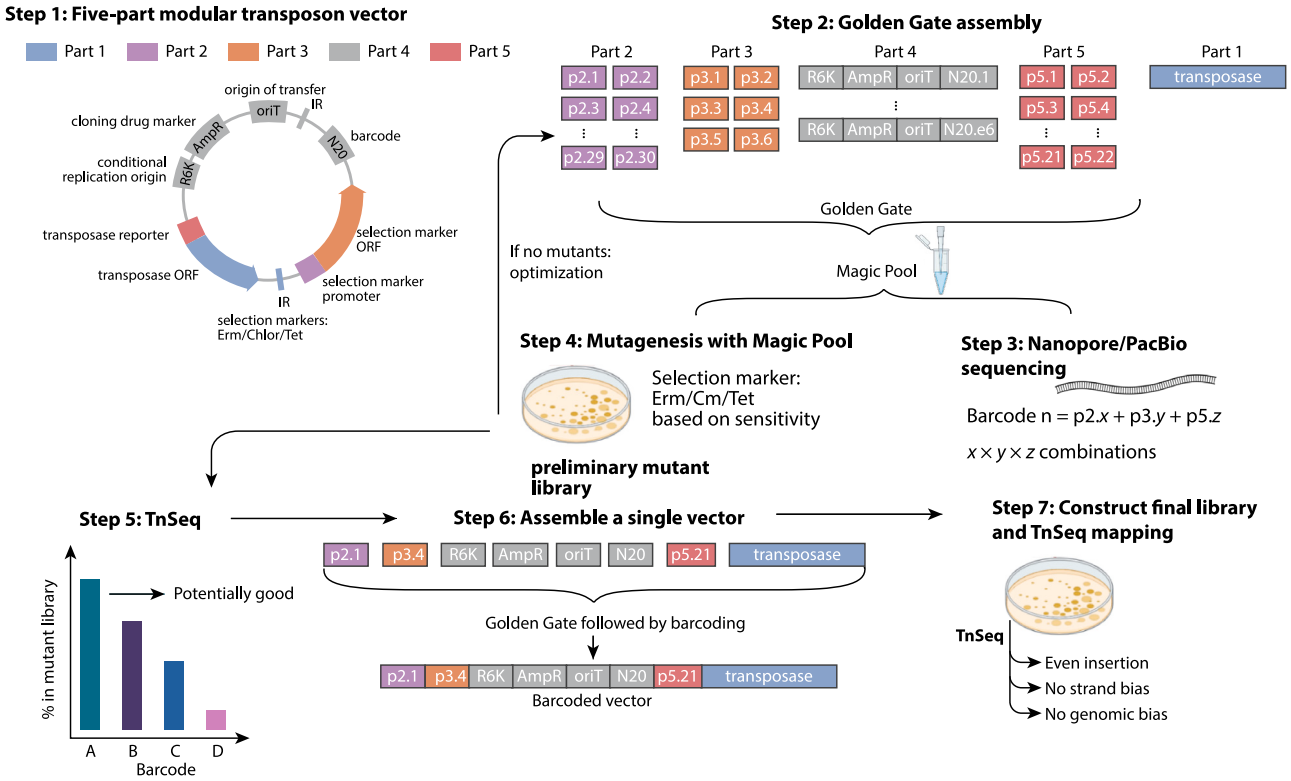

Figure 1. Magic pool-based vector design pipeline

"Magic pools" refers to a systematic strategy for constructing libraries of transposon delivery vectors from a collection of parts (promoters, antibiotic resistance genes; step 1) that can be tested for efficacy in parallel to identify the optimal vector for a given bacterial strain. A single Golden Gate reaction with all parts produces thousands of transposon vectors, each with a unique barcode (step 2). This library of thousands of transposon vectors is sequenced using long-read sequencing to map each barcode to its complete vector design (i.e., its component parts) (step 3). The magic pool is introduced into an *E. coli* conjugation donor strain and used to conjugate with a recipient, and the transconjugants are selected with appropriate selective media (step 4). DNA is purified from the transconjugants and transposon sequencing (Tn-Seq) is used to identify the vector designs most effective at generating transposon mutants without insertion biases (step 5). Single vectors with effective performance are designed based on the Tn-Seq data (step 6), and the final library is prepared by conjugating recipients with an *E. coli* donor carrying this single vector (step 7).

RNA-sequencing (RNA-seq) experiments to identify strong promoters from highly expressed genes.[62]

Potential issues with library diversity

It is important to measure the diversity of the library (barcodes and insertion sites), as it is a primary quality feature and helps to estimate the impact of bottlenecking during experiments.[7] Colony-forming unit (CFU) counts should be used only as a rough estimate of library diversity, since CFU counts can be misleading if the species typically exhibits a range of colony sizes during surface growth, such that the true diversity of mutants is larger than what is apparent from large colonies alone. To assess diversity and undesirable multiple integration events, the transposon insertion site and associated barcodes should be determined by either Sanger sequencing of a reasonable number of single colonies or Illumina sequencing of a small pool of a known number of colonies. Moreover, one way to rapidly assess barcode diversity is to sequence only the barcodes of the pooled library (by single-step PCR,[11,35] Bar-Seq), which reports on the unique number of barcodes found and the relative abundance of each barcode. Thus, Bar-Seq also reports on potential barcode biases in the library. Ultimately, deep sequencing of promising libraries via RB-Tn-Seq should be performed to characterize the number of unique barcodes and insertion locations as well as biases and barcode-gene linkages.

While a large number of mutants is typically desirable, if the library is too diverse (i.e., most genes are represented by >>10 barcodes), then population bottlenecking can become an issue. In addition, a very diverse library can contain multiple mutants that map to the same barcode if the conjugation donor library is insufficiently diverse, which should be avoided, because these reused barcodes cannot be uniquely assigned to one gene. If the number of cells representing each barcoded mutant in the inoculum of a given experiment is small, stochastic fluctuations in mutant abundance will prevent accurate comparison of barcode relative abundance in each condition with the reference.[63–65] The scale at which bottlenecks become important will depend on the nature of the experiment. For instance, a library that has appropriate diversity for monoculture experiments may be too diverse for culturing in combination with other species given the same culture volume that reduces the number of input cells from the

library and is low enough to result in a bottleneck. Bottleneck issues can be overcome by lowering the diversity of the library, which can be achieved by constructing the library in many pools of which a subset are combined to achieve the desired diversity or by repooling an arrayed subset of the initial pooled library, also referred to as condensing. Condensing into pooled subsets of mutants can also be a powerful approach to focus on specific functions such as carbohydrate metabolism in a pooled context.

METHODOLOGIES FOR CONSTRUCTION OF A POOLED TRANSPOSON MUTANT LIBRARY

The choice of method for introducing DNA is usually constrained by the organism being targeted. Phage transduction, transformation by natural competence (direct uptake), and conjugation (transfer from a donor organism) are naturally occurring mechanisms of horizontal gene transfer (HGT) among bacteria, including most phyla in the human gut microbiome.[66,67] The two most common approaches for constructing transposon mutant libraries are conjugation and electroporation; each requires optimization of method-specific parameters for DNA delivery, which we discuss below.

Introduction of DNA into the target organism

Generation of a library with hundreds of thousands of mutants requires that the target organism can be transformed with sufficiently high efficiency. Yet, transformation of the vast majority of gut commensals has not been systematically investigated. In addition, target organisms may express defensive restriction modification systems that can pose major limitations to the successful introduction of foreign DNA. Understanding the DNA methylation patterns in target organisms and utilizing this information to construct correctly methylated transposon vectors may be a viable strategy to target more gut bacterial species for library construction.[68,69] There are multiple ways to introduce DNA into bacteria; the two most common are conjugation and electroporation (Figure 2).

Conjugation-based transposon introduction

Conjugation is often used to introduce a transposon vector into the target species. In conjugation-mediated mutagenesis, the transposon vector is transferred from a donor cell into the cytoplasm of the target cell. Once in the target cell, the transposase is expressed from the specific target cell promoter and excises the transposon from the vector for insertion in the genome.

Conjugation: Beneficial features of the donor strain

Several features of the donor strain should be considered when using conjugation for transposon introduction. The donor (typically *Escherichia coli*) should have high transformation efficiency to achieve high barcode diversity in the plasmid pool. Typically, electroporation-based transformation of the donor strain with the barcoded transposon plasmid library (selecting for pools of transformants, not single colonies) results in high efficiency. Depending on the transposon vector design, the donor may need specific genes for plasmid maintenance. The γ origin of replication of the R6K replicon is useful in transposon vector design since replication occurs only in the donor strain, which carries the *pir* gene from lambda phage.[70] Most target species do not carry the *pir* gene, preventing the transposon vector from replicating in the recipient cell.

Along with genes for plasmid maintenance, the donor strain should also carry genes required for transfer of the transposon vector to recipient cells. A widely used system for conjugation-based DNA transfer is derived from the broad-host-range plasmid RP4.[71,72] Genes in the *tra1* and *tra2* regions of the RP4 plasmid encode conjugation machinery that is expressed by the donor cell and does not require specific features on the surface of the recipient for mating pair formation and transfer of DNA. As such, the RP4 conjugation system has an extremely broad range spanning gram-negative and gram-positive bacteria,[73] archaea,[74] and even eukaryotes.[75] Thus, unsuccessful conjugal transfer from an RP4-based system may indicate factors in the recipient that antagonize or kill the donor strain, suppress transfer genes through quorum-sensing mediators,[76] or encode defense systems such as CRISPR-Cas or restriction modification enzymes against foreign incoming DNA.[77] Deletion of such systems in target species can improve conjugation efficiency, but this action requires other genetic tools to be available in the target species, which is often not the case. It is also possible that conjugation failure could be remedied using a donor other than *E. coli*.[78,79]

Conjugation: Counter-selection of the donor strain

After conjugation, the donor strain must be removed from the donor-recipient mixture. One strategy for counter-selection is to grow the mixture on a medium that contains an antibiotic to which the recipient strain is resistant, but the donor strain is sensitive. Alternatively, use of a donor strain that is auxotrophic for the cell-wall component diaminopimelic acid (DAP) through deletion of the *dapA* gene allows for counter-selection by omitting DAP from the culture medium post-conjugation.[80] Combining these two approaches is the best counter-selection method in practice. The *E. coli* strains often used as conjugation donors, including *E. coli* S17-1(λpir), SM10(λpir),[81] Mu-phage-free donor (MFDpir),[82] and β2155,[83] all have the *pir* gene and RP4 transfer genes, and the latter two are DAP auxotrophs.

Conjugation: Features of the recipient strain

Prior to evaluating conjugation efficiency and quality, the sensitivity of the target organism to the applied selection (antibiotic and oxygen, if conjugation is performed aerobically) should be characterized. Whether the target organism is an efficient recipient of conjugal transfer must be established empirically, since strains of the same species can exhibit remarkably different conjugation efficiencies.[84] Studies with *E. coli* and *Bacillus subtilis* as recipients have shown that these species do not have particular genes strictly necessary for conjugation permissibility or resistance, although a small number of genes that enable modifications to the cell wall can affect conjugation efficiency[85,86]; thus, it may be possible to select for mutants with higher efficiency.

It has been proposed that conjugation with gram-positive target organisms and/or in strictly anaerobic conditions may be highly inefficient.[87] However, within the largely anaerobic mouse gut, the *E. coli* S17-1 strain can conjugate transposon

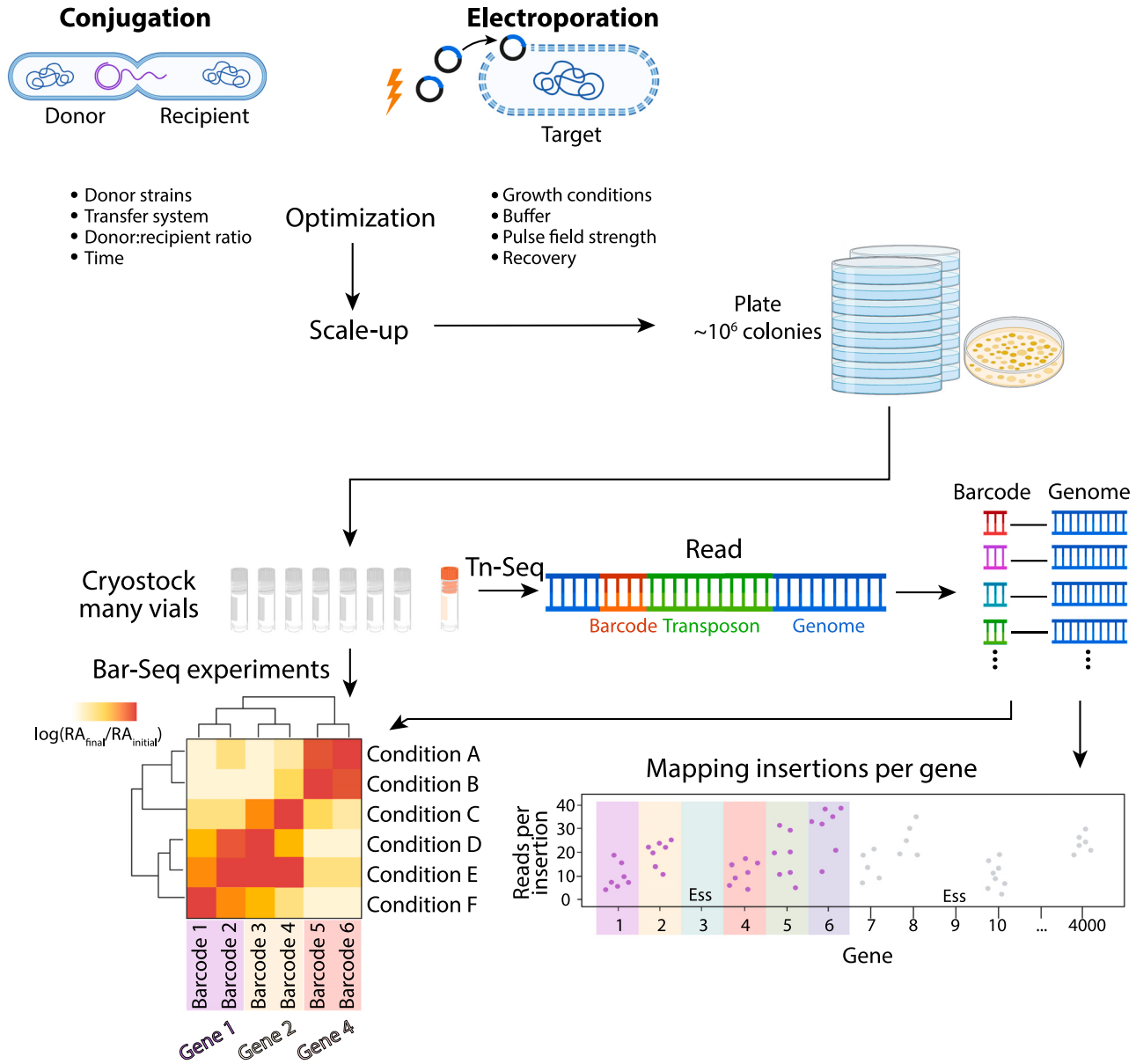

Figure 2. Transformation and construction of a transposon mutant library

Conjugation and electroporation are the most efficient ways to introduce a transposon vector into the target organism. Either method will require optimization of method-specific parameters for each target organism. Following optimization, a large number of transformants or transconjugants are plated on selective agar plates and grown to colony size. Colonies are collected *en masse* and cryostocked in many single-use vials. One representative vial is processed for transposon insertion sequencing (Tn-Seq), which produces short reads that capture the transposon-barcode-genome junction. A barcode-genome insertion site map is created for future Bar-Seq experiments. In addition, mapping of insertions across the genome reports on library density and uniformity, and on essential (Ess) genes that cannot tolerate insertions. Bar-Seq experiments should be performed from the same batch of cryovials as the ones that were sequenced. Bar-Seq experiments on the mutant library across various conditions report on the conditional fitness (log of the ratio of final to initial relative abundance [RA]) for all genes represented in the library. Barcodes from mutants with insertions in the same gene tend to have similar phenotypes within a given condition, increasing the statistical robustness of fitness analysis.

vectors with a phylogenetically diverse range of recipient organisms, including gram-positive species.[88] Moreover, the *E. coli* β2155 strain can conjugate *in vitro* with strict anaerobes that cannot be exposed to any oxygen to preserve viability.[89] Alternative conjugation donors such as *B. subtilis* can also be used to conjugate with gut anaerobes.[90] In addition, the mobile CRISPRi method[16] uses *B. subtilis* as a donor to deliver a CRISPRi vector to Firmicutes. The factors that influence conjugation efficiency under aerobic and anaerobic conditions remain unclear; deeper understanding could promote the engineering of

donor strains that can transfer transposon vectors and other genetic cargo into phylogenetically diverse and highly oxygen-sensitive gut bacterial species. Furthermore, optimization of the growth medium and growth phase (log versus stationary) for the donor and recipient should be considered for each donor-recipient pair.

Electroporation-based transposon introduction

Electroporation involves the application of an electric field to create a transmembrane potential that overcomes the hydrophobicity of the lipid bilayer membrane and creates multiple transient pores for passage of exogenous material such as nucleic acids or proteins. Electroporation can be used to transform a target organism for transposon mutagenesis by introducing either a transposon vector (DNA only) that is transcribed upon entry[91] or *in vitro*-assembled transposomes, which are complexes of a transposon and a transposase, the latter of which integrates the transposon into the chromosome upon entry.[35]

While electroporation has proven a powerful tool for the introduction of foreign DNA in certain organisms, most efforts to improve its efficiency have focused on model organisms such as *E. coli*. However, even different strains of *E. coli* exhibit efficiencies that can vary by more than six orders of magnitude.[92] Thus, it is likely that each organism will require individualized optimization focusing on details such as buffer composition, electrical-field strength, pulse shape/length, and growth conditions,[93] at least in part due to differences in how these variables affect the structure of the cell envelope. Using growth media in which cell growth is inhibited has been proposed to increase efficiency, potentially through weakening the cell envelope,[94] although the generality of this phenomenon has yet to be determined. An often-ignored variable is growth history, including the duration and phase of growth prior to electroporation, which may reduce the integrity of the cell envelope, like growth-inhibiting media. Moreover, heat shock prior to electroporation can increase electroporation efficiency,[95] potentially by swelling cells and/or changing membrane fluidity.

FUTURE PROSPECTS FOR TRANSFORMATION METHODOLOGIES

While powerful, the current approaches for conjugation and electroporation face some limits on their general applicability. Below, we discuss a starting point for promising new strategies to enhance the utility of these approaches.

Conjugation

Although the RP4 conjugal system has proven extremely effective in transposon mutagenesis for a wide phylogenetic range of species, and thus should be the first candidate system for generating mutant libraries, characterization and engineering of other conjugal systems may be necessary to target an even wider range of species. Like conjugative plasmids, integrative conjugative elements (ICEs) and conjugative transposons encode genes for their own transmission through conjugation but, unlike plasmids, need to integrate into the host genome for stable acquisition (similar to prophages). In a recent seminal study, the widespread ICE *Bs1* of *B. subtilis* was engineered and reintroduced into a *B. subtilis* donor strain that can transfer custom DNA with high efficiency into many gram-positive bacteria, including human gut commensals.[79] Recently, hybrid elements that combine features of two ICEs have been shown to increase conjugation efficiency and host range to a level higher than that of either individual ICE.[96] This finding suggests that construction of other hybrid elements could enable targeting of species in which efficiency is otherwise too low for library construction. Further investigation into the transmission and regulation of other mobile genetic elements will undoubtedly yield similarly powerful tools for the mutagenesis of species across the human gut microbiome.

Electroporation

Currently, developing a new electroporation protocol for a non-model organism typically involves blind exploration of parameter space, although hopefully it will become possible in the future to systematically adapt protocols from other organisms once a larger knowledge base has been established. Fortunately, new tools are being developed to expand the space of electroporation parameters that can be explored. Many commercial electroporators cannot adjust variables such as pulse time; hence, the time constant that dictates the duration of pore formation is based solely on the conductivity of the sample/cell type. However, some commercial devices can change capacitance to enable increasing the pulse time without using high field strength (which reduces viability).[97] Moreover, microfluidic devices can facilitate rapid screening of a large number of electroporation conditions[98] and high-volume transformation of hundreds of milliliters per minute.[92] Yet, even though systematic exploration of the parameter space is possible, it remains time consuming to produce competent cells; strategies for washing cells without centrifugation have thus far been limited to small volumes.[99] With improved competent cell preparation and inexpensive transposase enzyme, it may be possible to capitalize on high-volume devices[92] to generate a library even when the efficiency of transformation is low.

CONSTRUCTION OF A GENOME-SCALE POOLED LIBRARY

A suitably efficient transformation protocol for the target organism is necessary but not sufficient for the creation of a genome-scale mutant library. Ensuring that the library is unbiased, high quality, and suitable for long-term use requires attention to at least the following aspects (Figure 2).

Library outgrowth

After transformation, the library should be grown under selective pressure on a medium that can support the growth of most mutants. Hundreds of genes can be essential for growth in a particular medium[11,24]; thus, selection of the medium for library outgrowth introduces an intrinsic bias in terms of which genes will have viable insertions. A rich medium that supports high biomass yield is most likely to limit such biases, for instance, to avoid selecting against pathways for a specific carbon source.[100] Nonetheless, a few hundred genes are typically universally essential and required for the cell to survive in all conditions.[101] This set of core genes is usually devoid of insertions in

the library, although rare insertions in core essential genes can be found in the first or last part of the open reading frame (ORF), deleting non-essential parts of the protein. In general, transposon mutagenesis has been a very powerful tool for probing gene essentiality and even the essentiality of protein domains for sufficiently saturated libraries.[102]

After mutagenesis, the library is typically expanded via growth on selective agar plates and/or in liquid. Cells are then collected from the agar and grown in liquid to cryopreserve the library. Alternatively, the library can be outgrown only in liquid. Care should be taken to limit library outgrowth to avoid redistribution of mutant abundances and extinction of mutants with severe fitness defects. When grown on agar, the library has to be plated on tens or even hundreds of plates and grown until small colonies appear. Library expansion and thus the number of generations can be better controlled in liquid medium by tracking the optical density of the culture. It is often argued that liquid outgrowth carries the risk of selective expansion of mutants with higher fitness, although colony growth on a plate also introduces inevitable biases due to competition between neighboring colonies for nutrients.[103] After expansion of the library under selection, it is important to minimize further outgrowth by saving hundreds of single-use aliquots. For instance, by minimizing outgrowth, more insertions in genes that result in sick mutants can be maintained in the library and their phenotypes later measured reproducibly when the library is profiled across conditions.

Library sequencing

DNA from one of the single-use library aliquots should be prepared for insertion sequencing (Figure 2) using a method such as Tn-Seq,[32] TraDIS,[31] or semi-random PCR.[104] The transposon-inserted genome amplicons can then be sequenced on high-content, short-read Illumina platforms. Ideally, every insertion in the pool will be represented by multiple high-quality reads, but the required sequencing depth for a transposon mutant library is difficult to estimate prior to sequencing, since it is mainly influenced by library diversity. For libraries constructed with a mariner transposon, the maximal library diversity can be estimated, since transposons insert predominantly (>95%) at TA dinucleotide sites.[105] For example, for an organism with 300,000 genomic TA sites in non-essential genes, the expectation is up to 600,000 insertions (since transposons can enter a TA site on the sense and the antisense strand). For such a library, the target sequencing depth as computed using the Newman-Shepp generalization is $\geq$25 million reads per experiment; note that the required number of reads is often underestimated, leading to inadequate sequencing depth for some mutants. Following sequencing, the genomic sequence of each read must be mapped and linked to the barcode sequence of the same read to establish the barcode-genome references from which future Bar-Seq experiments are quantified using the unique barcodes (Figure 2). In the accompanying review,[45] we discuss data analysis pipelines in more detail.

CONSTRUCTION OF AN ARRAYED LIBRARY

A high-diversity and high-quality pooled library also forms the basis for creation of an ordered, non-redundant mutant collection. Such an arrayed library is an invaluable tool for mechanistic studies, providing ready access to non-essential gene mutants and enabling screening for diverse phenotypes unrelated to fitness. First, a progenitor collection is generated by arraying mutants isolated from the pooled library. This progenitor collection is sequenced to determine which mutants have been captured and their location in the collection. Next, a subset of mutants with desired features is rearrayed from the progenitor collection into an ordered, non-redundant library (Figure 3). Several important factors, including features of the pooled library, the magnitude of certain biases, and arraying methodologies, should be carefully considered during the planning stages of library arraying.

Properties of the pooled library that affect arraying: Density

The pooled library should be as dense as possible so that mutants in every non-essential gene can in principle be obtained. Density is particularly important for isolating mutants in short ORFs encoding small proteins,[47] which are naturally expected to have below-average numbers of insertions in the pool. In addition, the pooled library should ideally have an even distribution and abundance (estimated by read counts) of mutants in each gene, since the isolation strategy is likely to be largely unbiased. Minimizing the propagation of the pooled library limits such biases and preserves mutants with severe fitness defects (nearly essential). Note that the concept of essentiality need not apply only to units of whole genes: for a very dense pooled library, gaps in insertion frequency can identify regions of ORFs (protein domains) that are essential.[106]

During arraying, there is inevitable inefficiency due to the repeated isolation of some mutants by chance. Based on sequencing of the pooled library, the number of unique genes that will be captured by a given number of isolated mutants can be predicted with reasonable accuracy, taking into account any nonuniformity in the distribution of mutants across genes in the pooled library.[18] Desirable features such as having the transposon in the same reading direction as the gene (to avoid polar effects),[32,57] insertion in the central 60% of the gene, and having a unique barcode require isolation of a larger number of mutants to cover a given fraction of the genes (Figure 3). For a typical genome with $\sim$5,000 genes, the number of mutants that should be isolated is likely to range from 20,000 to 60,000 (influenced by library quality and desired saturation level as well as resources available); beyond this point feasibility becomes questionable. A pilot run with $\sim$20% of the desired number is useful for uncovering biases, such as the number of wells without growth or wells with more than one strain (if isolation is performed with a cell sorter), early in the process and adjusting the target.[18]

Properties of the pooled library that affect arraying: Multiple integrations

In any pooled library, some fraction of mutants will contain multiple integrations that occur during conjugation when the donor cell passes multiple copies of the vector to the same recipient cell or when one recipient cell forms mating pairs with more than one donor.[18] The presence of multiple-integration strains is of little consequence when estimating gene fitness using pooled libraries as long as multiple-integration mutants are vastly

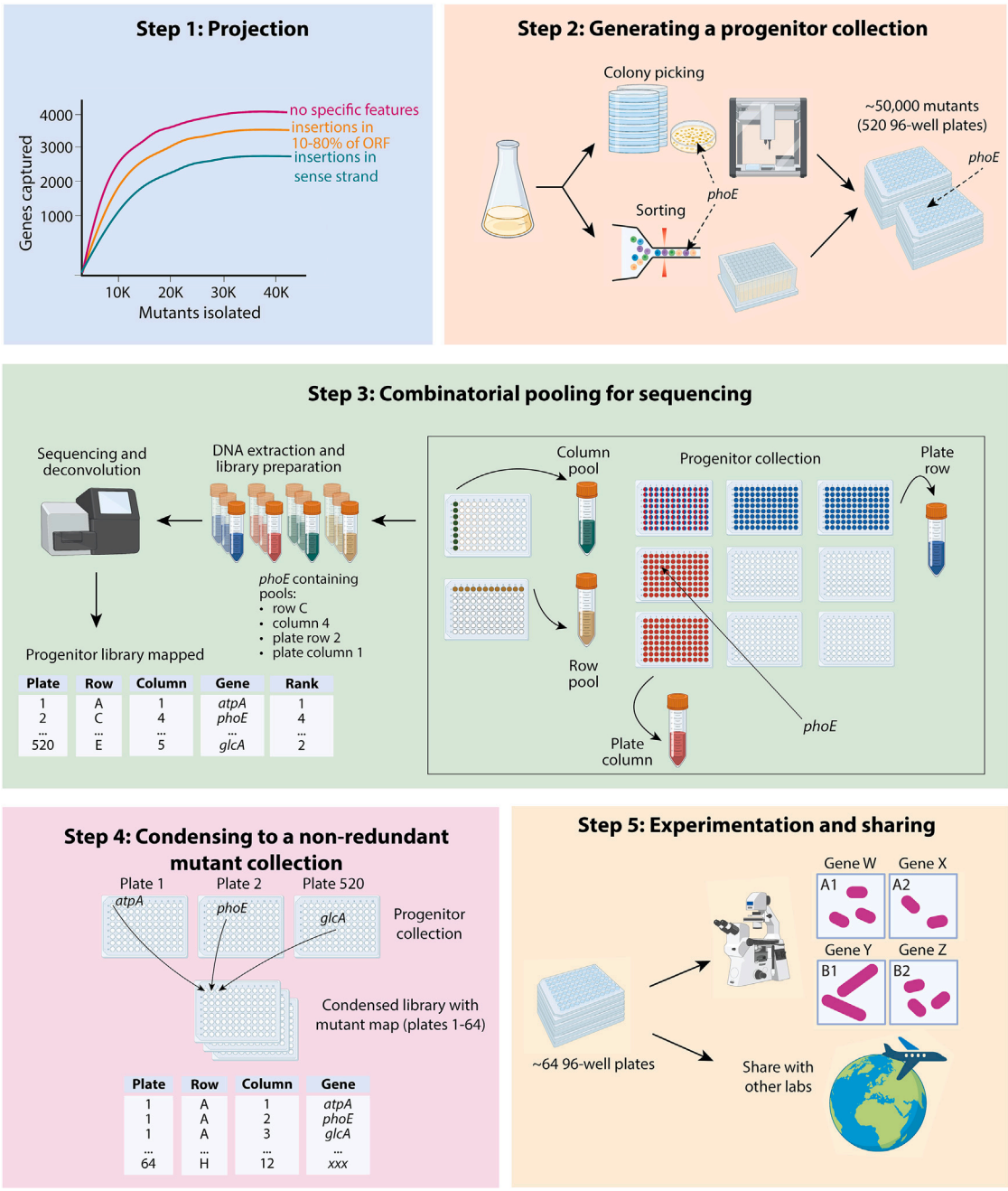

Figure 3. Construction of an arrayed mutant library

After creating and sequencing a pooled mutant library, the first step in constructing an arrayed library involves making a projection of the number of genes likely to be represented by a given number of mutants isolated (step 1). Aiming for specific insertion features will typically lower the predicted number of covered genes, since mutants that satisfy those requirements will be rarer. The projection defines the desired size of the progenitor collection of isolated mutants given a target number of covered genes. Starting from the pooled library, either colonies are picked or single cells are sorted into multi-well plates to isolate mutants (step 2). Specific mutants such as *phoE* are tracked as examples throughout the figure. After generating the progenitor collection, combinatorial pooling (in this case 4D) is used to make sequencing feasible (step 3). Every well is collected across four dimensions (column, row, plate column, and plate row). The mutants collected in each pool are sequenced using Bar-Seq, and computational deconvolution of pool compositions results in a map of mutant locations. Mutants can be ranked based on desirable features for generation of the final arrayed library. Desirable mutants from the progenitor collection are rearrayed into a smaller, non-redundant collection, and a final mutant map is generated (step 4). The condensed library should be sequenced again (after combinatorial pooling) to confirm correct placement of mutants. Arrayed libraries are a powerful tool for studying non-fitness-based phenotypes, such as cell morphology. Rapid dissemination and sharing with other labs facilitate gene function discovery (step 5).

Table 1. Comparison of mutant arraying methods

	Single-cell sorting	Robotic colony picking
Pros	fast, ~1 s/cell	fast (~1 s/colony)
	gating parameters allow for isolation of mutants with specific morphologies	settings may allow for picking aberrant colony shape, size, color, or height
	minimal library pre-growth is required	only viable mutants that form colonies are picked
	no consumables like Petri dishes or agar medium are required for pre-growth	
	very sick mutants can be sorted as single cells	
Cons	non-viable cells may be sorted, leading to empty wells	precise dilution is necessary to achieve optimal colony spreading
	substantial time may be required for outgrowth from a single cell	plating for colonies requires many agar plates, and colony development may require long incubation times
	cell sorter must be suitable for bacteria, which may be biosafety level 2 or higher	machines with needles require frequent, time-consuming cleaning and calibration
	cells with morphological defects such as filamentation or aggregation may not be sorted using standard gating parameters	

outnumbered by single-integration mutants of the same gene, and hence genetic interactions for the multiple random insertions in a given strain are extremely rare. However, the isolation of multiple-integration mutants during arraying is essentially wasted effort and means that one has to increase the number of mutants isolated to pick single-integration mutants, which ensures that phenotypes probed are due to disruption of only a single gene. The propensity of multiple integrations can be estimated by sequencing the transposon insertion and its barcode in a defined number of isolated mutants. For example, if insertion sequencing of a selection of 200 isolated mutants results in high-confidence identification of 200 unique barcodes and 240 unique genome insertion sites, then a reasonable estimate is that multiple integrations occurred ~20% of the time during library construction and such mutants are abundant in the pooled library.[18]

Properties of the pooled library that affect arraying: Unavoidable biases

If there is substantial bias in the pooled library (e.g., in mutant abundance or multiple integrations) and the source of that bias can be identified, it may be preferable to redesign library construction and regenerate a library for the sole purpose of arraying rather than tolerating the inevitable inefficiencies due to the bias. Note that certain biases such as gene size and the absence of TA sites in a gene, which will preclude insertion of mariner transposons,[105] are unavoidable and should be incorporated into the predicted outcome of library arraying.[18] Moreover, it is important to recognize the diminishing returns of isolating more mutants: the steepness of the saturation curve (Figure 3) indicates that it is practically infeasible to isolate a mutant for every non-essential gene, so the balance between time/resources and saturation must be considered.

METHODS OF ARRAYING

Arraying can be accomplished manually by plating the library on a solid medium and transferring individual colonies into multi-well plates in an arrayed format.[104] However, manual arraying is time consuming and can become increasingly error prone, especially for a large progenitor collection. The two most common robotic methods for arraying mutants into multi-well plates are flow cytometry-based single-cell sorting and robotic colony picking (Figure 3). Both methods are relatively fast, with specific pros and cons (Table 1). Most standard anaerobic chambers are not well suited to house either a flow cytometer or colony-picking robot. Fortunately, protection of strict anaerobes from oxygen exposure can be achieved by pre-reducing the culture medium, limiting time outside the chamber, and supplementing the medium with oxygen scavengers such as glutathione or cysteine.[107] In the future, the development of small-footprint microfluidics approaches for bacterial isolation[108,109] may facilitate arraying of extremely oxygen-sensitive strains inside an anaerobic chamber.

POOLING AND DECONVOLUTION

Pooling methods

Once mutants from the pooled library have been arrayed to form the progenitor collection, some form of combinatorial pooling is necessary to make sequencing of the library feasible. Methods of combinatorial pooling include *n*-bit binary string assignment,[19] DNA-Sudoku,[110] and Cartesian 2D, 3D, or 4D pooling[104] (Table 2). Pooling into more dimensions lowers the number of pools that must be prepared and sequenced but increases pool complexity. More complex pools are at greater risk of acquiring artifactual biases during library preparation (e.g., by PCR), which will amplify errors and lead to more ambiguous location calls. Furthermore, complex pools result in greater ambiguity when trying to find a mutant that occurs in more than one location. Conversely, creating more pools with lower complexity can increase accuracy but results in more laborious sample preparation. For barcoded libraries, Bar-Seq[35] can be used to identify mutant locations instead of the more laborious insertion sequencing.[18] This strategy reduces sequencing preparation

Table 2. Comparison of Cartesian pooling strategies

Cartesian axis	What to pool	Number of pools for 16 384-well plates
2D	every plate separately, each well from every plate	400 (384 well pools, 16 plate pools)
3D	every plate separately, rows, and columns	56 (16 row pools, 24 column pools, 16 plate pools)
4D	rows and columns of 4 × 4 plate grid, rows and columns of every plate	48 (16 row pools, 24 column pools, 4 plate rows, 4 plate columns)

time but requires the pooled library to have a high fraction of uniquely barcoded mutants for successful deconvolution.

Deconvolution

After the combinatorial pools are sequenced, deconvolution algorithms are used to predict the location of each mutant in the progenitor collection. There is currently no standard software for library deconvolution, although most authors have shared custom scripts for deconvolving their particular library.[104,107] The deconvolution process is based on the distribution of mutants within each pool. A mutant that is isolated only once and pooled in a 4D pooling scheme should be present in exactly four pools, resulting in an unambiguous predicted location. If a mutant isolated once appears in more than four pools, e.g., due to contamination during library preparation, PCR errors, or index hopping, its location will be ambiguous, although such cases can often be remedied by filtering pools to retain only high-quality reads above a threshold that depends on sequencing depth.

Any mutant isolated more than once will necessarily be present in more than four pools (a mutant isolated twice will typically be present in eight pools), and the number of possible locations consistent with these pools increases combinatorially with the number of isolations. This ambiguity in location calling decreases the power of the progenitor collection. When using binary pooling patterns with pre-defined allocations,[19] ambiguous locations can be partially resolved by calculating the Hamming distance between an observed mutant placement pattern and all pre-defined location patterns. The location with the smallest Hamming distance can then be selected as being the most likely location that contains that particular mutant[19,110] or, when using Cartesian pooling schemes, by identifying the most likely location through fitting onto a distribution constructed from the pool counts of unambiguous mutants.[104] Working with multiple batches of pools aids deconvolution because a mutant isolated in one batch will be deconvolved independent of its occurrence in other batches, and smaller batches will be more effective at identifying high-abundance mutants, while larger batches will identify low-abundance mutants.

The pooling method can be tested *in silico* by randomly selecting mutants according to their abundance in the pooled library (which is affected by library outgrowth just before arraying), pooling, and deconvolving a simulated sequencing dataset. The accuracy of deconvolution, and thus the number of mutants that can be confidently retrieved from the progenitor collection, is influenced by the abundance distribution of mutants in the pooled library, the level of cross-contamination during arraying and preparation for sequencing, pool complexity and evenness (ideally each pool contains a roughly similar number of mutants/genotypes), and sequencing depth. After the deconvolution algorithm predicts the location of each mutant in the progenitor collection, desired mutants can be rearrayed to create a condensed, non-redundant mutant library. As quality control, this condensed library should then be combinatorially pooled and sequenced (and possibly modified by partial rearraying) to create a final, high-confidence mutant collection ready for experimentation.

COSTS OF LIBRARY GENERATION AND DISTRIBUTION

With all these factors in mind, we estimate that the cost of generating a condensed library with ~6,000 mutants originating from a progenitor collection of ~50,000 mutants is ~$10,000 or less in consumables and sequencing. While such costs will not be prohibitive in many cases given the substantial expected impact of a high-quality arrayed mutant library, they are sufficiently high that libraries should be shared among the scientific community. To facilitate sharing and to maintain library quality long term, copies would ideally be available at well-known corporate or institutional culture collections (such as ATCC, DMSZ, or AddGene). In addition, the sharing of mutant pools alongside arrayed libraries will facilitate the use of RB-Tn-Seq. Such third parties can actively promote library distribution, which may also motivate academic laboratories to generate further libraries with the confidence that storage and distribution will be facilitated by this infrastructure. The cost to purchase a library would need to cover storage space, electricity for cold storage, quality control, and shipping; the *E. coli* Keio knockout collection serves as a template of affordability.[12] Detailed records of library construction and sharing of sequencing data and deconvolution code to underscore quality control will promote interest in and usage of libraries, although it will remain the end user's responsibility to confirm mutants of interest.

TARGET ORGANISMS FOR FUTURE LIBRARY CONSTRUCTION

What we do not know about the functional capacity of gut commensals is vast compared with our currently limited knowledge (Figure 4). Thus, the lineup of organisms potentially worthy of the effort of transposon mutant library construction is long (and frankly, somewhat overwhelming). Ultimately, the prospects for biological discovery should be the primary driver. Many host-microbe interactions with postulated health significance could potentially be addressed straightforwardly through screening of a mutant library, such as the mechanism through which *Fusobacterium nucleatum* is linked to colorectal cancer.[111] Certain species have already been developed to some degree into model organisms. For a variety of reasons, *B. thetaiotaomicron* is a model for polysaccharide metabolism,[112,113] which motivated the construction of a barcoded transposon library whose

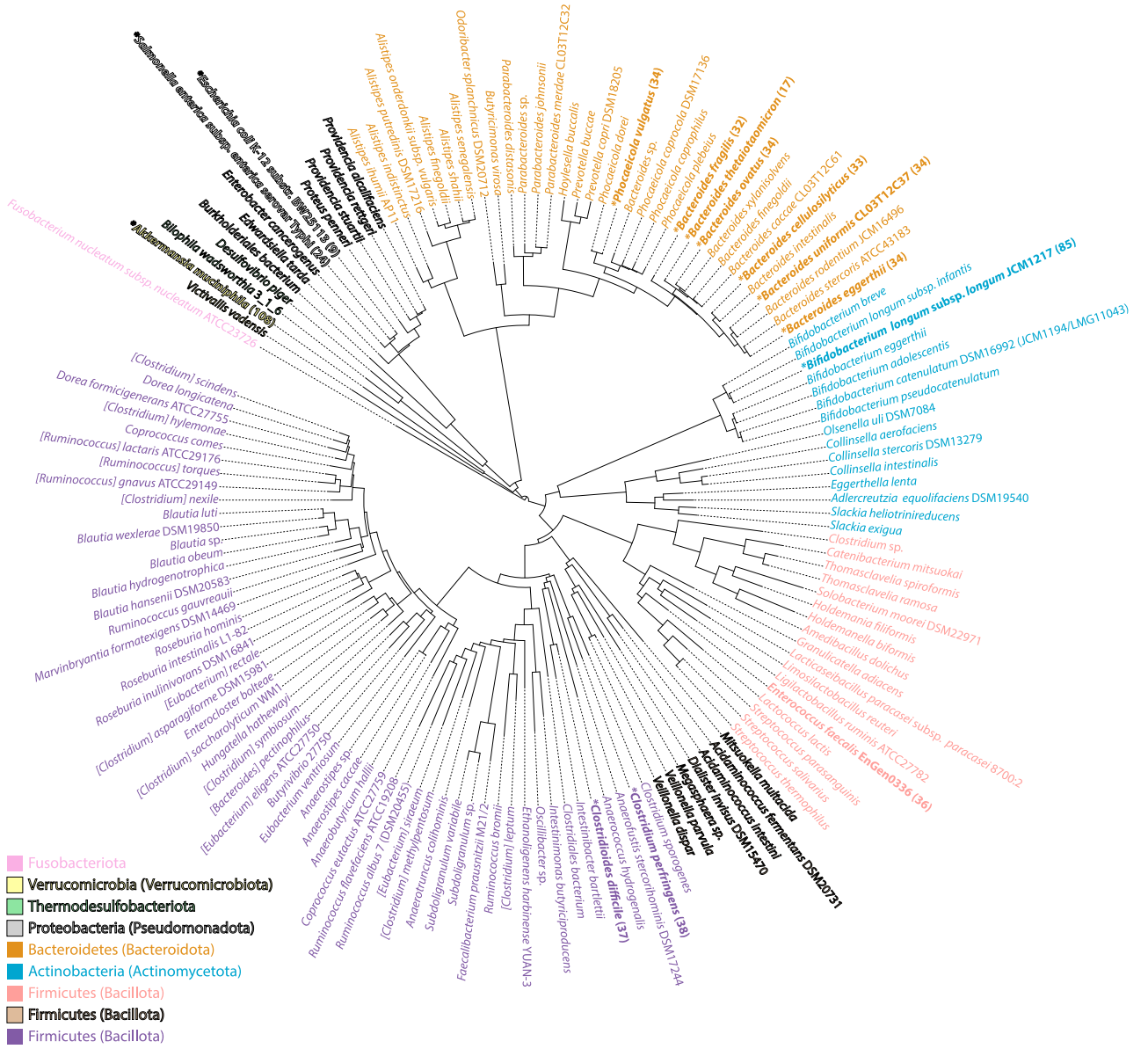

Figure 4. Candidates among the gut microbiome for library generation

Phylogenetic relationship of species from the most prevalent phyla in the human intestinal microbiota.[115] Among this substantial diversity, transposon mutant libraries have been generated in the asterisked species indicated in bold and labeled with a reference number in parentheses. Genome names from Cheng et al.,[115] Maier et al.,[116] and Zimmerman et al.[117] were aggregated, and the NCBI Taxonomy Name/ID Status Report tool was used to standardize their names and remove duplicates, resulting in a superset of 141 genomes. Representatives for these genomes were downloaded from NCBI using the datasets tool. GTDB-tk[118] classify-wf was used to find the maximum-likelihood placement of each genome in the full GTDB-Tk reference tree. The reference tree was then pruned to the 141 genomes using custom code. The final tree was visualized using iToL. All code used to generate the tree can be found at https://github.com/sunitj/paper_transposonLibrary. Branch lengths represent the number of expected substitutions per site.

functional genetic screening led to the discovery of new polysaccharide-related metabolic functions.[36] Metabolic comparisons across the Bacteroidetes phylum revealed a wide range of polysaccharide utilization capabilities.[114] It remains unknown the extent to which polysaccharide utilization locus (PUL) functionalities are conserved across species and strains, motivating the future construction of libraries in a wide range of *B. thetaiotaomicron* relatives in the *Bacteroides* genus.

Previous work characterizing the gene-phenotype landscape of 32 bacteria (primarily Proteobacteria species)[11] demonstrated that comparisons across related species can highlight conserved functions of certain genes, improving homology-based identification of gene function while also revealing new phenotypes. Related members of the gut microbiota are high-interest targets for forward-genetic screening. For instance, *Bilophila wadsworthia* from the Desulfovibrionaceae family plays

an important role in bile acid processing in the small intestine.[119,120] *B. wadsworthia* is likely highly distinct from Proteobacteria species; indeed, although previously classified as a Deltaproteobacteria member, *B. wadsworthia* was recently reclassified into the phylum Thermodesulfobacteriota.

Moving beyond commonly studied taxa, evolutionary distance may be a useful guide for discovering new genetic capabilities. Barcoded transposon mutant libraries currently exist for representative Proteobacteria, Bacteroidetes, and Actinobacteria, leaving Firmicutes and Verrucomicrobia as outstanding missing phyla. Among the Firmicutes, the Clostridia have diverse representation in the gut microbiota.[121,122] A TraDIS study defined gene essentiality for the pathogen *Clostridiodes* (formerly *Clostridium*) *difficile*,[42] demonstrating that barcoded library construction in this genus is feasible. *Clostridium* commensals are responsible for the production of many molecules in the gut, and genetic tools have been developed for *Clostridium sporogenes*.[123] *C. sporogenes* competes with *C. difficile* for vital nutrients[124] and interacts with other species such as *Lactococcus lactis*,[115] suggesting that *C. sporogenes* would serve as an attractive representative of the Clostridiaceae family, and perhaps the Firmicutes phylum at large. *A. muciniphila*, the model representative of the Verrucomicrobia phylum, has the ability to produce an extensive repertoire of mucin-degrading enzymes[125] and has been associated with both health and disease states[126]; non-barcoded transposon libraries have already been constructed in an *Akkermansia* species.[44] Taken together, we hope we are on the cusp of an explosion of functional information.

For some target organisms, genetic tools may be more readily available in a closely related species. A classic example is the development of a genome-wide knockout library of *Mycobacterium smegmatis* to model *Mycobacterium tuberculosis*.[127] However, for many families in the gut microbiome, there is no widely established model organism. This ambiguity could be beneficial: one could simply screen to identify the species in a given taxon with the highest transformation efficiency.[128] Given the extensive HGT among gut commensals, there is likely to be a wide range of transformability across strains, due for instance to the type of vector being used, variation in surface features that affects DNA entry, and/or the ability of the transposase to insert into the genome. In some cases, existing knowledge about strains of a target genus/species that are easier to transform should motivate their prioritization. For example, *Bifidobacterium breve* UCC2003 has been developed as a genetic model for the Bifidobacteriaceae family (Actinobacteria phylum)[100] due to its ease of transformation relative to other *Bifidobacterium* species. Moreover, it may be important to create libraries in multiple strains; for example, there are multiple clades of *A. muciniphila* with distinct growth behaviors, and some humans are inhabited concurrently by members of multiple clades.[129] Ultimately, a strong motivator in the choice of target species or strains will be the potential for novel metabolic or disease-related phenotypes.

CONCLUDING REMARKS

Over the past decade, dozens of transposon and knockout libraries have been constructed, covering members of diverse bacterial phyla.[12,13,18–22] These libraries have had enormous impact, revealing phenotypes, pathways, and functions important for bacterial physiology. However, there is still much work to be done, with many health-relevant gut commensals and entire phyla not represented. Efforts to develop genetic systems tend to focus on organisms that are easy to work with and fast growing, but it is also important to characterize fastidious and slow-growing species, which are able to coexist with and influence fast-growing species in the gut.[130] It is still unclear to what extent interstrain (genomic) variation affects the phenotypic landscape of a species and, hence, how many strains of a single species must be characterized before phenotypic knowledge approaches saturation.[131]

Several factors could accelerate the pace of transposon mutant library construction. Mutagenesis will be facilitated by synergy with synthetic biology to design better vectors.[16,54] Moreover, as sequencing becomes less expensive and more efficient, the time required for library analysis will decrease, enabling rapid feedback during optimization of construction strategies; reducing the time from vector design to finalized library to under a week would represent a dramatic, yet achievable, advance. It may be efficient to split library construction efforts among multiple research groups: for instance, groups could specialize in optimizing electroporation/conjugation efficiency in many organisms, designing highly effective vectors, and arraying libraries. Such a collaborative structure with centralized data management, shared methodologies, and standardized growth conditions could also be a model for other systems biology efforts.

While widespread interest and health relevance motivate focus on gut commensal function, there are many sources of potential species that have yet to be targeted. Most gut isolates have been obtained from stool, yet the small intestine is an environment highly distinct from the colon and microbes enriched in the small intestines have distinct functions compared with those enriched in stool.[120] Recently developed devices for collecting microbes and metabolites from the small intestine can provide both isolates and information about the environment.[120] Clearly, barcoded transposons can also be used to study the functional genetics of other microbiomes. Many of the tools and strategies developed for constructing pooled and arrayed libraries in gut commensals can likely be applied to soil and marine communities, as well as other human and animal communities, like the skin, oral, and vaginal microbiomes, and even other kingdoms such as the Archaea. However, there may be environment-specific considerations for library construction, for instance, related to the appropriate growth medium or atmospheric gas composition.

Although we have only scratched the surface of gene-phenotype mapping in gut commensals, the immediate future is bright. One of our goals for this review is to provide a knowledge base that enables other researchers to participate in this exciting endeavor. A reasonable expectation for the coming decade is the construction of a pooled library for at least one representative of each phylum present in the gut and, ideally, one representative of every major family. Arraying of these pooled libraries will provide the mutant strains necessary to test a wide range of mechanistic hypotheses, which should help to recruit new researchers

and approaches to study these organisms. We hope that it will be possible to translate protocols between closely related species, although even strains of the same species can exhibit differences in properties such as envelope structure[132,133] or CRISPR-Cas systems[134,135] that may substantially affect transformation efficiency. In any event, an expanded array of efforts will establish optimal strategies for library construction as well as parameters for which variability can be tolerated to increase throughput. In much the same way that AlphaFold has transformed many researchers into aspiring structural biologists,[136] we envision that barcoded transposon mutant libraries will facilitate the emergence of the inner geneticist in most microbiologists.

ACKNOWLEDGMENTS

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., J.L.S., C.B., and A.M.D.), ERC grant uCARE 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 Skłodowska-Curie Actions COFUND (grant 847543) and by an Add-On Fellowship for Interdisciplinary 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. A.T. and M.Z. are ERC investigators. K.C.H. and J.L.S. are Chan Zuckerberg Biohub Investigators. Figures were made in part using BioRender.com.

AUTHOR CONTRIBUTIONS

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

DECLARATION OF INTERESTS

The authors declare no competing interests.

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. Nat. Biotechnol. *39*, 105–114.
2. Dodd, D., Spitzer, M.H., Van Treuren, W., Merrill, B.D., Hryckowian, A.J., Higginbottom, S.K., Le, A., Cowan, T.M., Nolan, G.P., Fischbach, M.A., and Sonnenburg, J.L. (2017). A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. Nature *551*, 648–652.
3. Ivanov, I.I., Atarashi, K., Manel, N., Brodie, E.L., Shima, T., Karaoz, U., Wei, D., Goldfarb, K.C., Santee, C.A., Lynch, S.V., et al. (2009). Induction of intestinal Th17 cells by segmented filamentous bacteria. Cell *139*, 485–498.
4. Mazmanian, S.K., Round, J.L., and Kasper, D.L. (2008). A microbial symbiosis factor prevents intestinal inflammatory disease. Nature *453*, 620–625.
5. Venturelli, O.S., Carr, A.C., Fisher, G., Hsu, R.H., Lau, R., Bowen, B.P., Hromada, S., Northen, T., and Arkin, A.P. (2018). Deciphering microbial interactions in synthetic human gut microbiome communities. Mol. Syst. Biol. *14*, e8157.
6. Scher, J.U., Sczesnak, A., Longman, R.S., Segata, N., Ubeda, C., Bielski, C., Rostron, T., Cerundolo, V., Pamer, E.G., Abramson, S.B., et al. (2013). Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. Elife *2*, e01202.
7. van Opijnen, T., and Camilli, A. (2012). A fine scale phenotype–genotype virulence map of a bacterial pathogen. Genome Res. *22*, 2541–2551.
8. Bouhenni, R., Gehrke, A., and Saffarini, D. (2005). Identification of genes involved in cytochrome C biogenesis in *Shewanella oneidensis*, using a modified mariner transposon. Appl. Environ. Microbiol. *71*, 4935–4937.
9. Fey, P.D., Endres, J.L., Yajjala, V.K., Widhelm, T.J., Boissy, R.J., Bose, J.L., and Bayles, K.W. (2013). A genetic resource for rapid and comprehensive phenotype screening of nonessential *Staphylococcus aureus* genes. mBio. 4, e00537-e00512.
10. Jacobs, M.A., Alwood, A., Thaipisuttikul, I., Spencer, D., Haugen, E., Ernst, S., Will, O., Kaul, R., Raymond, C., Levy, R., et al. (2003). Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. Proc. Natl. Acad. Sci. USA *100*, 14339–14344.
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., Suh, Y., et al. (2018). Mutant phenotypes for thousands of bacterial genes of unknown function. Nature *557*, 503–509.
12. 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.
13. Koo, B.-M., Kritikos, G., Farelli, J.D., Todor, H., Tong, K., Kimsey, H., Wapinski, I., Galardini, M., Cabal, A., Peters, J.M., et al. (2017). Construction and analysis of two genome-scale deletion libraries for *Bacillus subtilis*. Cell Syst. *4*, 291–305.e7.
14. Porwollik, S., Santiviago, C.A., Cheng, P., Long, F., Desai, P., Fredlund, J., Srikumar, S., Silva, C.A., Chu, W., Chen, X., et al. (2014). Defined single-gene and multi-gene deletion mutant collections in *Salmonella enterica* sv Typhimurium. PLoS One *9*, e99820.
15. Hawkins, J.S., Silvis, M.R., Koo, B.M., Peters, J.M., Osadnik, H., Jost, M., Hearne, C.C., Weissman, J.S., Todor, H., and Gross, C.A. (2020). Mismatch-CRISPRi Reveals the Co-varying Expression-Fitness Relationships of Essential Genes in *Escherichia coli* and *Bacillus subtilis*. Cell Syst. *11*, 523–535.e9.
16. Peters, J.M., Koo, B.-M., Patino, R., Heussler, G.E., Hearne, C.C., Qu, J., Inclan, Y.F., Hawkins, J.S., Lu, C.H.S., Silvis, M.R., et al. (2019). Enabling genetic analysis of diverse bacteria with Mobile-CRISPRi. Nat. Microbiol. *4*, 244–250.
17. 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.
18. 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.
19. 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.
20. 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*. Proc. Natl. Acad. Sci. USA *105*, 8736–8741.
21. Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. Nature *418*, 387–391.
22. Liberati, N.T., Urbach, J.M., Miyata, S., Lee, D.G., Drenkard, E., Wu, G., Villanueva, J., Wei, T., and Ausubel, F.M. (2006). An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. Proc. Natl. Acad. Sci. USA *103*, 2833–2838.

23. Liu, X., Gallay, C., Kjos, M., Domenech, A., Slager, J., van Kessel, S.P., Knoops, K., Sorg, R.A., Zhang, J.R., and Veening, J.W. (2017). High-throughput CRISPRi phenotyping identifies new essential genes in *Streptococcus pneumoniae*. Mol. Syst. Biol. *13*, 931.

24. 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.

25. Brochado, A.R., and Typas, A. (2013). High-throughput approaches to understanding gene function and mapping network architecture in bacteria. Curr. Opin. Microbiol. *16*, 199–206.

26. Huang, Y.Y., Price, M.N., Hung, A., Gal-Oz, O., Ho, D., Carion, H., Deutschbauer, A.M., and Arkin, A.P. (2022). Functional screens of barcoded expression libraries uncover new gene functions in carbon utilization among gut Bacteroidales. Preprint at bioRxiv.

27. 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 Rep. e57369.

28. 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. Nat. Commun. *10*, 5729.

29. 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.

30. Gawronski, J.D., Wong, S.M.S., Giannoukos, G., Ward, D.V., and Akerley, B.J. (2009). Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. Proc. Natl. Acad. Sci. USA *106*, 16422–16427.

31. 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.

32. Van Opijnen, T., Bodi, K.L., and Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat. Methods *6*, 767–772.

33. Chao, M.C., Abel, S., Davis, B.M., and Waldor, M.K. (2016). The design and analysis of transposon insertion sequencing experiments. Nat. Rev. Microbiol. *14*, 119–128.

34. Smith, A.M., Heisler, L.E., Mellor, J., Kaper, F., Thompson, M.J., Chee, M., Roth, F.P., Giaever, G., and Nislow, C. (2009). Quantitative phenotyping via deep barcode sequencing. Genome Res. *19*, 1836–1842.

35. 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–e00315.

36. 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.

37. Veeranagouda, Y., Husain, F., Tenorio, E.L., and Wexler, H.M. (2014). Identification of genes required for the survival of *B. fragilis* using massive parallel sequencing of a saturated transposon mutant library. BMC Genom. *15*, 429–511.

38. 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.

39. Cullen, T.W., Schofield, W.B., Barry, N.A., Putnam, E.E., Rundell, E.A., Trent, M.S., Degnan, P.H., Booth, C.J., Yu, H., and Goodman, A.L. (2015). Antimicrobial peptide resistance mediates resilience of prominent gut commensals during inflammation. Science *347*, 170–175.

40. Ito, M., Kim, Y.-G., Tsuji, H., Takahashi, T., Kiwaki, M., Nomoto, K., Danbara, H., and Okada, N. (2014). Transposon mutagenesis of probiotic *Lactobacillus casei* identifies *asnH*, an asparagine synthetase gene involved in its immune-activating capacity. PLoS One *9*, e83876.

41. Dale, J.L., Beckman, K.B., Willett, J.L.E., Nilson, J.L., Palani, N.P., Baller, J.A., Hauge, A., Gohl, D.M., Erickson, R., Manias, D.A., et al. (2018). Comprehensive Functional Analysis of the *Enterococcus faecalis* Core Genome Using an Ordered, Sequence-Defined Collection of Insertional Mutations in Strain OG1RF. mSystems *3*, 000622-18.

42. Dembek, M., Barquist, L., Boinett, C.J., Cain, A.K., Mayho, M., Lawley, T.D., Fairweather, N.F., and Fagan, R.P. (2015). High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. mBio *6*, e02383.

43. Liu, H., Bouillaut, L., Sonenshein, A.L., and Melville, S.B. (2013). Use of a mariner-based transposon mutagenesis system to isolate *Clostridium perfringens* mutants deficient in gliding motility. J. Bacteriol. *195*, 629–636.

44. 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. *8*, 1450–1467.

45. Voogdt, C.G.P., Tripathi, S., Bassler, S.O., McKeithen-Mead, S.A., Guiberson, E.R., Koumoutsi, A., Bravo, A.M., Buie, C., Zimmerman, M., Sonnenberg, J., et al. (2023). Randomly barcoded transposon mutant libraries for gut commensals II: Applying libraries for functional genetics. Cell Rep. *42*, 113519.

46. Newman, D.J., and Shepp, L. (1960). The double dixie cup problem. Am. Math. Mon. *67*, 58–61.

47. Sberro, H., Fremin, B.J., Zlitni, S., Edfors, F., Greenfield, N., Snyder, M.P., Pavlopoulos, G.A., Kyrpides, N.C., and Bhatt, A.S. (2019). Large-Scale Analyses of Human Microbiomes Reveal Thousands of Small Novel Genes. Cell *178*, 1245–1259.e1214.

48. Reznikoff, W.S. (2008). Transposon Tn5. Annu. Rev. Genet. *42*, 269–286.

49. Brillet, B., Bigot, Y., and Augé-Gouillou, C. (2007). Assembly of the Tc1 and mariner transposition initiation complexes depends on the origins of their transposase DNA binding domains. Genetica *130*, 105–120.

50. Korem, T., Zeevi, D., Suez, J., Weinberger, A., Avnit-Sagi, T., Pompan-Lotan, M., Matot, E., Jona, G., Harmelin, A., Cohen, N., et al. (2015). Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. Science *349*, 1101–1106.

51. Cartman, S.T., and Minton, N.P. (2010). A mariner-Based Transposon System for *In Vivo* Random Mutagenesis of *Clostridium difficile*. Appl. Environ. Microbiol. *76*, 1103–1109.

52. Yang, Y., Stewart, P.E., Shi, X., and Li, C. (2008). Development of a transposon mutagenesis system in the oral spirochete *Treponema denticola*. Appl. Environ. Microbiol. *74*, 6461–6464.

53. Liu, H., Price, M.N., Waters, R.J., Ray, J., Carlson, H.K., Lamson, J.S., Chakraborty, R., Arkin, A.P., and Deutschbauer, A.M. (2018). Magic Pools: Parallel Assessment of Transposon Delivery Vectors in Bacteria. mSystems *3*, e00143-17.

54. Whitaker, W.R., Shepherd, E.S., and Sonnenburg, J.L. (2017). Tunable expression tools enable single-cell strain distinction in the gut microbiome. Cell *169*, 538–546.e12.

55. Mimee, M., Tucker, A.C., Voigt, C.A., and Lu, T.K. (2015). Programming a human commensal bacterium, *Bacteroides thetaiotaomicron*, to sense and respond to stimuli in the murine gut microbiota. Cell Syst. *1*, 62–71.

56. Meredith, T.C., Wang, H., Beaulieu, P., Gründling, A., and Roemer, T. (2012). Harnessing the power of transposon mutagenesis for antibacterial target identification and evaluation. Mob. Genet. Elements *2*, 171–178.

57. Hutchison, C.A., III, Merryman, C., Sun, L., Assad-Garcia, N., Richter, R.A., Smith, H.O., and Glass, J.I. (2019). Polar effects of transposon

insertion into a minimal bacterial genome. J. Bacteriol. *201*, e00185-19–e00119.

58. Kucho, K.-i., Kakoi, K., Yamaura, M., Iwashita, M., Abe, M., and Uchiumi, T. (2013). Codon-optimized antibiotic resistance gene improves efficiency of transient transformation in *Frankia*. J. Biosci. *38*, 713–717.
59. Berglund, F., Österlund, T., Boulund, F., Marathe, N.P., Larsson, D., and Kristiansson, E. (2019). Identification and reconstruction of novel antibiotic resistance genes from metagenomes. Microbiome *7*, 1–14.
60. Jiang, X., Hall, A.B., Xavier, R.J., and Alm, E.J. (2019). Comprehensive analysis of chromosomal mobile genetic elements in the gut microbiome reveals phylum-level niche-adaptive gene pools. PLoS One *14*, e0223680.
61. Yao, L., Seaton, S.C., Ndousse-Fetter, S., Adhikari, A.A., DiBenedetto, N., Mina, A.I., Banks, A.S., Bry, L., and Devlin, A.S. (2018). A selective gut bacterial bile salt hydrolase alters host metabolism. Elife *7*, e37182.
62. Wilson, E.H., Groom, J.D., Sarfatis, M.C., Ford, S.M., Lidstrom, M.E., and Beck, D.A.C. (2021). A computational framework for identifying promoter sequences in nonmodel organisms using RNA-seq data sets. ACS Synth. Biol. *10*, 1394–1405.
63. Pritchard, J.R., Chao, M.C., Abel, S., Davis, B.M., Baranowski, C., Zhang, Y.J., Rubin, E.J., and Waldor, M.K. (2014). ARTIST: high-resolution genome-wide assessment of fitness using transposon-insertion sequencing. PLoS Genet. *10*, e1004782.
64. Chaudhuri, R.R., Morgan, E., Peters, S.E., Pleasance, S.J., Hudson, D.L., Davies, H.M., Wang, J., van Diemen, P.M., Buckley, A.M., Bowen, A.J., et al. (2013). Comprehensive assignment of roles for *Salmonella typhimurium* genes in intestinal colonization of food-producing animals. PLoS Genet. *9*, e1003456.
65. Vohra, P., Chaudhuri, R.R., Mayho, M., Vrettou, C., Chintoan-Uta, C., Thomson, N.R., Hope, J.C., Hopkins, J., and Stevens, M.P. (2019). Retrospective application of transposon-directed insertion-site sequencing to investigate niche-specific virulence of *Salmonella* Typhimurium in cattle. BMC Genom. *20*. 20-12.
66. Groussin, M., Poyet, M., Sistiaga, A., Kearney, S.M., Moniz, K., Noel, M., Hooker, J., Gibbons, S.M., Segurel, L., Froment, A., et al. (2021). Elevated rates of horizontal gene transfer in the industrialized human microbiome. Cell *184*, 2053–2067.e18.
67. Jeong, H., Arif, B., Caetano-Anollés, G., Kim, K.M., and Nasir, A. (2019). Horizontal gene transfer in human-associated microorganisms inferred by phylogenetic reconstruction and reconciliation. Sci. Rep. *9*, 5953.
68. Zhang, G., Wang, W., Deng, A., Sun, Z., Zhang, Y., Liang, Y., Che, Y., and Wen, T. (2012). A mimicking-of-DNA-methylation-patterns pipeline for overcoming the restriction barrier of bacteria. PLoS Genet. *8*, e1002987.
69. Johnston, C.D., Cotton, S.L., Rittling, S.R., Starr, J.R., Borisy, G.G., Dewhirst, F.E., and Lemon, K.P. (2019). Systematic evasion of the restriction-modification barrier in bacteria. Proc. Natl. Acad. Sci. USA *116*, 11454–11459.
70. Rakowski, S.A., and Filutowicz, M. (2013). Plasmid R6K replication control. Plasmid *69*, 231–242.
71. Grahn, A.M., Haase, J., Bamford, D.H., and Lanka, E. (2000). Components of the RP4 conjugative transfer apparatus form an envelope structure bridging inner and outer membranes of donor cells: implications for related macromolecule transport systems. J. Bacteriol. *182*, 1564–1574.
72. Pansegrau, W., Lanka, E., Barth, P.T., Figurski, D.H., Guiney, D.G., Haas, D., Helinski, D.R., Schwab, H., Stanisich, V.A., and Thomas, C.M. (1994). Complete nucleotide sequence of Birmingham IncP alpha plasmids. Compilation and comparative analysis. J. Mol. Biol. *239*, 623–663.
73. Schäfer, A., Kalinowski, J., Simon, R., Seep-Feldhaus, A.H., and Pühler, A. (1990). High-frequency conjugal plasmid transfer from Gram-negative *Escherichia coli* to various Gram-positive coryneform bacteria. J. Bacteriol. *172*, 1663–1666.
74. Dodsworth, J.A., Li, L., Wei, S., Hedlund, B.P., Leigh, J.A., and de Figueiredo, P. (2010). Interdomain conjugal transfer of DNA from bacteria to archaea. Appl. Environ. Microbiol. *76*, 5644–5647.
75. Bates, S., Cashmore, A.M., and Wilkins, B.M. (1998). IncP plasmids are unusually effective in mediating conjugation of *Escherichia coli* and *Saccharomyces cerevisiae*: involvement of the tra2 mating system. J. Bacteriol. *180*, 6538–6543.
76. Lu, Y., Zeng, J., Wu, B., E, S., Wang, L., Cai, R., Zhang, N., Li, Y., Huang, X., Huang, B., and Chen, C. (2017). Quorum Sensing N-acyl Homoserine Lactones-SdiA Suppresses *Escherichia coli-Pseudomonas aeruginosa* Conjugation through Inhibiting traI Expression. Front. Cell. Infect. Microbiol. *7*, 7.
77. Virolle, C., Goldlust, K., Djermoun, S., Bigot, S., and Lesterlin, C. (2020). Plasmid Transfer by Conjugation in Gram-Negative Bacteria: From the Cellular to the Community Level. Genes *11*, 1239.
78. Christie, P.J., Korman, R.Z., Zahler, S.A., Adsit, J.C., and Dunny, G.M. (1987). Two conjugation systems associated with *Streptococcus faecalis* plasmid pCF10: identification of a conjugative transposon that transfers between *S. faecalis* and *Bacillus subtilis*. J. Bacteriol. *169*, 2529–2536.
79. Brophy, J.A.N., Triassi, A.J., Adams, B.L., Renberg, R.L., Stratis-Cullum, D.N., Grossman, A.D., and Voigt, C.A. (2018). Engineered integrative and conjugative elements for efficient and inducible DNA transfer to undomesticated bacteria. Nat. Microbiol. *3*, 1043–1053.
80. Allard, N., Garneau, D., Poulin-Laprade, D., Burrus, V., Brzezinski, R., and Roy, S. (2015). A diaminopimelic acid auxotrophic *Escherichia coli* donor provides improved counterselection following intergeneric conjugation with actinomycetes. Can. J. Microbiol. *61*, 565–574.
81. Simon, R., Priefer, U., and Pühler, A. (1983). A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in Gram-negative bacteria. Biotechnology *1*, 784–791.
82. Ferrières, L., Hémery, G., Nham, T., Guérout, A.M., Mazel, D., Beloin, C., and Ghigo, J.M. (2010). Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. J. Bacteriol. *192*, 6418–6427.
83. Dehio, C., and Meyer, M. (1997). Maintenance of broad-host-range incompatibility group P and group Q plasmids and transposition of Tn*5* in *Bartonella henselae* following conjugal plasmid transfer from *Escherichia coli*. J. Bacteriol. *179*, 538–540.
84. 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.
85. Johnson, C.M., and Grossman, A.D. (2014). Identification of host genes that affect acquisition of an integrative and conjugative element in *Bacillus subtilis*. Mol. Microbiol. *93*, 1284–1301.
86. Pérez-Mendoza, D., and de la Cruz, F. (2009). *Escherichia coli* genes affecting recipient ability in plasmid conjugation: are there any? BMC Genom. *10*, 71.
87. Trieu-Cuot, P., Carlier, C., Martin, P., and Courvalin, P. (1987). Plasmid transfer by conjugation from *Escherichia coli* to Gram-positive bacteria. FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Lett. *48*, 289–294.
88. 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.
89. Li, J., Gálvez, E.J.C., Amend, L., Almási, É., Iljazovic, A., Lesker, T.R., Bielecka, A.A., Schorr, E.M., and Strowig, T. (2021). A versatile genetic toolbox for *Prevotella copri* enables studying polysaccharide utilization systems. EMBO J. *40*, e108287.
90. Mullany, P., Wilks, M., Lamb, I., Clayton, C., Wren, B., and Tabaqchali, S. (1990). Genetic analysis of a tetracycline resistance element from

Clostridium difficile and its conjugal transfer to and from *Bacillus subtilis*. Microbiology *136*, 1343–1349.

91. Artiguenave, F., Vilaginès, R., and Danglot, C. (1997). High-efficiency transposon mutagenesis by electroporation of a *Pseudomonas fluorescens* strain. FEMS Microbiol. Lett. *153*, 363–369.
92. Huang, P.H., Chen, S., Shiver, A.L., Culver, R.N., Huang, K.C., and Buie, C.R. (2022). M-TUBE enables large-volume bacterial gene delivery using a high-throughput microfluidic electroporation platform. PLoS Biol. *20*, e3001727.
93. Luchansky, J.B., Muriana, P.M., and Klaenhammer, T.R. (1988). Application of electroporation for transfer of plasmid DNA to *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Listeria*, *Pediococcus*, *Bacillus*, *Staphylococcus*, *Enterococcus* and *Propionibacterium*. Mol. Microbiol. *2*, 637–646.
94. Zhang, G.-q., Bao, P., Zhang, Y., Deng, A.-h., Chen, N., and Wen, T.-y. (2011). Enhancing electro-transformation competency of recalcitrant *Bacillus amyloliquefaciens* by combining cell-wall weakening and cell-membrane fluidity disturbing. Anal. Biochem. *409*, 130–137.
95. van der Rest, M.E., Lange, C., and Molenaar, D. (1999). A heat shock following electroporation induces highly efficient transformation of *Corynebacterium glutamicum* with xenogeneic plasmid DNA. Appl. Microbiol. Biotechnol. *52*, 541–545.
96. Bean, E.L., Herman, C., Anderson, M.E., and Grossman, A.D. (2022). Biology and engineering of integrative and conjugative elements: Construction and analyses of hybrid ICEs reveal element functions that affect species-specific efficiencies. PLoS Genet. *18*, e1009998.
97. Swafford, A.J.M., Hussey, S.P., and Fritz-Laylin, L.K. (2020). High-efficiency electroporation of chytrid fungi. Sci. Rep. *10*, 15145–15149.
98. Garcia, P.A., Ge, Z., Moran, J.L., and Buie, C.R. (2016). Microfluidic screening of electric fields for electroporation. Sci. Rep. *6*, 21238–21311.
99. Moore, J.A., Nemat-Gorgani, M., Madison, A.C., Sandahl, M.A., Punnamaraju, S., Eckhardt, A.E., Pollack, M.G., Vigneault, F., Church, G.M., Fair, R.B., et al. (2017). Automated electrotransformation of *Escherichia coli* on a digital microfluidic platform using bioactivated magnetic beads. Biomicrofluidics *11*, 014110.
100. Ruiz, L., Motherway, M.O., Lanigan, N., and van Sinderen, D. (2013). Transposon mutagenesis in *Bifidobacterium breve*: construction and characterization of a Tn*5* transposon mutant library for *Bifidobacterium breve* UCC2003. PLoS One *8*, e64699.
101. Peters, J.M., Colavin, A., Shi, H., Czarny, T.L., Larson, M.H., Wong, S., Hawkins, J.S., Lu, C.H.S., Koo, B.-M., Marta, E., et al. (2016). A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. Cell *165*, 1493–1506.
102. Goodall, E.C.A., Robinson, A., Johnston, I.G., Jabbari, S., Turner, K.A., Cunningham, A.F., Lund, P.A., Cole, J.A., and Henderson, I.R. (2018). The Essential Genome of *Escherichia coli* K-12. mBio *9*, 020966-17.
103. Chacón, J.M., Möbius, W., and Harcombe, W.R. (2018). The spatial and metabolic basis of colony size variation. ISME J. *12*, 669–680.
104. Anzai, I.A., Shaket, L., Adesina, O., Baym, M., and Barstow, B. (2017). Rapid curation of gene disruption collections using Knockout Sudoku. Nat. Protoc. *12*, 2110–2137.
105. Lampe, D.J., Grant, T.E., and Robertson, H.M. (1998). Factors affecting transposition of the Himar1 mariner transposon *in vitro*. Genetics *149*, 179–187.
106. Christen, B., Abeliuk, E., Collier, J.M., Kalogeraki, V.S., Passarelli, B., Coller, J.A., Fero, M.J., McAdams, H.H., and Shapiro, L. (2011). The essential genome of a bacterium. Mol. Syst. Biol. *7*, 528.
107. 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.
108. Chijiiwa, R., Hosokawa, M., Kogawa, M., Nishikawa, Y., Ide, K., Sakanashi, C., Takahashi, K., and Takeyama, H. (2020). Single-cell genomics of uncultured bacteria reveals dietary fiber responders in the mouse gut microbiota. Microbiome *8*, 5.
109. Matuła, K., Rivello, F., and Huck, W.T.S. (2020). Single-Cell Analysis Using Droplet Microfluidics. Adv. Biosyst. *4*, e1900188.
110. Erlich, Y., Chang, K., Gordon, A., Ronen, R., Navon, O., Rooks, M., and Hannon, G.J. (2009). DNA Sudoku—harnessing high-throughput sequencing for multiplexed specimen analysis. Genome Res. *19*, 1243–1253.
111. LaCourse, K.D., Johnston, C.D., and Bullman, S. (2021). The relationship between gastrointestinal cancers and the microbiota. Lancet. Gastroenterol. Hepatol. *6*, 498–509.
112. Sonnenburg, E.D., Zheng, H., Joglekar, P., Higginbottom, S.K., Firbank, S.J., Bolam, D.N., and Sonnenburg, J.L. (2010). Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. Cell *141*, 1241–1252.
113. Martens, E.C., Chiang, H.C., and Gordon, J.I. (2008). Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell host & microbe *4*, 447–457.
114. Pudlo, N.A., Urs, K., Crawford, R., Pirani, A., Atherly, T., Jimenez, R., Terrapon, N., Henrissat, B., Peterson, D., Ziemer, C., et al. (2022). Phenotypic and Genomic Diversification in Complex Carbohydrate-Degrading Human Gut Bacteria. mSystems *7*, e0094721.
115. Cheng, A.G., Ho, P.-Y., Aranda-Díaz, A., Jain, S., Yu, F.B., Meng, X., Wang, M., Iakiviak, M., Nagashima, K., Zhao, A., et al. (2022). Design, construction, and *in vivo* augmentation of a complex gut microbiome. Cell *185*, 3617–3636.e19.
116. 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.
117. Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R., and Goodman, A.L. (2019). Mapping human microbiome drug metabolism by gut bacteria and their genes. Nature *570*, 462–467.
118. Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database *36*, 1925–1927.
119. Devkota, S., Wang, Y., Musch, M.W., Leone, V., Fehlner-Peach, H., Nadimpalli, A., Antonopoulos, D.A., Jabri, B., and Chang, E.B. (2012). Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in *Il10*$^{-/-}$ mice. Nature *487*, 104–108.
120. Shalon, D., Culver, R.N., Grembi, J.A., Folz, J., Treit, P.V., Shi, H., Rosenberger, F.A., Dethlefsen, L., Meng, X., Yaffe, E., et al. (2023). Profiling the human intestinal environment under physiological conditions. Nature *617*, 581–591.
121. 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.
122. Gevers, D., Knight, R., Petrosino, J.F., Huang, K., McGuire, A.L., Birren, B.W., Nelson, K.E., White, O., Methé, B.A., and Huttenhower, C. (2012). The Human Microbiome Project: a community resource for the healthy human microbiome. PLoS Biol. *10*, e1001377.
123. Guo, C.J., Allen, B.M., Hiam, K.J., Dodd, D., Van Treuren, W., Higginbottom, S., Nagashima, K., Fischer, C.R., Sonnenburg, J.L., Spitzer, M.H., and Fischbach, M.A. (2019). Depletion of microbiome-derived molecules in the host using *Clostridium* genetics. Science *366*, eaav1282.
124. Reed, A.D., and Theriot, C.M. (2021). Contribution of Inhibitory Metabolites and Competition for Nutrients to Colonization Resistance against *Clostridioides difficile* by Commensal *Clostridium*. Microorganisms *9*, 371.
125. Ottman, N., Davids, M., Suarez-Diez, M., Boeren, S., Schaap, P.J., Martins Dos Santos, V.A.P., Smidt, H., Belzer, C., and de Vos, W.M. (2017). Genome-scale model and omics analysis of metabolic capacities of

Akkermansia muciniphila reveal a preferential mucin-degrading lifestyle. Appl. Environ. Microbiol. *83*, e01014–e01017.

126. Ottman, N., Geerlings, S.Y., Aalvink, S., de Vos, W.M., and Belzer, C. (2017). Action and function of *Akkermansia muciniphila* in microbiome ecology, health and disease. Best Pract. Res. Clin. Gastroenterol. *31*, 637–642.
127. Majumdar, G., Mbau, R., Singh, V., Warner, D.F., Dragset, M.S., and Mukherjee, R. (2017). Genome-wide transposon mutagenesis in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*. In In Vitro Mutagenesis: Methods and Protocols, pp. 321–335.
128. Cuív, P.Ó., Smith, W.J., Pottenger, S., Burman, S., Shanahan, E.R., and Morrison, M. (2015). Isolation of genetically tractable most-wanted bacteria by metaparental mating. Sci. Rep. *5*, 1–11.
129. Becken, B., Davey, L., Middleton, D.R., Mueller, K.D., Sharma, A., Holmes, Z.C., Dallow, E., Remick, B., Barton, G.M., David, L.A., et al. (2021). Genotypic and phenotypic diversity among human isolates of *Akkermansia muciniphila*. mBio *12*. 4788–21.
130. Tramontano, M., Andrejev, S., Pruteanu, M., Klünemann, M., Kuhn, M., Galardini, M., Jouhten, P., Zelezniak, A., Zeller, G., Bork, P., et al. (2018). Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. Nat. Microbiol. *3*, 514–522.
131. Galardini, M., Koumoutsi, A., Herrera-Dominguez, L., Cordero Varela, J.A., Telzerow, A., Wagih, O., Wartel, M., Clermont, O., Denamur, E., Typas, A., and Beltrao, P. (2017). Phenotype inference in an *Escherichia coli* strain panel. Elife *6*, e31035.
132. Porter, N.T., Hryckowian, A.J., Merrill, B.D., Fuentes, J.J., Gardner, J.O., Glowacki, R.W.P., Singh, S., Crawford, R.D., Snitkin, E.S., Sonnenburg, J.L., and Martens, E.C. (2020). Phase-variable capsular polysaccharides and lipoproteins modify bacteriophage susceptibility in *Bacteroides thetaiotaomicron*. Nat. Microbiol. *5*, 1170–1181.
133. Simpson, B.W., Nieckarz, M., Pinedo, V., McLean, A.B., Cava, F., and Trent, M.S. (2021). *Acinetobacter baumannii* can survive with an outer membrane lacking lipooligosaccharide due to structural support from elongasome peptidoglycan synthesis. mBio *12*, e0309921.
134. Zhao, J., Fang, H., and Zhang, D. (2020). Expanding application of CRISPR-Cas9 system in microorganisms. Synth. Syst. Biotechnol. *5*, 269–276.
135. Wang, Y., Zhang, Z.-T., Seo, S.-O., Lynn, P., Lu, T., Jin, Y.-S., and Blaschek, H.P. (2016). Bacterial genome editing with CRISPR-Cas9: deletion, integration, single nucleotide modification, and desirable "clean" mutant selection in *Clostridium beijerinckii* as an example. ACS Synth. Biol. *5*, 721–732.
136. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589.