High-throughput bacterial functional genomics in the sequencing era

Andrew N. Gray1,4, Byoung-Mo Koo1,4, Anthony L. Shiver3,4, Jason M. Peters1, Hendrik Osadnik1, and Carol A. Gross1,2,*

1Department of Microbiology and Immunology, University of California, San Francisco, CA 94158, USA
2Department of Cell and Tissue Biology, University of California, San Francisco, CA 94158, USA
3Graduate Group in Biophysics, University of California, San Francisco, CA 94158, USA

Abstract

High-throughput functional genomic technologies are accelerating progress in understanding the diversity of bacterial life and in developing a systems-level understanding of model bacterial organisms. Here we highlight progress in deep-sequencing-based functional genomics, show how phenotyping based on whole genome sequencing is enabling phenotyping in organisms recalcitrant to genetic approaches, and recount the rapid proliferation of functional genomic approaches to non-growth phenotypes, and discuss how advances are enabling genome-scale resource libraries for many different bacteria.

Introduction

There is an increasingly urgent need for the broad application of high-throughput, genome-scale genetic approaches in bacteria. First, there are large gaps in our knowledge of gene function and pathway connections even in well-studied model bacterial organisms. Yet, functional annotation of these bacteria to understand gene function and pathway connections is foundational knowledge that is broadly used in understanding both prokaryotic and eukaryotic organisms. Second, the advent of rapid, inexpensive DNA sequencing has fuelled large-scale genomics and metagenomics projects, providing blueprints for a wide range of bacterial species and insights into complex communities such as the gut microbiome. This has catalyzed the study of many new organisms. Increasing emphasis on these new species, including both commensal and pathogenic microbiome organisms, environmental organisms, and a variety of organisms important for specialized studies and applications, requires us to rapidly characterize the gene functions and circuits of diverse new organisms, including those with limited or no genetic tools. Finally, unforeseen effects of cellular
context often undermine the promise of a knowledge-based bio-economy based on synthetic biology, because it remains so challenging to control or predict how engineered pathways will interact with the host cell [1]. This underlines the importance of probing in vivo relationships between pathways, an important outcome of unbiased genome-scale screens.

Efforts in many laboratories, carried out in diverse organisms, are rising to the challenge of addressing this need. Brochado and Typas [2]** provided an excellent review of this area in 2013, with an emphasis on reverse genetic approaches. Recent advances are dramatically increasing the power of deep-sequencing-based functional genomic experiments. This includes the use of whole-genome sequencing for forward genetic approaches in organisms recalcitrant to genetic tools. In addition, many non-growth and other phenotypes can now be investigated at the genomic level. Finally, recent innovations are facilitating the construction of arrayed libraries for diverse species that represent important community resources. In this review, we focus on advances in these areas. For reference, key terms and concepts underlying these approaches are discussed in Box 1.

### Box 1

**Glossary of key terms and concepts in high-throughput genetics**

**Metagenome**

Genetic material from multiple organisms obtained directly from an environmental sample. Metagenomics allows for genetic characterization of non-culturable bacteria.

**Library**

A large (usually genome-scale) collection of strains or plasmids. Here, used primarily to describe a collection of strains with mutant alleles covering a significant fraction of the genome.

**Arrayed library**

A library in which individual strains are grown and stored in pure culture. Arrayed libraries are typically stored in a format that allows manipulation with an automated liquid handling system. Because strains are grown individually in a defined array, phenotypes can be measured without needing to simultaneously track genotype.

**Pooled library**

A library in which strains are grown and stored together in a single mixed culture.

**Fitness**

Fitness is a general term for phenotypes that are related to strain growth. Colony size is the fitness parameter used in arrayed approaches while digital counting is the parameter most used in pooled screens.

**Multiplexing**

Coverage depth of NGS can be sacrificed to sequence multiple samples simultaneously. During library preparation, short indexing barcodes present in amplification primers are
added to individual samples, allowing them to be identified and parsed in the data processing step.

Next Generation Sequencing (NGS)
Refers to the current collection of sequencing technologies that provide large numbers of sequencing reads. A variety of technologies (reviewed in [80]) can be applied to the approaches described in this review.

Forward genetic screen
A library of unknown composition is screened to identify strains with a particular phenotype. The genotypes of these strains are then characterized to understand the genetic basis of the phenotype.

Reverse genetic screen
A library of known composition is assayed, and the phenotypes of each individual strain are measured.

Chemical phenotypic profiling
A library of mutants is assayed for fitness across a large set of chemical stresses. The resultant phenotypic signatures can be subjected to hierarchical clustering to group genes with similar phenotypes, suggesting shared or related functions.

Phenotypic signature
A set of quantitative phenotypes for an individual mutant strain.

Hierarchical clustering
A computational process for grouping objects based on similarity in a hierarchical manner, generally represented as a dendrogram or similar tree structure. Hierarchical clustering is used to mine profiling data for new insights by associating strains with similar phenotypic signatures.

Loss-of-function allele
A mutant allele of a gene that results in either a complete loss of gene function or reduced functionality. Common types include full gene knockouts and gene disruptions such as transposons (provided that the disruption results in loss of function). Gene knockdown (e.g., CRISPRi) can also be used to induce loss-of-function by blocking gene expression.

Gain-of-function allele
A mutant allele of a gene that results in increased expression, activity or gain of a novel function. The most common type of allele in large-scale libraries is a constitutive overexpression mutant that results in increased activity.

DNA barcode
A short DNA sequence that uniquely identifies a specific strain in a library. Barcodes can be generated randomly and then associated with a mutation via sequencing (e.g., RB-
TnSeq or barcoded deletion libraries), derived from flanking genomic sequence (e.g., Tn-seq) or rationally designed (e.g., CRISPRi).

**Deep sequencing**

A next-generation sequencing application where particular regions are sequenced many times (high coverage). Adequate depth of sequencing is critical for robust digital counting.

**Digital counting**

Uses the number of sequencing reads associated with a particular mutant strain as an estimate of cell count. Digital counting can be used to quantify fitness of strains in a pooled library under various growth/stress conditions.

**Type IIS restriction enzyme**

A restriction endonuclease that cleaves DNA downstream of its recognition sequence. For example, MmeI cleaves 20 bp downstream from last nucleotide in its recognition site.

**Synthetic lethal**

A combination of two or more mutations that results in cell death whereas each of these mutation does not.

---

## Deep-sequencing-based functional genomics

DNA sequence-based assays for determining the fitness of individual strains in a pooled library have been exceedingly valuable since their introduction two decades ago [3,4]**. Now, advances in methods for generating libraries, coupled with continued improvements in throughput and multiplexing for next generation sequencing (NGS), have increased the power and decreased the cost of these approaches. In addition to forward and reverse genetic screening to identify genes involved in particular microbial cell processes, functional genomic profiling across many conditions (chemical phenotypic profiling) using deep sequencing is now feasible [5,6]. Importantly, such studies enable hierarchical clustering of the resultant phenotypic signatures to reveal functional associations between genes as well as higher order characterization of interaction networks.

The evolution of transposon methodologies has played a pivotal role in enabling deep-sequencing-based functional genomics. Since their introduction as a genetic tool in the 1970s, transposons have been a driving force for analysis of mutant phenotypes. They were first used extensively in forward genetic selections, and later adapted for use in reverse genetic screens as well. Now, transposons are also likely to play a central role in the expanding functional genomics to new organisms, given the diversity of organisms amenable to random transposon insertion mutagenesis [3,7]. Further, developments in transposon engineering and delivery methods continue to expand both the repertoire of available applications and range of targetable organisms for transposon mutagenesis [4]**[8]. Transposon insertions generate loss-of-function (LOF) alleles via gene disruption. However, since transposons often carry outward-facing promoters to alleviate polar effects
on gene expression, they can also be used to generate gain-of-function (GOF) alleles via overexpression of downstream genes in an operon [9].

Historically, signature-tagged mutagenesis (STM) was the first transposon-based technology to systematically query gene function [10], and was especially important in identifying genes required for pathogen virulence [11]. Here, each transposon carried a DNA barcode flanked by common priming sites, serving as a unique identifier for each insertion mutant (Figure 1). Individual insertion mutants were arrayed, pooled in small groups (50–100), and used in various infection models.Virulence mutants were identified as those present in the input pool but not in the output pool, determined by hybridization to barcode arrays [12,13]. This first iteration of STM required use of an arrayed transposon library to generate corresponding barcode arrays. Later adaptations such as transposon site hybridization (TraSH) [14,15] and microarray tracking of transposon mutants (MATT) [16] enabled amplification of probes from flanking genomic DNA and hybridization to genomic oligonucleotide microarrays, obviating the need for arrayed libraries. However, both the difficulty of sample preparation and high levels of background hybridization in microarrays limited the utility of this approach.

The advent of next generation sequencing and its coupling to transposon mutagenesis revolutionized and revitalized transposon-based genetic approaches. Four contemporaneously published methods, reviewed extensively in [4] *, enabled NGS-based deep sequencing and digital counting of individual transposon mutants in a pooled population; in this review we refer to these methods collectively as transposon sequencing (Tn-seq). Tn-seq uses genomic DNA flanking the site of transposon insertion both to identify the site of mutation and as a barcode for quantification (Figure 1). Flanking DNA is obtained either using Type II restriction enzyme sites at the transposon ends, oriented such that cutting occurs about 20 bp downstream of the transposon [17,18], or DNA shearing followed by size selection [19,20]; preparation of the DNA for sequencing requires additional steps, including adapter oligonucleotide ligation and PCR amplification.

Deep sequencing provides a highly accurate method for quantifying all strains in a genome-scale pooled library, a vast improvement on the previous methodology. A typical NGS run (e.g., HiSeq 2500, single flow cell) produces up to 300 million sequence reads. Since accurate quantification requires e.g. somewhat over 100 reads/strain, fitness can be calculated at genome scale, even if some strains are underrepresented in the starting population. With this high depth of coverage, initial library sequencing can itself be highly informative, as genes that are missing (or severely underrepresented with statistical significance) in transposon libraries are candidates for essential genes. Indeed, the broad applicability of transposon mutagenesis has shed light on the essential gene sets in several bacterial species [21]. Hyper-saturated libraries can even yield information on protein domains, promoters, non-coding RNA, and intergenic regions [20,22,23].

The genome-scale evaluation of mutant fitness afforded by Tn-seq has illuminated biology in many organisms. Identifying S. aureus transposon insertion mutants sensitive to the wall teichoic acid (WTA) biosynthesis inhibitor tunicamycin revealed an interaction network connecting the WTA pathway with other cell envelope related genes [24]. Screening the S. aureus
pneumonia transposon library in multiple conditions both in vivo and in vitro resulted in a genotype-phenotype virulence map [25]. Similarly, assessing Vibrio cholerae transposon mutant fitness in both a model host and pond water, representing different stages of its life cycle, identified genes important for host infection and dissemination [26]. Tn-seq has also revealed unique requirements for host colonization in vivo, including biofilm development, particular metabolic requirements during nutrient restriction, and other novel factors [27–34]. Finally, Tn-seq has also been useful in a limited sense for genetic interaction analysis: Tn-seq mutagenesis of B. subtilis strain lacking 4 multidrug/oligosaccharidyl-lipid/ polysaccharide (MOP) exporter superfamily genes identified a novel lipid II flippase gene, amj, via a synthetic lethal interaction with the known lipid II flippase gene, murJ [35].

While enormously useful for screens involving single or small sets of conditions, the multistep process involved in Tn-seq sequencing library preparation is laborious enough to be prohibitive for larger-scale profiling across many conditions. A recent modification of Tn-seq, however, has greatly increased the scalability of the approach, and promises to overcome this limitation. Tn-seq mutagenesis of B. subtilis lacking 4 multidrug/oligosaccharidyl-lipid/ polysaccharide (MOP) exporter superfamily genes identified a novel lipid II flippase gene, amj, via a synthetic lethal interaction with the known lipid II flippase gene, murJ [35].

Tn-seq and RB-Tn-seq are powerful methodologies for exploring gene function, but have drawbacks. Transposon library composition can be biased against low fitness mutants after outgrowth [37], there is a constant selective pressure for compromised strains to accumulate second site suppressors, and essential genes cannot be studied. Additionally, an advantage of transposon mutagenesis–dense random coverage of the genome–itself introduces certain issues; there is no simple way to target a subset of genes, and full genome coverage results in multiple insertions in most genes, decreasing the throughput of deep sequencing. Some of these problems can be alleviated by instead pooling clones from an arrayed barcoded deletion library; this approach also allows all strains to be initially present in similar amounts.

A promising new technology, CRISPRi (Clustered Regularly Interspaced Short Palindromic Repeats interference), addresses many of these issues and promises to offer a new platform for functional genomics in bacteria. CRISPRi uses computationally designed single guide RNAs (sgRNAs) to direct a catalytically inactivate variant of the Cas9 endonuclease (dCas9) to target genes, repressing transcription [38], (reviewed in this issue [Peters et al., 2015]). The system is inherently well-suited to deep sequencing applications, because the
portion of the sgRNA that mediates gene targeting also serves as a unique DNA barcode. Components of the CRISPRi system can be expressed under the control of inducible promoters. Because repression is inducible and tunable, essential genes can be targeted [39], and the library can be maintained in the non-perturbed (uninduced) state, reducing pressure for accumulating suppressor mutations. Because sgRNA libraries are rationally designed, they can target subsets of genes for in-depth genetic exploration of particular pathways [40]*. Finally, CRISPRi in bacteria may require only a single sgRNA per gene for efficient knockdown leading to a null phenotype [Peters JM et al., unpublished], which would increase the throughput of sequencing based assays. Efficacy may first require optimization in other species [39], however, and has yet to be tested across diverse bacteria. A further potential weakness of CRISPRi is that the dCas9-sgRNA complex sterically blocks transcription elongation [38], causing a polar effect on downstream gene expression in an operon. Fortunately, this polar effect is quantitatively predictable for operons without internal promoters [Peters, JM, unpublished], and genes in the same operon are often functionally related, minimizing pleiotropy.

**Phenotyping approaches based on whole-genome sequencing**

Though Tn-seq is a powerful technology for functional genomics, some important organisms are refractory to transposon mutagenesis, and indeed to introduction of any foreign genetic elements. In these organisms, chemical or UV mutagenesis followed by selection and then whole gene sequencing (WGS) is proving an effective alternative for genetic interrogation, albeit at lower throughput than transposon analysis. Recent advances in sequencing technology and tools for data analysis allow the rapid identification of mutations with single nucleotide resolution. For example, chemical mutagenesis coupled with WGS was used to dissect genetic changes underlying the plaque phenotypes of *Chlamydia trachomatis* plated on a macrophage lawn [41]** and to identify novel magnetosome genes in *Desulfovibrio magneticus* RS-1 [42]*.

For centuries, genetic crosses were the gold standard for determining the association of phenotype with genotype. However, with the advent of WGS, Genome wide association studies (GWAS), which report on the correlation of single nucleotide polymorphisms (SNPs) to diseases states or other phenotypes, have been successfully used as a basis to narrow in on genes of interest for further functional studies [43]. In bacteria, these methodologies are now being applied to understand drug resistance mechanisms [44]* and compensatory mutations that improve the competitive fitness of drug-resistant strains [45,46]. For example, sequencing of over 100 drug-resistant *Mycobacteria tuberculosis* strains revealed genomic and intragenic SNPs that were highly associated with resistance [44]*. Of the top ~20 hits, 10 were already known to be associated with drug resistance, suggesting that these SNPs were high confidence leads for further study. Large-scale phenotyping of natural isolates is also being used to identify naturally-occurring determinants of fitness. For example, WGS in conjunction with phenotyping of *Pseudomonas aeruginosa* clinical isolates has been used to characterize evolution and diversification of the pathogen during chronic infection [47].
The rapid expansion of next generation sequencing has enabled culture-independent characterization of microbial communities such as the human microbiome, in particular via 16S rRNA gene deep sequencing and metagenome assembly. To improve our understanding of how microbial communities in different environments function and to identify useful proteins for biotechnology, gain-of-function approaches to discover the functions of genes from the metagenome, called functional metagenomics, have been developed [48].

Metagenomic libraries are expressed in host strains with high transformation efficiency (e.g., *E. coli*) and then screened for functions of interest such as antibiotic resistance, enzymatic degradation of dietary fiber, or aspects of bacteria-host interactions [49–52]. One of the major challenges in functional metagenomics is efficient expression and maintaining functionality of foreign genes in the host strain. Expression of libraries in other highly transformable bacteria, as well as shuttle vectors that allow screening of libraries in a range of transformable bacteria, are now being used to improve the recovery of gene functions [53,54]. A related approach termed “Temporal Functional Metagenomics sequencing (TFUMseq) identified genes improving fitness of a gut commensal bacterium, *Bacteroides thetaiotaomicron*, by transforming a *B. thetaiotaomicron* genome fragment library into *E. coli*, and determining the ability of the transformants to survive in germ-free mice by tracking the population over time using deep sequencing [55]∗. Similar approaches will be useful for mining metagenomic genes that contribute to fitness of bacteria both *in vitro* and *in vivo*. Finally, microfluidic platforms have been developed for cultivating and screening rare species isolated directly from the environment [56–58]∗ (see also [Morten Sommer review reference] in this issue).

**Phenotypes that are not measurable by growth differences**

Bacteria carry out many activities that cannot be assayed via growth fitness, and genome-wide screening methods have been developed for many of them. Visible phenotypes of developmental processes such as biofilms, motility and sporulation can be directly monitored using arrayed ordered deletion libraries [[59,60], Koo BM et al., *unpublished*], or from pooled transposon libraries [61–63]∗. Cell morphology phenotypes can also be screened with increasing throughput; for example, cell shape mutants in *Helicobacter pylori* were identified by sorting pooled mutants with differing cell morphology by flow cytometry followed by deep sequencing [64]∗. This approach can also be used in conjunction with fluorescent dyes or reporters for phenotyping based on pathway activation, cell damage or metabolic state [65].

Assays have also been developed to identify players in phenotypes that lack obvious visual or growth-related correlates. For example, genes important for acquisition of the *B. subtilis* integrative and conjugative element ICEBs1 were identified by mating a donor element with a kanamycin resistance marker into an existing transposon library marked with spectinomycin resistance, followed by deep sequencing of transconjugants [66]∗. *Campylobacter jejuni* genes required for motility and host invasion were identified by comparing strains present inside vs. outside of host cells after infection in cell culture [67]. Paradis-Bleau et al. identified mutants defective in the permeability barrier and envelope integrity using a colorimetric assay of the arrayed *E. coli* single gene deletion library [68]∗.

In this study, many of the 100 genes identified as important for envelope biogenesis were of...
unknown function, and had no corresponding fitness defect in the chemical genomics screen of the same library. The analytical tools developed to measure this phenotype has immediate application for ordered libraries and plate-based colorimetric assays, including Gram staining, pH indicators and biofilm indicators. Incorporation of colorimetric reporters for other phenotypes enables readout of other specific responses. High-throughput image analysis may allow for characterization of colony morphology and single cell imaging can generate multiple phenotypic indicators as shown in yeast [69]. High-throughput and high-content screening promises to increase the sensitivity and power of phenotypic analyses.

**Is it feasible to have community resource libraries for many bacteria?**

*E. coli* was the first bacterium with a dedicated single gene deletion library. Since its publication in 2006 [70], this library has had an enormous impact on research in the community. To date, Baba et al. has been cited >1400 times, and has spawned global chemical-genomic, gene-gene, and drug-drug interaction analyses providing leads for numerous mechanistic studies, most recently several related to the *E. coli* envelope [71–77].

There are many advantages of an arrayed and barcoded single gene deletion/antibiotic replacement library as a starting point for genome scale studies [2]**. The same library is available to all researchers, enabling cross comparison of results. Slow growing strains are retained in the library. Relevant strains are immediately available for directed follow-up experiments. Many single-cell phenotyping assays are currently feasible only with arrayed libraries. Additionally, since strains are physically isolated from one another, they cannot be complemented by other strains in the population, which is especially important when assaying secreted factors. Finally, when desirable, arrayed libraries can be used as pooled libraries, but input ratios of the initial library can be adjusted as necessary for equal representation of strains with different fitness. However, arrayed gene deletion libraries are expensive to construct, and many bacteria lack the genetic tools necessary to make construction feasible.

As an alternative, rapid methods are being developed to array transposon libraries. Pioneered in *B. thetaiotaomicron* [18], combinatorial pooled sequencing strategies can be used to uniquelly identify individual archived clones at fractional cost, and increasingly sophisticated pooling strategies [78][79] have made this cost-effective even with large initial libraries. As these technologies continue to be adapted, ordered libraries—either transposon or CRISPRi-based—will be feasible in a much wider set of bacteria.

**Summary and Prospects**

Sequencing based approaches are being carried out to perform functional genomic analyses of genetic circuits, phenotypes, and physiological responses. This advance relies on easier methods of preparation of pooled samples and increases in the power of deep sequencing—in addition to longer reads, newer machines can achieve greater depth of reads in a shorter time. Multiplexing samples—up to ~100/sequencing lane for RB-Tn-seq [36]**, decreases cost and enables sampling multiple conditions, moving beyond the analysis of a single phenotype to query functional associations among genes and gene networks. At the same time, CRISPRi technology is providing a new, streamlined method for creating a genome-
scale knockdown library based on small guide RNAs that target the dCas9 repression to each ORF. Because CRISPRi works in trans, it may be possible to deliver multiple guide RNAs to cells in a configuration where they can be detected by deep sequencing, which will enable us to develop the one method currently missing from our arsenal-- genetic interaction analysis of a pooled library at a genome or subgenome scale.

Bacterial life is incredibly diverse, having adapted to some of the harshest and most varied environments on earth. How do they thrive in so many niches? What biological functions underpin the solutions they have found, what adaptations to their profound lineage of ever-changing evolutionary pressures? If we can make sense of these blueprints in a systematic way, we stand to revolutionize our understanding of the diversity of bacterial life, at both a systems and mechanistic level. Genomic resources for high throughput phenotyping are constantly improving, and being applied to an increasingly large number of bacteria, as well as to an increasingly large number of phenotypes. For some bacteria, these efforts are at the beginning; for others, like *E. coli*, we are steadily moving towards unraveling the circuitry and functions that enable its success. It is exciting to be at the beginning of this vista.

Acknowledgments

We would like to thank members of the Gross lab and Jeffery Cox for critically reading the manuscript. C.A.G. is supported by National Institutes of Health grants (R01GM102790 and R01GM036278).

References and Recommended reading


37. van Opijnen T, Camilli A. Genome-wide fitness and genetic interactions determined by Tn-seq, a high-throughput massively parallel sequencing method for microorganisms. Curr Protoc Microbiol. 2010; Chapter 1(Unit1E 3)


41**. Kokes M, Dunn JD, Granek JA, Nguyen BD, Barker JR, Valdivia RH, Bastidas RJ. Integrating chemical mutagenesis and whole-genome sequencing as a platform for forward and reverse genetic analysis of Chlamydia. Cell Host Microbe. 2015; 17:716–725. The authors used an ordered mutant library of C. trachomatis using chemical mutagenesis and whole genome sequencing. C. trachomatis is refractory to genetic manipulation, and this study provides a blueprint for functional genomics in other refractory bacteria. [PubMed: 25307932]


55*. Yaung SJ, Deng L, Li N, Braff JL, Church GM, Bry L, Wang HH, Gerber GK. Improving microbial fitness in the mammalian gut by in vivo temporal functional metagenomics. Mol Syst Biol. 2015; 11:788. This study provided a new framework for using metagenomic DNA to construct large-scale heterologous expression libraries that are tracked in vivo over time by deep sequencing.


Highlights

- High-throughput phenotyping accelerates understanding of gene function and network.
- Tn-seq is enabling functional genomics in a diverse set of bacteria.
- Whole genome sequencing is accelerating forward genetic screens.
- New approaches are expanding the types of phenotypes assayed on a global scale.
Figure 1. Timeline and comparison of screening technologies used in functional genomics

All technologies result in a library of pooled mutant strains (middle), which are identified by different barcodes (e.g., red and blue for gene A and gene B). Libraries can then be screened under different test conditions. Differences in technologies lie in the construction of the strains (left) and steps during analysis after screening (right). In **Signature Tagged Mutagenesis** [10], transposon mutants with barcodes of known sequence are pooled. After screening, barcodes are amplified using invariant priming sequences (yellow), labeled and detected via filter binding. **Tn-seq** and related methodologies (including INSeq, HITS, TraDIS [4]) use flanking DNA-sequences as barcodes. Sequencing library preparation requires restriction cleavage or DNA shearing, followed by linker ligation and amplification using priming sequences in the transposon (yellow) and linker (color). Relative enrichment of mutants is then determined via deep sequencing and genome alignment (middle right).
Random barcode Tn-seq [36]** incorporates random unique barcodes within the transposon, which are linked with a corresponding site of transposon insertion via Tn-seq before screening. As in Signature Tagged Mutagenensis, these barcodes are flanked by invariant priming sequences for barcode amplification. The relative enrichment of mutants can then be determined by barcode amplification and sequencing. Instead of relying on random insertion of transposons into the genome, Random barcode deletion libraries consist of arrayed and pooled deletion strains, which allows more control over the composition of the initial mutant library. CRISPRi uses small guide RNAs (sgRNA) to repress a specific target gene when induced. The 20 bp variable portion of the sgRNA defines the target and also serves as a unique barcode.
Figure 2. Deep-sequencing dependent identification of protein functions in environmental samples

Metagenomic libraries are constructed with DNA from an environmental sample (left), and introduced into a transformable model organism, here E. coli, as host (middle). Pooled screening of the library followed by deep sequencing analysis then allows identification of ORFs that confer gain-of-function, increasing viability of the host strain under the conditions tested (right).