Critical Reviews in Biochemistry and Molecular Biology, 2011; 46(1): 41–66

informa
healthcare

REVIEW ARTICLE

Probing bacterial pathogenesis with genetics, genomics, and chemical biology: past, present, and future approaches

James E. Gomez[1], Anne Clatworthy[1,2], and Deborah T. Hung[1,2,3]

[1]*The Broad Institute of MIT and Harvard, Cambridge, MA, USA,* [2]*Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, USA, and* [3]*Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA, USA*

Abstract

Classical genetic approaches for studying bacterial pathogenesis have provided a solid foundation for our current understanding of microbial physiology and the interactions between pathogen and host. During the past decade however, advances in several arenas have expanded the ways in which the biology of pathogens can be studied. This review discussed the impact of these advances on bacterial genetics, including the application of genomics and chemical biology to the study of pathogenesis.

Keywords: *Virulence; STM; IVET; libraries; transposon; mutagenesis; small molecules*

Introduction

Living organisms are incredibly complex, dynamic systems. In many ways, bacteria are on the extreme end of the spectrum, due to their ability to rapidly evolve and adapt to numerous ecological niches, including the human body. While most of the bacteria living in association with humans are harmless or even beneficial commensals, a small subset of organisms are able to cause human disease. In order to obtain a better understanding of how such pathogenic bacteria infect and injure their hosts, advances that allow a more comprehensive, systematic approach provide a wonderful complement to well-established reductionist approaches.

The paired ability to perturb a system and measure its response has been the central principle underlying the study of pathogenesis. Classically, the perturbation involves an alteration of a bacterial gene on the DNA level, followed by a measurement of response that is manifest as a phenotype. Historically, such efforts have been powerfully applied both on small scales as well as on large scales in the execution of screens (where the readout phenotype is not death) and selections (where the readout phenotype is death) to identify genes that are linked to the phenotype of interest, such as virulence.

One of the major contributions of genomics and chemical biology to the field of bacterial pathogenesis is the significant expansion in the ability to both perturb and measure responses. Genomics has provided a blueprint for perturbing bacteria on the DNA level in much more comprehensive and systematic ways, making possible a gene-by-gene interrogation of the entire genome. At the same time, chemical biology has introduced the intentional use of small molecules to perturb bacterial behavior in a complementary and at times distinct way to link gene function with phenotype (Stanley and Hung, 2009). In addition, genomics has dramatically altered the ability to measure responses to perturbations by introducing the ability to interrogate changes in gene expression on a genome-wide level. Thus, it is not exaggeration to say that the field of genomics has exploded over the past 15 years, and with the tremendous progress made in this field, advances in our ability to dissect bacterial physiology and virulence are quickly following.

Address for Correspondence: Debroh T. Hung, Department of Molecular Biology, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA 02114, USA. E-mail: hung@molbio.mgh.harvard.edu

(Received 15 July 2010; revised 17 October 2010; accepted 05 November 2010)

ISSN 1040-9238 print/ISSN 1549-7798 online © 2011 Informa Healthcare USA, Inc.
DOI: 10.3109/10409238.2010.538663

An annotated genome sequence and thus the full-predicted repertoire of proteins that a bacterium uses to survive and cause disease are available for a representative of nearly every group of bacterial pathogens that is currently being studied. This incredible trove of information has in many ways altered our approach to studying bacterial physiology and pathogenicity. Not only do we now begin our investigations with a much better understanding of the biological potential of our organisms of interest, but we also have an ever-improving set of tools at our disposal to experimentally interrogate these organisms. These improvements, which include the ability to delete or manipulate every gene in the genome and the introduction of massively parallel sequencing as a potential readout for a multitude of assays, as well as a host of improvements to our high-throughput screening capabilities such as automated microscopy, comprehensive mutant libraries, and improved small molecule libraries, are changing the speed and scale of our experiments.

Advances in genetics, genomics, and chemical biology have also impacted host models used to study infection. The advent of RNAi and gene knockout animals, and the application of highly specific small molecule inhibitors of host functions have increased the sophistication of our ability to perturb the host, both within tissue culture systems and live animals. Such approaches are allowing us to better understand the host side of the pathogen-host relationship and to study bacterial behavior within a more physiologically relevant context that incorporates *in vivo* microenvironment and host response. In many respects, and for the purposes of this review, the host microenvironment and host response can simply be considered as another perturbation of bacterial physiology. The development of cheaper, higher-throughput host models is another important advance, enabling researchers to take full advantage of newer tools that allow comprehensive, systematic interrogation of bacterial genomes. Simpler *in vivo* systems such as *Caenorhabditis elegans*, *Drosophila melanogaster*, *Arabidopsis thaliana*, and the zebrafish (*Danio rerio*) have been developed for use in high throughput studies of bacterial pathogenicity, and these can effectively complement studies in more costly, lower-throughput mammals.

This review focuses on approaches that have been taken during the past two decades to elucidate the virulence mechanisms of bacterial pathogens, and discusses many important paradigms that have arisen from these studies through the selected examples. We restrict our discussions to experimental manipulations and measurements that focus on the pathogen rather than the host response. Excellent reviews of computational approaches for studying bacterial pathogenesis, comparative microbial genomics, and host responses have been published elsewhere (Raskin et al., 2006; Mohr et al., 2010). We also restrict ourselves to the discussion of predominantly forward-based genetic approaches, which are focused on the goal of identifying genes or functions involved in a given phenotype of interest, that is, virulence or survival, and not reverse genetic approaches for studying a particular gene of interest and finding-associated phenotypes.

No single experimental approach is sufficient to reveal the complete biology that underpins the pathogenic capabilities of a given bacterial pathogen. Some pathogens, such as *Pseudomonas aeruginosa* and varying *Streptococcus* pathogens, are capable of infecting multiple sites in the human body where they cause pathology that can vary tremendously in its presentation and its severity. Clearly, with such organisms, multiple animal models are needed to recapitulate as many of the aspects of these diverse pathogenic processes as possible. Moreover, as discussed later, some infection models may have features (low infectious dose, population bottlenecks) that negate the utility of pooled genetic approaches.

Similarly, no single experimental approach is able to equally elucidate the biology of all different bacterial pathogens. The pathogen itself can provide substantial barriers to some genetic approaches. At one extreme, *Chlamydia* and other obligate intracellular pathogens are completely refractory to standard genetic analysis because they cannot be propagated outside of their host cell, making mutant isolation and other genetic manipulations impossible. Some organisms such as *Helicobacter pylori* are naturally competent and readily take up and incorporate exogenously manipulated DNA, whereas others, like the mycobacteria, are difficult to transform. These many characteristics help steer the researcher toward an experimental approach or approaches that are best suited for the pathogen of interest.

We began with a discussion of how bacterial genetics is used to dissect microbial virulence by identifying gene functions that are integral to the pathogenic process. Two distinct approaches have been taken. **Correlative** approaches seek to identify genes that are specifically activated during infection (relative to an *in vitro* baseline), based on the assumption that if a gene is induced under a specific set of conditions, its product likely plays a significant role in bacterial behavior under those conditions. Here, genomics has played a major recent role in the ability to identify such genes by advancing our ability to measure gene expression responses in the bacteria upon their introduction into a host (which serves as a perturbation). In contrast, **functional** genetic approaches seek to define the genetic requirements for a particular process, in this case virulence, in a direct manner. Classically, such functions have been identified through the isolation and characterization of genetic mutants that are defective in the process of interest. More recently, genomics and chemical biology have changed the way in which such functions can be identified by providing new tools for perturbing bacterial physiology and pathogenesis.

Correlative approaches

During their eons of existence and expansion into nearly every environmental niche imaginable, bacteria have perfected the art of adaptation. Most pathogens need to move through a variety of microenvironments in order to successfully establish infection in one host and then be transmitted to a new host, often after an extended exposure to a nonhost environment such as soil or water. Thus, many bacterial species have a diverse repertoire of genes that enable them to survive and replicate under multiple different conditions, including highly artificial *in vitro* culture conditions used in the laboratory. While *in vitro*-based studies can be technically easier than *in vivo*-based studies and have contributed immensely to our understanding of bacterial physiology and virulence, the *in vitro* laboratory conditions typically used to propagate pathogenic bacteria rarely recapitulate the natural *in vivo* environment of the host. Therefore, when laboratory-grown organisms are introduced into a host (or other system designed to mimic the host environment), complex changes in patterns of transcription are needed to cope with this new environment. These changes lead to the expression of new subsets of proteins that are required for adaptation to the new environment, and in the case of infection, involve processes such as adhesion, invasion, replication, toxin production, immune evasion, and other functions needed for survival. By identifying the gene products that increase or decrease in abundance during infection, a great deal can be gleaned about the host–pathogen interaction and what bacterial functions are required in this relationship. Not only can candidate virulence factors be discovered, but also important features of the host environment can be inferred from the bacterial genes that are induced or repressed in this milieu. Importantly, correlative genetic approaches can be applied to the study of the most important host environment—the site of human infection—in ways that functional approaches cannot.

Promoter traps

Not long ago, the contents of a pathogen's genome were largely unknown, and so in the absence of genome sequences and microarrays, there was no straightforward way to catalog the large-scale changes in gene expression occurring within a pathogen as it entered into and adapted to the host environment. Nonetheless, several methods were developed to identify bacterial promoters activated *in vivo*. These methods are based on an approach known as promoter trapping, in which libraries are created that consist of individual bacterial clones carrying a reporter gene fused to a random fragment of genomic DNA. If the DNA fragment contains a promoter, the reporter gene is expressed under conditions when and where that promoter is active. β-Galactosidase or an antibiotic resistance marker often serves as the reporter, thus allowing for either screens or selections to be executed to identify differentially regulated promoters. Thus, simple reporter libraries have been incredibly useful tools for identifying such promoters, as long as the experimental conditions allow for the easy interrogation and recovery of clones of interest. Unfortunately for those who study bacterial pathogenesis, the *in vivo* environment rarely meets these criteria, and so more complex, ingenious promoter probe systems have been developed to address these issues.

In vivo expression technology

In vivo expression technology (IVET) is a promoter trapping method used to identify promoters that are induced *in vivo* specifically during infection (Angelichio and Camilli, 2002; Rediers et al., 2005; Hsiao and Zhu, 2009). This approach was first used in 1993 to identify *Salmonella enterica* serovar *typhimurium* promoters that were specifically induced during growth in mice (Mahan et al., 1993). This and other promoter trap-based strategies are illustrated in Figure 1. The promoter trap developed by Mahan consisted of a library of genomic DNA fragments cloned upstream of a transcriptional fusion of two reporters, *purA* and *lacZY*. *purA* encodes a protein required for purine biosynthesis, and *lacZY* encodes b-galactosidase. This plasmid library, built on an oriR6K plasmid, was introduced into a purine-auxotrophic *purA*-deleted strain of *Salmonella*, and single crossover mutants were isolated on purine-replete growth medium. Because the purine auxotroph is unable to survive *in vivo*, only clones that carried an active promoter fusion resulting in the expression of *purA* were capable of surviving and establishing infection in mice. Three days after infection with the pooled *Salmonella* clones, surviving bacteria (clones) were recovered from the mouse spleens and screened for β-galactosidase activity. Clones that reproducibly survived the *in vivo* environment but displayed a lac^- pur^- phenotype *in vitro* carried promoters that were specifically induced *in vivo* and that were thus presumed to control expression of candidate virulence genes.

IVET has been refined many times during the past two decades and used to identify virulence genes from a diverse collection of pathogens, including *Vibrio cholerae* (Merrell et al., 2002b), *Salmonella* spp. (Mahan et al., 1995; Heithoff et al., 1997), and *P. aeruginosa* (Wang et al., 1996). Because purine auxotrophs are not available for all bacterial pathogens, other selections can be used. In developing IVET, Mahan also used *thyA* to complement a *Salmonella* thymine auxotroph. Many other essential metabolic genes or established virulence factors required for survival in the host have also been incorporated into IVET systems for other bacterial species (Dubail et al.,

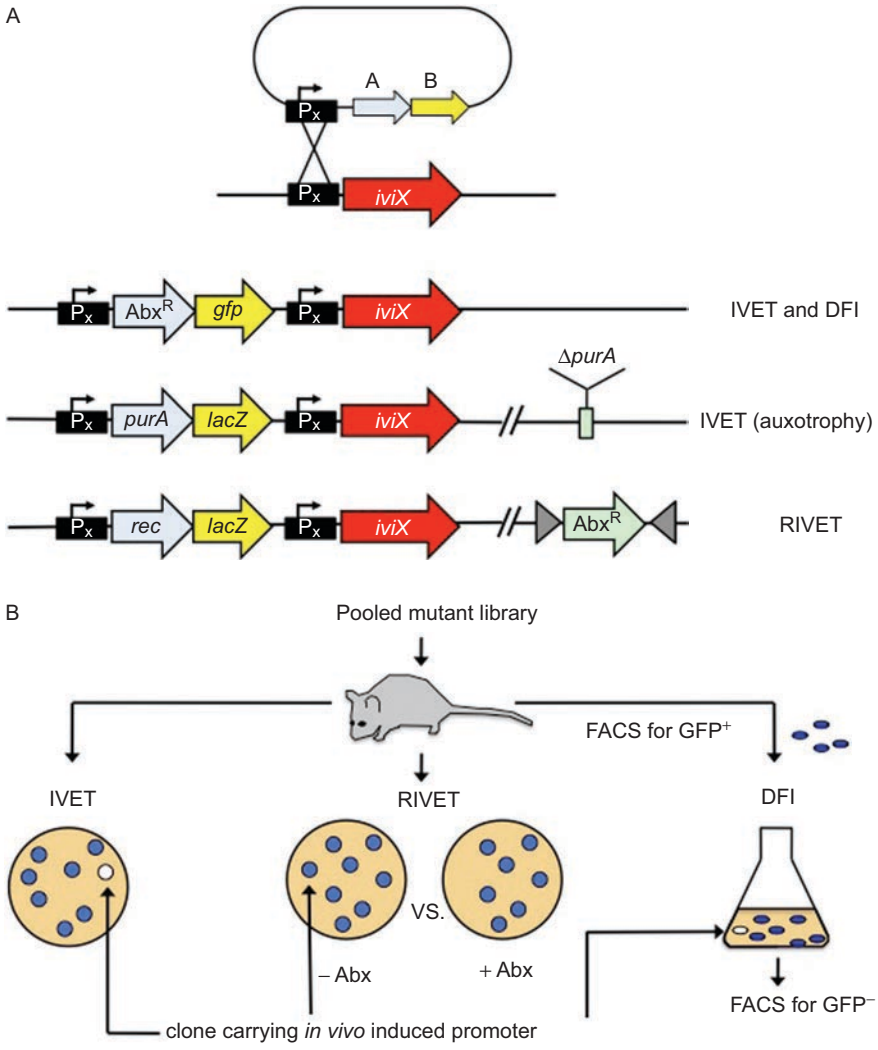

Figure 1. Strategies to identify *in vivo*-induced genes. (A) A schematic of the plasmids and strains used for promoter trapping assays. A candidate *in vivo*-induced gene is shown (*iviX*). A library is first created by cloning random genomic fragments (black box) into a promoter trap plasmid carrying selectable markers (genes A and B), which is then integrated into the chromosome of the screening strain by homologous recombination. This generates a library of reporter strains that can be used in selections and/or screens, and importantly maintain a functional copy the gene(s) associated with the promoter fragment in the plasmid. In IVET, gene A on the promoter probe construct encodes a selectable marker such as an antibiotic resistance gene or an essential gene such as *purA* in IVET, a recombinase in RIVET, and a FACS-optimized *gfp* in DFI. Gene B is an additional reporter such as *lacZ*. The relevant chromosomal screening backgrounds are illustrated on the right. (B) The workflow for promoter trap screening strategies, including IVET, RIVET, and DFI. These screens identify clones carrying *in vivo*-induced promoters driving the expression of reporter genes.

2000; Gahan and Hill, 2000). Antibiotic resistance markers can also be used as a selection marker (Mahan et al., 1995), as long as the antibiotic penetrates the relevant host tissues in a fairly uniform manner to provide selective pressure.

In order for IVET to be successful, the selection pressure *in vivo* needs to be in balance with desired levels of promoter activity. For example, clones carrying promoters that are weakly induced *in vivo* may not produce enough of the reporter antibiotic-inactivating enzyme to overcome the *in vivo* selective pressure and will be missed. Likewise, clones carrying strong promoters that are induced *in vivo* may be recovered after *in vivo* selection but discarded if they also have relatively high basal activity. Additionally, clones containing promoters that are only transiently expressed may not produce the selectable marker in enough abundance for a sufficient duration of time to avoid being selected against.

Recombinase-based IVET

In order to address these issues, recombinase-based IVET (RIVET) was developed (Lee et al., 1999). RIVET uses a library strain that carries a selectable marker (generally an antibiotic resistance cassette) flanked by recognition sites for a site-specific resolvase (*res* sites). Potential promoters are cloned upstream of a gene encoding the

resolvase enzyme, so that when the promoter is active, the resolvase is expressed and irreversibly excises the resistance cassette from the chromosome. RIVET differs from the previously described IVET approaches in that selection is not actually achieved during host infection; instead, after recovery from the animal, clones are screened in the presence and absence of antibiotic to reveal promoters that were activated *in vivo* (Figure 1B). Also, unlike standard IVET, during library construction, there is actually preselection against clones that have *in vitro* activity, since such clones would undergo rapid excision of their antibiotic resistance cassette. Because of these properties, RIVET can identify promoters that are expressed either weakly or transiently during infection. The sensitivity of a RIVET system can be fine-tuned by altering factors that contribute to the efficiency of marker excision, including the sequence of the *res* sites or the strength of the ribosome-binding site preceding the resolvase gene (Slauch and Camilli, 2000). Thus, by constructing multiple RIVET libraries with different levels of sensitivity, *in vivo*-induced promoters with a wide range of baseline and induced activities can be identified. Finally, methods have also been developed, which allow for *in vivo*-induced promoters to be identified by selection rather than screening; a counterselectable marker (such as *sacB*) can be incorporated into the cassette that is excised upon resolvase induction (Osorio et al., 2005), or the cassette can be engineered so that resolution leads to the loss of one selectable marker but regenerates another (Livny and Friedman, 2004).

RIVET in its many forms has been used to study gene expression in a wide variety of infections, including cell culture and mice (Lee et al., 1999; Osorio et al., 2005). Intriguingly, this approach has even been used in human volunteers who have been infected with a *V. cholerae* RIVET library (Lombardo et al., 2007). When these results were compared with those obtained in a screen carried out using the same library in an infant mouse model (Osorio et al., 2005), a strong correlation was observed, supporting the validity of the infant mouse model for the study of human cholera. Surprisingly, the overlap between these RIVET-identified genes and genes identified in other published studies of *V. cholerae* virulence was not large. These other studies include a microarray-based analyses of *V. cholerae* gene expression in human stool and vomitus (Merrell et al., 2002a; Larocque et al., 2005), an STM-based study of colonization and acid tolerance (Merrell et al., 2002b), and an IVET study in a rabbit model (Xu et al., 2003). The variability in results shows that, for the study of a given pathogen, there may be model-specific and approach-specific influences on the experimental results, thus underscoring the value of applying a collection of experimental approaches to the study of a single biological question.

One of the main advantages of RIVET is that promoter induction leads to a permanent phenotypic change (i.e. antibiotic sensitivity or sucrose insensitivity), and thus the timing of gene induction can be studied by isolating promoter trap clones at multiple time points during infection. Using this approach, Lee et al. examined a selected group of promoter–resolvase fusions of *V. cholerae* virulence genes in order to discern the timing of their induction *in vivo* (Lee et al., 1999). By examining the behavior of these fusions in mutants lacking some of the known *V. cholerae* virulence regulators, the authors were able to demonstrate a biphasic pattern of induction for the *tcpA* promoter (encoding the toxin-coregulated pilus), and found that the genetic requirements for virulence gene induction *in vivo* differed from what had been observed *in vitro* for the same genes; observations such as this serve to emphasize the importance of *in vivo* studies and the caution that must be applied in extrapolating *in vitro* phenomena to *in vivo* behavior. Similarly, to identify *V. cholerae* genes induced late in infection, Schild and Camilli administered kanamycin directly to the host 7 h after infection (Schild et al., 2007). The administration of drug killed off clones that had resolved within the first few hours after infection, as would be expected for clones carrying promoters for many of the known *V. cholerae* virulence factors. By limiting their kanamycin-sensitive output clones to those that resolved late, Schild and Camilli were able to show that *Vibrio* induces a previously unappreciated set of genes that promote the successful transition into the aquatic environment.

Differential fluorescence induction

Another promoter trapping strategy that has been applied to bacterial pathogenesis is differential fluorescence induction, or DFI (Figure 1). Like IVET, this approach fuses a random library of genomic fragments to a reporter gene, in this case a gene encoding a fluorescent protein such as GFP or one of the many GFP variants. As the respective promoters are induced, the corresponding clones begin to fluoresce, allowing clones carrying active promoters to be isolated by fluorescence-activated cell sorting (FACS). Valdivia and Falkow introduced a DFI library into *Salmonella typhimurium*, and used it to identify genes induced both in an *in vitro* model (acid stress) (Valdivia and Falkow, 1996) and during infection of macrophages (Valdivia and Falkow, 1997). To identify acid-induced genes, the library was exposed to growth medium at pH 4.5 and then sorted to isolate the most highly fluorescent bacteria. These selected clones were regrown at neutral pH and again sorted, this time based on the absence of fluorescence. After a second exposure to acid stress, the brightest cells were isolated and their insertions characterized. Fluorescence induction in these clones was then tested in macrophages, where interestingly, only a subset was induced. These results suggest

that the signals that regulate the expression of these genes *in vivo* may be more complex than expected, and not simply the result of the acidic milieu of the phagolysosome. To follow up on this observation, Valdivia and Falkow carried out a similar assay directly in cultured macrophages and went on to show that *S. typhimurium* genes induced within the macrophage are also induced in the spleens of infected mice.

DFI has been similarly applied to the identification of *Mycobacterium marinum* genes induced in both macrophages and frogs (Ramakrishnan et al., 2000). *M. marinum* is a close relative of *Mycobacterium tuberculosis* and causes granulomatous infections in both frogs and fish, including zebrafish. and because of this, it has become an intriguing model for human tuberculosis. Ramakrishnan et al. identified promoters activated in both macrophages and granulomas, as well as promoters activated in only one of these models. These specifically activated promoters have since been used as tools to evaluate granuloma development and bacterial behavior during infection of frogs and zebrafish (Chan et al., 2002), including zebrafish embryos (Davis et al., 2002), whose transparency allows for the direct visualization of fluorescent *M. marinum* in a living host.

Microarrays

Library-based, promoter trap methods for identifying *in vivo*-induced genes can be extremely labor-intensive and challenging to obtain saturation. One of the great technical advances that accompanied the new availability of complete genome sequences was the development and application of microarrays to study gene expression in all cell types, including bacteria (Figure 2) (Waddell et al., 2007). Microarrays are high-density arrays of DNA spotted onto a slide. These reporter probes correspond to all or many genes of an organism and can also encode regulatory elements such as small RNAs. Fluorescently labeled cDNA prepared from that organism grown in an experimental condition can be hybridized to the array, often in competition with a second "reference" sample labeled with a different fluorophore. The relative fluorescence at each spot is an indication of the abundance of transcripts for that gene relative to the reference.

While initially some believed that the speed and scale of this technology would quickly make IVET-like genetic approaches obsolete, the use of microarrays to directly evaluate bacterial gene expression during host infection has proven to be somewhat challenging. Transcripts from bacteria grown *in vitro* are ideally suited for microarray analysis because of the ease and rapidity with which abundant amounts can be isolated. Thus, many studies of bacteria cultivated under various *in vitro* conditions that attempt to mimic some aspect of the host microenvironment have been performed (Hinton et al., 2004).

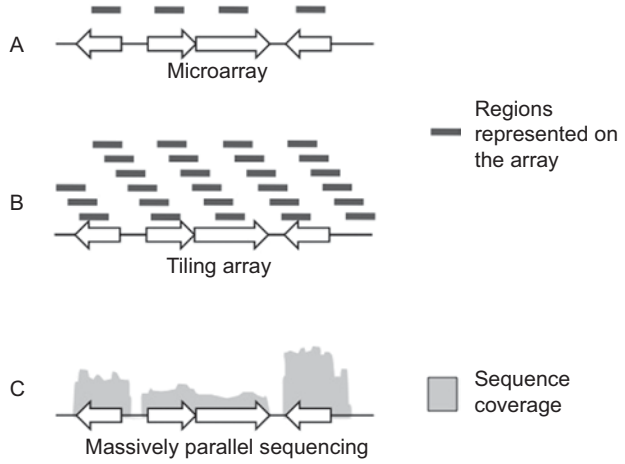

Figure 2. A comparison of microarrays, tiling arrays, and deep sequencing for the measurement of RNA levels. (A) Microarrays consist of arrayed, spotted oligonucleotides probes that typically provide resolution at the level of the individual gene. (B) Tiling arrays can have 10–30 bp resolution, due to overlapping oligonucleotide probes for each gene. (C) Deep sequencing using massively parallel sequencing technologies can provide resolution of transcript abundance down to the nucleotide level.

For bacteria that replicate intracellularly, tissue culture infection models provide the opportunity to recover large amounts of relatively pure RNA from bacteria growing in an environment that more closely reflects the genuine *in vivo* environment. Microarrays have thus been used to study gene expression in intracellular pathogens such as *M. tuberculosis* (Schnappinger et al., 2003; Voskuil et al., 2003), *Salmonella* (Eriksson et al., 2003), and *Shigella* (Lucchini et al., 2005), as well as in other organisms like *Escherichia coli* that do not primarily inhabit the host intracellular environment (Staudinger et al., 2002). Isolating high-quality mRNA from bacteria infecting a whole organism host model is much less straightforward. Bacterial transcriptional profiles can change extremely rapidly due to both mRNA degradation and new transcription during sample processing. Also, not all infection models allow for the easy recovery of bacteria from host tissues in quantities large enough to be amenable to microarray analysis. In these cases, amplification is required, which in itself, can cause bias. Nonetheless, by implementing some technical alterations to the sample preparation process, microarrays have been used to study bacterial gene expression in the host.

One intracellular bacterial pathogen that has been probed extensively using microarrays in order to define genes induced in a plethora of conditions including starvation, hypoxia, and other stressors is *M. tuberculosis* (Wilson et al., 1999; Boshoff et al., 2004). In addition, microarrays have been used to look at bacterial gene expression under potentially more relevant conditions, including infection of macrophages (Schnappinger et al., 2003) and mice (Talaat et al., 2004), and even in lung

tissue (Rachman et al., 2006) and sputum (Garton et al., 2008) isolated from actual human cases of tuberculosis. We briefly summarized some of results to demonstrate the utility of this approach as well as some of its limitations.

Although resident tissue macrophages are generally the first line of host defense against bacterial invaders, *M. tuberculosis* has evolved to persist and sometimes thrive within these cells by subverting, or at least tolerating, many key macrophage functions. Schnappinger et al. carried out an elegant microarray analysis of bacterial transcription in a series of environments, including bone-marrow-derived macrophages isolated from both immune competent and immunocompromised *NOS2*$^{-/-}$ mice (Schnappinger et al., 2003). From these studies, they were able to support the hypothesis that intracellular *M. tuberculosis* metabolizes fatty acids and to demonstrate that a large regulon of hypoxia-associated genes is induced by host-produced nitric oxide.

Investigators have been somewhat reluctant to study bacterial gene expression in animal models using microarrays, largely because of the perceived challenge of measuring bacterial transcripts amidst a sea of contaminating eukaryotic RNA derived from host tissues. Instead, quantitative real-time reverse transcription–PCR (qRT-PCR) has often been used to evaluate *in vivo* expression changes in small sets of genes that have been implicated in pathogenesis through other experimental approaches (Timm et al., 2003). qRT-PCR is both sensitive and quantitative, but is extremely difficult and costly to apply on a genome-wide scale.

Nevertheless, several studies have attempted to characterize the expression profiles of *M. tuberculosis* from host organisms using comprehensive microarray analysis. To overcome the challenge of obtaining sufficient bacterial cDNA for such analysis while minimizing interference from host-derived cDNA, Talaat et al. used gene-specific primers to direct cDNA synthesis against the mRNA of *M. tuberculosis* bacteria recovered from the lungs of mice (Talaat et al., 2004). Two other groups have also since characterized *M. tuberculosis* gene expression in human clinical samples. Rachman et al. (2006) utilized Talaat's gene-specific primer set to measure *M. tuberculosis* gene expression in resected lung tissue, and Garton et al. (2008) used an RNA amplification step to analyze transcription from bacteria in human sputum. While the behavior of some genes is similar in all three of these *in vivo* studies, a large number of genes behave quite differently in the various models. How to best interpret these differences remains unclear. One possibility is that the respective environments from where the bacilli were isolated in each case are actually more disparate than expected. Another possibility is that the transcriptional behavior of a population of bacteria in a complex environment such as the host may in fact be heterogeneous. Thus, there may be subpopulations of bacteria in all three models that share similar transcriptional profiles, whereas other subpopulations may coexist whose profiles reflect peculiarities of these three different environments. Since microarrays can only effectively measure a population-wide average, different coexisting behaviors may be obscured in a population-wide blur. Garton et al. tried to account for this possibility in their study by trying to fit their results to a model in which the microarray signal was the product of a mix of bacteria in two distinct, well-characterized physiological states, but were unable to do so. Thus, in situations where heterogeneity is indeed an issue, the use of RIVET-like approaches, which monitors the behavior of a single promoter within a single cell, may be more informative than an array-based approach.

Massively parallel sequencing, tiling arrays, and other recently developed tools

Two tools that are extending our ability to analyze the complete RNA content of cells are massively parallel sequencing (i.e. RNA-seq) and genomic tiling arrays (Figure 2). In addition to providing quantitative information on transcript abundance, these approaches reveal details of an organism's transcript composition that shed light on promoter structure, operon structure, and other important regulatory features. Genomic tiling arrays are extremely high-resolution microarrays composed of small, overlapping oligonucleotides that encompass the entire genome and allow for the fine definition of transcript architecture and abundance (Mockler et al., 2005). The need for fast, inexpensive ways to sequence large genomes has driven and continues to drive the development of new sequencing tools such as the Roche 454, Illumina GA, Helicos, and ABI SOLiD methodologies, to name a few (Ansorge, 2009). Each of these technologies can generate tremendous amounts of sequence data. Roche 454 produces the longest read lengths and is fast, but generates the smallest amount of overall data, making the cost per gigabase sequenced relatively high. Illumina is less expensive, generating many more reads, but these are shorter in length (35–120 bases); Helicos is a similar system, slower but capable of generating huge amounts of data; and ABI SOLiD also produces a large number of short (35–50 base), highly accurate reads. The computational effort required to align and analyze sequencing data generated by these new technologies is inversely related to read length. Nevertheless, depending on the objective, one or more of these technologies may be used to generate the requisite data used to reconstruct the transcriptome.

Recently, combinations of next generation sequencing and tiling arrays were used to define the comprehensive transcriptomes of three bacterial pathogens,

Salmonella typhi (Perkins et al., 2009), *Mycoplasma pneumoniae* (Güell et al., 2009), and *H. pylori* (Sharma et al., 2010), with each study revealing an unexpected level of transcriptome complexity, including an abundance of riboswitches, small regulatory RNAs, and *cis*-antisense transcripts. These studies on *in vitro* grown bacteria suggest that the regulation of virulence genes also involves similar complex regulatory mechanisms. Antisense transcripts had previously been known to be involved in bacterial gene regulation, but the large scale of their involvement has only recently been made clear (Waters and Storz, 2009). Conventional promoter trap approaches aimed at discovering *in vivo*-induced promoters have frequently identified clones that carry genomic fragments that would not be predicted to contain a promoter or that are oriented "the wrong way" in the promoter probe vector; these earlier results may, in fact, be harbingers of more recent results obtained from RNA sequencing. Deep sequencing and tiling arrays have not yet been directly applied to the identification or characterization of *in vivo* expressed transcripts, but the data coming out of simple *in vitro* experiments suggest that *in vivo* virulence regulation is likely to also involve a complex network of regulatory elements.

Transcriptomics also informs genomics by facilitating more accurate annotation of protein-coding regions, a better understanding of how transcription units are organized in operons, and a much broader cataloguing of genome features such as noncoding RNAs. More transcriptomic data obtained by sequencing leads to better genome annotation, which forms the basis for improved micro- and tiling-array design. Whether RNA sequencing will ultimately completely replace arrayed approaches to expression analysis is simply a question of scale, cost, and feasibility.

In their recent review of prokaryotic transcriptomics, Sorek and Cossart (2010) pointed out that a drawback of transcriptomic methods is that they currently require millions of cells as starting material. While some infection models can easily generate the bacterial populations needed for such analyses, not all can. Even in cases where enough bacteria are present in the tissues, methods need to be optimized to obtain mRNA from these bacteria quickly and cleanly so that the architecture of the transcripts is preserved, a concern that applies to microarray-based analyses as well. Also, this approach shares the same limitations as microarrays with regard to the ability to detect transcriptional heterogeneity within the bacterial population when large numbers of bacteria are included in the generation of a single cDNA sample. A method to determine the transcriptome of a single cell is a much-anticipated advance that would illuminate how pathogens coordinate their gene expression and the role that heterogeneous responses might play in pathogen survival inside the host.

An advantage to microarray and sequencing-based approaches to studying bacterial gene expression is the general applicability of the methodology. Aside from isolating RNA from bacterial cells, no other molecular manipulation of the bacterium is needed to determine which genes are up-regulated during infection, enabling its use in genetically intractable organisms. However, all such gene expression-based approaches to defining genes functions that are important for pathogenicity are based on temporal and spatial correlation between the induction of certain genes and the host environment. These correlative approaches are based on the underlying assumption that genes that are up-regulated under *in vivo* conditions must play a role in virulence and infection. This assumption can be challenged, leading to the need for a more direct approach to identifying genes that have functional significance during infection. To accomplish this, some perturbation of the bacterium—random or targeted, genetic or chemical—must occur to link a particular gene function with its requirement in infection.

Functional approaches

An early challenge to bacterial geneticists interested in acquiring functional information was the issue of scale and ability to retrospectively identify genetic lesions that account for a given phenotype. Any attempt to screen for bacterial mutants that show virulence-associated phenotypes clearly requires the generation of diversity (i.e. large numbers of random mutants), an ability to screen or select for the desired phenotype, followed by identification of the relevant genetic lesion in the mutants of interest.

Originally, diversity could be generated through exposure to mutagens such as UV irradiation or chemicals such as nitrosoguanidine. Such random mutations within the genome could be challenging to map and required extensive complementation efforts. The utilization of transposons to mutagenize bacteria dramatically altered this landscape by providing a mechanism to generate genomic diversity that could be more easily mapped by sequencing of insertion sites. Recently, efforts have been made to create comprehensive arrayed libraries of deleted genes (Baba et al., 2006) that provide a well-defined collection of mutants. Notably, such gene deletion libraries do not define all of genetic diversity since other lesions such as point mutations may provide very different phenotypes than gene deletion.

The type of mutagenesis used to construct the library, the phenotypes for which the library can be screened, and the way in which each mutant within the library is (or is not) tagged for subsequent identification set the limitations on the studies that can be performed.

However, recent advances in technologies, such as identification of mutants through whole genome sequencing, circumvent previously existing limitations. Numerous new methods have been developed to better interrogate the host–pathogen interaction and are described below.

Genetic-based identification of virulence factors

Genetic tools have been applied to the study of bacterial pathogenesis for the better part of four decades. Long before molecular biological tools became available in the 1970s, investigators were able to biochemically identify and characterize a collection of virulence factors and virulence-associated phenotypes, including a large number of bacterial toxins. With the introduction of modern genetic techniques, rapid progress was made toward understanding not only the factors that contribute to bacterial virulence, but also the regulators that control the expression of these genes and the stimuli that modulate the activity of these regulators. Many of the early studies of virulence relied on the use of *in vitro* systems to either model some aspect of the host environment or elicit the production of known virulence factors. As genetic systems improved, a move was made toward the direct identification of virulence mutants *in vivo*.

Large-scale screens to identify virulence factors and their regulators

Early genetic screens aimed at understanding *V. cholerae* virulence relied heavily on the creation of gene reporter fusions *in vitro*. For example, activators of cholera toxin production could be found by screening an *E. coli* reporter strain carrying a fusion of the cholera toxin promoter (*ctxAB*) to *lacZ* (Miller and Mekalanos, 1984). A genomic library from *V. cholerae* was transformed into this strain, and clones that expressed β-galactosidase activity could be isolated and characterized, leading to the identification of ToxR, key regulator of *V. cholerae* virulence. Similar efforts were used to elucidate a whole transcriptional regulatory cascade that is central to *V. cholerae* virulence (DiRita et al., 1991).

Alternative approaches have also been used to identify virulence factors through associated function. For example, TnphoA mutagenesis, which generates clones that produce alkaline phosphatase when the transposon inserts into genes encoding secreted proteins, was used to screen for secreted proteins that were coordinately regulated with cholera toxin (Taylor et al., 1989). This approach identified TcpA, a toxin-coregulated pilin that is required for virulence in both animals and humans. While such reporter-based studies paved the way for the development of correlative techniques such as IVET, these reporters could also served as reporters of function in *in vitro* screens and studies.

Genes identified in large scale, *in vitro* screens that were thought to play a functional role in virulence were historically tested for *in vivo* significance by challenging an animal model, typically mice, with individual mutants. However, in order to take a more comprehensive approach to defining genes essential for *in vivo* infection, methods to directly screen for virulence-associated phenotypes in animal models were needed. The need was met by the development of signature-tagged transposon mutagenesis and other negative selection-based pooled screens.

Before discussing the use of transposons in these *in vivo* screens, it should be noted that the construction of large transposon libraries provides the opportunity to identify genes that are required for growth and survival in culture (Judson and Mekalanos, 2000b). Many of these genes are likely to be required for *in vivo* growth as well, making them potentially attractive as targets for therapeutic intervention. Methods to identify indispensible genes, that is, genes that cannot sustain a transposon insertion, have been evolved, and currently high-throughput screening of transposon insertion sites provide this information in high resolution. Nonsequencing methods such as PCR-based footprinting remain useful for interrogating small regions of the genome. Regardless of how the insertion sites are mapped, the absence of transposon insertions in a gene does not prove essentiality; essentiality must be demonstrated on a gene-by-gene basis by targeted efforts to delete or disrupt candidate essential genes. A positive approach was developed by Judson and Mekalanos (2000a) that identifies essential genes by transcriptionally fusing them to a regulatable promoter located at the end of the transposon. Using this approach, they identified arabinose-dependent strains of *V. cholerae* that included characterized essential genes as well as previously uncharacterized genes encoding hypothetical proteins.

While *in vitro* essential genes may have value as candidate drug targets, clearly there is much to be gained by identifying genes that merely conditionally essential. For studies of bacterial pathogenesis, the most relevant condition for study is growth *in vivo*. Methods for identifying *in vivo* essential genes are discussed below, along with the advantages and pitfalls associated with the various experimental methodologies.

Negative selection screens

A challenge to bacterial geneticists interested in acquiring functional information is the issue of scale. Mutant hunts require the generation of thousands of random mutants, and most importantly, a well-conceived selection or screen. For a systematic study of virulence, screening conditions are best achieved by introducing the mutants into a host—ideally using an animal model relevant to human disease, but are also achievable for some organisms using

cell culture or other *in vitro* conditions mimicking the host environment. Since the best-characterized animal model for many bacterial pathogens is the mouse, this poses an obvious dilemma. How can the behavior of thousands of mutants be monitored within an animal host without the need to infect thousands of mice with individual mutants?

In 1995, Hensel et al. provided an elegant solution to this problem (Hensel et al., 1995). Their strategy was to infect a mouse simultaneously with a pool composed of 96 different bar-coded mutants, followed by identification of attenuated members of these mutant pools through a process of negative selection (Figure 3A). Their approach, dubbed signature-tagged mutagenesis, or STM, has become a standard methodology in the field, and has been used in dozens of laboratories to identify virulence factors in a broad variety of bacterial pathogens. STM led the way to the development of a series of transposon-based negative selection genetic techniques that allow for the simultaneous screening of increasingly large and complex pools of mutants, and that provide increasingly quantitative information on the relative importance of each gene to the success of the infecting bacterium.

Signature-tagged mutagenesis In their initial publication, Hensel and Holden used a collection of modified transposons to mutagenize *S. typhimurium*. These transposons each carried a unique, randomly generated oligonucleotide tag flanked by conserved PCR primer-binding sites, so that the tags could later be PCR-amplified and used to track mutants (Figure 3A). The random transposition of these tagged transposons into *S. typhimurium* led to the production of large numbers of mutants, each of which had two distinguishing features: a unique, random site of insertion, and a unique oligonucleotide tag. The disrupted locus determined the behavior of the mutant, and the tag allowed the mutant to be tracked and identified. Growing each mutant in an individual well of a 96-well plate and then combining aliquots of these individual wells to create inoculum pools, the authors were able to infect mice with complex mixtures of mutant strains. The census of the pool was taken both before and after infection by PCR amplifying the tags and then hybridizing the radiolabeled PCR products against a membrane spotted with the 96 transposons (and their unique tags) used to create the pool. Hensel and Holden were able to screen 1152 mutants in this manner, resulting in the isolation of 43 attenuated mutants, including several with insertions in genes with homology to components of a type III secretion system (located on the SPI1 pathogenicity island). This locus is known to be important for *Salmonella* invasion of epithelial cells. Since the complete genome sequence of any *Salmonella* species was still years away from publication, the discovery of this second type III secretion system required for virulence and conserved across pathogenic *Salmonella* species was a significant contribution to the current understanding of how *Salmonella* manipulates the biology of its host cell from within the phagosome.

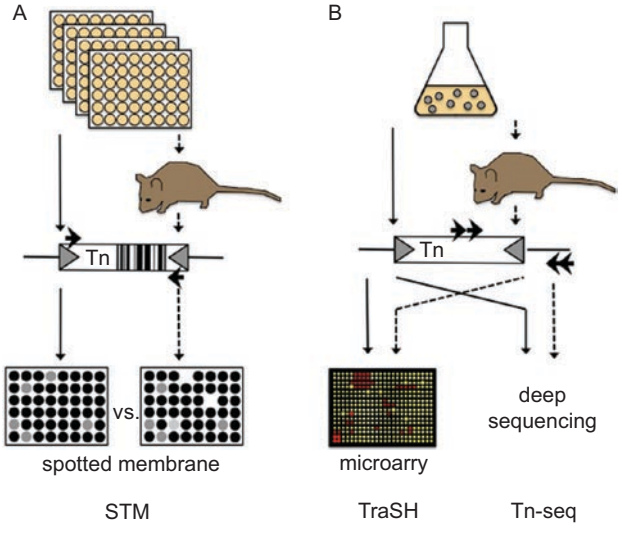

Figure 3. Negative selection screens. (A) STM uses bar-coded transposons to create limited pools of mutants. These pools are passed through a selection, that is, mouse model that filters out attenuated mutants. The mutants that disappear from the pool during mouse passage are detected by PCR amplifying the bar codes, and comparing the input pool and output pool that is hybridized to a spotted array. (B) TraSH and Tn-seq approaches involve probes made from amplified transposon junctions to monitor the fitness of a genome-wide library of mutants following selection. In TraSH, these probes are hybridized to a microarray, while in Tn-seq, they are directly sequenced to identify mutants that disappear from the output pool.

During the 1990s, STM was applied to the study of a wide variety of genetically tractable bacterial pathogens, including *Staphylococcus aureus* (Mei et al., 1997; Coulter et al., 1998; Schwan et al., 1998), *V. cholerae* (Chiang and Mekalanos, 1998), *Streptococcus pneumoniae* (Polissi et al., 1998), *Yersinia enterocolitica* (Darwin and Miller, 1999), *Legionella pneumophila* (Edelstein et al., 1999), and *M. tuberculosis* (Camacho et al., 1999; Cox et al., 1999). The functional genetic elements identified in these studies include clear-cut virulence factors like components of type III secretion systems and their associated effectors, genes required for the synthesis of cell surface-associated molecules involved in host–pathogen interactions, and genes encoding metabolic enzymes and damage–repair proteins.

By screening identical mutant pools in multiple infection models, STM can reveal whether virulence loci are universally needed to support infection or are specifically required for processes that are unique to a given model. With the objective of identifying *S. aureus* loci that fall into the former class, Coulter et al. at PathoGenesis Corporation utilized three different murine infection models, systemic bacteremia, abscess, and wound infection, to screen over

1500 transposon mutants (Coulter et al., 1998; Schwan et al., 1998). Since the long-range objective of this work was the identification of pathways that might serve as targets for therapeutic intervention, the group chose six mutants that showed significant attenuation in all three models for further evaluation. All six mutants were then found to be attenuated in yet another system, a rabbit endocarditis model. Notably, of the 237 mutants that were identified as being attenuated in at least one of the three screening models, only 23 (<10%) were attenuated in all three. Thus, the use of a single model for STM-based studies may very well result in an incomplete inventory of virulence loci, suggesting that performing such screens in multiple animal models may be important.

The differential behavior of mutants in multiple animal models can be used to deliberately identify genes involved in responding to defined host defense mechanisms. Hisert et al. rescreened *S. typhimurium* mutants that had been shown to be attenuated in immunocompetent mice in a series of gene knockout mice that lacked the genes encoding the inducible nitric oxide synthase (iNOS; NOS2), phagocyte oxidase (phox), or both of these phagocyte effector systems (Hisert et al., 2005). One mutant with an insertion in a conserved hypothetical protein was attenuated only in the presence of a functional phagocyte oxidase, implicating this locus in the bacterial defense against this host effector. The group went on to show that this previously uncharacterized locus, renamed *cdgR*, encodes a protein that regulates intrabacterial cyclic diGMP levels and is critical for survival under oxidative stress. While in this case, the *cdgR* locus could have been identified by screening a panel of *S. typhimurium* mutants for sensitivity to hydrogen peroxide *in vitro*, it is not always a simple task to generate *in vitro* screening conditions that accurately recapitulate the *in vivo* environment.

Although STM can clearly be a powerful tool, a number of issues need to be carefully evaluated prior to initiating a screen (Miller, 1999; Perry, 1999). As illustrated earlier, first and foremost among the concerns is the choice of infection model, since the genetic requirements for successful infection may vary depending on how the bacteria initially encounter the host. For example, when Hensel et al. inoculated their *Salmonella* mutant pools via intraperitoneal injection, they failed to isolate mutants in the previously characterized *inv/spa* virulence locus. This locus, which is required for intravenous infection, is thus dispensable when bacteria are introduced by the intraperitoneal route. Studies in other bacterial pathogens have also shown various route-dependent functions, as illustrated in other routes of infection including oral (Chiang and Mekalanos, 1998) or intranasal (Polissi et al., 1998) pathogen delivery.

Not only do the different routes of infection affect the host defenses that may be encountered upon inoculation, but they can also lead to potentially different fates for the individual members of an inoculating pool due to the stochastic loss of mutants that can occur at population bottlenecks. Such a population bottleneck was described for a rat model of *Streptococcus sanguinis* endocarditis (Paik et al., 2005). Despite infection with as many as 10^{10} bacteria, only a small fraction of the 40 mutants present in the input pool were able to be recovered from rat heart valves, suggesting that extremely small numbers of bacteria were involved in the establishment of this infection. The authors were able to overcome this problem by switching to a rabbit endocarditis model, which resulted in more reproducible representation of the 40 mutants in the output. The choice of an appropriate inoculum size and pool complexity, as well as the study of each pool in more than one animal, can help to minimize the stochastic variability that occurs with population bottlenecks (Miller, 1999).

Another major limitation of STM and other negative selection-based screens that rely on the introduction of pools of mutants is that certain gene functions cannot be identified due to *trans* effects. Since each mutant comprises only a small fraction of the total bacterial population introduced into the selection system, it is possible that some mutants that would be severely attenuated in a monotypic infection might be rescued by their pool mates in *trans*, particularly if the disrupted virulence gene encodes a secreted factor. Thus, the identification of gene functions that can operate in *trans* can be challenging to be identified by these methods.

STM continues to be a valuable tool for the identification of virulence determinants, as illustrated by the recent discovery of the Trw type IV secretion system in *Bartonella* (Vayssier-Taussat). While the introduction of STM has certainly improved upon the inefficiency of introducing individual mutants into individual host animal, newer technologies have recently improved upon the ability to assess the relative fitness of mutants during pooled infection. Such technologies now enable the simultaneous interrogation of genome-wide mutant pools.

Negative selection using microarray-based readouts STM as a gene discovery tool is limited by the small number of mutants that can be simultaneously inoculated into a given host (i.e. the complexity of a given pool) because of the challenges of tracking increasing numbers of mutants. Thus, correspondingly large numbers of STM pools must be screened to approximate a saturating screen that produces a comprehensive inventory of a pathogen's virulence genes. As a result of these limitations, very few of the STM studies published in the 1990s could claim to have approached saturation. As the readout was simplified by the use of oligonucleotide tagged microarrays, saturating or near saturating screens became a more realistic objective. At the same time, as the complete genome sequences of bacterial pathogens became increasingly available, new ways to functionally interrogate these organisms were developed.

Access to the complete genome sequence of a microbe makes possible the construction of microarrays where every gene is represented on the array. Although most early work with microarrays focused on evaluating changes in gene expression, it became clear that these tools could also be used for functional studies. Instead of relying on tags that had to be introduced on the transposons used to generate the mutant pools, the fitness and fate of mutants could instead be assessed by interrogating whole genome microarrays with a complex probe mixture derived from DNA adjacent to transposon insertion sites. When such probes are made from a transposon library that has undergone selection and are hybridized to an array, weak (or no) hybridization signal will be observed on the microarray at spots corresponding to transposon insertions into genes that are conditionally essential under the experimental condition (Figure 3B). Genes that are essential under all conditions, or at least under the conditions in which the transposon library was generated, will not be able to sustain a transposon insertion and will thus be absent from the starting pool.

Like STM, microarray-based negative selection approaches are limited by population bottlenecks that might occur during animal infections. However, when selection conditions do not generate these bottlenecks, microarrays enable the rapid screening of many thousands of mutants simultaneously. An additional benefit is that the readout for a particular gene is often determined by the behavior of multiple mutants within the pool, carrying transposon insertions in different regions of the gene.

Aside from the ability to screen larger and more complex pools simultaneously, whole genome microarrays allow for a rapid determination of the genes that are capable of sustaining a transposon insertion during library construction. Thus, after construction of the transposon library, hybridization signals of the "input" probes generated from this library, relative to control probes made against genomic DNA, provide an indication of which genes are essential for *in vitro* growth. This approach can be confounded by the fact that microarray-based mapping of transposon insertion sites is a fairly low resolution methodology, and thus some hybridization signal at a given spot on the array can be the result of signals from Tn insertions in adjacent genes. Similarly, the statistical likelihood of a transposon insertion into a given gene is related to the size of the gene. Thus, the absence of a transposon insertion in a particular gene may result in the false positive assignment of essentiality. Finally, because such approaches are transposon-based, polar effects on downstream genes can confound all such efforts to identify essential genes.

The first reports of microarray-based negative selection in bacteria were effectively proof of principle experiments, where mutant libraries were generated on rich growth medium and then subjected to selection in minimal medium. Not surprisingly, subsets of mutants defective in nutrient uptake, amino acid biosynthesis, and similar processes were identified as being negatively selected under nutrient limiting conditions. Badarinarayana et al. used a modified Tn10 transposase to mutagenize *E. coli* (Badarinarayana et al., 2001), and Sassetti et al. created a modified Mariner Tn system for use in mycobacteria, where they termed their methodology TraSH, for transposon site hybridization (Sassetti et al., 2001). TraSH was later applied to the study *M. tuberculosis* growth and survival in both macrophages (Rengarajan et al., 2005) and a murine model of tuberculosis (Sassetti and Rubin, 2003), allowing for a comparison of the genetic requirements for infection of these two distinct infection models. Somewhat surprisingly, the set of genes required for macrophage infection did not significantly overlap with the set of *M. tuberculosis* genes whose expression had previously been shown to be up-regulated during macrophage infection (Schnappinger et al., 2003), thus highlighting potential limitations of correlative approaches. Similar approaches to TraSH have been used to identify essential genes in *H. pylori* (Salama et al., 2004), and a modified approach using a library of defined mutants, called DeADMAn, was also used to identify requirements for *M. tuberculosis* survival in mouse lungs (Lamichhane et al., 2005).

Such negative selection approaches can be taken to increasingly sophisticated levels, including screens for interacting genes (Figure 4). In species that are easily mutagenized using transposons, once an interesting virulence gene has been identified, microarray-based negative selection can be used iteratively to reveal other genes that interact with it. Transposon libraries are generated in both a wild-type strain and a mutant strain lacking the gene of interest. Both pools of mutants are subjected to the same selective conditions. The recovered bacteria from both the wild-type pool and the mutant pool are compared by microarray hybridization. Insertions in genes that are functionally unlinked to the gene of interest will have a quantitatively similar phenotype in both genetic backgrounds. In contrast, insertions that disrupt genes that function within the same pathway as the gene of interest will display a decreased level of attenuation in the mutant (epistasis), and insertions that disrupt genes that are functionally redundant with the gene of interest will show an exacerbated phenotype in the mutant. Suppressor mutations (insertions) can also be identified since they will be overrepresented in the mutant background postselection.

Using this approach, a partial functional overlap between two transporters in *M. tuberculosis*, *mce1* and *mce4* was revealed (Joshi et al., 2006). Both genes had been previously shown by TraSH to be required during infection in mice. The *mce* operons of *M. tuberculosis* are composed of genes with homology to the transmembrane

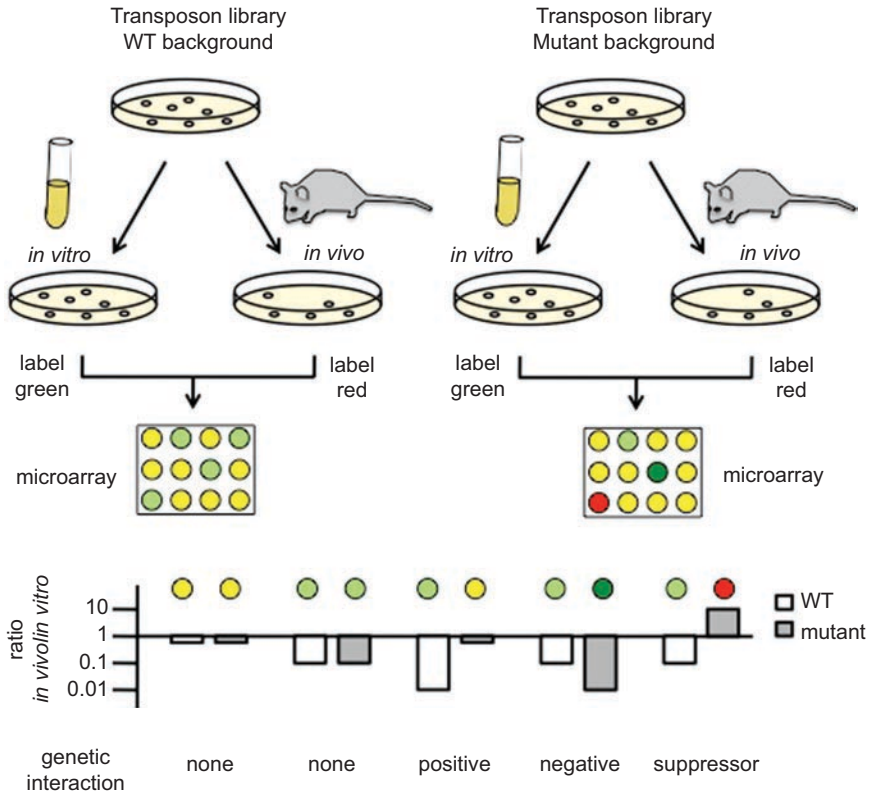

Figure 4. TraSH-based approach for genetic interaction mapping. Functional pathways can be delineated using parallel TraSH screens. Standard TraSH experiments are performed using saturated transposon libraries generated in the wild-type background or in the background of a mutant strain chosen to represent the pathway of interest for genetic interaction mapping. Both libraries are passaged through a condition of interest. For studying virulence, passage through a mouse or other animal model is compared to growth *in vitro*. Separate genomic libraries are generated from the pool of mutants recovered after growth *in vitro* or *in vivo*, and independent microarray experiments are performed to identify genes that are required for growth *in vivo* in both the wild-type and mutant backgrounds. Transposon mutations that cause the same phenotype in both genetic backgrounds have the same ratio in each microarray experiment, and do not interact genetically. Negative genetic interactions will produce smaller ratios in the mutant background, positive interactions with produce larger ratios, and suppressors of the mutation will produce ratios greater than one.

and periplasmic binding components of ABC transporters. Yet, neither of the *mce* operons encodes a protein that might serve as the energy providing ATP-binding component required by these transporters. Interacting TraSH analysis of *mce1* and *mce4* mutants led to the identification of this ATP-binding component (transporter), which is encoded at an unlinked locus and supplies the ATP-binding activity for all of the *M. tuberculosis mce* operons.

Clearly, a significant limitation of microarray-based negative selection is the quality and availability of the microarray. For many pathogens, arrays are commercially available, and if not, array design and synthesis are becoming increasingly easy. However, strain-to-strain variations may limit the effectiveness of this method if experimental strains are studied that contain genes that are not represented on the array. Thus, this strategy is entirely dependent on the availability of a well-annotated genome sequence for the pathogen being studied and the quality (completeness) of available arrays. The quality of such arrays is continuing to improve with the incorporation of spots on the microarrays to detect regulatory RNAs and other noncoding transcripts as well as mRNAs. Recently, powerful sequencing methods have been applied to obtain quantitative fitness measurements of members of the mutant pools, thus allowing these negative selections to circumvent many of the current limitations associated with using microarrays to measure fitness. This method has been variably termed Tn-seq (van Opijnen et al., 2009), TraDIS (transposon directed insertion site sequencing) (Langridge et al., 2009), or HITS (high-throughput insertion tracking by deep sequencing) (Gawronski et al., 2009).

Fitness and virulence assessment using sequencing-based readouts Massively parallel sequencing can precisely define transposon junctions and acquire an accurate census of individual mutants within the pool with an accuracy that far exceeds microarray-based approaches. Illumina sequencing is ideal for this purpose. Illumina generates relatively short sequence reads (36–120 nt), but this size is well-suited to cover a region that contains a short region of the transposon end, fused to enough genomic

sequence to allow the unambiguous identification of the insertion site (in most cases). Illumina is highly accurate, particularly in shorter reads, and can generate up to 5–10 giga base pairs (Gbp) per run (van Vliet, 2010). Although the run times may be as long as several days, the cost per Gbp of sequence is relatively inexpensive and multiplexing of samples is possible when less than 5–10 Gbp of data per sample is needed.

To map and count transposon insertion sites, the usual Illumina genome sequencing sample preparation protocol is modified slightly. After fragmentation of the pool of genomic DNA, adapters are ligated to the free ends and the library is PCR-amplified using an adapter-specific primer and a transposon-specific primer. Therefore, the vast majority of reads contain transposon sequence and adjacent genomic sequence of interest, allowing up to several million unambiguously mappable reads per flow cell lane (Langridge et al., 2009; van Opijnen et al., 2009). Since prokaryotic genomes generally contain ~1000–6000 genes, this means that many thousands of reads in a single run may map to a single gene. With this depth of coverage, much more subtle fluctuations in fitness related to the disruption of each gene can be measured. Further, the list of genes that are essential to the viability of a given bacterium can be determined with a much higher degree of confidence.

Tn-seq/TraDS/HITS have been applied to *S. typhi* (Langridge et al., 2009), *S. pneumoniae* (van Opijnen et al., 2009), and *Haemophilus influenzae* (Gawronski et al., 2009). In *S. typhi*, a pool of over 1 million transposon mutants was created and assayed for two phenotypes: growth in rich medium and tolerance to bile, which is important for *S. typhi* survival in the gall bladder. Three hundred and fifty six genes were deemed essential based on a statistically significant underrepresentation of insertions in their respective open reading frame, after normalizing for the size of the ORF. The mutant pool was subjected to six passages, with samples taken for analysis after each passage. By the sixth passage, the pool complexity had diminished, and the authors were able to define genes that were both advantageous and disadvantageous during this period of outgrowth. Interestingly, a set of 30 genes involved in flagellar synthesis appeared to carry a substantial cost during *in vitro* growth in nutrient-rich medium, as Tn insertions in these genes were substantially overrepresented in the pool after the sixth passage.

In *S. pneumoniae*, 398 genes were determined to be either possibly or likely essential, based on similar criteria to those described above. The depth of information provided by deep sequencing enabled more than the simple designation of genes as either essential or nonessential; by analyzing the pool composition before and after a period of expansion (about seven generations), the relative contribution of every gene to pneumococcal fitness was able to be determined. Using the interaction mapping application described above, genetic interactions were mapped for five genes of interest: three transcriptional regulators and two transporters. Tn libraries were constructed in five strains in which the corresponding genes of interest had been replaced by a drug-resistance marker. Genes were scored as interactors when the fitness cost of the two lesions in combination differed from the product of the fitness cost of disrupting the genes individually. Such studies provided the basis for an increasingly systems biology-scale analysis of gene networks in bacterial pathogens.

In *H. pylori*, deep sequencing was used not only to define essential genes *in vitro*, but also to evaluate the *in vivo* fitness of a 75,000 member mutant pool. From analysis of the insertion sites in the input pool, 1239 genes were not required for *in vivo* growth. After passage through mouse lungs, 136 of these genes had significantly fewer sequencing reads mapping within their boundaries, suggesting *in vivo* selection against mutants carrying insertions in these genes. The authors constructed deletion mutants in two genes identified in their screen, *galU* and *orfH* and showed that their attenuation levels in a monotypic infection matched, within error, the attenuation levels predicted from the sequence-based readout.

Despite its technical improvements over previous negative selection approaches, identification of underrepresented mutants in a pool by deep sequencing is still subject to the same limitations as all other negative selections. Primary among these is the mixed nature of the infection, which limits the ability to identify functions that can be complemented in *trans*, and prevents the discovery of virulence phenotypes that are not associated with a large fitness cost to the pathogen but may impact the disease process in the host. Only monotypic, mutant-by-mutant screens can provide this type of information. The next section will discuss how recent, parallel advances in currently available resources, such as full genome sequences, rationally designed minimal mutant libraries, and new animal models for screening make these complementary types of screens possible.

Nonpooled approaches to identify virulence genes

Once the genomic content of an organism is defined, interrogation and analysis of that content can be streamlined by creating a library that is nonredundant, or at least minimally redundant, so that the entire gene set can be assayed (as individual mutants) with minimal labor and cost. For example, for an average bacterial genome containing 4000 genes, a screen can now be performed on a library of ~3500 clones (minus the essential genes in a genome that cannot be constructed and thus interrogated in this manner) compared with screening of a nondefined, redundant library, which may require interrogation of ~20–30,000 clones in order

to obtain the same level of saturation. This reduction in numbers makes it possible to screen these libraries for increasingly complex phenotypes such as transcriptional alterations or even growth inside an animal model such as the mouse.

There are a few caveats one must consider when creating and using such defined, nonredundant libraries. First, this strategy relies on an accurate bioinformatic evaluation of the genome, including proper annotation of coding regions and the identification of other chromosomal regions of interest, such as noncoding RNAs. Second, in nonredundant, defined libraries, there are no longer multiple mutants per gene to internally validate the primary screening results. Third, if the library is assembled from transposon mutants, it is possible that some of the mutants chosen for inclusion in the pool do not represent true null alleles (as in the case of insertions at the 3'-end of a gene), although careful mutant selection from large pools of available insertion mutants can minimize this risk. Also, if the library is created from transposon mutants, the clones are still subject to confounding polar effects of the insertions. Despite these drawbacks, the advantage to using defined libraries—the ability to carry out complex screens on a genome-wide scale—is quite large.

Defined mutant libraries The construction of comprehensive mutant libraries requires a substantial upfront investment in time and effort, but a high-quality, ordered library is a valuable resource that can be shared throughout the research community. These libraries can take a variety of forms, and with enough foresight, the mutations can be introduced along with additional features that increase the versatility of the library by allowing for pooled use, measurement of gene expression, or further genetic manipulation. The potential utility for such bacterial libraries is exemplified by the earlier construction of a comprehensive, bar-coded mutant library in the budding yeast *Saccharomyces cerevisiae*, which has absolutely revolutionized eukaryotic systems biology.

Sequence-defined mutant libraries can be assembled from existing libraries of transposon mutants, as has been done with *Francisella novicida* (Gallagher et al., 2007), *V. cholerae* (Cameron et al., 2008), and two strains of *P. aeruginosa* (Jacobs et al., 2003; Liberati et al., 2006); alternatively, mutants can be generated in a targeted manner, as has been done in *Bacillus subtilis* (Kobayashi et al., 2003) and *E. coli* (Baba et al., 2006). To generate the Keio collection of *E. coli* mutants, every deletable ORF in the *E. coli* genome was individually replaced with a kanamycin-resistance cassette flanked by FLP recombinase recognition sites. When the resistance cassette in each mutant was resolved, the resulting mutant contained a clean, in-frame deletion of the target gene, thus eliminating concerns about polar effects on downstream genes.

Defined libraries are useful in several ways. First, they serve as a repository of easily accessible individual mutants, enabling the rapid follow-up to experiments that implicate specific genes in a process. For example, if a gene is found to be induced in a particular infection model, a mutant in that gene can be quickly retrieved and its functional significance in that model can be tested. Second, the library can be screened systematically and comprehensively for a phenotype of interest. The *P. aeruginosa* libraries have thus far been screened for phenotypes including biofilm formation (Müsken et al., 2010) and antibiotic sensitivity (Lee et al., 2009) and the *V. cholerae* library has been screened for motility-related genes (Cameron et al., 2008). *In vitro* phenotypes such as these are rapidly assayed at relatively low cost, in contrast to individually passing the nearly 6000 mutants present in the *P. aeruginosa* nonredundant library through an *in vivo* model such as the mouse, which is a much more daunting undertaking.

Such arrayed libraries can be parsed into sub-libraries containing a subset of genes that may share a common feature. The task of testing a sub-library in a whole organism, animal model is more feasible, as illustrated by the systematic study of 83 regulators that had been inferred to play a role in *S. typhimurium* virulence (Yoon et al., 2009). Testing each of these regulatory mutants in three different mouse infection models identified 35 regulators that were important in at least one infection model, and 14 that were needed for systemic infection. Each of these 14 regulatory mutants was then grown under four different conditions and the impact of the gene deletion on global gene expression was evaluated by microarray analysis, thus providing an in-depth, systems biology-based description of the regulatory cascade leading to the production of SPI-2-encoded virulence factors.

High-throughput infection models One of the remarkable experimental systems that can now be evaluated due to the creation of ordered, arrayed libraries is the single, gene-by-gene interrogation of a pathogen within a whole organism host model. While efforts are now underway to perform such daunting studies in higher-order vertebrates that more closely resemble the human host, that is, mice, coincidental efforts have been made to identify simpler host models that are easier to screen and may nevertheless provide insight into relevant infection biology.

The nematode *C. elegans* has been a workhorse for developmental biologists but is now also found use as a simple, inexpensive model to study the virulence mechanisms of several bacteria, including *P. aeruginosa* (Tan and Ausubel, 2000) and *S. aureus* (Sifri et al., 2003). *C. elegans* is a particularly easy to use model in that the nematodes naturally consume bacteria, thus infecting themselves during feeding. Ten or more worms can be

grown in a single well of a 96-well plate, allowing for the screening of large collections of bacterial mutants (or small molecules, as will be discussed later). The outcome of infection is also easily measured, since dead nematodes assume a straight, rigid morphology in contrast to the S-curved morphology displayed by viable organisms. This morphological change upon death is sufficiently robust such that worm viability can be scored by automated microscopy. Many, though not all, of the bacterial mutants defective for growth in *C. elegans* have been shown to be attenuated in mouse models of infection (Mahajan-Miklos et al., 2000).

Another favorite model organism of developmental biologists, the zebrafish (*D. rerio*), is also a suitable host for a variety of bacterial infections, including *S. typhimurium* (van der Sar et al., 2003), *P. aeruginosa* (Clatworthy et al., 2009), *S. aureus* (Prajsnar et al., 2008), *Streptococcus* (Phelps et al., 2009), and *M. marinum* (Cosma et al., 2006). The zebrafish has several attractive features. First, embryos are transparent, allowing the infectious process to be visualized by light or fluorescence microscopy in living embryos over the course of the infection. Second, zebrafish are genetically tractable, and morpholinos can be used to modify host gene expression before and during infection. Third, zebrafish possess both innate and adaptive immunity similar to mammals. Fourth, zebrafish fecundity and the small size of embryos, which may be housed in 96-well plates for up to 12 days, facilitate the generation of large numbers of host animals that can be housed in a small space, facilitating higher-throughput studies. Finally, as with *C. elegans*, mutants that are attenuated in vertebrate models have also been shown to be attenuated in zebrafish. The use of these and additional nonvertebrate infection models has been nicely reviewed elsewhere (Mahajan-Miklos et al., 2000; van der Sar et al., 2004).

Clearly, the application of genomics to the study of bacterial pathogenesis has dramatically changed the scale and scope of studies that can be performed. Increasingly, facile methods are being developed that allow systematic genetic perturbation of bacterial (and host) genomes. At the same time, more comprehensive methods are being developed to allow measurements of bacterial response to genetic interventions with increasing sensitivity and quantitative precision. The progress made in these areas is complemented by the development of new host models, as *in vitro* studies can only provide a limited view on true infection of relevant hosts. However, in the face of an increasingly endless number of exciting genetic and genomic developments, there continue to be inherent limitations to manipulations on the DNA level. Thus, the increasing application of chemical biology to the study of bacterial pathogenesis serves to fulfill a unique niche and is an important complementary approach to genetic methods.

Small molecule-based identification of essential genes

Functional genetic approaches to studying pathogenesis have been incredibly powerful, and the technological advancements that have come hand-in-hand with the application of genomic methods portend an even more exciting future for understanding pathogenesis in increasingly comprehensive and relevant ways to human infection. However, even as many technical barriers to understanding bacterial pathogenesis are being removed, the study of bacterial essential gene function and the study of genetically intractable pathogens continue to be challenging if one is restricted to utilizing only classical genetic approaches. To date, most functional studies are based on the ability to knockout or mutate a particular genetic locus, usually resulting in loss of function. Clearly, for genes that are essential *in vitro*, creating such lesions is not possible. While it is possible to perturb essential genes by creating genetic "knockdowns" that titrate the expression of gene products and thus regulate the magnitude of gene function (Wong and Akerley, 2008) on a genome-wide scale, this type of effort can be quite challenging. Likewise, conventional genetic approaches have either been limited or are impossible in certain organisms refractory to current genetic techniques for DNA manipulation, including such organisms as *Chlamydia*, *Rickettsia*, *Coxiella* spp., and *Mycobacterium leprae.*

The use of chemical genetics (the application of small molecules) to perturb bacterial behavior has become an increasingly valuable and complementary tool to examine the involvement of essential genes in pathogenesis and to interrogate genetically intractable pathogens. In addition, small molecules can also play complementary roles to genetic manipulations under specific circumstances, such as when protein turnover is slow and therefore genetic control of transcription levels has little bearing on protein expression levels and thus protein function, when redundant gene pairs and/or gene families exist in a genome complicating the relationship between single gene and gene function, or when compensations (mutations) can occur in a genome in response to a long-term, stable genetic lesion resulting in confounding results (Figure 5). In these and other cases, the ability of a small molecule to have rapid onset and to be reversible can be important (Stanley and Hung, 2009).

Many of the same phenotypes that can be screened through experimental genetic methods can also be screened using a chemical biological approach to identify small molecules that interfere with gene function, most often through inhibition of protein function. Most pathogenesis-related or bacterial chemical screens (or selections) fall into three main categories that are designed to find small molecule modulators of (1) *in vitro*

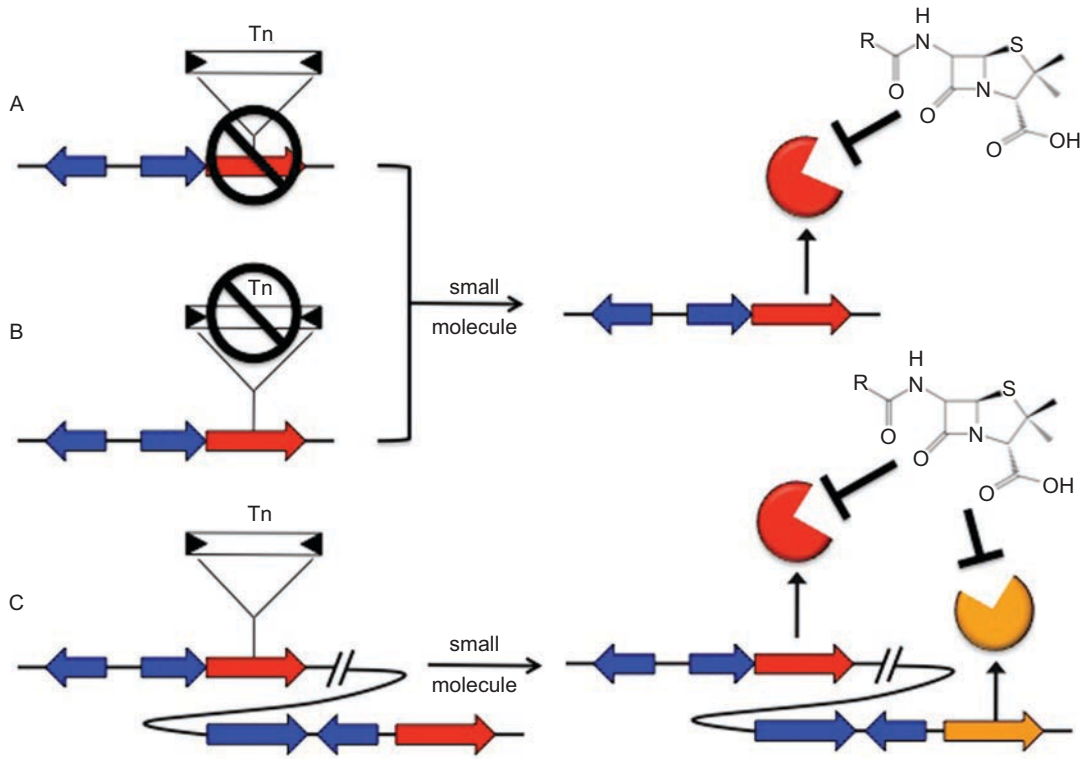

Figure 5. The use of small molecules to modulate bacterial behavior in situations where standard genetic approaches fail. (A) Genes that are required for *in vitro* growth cannot be disrupted by conventional genetic approaches including transposon mutagenesis; however, the functions of their gene products can be modulated by small molecules. (B). In cases where pathogens are refractory to genetic manipulation, small molecules can be used to perturb protein function. (C) Functionally redundant genes can complement each other and thus prevent the isolation of mutants during conventional genetic screens, although small molecules may be capable of interfering with the function of gene products that play similar functional roles and show good homology within their respective active sites.

virulence phenotypes, (2) bacterial physiology and thus survival, or (3) the complex host–pathogen interactions involved in infection of a whole host organism. Many chemical screens of these types have become of increasing interest because of the overlap that they have with antibiotic discovery efforts and establishing new paradigms for how one defines an antibiotic (Clatworthy et al., 2007; Barczak and Hung, 2009).

Inhibition of virulence phenotypes

Numerous virulence phenotypes have now been screened *in vitro* to identify small molecule inhibitors of virulence for both Gram-positive and Gram-negative organisms, with increasing numbers of reported successes in whole animal models (reviewed in Clatworthy et al., 2007; Escaich, 2008; Rasko and Sperandio, 2010). A number of small molecules have been identified, which probe a large and diverse set of bacterial functions associated with virulence including: adherence, as exemplified by *E. coli* pilicides (Cegelski et al., 2009); virulence regulation, in the case of cholera toxin expression in *V. cholerae* (Hung et al., 2005) or quorum sensing in both Gram-positive and Gram-negative organisms (Balaban et al., 2007; Rasko et al., 2008); toxin delivery, through the inhibition of secretion systems (Baron, 2006); and toxin function, for example, by inhibiting anthrax lethal toxin protease activity (Shoop et al., 2005).

The methods for identifying small molecule modulators of virulence are many and varied and have included biochemical assays, such as hemagglutination, which has been used to detect type 1 pilus formation in *E. coli*, or protease activity, which has been used to detect anthrax lethal toxin cleavage of a peptide substrate by FRET (Cummings et al., 2002). In addition, bacterial strains manipulated to express reporter genes (e.g. luciferase, fluorescent proteins, or antibiotic-resistant cassettes) under the control of virulence factor promoters have been used to screen for small molecule inhibitors of quorum sensing and type III secretion (Kauppi et al., 2003). Other whole cell, phenotypic assays have been performed, including assays to identify inhibitors of biofilm formation (Junker and Clardy, 2007) and inhibitors of secretion, as measured by the ability of pathogens to kill host cells in cell culture (Arnoldo et al., 2008). In the future, we would anticipate that two-hybrid systems will be used to identify inhibitors of protein–protein interactions that may be critical for virulence phenotypes; the use of such an assay is certainly plausible based on

precedence in yeast (Armour and Lum, 2005). In fact, a bacterial two-hybrid system was used to verify the requisite dimerization of the transcription factor ToxT in *V. cholerae* and its disruption by a small molecule inhibitor, virstatin (Shakhnovich et al., 2007).

The types of screens or selections that can be performed with small molecules are nearly limitless since there are few restrictions on the system (i.e. in principle, no genetic manipulations are required). One simply needs to have a robust phenotype that can be scored to identify small molecules of interest. One clever example of how chemical genetics was used to identify an essential virulence factor in the genetically intractable intracellular pathogen *Chlamydia pneumoniae* has been recently described (Huang et al., 2008). In this study, Huang and colleagues transformed yeast with a putative virulence factor from *C. pneumoniae*, CopN, and found that CopN overexpression in yeast resulted in mitotic spindle disruption and G2/M cell cycle arrest, similar to CopN overexpression in mammalian cells. The authors then performed a chemical screen to identify small molecule inhibitors of CopN-mediated cell cycle arrest in yeast and found two small molecules that not only rescued yeast from mitotic spindle disruption but also interfered with *C. pneumoniae* replication and inclusion body formation in cultured, mammalian cells. Thus, in this example a chemical genetic approach was paired with a more classical genetic approach (analysis of protein function through heterologous overexpression in yeast) with the result that the combination of both approaches led to the identification of an essential virulence factor in a genetically intractable pathogen as well as useful, small molecule tools with which to probe *Chlamydia* biology.

Modulators of physiology: death as the phenotype

It is difficult to consider the study of bacterial pathogenesis and the ultimate goal of intervening on human disease without considering the role of bacterial genes required for survival in axenic culture. The corresponding essential gene functions are the targets of all current antibiotics, which are active both *in vitro* and *in vivo*.

Current progress in genomics and chemical biology has contributed to the manner in which we can now identify and understand *in vitro* essential genes. In one extreme, the availability of sequenced, annotated genomes can be used to aid in the prediction of essential genes that might be targeted by small molecules using a "reverse chemical genetics" approach (Stockwell, 2000). At the other extreme, the application of sequencing and other genomic technologies can help in the identification of the mechanisms of action of newly identified small molecules, in some cases, even identifying the actual gene(s) targeted using a "forward chemical genetics" approach.

Genes to small molecules Genomics and chemical biology together have thus far proven to be a powerful combination for identifying small molecule inhibitors of essential gene functions. This is illustrated in the case of a small molecule natural product, platensimycin, which was discovered by Merck. Investigators used a shotgun genome-wide antisense RNAi approach to knockdown expression of genes in the *S. aureus* genome and thus identify genes whose expression was tied to growth of the bacterium (Forsyth et al., 2002). They demonstrated that knockdown in *fabF*, a β-ketoacyl synthase involved in type II fatty acid biosynthesis (FASII), sensitized *S. aureus* to inhibitors of the enzyme, in this case cerulenin. There has been considerable interest in targeting fatty acid biosynthesis as an essential function in cell wall biosynthesis in Gram-positive organisms, with isoniazid, which targets FabI in *M. tuberculosis*, being the exemplary case of an effective anti-infective. In an impressive effort, Merck screened 250,000 natural product extracts against a *S. aureus* strain in which *fabF* expression was knocked down using an antisense strategy. The screen resulted in the identification of the FabF/B inhibitor, platensimycin (Wang et al., 2006). They went on to confirm its mechanism of action and demonstrate its activity in a mouse model of methicillin-resistant *S. aureus* (MRSA) infection. Interestingly, a subsequent study reported the nonessentiality of the FASII pathway in *Streptococcus agalactiae* (Brinster et al., 2009) *in vivo* due to its ability to scavenge fatty acids from serum. This report called into question whether FASII is in fact essential and thus a reasonable target for intervention in Gram-positive organisms. However, a subsequent report (Balemans et al., 2010) has demonstrated that FASII is in fact essential in *S. aureus* and not *S. agalactiae*, thus raising a major caveat in comparative genomic efforts to define essentiality across multiple bacterial pathogens and the importance of defining essentiality within the context of relevant growth conditions.

Small molecules to genes Recent advances in genomic applications and technologies have also facilitated the identification of essential gene functions inhibited by small molecules (Figure 6). In these cases, a phenotype of interest can be generated (e.g. death of the bacteria) upon exposure to a newly identified small molecule. Using next generation sequencing, for a given molecule of interest, resistant mutants can be generated and whole genome sequencing can be used to identify the genetic lesion that confers resistance, thus revealing a candidate essential gene product that might be targeted by the small molecule (Figure 6A) (Nusbaum et al., 2009). In the particular case of *M. tuberculosis*, a high-throughput chemical screen was performed on a related nonpathogen, *Mycobacterium smegmatis*, resulting in the identification of a diarylquinoline TMC207 with potent activity against *M. tuberculosis* both *in vitro* and in a mouse

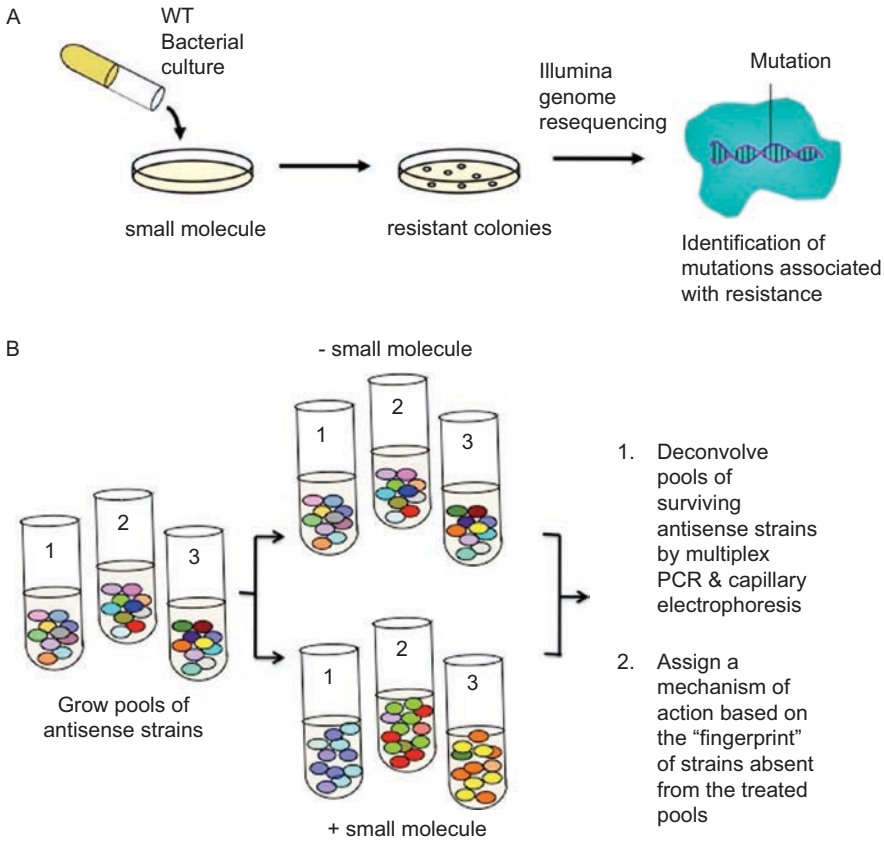

Figure 6. Genomic methods can be used to identify targets and/or mechanisms of action of small molecule modulators. (A) Whole genome resequencing of resistant mutants generated against a small molecule can be applied to identify genetic lesions within the bacterial genome that confer resistance. Such lesions can occur in the target of the small molecule. (B) Patterns of growth and fitness of an arrayed knockdown library of a given pathogen in the presence of a small molecule can result in canonical fingerprints that reveal the mechanism of action for the molecule.

model (Andries et al., 2005). The target of TMC207 was identified when mutations were identified through sequencing of resistant mutants in the F_0F_1 ATP synthase gene. Since this discovery, subsequent studies have gone on to characterize the synthase and its role in maintaining membrane potential and a proton gradient both in actively replicating bacteria as well as in quiescent bacteria (Koul et al., 2007, 2008).

In an alternative genomic approach to identifying essential mechanisms that are targeted by small molecules, Elitra Pharmaceuticals has extended the *S. aureus* knockdown approach described above to the development of a "TargetArray," which can be used to identify mechanisms of action based on the similar concept of haploinsufficiency in yeast (Giaever et al., 1999). The principle of the collection of 236 *S. aureus* antisense clones is that knockdown of specific genes/pathways should sensitize different clones to certain small molecules. Clones included in TargetArray were selected to provide unique patterns of response (or fingerprints) to distinguish between small molecules that perturb the bacteria with differing mechanisms of action. Thus, the growth pattern of the strains in TargetArray in the presence and absence of a small molecule can provide insight into that small molecule's mechanism of action (Xu et al., 2010). Specifically, the identification of clones (and thus the corresponding genes that have been knocked down) that cannot survive in the presence of a given small molecule can link the mechanism of action of the small molecule to functional pathways encoded by the corresponding genes.

In contrast to the knockdown approach described above, it is also possible to identify the cellular targets of small molecules that are required for bacterial viability by overexpression of the target. This approach, termed "high-copy suppression," involves the screening of ordered libraries of clones, where each independent clone overexpresses a gene of interest. Clones that overexpress the target of a small molecule that has growth inhibitory properties should be more resistant to growth inhibition by that compound as long as the abundance of the target is in excess of the concentration of the small molecule. Recently, an ordered, high expression clone set of essential genes in *E. coli* was screened to identify genes that suppressed the growth inhibitory properties of a compound named MAC14323, which has antibacterial

properties against a number of Gram-negative organisms including *E. coli* and *P. aeruginosa* (Pathania et al., 2009). Through this screen, Pathania and colleagues identified that LolA, a key protein involved in lipoprotein targeting, is the likely target of MAC14323. It is clear that new genomic technologies, including whole genome sequencing, antisense knockdown approaches, and construction of libraries that facilitate high-copy suppression screens, together with chemical biology, are a powerful combination to identifying small molecules that can modulate survival and essential gene functions.

Modulators of host–pathogen interaction

In a manner paralleled by advances in genetic screening, chemical biological approaches are being taken in more sophisticated ways that allow one to probe pathogen behavior in increasingly relevant biological systems, at the host–pathogen interface where the collective host response (which consists of both the host microenvironment with its associated nutrient limitations, pH, oxygen tensions, and host immunity) serves as a perturbation from which we can detect a change in bacterial behavior. Typically, the change in bacterial behavior of greatest interest is the ability to be pathogenic. Thus, we are often most interested in perturbations that result in the decreased ability of the pathogen to survive within the host or in decreased virulence, though perturbations can also be reflected in more subtle phenotypes such as improper trafficking within host cells or decreased activation of host responses.

On top of the bacterial response to the host, we can now introduce further perturbations of the system by intentionally altering individual bacterial gene functions (infection of the host with bacterial mutants) or individual host gene functions (infection of wild-type bacteria into a host that has been genetically manipulated using RNAi, morpholinos, or where host genes have been disrupted), and measuring bacterial and host responses in the collective interaction. An alternative method is to introduce small molecules into the pathogen–host interaction that may target either host or bacterial functions.

In one recent example, both chemical and RNAi-mediated perturbations of the host cell were used to yield valuable insights into *S. typhimurium* pathogenesis (Kuijl et al., 2007). In this study, Kuijl and colleagues screened a small collection of kinase inhibitors for their ability to restrict the growth of *S. typhimurium* in cultured macrophages. In parallel, they also used automated fluorescent microscopy to perform an siRNA screen targeting the human kinome to identify host kinases required for survival of *S. typhimurium* in cultured macrophages. Both complementary approaches identified that AKT1 is required for proliferation *S. typhimurium* in cultured macrophages. Moreover, the study extended the observation to other intracellular pathogens and found that inhibition of AKT1 also restricts the growth of *M. tuberculosis* in macrophages. Finally, they were able to demonstrate that small molecule inhibition of AKT1 in a mouse model resulted in increased survival of infected animals.

A second example illustrates the ability to identify small molecules that target the bacterial side of the pathogen–host interaction. Automated, high-content microscopy has become a powerful tool to measure the host–pathogen collective interaction as illustrated in a study for small molecules that restrict mycobacterial growth in cultured macrophages (Christophe et al., 2009). This study identified that dinitrobenzamide derivatives effectively restricted the growth of *M. tuberculosis*, including XDR strains, by inhibiting an enzyme involved in arabinan synthesis, decaprenyl-phosphoribose 2' epimerase. This study thus identified an essential gene function required by *M. tuberculosis* for survival within macrophages (as well as in axenic culture) by using high-content imaging of infected macrophages to identify compounds that interfere with mycobacterial cell wall synthesis.

Finally, the previously described developments in alternative animal model organisms that have been invaluable to higher-throughput genetic studies have equal applicability to small molecule screens for modulators that disrupt the pathogen–host interaction. Historically, whole organism screens for anti-infectives conducted in mammalian hosts have been few and far between due to cost and housing restraints and potentially, ethical considerations. One example of such an endeavor is the discovery of ivermectin, which was identified in a herculean screen performed by Merck in the 1980s. In this screen, thousands of natural product extracts were tested in nematode-infected mice to identify compounds with antihelminthic activity. Thus, while whole organism screens for anti-infectives do clearly work, the scale in terms of the number of mammalian hosts one must screen to identify an anti-infective is simply not feasible in an academic setting. However, if one uses an alternative model host, such as *C. elegans* or zebrafish, that is cheaper and where housing restraints are minimal, one can feasibly conduct chemical screens for anti-infectives. Unlike traditional screens for antibiotics, which are designed to identify compounds that kill or inhibit the growth of a bacteria *in vitro*, whole organism screens have the capacity to also identify small molecules that act by inhibiting either bacterial or host functions. An example of a small molecule screen for anti-infectives in an alternative host model was performed (Moy et al., 2006) on *Enterococcus faecalis*-infected *C. elegans*. In this screen, the authors identified several small molecules that rescue infected nematodes, but do not appear to work by the conventional antibiotic paradigm of targeting essential gene functions that are required regardless

of the environmental condition. Some of the compounds that were identified were more active in the infected host than against the pathogen alone in axenic culture, leading the authors to consider whether the compounds might be targeting other bacterial (or even host) gene functions that are involved in virulence or *in vivo* survival, but not required for *in vitro* survival. Thus, the array of phenotypes that can be examined using small molecules grows ever more complex, at the same time that methods to study pathogenesis allow us to study systems that are increasingly relevant and/or better mimic the true conditions of human infection.

Future

The increasing ability to obtain genome-wide descriptions of bacterial responses to increasing numbers of perturbations, both genetic and chemical, has significantly expanded our understanding of bacterial pathogenesis and behavior. With these opportunities however, often comes an accompanying set of new challenges that must be solved in order to continuously propel us forward in our understanding.

Challenges

One of the major challenges that has come with the dramatic increase in genomic data that can now be obtained is the ability to analyze such large amounts of data and to integrate them into coherent models of bacterial response and behavior. Such advances have spawned new fields such as computational biology and have resulted in the need for close collaborations between experimental microbiologists and bioinformaticists. Such interdisciplinary efforts are needed on both the technical level, for example, translating millions of sequencing reads into coherent biological information, as well as the analytical level, for example, translating the varying expression levels of all genes in the genome under a certain condition into relevant biological models of behavior. The unprecedented level of detail and sensitivity with which we can acquire data, for example, in gene expression using RNA sequencing or on single cells will require an additional revolution in our ability to analyze and interpret data.

With the increasing ease with which we can sequence bacterial genomes, we must discard basic assumptions or approximations that we previously took for granted, and create a new framework from which we must consider the concepts of pathogenesis. For example, the concept of a clone and/or the use of standard laboratory pathogenic strains begin to lose their meaning as sequencing now reveals heterogeneity within such strains. One lab's standard strain may in fact differ from another's, with each carrying numerous genetic variations relative to the other. The corollary of this fact is that the data resulting from every individual lab can be the result of strain specific behavior. Certainly, the optimist hopes that nevertheless, many principles that we have learned and continue to learn are conserved across strains and clones.

Recent lessons and caveats to the interpreting of genomic data are highlighted in the relatively unsuccessful GlaxoSmithKline campaign in postgenomic antibiotic discovery (Payne et al., 2007). Mining the increasing number of available genomes for new drug targets, efforts were made to identify genes that were essential across multiple pathogens, including *Streptococcus pneumococcus*, *S. aureus*, and *H. influenzae*, and that were conserved on the amino acid level, present in single copy, and had no host ortholog. Of 350 genes common to all three pathogens, 127 genes were described as being essential *in vitro* in at least one of the organisms. Yet, upon completion of their campaign, they found that many of the genes that had been predicted to be essential were in fact nonessential. Their work calls to attention one of the major caveats in our current ability to designate a gene as "essential"; it can be challenging to unambiguously assign essentiality, since essentiality is typically defined by the inability to isolate a knockout mutant (most often by transposon insertion) in a functional study. Thus one could inappropriately designate a gene as essential, when in fact it is simply statistically difficult to hit using a transposon (e.g. it is a small gene), resulting in a false positive assignment of essentiality. The creation of knockdown strains is one method to more rigorously define essentiality, by linking gene dosage to growth.

Further, their work demonstrated that predictions based on mining genomes can only be as good as the input genomic data. In fact, their efforts demonstrated that the limited number of available annotated genomes for a given pathogen only provide a small and inadequate window into the genetic diversity of the entire genus including all clinical isolates. For example, methionine tRNA synthetase (MetRS) was initially predicted to be a good essential gene candidate, potentially across many pathogens. However, subsequent genomic data and studies revealed that many clinical strains of *S. pneumoniae* carry two different copies of the gene or that alternative gene products can perform the same reaction, thus relegating any one gene product as unessential. Also, the comparative element of such studies is complicated by the fact that different genes can perform different functions in different genomes (i.e. FabI versus FabK), thus demonstrating the limits of purely genomic data without full functional characterization.

Their experience concludes that a lot more sequencing of pathogenic strains and clinical isolates will be necessary in order to understand the extent of genetic diversity and the functional consequences of this diversity for the behavior of a given bacterial pathogen.

Fortunately, this is now possible with current technologies and will be increasingly cheaper and easier with the advent of newer technologies. Thus, it is clear that genomics has both clearly highlighted the problem of strain specificity in pathogenesis as well as provided some of the tools that will be required to tackle the problem.

Opportunities

While in some cases, technical advances have provided a better method for obtaining the same set of data than could be obtained by less efficient technologies (i.e. the use of sequencing rather than microarrays for determining the results of pooled, negative selections), in many cases, technological advances have opened up new approaches that were previously inaccessible or logistically unfeasible. Thus, these new strategies can provide new, complementary insight into bacterial behavior.

As increasing efforts are invested to overcome the challenges of analyzing and interpreting the copious amounts of data that are being produced by experimental genomic studies of pathogenesis, increasing refinement of our current understanding will be possible as these results are reconciled with those obtained from well-established, reductionist ones. Genomics and chemical biology bring to pathogenesis an expansion of approaches that are largely complementary, with the numerous different approaches that are now possible each serving to play a unique and important role by the insight that they provide. For example, correlative approaches are able to illuminate bacterial behavior from infected humans, whereas functional studies are for the most part limited to *in vitro* or animal model studies. In contrast, functional studies provide a much more direct way to identify genes essential for a given phenotype. Similarly, small molecules may be a more facile way to perturb essential gene function, although conventional genetic manipulations may be a more straightforward way to interrogate *in vitro* nonessential virulence genes.

In the future, the limitations of current approaches will be challenged even further as we strive for increasingly comprehensive descriptions of the complex, dynamic behavior of bacteria within the context of the host. Exquisitely sensitive methods for measuring gene expression *in situ* on a genome-wide scale will allow us to understand the interactions between host and pathogen at incredibly high resolution, perhaps even revealing the transcriptional interplay between a single bacterium and the host cell or cells. Such studies might reveal a previously unrecognized level of individuality in bacterial behavior during infection; such individuality has already been proposed to account for bacterial persistence in the face of antibiotics, so it is not a big leap to imagine a role for individuality in the pathogenic process. Even currently, new technologies such as NanoString digital gene expression technology (Geiss et al., 2008) open the door to highly sensitive, rapid, and affordable ways to simultaneously measure the expression of hundreds of genes in very tiny populations of cells, without the need for sophisticated sample preparation techniques, in an unprecedented number of conditions. Single cell genome-wide expression data cannot be far behind.

These are exciting times in biology in general, and there is every reason to expect that our understanding of a wide range of host–pathogen interactions will grow by leaps and bounds during the next decade as new technology pairs with investigator creativity to devise new strategies for dissecting host–pathogen interactions.

Acknowledgement

The authors wish to thank Jonathan Livny and Sarah Stanley for helpful discussions and assistance with preparation of figures.

Declaration of interest

This work was funded by a Pew Scholars Award in Biomedical Sciences (D.T.H.).

References

Andries K, Verhasselt P, Guillemont J, Göhlmann HW, Neefs JM, Winkler H, Van Gestel J, Timmerman P, Zhu M, Lee E, Williams P, de Chaffoy D, Huitric E, Hoffner S, Cambau E, Truffot-Pernot C, Lounis N, Jarlier V. 2005. A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. Science 307:223–227.

Angelichio MJ, Camilli A. 2002. *In vivo* expression technology. Infect Immun 70:6518–6523.

Ansorge WJ. 2009. Next-generation DNA sequencing techniques. N Biotechnol 25:195–203.

Armour CD, Lum PY. 2005. From drug to protein: using yeast genetics for high-throughput target discovery. Curr Opin Chem Biol 9:20–24.

Arnoldo A, Curak J, Kittanakom S, Chevelev I, Lee VT, Sahebol-Amri M, Koscik B, Ljuma L, Roy PJ, Bedalov A, Giaever G, Nislow C, Merrill AR, Merrill RA, Lory S, Stagljar I. 2008. Identification of small molecule inhibitors of *Pseudomonas aeruginosa* exoenzyme S using a yeast phenotypic screen. PLoS Genet 4:e1000005.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2:2006.0008.

Badarinarayana V, Estep PW 3rd, Shendure J, Edwards J, Tavazoie S, Lam F, Church GM. 2001. Selection analyses of insertional mutants using subgenic-resolution arrays. Nat Biotechnol 19:1060–1065.

Balaban N, Cirioni O, Giacometti A, Ghiselli R, Braunstein JB, Silvestri C, Mocchegiani F, Saba V, Scalise G. 2007. Treatment of

Staphylococcus aureus biofilm infection by the quorum-sensing inhibitor RIP. Antimicrob Agents Chemother 51:2226–2229.

Balemans W, Lounis N, Gilissen R, Guillemont J, Simmen K, Andries K, Koul A. 2010. Essentiality of FASII pathway for *Staphylococcus aureus*. Nature 463:E3; discussion E4.

Barczak AK, Hung DT. 2009. Productive steps toward an antimicrobial targeting virulence. Curr Opin Microbiol 12:490–496.

Baron C. 2006. VirB8: a conserved type IV secretion system assembly factor and drug target. Biochem Cell Biol 84:890–899.

Boshoff HI, Myers TG, Copp BR, McNeil MR, Wilson MA, Barry CE 3rd. 2004. The transcriptional responses of *Mycobacterium tuberculosis* to inhibitors of metabolism: novel insights into drug mechanisms of action. J Biol Chem 279:40174–40184.

Brinster S, Lamberet G, Staels B, Trieu-Cuot P, Gruss A, Poyart C. 2009. Type II fatty acid synthesis is not a suitable antibiotic target for Gram-positive pathogens. Nature 458:83–86.

Camacho LR, Ensergueix D, Perez E, Gicquel B, Guilhot C. 1999. Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. Mol Microbiol 34:257–267.

Cameron DE, Urbach JM, Mekalanos JJ. 2008. A defined transposon mutant library and its use in identifying motility genes in *Vibrio cholerae*. Proc Natl Acad Sci USA 105:8736–8741.

Cegelski L, Pinkner JS, Hammer ND, Cusumano CK, Hung CS, Chorell E, Aberg V, Walker JN, Seed PC, Almqvist F, Chapman MR, Hultgren SJ. 2009. Small-molecule inhibitors target *Escherichia coli* amyloid biogenesis and biofilm formation. Nat Chem Biol 5:913–919.

Chan K, Knaak T, Satkamp L, Humbert O, Falkow S, Ramakrishnan L. 2002. Complex pattern of *Mycobacterium marinum* gene expression during long-term granulomatous infection. Proc Natl Acad Sci USA 99:3920–3925.

Chiang SL, Mekalanos JJ. 1998. Use of signature-tagged transposon mutagenesis to identify *Vibrio cholerae* genes critical for colonization. Mol Microbiol 27:797–805.

Christophe T, Jackson M, Jeon HK, Fenistein D, Contreras-Dominguez M, Kim J, Genovesio A, Carralot JP, Ewann F, Kim EH, Lee SY, Kang S, Seo MJ, Park EJ, Skovierová H, Pham H, Riccardi G, Nam JY, Marsollier L, Kempf M, Joly-Guillou ML, Oh T, Shin WK, No Z, Nehrbass U, Brosch R, Cole ST, Brodin P. 2009. High content screening identifies decaprenyl-phosphoribose 2' epimerase as a target for intracellular antimycobacterial inhibitors. PLoS Pathog 5:e1000645.

Clatworthy AE, Lee JS, Leibman M, Kostun Z, Davidson AJ, Hung DT. 2009. *Pseudomonas aeruginosa* infection of zebrafish involves both host and pathogen determinants. Infect Immun 77:1293–1303.

Clatworthy AE, Pierson E, Hung DT. 2007. Targeting virulence: a new paradigm for antimicrobial therapy. Nat Chem Biol 3:541–548.

Cosma CL, Swaim LE, Volkman H, Ramakrishnan L, Davis JM. 2006. Zebrafish and frog models of *Mycobacterium marinum* infection. Curr Protoc Microbiol Chapter 10:Unit 10B.2.

Coulter SN, Schwan WR, Ng EY, Langhorne MH, Ritchie HD, Westbrock-Wadman S, Hufnagle WO, Folger KR, Bayer AS, Stover CK. 1998. *Staphylococcus aureus* genetic loci impacting growth and survival in multiple infection environments. Mol Microbiol 30:393–404.

Cox JS, Chen B, McNeil M, Jacobs WR Jr. 1999. Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice. Nature 402:79–83.

Cummings RT, Salowe SP, Cunningham BR, Wiltsie J, Park YW, Sonatore LM, Wisniewski D, Douglas CM, Hermes JD, Scolnick EM. 2002. A peptide-based fluorescence resonance energy transfer assay for *Bacillus anthracis* lethal factor protease. Proc Natl Acad Sci USA 99:6603–6606.

Darwin AJ, Miller VL. 1999. Identification of *Yersinia enterocolitica* genes affecting survival in an animal host using signature-tagged transposon mutagenesis. Mol Microbiol 32:51–62.

Davis JM, Clay H, Lewis JL, Ghori N, Herbomel P, Ramakrishnan L. 2002. Real-time visualization of mycobacterium–macrophage interactions leading to initiation of granuloma formation in zebrafish embryos. Immunity 17:693–702.

DiRita VJ, Parsot C, Jander G, Mekalanos JJ. 1991. Regulatory cascade controls virulence in *Vibrio cholerae*. Proc Natl Acad Sci USA 88:5403–5407.

Dubail I, Berche P, Charbit A. 2000. Listeriolysin O as a reporter to identify constitutive and *in vivo*-inducible promoters in the pathogen *Listeria monocytogenes*. Infect Immun 68:3242–3250.

Edelstein PH, Edelstein MA, Higa F, Falkow S. 1999. Discovery of virulence genes of *Legionella pneumophila* by using signature tagged mutagenesis in a guinea pig pneumonia model. Proc Natl Acad Sci USA 96:8190–8195.

Eriksson S, Lucchini S, Thompson A, Rhen M, Hinton JC. 2003. Unravelling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica*. Mol Microbiol 47:103–118.

Escaich S. 2008. Antivirulence as a new antibacterial approach for chemotherapy. Curr Opin Chem Biol 12:400–408.

Forsyth RA, Haselbeck RJ, Ohlsen KL, Yamamoto RT, Xu H, Trawick JD, Wall D, Wang L, Brown-Driver V, Froelich JM, C KG, King P, McCarthy M, Malone C, Misiner B, Robbins D, Tan Z, Zhu Zy ZY, Carr G, Mosca DA, Zamudio C, Foulkes JG, Zyskind JW. 2002. A genome-wide strategy for the identification of essential genes in *Staphylococcus aureus*. Mol Microbiol 43:1387–1400.

Gahan CG, Hill C. 2000. The use of listeriolysin to identify *in vivo* induced genes in the Gram-positive intracellular pathogen *Listeria monocytogenes*. Mol Microbiol 36:498–507.

Gallagher LA, Ramage E, Jacobs MA, Kaul R, Brittnacher M, Manoil C. 2007. A comprehensive transposon mutant library of *Francisella novicida*, a bioweapon surrogate. Proc Natl Acad Sci USA 104:1009–1014.

Garton NJ, Waddell SJ, Sherratt AL, Lee SM, Smith RJ, Senner C, Hinds J, Rajakumar K, Adegbola RA, Besra GS, Butcher PD, Barer MR. 2008. Cytological and transcript analyses reveal fat and lazy persister-like bacilli in tuberculous sputum. PLoS Med 5:e75.

Gawronski JD, Wong SM, Giannoukos G, Ward DV, Akerley BJ. 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.

Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, Fell HP, Ferree S, George RD, Grogan T, James JJ, Maysuria M, Mitton JD, Oliveri P, Osborn JL, Peng T, Ratcliffe AL, Webster PJ, Davidson EH, Hood L, Dimitrov K. 2008. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol 26:317–325.

Giaever G, Shoemaker DD, Jones TW, Liang H, Winzeler EA, Astromoff A, Davis RW. 1999. Genomic profiling of drug sensitivities via induced haploinsufficiency. Nat Genet 21:278–283.

Güell M, van Noort V, Yus E, Chen WH, Leigh-Bell J, Michalodimitrakis K, Yamada T, Arumugam M, Doerks T, Kühner S, Rode M, Suyama M, Schmidt S, Gavin AC, Bork P, Serrano L. 2009. Transcriptome complexity in a genome-reduced bacterium. Science 326:1268–1271.

Heithoff DM, Conner CP, Hanna PC, Julio SM, Hentschel U, Mahan MJ. 1997. Bacterial infection as assessed by *in vivo* gene expression. Proc Natl Acad Sci USA 94:934–939.

Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E, Holden DW. 1995. Simultaneous identification of bacterial virulence genes by negative selection. Science 269:400–403.

Hinton JC, Hautefort I, Eriksson S, Thompson A, Rhen M. 2004. Benefits and pitfalls of using microarrays to monitor bacterial gene expression during infection. Curr Opin Microbiol 7:277–282.

Hisert KB, MacCoss M, Shiloh MU, Darwin KH, Singh S, Jones RA, Ehrt S, Zhang Z, Gaffney BL, Gandotra S, Holden DW, Murray D, Nathan C. 2005. A glutamate–alanine–leucine (EAL) domain protein of *Salmonella* controls bacterial survival in mice, antioxidant defence and killing of macrophages: role of cyclic diGMP. Mol Microbiol 56:1234–1245.

Hsiao A, Zhu J. 2009. Genetic tools to study gene expression during bacterial pathogen infection. Adv Appl Microbiol 67:297–314.

Huang J, Lesser CF, Lory S. 2008. The essential role of the CopN protein in *Chlamydia pneumoniae* intracellular growth. Nature 456:112–115.

Hung DT, Shakhnovich EA, Pierson E, Mekalanos JJ. 2005. Small-molecule inhibitor of *Vibrio cholerae* virulence and intestinal colonization. Science 310:670–674.

Jacobs MA, Alwood A, Thaipisuttikul I, Spencer D, Haugen E, Ernst S, Will O, Kaul R, Raymond C, Levy R, Chun-Rong L, Guenthner D, Bovee D, Olson MV, Manoil C. 2003. Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. Proc Natl Acad Sci USA 100:14339–14344.

Joshi SM, Pandey AK, Capite N, Fortune SM, Rubin EJ, Sassetti CM. 2006. Characterization of mycobacterial virulence genes through genetic interaction mapping. Proc Natl Acad Sci USA 103:11760–11765.

Judson N, Mekalanos JJ. 2000a. TnAraOut, a transposon-based approach to identify and characterize essential bacterial genes. Nat Biotechnol 18:740–745.

Judson N, Mekalanos JJ. 2000b. Transposon-based approaches to identify essential bacterial genes. Trends Microbiol 8:521–526.

Junker LM, Clardy J. 2007. High-throughput screens for small-molecule inhibitors of *Pseudomonas aeruginosa* biofilm development. Antimicrob Agents Chemother 51:3582–3590.

Kauppi AM, Nordfelth R, Uvell H, Wolf-Watz H, Elofsson M. 2003. Targeting bacterial virulence: inhibitors of type III secretion in *Yersinia*. Chem Biol 10:241–249.

Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, Asai K, Ashikaga S, Aymerich S, Bessieres P, Boland F, Brignell SC, Bron S, Bunai K, Chapuis J, Christiansen LC, Danchin A, Débarbouille M, Dervyn E, Deuerling E, Devine K, Devine SK, Dreesen O, Errington J, Fillinger S, Foster SJ, Fujita Y, Galizzi A, Gardan R, Eschevins C, Fukushima T, Haga K, Harwood CR, Hecker M, Hosoya D, Hullo MF, Kakeshita H, Karamata D, Kasahara Y, Kawamura F, Koga K, Koski P, Kuwana R, Imamura D, Ishimaru M, Ishikawa S, Ishio I, Le Coq D, Masson A, Mauël C, Meima R, Mellado RP, Moir A, Moriya S, Nagakawa E, Nanamiya H, Nakai S, Nygaard P, Ogura M, Ohanan T, O'Reilly M, O'Rourke M, Pragai Z, Pooley HM, Rapoport G, Rawlins JP, Rivas LA, Rivolta C, Sadaie A, Sadaie Y, Sarvas M, Sato T, Saxild HH, Scanlan E, Schumann W, Seegers JF, Sekiguchi J, Sekowska A, Séror SJ, Simon M, Stragier P, Studer R, Takamatsu H, Tanaka T, Takeuchi M, Thomaides HB, Vagner V, van Dijl JM, Watabe K, Wipat A, Yamamoto H, Yamamoto M, Yamamoto Y, Yamane K, Yata K, Yoshida K, Yoshikawa H, Zuber U, Ogasawara N. 2003. Essential *Bacillus subtilis* genes. Proc Natl Acad Sci USA 100:4678–4683.

Koul A, Dendouga N, Vergauwen K, Molenberghs B, Vranckx L, Willebrords R, Ristic Z, Lill H, Dorange I, Guillemont J, Bald D, Andries K. 2007. Diarylquinolines target subunit c of mycobacterial ATP synthase. Nat Chem Biol 3:323–324.

Koul A, Vranckx L, Dendouga N, Balemans W, Van den Wyngaert I, Vergauwen K, Göhlmann HW, Willebrords R, Poncelet A, Guillemont J, Bald D, Andries K. 2008. Diarylquinolines are bactericidal for dormant mycobacteria as a result of disturbed ATP homeostasis. J Biol Chem 283:25273–25280.

Kuijl C, Savage ND, Marsman M, Tuin AW, Janssen L, Egan DA, Ketema M, van den Nieuwendijk R, van den Eeden SJ, Geluk A, Poot A, van der Marel G, Beijersbergen RL, Overkleeft H, Ottenhoff TH, Neefjes J. 2007. Intracellular bacterial growth is controlled by a kinase network around PKB/AKT1. Nature 450:725–730.

Lamichhane G, Tyagi S, Bishai WR. 2005. Designer arrays for defined mutant analysis to detect genes essential for survival of *Mycobacterium tuberculosis* in mouse lungs. Infect Immun 73:2533–2540.

Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK. 2009. Simultaneous assay of every *Salmonella typhi* gene using one million transposon mutants. Genome Res 19:2308–2316.

Larocque RC, Harris JB, Dziejman M, Li X, Khan AI, Faruque AS, Faruque SM, Nair GB, Ryan ET, Qadri F, Mekalanos JJ, Calderwood SB. 2005. Transcriptional profiling of *Vibrio cholerae* recovered directly from patient specimens during early and late stages of human infection. Infect Immun 73:4488–4493.

Lee S, Hinz A, Bauerle E, Angermeyer A, Juhaszova K, Kaneko Y, Singh PK, Manoil C. 2009. Targeting a bacterial stress response to enhance antibiotic action. Proc Natl Acad Sci USA 106:14570–14575.

Lee SH, Hava DL, Waldor MK, Camilli A. 1999. Regulation and temporal expression patterns of *Vibrio cholerae* virulence genes during infection. Cell 99:625–634.

Liberati NT, Urbach JM, Miyata S, Lee DG, Drenkard E, Wu G, Villanueva J, Wei T, Ausubel FM. 2006. An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. Proc Natl Acad Sci USA 103:2833–2838.

Livny J, Friedman DI. 2004. Characterizing spontaneous induction of Stx encoding phages using a selectable reporter system. Mol Microbiol 51:1691–1704.

Lombardo MJ, Michalski J, Martinez-Wilson H, Morin C, Hilton T, Osorio CG, Nataro JP, Tacket CO, Camilli A, Kaper JB. 2007. An *in vivo* expression technology screen for *Vibrio cholerae* genes expressed in human volunteers. Proc Natl Acad Sci USA 104:18229–18234.

Lucchini S, Liu H, Jin Q, Hinton JC, Yu J. 2005. Transcriptional adaptation of *Shigella flexneri* during infection of macrophages and epithelial cells: insights into the strategies of a cytosolic bacterial pathogen. Infect Immun 73:88–102.

Mahajan-Miklos S, Rahme LG, Ausubel FM. 2000. Elucidating the molecular mechanisms of bacterial virulence using non-mammalian hosts. Mol Microbiol 37:981–988.

Mahan MJ, Slauch JM, Mekalanos JJ. 1993. Selection of bacterial virulence genes that are specifically induced in host tissues. Science 259:686–688.

Mahan MJ, Tobias JW, Slauch JM, Hanna PC, Collier RJ, Mekalanos JJ. 1995. Antibiotic-based selection for bacterial genes that are specifically induced during infection of a host. Proc Natl Acad Sci USA 92:669–673.

Mei JM, Nourbakhsh F, Ford CW, Holden DW. 1997. Identification of *Staphylococcus aureus* virulence genes in a murine model of bacteraemia using signature-tagged mutagenesis. Mol Microbiol 26:399–407.

Merrell DS, Butler SM, Qadri F, Dolganov NA, Alam A, Cohen MB, Calderwood SB, Schoolnik GK, Camilli A. 2002a. Host-induced epidemic spread of the cholera bacterium. Nature 417:642–645.

Merrell DS, Hava DL, Camilli A. 2002b. Identification of novel factors involved in colonization and acid tolerance of *Vibrio cholerae*. Mol Microbiol 43:1471–1491.

Miller VL. 1999. Signature-tagged mutagenesis and the hunt for virulence factors: response. Trends Microbiol 7:388.

Miller VL, Mekalanos JJ. 1984. Synthesis of cholera toxin is positively regulated at the transcriptional level by toxR. Proc Natl Acad Sci USA 81:3471–3475.

Mockler TC, Chan S, Sundaresan A, Chen H, Jacobsen SE, Ecker JR. 2005. Applications of DNA tiling arrays for whole-genome analysis. Genomics 85:1–15.

Mohr S, Bakal C, Perrimon N. 2010. Genomic screening with RNAi: results and challenges. Annu Rev Biochem 79:37–64.

Moy TI, Ball AR, Anklesaria Z, Casadei G, Lewis K, Ausubel FM. 2006. Identification of novel antimicrobials using a live-animal infection model. Proc Natl Acad Sci USA 103:10414–10419.

Müsken M, Di Fiore S, Dötsch A, Fischer R, Häussler S. 2010. Genetic determinants of *Pseudomonas aeruginosa* biofilm establishment. Microbiology (Reading, Engl) 156:431–441.

Nusbaum C, Ohsumi TK, Gomez J, Aquadro J, Victor TC, Warren RM, Hung DT, Birren BW, Lander ES, Jaffe DB. 2009. Sensitive, specific polymorphism discovery in bacteria using massively parallel sequencing. Nat Methods 6:67–69.

Osorio CG, Crawford JA, Michalski J, Martinez-Wilson H, Kaper JB, Camilli A. 2005. Second-generation recombination-based *in vivo* expression technology for large-scale screening for *Vibrio cholerae* genes induced during infection of the mouse small intestine. Infect Immun 73:972–980.

Paik S, Senty L, Das S, Noe JC, Munro CL, Kitten T. 2005. Identification of virulence determinants for endocarditis in *Streptococcus sanguinis* by signature-tagged mutagenesis. Infect Immun 73:6064–6074.

Pathania R, Zlitni S, Barker C, Das R, Gerritsma DA, Lebert J, Awuah E, Melacini G, Capretta FA, Brown ED. 2009. Chemical genomics in *Escherichia coli* identifies an inhibitor of bacterial lipoprotein targeting. Nat Chem Biol 5:849–856.

Payne DJ, Gwynn MN, Holmes DJ, Pompliano DL. 2007. Drugs for bad bugs: confronting the challenges of antibacterial discovery. Nat Rev Drug Discov 6:29–40.

Perkins TT, Kingsley RA, Fookes MC, Gardner PP, James KD, Yu L, Assefa SA, He M, Croucher NJ, Pickard DJ, Maskell DJ, Parkhill J, Choudhary J, Thomson NR, Dougan G. 2009. A strand-specific RNA-seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. PLoS Genet 5:e1000569.

Perry RD. 1999. Signature-tagged mutagenesis and the hunt for virulence factors. Trends Microbiol 7:385–388; discussion 388.

Phelps HA, Runft DL, Neely MN. 2009. Adult zebrafish model of streptococcal infection. Curr Protoc Microbiol Chapter 9:Unit 9D.1.

Polissi A, Pontiggia A, Feger G, Altieri M, Mottl H, Ferrari L, Simon D. 1998. Large-scale identification of virulence genes from *Streptococcus pneumoniae*. Infect Immun 66:5620–5629.

Prajsnar TK, Cunliffe VT, Foster SJ, Renshaw SA. 2008. A novel vertebrate model of *Staphylococcus aureus* infection reveals phagocyte-dependent resistance of zebrafish to non-host specialized pathogens. Cell Microbiol 10:2312–2325.

Rachman H, Strong M, Ulrichs T, Grode L, Schuchhardt J, Mollenkopf H, Kosmiadi GA, Eisenberg D, Kaufmann SH. 2006. Unique transcriptome signature of *Mycobacterium tuberculosis* in pulmonary tuberculosis. Infect Immun 74:1233–1242.

Ramakrishnan L, Federspiel NA, Falkow S. 2000. Granuloma-specific expression of *Mycobacterium* virulence proteins from the glycine-rich PE-PGRS family. Science 288:1436–1439.

Raskin DM, Seshadri R, Pukatzki SU, Mekalanos JJ. 2006. Bacterial genomics and pathogen evolution. Cell 124:703–714.

Rasko DA, Moreira CG, Li de R, Reading NC, Ritchie JM, Waldor MK, Williams N, Taussig R, Wei S, Roth M, Hughes DT, Huntley JF, Fina MW, Falck JR, Sperandio V. 2008. Targeting QseC signaling and virulence for antibiotic development. Science 321:1078–1080.

Rasko DA, Sperandio V. 2010. Anti-virulence strategies to combat bacteria-mediated disease. Nat Rev Drug Discov 9:117–128.

Rediers H, Rainey PB, Vanderleyden J, De Mot R. 2005. Unraveling the secret lives of bacteria: use of *in vivo* expression technology and differential fluorescence induction promoter traps as tools for exploring niche-specific gene expression. Microbiol Mol Biol Rev 69:217–261.

Rengarajan J, Bloom BR, Rubin EJ. 2005. Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. Proc Natl Acad Sci USA 102:8327–8332.

Salama NR, Shepherd B, Falkow S. 2004. Global transposon mutagenesis and essential gene analysis of *Helicobacter pylori*. J Bacteriol 186:7926–7935.

Sassetti CM, Boyd DH, Rubin EJ. 2001. Comprehensive identification of conditionally essential genes in mycobacteria. Proc Natl Acad Sci USA 98:12712–12717.

Sassetti CM, Rubin EJ. 2003. Genetic requirements for mycobacterial survival during infection. Proc Natl Acad Sci USA 100:12989–12994.

Schild S, Tamayo R, Nelson EJ, Qadri F, Calderwood SB, Camilli A. 2007. Genes induced late in infection increase fitness of *Vibrio cholerae* after release into the environment. Cell Host Microbe 2:264–277.

Schnappinger D, Ehrt S, Voskuil MI, Liu Y, Mangan JA, Monahan IM, Dolganov G, Efron B, Butcher PD, Nathan C, Schoolnik GK. 2003. Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: insights into the phagosomal environment. J Exp Med 198:693–704.

Schwan WR, Coulter SN, Ng EY, Langhorne MH, Ritchie HD, Brody LL, Westbrock-Wadman S, Bayer AS, Folger KR, Stover CK. 1998. Identification and characterization of the PutP proline permease that contributes to *in vivo* survival of *Staphylococcus aureus* in animal models. Infect Immun 66:567–572.

Shakhnovich EA, Hung DT, Pierson E, Lee K, Mekalanos JJ. 2007. Virstatin inhibits dimerization of the transcriptional activator ToxT. Proc Natl Acad Sci USA 104:2372–2377.

Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R, Stadler PF, Vogel J. 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature 464:250–255.

Shoop WL, Xiong Y, Wiltsie J, Woods A, Guo J, Pivnichny JV, Felcetto T, Michael BF, Bansal A, Cummings RT, Cunningham BR, Friedlander AM, Douglas CM, Patel SB, Wisniewski D, Scapin G, Salowe SP, Zaller DM, Chapman KT, Scolnick EM, Schmatz DM, Bartizal K, MacCoss M, Hermes JD. 2005. Anthrax lethal factor inhibition. Proc Natl Acad Sci USA 102:7958–7963.

Sifri CD, Begun J, Ausubel FM, Calderwood SB. 2003. *Caenorhabditis elegans* as a model host for *Staphylococcus aureus* pathogenesis. Infect Immun 71:2208–2217.

Slauch JM, Camilli A. 2000. IVET and RIVET: use of gene fusions to identify bacterial virulence factors specifically induced in host tissues. Meth Enzymol 326:73–96.

Sorek R, Cossart P. 2010. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. Nat Rev Genet 11:9–16.

Stanley SA, Hung DT. 2009. Chemical tools for dissecting bacterial physiology and virulence. Biochemistry 48:8776–8786.

Staudinger BJ, Oberdoerster MA, Lewis PJ, Rosen H. 2002. mRNA expression profiles for *Escherichia coli* ingested by normal and phagocyte oxidase-deficient human neutrophils. J Clin Invest 110:1151–1163.

Stockwell BR. 2000. Chemical genetics: ligand-based discovery of gene function. Nat Rev Genet 1:116–125.

Talaat AM, Lyons R, Howard ST, Johnston SA. 2004. The temporal expression profile of *Mycobacterium tuberculosis* infection in mice. Proc Natl Acad Sci USA 101:4602–4607.

Tan MW, Ausubel FM. 2000. *Caenorhabditis elegans*: a model genetic host to study *Pseudomonas aeruginosa* pathogenesis. Curr Opin Microbiol 3:29–34.

Taylor RK, Manoil C, Mekalanos JJ. 1989. Broad-host-range vectors for delivery of TnphoA: use in genetic analysis of secreted virulence determinants of *Vibrio cholerae*. J Bacteriol 171:1870–1878.

Timm J, Post FA, Bekker LG, Walther GB, Wainwright HC, Manganelli R, Chan WT, Tsenova L, Gold B, Smith I, Kaplan G, McKinney JD. 2003. Differential expression of iron-, carbon-, and oxygen-responsive mycobacterial genes in the lungs of chronically infected mice and tuberculosis patients. Proc Natl Acad Sci USA 100:14321–14326.

Valdivia RH, Falkow S. 1996. Bacterial genetics by flow cytometry: rapid isolation of *Salmonella typhimurium* acid-inducible promoters by differential fluorescence induction. Mol Microbiol 22:367–378.

Valdivia RH, Falkow S. 1997. Fluorescence-based isolation of bacterial genes expressed within host cells. Science 277:2007–2011.

van der Sar AM, Appelmelk BJ, Vandenbroucke-Grauls CM, Bitter W. 2004. A star with stripes: zebrafish as an infection model. Trends Microbiol 12:451–457.

van der Sar AM, Musters RJ, van Eeden FJ, Appelmelk BJ, Vandenbroucke-Grauls CM, Bitter W. 2003. Zebrafish embryos as a model host for the real time analysis of *Salmonella typhimurium* infections. Cell Microbiol 5:601–611.

van Opijnen T, Bodi KL, Camilli A. 2009. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat Methods 6:767–772.

van Vliet AH. 2010. Next generation sequencing of microbial transcriptomes: challenges and opportunities. FEMS Microbiol Lett 302:1–7.

Voskuil MI, Schnappinger D, Visconti KC, Harrell MI, Dolganov GM, Sherman DR, Schoolnik GK. 2003. Inhibition of respiration by nitric oxide induces a *Mycobacterium tuberculosis* dormancy program. J Exp Med 198:705–713.

Waddell SJ, Butcher PD, Stoker NG. 2007. RNA profiling in host–pathogen interactions. Curr Opin Microbiol 10:297–302.

Wang J, Mushegian A, Lory S, Jin S. 1996. Large-scale isolation of candidate virulence genes of *Pseudomonas aeruginosa* by *in vivo* selection. Proc Natl Acad Sci USA 93:10434–10439.

Wang J, Soisson SM, Young K, Shoop W, Kodali S, Galgoci A, Painter R, Parthasarathy G, Tang YS, Cummings R, Ha S, Dorso K, Motyl M, Jayasuriya H, Ondeyka J, Herath K, Zhang C, Hernandez L, Allocco J, Basilio A, Tormo JR, Genilloud O, Vicente F, Pelaez F, Colwell L, Lee SH, Michael B, Felcetto T, Gill C, Silver LL, Hermes JD, Bartizal K, Barrett J, Schmatz D, Becker JW, Cully D, Singh SB. 2006. Platensimycin is a selective FabF inhibitor with potent antibiotic properties. Nature 441:358–361.

Waters LS, Storz G. 2009. Regulatory RNAs in bacteria. Cell 136:615–628.

Wilson M, DeRisi J, Kristensen HH, Imboden P, Rane S, Brown PO, Schoolnik GK. 1999. Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. Proc Natl Acad Sci USA 96:12833–12838.

Wong SM, Akerley BJ. 2008. Identification and analysis of essential genes in *Haemophilus influenzae*. Methods Mol Biol 416:27–44.

Xu HH, Trawick JD, Haselbeck RJ, Forsyth RA, Yamamoto RT, Archer R, Patterson J, Allen M, Froelich JM, Taylor I, Nakaji D, Maile R, Kedar GC, Pilcher M, Brown-Driver V, McCarthy M, Files A, Robbins D, King P, Sillaots S, Malone C, Zamudio CS, Roemer T, Wang L, Youngman PJ, Wall D. 2010. *Staphylococcus aureus* TargetArray: comprehensive differential essential gene expression as a mechanistic tool to profile antibacterials. Antimicrob Agents Chemother 54:3659–3670.

Xu Q, Dziejman M, Mekalanos JJ. 2003. Determination of the transcriptome of *Vibrio cholerae* during intraintestinal growth and midexponential phase *in vitro*. Proc Natl Acad Sci USA 100:1286–1291.

Yoon H, McDermott JE, Porwollik S, McClelland M, Heffron F. 2009. Coordinated regulation of virulence during systemic infection of *Salmonella enterica* serovar *typhimurium*. PLoS Pathog 5:e1000306.

Editor: Michael M. Cox