Microbial Pathogenesis: Genomics and Beyond

Evelyn J. Strauss* and Stanley Falkow

The growing number of complete microbial genome sequences provides a powerful tool for studying the biology of microorganisms. In combination with assays for function, genomic-based approaches can facilitate efficient and directed research strategies to elucidate mechanisms of bacterial pathogenicity. As genomic information accrues, the challenge remains to construct a picture of the biology that accurately reflects how individual genes collaborate to create the complex world of microbial specialization.

Entering the Genomic Era— New Challenges

For many biologists, the late 1990s mark the beginning of the era of genomics. For the first time in history, we have access to the entire genetic content of a growing number and variety of living creatures. This information has the potential to open innovative and efficient research avenues. However, homology is not function, and the burden for the biologist is to translate inanimate DNA sequences into cellular activities. The availability of extensive genomic information frees us to develop new genetic and molecular approaches by which to examine microbial populations in host organisms and the environment.

In this article, we discuss some of the promises and limitations of using genomics—the combination of complete genome sequences and the informatic tools with which to analyze them—to study bacterial pathogenicity. We review some of the strategies that can be used to exploit genomics, and we point out caveats of various approaches. We also survey several techniques that are complementary to genomics, allowing direct examination of the genetic basis of virulence. Finally, we discuss challenges of the future and how genomics may help us surmount them. Although our focus is pathogenesis, many of the concepts we discuss can be applied to other areas of microbiology.

Providing More Clues and Saving Time

With the aid of a database, a small amount of DNA sequence can provide a view of an entire gene and its neighbors in minutes.

The authors are in the Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305–5402, USA.

*To whom correspondence should be addressed. E-mail: estrauss@cmgm.stanford.edu

This can assist in the immediate design of experiments to deduce gene function; the sequence may suggest possible biochemical activities, implied by the presence of motifs or homology to genes with known activities. Conversely, if there is interest in a particular function that is suggested by a conserved amino acid or nucleotide sequence, candidate genes can be identified simply by searching the complete sequence of the organism of interest. These motif-based approaches can provide an intellectual framework when information about biological function is sought.

General information about a microbe can also be derived from examining genome sequences. It is now possible to estimate how much of an organism's genomic content is invested in different biosynthetic or metabolic processes. This can yield insight into the organism's activities, if one assumes that the space allotted to a given function correlates with its importance to the organism. For example, the streamlined genome of Mycoplasma genitalium commits almost 5% of its content to encoding an adhesin gene and related sequences (1); it is thought that the organism uses these sequences to produce antigenic variants of the protein in order to evade the immune response. What is absent from a genome can be informative as well. For example, Haemophilus influenzae appears to be missing three of the enzymes needed for the tricarboxylic acid cycle (2), a finding that is surprising but might yield insight into key features of the metabolic strategy that this fastidious organism uses to survive exclusively in humans.

Using Genomics to Identify Genes That Are Important for Virulence

The complete H. influenzae sequence was finished in 1995 (2), and the potential of genomics is already being realized. The se-

quence has been used to identify candidate virulence genes and components of the biosynthetic machinery for lipopolysaccharide (LPS) (3, 4), a molecule known for decades to function in pathogenesis. LPS is critical in maintaining the integrity and function of the outer cell membrane of Gram negative bacteria and is an essential component of a successful infection. However, LPS is such a potent activator of the immune system that the response against it can have devastating consequences for the host as well as the microorganism.

Many pathogens, including M. genitalium and H. influenzae, are able to confound the immune system by varying the expression of bacterial envelope molecules. This process, called phase variation, can be mediated by the loss or gain of one or more nucleotide repeating units, which alters transcription signals or changes the reading frame of genes encoding cell surface molecules. The Haemophilus genome was searched and nine novel loci with multiple tandem tetranucleotide repeats were identified (3). Several of these genes encoded homologs of bacterial envelope factors, and another (lgtC) was shown to be involved in the phenotypic switching of an LPS epitope. Moreover, an H. influenzae strain carrying a disruption in lgtC is less virulent than the parental strain in an animal model of disease.

Despite long familiarity with the importance of LPS, many aspects of its biosynthetic pathway in H. influenzae have remained elusive. The Haemophilus genome was searched for DNA and amino acid similarity to genes known to be involved in LPS production in other organisms (4). Targeted gene disruptions, immunochemical techniques, tricine-T–SDS–polyacrylamide gel electrophoresis, and mass spectrometry were used to examine potential roles in LPS biosynthesis for the 25 genes identified, and strains carrying mutations in many of these genes were shown to have altered LPS structures. The analysis has already provided information of direct relevance to pathogenesis: The behavior of strains containing mutations in these genes allowed predictions of the minimal LPS structure required for efficient intravascular dissemination of H. influenzae in the infant rat.

There is a more general lesson here as well. In the case of LPS biosynthesis, it is

not uncommon for proteins of related function to be encoded by genes of homologous amino acid, but divergent nucleotide, sequence. Of the genes that were described, 60% could not have been reliably identified by DNA sequence alone and, therefore, would not have been found by nucleic acid hybridization experiments, with the use of either conventional blotting techniques or a computer.

Using Clues About the Essentials of Life to Study Microbial Pathogenesis

Genomics has the potential for informing us about the minimal gene set required for cellular life. The first two publicly available whole bacterial genomes (H. influenzae and M. genitalium) were compared (5), and a catalog of genes conserved in both organisms was compiled, adding genes to fill in the missing steps of critical metabolic pathways that are encoded by dissimilar genes in the two bacteria and subtracting apparently redundant and organism-specific genes. Because such a list may represent a minimal metabolic core of reactions for all cells (or at least for cells that live in similar environments), subtracting it from a test organism's total genetic inventory should reveal factors that specify the phenotypic characters that make the organism unique. If this exercise is carried out on a pathogen, the resulting gene set should include traits required for virulence.

Many bacterial virulence genes are found as discrete segments, present in pathogenic organisms but absent from nonpathogenic members of the same genus or species (6). These pathogenicity islands often display distinct codon usage and a different overall base composition from the core chromosomal elements, suggesting that they were acquired from a foreign source. These genetic features provide possible criteria for identifying potential virulence genes when scanning a complete chromosomal sequence.

Although pathogenicity is often multifactorial, the inheritance or alteration of a single virulence factor can profoundly change the infectious process that a pathogen causes in its host. For example, infection by enteropathogenic Escherichia coli (EPEC) is characterized by a watery, often chronic, diarrheal disease. A distinct E. coli strain, enterohemorrhagic E. coli (EHEC) causes bloody colitis and sometimes hemolytic uremia syndrome. Yet, EHEC shares many of the major virulence genes of EPEC. In addition, EHEC carries a Shiga-like toxin that is not present in EPEC; it has been proposed that this mol-

ecule causes the EHEC-specific symptoms (7). Indeed, when Shiga-like toxin I was expressed in the rabbit pathogen, rabbit diarrheal E. coli-1 (RDEC-1), the bacterium caused an illness in rabbits that closely resembles an EHEC-like colitis rather than the diarrheal disease normally seen with RDEC-1 (8). In this case, the presence or absence of the toxin defined these strains before elucidation of the relevant genetic determinant. However, if the entire sequences of both organisms had been complete before the initiation of these studies, the significance of the Shiga-like toxin I gene might have been deduced by employing subtractive hybridization with computer informatics.

Sequence not only provides information useful for distinguishing closely related sets of organisms, but it also raises some broad questions about the nature of pathogenicity. Although pathogenesis is rare in general (when one considers the low fraction of extant organisms that are virulent), ample numbers of pathogenic eukaryotes and prokaryotes exist. In contrast, there are no known members of the archaea kingdom that are virulent, despite the fact that some archaeans share our environment and are part of our normal bowel flora. We do not

Table 1. Genome sequencing projects cover a wide range of organisms [courtesy of C. Venter, The Institute for Genome Research (TIGR)]. Updated information can be found at the TIGR worldwide web site, www.tigr.org   HGE indicates human granulocytic ehrlichiosis.

Organism	Organization	Size (Mb)
Bacteria		
Actinobacillus actinomycetemcomitans	University of Oklahoma	2.20
Aquifex aeolicus	Recombinant BioCatalysis Inc.	1.50
Bacillus subtilis	European and Japanese consortium	4.20
Borrelia burgdorferi*	TIGR	1.30
Caulobacter crescentus	TIGR	3.80
Deinococcus radiodurans	TIGR	3.00
Ehrlichia species (HGE agent)	TIGR	1.40
Enterococcus faecalis†	TIGR	3.00
Escherichia coli*	University of Wisconsin	4.60
Haemophilus influenzae‡	TIGR	1.83
Helicobacter pylori*	TIGR	1.66
Legionella pneumophila	TIGR	4.10
Mycobacterium avium	TIGR	4.70
Mycobacterium tuberculosis	TIGR (clinical isolate)	4.40
Mycobacterium tuberculosis	Sanger Center (lab strain)	4.40
Mycoplasma genitalium‡	TIGR	0.58
Mycoplasma pneumoniae‡	University of Heidelberg	0.81
Neisseria gonorrhoeae	University of Oklahoma	2.20
Neisseria meningitidis	TIGR and Oxford University	2.30
Porphyromonas gingivalis	TIGR	2.20
Rickettsia prowazekii	University of Uppsala	1.10
Salmonella typhimurium†	TIGR	4.50
Shewanella putrefaciens†	TIGR	4.50
Streptococcus pneumoniae	TIGR	2.20
Streptococcus pyogenes	University of Oklahoma	1.98
Synechocystis sp. ‡	Kazusa DNA Research Institute	3.57
Thermotoga maritima	TIGR	1.80
Treponema denticola	TIGR and University of Texas	3.00
Treponema pallidum*	TIGR and University of Texas	1.05
Ureaplasma urealyticum	University of Alabama, Birmingham	0.76
Vibrio cholerae	TIGR	2.50
Archaea		
Archaeoglobus fulgidus*	TIGR	2.20
Halobacterium salinarium	Max Planck Institute for Biochemistry	4.00
Methanococcus jannaschii‡	TIGR	1.66
Methanobacterium thermoautotrophicum	Genome Therapeutics and Ohio State University	1.70
Pyrobaculum aerophilum	University of California Los Angeles and Recombinant Biocatalysis Inc.	1.80
Pyrococcus furiosus	Center of Marine Biotechnology and University of Utah	2.10
Pyrococcus shinkaj	MITI and University of Tokyo	2.00
Sulfolobus solfataricus	Canadian and European consortium	3.05
Thermoplasma acidophilum	Max Planck Institute for Biochemistry	1.70

*Expected to be completely sequenced by Spring 1997.   †Application submitted, funding pending.   ‡Completely sequenced and published.

know why the archaea are absent from the group of pathogenic microbes, but as we continue to collect and examine microbial genome sequences, informative patterns may emerge that reveal fundamental distinctions between commensals and pathogens. A popular paradigm in the field of pathogenesis is one in which normally harmless bacteria acquire virulent properties; the archaeans may be an indication that, in some cases, there are more fundamental differences between organisms that cause us harm and those that live peacefully within us.

Number and Types of Sequences

Table 1 lists the organisms that have a completed genome sequence and those whose genomes are currently being sequenced. It contains a wide range of microbes, including many pathogens. The pace of research in this area is increasing rapidly. Although several of these sequences are currently proprietary, the amount of publicly available genomic information is growing quickly. The political issues and practical implications surrounding private ownership of genome sequences are subjects of great importance, but ones that are more appropriate for a news article (9).

The choice of genomes to examine has significant impact on the potential use of sequence information. It is sobering to realize that the full sequence of E. coli K-12, when available, is likely to have no more than a handful of the known virulence determinants of pathogenic Escherichia. Even carefully choosing one pathogenic member of a species is insufficient. Diversity within species produces profoundly different outcomes. The complete sequence of one H. influenzae strain is available. Yet, of the 104 H. influenzae clones that exist, six cause about 80% of all human disease (10). Why are these clones different from any other clone? It is not immediately obvious from the genomic sequence of one organism why only a few emerged in recent history as major human pathogens. It is impractical and impossible, of course, to sequence every organism in existence. Technological innovations (11) promise to allow largescale genome comparisons, thus expanding the usefulness of information from the genomes that are sequenced. Once strain- or pathogen-specific genes are identified, however, the challenge remains to devise ways to study their biological functions.

Using Care When Analyzing Sequences

Once candidate virulence genes are identified, the sequence information must be interpreted. The statistical information that is currently available from database searches is often underutilized. Fortunately, there are numerous resources, including the Internet, that can facilitate wise use of available sequence information (12). It is critical to determine whether the gene of interest is a true functional homolog of a gene with known activities, or if they share only some features. Finally, even if a gene contains a motif that always signifies a particular enzymatic behavior, the gene's function may remain unclear. For example, in vitro activity often does not reveal the relevant biological substrate or substrates.

Finally, even if a gene contains a recognizable motif and the motif can function as predicted, it is still possible that the molecule's biological relevance lies in a different arena or is more subtle than expected. An example of this is the case of calmodulin, a calciumbinding protein required for viability in Saccharomyces cerevisiae. Although calmodulin possesses calcium-dependent functions in yeast, a calmodulin derivative that no longer binds calcium is still able to perform its essential activities. This was a surprising finding and one antithetical to functional predictions based on sequence (13).

Uninformative and Even Misleading Sequence Comparisons

Approximately 42% of the Haemophilus genes are either orphans, not belonging to any established gene family, or are similar to genes of unknown function (2). These sequences, therefore, do not immediately lend insight into biological activity.

Despite low or undetected sequence similarities (even using the relatively lax requirement of amino acid sequence similarity), proteins still may share functions. For example, Listeria monocytogenes and Shigella flexneri display strikingly similar behavior inside host cells. They both harness actin, using the force generated by actin polymerization to infect adjacent cells (Fig. 1) (14). In each case, one bacterial molecule is used to initiate actin assembly (ActA in L. monocytogenes and IcsA in S. flexneri). However, ActA and IcsA have virtually no sequence similarity. This theme of evolutionarily unrelated molecules performing similar functions is reiterated and strengthened by the observation that vaccinia virus, which shares the ability to recruit host cell actin, contains no genes with resemblance to either ActA or IcsA (15). The genes used to harness actin would not be identified in a
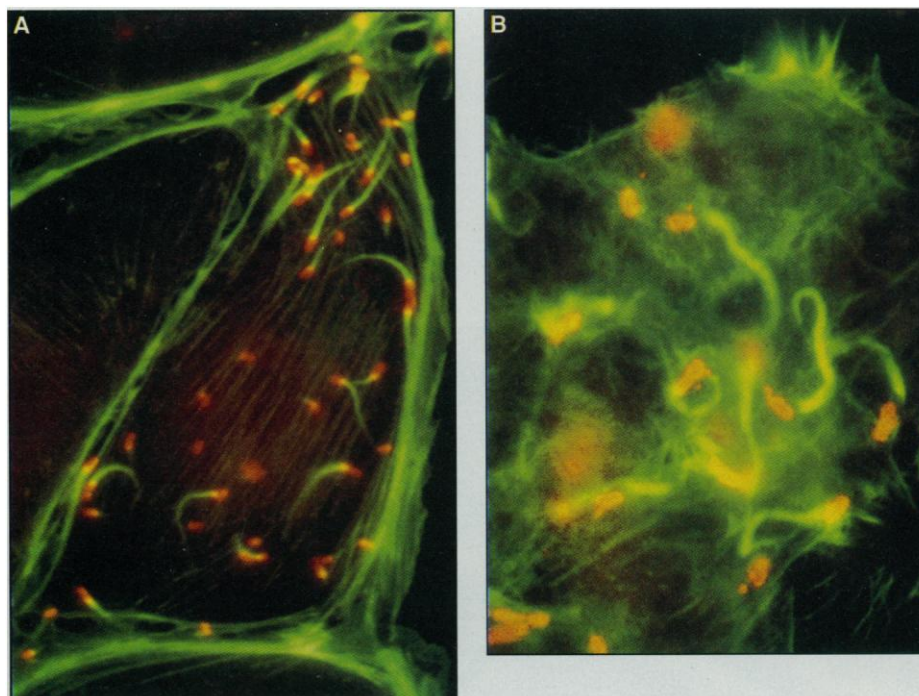

Fig. 1. Listeria monocytogenes and S. flexneri use actin to propel themselves within host cells. (**A**) Listeria monocytogenes in Ptk-2 cells were labeled with rabbit α-ActA polyclonal serum, followed by Texas red–labeled α-rabbit antibody (red), and actin was labeled with fluorescein-phalloidin (green) (courtesy of J. Theriot). (**B**) Shigella flexneri in HeLa cells were labeled with rabbit α-LPS polyclonal serum, followed by rhodamine-labeled α-rabbit antibody (orange), and actin was labeled with NBD-phalloidin (green) (reprinted with permission from Molecular Microbiology).

genomics-based approach.

Sequence similarity can be misleading in some cases. For example, *Salmonella* and *S. flexneri* share similar virulence factor secretion systems (*16*), and some of the secreted molecules functionally complement each other. The *S. flexneri ipaB* gene product plays an essential role in host cell invasion and subsequent escape from the endosomal vacuole. The *Salmonella* homolog, SipB, is required for *Salmonella* invasion, and it can complement the invasive defect of a *S. flexneri ipaB*-strain (*17*). This functional complementation is not surprising, because both organisms invade epithelial cells. What is notable is that *sipB* also overcomes the inability of the mutant *Shigella* strain to escape from the vacuole into the cytoplasm: In contrast to *Shigella*, *Salmonella* remain inside the vacuole throughout their intracellular phase. The *Salmonella* gene is thus performing a function in *Shigella* that it does not normally display in its parental organism. Presumably SipB's inherent ability to dissolve the membrane is blocked in *Salmonella*. In this case, sequence scanning and functional complementation might lead to the conclusion that SipB and IpaB perform identical functions in their native strains. Just because a factor can do something does not mean that it does.

A gene's sequence does not necessarily reveal when its corresponding product is expressed during the life of the microorganism. A critical feature of any successful infection, and indeed any microbial life cycle, is responsiveness to extracellular cues (*18*). Such conditions can be complex, and the microbial response must be precise to ensure successful infection. Thus, it is not sufficient to know biochemical activity; one must also uncover the circumstances under which a particular molecule acts. In *Salmonella typhimurium*, expression of genes required for invasion is keyed to a precise combination of pH, osmolarity, and oxygen concentration (*19*). Presumably, these conditions reflect signals that the bacteria experience when they encounter the specific type of epithelial cell in the Peyer's patch that they invade. Understanding when and where gene products act is an essential component of elucidating mechanisms of pathogenicity, and one that illustrates some of the limitations of using sequence information alone to discern function.

Getting at Function

Genomics can provide clues, but scientists in the age of complete genomic sequences are still left with the challenge of determining function. Conventional genetic and biochemical approaches remain vital to these types of endeavors. In addition, new tools are being used to probe microbial interactions within the host or environmental reservoir. The coordinate expression of a subset of genes in the host is often necessary for an organism's ability to colonize, survive, and replicate successfully. Identification of bacterial genes expressed preferentially within infected cells and animals is critical in understanding how bacterial pathogens circumvent the immune system and cause disease.

Selected Modern Approaches

With the complete sequence of an organism in hand, it is possible in principle, to precisely define which genes are expressed under any particular set of conditions. This can be accomplished in a number of ways, and the efficiency of these methods can be significantly enhanced by use of information from genomics.

One technique is to survey which messages are produced in the experimental system of choice. An approach was devised to study globally regulated genes in *E. coli*, using overlapping lambda clones spanning the genome to map and clone genes whose transcription levels change in response to various conditions (*20*). New methods that employ high-density nucleic acid probe arrays on chips have been used to examine mRNA levels in several systems (*21*). The power of these techniques lies in their ability to assay expression of thousands of genes in parallel and their semiautomated and quantitative properties. With such approaches, a series of message snapshots can be compiled that documents gene expression temporally, such as during an infection, or spatially, at different sites in the host's body.

A genetic approach was devised for identifying genes that are specifically induced during infection (*22*). The general method of in vivo expression technology (IVET) is to make a library in which random genomic fragments are ligated to a gene for a selectable marker that is required for survival in the host animal. Only those bacteria harboring a fusion that contains an active promoter will survive passage through the host. Fusions bearing promoters with constitutive activity can be identified and discarded by examining reporter activity on laboratory medium. By harvesting bacteria from different sites in the body, a list of genes required for different stages of infection can be compiled. The general method of IVET can be applied to a variety of situations where a pair of experimental conditions can be compared. The IVET system can also be altered to vary the selective stringency. One strategy demands only that the promoter be activated briefly in the host (*23*). This will yield those genes that are required at a specific time during, rather than throughout, infection.

Identifying host-induced bacterial genes with fluorescence-based methods is especially powerful because it allows high throughput and semiautomation. Green fluorescent protein (Gfp) can be expressed in a variety of microorganisms without adversely affecting their pathogenicity. A procedure called differential fluorescence induction (DFI) has been developed, in which bacteria bearing random transcriptional fusions to *gfp* are sorted by a fluorescence-activated cell sorter on the basis of stimulus-dependent synthesis of Gfp (*24*). The methodology has been employed to identify genes of *Salmonella* that respond to an acidic environment as well as those genes that are exclusively expressed within macrophages (Fig. 2). With the use of a fluorescence-activated cell
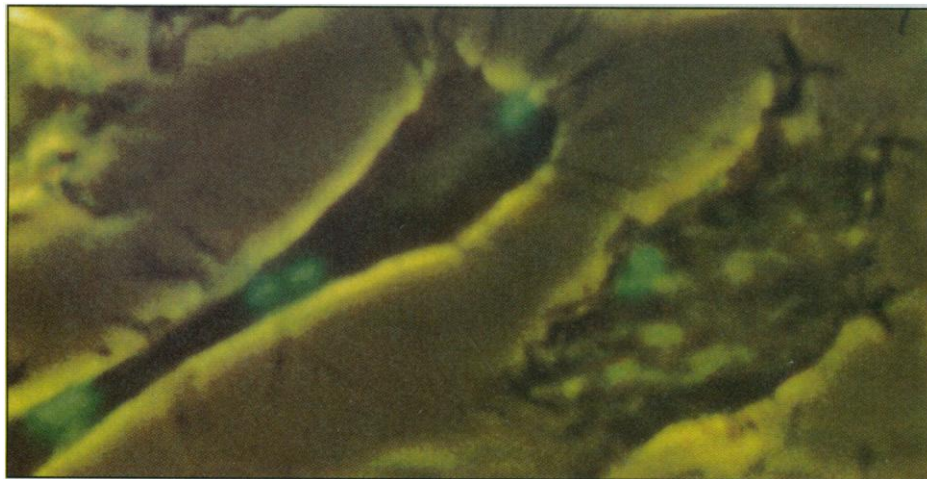

Fig. 2. RAW 264.7 cells were infected with *S. typhimurium* expressing Gfp fused to a macrophage-inducible promoter (courtesy of R. Valdivia).

sorter to select for Gfp expression, the selection parameters can be adjusted such that bacterial clones can be separated on the basis of small differences in expression levels. Because selections are independent of nutritional requirements or drug susceptibility, this method uncouples metabolic requirements from selection parameters.

A powerful screening method has been used to identify genes that are required for survival in a host (25). This signature-tagged transposon method (STM), like IVET, requires passage through a host, but it does not measure promoter activity; rather, it determines whether disrupting a particular gene adversely affects survival. In STM, each member of a complex library of mutants is marked with a unique oligonucleotide sequence. If a mutant is absent after passage of the library through an infected animal or another selective environment, the mutation it harbored may be in a gene essential for survival. The first application of STM was highly successful and led to the discovery of a previously unknown pathogenicity island necessary for *Salmonella* survival in the spleen (26).

IVET, STM, and DFI represent the first experimental forays to detect and follow specific virulence factors at discrete stages of interaction between the mammalian host and the invading microorganism. Complete genomic sequence information will be helpful in using these methods because, once a promoter or disrupted gene is identified, the entire gene to which it corresponds can be easily isolated. Furthermore, identifying candidate virulence genes by comparative approaches (such as those described) may increase the efficiency of these genetic methods. The availability of such gene subsets should allow investigators to use directed, rather than open-ended, strategies, at least for the many pathogens in which targeted gene disruption technology is available. Eventually, it may be possible to build upon the genetic methods, adapting them for use in conjunction with the high-density probe array systems and exploiting the power of semiautomation inherent to these techniques. These methodologies and others that will inevitably follow will help identify many genes that are activated or required during infections in various model systems.

Findings from such studies, when melded with the information of genomics, will begin to provide us with discrete signature motifs that are characteristic of particular constellations of genes involved in host-parasite relationships. The instructions that dictate pathogenic relationships are encrypted in the DNA. Genomics provides us

with all the pieces of a complex jigsaw puzzle; as we decipher function, new pieces will suddenly fit, and the picture will take on a recognizable form.

Future Challenges

The techniques discussed here are most powerful when used in conjunction with experimental systems that include a host or host cells that accurately reflect conditions of a natural infection. Virulence genes and their functions will be completely understood only when we have elucidated the signals provided by these environments. In the meantime, however, IVET, STM, and DFI allow us to probe bacterial-host interactions in the absence of full knowledge about the relevant chemical and physical cues. These methods facilitate identification of expressed sequences or genes important under a given set of circumstances without requiring that the conditions be replicated accurately in a test tube. The approaches do not necessarily elucidate roles of the gene products. In some cases, however, knowing when or where a factor is required, in combination with information about its sequence, can lead to reasonable and testable hypotheses of gene function.

Although we have developed some tools with which to probe pathogen-host interactions, we have barely begun to study what happens to pathogens in the environment. Events during the environmental stage of a pathogen's life cycle can have profound effects on their potential for causing infection. For example, *Vibrio cholerae* enters a viable but not culturable state (VBNC) under particular conditions (27). Organisms in this state remain a serious threat to human populations because certain cues can trigger exit from VBNC to the infectious form of *V. cholerae*. It is essential to understand such transitions in order to gain a coherent view of this pathogen's life cycle and to hope to control cholera epidemics.

Studies of entire microbial populations in the wild aimed at identifying the inhabitants of particular environments (28) may segue nicely into the pathogenicentric world presented in this review. For example, such studies may identify nonculturable phases of human pathogens for which we do not currently know the environmental reservoir, or they may reveal obligate syntrophic relationships that exist in the host as well as in the external environment. Phylogenetic analysis from these studies has produced information that is useful for classifying new or unclassified pathogens. It is advantageous to increase our familiarity with characteristic properties of the wide variety of microbial families present on Earth. This knowledge is

helpful in both assigning a pathogen to a particular group and in inferring properties based on that classification. Such information may suggest methods by which to culture the pathogen or interfere with its growth.

Genomics provides us with a snapshot of a microorganism in its current state of evolution. But bacteria evolve significantly within a single human lifetime. One need only look at the impact of worldwide antibiotic use on natural selection to see the sudden appearance of new bacterial clones that differ significantly from those seen in historical memory. In 50 years, the sequence of a representative pathogen could be significantly different than it is now. One benefit of collecting microbial sequences will be the opportunity to document evolutionary changes at a level of precision previously unattainable.

Counterintuitive as it may seem, many of the most notorious diseases in recent decades are illnesses of human progress. For example, the shift in the West from baths to showers has had a profound effect on microorganisms such as *Legionella pneumophila*, which can be transmitted by aerosolization in potable water supplies. We have created new environmental niches for many organisms, thus introducing them into human proximity (29). As the new scientific discipline of genomic analysis of bacteria matures and our database grows, it will permit us to better understand the nature of the delicate balance that exists between us and the microorganisms that inhabit our bodies.

Just as bacteria have served as model systems in which to develop genetic techniques, their small and manipulable genomes promise to serve the global scientific community well as we embark on this new genomic path. The 21st century will no doubt usher in ways to tap the vast resource of genomic information as we use our ingenuity to understand microbes and their intimate relationship with us and all life on the planet.

REFERENCES AND NOTES

1. C. M. Fraser et al., Science 270, 397 (1995).
2. R. D. Fleischmann et al., ibid. 269, 496 (1995).
3. D. W. Hood et al., Proc. Natl. Acad. Sci. U.S.A. 93, 11121 (1996).
4. D. W. Hood et al., Mol. Microbiol. 22, 951 (1996).
5. A. R. Mushegian and E. V. Koonin, Proc. Natl. Acad. Sci. U.S.A. 93, 10268 (1996).
6. J. Mecsas and E. J. Strauss, Emerg. Infect. Dis. 2, 271 (1996).
7. T. S. Whittam et al., Infect. Immun. 61, 1619 (1993).
8. R. Sjogren et al., Gastroenterology 106, 306 (1994).
9. E. Marshall, Science 275, 777 (1997).
10. D. A. Relman and S. Falkow, in Principles and Practice of Infectious Diseases, G. L. Mandell, J. E. Bennett, R. Dolin, Eds. (Churchill Livingstone, New York, 1995).
11. R. J. Lipshutz et al., Biotechniques 19, 442 (1995);

M. Chee et al., Science **274**, 610 (1996).
12. cmgm.stanford.edu/classes/genefind/
13. J. R. Geiser, D. van Tuinen, S. E. Brockerhoff, M. M. Neff, T. N. Davis, Cell **65**, 949 (1991).
14. J. A. Theriot, Ann. Rev. Cell Dev. Biol. **11**, 213 (1995).
15. S. Cudmore, P. Cossart, G. Griffiths, M. Way, Nature **378**, 636 (1995).
16. J. E. Gálan, Mol. Microbiol. **20**, 263 (1996).
17. D. Hermant, R. Menard, N. Arricau, C. Parsot, M. Y. Popoff, ibid. **17**, 781 (1995).
18. M. J. Mahan, J. M. Slauch, J. J. Mekalanos, in Escherichia coli and Salmonella, F. C. Neidhardt, Ed. (American Society for Microbiology, Washington,

DC, 1996).
19. V. Bajaj, R. L. Lucas, C. Hwang, C. A. Lee, Mol. Microbiol. **22**, 703 (1996).
20. S.-E. Chuang, D. L. Daniels, F. R. Blattner, J. Bacteriol. **175**, 2026 (1993)
21. M. Schena et al., Proc. Natl. Acad. Sci. U.S.A. **93**, 10614 (1996); D. J. Lockhart et al., Nature Biotechnol. **14**, 1675 (1996).
22. M. J. Mahan, J. M. Slauch, J. J. Mekalanos, Science **259**, 686 (1993).
23. A. Camilli et al., Mol. Microbiol. **18**, 671 (1995).
24. R. H. Valdivia and S. Falkow, ibid. **22**, 367 (1996).
25. M. Hensel et al., Science **269**, 400 (1995).

26. J. E. Shea, M. Hensel, C. Gleeson, D. W. Holden, Proc. Natl. Acad. Sci. U.S.A. **93**, 2593 (1996).
27. R. R. Colwell, Science **274**, 2025 (1996).
28. N. Pace, ibid. **276**, 734 (1997).
29. B. B. Finlay and S. Falkow, Microbiol. Rev., in press.
30. We thank D. Botstein, D. Brutlag, B. Cormack, D. Frank, T. McDaniel, J. F. Miller, K. Ottemann, B. Raupach, and R. Valdivia for helpful discussions and critical comments on the manuscript. We also thank P. Sansonetti for providing Fig. 1B. Supported by an American Cancer Society postdoctoral fellowship (grant PF-4120) to E.J.S. and USPHS grant AI 26195 to S.F.

Protein Localization and Cell Fate in Bacteria

Lucy Shapiro and Richard Losick

A major breakthrough in understanding the bacterial cell is the discovery that the cell is highly organized at the level of protein localization. Proteins are positioned at particular sites in bacteria, including the cell pole, the incipient division plane, and the septum. Differential protein localization can control DNA replication, chromosome segregation, and cytokinesis and is responsible for generating daughter cells with different fates upon cell division. Recent discoveries have revealed that progression through the cell cycle and communication between cellular compartments are mediated by two-component signal transduction systems and signaling pathways involving transcription factor activation by proteolytic processing. Asymmetric cell division in Caulobacter crescentus and sporulation in Bacillus subtilis are used as paradigms for the control of the cell cycle and cellular morphogenesis in bacterial cells.

Temporal and spatial constraints restrict how a bacterium gleans information from its genome. These constraints dictate the ordered expression of cell cycle events, the three-dimensional organization of the cell, and the biogenesis of subcellular organelles and altered cellular forms. Each of these general classes of cellular functions is programmed by a genetic network that includes a pathway of regulatory events. The combination of bacterial genetics and new ways of using cytological methods to examine bacterial cell organization has not only led to the identification of the components of these regulatory pathways, but has placed the pathway outcomes at specific sites in the cell. Analysis of two bacteria separated by large evolutionary distance, the Gram-negative, aquatic bacterium Caulobacter crescentus and the Gram-positive, soil bacterium Bacillus subtilis, has elucidated common mechanisms that serve as paradigms for prokaryotic cell cycle control and cellular morphogenesis.

The use of immunogold electron micros-

copy and fluorescence microscopy to study the subcellular organization of bacterial cells has revealed a surprising extent of protein compartmentalization and localization. In some instances, localized proteins are not essential for viability, as is the case for the flagellum and the chemosensory apparatus (1), actin-recruiting virulence proteins in certain pathogenic bacteria (2), and certain morphogenetic and regulatory proteins involved in the process of sporulation (3–5). In other cases, such as the proteins involved in DNA replication, chromosome segregation, or cell division, the site of protein deposition is an integral part of the regulatory machinery that mediates these essential functions. A critical question now facing the prokaryotic community is how proteins are deployed in a targeted three-dimensional array in the absence of known cytoskeletal components. Notwithstanding free diffusion in the microenvironment of the bacterial cell, how are proteins directed to specific sites in the cell, and how are they kept there? What controls the rearrangement of proteins during morphogenesis? As we will show, C. crescentus and B. subtilis offer numerous and striking examples of subcellular localization of proteins. These, and examples from other bacteria, have changed our view of the bacterial cell and

demonstrated that it is important not only to understand what a protein does but where and when it acts.

Asymmetric Cell Division

Differentiation in C. crescentus and B. subtilis results from a modified process of cell division (6, 7). Unlike the familiar process of binary fission in which a dividing cell generates identical progeny, C. crescentus and sporulating B. subtilis undergo cell division asymmetrically, each generating dissimilar progeny with distinct fates (Fig. 1). Asymmetric division in C. crescentus is an obligatory feature of its cell cycle. Thus, a dividing cell generates a motile, swarmer cell, which is propelled by a rotary flagellum, and a nonmotile, stalked cell. The stalked cell replicates its chromosome and undergoes differentiation into a predivisional cell with a nascent swarmer cell on one side of the division plane and the stalked cell on the other. Meanwhile, after an interval in which DNA replication is repressed, the swarmer sheds its flagellum and undergoes metamorphosis into a stalked cell, which repeats the cycle.

Asymmetric division in sporulating B. subtilis, in contrast, is an environmentally induced adaptation to conditions of nutritional stress. Bacillus subtilis normally propagates by binary fission. However, when nutrients become limiting, it undergoes a switch from symmetric to asymmetric division by the formation of a polar septum. The polar septum partitions the developing cell into a small, forespore cell and a large, mother cell. Initially, the forespore and the mother cell lie side by side in the sporangium. Later, the forespore becomes wholly engulfed by the mother cell in a phagocytic-like process to create a cell within a cell. The forespore ultimately becomes a spore and gives rise to future progeny, whereas the mother cell is discarded by lysis when maturation of the spore is complete.

Polarity and Protein Localization

In C. crescentus, the cell cycle perpetuates asymmetry that is present before cell division. The predivisional cell is itself asym-

L. Shapiro is in the Department of Developmental Biology, Beckman Center, Stanford University School of Medicine, Stanford, CA 94305–5427, USA. R. Losick is in the Department of Molecular and Cellular Biology, The Biological Laboratories, Harvard University, Cambridge, MA 02138, USA.