proteins can be expressed recombinantly and tested for the ability to elicit a serologic response and protective immunity. This process, termed *reverse vaccinology*, has proved particularly useful for pathogens that are difficult to culture or poorly immunogenic, as was the case with the development of a vaccine for *Neisseria meningitidis* serogroup B.

Large-scale gene content analysis from sequencing or expression profiling enables new research directions that provide novel insights into the interplay of pathogen and host during infection or colonization. One important goal of such research is to suggest new therapeutic approaches to disrupt this interaction in favor of the host. Indeed, one of the most immediate applications of next-generation sequencing technology has come from simply characterizing human pathogens and related commensal or environmental strains and then finding genomic correlates for pathogenicity. For instance, as *Escherichia coli* varies from a simple nonpathogenic, lab-adapted strain (K-12) to a Shiga toxin–producing enterohemorrhagic gastrointestinal pathogen (O157:H7), it displays up to a 25% difference in gene content, even though its phylogenetic classification stays the same. Although this is an extreme example, it is not an isolated case. Some isolates of *Enterococcus*—notorious for its increasing incidence of resistance to common antibiotics such as ampicillin, vancomycin, and aminoglycosides—also contain recently acquired genetic material comprising up to 25% of the genome on mobile genetic elements. This fact suggests that horizontal gene transfer may play an important role in the organism's adaptation as a nosocomial pathogen. On closer study, this genome expansion has been demonstrated to be associated with loss of regulatory elements called CRISPRs (*c*lustered, *r*egularly *i*nterspaced *s*hort *p*alindromic *r*epeats). Loss of CRISPR elements, which protect the bacterial genome from invasion by certain foreign genetic materials, may thus facilitate the acquisition of antibiotic resistance–conferring genetic elements. While loss of this regulation appears to impose a competitive disadvantage in antibiotic-free environments, these drug-resistant strains thrive in the presence of even some of the most useful antienterococcal therapies. In addition to insights gained from genome sequencing, extension of unbiased whole-transcriptome sequencing (RNA-Seq) efforts to bacteria is beginning to identify unexpected regulatory, noncoding RNAs in many diverse species. While the functional implications of these new transcripts are as yet largely unknown, the presence of such features—conserved across many bacterial species—implies evolutionary importance and suggests areas for future study and possible new therapeutic avenues.

Thus, genomic studies are already beginning to transform our understanding of infection, offering evidence of virulence factors or toxins and providing insight into ongoing evolution of pathogenicity and drug resistance. One goal of such studies is to identify therapeutic agents that can disrupt the pathogenic process; there is currently much interest in the theoretical concept of antivirulence drugs that inhibit virulence factors rather than killing the pathogen outright as a means to intervene in infection. Further, as sequencing becomes increasingly accessible and efficient, large-scale studies with unprecedented statistical power to associate clinical outcomes with pathogen and host genotypes and thus to further reveal vulnerabilities in the infection process that can be targeted for disruption are being initiated. Although this is just the beginning, such studies point to a tantalizing future in which the clinician is armed with genomic predictors of infection outcome and therapeutic response to guide clinical decision-making.

EPIDEMIOLOGY OF INFECTIOUS DISEASES

Epidemiologic studies of infectious diseases have several main goals: to identify and characterize outbreaks, to describe the pattern and dynamics of an infectious disease as it spreads through populations, and to identify interventions that can limit or reduce the burden of disease. One classic, paradigmatic example is John Snow's elucidation of the origin of the 1854 London cholera outbreak. Snow used careful geographic mapping of cases to determine that the likely source of the outbreak was contaminated water from the Broad Street pump, and, by removing the pump handle, he aborted the outbreak. Whereas that intervention was undertaken without knowledge of the causative agent of cholera, advances in microbiology and genomics have expanded the purview of epidemiology, which now considers not just the disease but also the pathogen, its virulence factors, and the complex relationships between microbial and host populations.

Through the use of novel genomic tools such as high-throughput sequencing, the diversity of a microbial population can now be rapidly described with unprecedented resolution, with discrimination between isolates that have single-nucleotide differences across the entire genome and advancement beyond prior approaches that relied on phenotypes (such as antibiotic resistance testing) or genetic markers (such as multilocus sequence typing). The development of statistical methods grounded in molecular genetics and evolutionary theory has established analytical approaches that translate descriptions of microbial population diversity and structure into insights into the origin and history of pathogen spread. By linking phylogenetic reconstruction with epidemiologic and demographic data, genomic epidemiology provides the opportunity to track transmission from person to person, to infer transmission patterns of both pathogens and sequence elements that confer phenotypes of interest, and to estimate the transmission dynamics of outbreaks.

DECIPHERING PERSON-TO-PERSON TRANSMISSION

The use of comparisons of whole-genome sequencing to infer person-to-person transmission and identify point-source outbreaks of pathogens has proved useful in hospital infection control settings. As reported in a seminal paper in 2010, a study of MRSA in a Thai hospital demonstrated that whole-genome sequencing can be used to infer transmission of a pathogen from patient to patient within a hospital setting through integration of the analysis of accumulation of mutations over time with dates and hospital locations of the infections. Since that time, multiple instances of the use of whole-genome sequencing to define and motivate interventions aimed at interrupting transmission chains have been reported. In another MRSA outbreak in a special-care baby unit in Cambridge, United Kingdom, whole-genome sequencing extended the traditional infection control analysis, which relies on typing organisms by their antibiotic susceptibilities, to sequencing of isolates from clinical samples. This approach identified an otherwise unrecognized outbreak of a specific MRSA strain that was occurring against a background of the usual pattern of infections caused by a diverse circulating population of MRSA strains. The analysis showed evidence of transmission among mothers within the special-care baby unit and in the community and demonstrated the key role of MRSA carriage in a single health care provider in the persistence of the outbreak. MRSA decolonization of the health care provider terminated the outbreak. In yet another example, in response to the observation of 18 cases of infection by carbapenemase-producing *Klebsiella pneumoniae* over 6 months at the National Institutes of Health Clinical Research Center, genome sequencing of the isolates was used to discriminate between the possibilities that these cases represented multiple, independent introductions into the health care system or a single introduction with subsequent transmission. On the basis of network and phylogenetic analysis of genomic and epidemiologic data, the authors reconstructed the likely relationships among the isolates from patient to patient, demonstrating that the spread of resistant *Klebsiella* infection was in fact due to nosocomial transmission of a single strain.

Uncovering of unexpected transmission events by genomic epidemiology studies is motivating renewed questioning of pathogen ecology and modes of transmission. For example, the rise in prevalence of infections with nontuberculous mycobacteria, including *Mycobacterium abscessus*, among patients with cystic fibrosis (CF) has led to speculation about the possible role of patient-to-patient transmission in the CF community; however, conventional typing approaches have lacked the resolution to define population structure accurately, a critical component of inferring transmission. Past infection control guidelines discounted the possibility of acquisition of nontuberculous mycobacteria in health care settings, as no strong evidence for such transmission had been described. In a whole-genome sequencing study of *M. abscessus* isolates from patients with CF, an analytical approach using genome sequencing, epidemiology, and