CHAPTER 23.

Infectious diseases

Betsy Foxman

Summary

Molecular tools have enhanced our understanding of the epidemiology of infectious diseases by describing the transmission system, including identifying novel transmission modes and reservoirs, identifying characteristics of the infectious agent that lead to transmission and pathogenesis, identifying potential vaccine candidates and targets for therapeutics, and recognizing new infectious agents. Applications of molecular fingerprinting to public health practice have enhanced outbreak investigation by objectively confirming epidemiologic evidence, and distinguishing between time-space clusters and sporadic cases. Clinically, molecular tools are used to rapidly detect infectious agents and predict disease course. Integration of molecular tools into etiologic studies has identified infectious causes of chronic diseases, and characteristics of the agent and host that modify disease risk. The combination of molecular tools with epidemiologic methods provides essential information to guide clinical treatment, and to design and implement programmes to prevent and control infectious diseases. However, incorporating molecular tools into epidemiologic studies of infectious diseases impacts study design, conduct, and analysis.

Historical perspectives

The development of epidemiology as a discipline was roughly simultaneous with the development of microbiology. As the presence of infectious agents was linked to disease, laboratory methods were incorporated into epidemiologic studies. One epidemiologic hero is John Snow, who identified a strong epidemiologic association between sewage-contaminated water and cholera. Despite extremely well-documented evidence supporting thoroughly researched and reasoned arguments, his findings remained in doubt for some time. Max Von Pettenkofer, 1818–1901, a contemporary of Snow and also an early epidemiologist, is

related (in a perhaps apocryphal story) to have drunk a glass of the stool of someone with cholera to test the hypothesis; Pettenkofer remained disease-free. Snow's conclusions were not generally accepted until 25 years after his death, when the cholera vibrio was discovered by Joseph Koch, who definitively demonstrated the causal relationship between the vibrio and cholera (1). The strategy of isolating an organism from an ill individual, showing it can cause disease in a disease-naïve individual, and then be re-isolated as described in the landmark postulates of Henle and Koch reflects how incorporating laboratory methods enhances our ability to make causal inferences about disease transmission and pathogenesis from even the most carefully researched epidemiologic evidence.

Early epidemiologists made tremendous strides with what are now relatively simple molecular tools: using microscopy for identification, which showed that agents not visible by microscope caused disease ("filterable"); and detection of protective antibodies with haemagglutination assays. For example, Charles Nicolle and Alphonse Laveran showed that a protozoan caused malaria (2,3), and Charles Nicolle demonstrated, by injecting a monkey with small amounts of infected louse, that lice transmitted typhus. He also observed that some animals carry infection asymptomatically (3). Wade Hampton Frost used the presence of protective antibodies in the serum of polio patients to explain the emergence of polio epidemics (4). These early, dramatic successes combined with the successful development and implementation of vaccines against major childhood diseases, including smallpox, measles, diphtheria, whooping cough and polio, and the identification of antibiotics, led to a rather simplistic view of infectious disease, and ultimately to the incorrect impression that we might "close the book" on infectious disease during the 20th century. This assertion was quickly undermined in the last quarter of the 20th century by the emergence of new infectious agents such as human immunodeficiency virus (HIV), Ebola and Hantavirus, the re-emergence of tuberculosis and malaria, and the transcontinental transmission of agents such as West Nile Virus and Dengue.

Infectious disease epidemiologists were early adapters of modern molecular biologic techniques to epidemiology, such as those used in genomics. Indeed, the term molecular epidemiology came from infectious disease work (5). Modern molecular techniques have fundamentally changed our understanding of the epidemiology of infectious agents. Characterizing the genetics of human pathogens has revealed the tremendous heterogeneity of various infectious agents, and the rapidity with which they evolve. This heterogeneity and rapid evolution helps explain our difficulties in creating successful vaccines for the more heterogeneous organisms, such as *Neisseria gonorrhoeae*. With the increased ability to detect host immunologic response to infectious agents, we increase our ability to test and refine our theories about the extent and duration of immunity, a key parameter in disease spread. Molecular analysis has also revealed the role of infectious agents in the initiation and promotion of previously classified chronic diseases. Further, molecular tools have enhanced our understanding of the epidemiology of infectious diseases by describing the transmission systems, identifying novel transmission modes and reservoirs, identifying characteristics of the infectious agent that lead to transmission and pathogenesis, revealing potential targets for vaccines and therapeutics, and recognizing new infectious agents. The combination of molecular tools with epidemiologic methods thus provides essential information to design and implement programmes to prevent and control infectious diseases.

Outbreak investigations

Molecular tools have substantially improved inferences from outbreak investigations. We can more rapidly provide laboratory confirmation of disease diagnosis, and detect the presence of difficult-to-culture or uncultivable agents, enhancing the sensitivity and specificity of case definitions. Molecular tools make it possible to determine if the epidemiologically identified outbreak source, such as a food item, contains the infecting organism, and if the identified organism has the same genotype causing the outbreak, enhancing causal inference. Molecular typing can also be used to determine order of transmission (Table 23.1). It is not surprising that molecular typing has become a standard tool in outbreak investigation.

Case definition is an essential component of successful outbreak investigation. The detection of the infectious agent from all cases more accurately classifies cases than clinical diagnosis alone; molecular typing further minimizes misclassification. For point-source outbreaks, all individuals are expected to be infected with the same strain of a particular genus and species; for propagated outbreaks similar molecular fingerprints are expected that vary only in the

Table 23.1. Applications of molecular tools in outbreak investigations

• Enhance case definitions
• Determine whether cases occurring in the same time frame are part of the same outbreak
• Confirm or refute epidemiologic inferences regarding etiologic pathways
• Determine the order of transmission

mutations accrued over repeated transmission events. The resulting reduction of misclassification of disease status increases the study power and the validity of inferences.

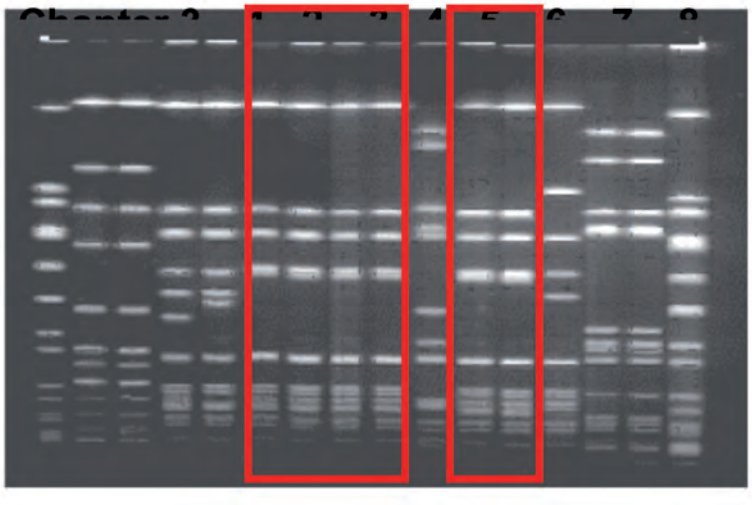
The first and most typical application of molecular tools in an outbreak situation is to confirm or refute epidemiologic information. For example, in a foodborne outbreak, isolates might be collected from all those with disease, the person suspected to have introduced the infected agent into a food item, and the suspected food item. Figure 23.1 shows the molecular fingerprints, determined using pulsed-field gel eletrophoresis (PFGE), of a foodborne *Staphylococcus aureus (S. aureus)* outbreak isolated from the suspected food that had the same molecular type as *S. aureus* found in the food handler, and in those with disease.

In most foodborne outbreaks, neither specimens from all individuals that meet the case definition, nor the putative food is available for testing by the time it is identified. With clear epidemiologic evidence, the demonstration of genetic relatedness between cases and the putative item is solely confirmatory, and adds little unless the food vehicle has not been previously identified. For example, epidemiologic evidence linked consumption of toasted oats cereal with a multistate outbreak of Salmonella in the USA. A culture of the cereal from one opened and two unopened boxes found *Salmonella agona (S. agona)* with the same PFGE pattern as the *S. agona* causing the outbreak (7). This was the first time a commercial cereal product was implicated in a Salmonella outbreak, so the PFGE evidence was particularly compelling.

A second application in an outbreak investigation is determining whether cases occurring in the same time frame are part of the same outbreak. With disseminated outbreaks, there can be apparent sporadic cases that are actually linked. For example, in 1997 a large (n = 126) foodborne outbreak of hepatitis A occurred in Michigan (8). Epidemiologic evidence implicated frozen strawberries from a single processor. During the same time period, a much smaller outbreak (n = 19) occurred in Maine, and sporadic cases occurred in three other states among individuals suspected to have consumed frozen strawberries from the same processor. The genetic sequences of the virus from all tested individuals were the same, confirming a common source of infection for all cases. Without molecular fingerprinting, it would have been very difficult, if not impossible, to link these apparently disparate cases to one common source using solely epidemiologic methods. This example also highlights the importance of using molecular tools with surveillance, discussed in detail in the next section.

A third application is to determine the order of transmission. This application has been particularly useful in forensic cases. Understanding the evolution of an organism and the ability to trace the order of that evolution has made it possible to detect cases where an individual deliberately infected others. For example, a gastroenterologist was convicted of infecting a former girlfriend with the blood of an HIV patient. A variety

Figure 23.1. Example of molecular typing using pulsed-field gel electrophoresis (PFGE). In this foodborne outbreak of methicillin sensitive *Staphylococcus aureus,* isolates from food handler (lane 6), cases of food poisoning (lanes 7-9, 11) and infected food (lane 12) had the same PFGE type (6).



of molecular analyses, including phylogenetic analyses of HIV-1 reverse transcriptase and *env* DNA sequences isolated from the victim, the patient, and a local population sample of HIV-1-positive individuals, strongly supported not only that there was transmission between the two individuals, but who had infected whom (9).

Surveillance

Surveillance is an essential component of a successful public health infrastructure. Laboratories are key components of many surveillance systems; hospital laboratories may be part of regional surveillance networks, as well as part of a local surveillance system. Monitoring of infectious disease isolates identifies time-space clusters of infection; molecular typing distinguishes between infectious agents of the same species, allowing differentiation among clusters of disease occurring by chance and true outbreaks. True outbreaks and clusters of the same strain can be traced back to a common source and presumably are amenable to public health intervention. Spurious clusters cannot, and their investigation wastes time and resources. Applying molecular tools to surveillance isolates can also identify new strains with increased virulence or changing patterns of resistance (Table 23.2).

Table 23.2. Applications of molecular tools in surveillance

• Distinguish between time-space clusters and sporadic cases of the same infection
• Identify clusters requiring further investigation
• Detect the emergence of strains with new resistance profiles
• Estimate prevalence of infection and observe trends over time

Hospitals have high endemic rates of bacterial infection, but the infections are often due to a bacterial strain that was colonising an individual before entering the hospital, for example, *S. aureus.* The prevalence of *S. aureus* colonization among the general population is 32% in the nares (10), but much higher in patients and personnel in hospitals and long-term care facilities. By typing strains causing infection among patients, a distinction can be made between a strain from the community and one circulating endemically or causing an outbreak within the hospital. The prevention and control strategies are different in each case, and thus it is important to make a distinction between them.

A second application of molecular tools in surveillance is to identify clusters requiring further investigation. By monitoring isolates from time-space clusters for the presence of a common molecular type, one can distinguish between common-source outbreaks that are local and those that are widely disseminated. Processed foods may be distributed widely, as demonstrated in the earlier example of the Michigan hepatitis A outbreak caused by frozen strawberries (8), so adding molecular typing to laboratory monitoring of specimens is essential. The US Centers for Disease Control and Prevention's PulseNet, a molecular subtyping surveillance system for foodborne bacterial disease, monitors *Escherichia coli (E. coli)* O157:H7, *Salmonella, Shigella*, and *Listeria monocytogenes*, and other bacterial pathogens (11) causing disease throughout the United States. In 2006, clusters of a common *E. coli* O157:H7 pulsed-field type were observed at several monitoring sites. An investigation revealed the source of the outbreak to be washed, pre-packaged, fresh spinach. Once the epidemiologic investigation identified spinach, the public was notified and *E. coli* O157:H7 with the putative pulsed-field type was isolated from an unopened package of spinach from an individual's home. Molecular typing enabled rapid linkage of cases occurring across several states, the identification of the disease source, and facilitated quick public health intervention.

A third application is the detection of infectious agents resistant or insensitive to prevailing therapies. A cluster of drug-resistant agents is often the first indication of an outbreak, particularly in a hospital setting. Mobile genetic elements that confer resistance can be exchanged between bacteria, even across species, complicating outbreak investigation. Molecular tools can distinguish between a common strain of a single bacterial species or a mobile genetic element, conferring antibiotic resistance across strains of the same or even different species. Outbreak control must take into account whether a mobile genetic element is being exchanged between species or if there is clonal spread of a single organism.

In the United States there is selective culture of organisms. In outpatient settings the most common bacterial infections, urinary tract infection and pneumonia, are generally treated empirically. Only if treatment fails is a culture taken. Thus surveillance for antimicrobial resistance reflects a biased sample, suggesting the subset most likely to have a resistant infection. The inherent selection biases should be taken into consideration when

suggesting policy changes in therapies based on surveillance data.

Molecular tools have also been applied to screen biological specimens collected as part of ongoing national databases for the presence of known and newly discovered infectious agents. For example, blood samples are collected as part of the National Health and Nutrition Examination Survey, a multistage probability sample of the United States conducted every 10 years. This has enabled the estimation of the prevalence of various infectious agents, including hepatitis B and C viruses, human herpes virus 8 (which causes Kaposi sarcoma) and herpes simplex viruses 1 and 2. These studies provide insight into the frequency of new agents, and the distributions of agents by spatial-temporal and host characteristics. Such studies are extremely useful for generating hypotheses about transmission systems, potential prevention and control strategies, evaluating the effectiveness of ongoing prevention and control programmes, and observing time trends.

Describe the transmission system

The transmission system of an infectious agent determines how infectious agents are circulated within a population, and includes the transmission mode, interactions between the infectious agent and the host, the natural history of the infection, and interactions between hosts that lead to infection. The emergence and re-emergence of a variety of infectious agents highlights the utility of understanding the various transmission systems, as this understanding is central to identifying effective prevention and control strategies. Combining molecular typing methods with questionnaire data can confirm self-reported behaviours, especially important when the validity of self-report may be in doubt, such as contact tracing of sexually transmitted diseases. As described in detail below, molecular tools facilitate estimating parameters key to understanding the transmission system, including the incidence, prevalence, transmission probability, duration of carriage, effective dose, and probability of effective contact.

Estimation of key parameters

When using simple transmission models to estimate R_0, the average number of new cases generated from each infectious case in a fully susceptible population, the transmission probability per effective contact is needed, as well as the duration of infectivity and the rate of effective contact. Molecular tools can usefully be applied to estimate each of these parameters. Prior to the availability of modern molecular tools, our ability to empirically estimate transmission probabilities was limited. For example, the transmission of a sexually transmitted infection can be estimated by following couples where one is infected and the other is susceptible; however, without molecular tools it is difficult to ensure that the transmission event is not attributable to a person outside the partnership. For respiratory infections, such as pulmonary tuberculosis, our estimates of the transmission probability and natural history have been based on careful documentation of outbreaks. However, as we have been able to type individual strains, it has been determined that tuberculosis cases that previously were considered sporadic, and not part of apparent time-space clusters (because the exposure to the index case was very limited), were indeed part of the same outbreak (12).

Key transmission system parameters are incidence, prevalence, duration of infection, and transmission probabilities. The estimation of these parameters assumes the accurate measure of identical strains or subtypes of an infectious agent. As we have increased our ability to type infectious agents, we have been forced to re-evaluate many of our previous assumptions. One key assumption is that pathogens are clonal; that is, during active infection all infecting organisms are the same. A second, parallel assumption is that during an infectious process the pathogenic organism will be the one most frequently isolated from the infected site. For many diseases, we now know these assumptions

Table 23.3. Components of the transmission system and associated parameters

Component	Parameter
Occurrence in a population	Incidence Prevalence
Transmission mode	Probability of transmission given contact
Natural history of infection	Duration of infection
Interactions between agent and host	Effective dose
Interactions between hosts leading to transmission	Probability of contacting an infected individual

are false. For example, individuals can be infected with different strains of human papillomavirus (HPV), gonorrhea and even tuberculosis. During a diarrheal episode, the predominant organism isolated from the stool may not be the one causing the symptoms: a toxin-secreting organism occurring at low frequency may be the culprit. For infectious agents that also are human commensals, such as Streptococcus agalactiae, strains causing disease may be different from normal inhabitants and different strains may have different transmission systems.

These observations have profound impacts on the conduct of future studies. If the population genetic structure of pathogens is not clonal and the pathogen is not readily isolated, this must be reflected in the sampling of isolates for study. Multiple isolates must be sampled and tested from an individual. For example, if there is a second strain only 5% of the time and the pathogen is uniformly distributed in the sample, 28 different isolates must be sampled from an individual to reliably detect the second strain. Further, if an infectious agent mutates rapidly within a host, such as HIV, determining the mutation rate will be essential for accurately estimating transmission probabilities and following transmission chains.

As molecular tools can detect the presence of the organism or host response to a specific organism, in some cases, for a particular strain (13), studies can detect both incidence and prevalence of asymptomatic infection and clinical disease. Understanding the full extent of the circulation of a particular infectious agent is essential for making accurate predictions and determining appropriate prevention and control strategies.

The duration of carriage can be estimated from the prevalence and incidence, presuming that the average duration across strain type is of interest. However, if duration is short but incidence is high, an individual might become re-infected with a different strain type, suggesting a longer duration if strain types are not determined. By contrast, if a strain mutates rapidly within the human host, duration of carriage might be underestimated. Thus, strain-specific estimates of prevalence and incidence are essential to our understanding of disease etiology, especially if different strains have different propensities to cause diseases.

Using molecular tools to estimate contact patterns

Molecular tools can also assist in the estimation of contact patterns by identifying asymptomatic and low levels of infections. Asymptomatic infection is often a key component in maintaining disease transmission. For example, in a study of intra-family transmission of *Shigella*, asymptomatic carriage increased risk of a symptomatic episode within 10 days by nine-fold (14). Molecular typing can also be used to enrich and validate contact tracing information. The addition of molecular typing to epidemiologic information on gonorrhea cases in Amsterdam identified large clusters of individuals with related strains, individuals infected with different strains at different anatomical sites, and persons with high rates of re-infection (15). The results suggested that the transmission networks for men who have sex with men and for heterosexuals were essentially separate—a key public health insight for planning interventions.

Increase understanding of the epidemiology of infectious diseases

While the contributions of molecular tools to outbreak investigation and surveillance have been substantial, there have also been significant contributions to our understanding of the epidemiology of infectious diseases. Molecular tools enable us to trace the dissemination of a particular subtype across time and space, and thus develop theories of transmission and dissemination; determine the origin of an epidemic, and therefore test theories about reservoirs and evolution of a particular agent; follow the emergence of new infections as they cross species, testing our hypotheses about the apparent transmissibility and rate of evolution; and follow mobile genetic elements conferring antimicrobial resistance or virulence between strains within a species or between species, and so develop theories about evolution and transmission within the populations of infectious agents.

Tracing the dissemination of infectious agents across time and space

Infectious agents are constantly emerging and re-emerging. Some agents, like influenza, have a well-understood pattern, where new strains generally emerge from southeastern Asia. This allows not only set up of sentinel surveillance points, but prediction, with some accuracy, of which influenza strain type(s) are most likely to cause the next epidemic. As the genetics of influenza is fairly well-understood, appropriate vaccines for known variants can be prepared. The difficultly is when the virus undergoes an antigenic shift. At this writing, an influenza A strain

Table 23.4. Ways molecular tools increase understanding of the epidemiology of infectious diseases

• Tracing the dissemination of infectious agents across time and space
• Determine the origin of an epidemic
• Follow emergence of new infections
• Follow mobile genetic elements conferring virulence of antimicrobial resistance

Table 23.5. Applications of evolutionary theory to infectious disease epidemiology

• Identify genetic lineages
• Estimate rate of evolution
• Generate theories about the emergence and maintenance of specific lineages of infectious agents

(H5N1, also known as avian or bird influenza) has repeatedly caused human infection with very high case fatality rates (~50%), although the chains of transmission have been relatively short and the total number of cases is relatively small. However, there is an ongoing widespread epidemic among wild birds, and there have been several outbreaks among domestic birds, resulting in large-scale culling of birds and considerable adverse economic impact.

Other infectious agents have hit by surprise, such as the emergence of HIV, and the migration of West Nile virus to the United States. Further, some infectious agents have mutated in unpredicted ways, such as the emergence of multidrug-resistant tuberculosis, penicillin-resistant *Streptococcus pneumoniae,* and community-acquired methicillin resistant *S. aureus*. In addition to understanding the transmission system, the origin and source of entry of infectious agents into the population must be traceable to prevent and control the spread of infection. By comparing strains, it can be determined if there has been single or multiple points of entry, and if emerging resistance was from multiple spontaneous mutations or from dissemination of a single clone. For example, until 2004, only occasional isolates of gonorrhea found in Sweden were resistant to azithromycin, and these cases were attributed to acquisition elsewhere (16). However, in 2004, epidemiologic evidence suggested that domestic transmission might have occurred; this was confirmed by molecular typing. The ongoing transmission of the azithromycin-resistant strain in Sweden has short-term implications for surveillance and long-term implications for treatment recommendations.

Streptococcus pneumoniae (S. pneumonia) is a major cause of pneumonia, but also causes meningitis and otitis media. A major human pathogen, it is one of the most common indications for antibiotic use. Resistance to penicillin emerged relatively slowly, but once it emerged it was widely disseminated in relatively few clones as defined by multilocus sequence typing. By contrast, the recent emergence of *S. pneumonia* resistant to fluoroquinolones has been due to a diverse set of genetic mutations (17), suggesting spontaneous emergence following treatment. As *S. pneumonia* resistant to fluoroquinolones rapidly followed the introduction of fluoroquinolones, alternative antibiotics will be needed in relatively short order to treat S. pneumonia infections.

Determine the origin of an epidemic

Molecular tools enable us to trace an outbreak or epidemic back in time to its origin, and back in space to its reservoir. Knowing the origin in time is essential for predicting future spread and identifying the reservoir for infection is central for controlling disease spread. The use of molecular techniques has solved long-standing mysteries, such as cholera's reservoir betweencholera epidemics. The same strains of cholera that infect humans also thrive in aquatic environments (19). While the importance of pigs and fowl as the origin of antigenic shifts in the genetics of influenza is understood, molecular tools have clarified that avian influenza need not first pass through the pig before jumping to humans, and that direct transmission from birds to humans is often more virulent (20). Molecular tools can also provide insight into the origins of infection in highly endemic populations, such as hospitals. The prevalence of methicillin resistant *Staphylococcus aureus* (MRSA) has been steadily increasing in hospitals in the United States; in 2004 the prevalence among some intensive care units was as high as 68% (21). However, in the early 2000s, new strains of MRSA emerged among individuals in the community that could not be traced back to hospitals. Genetic typing of the strains confirmed that strains

isolated from those who had no epidemiologic linkage with hospitals had genotypically different strains (Figure 23.2) (18). More recently, community-acquired MRSA has joined hospital-acquired strains in causing infection in hospital settings (22).

Emergence of new infectious agents

Surveillance, outbreak investigation, sentinel networks, and the astute healthcare worker are keystones for identifying the presence of new disease syndromes. While a clearly defined clinical syndrome facilitates epidemiologic investigation, the potential for misclassification and associated bias can be high for non-specific syndromes. Molecular tools, such as non-culture techniques, have dramatically improved our ability to rapidly identify the etiologic agent and develop diagnostic tools. In addition, detection of the agent improves our ability to predict transmission routes, and identify potential therapies and prevention strategies by analogy to similar organisms.

Severe acute respiratory syndrome (SARS) was the first emerging disease identified this century. The story of the rapid isolation, identification, and sequencing of the coronavirus causing SARS, is illustrative of the synergistic effects of the marriage of molecular methods with epidemiology. SARS was first reported in southern China in 2002 and rapidly spread worldwide (Figure 23.3). Basic epidemiologic methods were essential for tracking the outbreak; a carefully collected epidemiologic case definition was sufficient for case ascertainment, clinical management, infection control, and identifying chains of transmission (23). However, key to characterizing and ultimately preventing and controlling the outbreak was the ability to detect mild cases and confirm that widely disseminated cases were caused by the same agent, which required a validated antibody test (24). Early in the epidemic there were many possible candidates identified as the causative agent, but these agents were not found in all SARS patients. A variety of state-of-the-art and standard molecular techniques were used to identify the viral agent, a new coronavirus. Molecular techniques established that the genetic sequences were the same throughout the world, and a rapidly developed test demonstrated that SARS patients had antibodies to the new coronavirus. Further, healthy controls not having SARS had no evidence of either past or present infection (25).

Figure 23.2. Pulsed-field gel electrophoresis pattern relatedness of community-associated and health care-associated methicillin resistant Staphylococcus aureus (MRSA) isolates to a reference strain. The reference strain was MR14, which was the most commonly identified pattern among Minnesota MRSA isolates with a community-associated case definition (18).

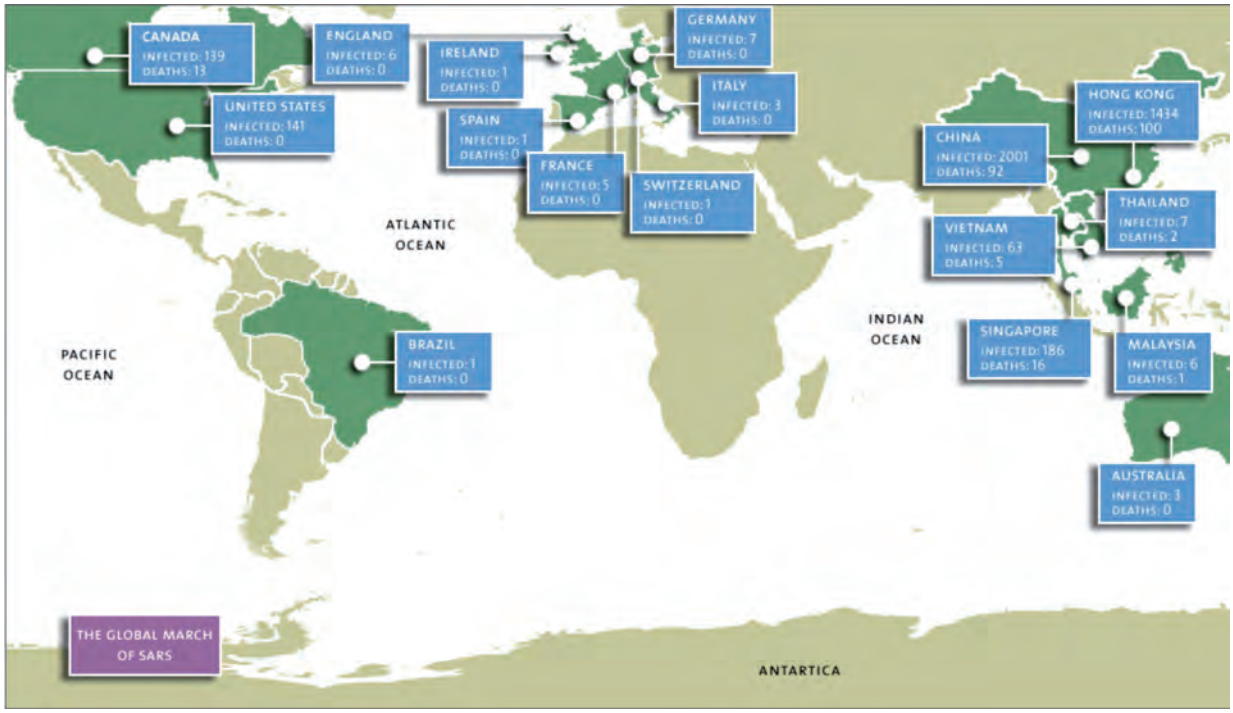
Figure not available

Trace mobile genetic elements

Mobile genetic elements are sequences of genetic material that can change places on a chromosome, and be exchanged between chromosomes, between bacteria, and even between species. A type of mobile genetic element, known as a plasmid, can integrate directly into the chromosome or extra-chromosomal in the cytoplasm of bacteria and still code for proteins. The recognition of mobile genetic elements, and the ability to trace these genetic elements as they move within and between species, has caused a re-thinking of the rate of, and potential for, evolution of infectious agents. For example, Shiga toxin-producing *E. coli* probably emerged from the transfer of genes coding for Shiga toxin from *Shigella* into *E. coli.*

Antibiotic resistance is often spread via a mobile genetic element. These elements tend to code for genes providing resistance against multiple antibiotics. This explains

Figure 23.3. The rapid dissemination of severe acute respiratory syndrome (SARS) (http://yaleglobal.yale.edu/reports/images/SARS_MAP1.jpg, permission given by Yale Center for the Study of Globalization and YaleGlobal Online).



several apparent mysteries, such as the spread across several bacterial species within a hospital of the same antibiotic resistance profile, and why treating an individual with one antibiotic can result in resistance to multiple unrelated antibiotics.

Determine phylogenetic relationships

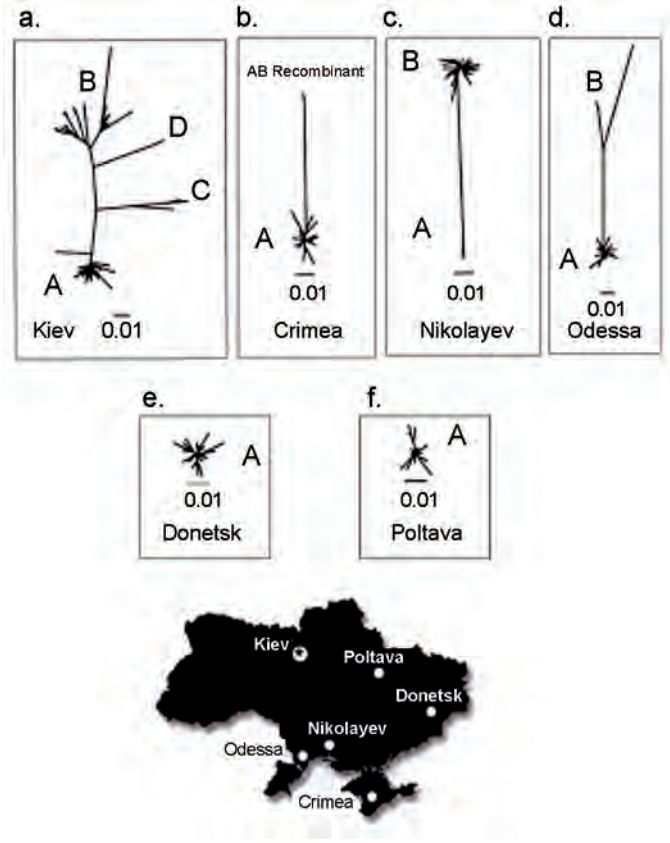
Genetic sequence and other molecular typing methods enable the construction of phylogenetic trees. Phylogenetics enables the use of evolutionary theory to explain epidemiologic phenomena, particularly emergence and transmission of more (or less) virulent strains, strains resistant to antimicrobials, simply to trace the transmission of a rapidly evolving species, or, in an outbreak situation, determine order of transmission. Separate phylogenies can be constructed for mobile genetic elements, or conserved elements on the chromosome.

Human immunodeficiency virus (HIV), which causes acquired immunodeficiency syndrome (AIDS), evolves quite rapidly even with a single host. Thus, the strain that infects an individual is not genetically identical to the strains that the individual might transmit to others. This property of HIV has made it possible to confirm the deliberate infection of one individual by another using a single blood sample from an individual (9) and to gain insight into the origin and spread of HIV worldwide.

There are three primary applications of phylogenic analyses in an epidemiologic context. First, phylogenic analysis enables us to determine genetic lineages. Using this type of analysis, researchers traced the introduction and spread of HIV in the Ukraine (Figure 23.4) (26). They were able to demonstrate that two subtypes introduced into drug networks in the 1990s still contributed to the epidemic in 2001 and 2002, and that one subtype spread widely throughout the Ukraine and into the Russian Federation, the Republic of Moldova, Georgia, Uzbekistan and Kyrgyzstan. Further studies to determine the biologic and social contributions to the success of the one subtype over another will provide important insights into how to control HIV.

A second application is to determine the rate of evolution. This is a standard application in biology, but understanding how fast infectious agents evolve has profound implications for choosing a molecular typing technique and interpreting epidemiologic data. For example, some agents change very rapidly, so that the agent infecting an individual is different from the agent that is transmitted to another, for

Figure 23.4. Phylogenetic analysis of strains from different cities in Ukraine. Phylogenetic trees of strains from Kiev (a), Crimea (b), Nikolayev (c), Odessa (d), Donetsk (e), and Poltava (f). A, B, C, D (capital letters) shown on tree branches are the HIV-1 subtypes. A scale bar of 0.01 substitutions per site is shown under each tree (26). The publisher for this copyrighted material is Mary Ann Liebert, Inc. publishers.



example HIV. Other agents change very slowly, such as tuberculosis. Thus, the appropriate typing technique must be chosen, so that phylogeny can be used to determine if rapidly changing agents evolved from a common ancestor, and to be able to distinguish between slowly evolving isolates. A typing technique for a rapidly evolving agent might focus on a region of the genome that evolves relatively slowly; a typing technique for a slowly evolving agent might focus on a genetic region that evolves fairly quickly, so that the investigator can distinguish between outbreak and non-outbreak strains.

A final application is to generate theories about the emergence and maintenance of specific genetic lineages. This application is a by-product of studies of genetic lineages and the rate of evolution. A key insight from the study in the Ukraine was the differential spread of different HIV subtypes (26). A study of the molecular epidemiology of norovirus outbreaks in Norway demonstrated the emergence of a new variant that accounted for a change in both the seasonal distribution and common transmission mode (27). A next step for furthering our understanding of the epidemiology of HIV and norovirus would be to generate theories to explain these phenomena.

Test hypotheses about transmission systems

Applying molecular typing to ongoing or endemic disease transmission increases our understanding of how contact patterns produce observed patterns of disease, revealing novel prevention and control strategies. In addition to characterizing ongoing chains of transmission, molecular typing can clarify who had contact with whom, and who was the source of infection, and thus identify a transmission network. Identifying transmission networks provides essential information for targeting intervention programmes, particularly when designing and implementing vaccine programmes.

Using polymerase chain reaction (PCR)-restriction length polymorphism typing of the porin and opacity genes of *Neisseria gonorrhoreae* and questionnaire data, a study of successive gonorrhea cases in Amsterdam identified several ongoing transmission chains. The epidemiologic characteristics, including number of sexual partners and choice of same or opposite partners of patients with different molecular types differed, suggesting that the transmission chains represented different transmission networks (15). Molecular typing has also improved our understanding of tuberculosis transmission. Until confirmed by molecular typing, tuberculosis was not believed to be transmitted by short-term casual contact. Several investigations have demonstrated that this assumption is incorrect, such as clusters associated with use of services at day shelters (28), and even linked to only a few brief visits to an infected individual's worksite (12). Molecular typing has also demonstrated linkage between apparently sporadic tuberculosis cases, and determined that at least some recurrent tuberculosis

is attributable to exogenous re-infection (reviewed by (29)).

Identify agent characteristics that lead to transmission and pathogenesis

The Microbial Genome Program of the US Department of Energy has sequenced more than 500 microbial genomes (http://microbialgenomics.energy.gov/brochure.pdf); the genetic sequence of numerous human pathogens have already been published, and many more are ongoing, as well as experiments to compare the sequences of other strains to a sequenced strain. A great deal can be learned from sequence data; of particular interest here is the identification of new open reading frames (ORF) which correspond to gene sequences. Although inferences can be made about a particular ORF based on the genetic sequence by comparing it to other gene sequences of known function, we cannot be certain of the gene's function or its importance to disease transmission or pathogenesis. However, conducting epidemiologic studies on appropriately collected samples can be done to increase understanding of the potential function of the genes and their relative prevalence using a molecular epidemiologic strategy (Table 23.6) (30). Many bacterial species are found in both diseased and healthy individuals and have highly diverse genomes. For example, *E. coli*, the most common cause of urinary tract infection and diarrhoea, is found in the normal bowel flora of virtually all humans and animals. When disease and commensal isolates are compared, the genome of *E. coli* is quite diverse, even when limited to human isolates. *E. coli* O157:H7, which causes diarrhoea and haemolytic urea syndrome, is substantially different in genetic content from the well-studied *E. coli* K12 strain, which was originally isolated from human feces in 1922 (http://www.sgm.ac.uk/pubs/micro_today/pdf/080402.pdf). Epidemiologic studies can take advantage of this variation in genetic content to compare the frequency of a putative virulence factor present among strains isolated from individuals with a specific pathology with the frequency among commensal isolates. For infectious agents with less diverse genomes, studies can determine differences in genetic alleles or in gene expression.

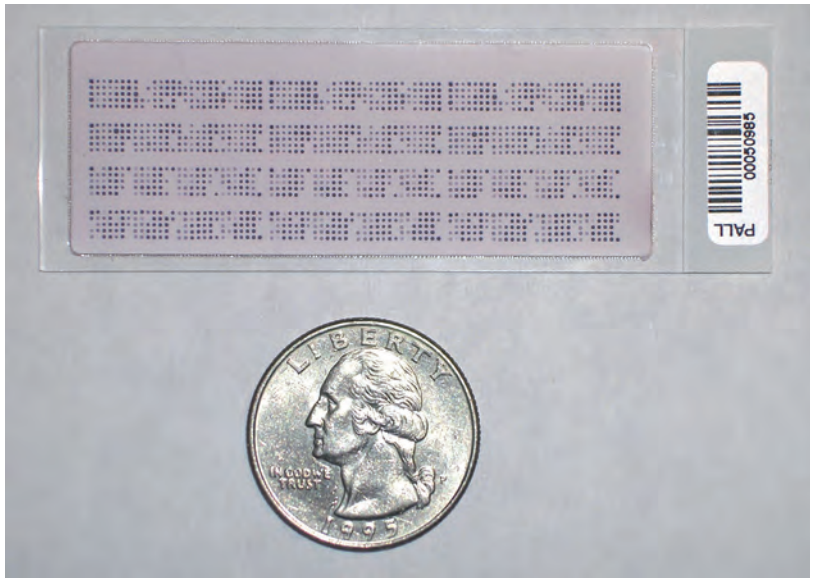
Table 23.6. Molecular epidemiologic strategy for gene discovery

• Identify candidate genes by combining bioinformatics information with molecular data
• Screen well-characterized representative samples of isolates causing different pathologies and asymptomatic infection
• Analyse to determine relative frequency of selected characteristics in various populations

Studies using a molecular epidemiologic strategy can be done using high-throughput methods, such as multiplex PCR or microarrays. For example, Library on a Slide is a novel microarray platform that enables the screening of thousands of bacterial isolates for the presence of a putative virulence gene in a single experiment. The genomes of up to 5000 bacterial isolates can be arrayed, in duplicate, on a single array, and screened for the presence or absence of a single gene using dot blot hybridization (Figure 23.5) (31). Library on a Slide has been created for *Mycobacterium tuberculosis, Haemaphilis influenza, E. coli, S. pneumoniae,* Group B Streptococcus and *Streptococcus mutans.*

The molecular epidemiologic strategy has been productively applied to many bacteria species. For example, a study screening collections of middle ear and throat non-typeable Haemaphilis influenzae isolates, a major cause of

Figure 23.5. Library on a slide microarray platform compared with a United States quarter. Each spot on the slide contains the total genomic DNA of a strain of *E. coli.* Photograph and slide by Dr. Lixin Zhang. Reprinted from (32), Copyright (2007), with permission from Elsevier.



otitis media, identified a gene found significantly more frequently among middle ear isolates, lic2B. lic2B was found 3.7 times more frequently among middle ear isolates than in throat isolates from children attending day care (33,34).

Identify infectious agents that are adapted to particular human hosts

Humans and infectious agents are extremely well adapted to each other. Some infectious agents are essential to human development, such as for digestion of foods and production of nutrients such as vitamins (35). Thus, it is not surprising that recent evidence suggests that normal microbiota co-evolved with their human hosts (36). Further, several studies suggest certain pathogenic species are adapted to certain human populations, and that infectious agents may select for mutations in human lineages.

Evidence that a particular genetic lineage of a pathogen may be better adapted to certain human populations, has been gathered by comparing the epidemiology of specific lineages in human populations of mixed genetic origin. For example, HPV variants of African origin persist longer among African American women than among white women, and European variants persist longer in white women than among African American women (37). A similar association between the infected host's region of origin and the infecting strain has been observed for tuberculosis (38). Social and behavioural factors have been associated with acquisition and persistence of each of these infectious agents, but tangible evidence of host susceptibility tied to specific agent characteristics implies that strategies based on the identification and development of more effective and specific therapies and prevention strategies may be more successful than attempts at changing human behaviour.

Infectious agents may also contribute to the evolution of humans. A well-documented case is the impact of malaria on the human host: there are several different genetic variations that protect against malaria. These variations are found in countries that either currently or in the past had endemic malaria. The effectiveness of the human genetic variant at reducing malarial disease varies with *Plasmodium* species. The most well-known variant is the sickle cell trait, but there are others, such as the Duffy blood group. When the Duffy blood group is absent, *Plasmodium vivax* is unable to enter the red blood cells (39). Using different molecular techniques, other human adaptations can be identified. These may provide insight into potential therapeutics, such as understanding the role of CCR5 in blocking HIV, or generate theories to explain observed human variation.

Identify new infectious agents causing disease

Epidemiologic studies have often suggested a possible infectious origin for a clinical syndrome. However, our ability to detect the etiologic agent has been hampered by our inability to culture most infectious agents. The development of the polymerase chain reaction (PCR) has dramatically increased our ability to detect infectious agents, both those cultivable and uncultivable. PCR has been essential to such public health triumphs as the development of a vaccine against HPV (the primary cause of cervical cancer), the identification of human herpes virus 8 as the infectious cause of Kaposi sarcoma, and the rapid identification of the coronavirus as the cause of SARS.

The development of a vaccine against cervical cancer was a direct result of our ability to use molecular tools in an epidemiologic context. Cervical cancer has an epidemiology that strongly suggests a sexually transmitted infection. The disease is associated with a greater lifetime number of sex partners, early age at first intercourse, and history of a sexually transmitted infection. The precursors of cervical cancer (cervical dysplasia detectable via PAP smear), were studied extensively, but widespread misclassification obscured the results. Cervical infection with different HPV types have different propensities to progress to cervical cancer, but the clinical presentation at the initial stages of infection, cervical dysplasia, is indistinguishable among types. It was not until the tools were available to identify HPV and to determine the different HPV types that the epidemiology was truly understood and an effective vaccine developed (40).

The Kaposi sarcoma (KS) story demonstrates the potential of combining an exquisitely sensitive molecular detection technique with epidemiologic study design. Using representational difference analysis to identify DNA sequences present in KS lesions but absent or present in low copy number in non-diseased tissue obtained from the same patient, researchers identified non-human DNA sequences in KS lesions of HIV patients (41). The sequences were determined to be herpes-like, subsequently designated human herpes virus 8. A randomized, blind, evaluation of tissue from patients with KS of different origin was then performed: AIDS-associated, classic, and

among homosexual men who were HIV-seronegative (42). This confirmed that the DNA sequences were present in all types of KS, suggesting that the sequences were not found only among AIDS patients. Seroepidemiology studies confirmed that seroprevalence was correlated with risk of KS, and that seroconversion and seropositivity predicted development of KS (43).

The ability to detect non-culturable infectious agents provides new strategies to more rapidly identify emerging infections. A surveillance system has been established in the United States to identify the infectious components of unexplained deaths and critical illnesses possibly due to infectious causes (http://www.cdc.gov/ncidod/eid/vol8no2/01-0165.htm). This system enabled the rapid detection of West Nile Virus encephalitis when it first appeared in the United States (44).

Identify infectious agents involved in the initiation and promotion of chronic disease

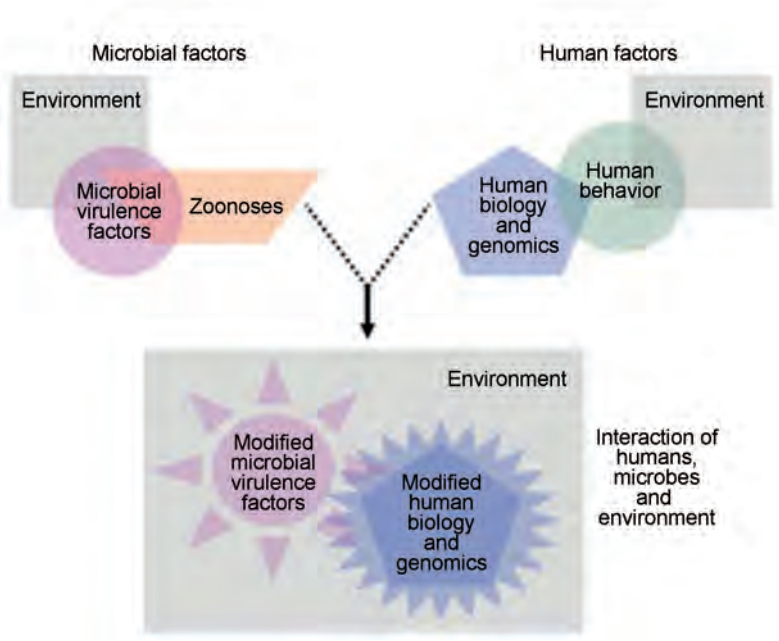
Infectious agents are popularly associated with acute disease processes, but many well-studied infectious processes lead to chronic diseases, such as tuberculosis and AIDS. However, other diseases that were previously attributed to genetics, behavioural, or lifestyle factors are now known to have an infectious component, including stomach ulcers, chronic liver disease, and arthritis (Table 23.7). Pathogenesis occurs at the infectious-chronic axis complex (Figure 23.6), and includes interactions between the agent, host and environment. Some of these diseases result from molecular mimicry, that is, the infectious agent has epitopes so similar to the host that the host response attacks itself, such as in reactive arthritis or rheumatic fever following infection.

Table 23.7. Selected infectious causes of chronic diseases

Chronic disease	Infectious agent
Arthritis	*Borrelia burgdorferi,* Epstein-Barr Virus, *Salmonella spp, Campylobacter spp, Yersinia spp, Chlamydia spp*
Bladder cancer	*Schistosoma spp.*
Cervical, anal, penile, head and neck cancers	Human papillomavirus
Chronic liver diseases, hepatocellular carcinoma	Hepatitis B, Hepatitis C
Creutzfeldt-Jakob disease	Variant Creutzfeldt-Jakob disease
Gastric cancer	*Helicobacter pylori*
Heart disease	*Chlamydia pneumoniae*
Kaposi sarcoma	Human herpes virus 8
Leukemia	Human T-lymphotropic virus type 1
Lymphoma	Epstein-Barr virus, Human T-lymphotropic virus type 1
Peptic ulcer disease, chronic gastritis	*Helicobacter pylori*
Whipple disease	*Tropheryma whipplei*

Modified and expanded from (45).

Figure 23.6. Schematic showing how multiple factors interact leading to chronic sequelae of infectious diseases (45).



Alternatively, there may be disease that results from a host primed for infection that does not happen.

Molecular tools have definitively demonstrated that there are infectious causes of cancer: hepatitis B and C can cause liver cancer; serotypes of human papillomavirus cause cervical, anal, penile, and head and neck cancers; and herpes virus 8 causes KS (45). Infectious diseases, such as *Chlamydophila pneumoniae (C. pneumoniae)*, are hypothesized to lead to the promotion of atherosclerotic plaques and thus coronary artery disease. Numerous studies have demonstrated C. pneumoniae in atherosclerotic tissue, and severity of disease has been positively associated with antibodies to *C. pneumoniae* after adjustment for other known risk factors. However, results of antibiotic therapy in preventing progression of cardiovascular disease among those who already have disease have been disappointing. The totality of the evidence suggests that C. pneumoniae is neither a necessary nor sufficient cause of coronary artery disease, but is likely a modifiable risk factor (46).

Determining an infectious cause of a chronic disease is difficult: active infection often has ceased by the time the chronic disease is manifest. Disease clusters may be quite informative: Lyme disease was identified as a cause of juvenile arthritis because of careful epidemiologic investigation of a disease cluster (47). As the disease process is not solely a function of the presence of the infectious agent, but an interaction of the agent with the host, it is as likely to detect the presence of specific genes associated with the disease as the infectious agent. Moreover, there is evidence that infectious agents can incorporate their genetic material into the human genome: human endogenous retroviruses are the remnants of ancient germ cell infections (48). However, as we increase our ability to detect specific host response to infectious agents and detect traces of infection within the host, we will likely increasingly detect infectious causes of many diseases of unknown etiology.

Guide clinical treatment and intervention strategies

Knowledge of the molecular genetics of infectious agents and the interaction of the infectious agent with the host gained from molecular epidemiologic studies can be used to more rapidly detect infectious agents and thus improve patient diagnosis, predict disease course and identify potential vaccine candidates. Molecular techniques can also be applied to characterize the ecology of normal human flora, detect disruptions in the flora that lead to disease, and to detect the presence of biofilms (microbial structures which often contain multiple species that can initiate or promote disease or protect the host from disease) (Table 23.8).

Rapid detection of infectious agents

Increasingly, there are rapid methods for the detection of infectious agents. These methods have profound implications for public health practice, clinical diagnosis and epidemiologic studies. Rapid detection methods are based on either PCR, the amplification of genetic sequences of the infectious agent that can identify the agent present, or an antigen-antibody reaction that detects the host response to the infectious agent or a metabolite of the agent. Rapid detection means faster and more accurate diagnosis. For example, intrapartum prophylaxis with antibiotics has reduced by 50% the incidence of neonatal Group B streptococcal (GBS) disease. However, GBS colonization is often transient, so many women may be treated unnecessarily. Detection of GBS colonization at time of labour and delivery would minimize inappropriate antibiotic use, decreasing unnecessary pressure for the development of antibiotic resistance. Further, GBS is increasingly resistant to the second-line antibiotics used for women sensitive to the first-choice antibiotic. Rapid detection of antibiotic resistance, based on the detection of resistance genes and the ability to discriminate between more virulent GBS strains, will potentially improve medical care.

Rapid techniques are often extremely sensitive, so when applied in an epidemiologic context

Table 23.8. Using molecular tools in a clinical epidemiologic context

• Rapidly detect infectious agents
• Distinguish between pathogens and commensals
• Distinguish between relapse and re-infection
• Predict disease course
• Evaluate potential vaccine candidates
• Guide intervention

may be cost saving. For example, epidemiologic studies of colonization with MRSA, which has emerged as a community-acquired pathogen of some significance, can be screened for using rapid techniques; only specimens screening positive by rapid methods might be cultured. Culture is not only more time-consuming, but costly in terms of reagents and personnel. However, the ability to propagate an infectious agent is highly desirable, as it facilitates more detailed studies at the molecular level.

Distinguish between commensals and pathogens

Increasingly, it is recognized that many species of infectious agents previously thought to be harmless commensals can, under certain circumstances, cause disease (49). For example, fungal infections are only a problem among immune-suppressed patients. Acquired immunosuppression can result either from medical therapy, such as chemotherapy for cancer or immunosuppressive drugs prescribed to transplant patients, or as a result of infection, such as HIV. It has been discovered that all strains within a species do not have equal disease potential; indeed, an opportunistic infection may arise because of special characteristics of the infectious agent itself. This makes the identification of the causal agent in the laboratory difficult, as there are cases where basic clues such as quantity of the agent or agent type may be insufficient to identify the cause. Molecular tools can be used to identify and characterize the specific virulence potential, as well as identify humans particularly susceptible. This ability should eventually translate into improved laboratory tests, making it much easier for the laboratorian to determine the causal agent and the physician to prescribe appropriate therapy.

Distinguish between relapse and re-infection

Some infections have a chronic, recurring nature. In this situation, it is extremely useful to distinguish between a relapse, which implies treatment failure, and a new infection. Strain typing can be extremely useful in this situation, as typing allows us to distinguish between strains of the same species. For example, molecular typing demonstrated that individuals can be infected with more than one strain of tuberculosis (50) and of HIV (51).

Predict disease course

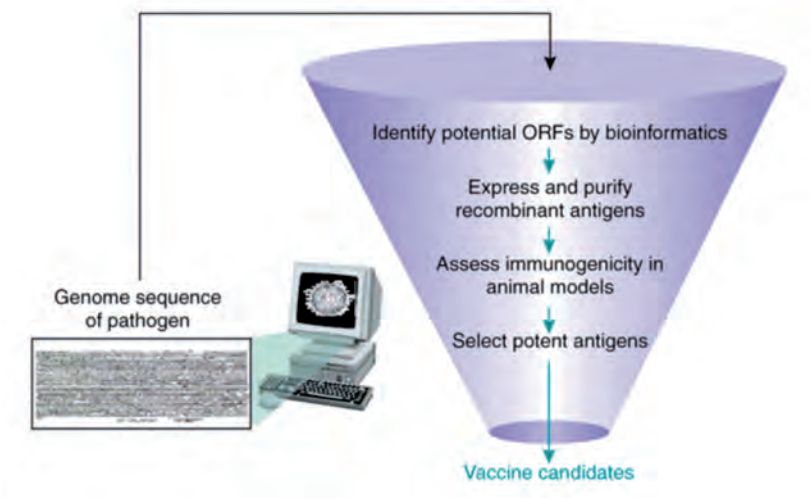
Disease course is a function of both host and agent factors. While an individual who receives a larger infectious dose of an infectious agent is, on average, more likely to become ill and to manifest symptoms more rapidly, this may not always be the case. The virulence of the infectious agent, whether the host has had previous exposure to the same or a similar strain of the infectious agent, or if the host has an underlying genetic predisposition or presence of predisposing factors, such as co-morbidities, all influence the infectious course. For example, initial viral load in HIV patients has been demonstrated as a good predictor of disease prognosis, as well as potential to transmit to others (52). Similar predictors for other infections are sure to follow.

Identify potential vaccine candidates

Early microbiology led to the development of several vaccines that have dramatically improved public health worldwide: smallpox has been eradicated, and measles and polio are largely under control in developed countries. Tetanus, influenza, diphtheria, mumps, rubella, chickenpox, hepatitis A and B, and yellow fever can be prevented, and the Bacille Calmette–Guérin vaccine for tuberculosis minimizes the most adverse manifestations of tuberculosis in children. Most recently, a vaccine against the HPV serotypes 16 and 18, which are most likely to cause cancer, was licensed. Nonetheless, many other infectious diseases that cause significant morbidity and mortality have remained intractable to prevention via vaccine using conventional strategies, but not from lack of trying. Some of these infectious agents, such as the bacteria that cause gonorrhea and bacterial meningitis, *Neisseria gonorrhoeae* and *Neisseria menigitidis*, can rapidly vary their surface antigens (53,54) making it difficult to identify an appropriate target.

The human pathogen sequencing project has resulted in a new strategy for the identification of vaccine candidates based on the predicted protein products based on the genetic sequence (Figure 23.7) (55). *In silico* analyses using bioinformatics enable the detection of potential epitopes. Many species are very diverse at the genetic level, thus not only must appropriate epitopes be identified, but epitopes that are found across the range of potentially diverse members of a particular species. Thus, epidemiologic screening of population-based samples of the species of interest for the presence of candidate epitopes can assist in selecting between potential candidates by ruling out those with limited geographic distribution (56).

Figure 23.7. Reverse vaccinology for identification of novel vaccine antigens (55). Reprinted by permission from Macmillan Publishers Ltd: Nature Biotechnology, copyright (2006).



Guide intervention strategies

The ability of molecular tools to detect asymptomatic infection not only increases our understanding of the transmission system, but also has implications for clinical practice. For example, using culture, the protozoan parasite *Trichomonas vaginalis (T. vaginalis)* is detected in only 8–20% of male partners of women infected with *T. vaginalis;* testing urine with PCR detects *T. vaginalis* in up to 70% of partners (57). This has profound implications for preventing the spread of this common infection, suggesting that routine PCR testing of sex partners is in order.

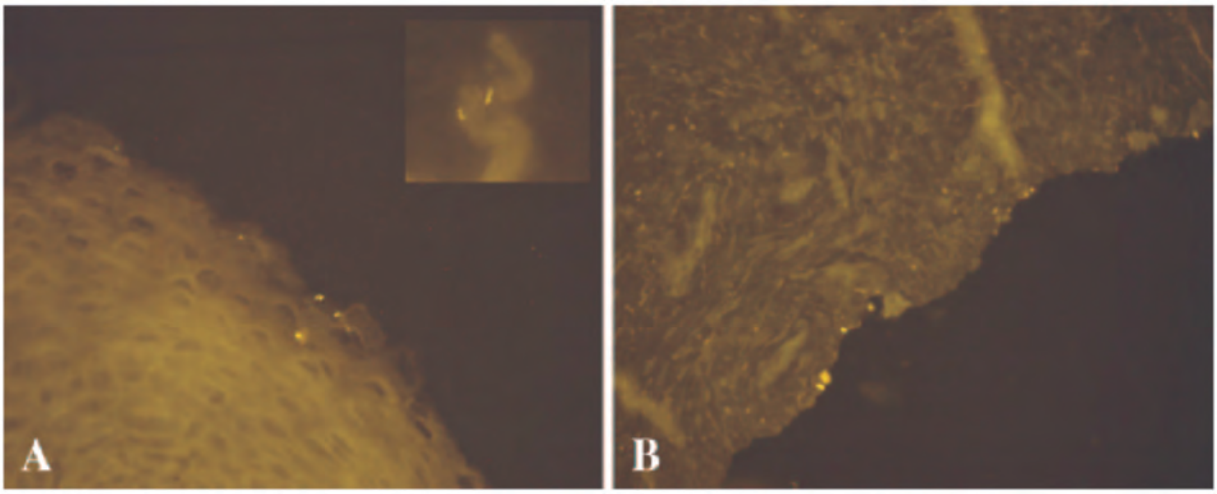
Detecting the emergence and spread of resistance to therapy should lead to changes in clinical practice. Resistance genes are often carried on the same mobile genetic elements, so that resistance to one drug often implies resistance to others. Further, not only does treating an individual with antibiotics select for resistant organisms within that host, it also increases risk of acquiring resistant organisms in their contacts (58). A better understanding of the transmission of resistance genes between bacteria, and of resistant bacteria between individuals, will aid in designing effective policies. New therapies or combination therapies may be introduced or steps taken to minimize the spread of resistance.

Polymicrobial infections and interactions

Many diseases result from infection not with a single infectious agent, but as a result of a change in the microbial community that results from microbial activities (35). Microbes have certain nutrient and other requirements, but as they grow and die, they themselves serve as a source of nutrients for additional microorganisms. For example, an upper respiratory infection caused by a virus can enhance the environment for bacterial growth. Thus, upper respiratory infections are often a precursor to otitis media or bacterial pneumonia; influenza vaccination of children reduces rates of otitis media. Further, by-products of microbial growth can create local variations in pH, and in the presence of oxygen result in the growth of biofilms that enable colonization by other infectious agents. Biofilms are complex, often polymicrobial structures formed on a variety of surfaces in the environment and human body. The scum that forms on the insides of water pipes is a biofilm, but so is plaque on teeth, and the slimy surface on the tongue, inside the nose and throat, in the vaginal cavity and on other surfaces. There are also natural synergies or anergies between organisms of the same or different species; for example, the anergy between *S. pneumoniae* and *S. aureus* colonization in the nasal cavity (59). There are also synergies, such as is observed in the vaginal flora with lactobacillus modifying the pH and enabling the growth of other species.

PCR and high-throughput sequencing have enabled the description of the complex microbiota found on and in the human body. These studies use the fact that all cells have a ribosome, and that the sequence of genes that code for the ribosome can be used for taxonomy (61). These techniques have a great advantage over detection by culture, as culture requires at least a rough idea of what organisms might be present and their growth requirements. For example, applications of non-culture techniques to microorganisms in the human gut suggest that as many as 93% of the rRNA sequences identified are from uncultured organisms (62). Most non-culture techniques are based on some type of PCR and detect highly-conserved genes, such as those coding for 16sRNA, which vary at the species level and are semiquantitative. Others involve *in situ* hybridization techniques, enabling both detection

Figure 23.8. Vaginal epithelium from a healthy premenopausal woman hybridized with a universal probe (x400) and lactobacillus probe (inset x1000). Only a small number of bacteria are scattered over the surface of intact epithelium (A). Long rods can be seen with high magnification (inset). Bacteria are found in similar concentrations on the subepithelial surface of the biopsy that was exposed by mechanical trauma of the tissue (B) (60).



of organism presence and visualization of structure, such as vaginal epithelium shown in Figure 23.8 (60). There is much to learn about what constitutes normal flora; the dynamics of colonization; how colonization varies with normal biovariations such as the menstrual cycle, pregnancy and aging; and in the face of antibiotic therapy and disease.

Implications of using molecular tools for the design, conduct and analysis of epidemiologic studies

Modern molecular techniques combined with epidemiologic methods allow us to identify novel methods of disease prevention and control, markers of disease diagnosis and prognosis, and fertile research areas for potential new therapeutics and/or vaccines. However, the success of these studies depends not only on the molecular measure chosen, but also on whether the strengths and limitations of the chosen measure are considered in the design, conduct, analysis, and interpretation of the study results (Table 23.9).

Table 23.9. Impact of using molecular tools on the design, conduct and interpretation of epidemiologic studies

- Study design: design choice, sampling
- Conduct: specimen collection and handling
- Analysis: translating laboratory measures to interpretable variables
- Interpretation: limitations of measures

Some molecular measures are relatively invariate with time, such as human genes. Studies associating genetic susceptibility to an infectious agent might be conducted using genetic material collected long before or after the disease occurred. By contrast, studies of host response to infection must be collected within a fairly tight time frame. Antibodies to an infectious agent may not appear until a defined period after infection, such as for HIV, or the infectious agent may be present only for a short duration, such as for *Streptococcus agalactiae*. Thus, some studies might be nested in large cohort studies or conducted using a case–control technique, while others require a prospective design.

The requirements of the molecular tool also affect sampling. For example, if a test must be conducted on fresh samples, the sampling of cases and controls in a case–control study should be done so that the groups are sampled and tested in similar time periods to minimize potential biases resulting from assay drift, which is where a method gives increasingly higher or lower results with time. For nested case–control studies, how specimens are collected may determine whether controls can be sampled from the base population (case-based, also called case–cohort, sampling) or at time

of incidence disease (incidence density sampling) (63).

The conduct of the study must take into account the requirements of molecular testing. Some tests are sensitive to freezing and thawing and must be tested immediately, while others degrade over time, even if stored properly. Multiple different strains of a single infectious agent might be isolated from one individual, such as from different body sites. Labelling should make it possible to identify the appropriate strain and link it back to the appropriate individual and isolation site. Further, infectious agents may be grown in different media, passed in multiple cultures that might change phenotypic characteristics, or a plasmid might be lost. Thus, noting the number of times an isolate is cultured and on what media is important.

Epidemiologic analyses often use simple cut-offs: e.g. diseased versus not diseased. Laboratory measures often are continuous, but the scales may be ordinal, that is, the differences between values are not consistent. The interpretation of the measures might vary with the study population or the presence of other ancillary information. For example, >100 000 of a single bacterial species in the urine of an asymptomatic, healthy, non-pregnant individual has no clinical significance. If the individual has symptoms referable to the urinary tract, such as urgency and frequency, the individual probably has a urinary tract infection (64). By contrast, >100 000 of a single bacterial species in the urine of an asymptomatic, healthy, pregnant woman, is a treatable condition, because of the increased risk of pyelonephritis due to physiologic changes that occur during pregnancy.

Conclusions and future challenges

The applications of molecular tools to the study of infectious disease are varied, including applications to public health practice, diagnostics, and understanding of the transmission, evolution and pathogenesis. To date, the major potentials have been explored using genomics, but applying the power of proteomics and transcriptomics to the understanding of disease transmission and pathogenesis and host-agent interactions will open new avenues to understanding.

Much remains to be learned about infectious agents. Some future challenges are listed in Table 23.10. One area that is particularly amenable to study using molecular techniques is the normal human flora or microbiota. Extremely little is known about normal human microbiota, its response to invasion by pathogens, and its response to therapeutic treatment. Disruptions of normal microbiota are associated with a variety of pathogenic syndromes, such as bacterial vaginosis, that put the affected host at increased risk of acquiring other, often more serious, infectious agents. Interactions between disrupted normal microbiota and the host may also be important in explaining chronic recurring infections. Relatively little is understood about the structures formed by microbes within the human body; there are also structures that microbes stimulate the human host to form, such as pedestals on which *E. coli* O157:H7 sit.

Another area to explore is the interaction between the host and the agent. With molecular tools it was possible to identify why some individuals are repeatedly exposed to HIV but do not develop disease: these individuals have a variant in their CCR5 receptor that makes it difficult for HIV to invade the cell (65). It is also possible to identify human genes that explain why some individuals infected with HIV do not progress to AIDS. For example, a genome-wide association study (GWAS) identified the *HCP5* gene of the HLA region in chromosome 6 (66). GWAS have been applied to hepatitis C virus to identify why the infection spontaneously resolves in some individuals and treatment of chronic disease only eradicates infection in 40% of cases (67). Other human genetic variants likely modify risk of infection and response to infection, both positively and negatively, for many other infectious agents.

We have only begun to explore these interactions, and many challenges, both technological and methodological, remain (68). By combining modern molecular tools with epidemiologic methods, we have a powerful means to understand host agent interactions—an understanding essential for us to learn to live peacefully with microbes within our bodies, which, after all, outnumber the human cells that comprise us.

Table 23.10. Future challenges in the study of infectious disease

• Normal flora
• Biofilm formation
• Host agent interactions
• Successive infection

References

1. Kaufmann SHE, Schaible UE (2005). 100th anniversary of Robert Koch's Nobel Prize for the discovery of the tubercle bacillus. *Trends Microbiol,* 13:469–475.doi:10.1016/j.tim.2005.08.003 PMID:16112578

2. Laveran A. The Nobel Prize in physiology or medicine 1907. In: Nobel lectures, physiology or medicine 1901–1921. Amsterdam: Elsevier Publishing Company; 1967 (http://nobelprize.org/nobel_prizes/medicine/laureates/1907/laveran-bio.html).

3. Nicolle C. The Nobel Prize in physiology or medicine 1928. In: Nobel lectures. Physiology or medicine 1922–1941. Amsterdam: Elsevier Publishing Company; 1965 (http://nobelprize.org/nobel_prizes/medicine/laureates/1928/nicolle-lecture.html).

4. Daniel TM, editor. Wade Hampton Frost, pioneer epidemiologist 1880–1938: up to the mountain. Rochester (NY): University of Rochester Press; 2004.

5. Kilbourne ED (1973). The molecular epidemiology of influenza. *J Infect Dis,* 127:478–487.doi:10.1093/infdis/127.4.478 PMID:4121053

6. Jones TF, Kellum ME, Porter SS *et al.* (2002). An outbreak of community-acquired foodborne illness caused by methicillin-resistant Staphylococcus aureus. *Emerg Infect Dis,* 8:82–84.doi:10.3201/eid0801.010174 PMID:11749755

7. Centers for Disease Control and Prevention (CDC) (1998). Multistate outbreak of Salmonella serotype Agona infections linked to toasted oats cereal–United States, April-May, 1998. *MMWR Morb Mortal Wkly Rep,* 47:462–464. PMID:9639368

8. Hutin YJF, Pool V, Cramer EH *et al.*; National Hepatitis A Investigation Team (1999). A multistate, foodborne outbreak of hepatitis A. *N Engl J Med,* 340:595–602.doi:10.1056/NEJM199902253400802 PMID:10029643

9. Metzker ML, Mindell DP, Liu XM *et al.* (2002). Molecular evidence of HIV-1 transmission in a criminal case. *Proc Natl Acad Sci USA,* 99:14292–14297.doi:10.1073/pnas.222522599 PMID:12388776

10. Mainous AG 3rd, Hueston WJ, Everett CJ, Diaz VA (2006). Nasal carriage of Staphylococcus aureus and methicillin-resistant S aureus in the United States, 2001–2002. *Ann Fam Med,* 4:132–137.doi:10.1370/afm.526 PMID:16569716

11. Swaminathan B, Barrett TJ, Hunter SB, Tauxe RV; CDC PulseNet Task Force (2001). PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. *Emerg Infect Dis,* 7:382–389. PMID:11384513

12. Golub JE, Cronin WA, Obasanjo OO *et al.* (2001). Transmission of Mycobacterium tuberculosis through casual contact with an infectious case. *Arch Intern Med,* 161:2254–2258.doi:10.1001/archinte.161.18.2254 PMID:11575983

13. Gieseler S, König B, König W, Backert S (2005). Strain-specific expression profiles of virulence genes in Helicobacter pylori during infection of gastric epithelial cells and granulocytes. *Microbes Infect,* 7:437–447.doi:10.1016/j.micinf.2004.11.018 PMID:15788154

14. Khan AI, Talukder KA, Huq S *et al.* (2006). Detection of intra-familial transmission of shigella infection using conventional serotyping and pulsed-field gel electrophoresis. *Epidemiol Infect,* 134:605–611.doi:10.1017/S0950268805005534 PMID:16288683

15. Kolader ME, Dukers NH, van der Bij AK *et al.* (2006). Molecular epidemiology of Neisseria gonorrhoeae in Amsterdam, The Netherlands, shows distinct heterosexual and homosexual networks. *J Clin Microbiol,* 44:2689–2697.doi:10.1128/JCM.02311-05 PMID:16891479

16. Lundbäck D, Fredlund H, Berglund T *et al.* (2006). Molecular epidemiology of Neisseria gonorrhoeae- identification of the first presumed Swedish transmission chain of an azithromycin-resistant strain. *J Clin Microbiol*, 114:67–71.doi:10.1111/j.1600-0463.2006.apm_332.x PMID:16499664

17. Doern GV, Richter SS, Miller A *et al.* (2005). Antimicrobial resistance among Streptococcus pneumoniae in the United States: have we begun to turn the corner on resistance to certain antimicrobial classes? *Clin Infect Dis,* 41:139–148.doi:10.1086/430906 PMID:15983908

18. Naimi TS, LeDell KH, Como-Sabetti K *et al.* (2003). Comparison of community- and health care-associated methicillin-resistant Staphylococcus aureus infection. *JAMA*, 290:2976–2984.doi:10.1001/jama.290.22.2976 PMID:14665659

19. Reidl J, Klose KE (2002). Vibrio cholerae and cholera: out of the water and into the host. *FEMS Microbiol Rev,* 26:125–139.doi:10.1111/j.1574-6976.2002.tb00605.x PMID:12069878

20. Van Reeth K (2007). Avian and swine influenza viruses: our current understanding of the zoonotic risk. *Vet Res,* 38:243–260.doi:10.1051/vetres:2006062 PMID:17257572

21. National Nosocomial Infections Surveillance System (2004). National Nosocomial Infections Surveillance (NNIS) System Report, data summary from January 1992 through June 2004, issued October 2004. *Am J Infect Control,* 32:470–485.doi:10.1016/j.ajic.2004.10.001 PMID:15573054

22. Sabol KE, Echevarria KL, Lewis JS 2nd (2006). Community-associated methicillin-resistant Staphylococcus aureus: new bug, old drugs. *Ann Pharmacother,* 40:1125–1133.doi:10.1345/aph.1G404 PMID:16735661

23. Weinstein RA (2004). Planning for epidemics–the lessons of SARS. *N Engl J Med,* 350:2332–2334.doi:10.1056/NEJMp048082 PMID:15175434

24. Berger A, Drosten Ch, Doerr HW *et al.* (2004). Severe acute respiratory syndrome (SARS)–paradigm of an emerging viral infection. *J Clin Virol,* 29:13–22.doi:10.1016/j.jcv.2003.09.011 PMID:14675864

25. Ksiazek TG, Erdman D, Goldsmith CS *et al.*; SARS Working Group (2003). A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med,* 348:1953–1966.doi:10.1056/NEJMoa030781 PMID:12690092

26. Saad MD, Shcherbinskaya AM, Nadai Y *et al.* (2006). Molecular epidemiology of HIV Type 1 in Ukraine: birthplace of an epidemic. *AIDS Res Hum Retroviruses,* 22:709–714.doi:10.1089/aid.2006.22.709 PMID:16910825

27. Vainio K, Myrmel M (2006). Molecular epidemiology of norovirus outbreaks in Norway during 2000 to 2005 and comparison of four norovirus real-time reverse transcriptase PCR assays. *J Clin Microbiol,* 44:3695–3702.doi:10.1128/JCM.00023-06 PMID:17021099

28. DeRiemer K, Daley CL (2004). Tuberculosis transmission based on molecular epidemiologic research. *Semin Respir Crit Care Med,* 25:297–306.doi:10. 1055/s-2004-829502 PMID:16088471

29. Mathema B, Kurepina NE, Bifani PJ, Kreiswirth BN (2006). Molecular epidemiology of tuberculosis: current insights. *Clin Microbiol Rev,* 19:658–685.doi:10.1128/CMR.00061-05 PMID:17041139

30. Zhang L, Foxman B, Manning SD *et al.* (2000). Molecular epidemiologic approaches to urinary tract infection gene discovery in uropathogenic Escherichia coli. *Infect Immun,* 68:2009–2015.doi:10.1128/IAI.68.4.2009-2015.2000 PMID:10722596

31. Zhang L, Srinivasan U, Marrs CF *et al.* (2004). Library on a slide for bacterial comparative genomics. *BMC Microbiol*, 4:12.doi:10.1186/1471-2180-4-12 PMID:15035675

32. Foxman B (2007). Contributions of molecular epidemiology to the understanding of infectious disease transmission, pathogenesis, and evolution. *Ann Epidemiol,* 17:148–156.doi:10.1016/j.annepidem.2006.09.004 PMID:17175168

33. Pettigrew MM, Foxman B, Marrs CF, Gilsdorf JR (2002). Identification of the lipooligosaccharide biosynthesis gene lic2B as a putative virulence factor in strains of nontypeable Haemophilus influenzae that cause otitis media. *Infect Immun,* 70:3551–3556.doi:10.1128/IAI.70.7.3551-3556.2002 PMID:12065495

34. Xie J, Juliao PC, Gilsdorf JR *et al.* (2006). Identification of new genetic regions more prevalent in nontypeable Haemophilus influenzae otitis media strains than in throat strains. *J Clin Microbiol,* 44:4316–4325. doi:10.1128/JCM.01331-06 PMID:17005745

35. Wilson M, editor. Microbial inhabitants of humans: their ecology and role in health and disease. New York (NY): Cambridge University Press; 2005.

36. Oh PL, Benson AK, Peterson DA *et al.* (2010). Diversification of the gut symbiont Lactobacillus reuteri as a result of host-driven evolution. *ISME J,* 4:377–387.doi:10.1038/ismej.2009.123 PMID:19924154

37. Xi LF, Kiviat NB, Hildesheim A *et al.* (2006). Human papillomavirus type 16 and 18 variants: race-related distribution and persistence. *J Natl Cancer Inst,* 98:1045–1052.doi:10.1093/jnci/djj297 PMID:16882941

38. Hirsh AE, Tsolaki AG, DeRiemer K *et al.* (2004). Stable association between strains of Mycobacterium tuberculosis and their human host populations. *Proc Natl Acad Sci USA,* 101:4871–4876.doi:10.1073/pnas.0305627101 PMID:15041743

39. Kwiatkowski DP (2005). How malaria has affected the human genome and what human genetics can teach us about malaria. *Am J Hum Genet,* 77:171–192.doi:10.1086/432519 PMID:16001361

40. Trottier H, Franco EL (2006). The epidemiology of genital human papillomavirus infection. *Vaccine*, 24 Suppl 1;S1–S15.doi:10.1016/j.vaccine.2005.09.054 PMID:16406226

41. Chang Y, Cesarman E, Pessin MS *et al.* (1994). Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science*, 266:1865–1869. doi:10.1126/science.7997879 PMID:7997879

42. Moore PS, Chang Y (1995). Detection of herpesvirus-like DNA sequences in Kaposi's sarcoma in patients with and without HIV infection. *N Engl J Med,* 332:1181–1185. doi:10.1056/NEJM199505043321801 PMID:7700310

43. Cathomas G (2000). Human herpes virus 8: a new virus discloses its face. *Virchows Arch,* 436:195–206.doi:10.1007/s004280050031 PMID:10782877

44. Hajjeh RA, Relman D, Cieslak PR *et al.* (2002). Surveillance for unexplained deaths and critical illnesses due to possibly infectious causes, United States, 1995–1998. *Emerg Infect Dis,* 8:145–153.doi:10.3201/eid0802.010165 PMID:11897065

45. O'Connor SM, Taylor CE, Hughes JM (2006). Emerging infectious determinants of chronic diseases. *Emerg Infect Dis,* 12:1051–1057. PMID:16836820

46. Watson C, Alp NJ (2008). Role of Chlamydia pneumoniae in atherosclerosis. *Clin Sci (Lond),* 114:509–531.doi:10.1042/CS20070298 PMID:18336368

47. Steere AC, Malawista SE, Snydman DR *et al.* (1977). Lyme arthritis: an epidemic of oligoarticular arthritis in children and adults in three connecticut communities. *Arthritis Rheum,* 20:7–17.doi:10.1002/art.1780200102 PMID:836338

48. Bannert N, Kurth R (2004). Retroelements and the human genome: new perspectives on an old relation. *Proc Natl Acad Sci USA,* 101 Suppl 2;14572–14579.doi:10.1073/pnas.0404838101 PMID:15310846

49. Casadevall A, Pirofski LA (2000). Host-pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease. *Infect Immun,* 68:6511–6518.doi:10.1128/IAI.68.12.6511-6518.2000 PMID:11083759

50. Warren RM, Victor TC, Streicher EM *et al.* (2004). Patients with active tuberculosis often have different strains in the same sputum specimen. *Am J Respir Crit Care Med,* 169:610–614.doi:10.1164/rccm.200305-714OC PMID:14701710

51. Janini LM, Tanuri A, Schechter M *et al.* (1998). Horizontal and vertical transmission of human immunodeficiency virus type 1 dual infections caused by viruses of subtypes B and C. *J Infect Dis, 1*77:227–231. doi:10.1086/517360 PMID:9419195

52. Pilcher CD, Tien HC, Eron JJ Jr *et al.*; Quest Study; Duke-UNC-Emory Acute HIV Consortium (2004). Brief but efficient: acute HIV infection and the sexual transmission of HIV. *J Infect Dis,* 189:1785–1792.doi:10.1086/386333 PMID:15122514

53. Seifert HS (1996). Questions about gonococcal pilus phase- and antigenic variation. *Mol Microbiol,* 21:433–440. doi:10.1111/j.1365-2958.1996.tb02552.x PMID:8866467

54. Davidsen T, Tønjum T (2006). Meningococcal genome dynamics. *Nat Rev Microbiol,* 4:11–22.doi:10.1038/nrmicro1324 PMID:16357857

55. Ulmer JB, Valley U, Rappuoli R (2006). Vaccine manufacturing: challenges and solutions. *Nat Biotechnol,* 24:1377–1383. doi:10.1038/nbt1261 PMID:17093488

56. Hebert AM, Talarico S, Yang D *et al.* (2007). DNA polymorphisms of Mycobacterium tuberculosis pepA and PPE18 genes among clinical strains: implications for vaccine efficacy. *Infect Immun,* 75:5798–5805. doi:10.1128/IAI.00335-07 PMID:17893137

57. Seña AC, Miller WC, Hobbs MM *et al.* (2007). Trichomonas vaginalis infection in male sexual partners: implications for diagnosis, treatment, and prevention. *Clin Infect Dis,* 44:13–22.doi:10.1086/511144 PMID:17143809

58. Samore MH, Lipsitch M, Alder SC *et al.* (2006). Mechanisms by which antibiotics promote dissemination of resistant pneumococci in human populations. *Am J Epidemiol,* 163:160–170.doi:10.1093/aje/kwj021 PMID:16319292

59. Teti G, Tomasello F, Chiofalo MS *et al.* (1987). Adherence of group B streptococci to adult and neonatal epithelial cells mediated by lipoteichoic acid. *Infect Immun,* 55:3057–3064. PMID:3316030

60. Swidsinski A, Mendling W, Loening-Baucke V *et al.* (2005). Adherent biofilms in bacterial vaginosis. *Obstet Gynecol,* 106:1013–1023. doi:10.1097/01.AOG.0000183594.45524.d2 PMID:16260520

61. Clarridge JE 3rd (2004). Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev,* 17:840–862.doi:10.1128/CMR.17.4.840-862.2004 PMID:15489351

62. Bäckhed F, Ley RE, Sonnenburg JL *et al.* (2005). Host-bacterial mutualism in the human intestine. *Science*, 307:1915–1920. doi:10.1126/science.1104816 PMID:15790844

63. Rundle AG, Vineis P, Ahsan H (2005). Design options for molecular epidemiology research within cohort studies. *Cancer Epidemiol Biomarkers Prev,* 14:1899–1907. doi:10.1158/1055-9965.EPI-04-0860 PMID:16103435

64. Warren JW, Abrutyn E, Hebel JR *et al.*; Infectious Diseases Society of America (IDSA) (1999). Guidelines for antimicrobial treatment of uncomplicated acute bacterial cystitis and acute pyelonephritis in women. *Clin Infect Dis,* 29:745–758.doi:10.1086/520427 PMID:10589881