Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Applications of Molecular Tools to Infectious Disease Epidemiology

The application of molecular techniques to epidemiology gives epidemiologists the tools to move beyond risk factor epidemiology and gain insight into the overall system of the disease. For infectious diseases, the system includes the transmission system, pathogenesis and virulence of the microbe, and the interaction of the microbe with the human (and other) host(s) and with the microbiota of the host (microbes that normally live on and in the human body). Thus, when dealing with an infectious disease, there are at least two genomes (sets of transcripts, proteins, and metabolites), that of the microbe causing disease and that of the host(s). Molecular tools enhance outbreak investigation and surveillance, facilitate description of the transmission system, increase understanding of the epidemiology, enable detection of previously unknown microbes, and provide insight into pathogen gene function and host–microbe interaction (Table 3.1). This chapter describes, with examples, each of these applications.

3.1 OUTBREAK INVESTIGATION

An important step in all outbreak investigations is setting the definition of what constitutes a case (Table 3.2). Molecular tools enhance case definitions, increasing specificity and reducing misclassification, and are now a standard tool in outbreak investigations. Although it is assumed during an outbreak that a single microbe is causing the clinical symptoms, it is possible that a microbe of the same genus and species but different strain is causing disease during the same time period. Molecular typing can distinguish between outbreak and nonoutbreak strains. Laboratory testing also distinguishes between syndromes with a similar clinical presentation. Laboratory screening can minimize misclassification of asymptomatic cases or cases in an early disease stage as nondiseased. Case definitions can be refined by including the molecular type as part of the case definition; this increases the specificity, reduces misclassification of nonoutbreak cases with outbreak cases, and thus increases the potential for identifying the outbreak source. There are a number of different types

| TABLE 3.1 Applications of Molecular Techniques to Infectious Disease Epidemiology |
|----------------------------------|
| Outbreak investigation           |
| Surveillance                     |
| Describe the transmission system |
| Increase understanding of the epidemiology |
| Identify previously unknown or uncultivable microbes |
| Provide insight into pathogen gene function and host–microbe interaction |

Molecular Tools and Infectious Disease Epidemiology. © 2012 Elsevier Inc. All rights reserved.
| TABLE 3.2 Core Steps of an Outbreak Investigation |
|------------------------------------------------|
| 1. Confirm that it meets epidemiologic definition of an outbreak (any one of the following).  |
|   a. There are more cases than expected (surveillance).  |
|   b. Cases are epidemiologically clustered by time, space, or common behaviors.  |
| 2. Consider whether there is ongoing transmission (one of the following).  |
|   a. Did regular contact investigations reveal epidemiologic links or similarities among cases?  |
|   b. Did the laboratory identify a genotyping cluster that confirms the epidemiologic links identified by regular  |
|       contact investigation?  |
|   c. Did the laboratory identify a genotyping or epidemiologic cluster of lab isolates clustered in time and space  |
|       where there is discordance between the clinical course of the patient and the laboratory results (false-positive  |
|       culture)?  |
| 3. Define an outbreak-related case.  |
| 4. Confirm existing number of outbreak-related cases.  |
| 5. Investigate existing outbreak-related cases by reviewing:  |
|   a. Medical records (history, physical, clinical chart, and notes);  |
|   b. Laboratory records (serial results of smears, cultures, drug sensitivities, and other testing);  |
|   c. Genotyping results for all culture-positive cases (if not already done, submit isolates for genotyping).  |
| 6. Determine the infectious period for each outbreak-related case based on:  |
|   a. Laboratory results, and  |
|   b. Results of screening of named contacts.  |
| 7. Determine the sites and facilities frequented and family and social groups exposed by outbreak-related patients  |
|   during their infectious periods.  |
|   a. Review information from case-patient interviews and contact investigations.  |
|   b. Review information from medical and public health records.  |
|   c. Review information from the facility logs or records.  |
| 8. Determine the exposed cohort of persons at each site/facility who may have been present when an outbreak-  |
|   related case-patient was present during his/her infectious period.  |
|   a. Review information from case-patient interviews and contact investigations.  |
|   b. Review information from medical and public health records.  |
|   c. Review information from the facility logs or records.  |
| 9. Determine the duration by number of hours, days, or weeks for the exposed cohort of persons who may have  |
|   spent time around an infectious outbreak-related patient.  |
|   a. Obtain information from case-patient interviews and contact investigations.  |
|   b. Obtain information from medical and public health records.  |
|   c. Obtain information from the facility logs or records.  |
| 10. Prioritize exposed cohorts for screening.  |
| 11. Define elements of and action plan for screening, implementation, and follow-up.  |
| 12. Identify resources necessary for action plan to be carried out.  |
| 13. Create a media plan to respond to possible inquiries.  |
| 14. Assign responsibilities and set deadlines.  |
| 15. If necessary, expand screening to include low-priority cohorts after screening high-priority cohorts based on  |
|   evidence of transmission.  |
| 16. Evaluate, treat, and follow up additional infected persons associated with this outbreak.  |
| 17. Make and implement recommendations to prevent future outbreaks for particular populations or settings  |
|   involved.  |
| 18. Evaluate outbreak response.  |
| 19. Determine whether interventions have effectively stopped transmission in this situation.  |
| 20. Identify the lessons learned that could improve the public health response to the next outbreak.  |

Source: Adapted from the Guide to the Application of Genotyping to Tuberculosis Prevention and Control: Appendix B. 

2
of laboratory tests that provide a molecular fingerprint. Most are based on the microbial genotype (see Chapter 5 for a description of different molecular fingerprinting methods), although phenotypic characteristics – such as serotype and antibiotic resistance profile – are also used. One typing method is called pulsed-field gel electrophoresis (PFGE), and in 2010 it is the standard for typing food-borne outbreaks.

The first step in an outbreak investigation is to confirm that an outbreak occurred, that is, there are more cases than expected or a space–time or behavioral cluster (Table 3.2). Space–time clusters can occur by chance alone, and molecular tools make it possible to distinguish between a cluster caused by different strains of the same species and one caused by a single strain. A cluster caused by a single strain likely indicates an outbreak. *Salmonella* is a common cause of diarrhea, transmitted by the fecal–oral route. An outbreak of a particular strain of *Salmonella* was first identified after a case of salmonellosis was reported to the health department in South Carolina. The first case identified a second case and the putative source – turtles. The turtles and cases all carried the same strain, *Salmonella Paratyphi B* var Java; a comparison of the molecular type with strains previously reported to the surveillance system identified additional cases. The case definition was thus “illness with onset from May 1, 2007, through January 31, 2008, in a U.S. resident yielding a *Salmonella Paratyphi B* var Java isolate with the outbreak PFGE pattern.”

Molecular typing confirms that there is ongoing transmission, and verifies epidemiologic linkage identified by contact investigation. Examples of epidemiologic linkage are that the individual had the same disease or syndrome during the appropriate time period, and was linked to other cases in some way. In the salmonellosis outbreak, all cases had contact with turtles; the first two cases had swum together with the pet turtles in a swimming pool. Typing can suggest linkage based on a common molecular fingerprint, and can confirm if the epidemiologically identified outbreak source, such as a food item, contains a microbe of the same genotype causing the outbreak, enhancing causal inference. In the outbreak of *Salmonella* associated with turtles, a total of 107 cases were identified that met the case definition; 72% of cases compared to 4% of controls reported turtle exposure in the week before illness. As confirmation of turtles as the source of the outbreak, samples were collected from six turtles (or water from turtle habitats) belonging to cases in four different states, and all the samples were positive for *Salmonella* of the outbreak type and PFGE pattern.

It is often difficult to distinguish between diseases based on clinical presentation alone. This can complicate outbreak investigations, especially if the symptoms are not very specific. There are many viruses that cause flulike symptoms; classification as influenza based on clinical presentation is specific only during an epidemic when the majority of flulike illnesses are caused by influenza. Even during an epidemic, laboratory confirmation is required, as there can be more than one strain of influenza in circulation. In 2008, there were two predominant influenza A strains in circulation: H1N1 and H3N2. Laboratory confirmation is particularly helpful for classifying individuals with mild or atypical symptoms, and determining the specific type.

Outbreak investigations undertaken by the Centers for Disease Control and Prevention (CDC) are routinely reported in the Morbidity and Mortality Weekly Report, which is available at the CDC website. One such investigation investigated a prolonged multistate outbreak of *Salmonella enterica* serotype Schwarzengrund infections in humans that was associated with dog food. The outbreak was first detected by local surveillance of PFGE types of *Salmonella*, which identified a cluster of three cases; CDC was notified. At CDC, a comparison with other reports identified isolates from multiple states with the same PFGE type, which stimulated an investigation. The case definition specifically included infection with the outbreak PFGE type; when epidemiologic evidence pointed to dog food as the source of infection, molecular tools confirmed presence of the outbreak strain in unopened bags of dog food and in
environmental samples from the implicated manufacturing plant. While not mentioned in the report, continued surveillance of the PFGE type of *Salmonella enterica* serotype Schwarzengrund infections could be used to confirm that the public health intervention – temporary closure of the plant for cleaning and disinfection – was successful in ending the outbreak.

### 3.2 SURVEILLANCE

Public health surveillance is the “ongoing, systematic collection, analysis, and interpretation of data (e.g., regarding agent/hazard, risk factor, exposure, health event) essential to the planning, implementation, and evaluation of public health practice, closely integrated with the timely dissemination of these data to those responsible for prevention and control.”

Public health surveillance is a cornerstone of public health infrastructure. Surveillance includes collection, analysis, and dissemination components (Figure 3.1). Data collected include incidence, morbidity, mortality, vaccination, clinical, behavioral, and laboratory data. Collected data are analyzed to monitor disease trends, giving a baseline for detection of outbreaks and epidemics, and to evaluate the effectiveness of public health interventions. Reports are disseminated regularly to decision-makers.

Laboratories are key components of many surveillance systems and essential for infectious diseases. Hospital laboratories may be part of both regional and local surveillance networks. Monitoring of infectious disease isolates identifies time–space clusters of infection; molecular typing distinguishes between microbes of the same species, allowing differentiation between clusters of disease occurring by chance and true outbreaks. Spurious clusters do not have a common source, so their investigation wastes time and resources. Molecular typing helps link disease reports from different geographic areas, as we saw in the outbreak of *Salmonella* associated with turtles presented in Section 3.1. True outbreaks and clusters of the same strain can be traced back to a common source and presumably are amenable to public health intervention. Applying molecular tools to surveillance isolates can also identify new strains with increased virulence or changing patterns of resistance.

Hospitals have high endemic rates of bacterial infection, but the infections are often due to a bacterial strain that was colonizing an individual before entering the hospital, for example, *Staphylococcal aureus*. The prevalence of *S. aureus* colonization among the general population is 32% in the nares, but much higher in patients and personnel in hospitals and long-term care facilities. By typing strains causing infection among patients we can distinguish 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 distinguish between them. Screening new hospital patients for the presence of *S. aureus* and then intervening to prevent spread or self-inoculation can reduce introduction of new strains. For strains already circulating in the hospital, screening hospital personnel and retraining personnel on proper hygiene may be in order. In hospitals there are a variety of surfaces that may be contaminated with potential pathogens that might serve as a source of infection. A study conducted in a 1600-bed hospital in Taiwan explored whether computer keyboards and mice might be a source of pathogens. Though one of three major pathogens was cultured from 17% of the 282 computer keyboards or mice, the PFGE types were all different from the PFGE types of clinical specimens obtained from the same wards, suggesting that good hygiene was sufficient to keep these devices from being a source of infection.

Surveillance can occur on multiple levels: in the United States there are surveillance systems within hospitals, cities, counties, states, and the entire country. By monitoring isolates from time–space clusters for the presence of a common molecular type at multiple levels, we can distinguish
between common-source outbreaks that are local and those that are widely disseminated. Food is frequently distributed widely; molecular tools enable identification of a common cause of disease disseminated through complex food distribution networks that cross large geographic areas. The CDC PulseNet, a molecular subtyping surveillance system for foodborne bacterial disease, monitors *Escherichia coli* O157:H7, *Salmonella*, *Shigella*, and *Listeria monocytogenes*, and other bacterial pathogens\(^8\) 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 as prepackaged, 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 and the identification of the disease source, and it facilitated quick public health intervention – recall of the spinach.\(^9\)

There is also ongoing surveillance for microbes resistant or insensitive to prevailing therapies. Surveillance monitors the emergence and spread of resistance, and provides essential information for effectively treating infection. A cluster of drug-resistant infections 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. Appropriate intervention should take into account whether a mobile genetic element is being exchanged between species or if there is clonal spread of a single organism. In an outbreak of multidrug-resistant *Pseudomonas aeruginosa* that occurred in a university hospital in Greece, isolates had a novel gene variant coding for resistance. The outbreak was primarily due to clonal spread of the same strain, but a second strain was found to carry the same novel gene variant, suggesting that some of the outbreak was due to the transfer of resistant genes to sensitive strains.\(^10\)

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 microbial pathogens. 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. Screening of the samples collected enabled the estimation of the prevalence of various human pathogens, 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 pathogens, and the distributions of pathogens by spatial–temporal and host characteristics. Such studies are extremely useful for generating hypotheses about transmission systems and potential prevention and control strategies, for evaluating the effectiveness of ongoing prevention and control programs, and observing time trends.

### 3.3 TRANSMISSION SYSTEM

The transmission system of a pathogen determines how pathogens are circulated within a population, and includes the transmission mode, interactions between the pathogen 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 pathogens highlights the utility of understanding the various transmission systems, because this understanding is central to identifying effective prevention and control strategies. Combining molecular typing methods with questionnaire data can confirm self-reported behaviors, especially important when the validity of self-report may be in doubt, such as might occur during contact tracing for sexually transmitted diseases. Molecular tools also 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.
When using simple transmission models to estimate the basic reproductive number ($R_0$), the average number of new cases generated from each infectious case in a fully susceptible population, we need the transmission probability per effective contact, the duration of infectivity, and the rate of effective contact. Molecular tools can usefully be applied in studies estimating each of these parameters. Before the availability of modern molecular tools, our ability to empirically estimate transmission probabilities was limited. Although the transmission of a sexually transmitted infection can be estimated by following couples where one is infected and the other is susceptible, without molecular tools, it is difficult to verify that the transmission event came from within 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, we have 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.1

Key transmission system parameters are incidence, prevalence, duration of infection, and transmission probabilities (Table 3.3). The accurate estimation of these parameters assumes that we can distinguish between subtypes (strains) of a microbe. As we have increased our ability to type microbes, we have been forced to re-evaluate many of our previous assumptions, because microbes are much more diverse than previously imagined. One prior key assumption was that pathogens are clonal, that is, during active infection all infecting organisms are the same. Genetic studies have emphasized that bacteria exist in populations; there is enough variability within even a clone that the genomic sequence obtained is an average of the population sequence rather than that of a specific individual.12 Some pathogens, RNA viruses for example, mutate quickly, so that the genetic type of a strain changes over the course of an epidemic, potentially complicating how a case is defined. A second prior assumption was that during an infectious process the pathogenic organism was the one most frequently isolated from the infected site. For many diseases we now know this assumption is false. 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 microbes 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. A third assumption was that individuals are only infected with one strain of a pathogen. This is also false. Individuals can be concurrently infected with different strains of human papillomavirus (HPV), gonorrhea, and tuberculosis.

| Parameter          | Definition                                                                 | Comments                                                                 |
|--------------------|----------------------------------------------------------------------------|--------------------------------------------------------------------------|
| Incidence          | New cases/individuals at risk/time                                         | Incidence is a rate                                                      |
| Prevalence         | All cases (new and existing)/all individuals                               | Prevalence is a proportion                                               |
| Duration of infection | Time from infected until infection is lost                                 | Duration is the reciprocal of the rate that infected individuals lose their infection |
| Transmission probability | The probability (chance) that disease will be transmitted when there is effective contact of a susceptible individual with an infected individual | Can be estimated from the number of transmission events/effective contacts between a susceptible and an infected individual |
| Rate of effective contact | Number of contacts sufficient to transmit disease/total contacts/time |                                                                         |
These observations have profound impacts on the conduct of future molecular epidemiologic studies of infectious diseases. 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, we must sample 28 different isolates from an individual to reliably detect the second strain. Further, if a pathogen mutates rapidly within a host, such as HIV, determining the mutation rate will be essential for accurately estimating transmission probabilities and following transmission chains. Moreover, laboratory analyses should take into account the heterogeneity of the organism when selecting isolates for analysis.

Understanding the full extent of the circulation of a particular pathogen is essential for making accurate predictions and determining appropriate prevention and control strategies. The two parameters of greatest interest are incidence, number of persons newly infected during a defined time period, and prevalence, the number of persons infected (Table 3.3). Prevalence includes both new and existing cases and can be measured at a single time point (point prevalence) or over a defined period (period prevalence). Incidence can be estimated by following a cohort and measuring the occurrence of new infections during a defined time period. 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. There are many types of HPV; only a few cause cancer. Before the development of the HPV vaccine, the incidence of infection with HPV type 16, which is a type strongly associated with cervical cancer, was estimated by measuring antibody to HPV 16 in the blood (serology). Women were tested for antibodies to HPV type 16 at the time of their first pregnancy and retested at the time of a second pregnancy. Incident cases were all women who tested negative for HPV type 16 during their first pregnancy that tested positive during their second pregnancy. Seroconversion rates by age (number seroconverted in a specific age group divided by total seronegative at the first pregnancy in that age group) were highest for women younger than 18 years (13.8%); this fell to 2.3% for women age 21 years.13

For some diseases, it is possible to distinguish between new and existing cases using molecular tests, so that it is not necessary to follow a cohort over time to estimate incidence. HIV is diagnosed using a test for host response to infection (antibodies), but antibodies do not appear until several weeks after infection. Once antibodies appear, infection can be diagnosed, but the standard test cannot determine how long it has been since infection occurred, so only prevalent cases are detected. Having a test to detect incident cases would be extremely useful for estimating the rate of ongoing HIV transmission with a single survey or as part of a screening program or intervention trial. Molecular tools have been developed to make this possible, and there are several tests available. After HIV status is identified, an additional test is conducted for a substrate that either is present only in those with recent infection or is a marker of extended infection.14 One such assay uses a branched synthetic peptide (called BED) that enables detection of multiple HIV subtypes; the assay quantitates the proportion of anti-HIV antibody in the serum that increases with time infected. There are still a number of issues with the interpretation of these tests, and it has been noted that biological changes in persons who have been infected for a long time may lead to a high false-positive rate. This is of particular concern in populations with high prevalence. Nonetheless, these tests are an important advance for monitoring HIV and evaluating HIV interventions.14

HIV almost invariably leads to detectable disease, but other infections do not. Many common infections, for example, diarrheal diseases, cause disease in only a subset of all infected, or the disease presentation is sufficiently mild that it does not require medical intervention. However, from a public health perspective it would be extremely useful to know the extent of the population infected, because this would enable a more accurate evaluation of the effectiveness of prevention strategies compared to an assessment based on
reduction in the number of outbreaks identified. Refinements of molecular tools combined with statistical approaches make it possible to estimate seroincidence. Seroincidence is the number of infections that lead to seroconversion within a defined time interval. The method uses the distribution of known changes in antibody titer since infection to estimate time since last infection, and then converts the time estimates to incidence. This approach has already been used to estimate seroincidence for Salmonella, a major cause of diarrheal disease. Though seroincidence cannot be used to estimate disease burden, because not all infections lead to disease, it does provide a useful estimate of the occurrence of new infections in the population.

The duration of infection can be estimated from the prevalence and incidence, presuming that the average duration across strain type is of interest, as prevalence divided by prevalence incidence times duration. If the disease is rare, prevalence is approximately incidence times duration. For many pathogens that can cause disease, pathogen presence does not have a one-to-one correspondence with symptoms: the pathogen can be carried asymptptomatically. The duration of carriage is important for understanding the transmission system; the duration of disease is useful for estimating disease burden. If duration of carriage is short but incidence is high, an individual might become reinfected with a different strain type, suggesting a longer duration than if strain types were determined. By contrast, if a strain mutates rapidly within the human host, we might underestimate the duration of carriage, an essential parameter for predicting ongoing circulation. Different types of the same species may vary in duration of carriage. S. agalactiae is an emerging pathogen that also frequently colonizes the rectum and the vagina. It has nine known serotypes that are also detected by gene sequence (capsular type). The duration of carriage varies by capsular type, and the duration is longer in women than men, presumably because of affinity of the pathogen for the vaginal cavity.

Transmission probabilities are difficult to estimate directly, unless the microbe is transmitted by person-to-person direct contact. For sexually transmitted diseases, the usual design is a couple study where one member of the couple is infected and the other is susceptible. Couples are followed until transmission occurs. Molecular tools are used to verify that both members of the couple were infected with the same strain of the microbe of interest. A couples study design estimated the probability of transmission of herpes simplex virus 1 (HSV1) during pregnancy; among 582 women initially seronegative for HSV1 with an HSV1-positive partner, 3.5% acquired HSV1. If sexual transmission is hypothesized, studying couples at a single point in time, a cross-sectional study, can also be informative. If couples carry the identical molecular type of a microbe more frequently than expected based on the population distribution of the microbe, it is strong evidence for sexual transmission. Sex partners of women with a urinary tract infection are more likely to be co-colonized with the E. coli strain that caused the urinary tract infection than a commensal E. coli colonizing the woman’s rectum, supporting the hypothesis that urinary tract infections can be sexually transmitted. However, dormitory roommates of men and women carrying S. agalactiae were no more likely to carry the same strain than expected by chance alone, where chance was estimated based on the distribution of molecular types in the study population.

Molecular tools can also assist in the estimation of contact patterns, by identifying asymptomatic and low level of infections. Contact patterns are a description of the interactions occurring between individuals sufficient to transmit the microbe. For a sexually transmitted disease, contact patterns of a population describe who has sex with whom. Asymptomatic infection is often a key component in maintaining disease transmission. For example, in a study of intrafamily transmission of shigella, asymptomatic carriage increased risk of a symptomatic episode within 10 days by ninefold. 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 reinfection.21

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 programs, particularly when designing and implementing vaccine programs. Using PCR–restriction fragment length polymorphism typing of the porin and opacity genes of Neisseria gonorrhoeae 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.21 Further analysis revealed that the transmission networks for men who have sex with men, and heterosexuals were essentially separate – a key public health insight for planning interventions. 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, because clusters have been associated with use of services at day shelters,22 and even linked to only a few brief visits to an infected individual’s work site.11 Molecular typing has also demonstrated linkage between apparently sporadic tuberculosis cases, and determined that at least some recurrent tuberculosis is attributable to exogenous reinfection.23

3.4 INCREASE 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 test theories about reservoirs and evolution of a particular pathogen; 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 pathogens (Table 3.4).

3.4.1 Trace Dissemination of a Particular Subtype Across Time and Space

Microbes that cause human disease are constantly emerging and re-emerging. To prevent and control the spread of infection, we must be able to trace the origin and source of entry of pathogens into the population. By comparing strains we can determine if there have been single or multiple points of entry, and if emerging resistance was from multiple spontaneous mutations or from dissemination of a single clone. Until 2004, only occasional

| TABLE 3.4 Ways Molecular Tools Increase Understanding of the Epidemiology of Infectious Diseases |
| --- |
| Trace dissemination of a particular subtype across time and space |
| Determine the origin of an epidemic |
| Follow emergence of new infections |
| Trace mobile genetic elements |
isolates of gonorrhea found in Sweden were resistant to azithromycin, and these cases were attributed to acquisition elsewhere. 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* 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 (MLST). By contrast, the recent emergence of *S. pneumoniae* resistant to fluoroquinolones has been due to a diverse set of genetic mutations, suggesting spontaneous emergence following treatment. Because *S. pneumoniae* resistant to fluoroquinolones rapidly followed the introduction of fluoroquinolones, alternative antibiotics will be needed in relatively short order to treat *S. pneumoniae* infections.

### 3.4.2 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 between epidemics. The same strains of cholera that infect humans also thrive in aquatic environments, living in zooplankton. During zooplankton blooms the population of cholera vibrio also grows and is more likely to invade the human population.

Molecular tools 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%. 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 were genotypically different from hospital strains. More recently, community-acquired MRSA has been introduced into hospitals. Because community-acquired MRSA has, to date, different patterns of antibiotic resistance than hospital-acquired MRSA as well as different virulence factors, there is a clinical benefit in being able to distinguish between the two.

Influenza season comes every year, and we can predict, with reasonable accuracy, which strains will be circulating, enabling preparation and distribution of vaccine. Prediction is based on surveillance of influenza worldwide. The influenza virus mutates as it circulates; the mutations can be modest, known as “drift,” such that there is cross immunity with the previous strain; or mutations may be dramatic, where the virus acquires genes by recombining with other strains, known as “shift.” Antigenic shifts can occur when human influenza recombines (exchanges genetic material) with other influenza strains. There are influenza viruses that infect humans, fowl, and swine. Pigs have cell receptors that make them susceptible to both avian and human influenza as well as swine influenza, so genetic reassortment between different influenza viruses can occur. It was previously thought that recombination between human and bird influenza within a pig was necessary before an avian flu could infect humans. However, this is not the case. Molecular tools have clarified that avian influenza need not first pass through the pig before jumping to humans, and that strains directly transmitted from birds to humans are often more virulent.

*An infection is endemic if there is continued transmission within the population. By contrast, an epidemic is when there are more cases of a specific disease than expected.*
In 1918 there was a very severe epidemic of influenza, which, unlike seasonal flu, was most severe in young adults. This virus was different from those seen previously and probably originated in birds. This founding virus, an influenza A H1N1, remains with us (Figure 3.2), and its descendents plague us to this day. The founding virus was introduced to pigs by humans; in 2009 an H1N1 virus was transmitted from pigs to humans. As of yet, we are unable to predict when an antigenic shift will occur or when an avian virus will jump into humans or pigs. However, our ability to trace the flow of specific viral types and their mutations over time provide important information for predicting disease spread and hence for developing effective prevention strategies.

### 3.4.3 Follow the Emergence and Spread of New Infections

SARS was the first new disease to emerge this century. Before its identification, coronaviruses were not considered major pathogens – only 12 coronaviruses were known to infect humans or other animals. The SARS identification led to a search for additional coronavirus pathogens, and ultimately horseshoe bats were identified as the reservoir and civets as the amplification hosts. The time from the initial observation to the sequencing of the virus and development of a diagnostic test was 5 months. 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. This powerful combination enabled scientists to follow the emergence and spread, and to identify ways to prevent transmission and further introductions of the virus into human populations.

| Years          | Circulating Virus (Genetic Mechanism)                     | Excess Deaths from Any Cause no. per 100,000 persons/yr |
|----------------|----------------------------------------------------------|----------------------------------------------------------|
| 1918–1919      | H1N1 (viral introduction), pandemic                       | 598.0                                                    |
| 1928–1929      | H1N1 (drift)                                             | 96.7                                                     |
| 1934–1936      | H1N1 (drift)                                             | 52.0                                                     |
| 1947–1948      | H1N1 A’ (intrasubtypic reassortment)                     | 8.9                                                      |
| 1951–1953      | H1N1 (intrasubtypic reassortment)                        | 34.1                                                     |
| 1957–1958      | H2N2 antigenic shift, pandemic                           | 40.6                                                     |
| 1968–1969      | H3N2 (antigenic shift), pandemic                         | 16.9                                                     |
| 1972–1973      | H3N2 A Port Chalmers (drift)                             | 11.8                                                     |
| 1975–1976      | H3N2 (drift) and H1N1 ("swine flu" outbreak)             | 11.8                                                     |
| 1977–1978      | H3N2 (drift) and H1N1 ("swine flu" outbreak)             | 12.4                                                     |
| 1977–1999      | H3N2 (drift) and H1N1 (viral introduction)               | 49.5                                                     |
| 2003–2004      | H3N2 Fujian (intrasubtypic reassortment and H1N1 (drift) | 17.1                                                     |
| 2009           | H3N2 and H1N1 (drift) and swine-origin H1N1 (viral introduction), pandemic | ?                                                        |

*Mortality data include deaths associated with all influenza A and B viruses combined. Many of these data have been calculated with the use of differing methods and may not be strictly comparable. The 1934, 1951, and 1977 data span 2 years.

**FIGURE 3.2**

Mortality associated with influenza pandemics and selected seasonal epidemic events, 1918 to 2009. 

*Source: Reproduced, with permission, from Morens et al.*
SARS was first reported in southern China in 2002 and rapidly spread worldwide (Figure 3.3). 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 identification of chains of transmission. 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 pathogen, which required a validated antibody test. Early in the epidemic there were many possible candidates identified as the cause, but these microbes 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 coronavirus isolated from SARS cases identified worldwide were caused by the same virus, confirming transmission, and a rapidly developed test demonstrated that SARS patients had antibodies to the new coronavirus. That SARS was a virus newly introduced to humans was confirmed by demonstrating that healthy controls not suffering from SARS had no evidence of either past or present infection.

Most microbes have a relatively short life span compared to humans. As they reproduce, mutations occur; further, microbes may exchange genetic material. Therefore, it is not sufficient to compare genetic sequence, because we anticipate there will be changes in the genetic sequence over time. To make sense of genetic changes requires an analysis that takes into account evolutionary relationships, a field called phylogenetics. Phylogenetics enables epidemiologists to trace the emergence and transmission of a rapidly evolving species and, in an outbreak situation, determine order of transmission. HIV, which causes 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 and to gain insight into the spread of HIV.
worldwide. Not all HIV subtypes spread at the same rate. Using phylogenetic analysis Saad and colleagues\textsuperscript{38} traced the introduction and spread of HIV in the Ukraine. The analysis revealed that two HIV 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 Russia, Moldova, Georgia, Uzbekistan, and Kyrgyzstan. Further studies to determine the biological and social contributions to the success of the one subtype over another will provide important insights into how to control HIV.

\subsection*{3.4.4 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 survive as extrachromosomal material in the cytoplasm of bacteria and 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 rethinking of the rate of and potential for microbial evolution. For example, Shiga toxin–producing \textit{E. coli} probably emerged from the transfer of genes coding for Shiga toxin from \textit{Shigella} into \textit{E. coli}. Antibiotic resistance is often spread via mobile genetic elements, which tend to code for genes providing resistance against multiple antibiotics. This explains several apparent mysteries, such as the spread across many 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. Because mobile genetic elements evolve separately from their microbial hosts, separate phylogenies can be constructed, giving insight into the emergence and evolution of these elements.

\section*{3.5 IDENTIFY PREVIOUSLY UNKNOWN OR UNCULTIVABLE INFECTIOUS MICROBES}

The vast majority of microbes cannot be cultured using standard laboratory techniques; the ability to make a copy of genetic material and determine the genetic sequence, which can then be compared to known genetic sequence, has led to a radical reassessment of the amount of life around, in, and on us. Nonculture techniques have enabled us to characterize the microbial communities living in the mouth, gut, vagina, and other body sites, and the detection of microbes in body sites previously thought to be sterile, such as the blood. Epidemiologic data may suggest an infectious origin for a disease; in the past, if an organism could not be cultured, it remained only a suggestion. Molecular tools have changed this by enabling the detection of uncultivable microbes. Two examples of the power of combining epidemiologic principles with molecular techniques are the identification of the causative agent of Kaposi sarcoma, and the identification of HPV as the cause of cervical cancer, which led to the development of an effective vaccine.

The epidemiology of Kaposi sarcoma among persons with AIDS strongly suggested that it was caused by an infectious agent. Not all persons with AIDS had Kaposi sarcoma, and AIDS patients with specific characteristics were more likely to get the sarcoma. Kaposi sarcoma occurred much more frequently among men who have sex with men compared to heterosexual populations or those who acquired HIV from blood or blood products, and among those engaging in specific sexual practices. Reasoning that Kaposi sarcoma might be caused by a virus, Chang and associates\textsuperscript{39} used representative difference analysis, a technique that identifies DNA sequences present in one set of tissue but absent in another, to identify and characterize unique DNA sequences in Kaposi sarcoma tissue that were either absent or present in low copy number in nondiseased tissue obtained from the same patient. Chang and colleagues\textsuperscript{39} then went on to demonstrate that the sequences, which were similar to herpes virus, were not present in tissue DNA from non-AIDS patients, but were present in
AIDS and non-AIDS patients with Kaposi sarcoma worldwide, and that infection with the virus preceded development of the sarcoma.

HPV types 16 and 18 are now known to cause cervical and other cancers, and a vaccine is licensed to prevent acquisition. HPV 16 was first identified in 1983 before the virus could be grown. When HPV 16 was discovered, we were aware that papillomavirus could cause cancer in rabbits, cows, and sheep, but it was not clear that the HPV caused cancer in humans. HPV was a suspected cause of genital cancer, because – similar to Kaposi’s sarcoma – the epidemiology suggested that an infectious agent was involved. However, other genital infections, particularly herpes simplex virus, were also suspects. HPV had been ruled out by many, but a new molecular technique, the hybridization assay, detected in cancerous tissue a new subtype, HPV 16, which was specifically associated with cervical and other cancers. The association of HPV 16 with cancer was confirmed by comparing presence of HPV 16 among cancer patients to controls. Though this evidence can be very suggestive, it does not demonstrate temporal order, because the cancer might have preceded HPV 16 infection. Demonstrating temporal order required large-scale prospective cohort studies. These studies also provided key insights supporting the possibility that HPV could be prevented by vaccination, because reinfection with the same HPV subtype rarely occurred, and antibody could protect against reinfection and persistence of low-grade lesions.

### 3.6 PROVIDE INSIGHT INTO PATHOGEN GENE FUNCTION AND HOST–PATHOGEN INTERACTION

Many major human pathogens have been genetically sequenced, and hundreds of microbial genomes will be sequenced in the near future. (For a listing, see the website, Pathogen Genomics, at http://www.sanger.ac.uk/Projects/Pathogens.) Unlike genes from multicellular organisms, single-celled organisms often vary greatly in genetic content and expression – that is, genes may be present or absent as well as expressed or silent. Once a single strain of a microbe has been sequenced, the sequence can be used as a reference for comparison with others in the same species, providing insight into the heterogeneity of the species. Sequence information can be mined for potential virulence genes, by identifying genes of unknown function with structures similar to genes whose function is known in the same or other species. However, until the presence or expression of the gene is associated with transmission, pathogenesis, or virulence at a population level, the relative importance of a particular gene cannot be discerned. A gene that is associated with virulence in the laboratory may occur so rarely in vivo that it is an inappropriate target for vaccination or therapeutics. Epidemiologic screening of collections of human pathogens for the prevalence of genes that alter the transmission, pathogenesis, and virulence of the microbe provides insight into the potential importance and putative function of genes identified using genomic analyses. A gene found more frequently in strains that cause severe disease (virulent strains) than among strains that colonize without causing symptoms (commensal strains) suggests that the gene is worthy of more detailed laboratory analyses of its function. A genomic subtraction of a *Haemophilus influenzae* causing middle ear infection (otitis media) from a laboratory strain of *H. influenzae* identified several genes found only in middle ear isolates. Upon screening of collections of human isolates, one gene, *lic2B*, occurred 3.7 times more frequently among middle ear than colonizing isolates, suggesting that *lic2B* is involved in the pathogenesis of otitis media.

Not all genes are expressed at all times; expression profiling (transcriptomics) can identify which genes a pathogenic microbe expresses at different stages of pathogenesis, and which genes the human host expresses in response. Using expression arrays, expression of distinct sets of genes associated with acute, asymptomatic, and the AIDS stages of HIV-1 infection were detected. While only a first step, this type of descriptive analysis provides a basis for understanding the role of these genes in HIV-1 pathogenesis.
Some parasites live in very different environments during their life cycle; understanding which genes are expressed during different points of development is essential for identifying targets for control either via therapeutics or vaccination. The schistosome, a parasite that infects ~200 million people worldwide and causes a variety of adverse health outcomes, at different points lives in freshwater, snails, and vertebrate hosts. Gene expression studies have identified hundreds of genes that are differentially expressed over the life cycle, many that have potential to be targets for intervention.\(^4\) The bacteria *Porphyromonas gingivalis* inhabits the human mouth and is associated with periodontal disease. Smokers have higher rates of periodontal disease and of persistent *P. gingivalis* infection, suggesting that smoking may modify *P. gingivalis* interaction with the host. This was shown to be the case in an experiment that used a microarray representative of the *P. gingivalis* genome to monitor expression. Exposure to cigarette smoke extract changed regulation of a number of *P. gingivalis* genes, and macrophages and peripheral blood mononuclear cells had a lower proinflammatory response when *P. gingivalis* was exposed to cigarette smoke extract.\(^5\)

**References**

1. J.R. Harris, D. Bergmire-Sweat, J.H. Schlegel, et al., Multistate outbreak of *Salmonella* infections associated with small turtle exposure, 2007–2008, Pediatrics 124 (5) (2009) 1388–1394.
2. Centers for Disease Control and Prevention, Division of Tuberculosis Elimination. Guide to the Application of Genotyping to Tuberculosis Prevention and Control. [Internet]. Available at: http://www.cdc.gov/tb/programs/genotyping/Append/AppendixB_4_Core.htm, 2009 (Updated 1.6.09; accessed 5.8.10).
3. Centers for Disease Control and Prevention, Multistate outbreak of human *Salmonella* infections caused by contaminated dry dog food – United States, 2006-2007, MMWR Morb. Mortal. Wkly. Rep. 57 (19) (2008) 521–524.
4. Centers for Disease Control and Prevention: Department of Informatics Shared Services. Public Health Surveillance Slide Set [Internet]. Available at: http://www.cdc.gov/ncphi/disss/nndss/phs/overview.htm, 2009 (Updated 5.1.09; accessed 14.12.09).
5. M. Trostle, USAID: Infectious Diseases: Disease Surveillance: Overview. [Internet]. Washington, DC: United States Agency for International Development. Available at: http://www.usaid.gov/our_work/global_health/id/surveillance/ (accessed 25.2.10).
6. A.G. Mainous, 3rd, W.J. Hueston, C.J. Everett, et al., Nasal carriage of *Staphylococcus aureus* and methicillin-resistant *S. aureus* in the United States, 2001–2002, Ann. Fam. Med. 4 (2) (2006) 132–137.
7. P.L. Lu, L.K. Siu, T.C. Chen, et al., Methicillin-resistant *Staphylococcus aureus* and *Acinetobacter baumannii* on computer interface surfaces of hospital wards and association with clinical isolates, BMC Infect. Dis. 9 (2009) 164.
8. B. Swaminathan, T.J. Barrett, S.B. Hunter, et al., PulseNet: The molecular subtyping network for foodborne bacterial disease surveillance, United States, Emerg. Infect. Dis. 7 (3) (2001) 382–389.
9. Centers for Disease Control and Prevention, Ongoing multistate outbreak of *Escherichia coli* serotype O157:H7 infections associated with consumption of fresh spinach – United States, September 2006, MMWR Morb. Mortal. Wkly. Rep. 55 (38) (2006) 1045–1046.
10. V.I. Siarkou, D. Vitti, E. Protonotariou, et al., Molecular epidemiology of outbreak-related *Pseudomonas aeruginosa* strains carrying the novel variant blaVIM-17 metallo-beta-lactamase gene, Antimicrob. Agents Chemother. 53 (4) (2009) 1325–1330.
11. J.E. Golub, W.A. Cronin, O.O. Obasanjo, et al., Transmission of *Mycobacterium tuberculosis* through casual contact with an infectious case, Arch. Intern. Med. 161 (18) (2001) 2254–2258.
12. D. Medini, D. Serruto, J. Parkhill, et al., Microbiology in the post-genomic era, Nat. Rev. Microbiol. 6 (6) (2008) 419–430.
13. M. Kibur, V. af Geijerstam, E. Pukkala, et al., Attack rates of human papillomavirus type 16 and cervical neoplasia in primiparous women and field trial designs for HPV16 vaccination, Sex Transm. Infect. 76 (1) (2000) 13–17.
14. R. Guy, J. Gold, J.M. Calleja, et al., Accuracy of serological assays for detection of recent infection with HIV and estimation of population incidence: A systematic review, Lancet Infect. Dis. 9 (12) (2009) 747–759.
15. J. Simonsen, K. Molbak, G. Falkenhorst, et al., Estimation of incidences of infectious diseases based on antibody measurements, Stat. Med. 28 (14) (2009) 1882–1995.
16. B. Foxman, B. Gillespie, S.D. Manning, et al., Incidence and duration of group B *Streptococcus* by serotype among male and female college students living in a single dormitory, Am. J. Epidemiol. 163 (6) (2006) 544–551.
17. C. Gardella, Z. Brown, A. Wald, et al., Risk factors for herpes simplex virus transmission to pregnant women: A couples study. Am. J. Obstet. Gynecol. 193 (6) (2005) 1891–1899.
18. B. Foxman, S.D. Manning, P. Tallman, et al., Uropathogenic Escherichia coli are more likely than commensal E. coli to be shared between heterosexual sex partners. Am. J. Epidemiol. 156 (12) (2002) 1133–1140.
19. S.D. Manning, K. Neighbors, P.A. Tallman, et al., Prevalence of group B streptococcus colonization and potential for transmission by casual contact in healthy young men and women. Clin. Infect. Dis. 39 (3) (2004) 380–388.
20. A.I. Khan, K.A. Talukder, S. Huq, et al., Detection of intra-familial transmission of shigella infection using conventional serotyping and pulsed-field gel electrophoresis. Epidemiol. Infect. 134 (3) (2006) 605–611.
21. M.E. Kolader, N.H. Dukers, A.K. van der Bij, et al., Molecular epidemiology of Neisseria gonorrhoeae in Amsterdam, The Netherlands, shows distinct heterosexual and homosexual networks. J. Clin. Microbiol. 44 (8) (2006) 2689–2697.
22. K. DeRiemer, C.L. Daley, Tuberculosis transmission based on molecular epidemiologic research. Semin. Respir. Crit. Care Med. 25 (3) (2004) 297–306.
23. B. Mathema, N.E. Kurepina, P.J. Bifani, et al., Molecular epidemiology of tuberculosis: current insights. Clin. Microbiol. Rev. 19 (4) (2006) 658–685.
24. D. Lundback, H. Fredlund, T. Berglund, et al., Molecular epidemiology of Neisseria gonorrhoeae: identification of the first presumed Swedish transmission chain of an azithromycin-resistant strain, APMIS 114 (1) (2006) 67–71.
25. G.V. Doern, S.S. Richter, A. Miller, et al., 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 (2) (2005) 139–148.
26. J. Reidl, K.E. Klose, Vibrio cholerae and cholera: out of the water and into the host. FEMS Microbiol. Rev. 26 (2) (2002) 125–139.
27. NNIS System, National Nosocomial Infections Surveillance (NNIS) System Report, data summary from January 1992 through June 2004. Am. J. Infect. Control 32 (8) (2004) 470–485.
28. T.S. Naimi, K.H. LeDell, K. Como-Sabetti, et al., Comparison of community- and health care-associated methicillin-resistant Staphylococcus aureus infection, JAMA 290 (22) (2003) 2976–2984.
29. K.E. Sabol, K.L. Echevarria, J.S. Lewis, 2nd. Community-associated methicillin-resistant Staphylococcus aureus: new bug, old drugs. Annals of Pharmacotherapy 40 (6) (2006) 1125–1133.
30. K. Van Reeth, Avian and swine influenza viruses: our current understanding of the zoonotic risk. Vet. Res. 38 (2) (2007) 243–260.
31. D.M. Morens, J.K. Taubenberger, A.S. Fauci, The persistent legacy of the 1918 influenza virus. N. Engl. J. Med. 361 (3) (2009) 225–229.
32. V.C. Cheng, S.K. Lau, P.C. Woo, et al., Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. Clin. Microbiol. Rev. 20 (4) (2007) 660–694.
33. World Health Organization. SARS: Cumulative Number of Reported Cases: Total Number of Cases. 2671 as of 8 April, 2003 14:30 GMT. Available at: http://www.who.int/csr/sars/SARS2003_4_8.jpg, 2003 (accessed 5.8.10).
34. R.A. Weinstein, Planning for epidemics: the lessons of SARS. N. Engl. J. Med. 350 (23) (2004) 2332–2334.
35. A. Berger, C. Drosten, H.W. Doer, et al., Severe acute respiratory syndrome (SARS): paradigm of an emerging viral infection, J. Clin. Virol. 29 (1) (2004) 13–22.
36. T.G. Ksiazek, D. Erdman, C.S. Goldsmith, et al., A novel coronavirus associated with severe acute respiratory syndrome, N. Engl. J. Med. 348 (20) (2003) 1953–1966.
37. M.L. Metzker, D.P.Mindell, X.M. Liu, et al., Molecular evidence of HIV-1 transmission in a criminal case, Proc. Natl. Acad. Sci. USA 99 (22) (2002) 14292–14297.
38. M.D. Saad, A.M. Shcherbinskaya, Y. Nadai, et al., Molecular epidemiology of HIV Type 1 in Ukraine: birthplace for transmission by casual contact in healthy young men and women. Clin. Infect. Dis. 39 (3) (2004) 380–388.
39. A.I. Khan, K.A. Talukder, S. Huq, et al., Detection of intra-familial transmission of shigella infection using conventional serotyping and pulsed-field gel electrophoresis. Epidemiol. Infect. 134 (3) (2006) 605–611.
40. P.S. Moore, Y. Chang, Detection of herpesvirus-like DNA sequences in Kaposi’s sarcoma in patients with and without HIV infection, N. Engl. J. Med. 332 (18) (1995) 1181–1185.
41. Y. Chang, J. Ziegler, H. Wabinga, et al., Kaposi’s sarcoma-associated herpesvirus and Kaposi’s sarcoma in Africa. Uganda Kaposi’s Sarcoma Study Group, Arch. Intern. Med. 156 (2) (1996) 202–204.
42. S.J. Gao, L. Kingsley, D.R. Hoover, et al., Serocoversion to antibodies against Kaposi’s sarcoma-associated herpesvirus-related latent nuclear antigens before the development of Kaposi’s sarcoma, N. Engl. J. Med. 335 (4) (1996) 233–241.
43. P.S. Moore, L.A. Kingsley, S.D. Holmberg, et al., Kaposi’s sarcoma-associated herpesvirus infection prior to onset of Kaposi’s sarcoma, AIDS 10 (2) (1996) 175–180.
44. M. Durst, L. Gissmann, H. Ikenberg, et al., A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions, Proc. Natl. Acad. Sci. USA 80 (12) (1983) 3812–3815.

45. L. Koutsky, The epidemiology behind the HPV vaccine discovery, Ann. Epidemiol. 19 (4) (2009) 239–244.

46. Wellcome Trust Sanger Institute. Pathogen Genomics [Internet]. Available at: http://www.sanger.ac.uk/Projects/Pathogens/ (accessed 5.8.09).

47. M.M. Pettigrew, B. Foxman, C.F. Marrs, et al., 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 (7) (2002) 3551–3556.

48. Q. Li, A.J. Smith, T.W. Schacker, et al., Microarray analysis of lymphatic tissue reveals stage-specific, gene expression signatures in HIV-1 infection, J. Immunol. 183 (3) (2009) 1975–1982.

49. J.M. Fitzpatrick, E. Peak, S. Perally, et al., Anti-schistosomal intervention targets identified by lifecycle transcriptomic analyses, PLoS Negl. Trop. Dis. 3 (11) (2009) e543.

50. J. Bagaitkar, L.R. Williams, D.E. Renaud, et al., Tobacco-induced alterations to Porphyromonas gingivalis-host interactions, Environ. Microbiol. 11 (5) (2009) 1242–1253.