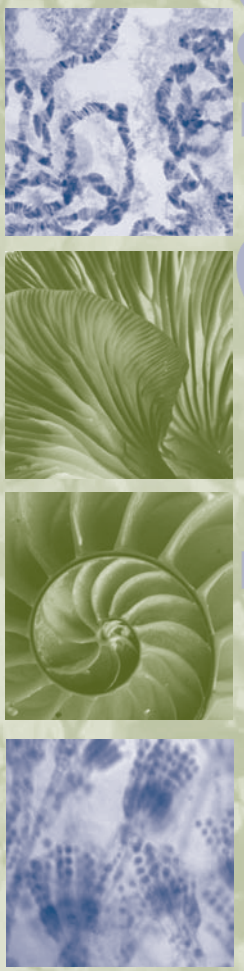
Part 4



Teaching Evolution's Importance for Public Health

Chapter 12

Health Applications of the Tree of Life

David M. Hillis

Scientific papers that use phylogenetic methods have been increasing at an exponential rate for the past 25 years (Hillis, 2004), and now virtually all biological journals contain applications of phylogenetic analyses. A quarter of a century ago, the tree of life was primarily of academic interest to systematists and evolutionary biologists, and its principal application was the organization and classification of living organisms. Although that use continues to this day, the applications of phylogenetics have grown rapidly, and now virtually all biologists need to understand how to use and interpret phylogenetic trees.

Why has phylogenetics become so critical to an understanding of biology in general? First, it has become widely appreciated that none of the things that we study in biology (genes, cells, individual organisms, populations, species, communities, ecosystems, etc.) are independent and identical entities. This sets biology apart from most of the physical sciences, where (for instance) a hydrogen atom of a given isotope is the same as every other hydrogen atom of that isotope. In contrast to understanding a chemical reaction, understanding the similarities, differences, and relationships among the entities that biologists study is critical to understanding how those entities work and interact. Biology can only be predictive if these relationships are taken into account. But then why has phylogenetics only been such an influential factor in biology for the past 25 years? This is largely attributable to scientific breakthroughs in phylogenetic analysis: vast increases of comparative data sets (especially DNA sequences), rapid increases in computational power, and parallel development of phylogenetic algorithms and theory. As the methods have become available and feasible, they have been rapidly applied by biologists to problems throughout biology. Nowhere has this been more apparent than in applications to biomedicine and human health, and in particular to the study of human pathogens. Here I discuss several examples of human health applications of phylogenetics and the reasons why evolutionary principles in general need to be understood by anyone who is concerned about human health.

Pathogens Evolve, Often Very Rapidly

It is an observable fact that pathogens evolve. Since many human pathogens have very short generation times and large population sizes, evolution by natural selection is often extremely rapid. Therefore, evolution of pathogens is often observable over the course of the infection of a single human individual. For instance, an individual human who becomes infected with HIV typically is infected with a single HIV virus, of just one genotype. This virus quickly replicates inside the infected individual, and this replication occurs with a relatively high error rate, so the virus evolves quickly. The human immune system mounts an attack on the infection, but the rate of evolution of HIV is so high that some of the evolving viruses escape detection by the immune system, and the virus population quickly increases in genetic diversity (e.g., see Nowak, May, & Anderson, 1990; Nowak, Anderson, McLean, Wolfs, Goudsmit, & May, 1991; Nowak & Bangham, 1996). If drug treatments are used, then there is rapid selection for resistant strains, which invariably exist because of the high population diversity (e.g., see Larder & Kemp, 1989; Leigh-Brown & Cleland, 1996). Thus, every HIV infection demonstrates evolution by natural selection, and an understanding of evolution and selection is critical to developing effective treatments of the disease (for reviews of the importance of evolutionary biology to understanding HIV, see Crandall, 1999). Ignorance of the fact of the evolving nature of the pathogen would lead to treatments that would worsen the course of the disease in the infected individual and in human populations as a whole.

The fact that human pathogens evolve does not just affect the way we develop treatment regimes. Because pathogens evolve, they do not have fixed genomes that can be identified by simple matching

methods. Instead, their identification relies on the same phylogenetic methods that are used to identify and classify all life. However, most organisms with longer generation times evolve slowly enough that we can use fixed features of their genotypes or phenotypes for identification at one place and time. Not so for many pathogens, which may evolve so quickly that phylogenetic placement is the only means available to identify them. In addition, the study of the spread of pathogens among human populations (the field of epidemiology) has been greatly aided by phylogenetic methods. Using these methods, it is now possible to follow a given pathogen through human populations in space and time, and thereby identify how the pathogen is spread and develop methods to curtail the epidemic. Development of effective vaccines also depends on an understanding of the past evolution and the future potential of the target pathogen to evolve, and phylogenetic methods are now routinely used to identify whether new cases of polio have resulted from back mutations of viruses used in vaccines or from naturally occurring reservoirs of the virus. These same methods are also used to determine the origins and timing of emergence of new diseases into human populations from nonhuman hosts. This information, in turn, is critical to blocking future diseases from moving into human populations, as well as to identifying appropriate animal models for studying human diseases. Therefore, an understanding of evolution and the application of phylogenetic methods has become essential for anyone with an interest in human health.

Identification of Pathogens, Now and in the Future

In 1993, there was an outbreak of a strange respiratory illness in the Four Corners region of the southwestern United States. In previous years, this disease would probably have gone unidentified, or at the least, isolation and identification of the viruses would have taken many years. However, by the early 1990s, the biologists from the Centers for Disease Control and Prevention who investigated the outbreak were armed with a relatively new tool for investigations of emerging diseases: phylogenetic analysis. By amplifying viral DNA from the infected individuals, and conducting a phylogenetic analysis of the sequences with sequences from other sequenced viruses, they were quickly able to identify the "new" virus as a hantavirus (Nichols et al., 1993). Armed with this information, biologists quickly traced the source of the infection to host populations of mice, which had recently increased to large population sizes in the region as a result of a wet El Niño year. The epidemic was quickly stemmed as health officials learned of the source of the infections and were able to recommend relatively simple measures to reduce infection rates. This incident led to nationwide studies of related hantaviruses in rodent populations, and it quickly became clear that these viruses are a common source of moderate to severe respiratory illnesses in many areas of the country (Monroe et al., 1999). Thus, a major source of respiratory illness was identified, and now phylogenetic investigations are used to track areas with high infection rates, identify the source rodent populations, and develop control programs. The phylogenetic analyses that were used so successfully in the case of the hantavirus outbreak are now used routinely to identify outbreaks of "new" diseases. For instance, these same methods were used in 2003 to rapidly identify the coronavirus that is responsible for severe acute respiratory syndrome (SARS; Peiris et al., 2003; Ksiazek et al., 2003).

Despite the success of cases such as hantaviruses and SARS, we are still unable to rapidly identify many common pathogens, such as the many viruses that cause coldlike symptoms in billions of people each year. When a sick person visits a physician's office, he or she wants treatment that will result in quick recovery. However, one of the most common reasons for illness is a viral infection, and most viral infections cannot be identified using current technology in a physician's office. The best physicians will recommend general, sensible measures that often help (get plenty of rest, drink lots of fluids, etc.) and tell the patient that there isn't much else that they can do. Patients hate this, of course, and often demand an antibiotic. Of course, the antibiotic does nothing to help fight the viral infection, and inappropriate prescription of antibiotics leads to the selection of antibiotic-resistant bacteria. Thus, the ignorance by the patient (and sometimes the physician) of simple evolutionary principles leads not only to a waste of money for an antibiotic that does no good, but also to a potentially much worse problem when a future bacterial infection cannot be treated with the antibiotic. In truth, this problem is not limited to viral infections: the vast majority of pathogens cannot be identified quickly enough (or specifically enough) in an infected individual to result in appropriate

treatment. Why can't these pathogens be quickly identified and appropriate treatments developed to treat the specific infection?

The short answer is that rapid identification of pathogens is technically possible, and that treatments can probably be developed for most or all of these infections. Humans have simply not made this a priority. This problem is a small part of a much larger problem: namely, the general human ignorance about the biological diversity of the world in which we live. This ignorance is not the result of limiting technology or resources; we have simply chosen to use our existing resources for other purposes. As of this writing, biologists have identified 1.7 million extant species on Earth. Estimates of the total number of living species vary widely, but most biologists place the number at 5–100 million species, so in any case we know only a small fraction of the total. Of the 1.7 million species that have been identified, we know the complete genomes of only a few hundred, and we know a fragment of a gene sequence from only a few tens of thousands. In recent years, many biologists have called for a systematic study of Earth's biota, so that we can move beyond this obstacle (see, for example, Wilson, 2004). Therefore, let's imagine that such a survey were to take place and that biologists could build a database of DNA sequences from a collection of genes sampled from every species on Earth. How could this database change a visit to the physician's office?

A phylogenetic tree of organisms sampled from throughout life can be down-loaded from the University of Texas Web page http://www.zo.utexas.edu/faculty/antisense/Download.html. This 2 x 2 meter wall poster depicts a phylogenetic tree that was built from the analysis of ribosomal RNA genes sampled from about 3,000 species. These genes evolve very slowly, because rRNA is the backbone of the ribosome, the site of protein translation. But rRNA genes do evolve, albeit very slowly, and they can be used to reconstruct the evolutionary relationships across all cellular life. (Viruses are not cellular, and they use the ribosomes of their hosts for protein translation). If there are 9 million of species life on Earth, then this sample of 3,000 species represents approximately the square root of the total number of living species. Therefore, we could represent the complete tree of cellular life by expanding each tip of this tree into a tree of similar size. We would probably want to use other genes to do this (genes that evolved more quickly would provide more resolution among closely related species), and we would also want to sequence other genes in viruses that lack rRNA genes. In fact, biologists are now building exactly such databases. In addition, technology is being developed to rapidly isolate and amplify DNA, sequence appropriate genes, and then place these sequences into a phylogenetic context in the tree of life. When most of life has been sampled, it will be possible to identify any species, anywhere, anytime by placing gene sequences of the unknown sample into the phylogenetic framework of the rest of life. Even a new pathogen, never before encountered, can be identified by its phylogenetic relationships with other species, which will provide immediate information about the treatment and biology of related pathogens. Thus, phylogenetic methods form the basis of the technology that will make a visit to the physician a much more positive experience in years to come: the source of the illness will be rapidly identified using phylogenetic methods, and then a specific treatment can be identified that targets the particular problem. Phylogenetic methods are therefore of great practical importance. Once this technology has been fully implemented, it will allow the implementation of specific and useful treatments for common diseases. Moreover, it also will allow us to predict the most successful treatments for new diseases never before encountered, based on the relationships of the newly encountered pathogens to other, known pathogens.

Epidemiological Investigations

Phylogenetic analysis has also become an important tool for studying the transmission of viruses throughout human populations. These analyses are used to determine risk groups for certain diseases, to identify source populations and source host species, and to study transmission dangers in various health settings. As one example, phylogenetic analyses have become the principal means for studying the infection of patients by health care workers, whether intentional or unintentional (e.g., Ou et al., 1992; Hillis & Huelsenbeck, 1994).

In one of the more dramatic cases of studying an infection pathway, a Louisiana physician was found to have purposefully infected his former mistress with HIV from an HIV-positive patient of the physician (Metzker, Mindell, Lin, Ptak, Gibbs, & Hillis, 2002). In this case, viruses from local HIV-positive individuals

were compared with HIV isolates obtained from the patient and the victim in the case. Phylogenetic analysis of the HIV sequences was consistent with a transmission from the patient to the victim, although these individuals had no known contact other than through the physician, who apparently injected the victim with blood drawn from the patient. The phylogenetic analyses were used as evidence in the court case, together with evidence that the physician had drawn blood from the patient and then had injected the victim against her will. In this case, the physician was convicted of attempted murder. In other court cases, phylogenetic analyses have been used to convict individuals of rape and aggravated assault (e.g., Leitner et al., 1996).

For some diseases, such as rabies, it is critical to identify the particular source host of the virus that has been transmitted to humans. Rabies occurs naturally in many mammalian hosts, some of which do not regularly transmit the virus to humans. To control the spread of rabies, it is important to identify which hosts are likely to transmit the virus to humans; these hosts can then be targeted for rabies control programs. The virus coevolves in several natural hosts, so a phylogenetic analysis can be used to identify which strain is involved in a particular transmission event, or across many transmission events in a population. In some cases, this information may be used to design oral vaccination programs for wildlife species that represent significant reservoirs of rabies virus (e.g., Rupprecht, Hanlon, & Slate, 2004).

Vaccine Development and Use

The development and worldwide use of vaccines requires information about the variation and evolution of the disease-causing organism that the vaccine is meant to target (Halloran et al., 1998). As an example, consider the effort to eliminate polio on a worldwide basis through a vaccination program. Oral polio vaccines (OPVs) are based on an attenuated form of the polio virus … in other words, an evolved form of the polio virus that does not cause disease in people, and yet still produces an immune response that is effective in providing protection against dangerous forms of the polio virus. These vaccines have prevented many millions of cases of polio since their introduction in 1961. Unfortunately, the attenuated viruses that are used for the vaccines also continue to evolve, and, rarely, they undergo spontaneous mutations that result in virulent forms of polio virus. Polio workers need to identify outbreaks of polio around the world and determine if they are caused by pockets of wild polio virus that have not yet been eradicated, or by viruses that have been introduced to human populations in vaccination programs and have reverted to virulent forms (for a review, see Dowdle, De Gourville, Kew, Pallansch, & Wood, 2003). In the case of human populations that are only exposed to polio through the vaccination programs, the vaccination programs may be terminated to eliminate polio (or the vaccination protocols may be modified to include other forms of vaccine; Alexander et al., 2004; Korotkova et al., 2003). On the other hand, where human populations are still exposed to wild polio virus, then the vaccination programs must be continued. Wild versus reverted polio viruses are easily identified through the use of phylogenetic analyses (Kew et al., 2004). By reconstructing the evolutionary history of the viruses, investigators can tell if the virulent viruses are derived from wild or laboratory stocks, and therefore determine where the vaccination programs should continue and where they should be terminated.

For some viral diseases, the rate of evolution is so high that a single vaccine is not likely to be effective. Many different vaccines may have to be developed for some phylogenetically diverse viruses. In these cases, phylogenetic analyses are useful at several levels. A phylogenetic analysis is used to study the worldwide geographic variation of the virus (for instance, see McCutchan, 1999, for an analysis of geographic variation of HIV, or Twiddy et al., 2002, for an analysis of geographic variation of dengue virus). For some diseases, a phylogenetic analysis of the virus present in a given patient informs health care providers with the information they need to determine which vaccine is needed (or whether a vaccination is needed at all).

In some cases, phylogenetic analyses can be used to predict which of the currently circulating strains of a pathogen is likely to lead to the epidemics of tomorrow (Bush, Bender, Subbarao, Cox, & Fitch, 1999; Hillis, 1999). Such information can be important for selecting strains of viruses to use in vaccine production. In the case of influenza, there is strong selection to escape detection by the human immune system, so through time, the lineages that are best able to escape detection are the ones that are likely to survive (Bush, Subbarao, Cox, & Fitch, 1999). By sequencing the genes for hemagglutinin (one of the

protein spikes on the surface of an influenza virus that is detected by the human immune system) and then conducting a phylogenetic analysis, biologists can assess which of the currently circulating strains of influenza virus has the greatest number of amino acid replacements in the target areas for immunoselection. Retrospective studies (e.g., Bush, Bender, Subbarao, Cox, & Fitch, 1999) have shown that these maximal escape strains are most closely related to the viruses that are present in epidemics of subsequent years. In other words, this information can be used to predict the future course of evolution of influenza viruses, and this information can then be used to select the most appropriate strains of virus for the development of future flu vaccines.

Origins of Emerging Diseases

New diseases appear with regularity in human populations. In some cases, these may be old diseases that have only recently been recognized in humans (as in the hantavirus example discussed above), and in other cases, they are actually diseases that have never before occurred in human populations. Usually, these are diseases that occur naturally in some non-human host and move (from once to many times) into human populations. It is important to know where these diseases come from and how often they are transferred into human populations if we are to control or stem the transfer of such diseases to humans.

HIV presents a good example of a disease-causing virus that has been studied extensively by phylogenetic methods to answer the where, when, and how questions about the origins of this virus (Hillis, 1999b). Phylogenetic studies have clarified that immunodeficiency viruses have moved into human populations from two different primate hosts, and that they have been transferred from both of these hosts on more than one occasion (Sharp, Robertson, Gao, & Hahn, 1994; Hahn, Shaw, De Cock, & Sharp, 2000). The viruses appear to have moved into human populations through the hunting and eating of the host primate species (Hahn, Shaw, De Cock, & Sharp, 2000). HIV-1 has its origins in chimpanzee populations in central Africa, whereas HIV-2 originated from sooty mangabey populations in western Africa. Both HIV-1 and HIV-2 have been transmitted to human populations multiple times, and it is likely that these viruses have been entering human populations for centuries or even millennia (for as long as humans have been hunting and eating the host species). Phylogenetic analyses can also be used to date the origins of these viruses into human populations; for instance, the M-subgroup of HIV-1 (the strain of HIV that is most prevalent in North America and western Europe) appears to date to between 1915 and 1941 (Korber et al., 2000).

If HIV has been transmitted repeatedly to human populations for centuries, then why have HIV and its resultant disease, AIDS, only become such global issues since the 1970s? It appears that these viruses were present in localized epidemics in Africa well before that time, but that they quickly spread in and out of Africa because of major social changes in Africa (and the rest of the world) throughout the 1950s and 1960s. Many factors have conspired to make HIV and AIDS global problems. Rapid population growth and upheaval, major movement of populations following years of civil wars, the rapid growth of large urban areas, increased movement of people within Africa and between Africa and the rest of the world, the reuse of hypodermic needles in vaccination campaigns and in illegal drug use, and increased sexual freedom and prostitution all combined to change local epidemics into global epidemics (Hahn, Shaw, De Cock, & Sharp, 2000). Phylogenetic analyses are now necessary to track the spread of HIV around the world and to identify the prevalent transmission pathways. These studies are critical for slowing the transmission of HIV (by identifying the important risk factors in different cultures around the world) and for identifying the growing divergence of the viruses (for producing effective means of control and treatment).

The factors that have resulted in the global HIV epidemic are not unique to HIV. Many new viruses are appearing in human populations as a result of these (and other) social changes. The large number of emerging diseases has given rise to entire new journals dedicated to studying these problems; for instance, the journal *Emerging Infectious Diseases* began publication in 1995. The pages of this journal are filled with phylogenetic analyses that are used to study the spread of new diseases into and among human populations around the world. Thus, evolutionary biology has become critical to the study of human health. The fact of the matter is that pathogens evolve, and so humans must study the evolution of these disease-causing organisms if they are to understand how to treat and control them.

The study of evolution and phylogeny is critical to a modern understanding of all aspects of biology, and nowhere is this dependence on evolutionary biology clearer than in the study of human health.

REFERENCES

Alexander, L. N., Seward, J. F., Santibanez, T. A., Pallansch, M. A., Kew, O. M., & Prevots, R. (2004). Vaccine policy changes and epidemiology of poliomyelitis in the United States. *Journal of the American Medical Association, 292*, 1696–1701.

Bush, R. M., Bender, C. A., Subbarao, K., Cox, N. J., & Fitch, W. M. (1999). Predicting the evolution of human influenza A. *Science, 286,* 1921–1925.

Crandall, K. A. (Ed.). (1999). *The evolution of HIV*. Baltimore: Johns Hopkins University Press.

Dowdle, W. R., De Gourville, E., Kew, O. M., Pallansch, M. A., & Wood, D. J. (2003). Polio eradication: The OPV paradox. *Reviews in Medical Virology, 13*, 277–291.

Hahn, B. H., Shaw, G. M., De Cock, K. M., & Sharp, P. M. (2000). AIDS as a zoonosis: Scientific and public health implications. *Science, 287*, 607–614.

Halloran, M. E., Anderson, R. M., Azevedo-Neto, R. S., Bellini, W. J., Branch, O., & Burke, M. A. (1998). Population biology, evolution, and immunology of vaccination and vaccination programs. *American Journal of Medical Sciences, 315*, 76–86.

Hillis, D. M. (1999a). Predictive evolution. *Science, 286*, 1866–1867.

Hillis, D. M. (1999b). Phylogenetics and the study of HIV. In K. A. Crandall (Ed.), T*he Evolution of HIV* (pp. 105–121). Baltimore: Johns Hopkins University Press.

Hillis, D. M. (2004). *The tree of life and the grand synthesis of biology*. In J. Cracraft & M. J. Donoghue (Eds.), Assembling the Tree of Life (pp. 545–547). New York: Oxford University Press.

Hillis, D. M., & Huelsenbeck, J. P. (1994). Support for dental HIV transmission. *Nature, 369*, 24–25.

Kew, O. M., Wright, P. F., Agol, V. I., Delpeyroux, F., Shimizu, H., & Nathanson, N. (2004). Circulating vaccine-derived polioviruses: Current state of knowledge. *Bulletin of the World Health Organization, 82*, 16–23.

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., & Lapedes, A. (2000). Timing the ancestor of the HIV-1 pandemic strains. *Science, 288*, 1789–1796.

Korotkova, E. A., Park, R., Lipskaya, E. A., Chumakov, K. M., & Feldman, E. V. (2003). Retrospective analysis of a local cessation of vaccination against poliomyelitis: A possible scenario for the future. *Journal of Virology, 77*, 12460–12465.

Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., & Emery, S. (2003). A novel coronavirus associated with severe acute respiratory syndrome. *New England Journal of Medicine, 348*, 1953–1966.

Larder, B. A., & Kemp, S. D. (1989). Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science, 246*, 1155–1158.

Leigh-Brown, A. J., & Cleland, A. (1996). Independent evolution of the env and pol genes of HIV-1 during zidovudine therapy. *AIDS, 10*, 1067–1073.

Leitner, T., Escanilla, D., Franzen, C., Uhlen, M., & Albert, J. (1996). Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proceedings of the National Academy Sciences of the United States of America, 93*, 10864–10869.

McCutchan, F. E. (1999). Global diversity in HIV. In K. A. Crandall (Ed.), *The Evolution of HIV* (pp. 41–101). Baltimore: Johns Hopkins University Press.

Metzker, M. L., Mindell, D. P., Liu, X-M., Ptak, R. G., Gibbs, R. A., & Hillis, D. M. (2002). Molecular evidence of HIV-1 transmission in a criminal case. *Proceedings of the National Academy Sciences of the United States of America, 99*, 14292–14297.

Monroe, M.C., Morzunov, S. P., Johnson, A. M., Bower, M. D., Artsob, H., & Yates, T. (1999). Genetic diversity and distribution of Peromyscus-borne hantaviruses in North America. *Emerging Infectious Diseases, 5*, 75–86.

Nichol, S. T., Spiropoulou, C. F., Morzunov, S., Rollin, P. E., Ksiazek, T. G., Feldmann, H., et al. (1993). Genetic identification of a hantavirus associated with an outbreak of acute respiratory illness. *Science, 262*, 914–917.

Nowak, M. A., Anderson, R. M., McLean, A. R., Wolfs, T. F., Goudsmit, J., & May, R. M. (1991). Antigenic diversity thresholds and the development of AIDS. *Science, 254*, 963–969.

Nowak, M. A., & Bangham, C. R. (1996). Population dynamics of immune responses to persistent viruses. *Science, 272*, 74–79.

Nowak, M. A., May, R. M., & Anderson, R. M. (1990). The evolutionary dynamics of HIV-1 quasispecies and the development of immunodeficiency disease. *AIDS, 4*, 1095–1103.

Ou, C-Y., Ciesielski, C. A., Myers, G., Bandea, C. I., Luo, C. C., & Korber, B. T. (1992). Molecular epidemiology of HIV transmission in a dental practice. *Science, 256*, 1165–1171.

Peiris, J. S. M., Lai, S. T., Poon, L. L. M ., Guan, Y., Yam, L. Y., & Lim, W. (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet, 361*, 1319–1325.

Rupprecht, C. E., Hanlon, C. A., & Slate, D. (2004). Oral vaccination of wildlife against rabies: Opportunities and challenges in prevention and control. *Developmental Biology (Basel), 119*, 173–184.

Sharp, P. M., Robertson, D. L., Gao, F., & Hahn, B. H. (1994). Origins and diversity of human immunodeficiency viruses. *AIDS, 8*, S27–S43.

Twiddy, S. S., Farrar, J. J., Chau, N. V., Wills, B., Gould, E. A., & Gritsun, T. (2002). Phylogenetic relationships and differential selection pressures among genotypes of dengue-2 virus. *Virology, 298*, 63–72.

Wilson, E. O. (2004). The meaning of biodiversity and the tree of life. In J. Cracraft & M. J. Donoghue (Eds.), *Assembling the tree of life* (pp. 539–542). New York: Oxford University Press.