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.

CHAPTER

7

Long-term virus evolution in nature

Abbreviations

BCE Before Christian Era
CAT Colonization-Adaptation Trade-off
cccDNA Covalently Closed Circular DNA
CTL Cytotoxic T Lymphocyte
CRF Circulating Recombinant Forms
ELISA Enzyme-Linked Immunosorbent Assay
EMCV Encephalomyocarditis Virus
ET Evolutionary Trace
FAO Food and Agriculture Organization of the United Nations
FMD Foot-and-Mouth Disease
FMDV Foot-and-Mouth Disease Virus
HAV Hepatitis A Virus
HBV Hepatitis B Virus
HCV Hepatitis C Virus
HIV-1 Human Immunodeficiency Virus Type 1
HLA Human Leukocyte Antigen
HTLV-1 Human T-Cell Lymphotropic Virus Type 1
HTLV-2 Human T-Cell Lymphotropic Virus Type 2
HRV Human Rhinovirus
ICTV International Committee on Taxonomy of Viruses
IV Influenza Virus
MARM Monoclonal Antibody-Resistant Mutant
ML maximum likelihood
MV Measles Virus
PAM Percent Accepted Mutation
PIR Protein Information Resource
PV Poliovirus
RV Rabies Virus
SARS Severe Acute Respiratory Syndrome
SIVcpz Chimpanzee Simian Immunodeficiency Virus
s/nt/y Substitutions Per Nucleotide and Year
URF Unique Recombinant Form
URL Uniform Resource Locator
WEEV Western Equine Encephalitis Virus
WNV West Nile Virus

7.1 Introduction to the spread of viruses. Outbreaks, epidemics, and pandemics

Intrahost virus replication and evolution are the first steps in the process of virus diversification that continue with successive virus transmission events that are a condition for long-term survival in nature. Viruses are perpetuated as a consequence of many rounds of persistent or acute infections, with possible extracellular stages in which genomes remain basically invariant. Despite lacking direct evidence, we presume that multitudes of successive transmissions have allowed viruses to survive at least for thousands of years, probably undergoing continuous genetic change. Picornavirologists are familiar with an Egyptian stela dated 1550–1333 BCE. (18th Egyptian dynasty) that portrays the image of a man with an atrophic leg probably a consequence of infection with poliovirus (PV) or a related virus (Eggers, 2002). In this chapter, we deviate from the focus on how viral population numbers affect short-term survival and evolution, and we turn to features of viruses as they infect successive hosts to permit virus perpetuation in nature.

Virus as Populations
https://doi.org/10.1016/B978-0-12-816331-3.00007-6

Viruses can be transmitted vertically or horizontally. Vertical transmission occurs from parental organisms to their offspring, and it includes infection through the germline in animals and plants, from the mother to the embryo during fetal development, and also postnatal transmission to the newborn via blood, milk, or contact (Mims, 1981; Nash et al., 2015). In horizontal transmission, a virus spreads from infected individuals to susceptible recipients. We are most familiar with this type of transmission. It frequently gives rise to disease outbreaks (infection episodes localized in space and time that affect a few individuals), epidemics (that affect an ample geographical area and are often extended in time), and pandemics (that affect most areas of our planet), typically the periodic influenza pandemics.

All transmission modes have probably contributed to the maintenance of viruses in our biosphere. Persistent infections are likely to have played a major role when the number of individual humans or animals living in close contact was limited throughout the pre-agricultural era, earlier than 10,000 years ago. A favorable climate change during the Holocene (the geological epoch that began at the end of the Pleistocene, around 12,000 years before the present; compare with Chapter 1) was probably an important driver toward large-scale domestication of plants and animals, 10,000–7000 years BCE. From the behavior of current viruses, there might have always been a dynamics of virus change for adaptability within individual hosts, and transmissions among animals or plants. The probability of transmission increased as host population numbers rose with agricultural practices and urban life in the last several thousand years. Intensive agriculture must have contributed to accelerated sequence space exploration by viruses with consequences for the emergence of viral disease (Section 7.7). Probably, there has been a continuous dynamics of viral emergences, reemergences, and extinctions with patterns that may be parallel to those observed in present-day viruses. Virology has existed as an organized scientific discipline with the possibility to isolate, store, and study viruses only for about one century. The challenge to reconstruct the events that might have led to viruses similar to the ones we isolate today was addressed in Chapter 1, with a critical first question being if viruses originated 4000 million years ago, or "only" 2000 million years ago (diagrams in Figs. 1.3 and 1.4, and Section 1.5 in Chapter 1). In this chapter, we are more modest in our aspirations, and we will analyze, with the tools of genomics, what happens when viruses evolve for months or years in what we call inter-host virus evolution.

Unfortunately, viruses have not left a fossil record (at least one that we can uncover with the available tools), since according to current paleontology, nitrogen- and phosphorus-rich molecules are unlikely to be protected in fossils older than 1 million years. At most, hundred to thousand years-old skeletal remains or frozen bodies that contain viruses have been analyzed, and sequences retrieved [(Muhlemann et al., 2018) and references therein]. Fortunately, there is a different historical record of ancient viral sequences in the DNA of differentiated organisms, in the form of integrated virus-like genetic elements. The research area that consists in rescuing ancestral viral sequences is termed paleo-virology, and it is providing information on viruses that circulated thousands of years or even millions of years ago in the case of viral sequences integrated into cellular DNA (Aswad and Katzourakis, 2012).

The presence of recognizable viral genomic sequences in present-day cellular DNA (Tomonaga et al., 2019) suggests a history of long-term interaction between viruses and cells, in support of some models of virus origins that propose long coevolution between precellular and cellular entities with virus-like elements (Chapter 1). Despite the current capacity to amplify tiny amounts of viral nucleic acids for nucleotide sequence determinations, proposals

on how long-term viral evolution might proceed have to be based mainly on the comparison of viral genomes and the structure of viral proteins from modern representatives of different virus groups. First, we should understand the basic concepts related to virus transmission, keeping in mind viral population numbers and the complexity of viral populations.

7.2 Reproductive ratio as a predictor of epidemic potential. Indeterminacies in transmission events

The basic reproductive ratio or basic reproduction number (R_0) is the average number of infected contacts per infected individual. At a population level, a value of R_0 larger than one means that a virus will continue its propagation among susceptible hosts if no environmental changes or external influences intervene. An R_0 value lower than one means that the virus is doomed to extinction at the epidemiological level under those specific circumstances. The basic models of infection dynamics were developed by R.M. Anderson, R.M. May, and M.A. Nowak, with inclusion of the following key parameters: rate k at which uninfected hosts enter the population of susceptible individuals (x), their normal death rate (u) (so that the equilibrium abundance of uninfected hosts is k/u), number of infected hosts (y), mortality due to infection (v) (so that $1/u + v$ is the average lifetime of an infected host), a rate constant (β) that characterizes parasite infectivity (so that βx is the rate of new infections and βxy is the rate at which infected hosts transmit the virus to uninfected hosts). These parameters are schematically indicated in Fig. 7.1 and they provide a theoretical value for R_0 (Anderson and May 1991; Nowak and May 2000; Nowak, 2006; Woolhouse, 2017).

R_0 values are not a universal constant for viruses because, as discussed in Chapters 3 and 4, virus variation may affect viral fitness and

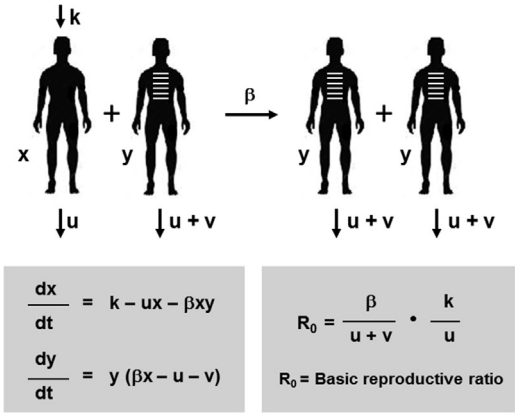

FIGURE 7.1 A schematic representation of the main parameters of viral dynamics that enter the equations that predict the rate of variation of uninfected and infected (internal horizontal lines in the human figure) individuals (shaded box on the left) and the R_0 value (shaded box on the right). The meaning of parameters and literature references are given in the text.

viral load in infected individuals, and the latter, in turn, may influence the amount of virus that surfaces in a host to permit transmission (Delamater et al., 2019). Despite uncertainties, consistent R_0 values have been estimated for different viral pathogens based on field observations. Values of R_0 for human immunodeficiency virus type 1 (HIV-1) and severe acute respiratory syndrome (SARS) coronavirus range from two to five; for PV the range is 5–7, and for Ebola virus is 1.5–2.5. For the measles virus (MV), which is one of the most contagious viruses described to date, the R_0 reaches 12–18 (Heffernan et al., 2005; Althaus, 2014). Most isolates of the SARS coronavirus that circulated months after the emergence of this human pathogen had modest R_0 values, and this is consistent with SARS not having reached the pandemic proportions that were feared immediately following its emergence. In contrast, MV is highly transmissible, thus explaining frequent outbreaks as soon as a sizable population stops vaccinating its infants. This is an important problem, fueled by antivaccination campaigns without scientific basis. Since some of the parameters that enter the basic

equations of viral dynamics depend on the nucleotide sequence of the viral genome, mutations may alter R_0 values, allowing some virus variants to overtake those that were previously circulating in the population (Fig. 7.2). Viral replication, fitness, load, transmissibility, and virulence are all interconnected factors that contribute to virus persistence in its broader sense of virus being perpetuated in nature. These parameters can affect both disease progression in an infected individual and transmissibility at the epidemiological level.

The difference between the number of infectious particles that participate in transmission and the total number of virus in an infected, donor organism provides a first picture of the indeterminacies involved in viral transmissions. The larger the population size and genetic heterogeneity of the virus in an infected individual, the higher will be the likelihood that independent transmission events have different outcomes. Individual susceptible hosts will receive subsets of related but nonidentical genomes. In a bright article that emphasized

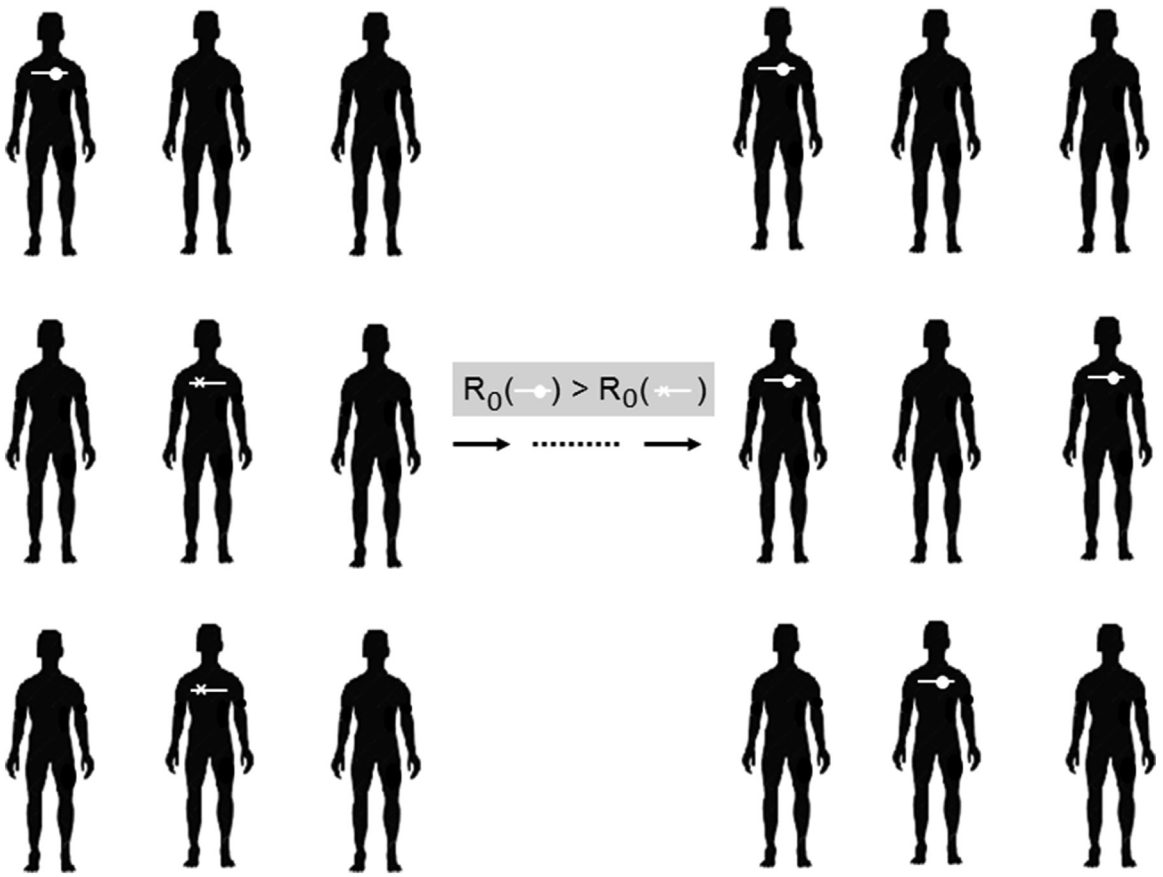

FIGURE 7.2 Displacement of a virus variant by another in the field by virtue of the latter displaying a higher R_0 value. The competing viruses are depicted as horizontal lines with a distinctive symbol. Differences in R_0 recapitulate part of the determinants of epidemiological fitness (Section 5.9 in Chapter 5). Concepts of competition among clones or populations within infected host organisms or cell cultures, treated in previous chapters, can be extended at the epidemiological level, with the appropriate choice of the key parameters. References are given in the text.

the molecular evidence and medical implications of quasispecies in viruses, J.J. Holland and colleagues wrote the following statement: "Therefore, the acute effects and subtle chronic effects of infections will differ not only because we all vary genetically, physiologically, and immunologically, but also because we all experience a different array of quasispecies challenges. These facts are easily overlooked by clinicians and scientists because disease syndromes are often grossly similar for each type of virus, and because it would appear to make no difference in a practical sense. However, for the person who develops Guillain-Barré syndrome following a common cold, or for the individual who remains healthy despite many years of HIV-1 infection, for example, it may make all the difference in the world" (Holland et al., 1992). The ever increasing number of identified human genes whose allelic forms influence viral infections provides strong support for the predictions of Holland et al. Indeterminacies in the process of virus spread can be viewed as an extension of the diversification due to bottleneck events in the case of virus transmission, as visualized in Figs. 6.1 and 6.2 in Chapter 6, when dealing with the limitations of the virus samples retrieved from an infected host as the starting material for experimental evolution approaches.

7.3 Rates of virus evolution in nature

Despite a necessarily approximate and imprecise knowledge of how many and which types of genomes participate in successive horizontal and vertical transmissions, we can obtain an overall estimate of the rate at which viruses evolve in nature. This is commonly done by comparing consensus genomic nucleotide sequences of viruses isolated at different times during an outbreak, epidemic, or pandemic.

The rate of evolution (also termed as the rate of fixation or rate of accumulation of mutations) is generally expressed as substitutions per nucleotide and year (s/nt/y). The term fixation is not the most adequate when dealing with virus evolution given that the term refers to a consensus that, in addition to being an average of the real sequences, has a fleeting dominance. However, the term is frequently used in the literature of general genetics and virus evolution. The rate of evolution is calculated from genetic distances between consensus viral genomic sequences of successive viral samples from a single persistently or acutely infected host, or from different host individuals infected at different times. Rates of evolution are only indirectly related to mutation rates and mutation frequencies that do not include a time factor in them. There have been several comparisons of rates of evolution for viruses that document the differences between RNA and DNA viruses (Jenkins et al., 2002; Hanada et al., 2004; Domingo, 2007). As was the case for mutant spectra (Section 3.3 in Chapter 3), some DNA viruses attain rates of evolution in nature that are very similar to those typical of RNA viruses (Duffy and Holmes, 2008). A few comparative values are given in Table 7.1.

Herpes simplex virus constitutes an example of a complex DNA virus for which, despite uncertainties (Firth et al., 2010) a calculated rate of evolution was 10^{-8} s/nt/y (Sakaoka et al., 1994), which is actually closer to the rate estimated for

TABLE 7.1 Some representative rates of virus evolution in nature.

Virus or organism	Range of values
RNA viruses (riboviruses)	10^{-2} to 10^{-5}
Retroviruses	10^{-1} to 10^{-5}
Single-stranded DNA viruses	10^{-3} to 10^{-4}
Double-stranded DNA viruses	10^{-7} to 10^{-8}
Cellular genes of host organisms	10^{-8} to 10^{-9}

Values are expressed as substitutions per nucleotide and year. The range of values is based on several studies. Values depend on the virus or organism under study, the genomic region analyzed, and several factors discussed in the text.

cellular genes than for most viruses. However, its mutation frequencies, measured by independent procedures, are in the range of 7×10^{-3} to 1×10^{-5} (see also Section 7.4.2). The latter values may result from the selective agent targeting a replicating herpes simplex virus that has produced multiple variants, while the overall slow rate of evolution may be influenced by periods of latency. Slow evolution is expected for retroviruses such as human T-cell lymphotropic virus types 1 and 2 (HTLV-1 and HTLV-2) whose life cycles are dominated by the integrated provirus stage, with the viruses following the clonal expansion of their host cells (Melamed et al., 2014; Kulkarni and Bangham, 2018). Some single-stranded DNA viruses display rates of evolution typical of the rapidly evolving RNA viruses (Table 7.1).

Different genes of the same virus set may show different rates of evolution (i.e., the polymerase and other nonstructural proteins may evolve more slowly than structural proteins). Thus, a rate of evolution is far from being a universal feature of a virus. A comparison of rates of synonymous substitutions (under the assumption that synonymous substitutions do not affect protein function; see Chapter 2 for limitations of considering synonymous mutations as neutral) for several RNA viruses, yielded a range of evolutionary rates of 6×10^{-2} to 1×10^{-7} synonymous substitutions per synonymous site per year (Hanada et al., 2004). The values were recalculated from primary phylogenetic data using maximum likelihood (ML) (Section 7.6), under the assumption of the molecular clock, and inference of the ancestral nucleotide sequences at the tree nodes. The five orders of magnitude variation were attributed mainly to the degree of virus replication rather than to differences in error rate. We will deal with the molecular clock hypothesis (constant rate of accumulation of mutations) in Section 7.3.3, but the major features of virus evolution studied in previous chapters (mainly those typical of mutant swarm-forming RNA and DNA viruses) should make us skeptical of similar evolution rates in different biological contexts. Rate variations were documented with HIV-1 subpopulations in different compartments of the human brain (Salemi et al., 2005). The data did not fit a "global" molecular clock for the virus in the brain, and "local" clocks showed that meninges and temporal lobe HIV-1 subpopulations evolved 30 and 100 times faster, respectively, than other HIV-1 populations in the brain. It is believed that these differences were due to random drift of viral sequences rather than selection for some genome types. An additional complication is that even restricting virus isolations to the same biological material in a standard epidemiological setting, several measurements indicated discontinuities in evolutionary rates. The discontinuities had at least two origins: the nonlinear effect of time, and some unique features of evolution occurring inside an infected host. These points are examined next.

7.3.1 Influence of the time of sampling

It has been known for decades that the calculated rates of virus evolution in nature depend on the genomic region analyzed and the time interval between the isolation of the viral samples on which the calculations are based. Noncumulative sequence changes in the hemagglutinin of influenza virus (IV) type C were found in an early study by Buonagurio et al. (1985). The authors proposed a cocirculation of variants that belonged to different evolutionary lineages. If multiple evolutionary pathways coexist in a given geographical area, and they establish a network of lineages that evolve with time, variations of calculated rates of evolution are expected, and they may distort the actual rate of evolution of individual lineages.

A second early observation was made during an episode of foot-and-mouth disease (FMD) in Spain. Estimates of the rate of evolution of the

virus ranged from $<4 \times 10^{-4}$ to 4×10^{-2} s/nt/y, depending on the genomic region analyzed, and the time period between isolations (Sobrino et al., 1986). Cocirculation of multiple heterogeneous foot-and-mouth disease virus (FMDV) samples ("evolving quasispecies") was proposed. The result to be emphasized here is that the calculated rates of evolution were extremely high (higher than 10^{-2} s/nt/y) if the two FMDVs compared were isolated at close time points, while lower values were calculated when the viruses were sampled from different animals at distant time points.

The dependence of the calculated rate of evolution during the epidemic spread of the virus on the time interval between virus isolations for sequence determination is expected for viruses that need not be transmitted by direct contact between an infected and a susceptible host. Some viruses remain infectious in the environment for prolonged time periods, until they reach a susceptible host in which to initiate replication rounds. This is the case of viruses transmitted by the fecal-oral route, such as enteroviruses. FMDV can adhere and remain infectious on many objects (fomites), including dust particles, food products with neutral pH, or insects that can transport the virus mechanically. Infectious FMDV can traverse long distances (many kilometers) on dust particles, people, trains, and the like. Even if some infectivity is lost, a few infectious particles are sufficient to infect an animal (Sellers, 1971, 1981). There are some classic examples of long-distance transport of FMDV, a virus subjected to close scrutiny due to its economic impact. One is the spread of SAT1 and A22 FMDV during the 1960s in Turkey along the railway line from the cattle-raising region of Lake Van to slaughterhouses in Istanbul [this and other examples are described in (Brooksby, 1981)]. Computer models have been developed to explain and predict possible airborne FMDV transmission in different geographical areas, the origin of epidemics, and the effectiveness of contingency plans (Sorensen et al., 2000; Tidesley et al., 2017). [As an anecdote, in my experience as a member of the Research Group of the Standing Technical Committee for the Control of FMD of Food and Agriculture Organization of the United Nations (FAO) in the 1980s, FMD outbreaks in any country always came from somewhere else]. A time-dependent bias in evolutionary rates for viruses has been amply documented (Duchene et al., 2014; Aiewsakun and Katzourakis, 2016), including when rates are compared between present-day isolates and those that circulated thousands of years ago (Muhlemann et al., 2018).

For viruses that can remain infectious outside their hosts, and that do not need donor-recipient host contacts to perpetuate transmission chains, the time between isolations will influence the calculated rate of evolution based on genomic nucleotide sequences. The reason is that during the extracellular stages, the virus will not undergo genetic change, at least to the extent of variation during intracellular replication (possible mutations due to chemical damage in viral genomes is indicated in Section 2.2. of Chapter 2). The effect of nonreplicative time intervals in the rate of evolution is illustrated in Fig. 7.3.

Some complications should be considered in the interpretation of the analyses depicted in Fig. 7.3: (i) the consensus sequences determined to characterize the virus shed by each animal represent a simplification of the real genome composition of the virus. (ii) Individual animals vary in physiological and immunological status, and, obviously, they are not in line waiting to be infected; they move, gather around water and food sources, some are isolated, others in close contact with their peer, and so on. (iii) In this case, virus transport is assumed to be mechanical (on dust particles carried by wind, aerosols, insects, etc.) without additional viral replication during transport. However, subpopulations of

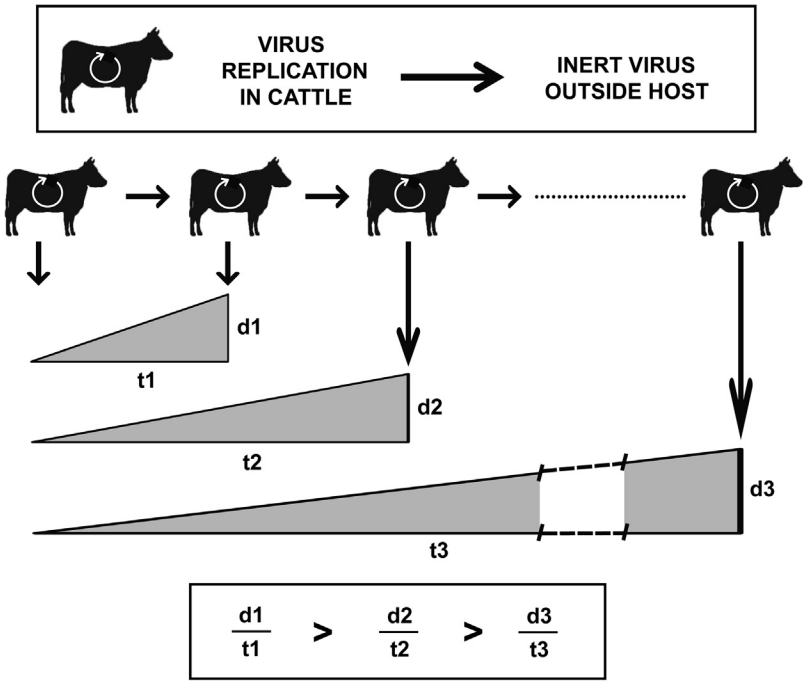

FIGURE 7.3 Inverse correlation between the time between viral isolations for consensus sequence determination, and the calculated rate of evolution. Animals sustain the replication (inside curved arrow) of a virus that will be transmitted to a susceptible animal. The time that the virus spends outside an animal (absence of replication) is depicted by a horizontal arrow. The time of virus isolation is given by $t1$, $t2$, and $t3$. The number of nucleotide differences in the virus isolates relative to the sequence of the initial reference virus (from the animal on the left) is given by $d1$, $d2$, and $d3$ (vertical arrows). Because of the increasing periods of stasis (addition of horizontal arrows), calculated rates of evolution given by the d/t ratio will be higher the shorter the time interval between isolations. See text for additional models of time effects of evolutionary rates, and references.

the most environment-resistant particles, or particles that adhere best to the transporter object, may bias the composition of the virus that will reach an animal to pursue replication. Such events, occurring for 10 to 100 rounds of host infections, render the appalling virus diversity described in Chapter 1 a bit less appalling. Since several additional environmental circumstances are changeable and unpredictable, it is unlikely that rates of viral evolution in nature can remain invariant on the basis of some internal principle of constant mutation occurrence (as if the accumulation of mutations was as monotonous as radioactive decay!).

7.3.2 Interhost versus intrahost rate of evolution

Additional observations against constant mutational input with time have been made with HIV-1 and human and avian hepatitis B virus (HBV). The main finding is that interhost rates of evolution are lower than intrahost rates, even under a comparable set of epidemiological parameters. Several proposals have been made to account for this difference. A.J. Leslie and colleagues described cytotoxic T lymphocyte (CTL)-escape mutants of HIV-1 from infected patients. Some of the mutants reverted to the

wild-type sequence after transmission to individuals negative for the human leukocyte antigen (HLA) alleles associated with long-term HIV-1 control (Leslie et al., 2004). Strong intrahost selective pressures and reversion of a part of the selected mutations upon transmission to a susceptible individual is one of the possible mechanisms behind diminished evolutionary rates when viruses from multiple host individuals are compared (Fig. 7.4, Box 7.1).

J.T. Herbeck, J.I. Mullins, and colleagues systematically observed lower nucleotide sequence divergence between HIV-1 isolates from different individuals sampled in primary infection than between isolates from individuals with advanced illness. HIV-1 regained some ancestral features when infecting a new host, again explaining a higher intrahost than the interhost evolutionary rate (Herbeck et al., 2006). In a study of HIV-1 transmission between several pairs of individuals over an 8-year period, A.D. Redd and colleagues reported that the viral populations found in the newly infected recipients were more closely related to ancestral sequences from the donor than to the sequences found in the donor near the time of transmission (Redd et al., 2012). Preferential transmission of ancestral sequences may also contribute to lower interhost than intrahost rates of evolution (Box 7.1).

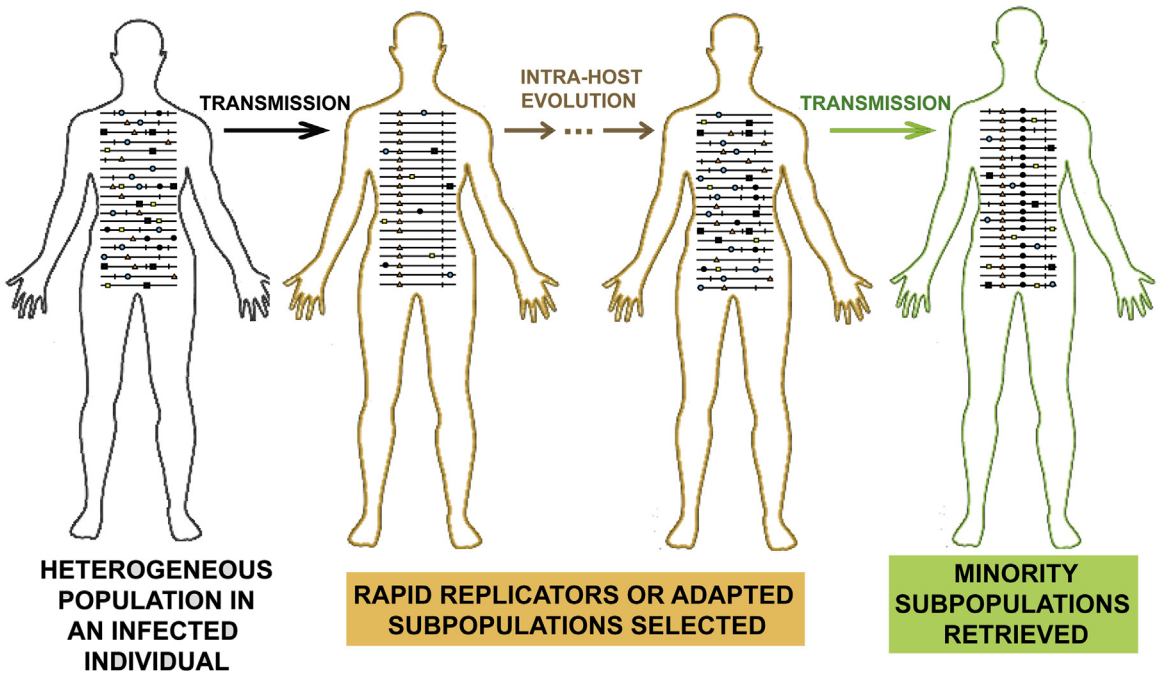

FIGURE 7.4 Scheme of a possible mechanism for faster intrahost than interhost virus evolution. Transmission events are represented by long arrows and intrahost evolution by short arrows (middle of the picture). The virus in the person on the left (black outline) has evolved to generate a complex mutant spectrum. However, only a subset of genomes are efficiently transmitted to the recipient person (brown outline). The virus in the recipient person evolves toward a complex mutant spectrum. Again, in this new mutant spectrum, only a minor set of genomes that resemble the ones in the first transmission is efficiently transmitted to the third person (green outline). The net result is that because at each transmission the genomes related to those that first entered the previous host have an advantage, rates of evolution will appear as slower than those within each host. Boxes at the bottom summarize the major event at each step. See text for additional related mechanisms and references.

BOX 7.1

MODELS FOR NONLINEAR RATES OF EVOLUTION

- For viruses that remain infectious in the extracellular environment, stasis due to the absence of replication will result in rates of evolution inversely correlated with the time between the isolation of the compared viruses.
- Adapt and revert. Mutants that permit adaptation to a new host individual revert upon transmission.
- Preferential transmission of ancestral sequences. Despite diversification in any host, ancestral sequences have a selective advantage in transmission. They may be retrieved from cellular memory (integrated provirus in HIV-1 or cccDNA in HBV).
- Colonization-adaptation trade-off. Sequential changes in the intensity of the host immune response favor dominance of some genome subpopulations over others. Upon transmission, ancestral minority subpopulations may become dominant.

K.A. Lythgoe and C. Fraser provided evidence that cycling of HIV-1 through long-lived memory $CD4^{+}T$ cells is probably the main contributing factor to slower HIV-1 evolution at the epidemic level (Lythgoe and Fraser, 2012). Ancestral sequences of HIV-1 in infected individuals may arise by the activation of proviral sequences kept in the form of quasispecies memory. In this case, we refer to the type of molecular memory defined as a reservoir, anatomical, or cellular memory in Section 5.5 of Chapter 5. A related type of reservoir memory is found in HBV, in the form of covalently closed circular DNA (cccDNA) that persists in the nuclei of infected hepatocytes, and acts as a template for the synthesis of pregenomic RNA and viral mRNAs (Kay and Zoulim, 2007). In this case, a record of ancient sequences is registered in the cccDNA. It should be noted that memory levels are dependent on fitness values, as evidenced experimentally with FMDV and expected from the theoretical basis of memory implementation (Chapter 5). In consequence, the most abundant memory genomes established early in an infection might be those displaying the highest fitness early in infection, and they might be better adapted to initiate infections than to sustain them (Fig. 7.4).

Additional mechanisms for the time dependence of evolutionary rates have been suggested for HBV. In a 17 years follow-up of several patients, HBV diversity increased during periods of active host immune response, and viral copy numbers decreased. When the immune response was weak, viral genome diversity decreased, and viral copy numbers increased; these periods are expected to be those of high transmissibility (Wang et al., 2010).

Endogenous hepadnaviruses are present in the genomes of several organisms. There is evidence that some of the integration events in avian hosts are at least 19 million years old. These integrated hepadnaviruses maintain about 75% nucleotide sequence identity with present-day hepadnaviruses, and the comparisons suggest that the long-term substitution rates are 10^{3}-fold lower than those for circulating avian HBVs (Gilbert and Feschotte, 2010). The permanence of viral genomic sequences in cellular DNA is a mechanism of evolutionary stasis, as it was emphasized in Chapter 3 with the comparison of the evolutionary rate of the

retroviral v-*mos* gene and its cellular counterpart c-*mos* (Gojobori and Yokoyama, 1985), among other evidence. Considerable evolutionary stasis is also observed by comparing isolates of HTLV-1 and HTLV-2, whose replication displays a preference for maintaining its integration in cellular DNA (Melamed et al., 2014). For viruses that have a dual potential of error-prone replication and cellular DNA-like stasis, the permanence in cellular DNA may also contribute to reduced long-term evolutionary rates.

HBV quasispecies dynamics was examined in a virus that infected members of the same family that presumably acquired the virus through mother-to-infant transmission (Lin et al., 2015). Again, the intrahost evolutionary rate was higher than the interhost rate, and the latter decreased with the number of transmissions. The differences were mainly due to nonsynonymous substitutions at limited sites. These observations were interpreted as a rapid switch of HBV between colonization (invasion of new host) and adaptation (quasispecies optimization in the new host). The authors referred to the colonization-adaptation trade-off (CAT) model or alternations of virus facing an environment marked by a limited host immune response followed by a period of active immune response. In the former environment, viruses displaying rapid replication are selected, while in the latter environment, HBV escape mutants with lower productivity are selected. In each transmission, when the virus reaches a new host, the previously adapted subpopulations are overgrown by the rapidly replicating ones. Again, cccDNA can serve as a reservoir of ancient sequences.

In agreement with these proposals, rates of evolution measured in a single infected individual persistently infected with a continuously replicating virus tend to be higher than those observed with the same viruses isolated from different individuals (Morse, 1994; Domingo et al., 2001; Domingo, 2006). Slowly evolving viral genes may nevertheless undergo episodes of rapid evolution and, vice versa, a rapidly evolving gene may be transiently static. This should be considered in statistical approaches to evolution (Gaucher et al., 2002).

The viruses that colonize the human gut are extremely diverse and evolving at different rates. From all the evidence, the gut virome composition is unique to each individual human. It includes single-stranded DNA viruses that evolve at rates $>10^{-5}$ substitutions per nucleotide and day (an extremely high rate!); other genomes, particularly those of temperate bacteriophages, display lower evolutionary rates, partly due to their being replicated by high fidelity bacterial polymerases when the prophage is integrated into host DNA (Minot et al., 2013). Thus, both the viral populations that infect our tissues and organs and those that infect the microbes that colonize us (1%–3% of the mass of a healthy human is composed of microorganisms) are diverse, unique to each of us, and evolving at different rates.

The reader will find in the literature additional proposals to account for the contrast between rapid intrahost evolution versus relative long-term conservation when comparing ancestral sequences [(Ali and Melcher, 2019; Simmonds et al., 2019), among other studies; see also Section 7.3.1]. It would not come as a surprise if several factors acted conjointly to produce the observed rate discrepancies. There is a need to integrate the approaches and conclusions of short-term evolution centered on quasispecies dynamics, and of long-term evolution based on the phylodynamics methodology (Geoghegan and Holmes, 2018).

7.3.3 Rate discrepancies and the clock hypothesis

The molecular clock hypothesis dwindles as a conceptual framework because, from all evidence, virus evolution is far from being dictated by a steady accumulation of mutations in viral genomes. The major event that we have learned by comparing the genomic composition of

individual viral populations is that evolution is guided by "replacement of subpopulations" rather than "accumulation of mutations in genomes." This conceptual change is as important for the understanding of long-term evolution of viruses as it is the consideration of the wild type as a cloud of mutants in the definition of a viral population, a major input from quasispecies theory (Chapter 3).

There are additional arguments against the operation of a molecular clock in virus evolution. According to the clock hypothesis, the rate of accumulation of mutations coincides with the rate at which mutations arise in infected individuals. This holds for neutral mutations, and as mentioned in several chapters of this book, very few mutations occurring in highly compact viral genomes are truly neutral (with no functional consequences in any environment; see Section 2.3 in Chapter 2). Even for neutral evolution, spatial asymmetries in populations are sufficient to perturb the molecular clock rate, as documented with a theoretical model of broad applicability (Allen et al., 2015). However, quasispecies and the operation of a clock are not totally irreconcilable. An epidemiological study with FMDV suggested that viral quasispecies could produce a transient molecular clock due to the periodic sampling of components of the mutant spectrum in transmission (Villaverde et al., 1991). In this case, both time differences between transmissions and spatial heterogeneities would blur the transiently observed regularity.

7.4 Long-term antigenic diversification of viruses

Viruses can change their antigenic properties gradually, in a process termed antigenic drift, or suddenly, in a process termed antigenic shift. The distinction between antigenic drift and shift was established with IV (Gething et al., 1980; Webster, 1999; Parrish and Kawaoka, 2005). The shift in IV is due to genome segment reassortment that incorporates new hemagglutinin or neuraminidase genes. In monopartite viruses, the difference between gradual and drastic antigenic change has also been established (Martínez et al., 1991b).

The antigenic diversification of one FMDV serotype was examined over a six-decade period by comparing amino acid sequences of the major antigenic sites of the virus isolated in three continents (Martínez et al., 1992). The evolution of the capsid genes was associated with linear accumulation of synonymous mutations, but not of amino acid substitutions. Remarkably, the antigenic variation over 6 decades was due to fluctuations among limited combinations of amino acid residues without net accumulation of amino acid substitutions over time (Fig. 7.5). This result suggests that constraints at the protein level may maintain a long-term virus identity at the antigenic level. In a related observation on the inter-host evolution of HIV-1 mentioned in Section 7.3.2, HIV-1 recovered ancestral features when infecting a new host (Herbeck et al., 2006). Thus, multiple constraints in viruses may limit the rate and mode of long-term diversification, resulting in different numbers of circulating serotypes among viruses of the same family.

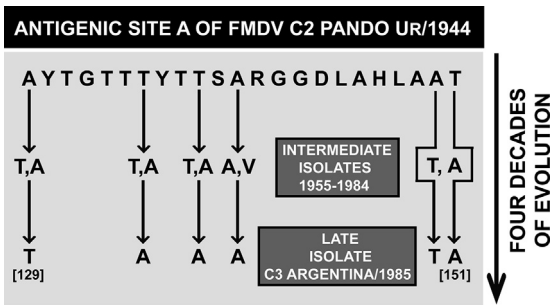

FIGURE 7.5 Evolution of a major antigenic site of FMDV over 4 decades. The sequence at the top is that of amino acid residues 129–151 of capsid protein VP1 of FMDV C2 Pando, isolated in Uruguay in 1944. Key amino acid positions did not diverge in a linear fashion and isolates over 4 decades displayed only two types of amino acid residues at each position. See text for possible mechanisms and references.

7.4.1 Widely different number of serotypes among genetically variable viruses

A puzzling question in evolutionary virology is that despite sharing high mutation rates, some viruses display extensive antigenic diversity in nature reflected in multiple serotypes, while other viruses maintain a relatively invariant antigenic structure, with only one serotype recorded. For the latter group of viruses, the same vaccine can maintain its efficacy over many decades; examples are rabies virus (RV) and MV, two RNA viruses that show remarkable genetic diversity in nature, and estimates of mutation rates and frequencies comparable to other RNA viruses. Antigenic constancy is a determinant of long-lasting immunity after infection or vaccination. MV infection produces lifelong immunity (probably as a result of several factors) while patients that have cleared the hepatitis C virus (HCV) can be reinfected by the same virus. Cases of patients infected with HCVs of different genotypes and subtypes are increasingly identified, as more refined diagnostic tests are utilized, and the virus diversifies in nature (Hedskog et al., 2019).

No correlation between virus structure (or its morphotype) and antigenic diversity has been found. Among structurally closely related viruses, differences in antigenic diversity are apparent. A dramatic case is that of the picornaviruses, since encephalomyocarditis virus (EMCV) or hepatitis A virus (HAV) have a single serotype, while human rhinoviruses (HRVs) have been divided into more than 100 serotypes. Other picornaviruses have intermediate numbers of serotypes: three in the case of PV and seven in the case of FMDV. Although it may seem that a diverse antigenic structure may predict a broad host range, this is actually not the case. HAV is highly specialized for the human host, while EMCV infects more than 30 species, including mammals, birds, and invertebrates (Knowles et al., 2010).

Several, not mutually exclusive models, have been proposed to account for differences in the antigenic stability (number of serotypes) among viruses:

- Differences in mutation rate, either the average value for the entire genome or the local mutation rate at the genomic sites that encode antigenic determinants.
- The presence of some dominant and invariant antigenic sites that evoke long-lasting antibodies in the infected hosts, and that obscure other antigenic sites that produce different antibodies that have a limited impact on the antigenic profile of the virus.
- Differences among the assays used for serotype classification. If a universal and standard procedure to classify virus isolates in different serotypes were applied, differences among viruses would be largely lost.
- Difference in the history of virus circulation. Ancient viruses that undergo many rounds of genome replication in each infected host have had an opportunity to diversify antigenically in a manner not possible with viruses that have a more limited history of circulation among susceptible hosts. According to this model, antigenic diversification of some viruses currently viewed as antigenically invariant will take place during the next hundreds of years if their circulation continues.
- Some viruses have antigenic sites that cannot vary because they are under severe constraints to accept amino acid substitutions. Antigenic variants may exist as low-fitness subpopulations, but their frequency is too low to modify the results of the diagnostic tests used for serological classifications.

Consideration of these possibilities requires examining some experimental data on virus antigenicity. First, as a conceptual precision, we assume that the number of serotypes is essentially determined by amino acid sequences

located in the virus particle and that either directly or indirectly can affect the interaction of the virus with antibodies. Neutralizing and nonneutralizing antibodies may contribute to serological distinctions, depending on the assays performed for serotyping. Serum neutralization tests will identify differences in sensitivity to neutralization, while Enzyme-Linked Immuno-Sorbent Assay (ELISA) tests will capture reactivity by all raised antibodies.

Antibodies can be obtained from infected natural hosts, or from some laboratory animals which are not a natural host for the virus. An ensemble of amino acid residues forms an antigenic determinant which is usually composed of multiple epitopes [defined here as a unit of interaction with a monoclonal antibody (MAb)]. Epitopes can be either continuous (also termed linear) or discontinuous (also termed structured). Continuous epitopes are those whose primary amino acid sequence has the information to react with the cognate antibody. Discontinuous epitopes are those whose reactive residues come from distant positions of the same protein or residues of different proteins. Many overlapping epitopes can be found within the same antigenic site. Epitopes can include modified amino acid residues such as glycosylated amino acids. Reactivity of discontinuous epitopes with the cognate antibody is generally lost as a consequence of denaturation of the proteins that form the epitope.

With these introductory clarifications, we are now in a position to examine the different possibilities listed above to account for differences in the number of serotypes among genetically variable viruses.

There is no correlation between limited antigenic diversity and low average mutation rate. Mutation rates and frequencies for RNA viruses fall in the range of 10^{-5} to 10^{-3} substitutions per nucleotide (Chapter 2). However, mutation rates along a viral genome are not uniform, as evidenced by the occurrence of hot spots for variation. Influences such as nucleotide sequence context or RNA structure may conceivably alter mutation rates. It was proposed that a predicted double-stranded RNA at the region encoding the major antigenic site of FMDV might increase the polymerase error rate locally and give rise to multiple amino acid substitutions (Weddell et al., 1985). While at some specific sites polymerases may be more error-prone than average, subsequent evidence for FMDV indicated that antigenic variation is due to amino acid substitutions at different antigenic sites and that even variation at the major site can be mediated by distant amino acids on the viral capsid (Rowlands et al., 1983; Geysen et al., 1984; Mateu et al., 1990; Feigelstock et al., 1996). Later molecular studies have not provided evidence that viruses may have a large number of serotypes because their polymerases are more error-prone when copying regions encoding amino acids that belong to antigenic sites. Therefore, the possibility that differences in mutation rates can determine a different number of circulating serotypes is unlikely.

Most viruses include multiple antigenic sites, and antibodies are raised against several surface proteins to produce an array of neutralizing and nonneutralizing antibody molecules. Taking picornaviruses again as an example, the number of antigenic domains (each composed of multiple epitopes) varies between one and four (Mateu, 1995, 2017; Fry and Stuart, 2010). There is no evidence that a restriction on the number of sites or epitopes or that the expression of a salient class of antibody molecule may explain a 100-fold difference in the number of serotypes among picornavirus genera. Thus, the second proposal is unlikely to be correct.

The difference among classification assays argument does not have an easy response. Indeed, there is no universal procedure used to classify viruses serologically, and therefore, strictly speaking, there is the possibility that a different number of serotypes could be obtained using alternative classification procedures.

FMDV is a pertinent example. Its seven serotypes are defined on the basis of a very stringent test that cannot be performed with human viruses for obvious reasons: the absence of cross-protection resulting from vaccination or infection with a given FMDV. Infection or vaccination with FMDV of one serotype does not confer protection against FMDV of a different serotype. In contrast, the subtype classification of FMDV was based on serological assays, such as cross-neutralization or complement fixation tests, usually using sera raised in guinea pigs. These assays allowed classification of FMDV in more than 65 serological subtypes. Subtyping was stopped when it was realized that using increasingly discriminatory assays such as reactivity with MAbs, virtually any new isolate could define a new subtype [(Mateu et al., 1988); see (Domingo et al., 1990; Sobrino and Domingo, 2004; Mateu, 2017) for review of serotype and subtype classification of FMDV]. Despite these considerations, it is unlikely that serological assays using in vitro tests would be responsible for a 100-fold difference between two human pathogens such as HRV and HAV. Thus, it does not seem justified to attribute antigenic constancy to an artifact derived from diagnostic procedures.

More extensive virus circulation will favor genetic and antigenic diversification, and a single serotype may evolve into multiple serotypes. What we describe today as the antigenic profile of virus groups is a snapshot of an evolving process. Genotype differentiation is actually being witnessed during the expansion of HCV pandemics, partly due to a true genetic diversification of the virus as it circulated over the last decades, and partly due to increasing capacity of virus surveillance, and molecular and phylogenetic tools for genome analysis. The reader can find an illustration of this point by comparing the expanded phylogenetic HCV tree from six to seven genotypes and the subtype ramifications, published by P. Simmonds and colleagues in 1993 and 2014 [compare (Simmonds et al., 1993) and (Smith et al., 2014)], and continuing new subtype identification (Hedskog et al., 2019). Although it cannot be excluded that time might tend to equalize the number of serotypes among viruses, current evidence does not justify blaming differences in the extent of virus circulation to settle this issue.

We come to constraints at antigenic sites that limit the number of accepted amino acid substitutions as a model for antigenic invariance. It is the preferred model of molecular virologists. The initial concept was proposed by M.G. Rossmann, in his canyon hypothesis (Rossmann, 1989), based on studies with HRV14. A canyon in the virus preserves the receptor-binding site inside while permitting amino acid substitutions that affect antigenicity, without consequences for receptor recognition. A physical and functional separation between receptor and antibody binding allows extensive antigenic variation. Clearly, in many viruses, there is an overlap between antigenic and receptor recognition sites (Section 4.5 in Chapter 4), that could limit antigenic variation. A difference in constraints is also supported by a structural comparison carried out by J.M. Casasnovas and his colleagues of the interaction of PV and HRV16 with their respective cellular receptors that revealed a receptor-binding site more accessible in PV than in HVR16, rendering the latter suited to escape antibody neutralization (Xing et al., 2000). This would render HRV a picornavirus prone to antigenic variation, as indeed found in nature. Thus, constraints imposed by the requirement to interact with the cellular receptor may explain the limited capacity for antigenic diversification, and perhaps with the contribution of other influences, the puzzle of widely different antigenic types despite similarly high genome mutability. Additional structural and functional studies with viruses of different families are necessary to substantiate this proposal.

7.4.2 Similar frequencies of monoclonal antibody-escape mutants in viruses differing in antigenic diversity

Fitness cost is expected to limit antibody escape (Louie et al., 2018) but this may be reflected in mutant stability but not in the frequency of MAb-escape mutants [monoclonal antibody-resistant mutant (MARM) frequencies] in laboratory experiments. Comparison of MARM frequencies of different viruses shows comparable values independently of the number of circulating serotypes of the virus (Table 7.2). In particular, the cardiovirus Mengo virus (one serotype) displays similar MAR frequencies than HRV (100 serotypes). In fact, none of the RNA and DNA viruses listed in Table 7.2 deviate from a broad range of MARM frequencies of 10^{-3} to 10^{-5}, except for substitutions at some discontinuous epitopes of FMDV (Lea et al., 1994). In some studies, the stability of the selected escape mutants was tested after a few passages in cell culture, but in other studies, the lack of reversion of the antigenic change was not ascertained. Two FMDV escape mutants showed a selective disadvantage over the parental wild-type virus (fitness decrease); upon continued replication, the mutants acquired fitness-enhancing mutations without reversion of the antigenic change (Martínez et al., 1991a).

Unless the escape mutations are selectively neutral, the expectation is that MARM frequencies may be an underestimate of the real rate at which the amino acid substitutions occur. Thus, it is possible that following selection by an antibody, some mutants may decrease in frequency due to a fitness cost, or that their level is maintained due to additional compensatory mutations acquired by the replicating genomes (Fig. 7.6). Viruses that are highly constrained for antigenic variation may be diagnosed through fitness decrease of MARM mutants despite their occurring at similar rates as those that affect unconstrained sites. This is a concept similar to the distinction between fitness and function that we made in Section 5.8 of Chapter 5. That is, the occurrence of an antigenic change

TABLE 7.2 Frequency of monoclonal antibody-resistant mutants (MARMs) for some viruses.

Virus	Monoclonal antibody-resistant mutant (MARM) frequencies	References
Poliovirus	10^{-4} to 10^{-5}	Emini et al. (1982) and Minor et al. (1983, 1986)
Mengovirus	3×10^{-3} to 5×10^{-5}	Boege et al. (1991)
Foot-and-mouth disease virus	10^{-4} to 10^{-5} (continuous epitopes)	Martínez et al. (1991a)
	10^{-4} to 10^{-7} (discontinuous epitopes)	Lea et al. (1994)
Rhinovirus	10^{-4} to 10^{-5}	Sherry et al. (1986)
Hepatitis A virus	3×10^{-3}	Stapleton and Lemon (1987)
Vesicular stomatitis virus	0.5×10^{-4} to 1×10^{-4}	Holland et al. (1990)
Rabies virus	10^{-4}	Wiktor and Koprowski (1980)
Measles virus	9×10^{-5}	Schrag et al. (1999)
Sindbis virus	10^{-3} to 10^{-5}	Stec et al. (1986)
Canine parvovirus	$10^{-3.4}$ to $10^{-5.4}$	Smith and Inglis (1987)
Herpes simplex virus	1×10^{-5}	Smith and Inglis (1987)

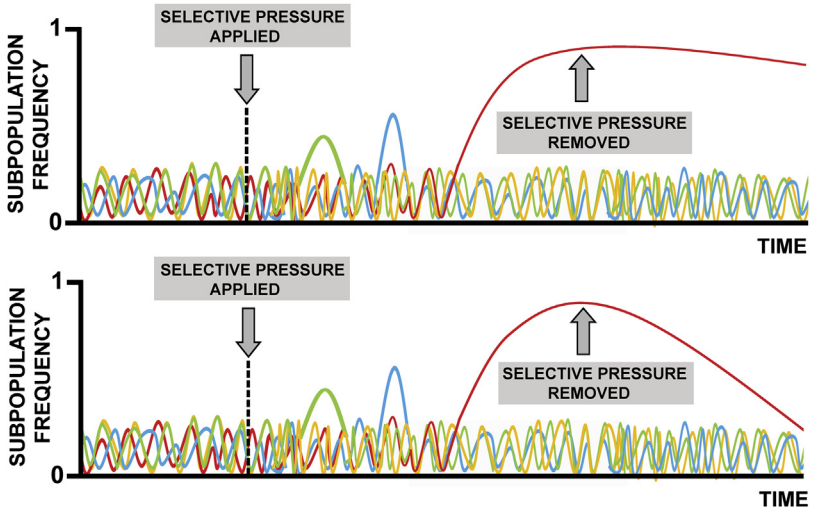

FIGURE 7.6 Stability of selected mutant subpopulations when selective pressure is removed. This scheme is the same portrayed in Fig. 3.6 of Chapter 3, but shifted toward the right of the time scale. Five different colors have been chosen to depict fluctuations of four genomic classes. In a real population, thousands of genomes may be involved in each infected cell. In the present diagram, the red line represents genomes selected for their resistance to neutralizing monoclonal or polyclonal antibodies. Once the antibody pressure is removed, the mutant genomes may remain dominant either because the relevant substitutions do not affect viral fitness or because compensatory mutations have been acquired (top diagram). In contrast, upon removal of the antibody pressure, the proportion of selected genomes may fade with time due to a fitness cost and absence of compensatory mutations (bottom diagram). In the latter case, the antibody-resistant mutant will not contribute to the long-term antigenic diversification of the virus.

does not guarantee that the change will be perpetuated in nature and contribute to natural antigenic diversification. Again, fitness should be considered as a relevant parameter, and fitness effects on antigenic stability have been largely unexplored.

7.5 Comparing viral genomes. Sequence alignments and databases

The likely multiple origins of viruses, followed by extended events of interaction with evolving host organisms of all phyla have produced myriad viral particles that at the present time outnumber cells by a factor of 10 (Chapter 1). The way to put order into such diversity is to classify viruses as done periodically by the International Committee on Taxonomy of Viruses (ICTV) (http://www.ictvonline.org/).

Computational procedures developed to study phylogenetic relationships in evolutionary biology are routinely applied to virology to establish relationships among closely or distantly related viruses (Page and Holmes, 1998; Hall, 2001; Felsenstein, 2004; Salemi and Vandamme, 2004; Yang, 2006; Russell, 2014). No phylogenetic tree that connects the viruses that have been characterized to date can be derived in a reliable way, not even a tree for DNA or for RNA viruses. What we can do is to produce trees for related viruses that probably share a common ancestor. Many data banks are available for viruses to retrieve sequences for comparison with new isolates. Despite the fact that data banks are periodically updated, some are listed in Table 7.3, and can serve as the starting point to reach the desired uniform resource locator (URL) to implement a procedure for genome characterization. Prior to any

TABLE 7.3 Information on nucleotide and amino acid sequence alignment programs, Data Banks, and phylogenetic procedures.

Identification	URL (or reference)	Contents
EMBL Nucleotide Sequence Database	http://www.ebi.ac.uk/ena	All reported sequences. General database
GenBank, the NIH Genetic Sequence Database	http://www.ncbi.nlm.nih.gov/genbank	All reported sequences. General database.
DNA Data Bank of Japan	http://www.ddbj.nig.ac.jp	All reported sequences. General database.
UNIPROT	http://www.ebi.ac.uk/uniprot	Protein sequences.
Protein Data Bank (PDB)	http://www.rcsb.org/	Protein structure data.
Virus Particle Explorer	http://viperdb.scripps.edu/	Virus structures and structure-derived properties: capsid interactions, residue contributions to protein-protein interactions. Links to sequences and taxonomic information.
Viral Genomes Project	http://www.ncbi.nlm.nih.gov/genome/viruses	Complete or nearly complete viral genome sequences. Additional information. Includes Pairwise Seqeunce Comparisons (PASC) within viral families.
The Influenza Sequence Database	http://www.fludb.org/	Sequences, tools for the analysis of hemagglutinin and neuraminidase sequences.
Picornavirus Sequence Database	http://www.viprbrc.org/brc/home.spg?decorator=picorna	Sequence and specific references for different picornavirus isolates.
Plant viruses	http://www.dpvweb.net	Description of Plant Viruses (DPV). Expanding data bank of viruses, viroids and satellites of plants, fungi and protozoa.
Potyvirus Database	http://www.danforthcenter.org/iltab/potyviridae/	Taxonomy, references, and sequence databases of members of the *Potyviridae* family.
Calicivirus Sequence Database	http://www.viprbrc.org/brc/home.spg?decorator=calici	Sequences, information, and specific references for different calicivirus isolates.
HPV Sequence Database	http://pave.niaid.nih.gov/	Human papilomavirus, sequences, analysis, and alignment tools.
HIV Sequences Database	http://www.hiv.lanl.gov http://www.hiv.lanl.gov/content/sequence/RESDB/	Sequences, drug resistance. Molecular immunology and vaccine trials. Analysis tolls.
HIV Drug Resistance Database	http://hivdb.stanford.edu/	Sequence, correlations genotype-phenotype, and genotype-antiretroviral treatment. Sequence analysis tools.
Hepatitis C virus Database	http://hcv.lanl.gov/ http://www.hcvdb.org	Sequences and genome analysis tools.
Hepatitis virus Database	http://s2as02.genes.nic.ac.jp/	Hepatitis B, C and E virus sequences.

TABLE 7.3 Information on nucleotide and amino acid sequence alignment programs, Data Banks, and phylogenetic procedures.—cont'd

Identification	URL (or reference)	Contents
Poxvirus Bioinformatics Research Center	http://www.poxvirus.org/	Poxvirus genomes.
Viral Bioinformatics Resource Center	http://athena.bioc.uvic.ca/	Large DNA viruses (Poxviruses, African Swine Fever Viruses, Iridoviruses, Baculoviruses).
Human Endogenous Retroviruses Database	http://herv.img.cas.cz	Human endogenous retroviruses, and genome analysis tools.
VIDA, Virus Database at University College London	http://www.biochem.ucl.ac.uk/bsm/virus_database/VIDA.html	Homologous protein families from herpes, pox, papilloma, corona and arteriviruses.
Subviral RNA Database	http://subviral.med.uottawa.ca/cgi-bin/home.cgi	Sequences and prediction of RNA secondary structures.
Vir Oligo Compilation Lab	http://viroligo.okstate.edu/	Database of oligonucleotides used in virus detection and identification. Technical information and several links to original sequence information.
ViPR Virus Pathogen Resource	http://viprbrc.org/	Sequences of multiple virus families.
Chromas	http://technelysium.com.au/?page_id=13	Software for DNA sequencing.
FORMAT CONVERSION	http://hcv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html	Program that converts the sequence(s) to a different user-specified format.
BEAST, Tracer	http://beast.bio.ed.ac.uk/Tracer	Phylogenetic inferences. Bayesian methods.
MODELTEST	http://darwin.uvigo.es/software/modeltest.html	Determination of nucleotide substitution model. Phylogenetic derivations.
PHYLIP package	http://evolution.genetics.washington.edu/phylip.html	Programs for inferring phylogenies.
PAUP	http://paup.csit.fsu.edu/	Software for inference of evolutionary trees.
BiBiServ	http://bibiserv.techfak.uni-bielefeld.de/splits	Bielefeld Bioinformatics Service.
Bowtie2	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml	It is a tool for aligning sequencing reads to long reference sequences.
SAM tools (Sequence Alignment/Map)	http://samtools.sourceforge.net/	A generic format for storing large nucleotide sequence alignments.
Free Bayes	https://github.com/ekg/freebayes	FreeBayes is a Bayesian genetic variant detector designed to find small polymorphisms.
EMBOSS	http://www.ebi.ac.uk/Tools/emboss/	Sequence analysis.
T-Coffee	http://www.tcoffee.org/	Tools for Computing, Evaluating, and Manipulating Multiple Alignments.

(*Continued*)

TABLE 7.3 Information on nucleotide and amino acid sequence alignment programs, Data Banks, and phylogenetic procedures.—cont'd

Identification	URL (or reference)	Contents
IGV, Integrative Genomics Viewer	http://www.broadinstitute.org/igv/	The Integrative Genomics Viewer (IGV) is a high-performance visualization tool for interactive exploration of large, integrated genomic datasets. It supports a wide variety of data types, including array-based and next-generation sequence data, and genomic annotations.
CLC Genomics Workbench	https://www.qiagenbioinformatics.com/	Analysis of deep sequencing data
RDP4, version 4.36 beta	P.D. Martin et al. Virus Evol. 1, vev003, 2015	Detection of recombination
Dintor	https://dintor.eurac.edu/	Tools for the analysis of genomic and proteomic data sets
Geneious Basic	https://www.geneious.com/	Organization of biological data, including nucleotide and amino acid sequences
G-large−INS−1 in MAFFT	https://mafft.cbrc.jp/alignment/software/mpi.html	Large-scale sequence alignments
WIT	http://www.algorithm-skg.com/wit/home.html	Alignment of large number of reads obtained by deep sequencing
MEGA 7	https://www.megasoftware.net/	Extended MEGA version for the analysis of large data sets
Quasispecies Diversity	https://www.bioconductor.org/packages/release/bioc/html/QSutils.html	Set of utility functions for viral quasispecies analysis with deep sequencing data
Meta PGN	https://github.com/peng-ye/MetaPGN	Visualization of pangenome networks

comparative study of nucleotide or amino acid sequences (not only to establish phylogenetic relationships, but also to calculate genetic distances, to identify regulatory regions, functional domains, and structural motifs, to design oligonucleotide primers for amplification, or other applications) it is essential to align sequences accurately, and some programs for sequence alignments are also given in Table 7.3, including processing of large data sets obtained by deep sequencing (Kearse et al., 2012; Kumar et al., 2016, 2019; Lai and Verma, 2017; Nakamura et al., 2018), and visualization of pangenome networks (Peng et al., 2018).

Databases differ in format and contents, which may include prediction of traits derived from sequence information (RNA secondary structures, antiviral drug sensitivity levels, assignments to homologous protein families, etc.). Some of them offer a link with the *web* page of the ICTV, thus providing background information to assign newly determined sequences to current taxonomic groups. A structure-based amino acid sequence alignment of protein homologs can be carried out based on three-dimensional structures of proteins. Such types of amino acid sequence alignments may help in the identification of relevant

structural and functional motifs. Sequence variability among a set of aligned sequences can be quantitated by the number of variable sites, mean pairwise diversity, mutation frequency, and other estimators (i.e., the Watterson's estimator) (Page and Holmes, 1998; Mount, 2004; Salemi and Vandamme, 2004) (see also Chapter 3 for parameters used to quantify mutant spectrum complexity).

Relevant information on protein evolution can be derived from alignment of the protein sequence of related viruses (or of isolates from one virus, or for components of the same mutant spectrum) and analyzing the statistical acceptability of the divergent amino acids at each position (Feng and Doolittle, 1996) with consideration of amino acid location when a three-dimensional structure is known (Farheen et al., 2017). Statistical acceptability derives from the chemical nature and shape of the amino acid side chains, their structural context, and also the limitations that the genetic code imposes on amino acid replacements (Porto et al., 2005). The basic assumption is that the more conserved the amino acid sequences, and the more similar are the variant amino acids, it is more likely that the proteins are derived from a common ancestor. M. Dayhoff pioneered the early comparison of protein sequences establishing a protein information resource (PIR) in the middle of the 20th century. Tables named PAM (percent accepted mutation) were constructed, and several evolved versions such as BLOSUM matrices, based on the BLOCKS database, are used to compare protein sequences. The BLOSUM62 amino acid substitution matrix groups amino acids according to their chemical structure and provides a probability of occurrence of each amino acid replacement: zero, amino acid replacement expected by chance; positive number, replacement found more often than by chance; and negative number, replacement found less often than by chance.

7.6 Phylogenetic relationships among viruses. Evolutionary models

The URLs listed in Table 7.3 give access to computational analyses that allow sequence alignments and derivation of phylogenetic trees, which are extremely informative of middle- and long-term evolutionary change of viruses (Page and Holmes, 1998; Notredame et al., 2000; Mount, 2004; Salemi and Vandamme, 2004; Holmes, 2008, 2009). Application of phylogenetic methods to virus evolution requires careful consideration of the evolutionary models to be used, including probabilities of the different types of nucleotide and amino acid replacements, and the rates at which they may occur. Statistical methods (i.e., likelihood ratio tests) are available to select an adequate model for a given data set (Salemi and Vandamme, 2004). At the nucleotide sequence level, it is often assumed that when transitions are more frequent than transversions in a set of related sequences, no saturation of mutation took place. In contrast, when transversions are more frequent than transitions, saturation is presumed (Xia and Xie, 2001). Parameter α applied to amino acid sequence alignments (e.g., using the program AAml from PAML package, version 3.14) takes into account multiple amino acid replacements per site, as well as unequal substitution rates among sites (Yang et al., 2000). Parameter α can be calculated using the amino acid replacement matrix WAG available in the program MODELTEST (Posada and Crandall, 1998). Despite their obvious utility, it is unlikely that these statistical procedures which were developed on the assumption of successions of defined sequences (rather than mutant clouds) can capture the complexities underlying long-term evolution of viruses in nature.

Phylogenetic reconstructions based on nucleotide (and deduced amino acid) sequence alignments are generally possible with selected genes

of relatively close viruses (i.e., that belong to the same family). The main methods used to derive evolutionary trees are: maximum parsimony, distance, ML, Bayesian methods of phylogenetic inference, and splits-tree analysis [reviewed in (Eigen, 1992; Page and Holmes, 1998; Mount, 2004; Salemi and Vandamme, 2004; Sullivan, 2005; Holmes, 2008; Beale et al., 2018; Geoghegan and Holmes, 2018)] (Table 7.3).

Maximum parsimony predicts the minimal mutation steps needed to produce the observed sequences from ancestor sequences. It is most suitable for closely related sequences. Often, all possible trees are examined before a consensus tree is produced, and, therefore, the method is time-consuming. Most programs based on maximum parsimony assume the operation of a molecular clock, with the limitations that were discussed in Section 7.3.

Distance methods are based on the calculation of genetic distances between any two sequences of a multiple sequence alignment. Large genetic distances require a correction for multiple mutational steps (i.e., Kimura 2-parameter distance). Most distance methods can handle large numbers of sequences, and results are relatively reliable even when a molecular clock does not operate. Commonly applied distance methods include neighbor-joining (NJ) (that does not assume a molecular clock and yields an unrooted tree), several variant versions of NJ, and the unweighted pair group method with arithmetic mean (UPGMA, a clustering method that assumes a molecular clock and produces a rooted tree). The software package TREECON was developed to derive NJ trees.

ML methods use probability calculations to derive a branching pattern from the mutations at different positions of the nucleic acids under study. They can estimate both distances and the most accurate mutational pathway between sequences. Generally, supercomputers are needed when many sequences are compared since all possible trees are examined. ML methods are included in several programs listed in Table 7.3. Bayesian methods (based on conditional probabilities derived by Baye's rule) (Huelsenbeck et al., 2001; Ronquist and Huelsenbeck, 2003; Huelsenbeck and Dyer, 2004; Tonkin-Hill et al., 2019) have the advantage of increased speed of data processing, but they still require time to avoid incorrect inferences.

Splits-tree procedures are based on split-decomposition theory or statistical geometry, and they provide a geometrical representation of the distance relationships in sequence space (Eigen, 1992; Dopazo et al., 1993; Salemi and Vandamme, 2004) (Chapter 3). The procedure has been used to analyze rapidly evolving viral sequences (http://bibiserv.techfak.uni-bielefeld.de/splits/); methods that allow the inclusion of insertions and deletions have been adapted to the splits-tree program (Cheynier et al., 2001). Phylogenetic trees can be presented as rooted trees (with a reference out-group) and unrooted trees.

When possible it is advisable to apply different phylogenetic procedures to compare tree topologies. Resampling methods (i.e., bootstrapping, jackknifing, etc.) are used to assess the statistical reliability of the trees (Page and Holmes, 1998; Salemi et al., 1998; Mount, 2004; Salemi and Vandamme, 2004). A tree defines clades or lineages of a virus attending to groupings by relatedness. Different tree topologies can be obtained when analyzing different genes of the same virus set. Discordant phylogenetic positions of two different genes of the same virus are suggestive of recombination that should be evaluated statistically (Worobey, 2001; Salemi and Vandamme, 2004; Martin et al., 2005). Recombination is frequent, and in some viruses recombination is intimately linked to the replication mechanism (Chapters 2 and 10).

The more conserved genes (i.e., those encoding the polymerase or other nonstructural proteins) may permit the establishment of phylogenetic relationships among some distant virus groups. Examples are the clustering of a number of animal and plant RNA viruses as

supergroups (Morse, 1994). Families of DNA-dependent DNA polymerases group some bacterial and bacteriophage DNA polymerases with some eukaryotic polymerases (Morse, 1994; Villarreal, 2005), in support of the active exchange of modules during coevolution of viruses and their hosts (Botstein, 1980, 1981; Zimmern, 1988). In contrast to conserved genes, variable genes (typically, capsid proteins and surface glycoproteins) serve to establish short-term evolutionary relationships within the same virus group, including the survey of virus variation during outbreaks, epidemics, and pandemics (Gorman et al., 1992; Martínez et al., 1992; Morse, 1994; Gavrilin et al., 2000).

Distantly related viruses, with no discernible nucleotide or amino acid sequence identity, can sometimes be grouped on the basis of the three-dimensional structures of viral proteins. The evolutionary trace (ET) clustering method combines phylogenetic partition of sequences with structural information (Chakravarty et al., 2005), and it may help in identifying functionally relevant domains shared by divergent isolates in particular highly variable capsid and surface viral proteins. ET can be applied to proteins and nucleic acids, and its clustering features may reveal conserved structures that are overlooked when all sequences are compared together. As explained in Chapter 1, the great diversity of amino acid sequences recorded among viral structural proteins (several URL links in Table 7.3) are actually reduced to a limited number of morphotypes at the structural level. In another approach, the probabilities of equivalence between pairs of residues in viral proteins are converted into evolutionary distances (Bamford et al., 2005; Ravantti et al., 2013). The structure-based classification has grouped the coat protein of icosahedral viruses in separate classes, each of which, interestingly, embraces different domains of life (Archaea, Bacteria, and Eukarya) (McTavish et al., 2017). A lineage of structurally related viruses includes tailed bacteriophages and the herpesviruses, suggesting that parts of the genomes of complex viruses may have a very ancient origin. They might have belonged to viruses that infected primitive cells before the latter diverged into the domains of life that we identify in our biosphere (Bamford et al., 2005; Villarreal, 2005) (compare with models of virus origins in Chapter 1).

Viral clades may cluster with clades of their host species, suggesting either virus-host coadaptation or an extended parasite-host relationship, with limited possibilities of jumping the host barrier (Section 7.7). Hantaviruses and their rodent hosts (Plyusnin and Morzunov, 2001), lyssaviruses and bat species, spumaviruses and their primate hosts, and herpesviruses and their vertebrate hosts, are some among other examples of long-term host-virus coevolution (McGeoch and Davison, 1999; Woolhouse et al., 2002; Switzer et al., 2005; Voskarides et al., 2018). [See, however, a discussion on time scale discrepancies of coevolutionary rates (Sharp and Simmonds, 2011), and compare with section 7.3.3].

7.7 Extinction, survival, and emergence of viral pathogens. Back to the mutant clouds

The viral groups defined by phylogenetic methods may or may not occupy a defined geographical location. It will depend on whether viral vectors or infected individuals carry the virus over long distances or not. A defined phylogenetic group may include viruses that produce similar or different pathology. This is because the capacity of a virus to cause disease may depend on modest genetic change (i.e., one or a few amino acid substitutions) that does not alter its position in a phylogenetic tree. It is important to emphasize that, independently of the time frame considered, the tips of phylogenetic trees are a cloud of mutants, that genomes within the cloud are the origin of future diversification pathways, and that individual

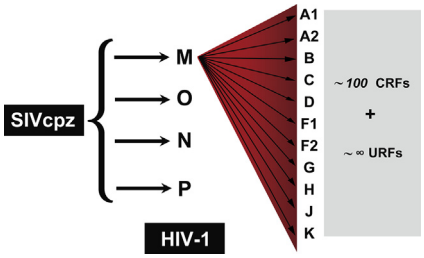

FIGURE 7.7 Diversification of HIV-1 from the time of introduction into the human population of retroviral simian ancestors SIVcpz from chimpanzees. Group M diversified into at least nine subtypes plus about 100 circulating recombinant forms (CRF), and multitudes of unique recombinant forms (URFs) that have not reached epidemiological relevance (box on the right). Genetic and antigenic diversifications are discussed in the text.

cloud components may differ in pathogenic potential. Fig. 7.7 summarizes the diversification of HIV-1, since it entered the human population. Once HIV-1 originated from multiple introductions of a chimpanzee simian immunodeficiency virus (SIVcpz), the four major HIV-1 groups M, O, N, and P were generated, and group M evolved into the multiple subtypes and recombinant forms that circulate at present. Many factors determine the pathogenic potential of any of the HIV-1 subtypes and the newly arising recombinant forms. HIV-1 is a notorious case of successful emergence of a new viral pathogen from a zoonotic reservoir of a related virus. HIV-1 populations are extremely complex and in continuous evolution (Hamelaar et al., 2019).

The relevance of the mutant cloud in determining viral fitness and survival was documented by comparing five isolates of West Nile virus (WNV) that had identical consensus sequences and differed in the mutant spectrum, as analyzed by dep sequencing (Kortenhoeven et al., 2015) (Fig. 7.8). The study concerned a WNV lineage 2 that circulated in Europe during the beginning of the 21st century. Environmental changes modified the haplotype composition while maintaining an invariant consensus sequence, an example of "perturbation" manifested only at the level of the mutant spectrum (see Section 6 in Chapter 6).

There is evidence that some viruses that once produced human disease might be now extinct. One example is provided by a putative viral agent of Economo's disease (also termed lethargic encephalitis or epidemic encephalitis), a degenerative disease of the brain that produced a loss of neurons. The disease had an acute phase of variable duration and intensity, followed by a chronic phase, sometimes with a late onset of symptoms. The disease showed a seasonal character with a maximum incidence in late winter. The first cases were recorded in Eastern Europe in 1915, and the disease was first described by Baron C. Von Economo in Vienna in 1917. During 1920–23 the disease attained pandemic proportions, although the number of cases and mortality were limited. It was estimated that between 1917 and 1929, about 100,000 cases occurred in Germany and Great Britain, and then mysteriously, the number of cases decreased, and the disease disappeared (Ford, 1937). At the time it was suspected that a virus similar to IV or some picornavirus might have been the etiological agent of this disease, but no proof could be provided. Occasional cases of lethargic encephalitis are diagnosed, and there is evidence of a possible connection of the once epidemic disease agent with postencephalitic Parkinsonism (Vilensky et al., 2010; Bigman and Bobrin, 2018).

There are additional examples of disease incidence decline that may be the prelude of virus extinction in some geographical areas. The alphavirus western equine encephalitis virus (WEEV) was an important human and animal pathogen in the Americas during the first part of the 20th century. The number of isolations of WEEV both from vertebrate and insect species declined during the second half of the 20th century, and two main mechanisms have been proposed: (i) fitness decrease due to operation of Muller's ratchet (concept described in Section 6.5 of Chapter 6), with elimination of

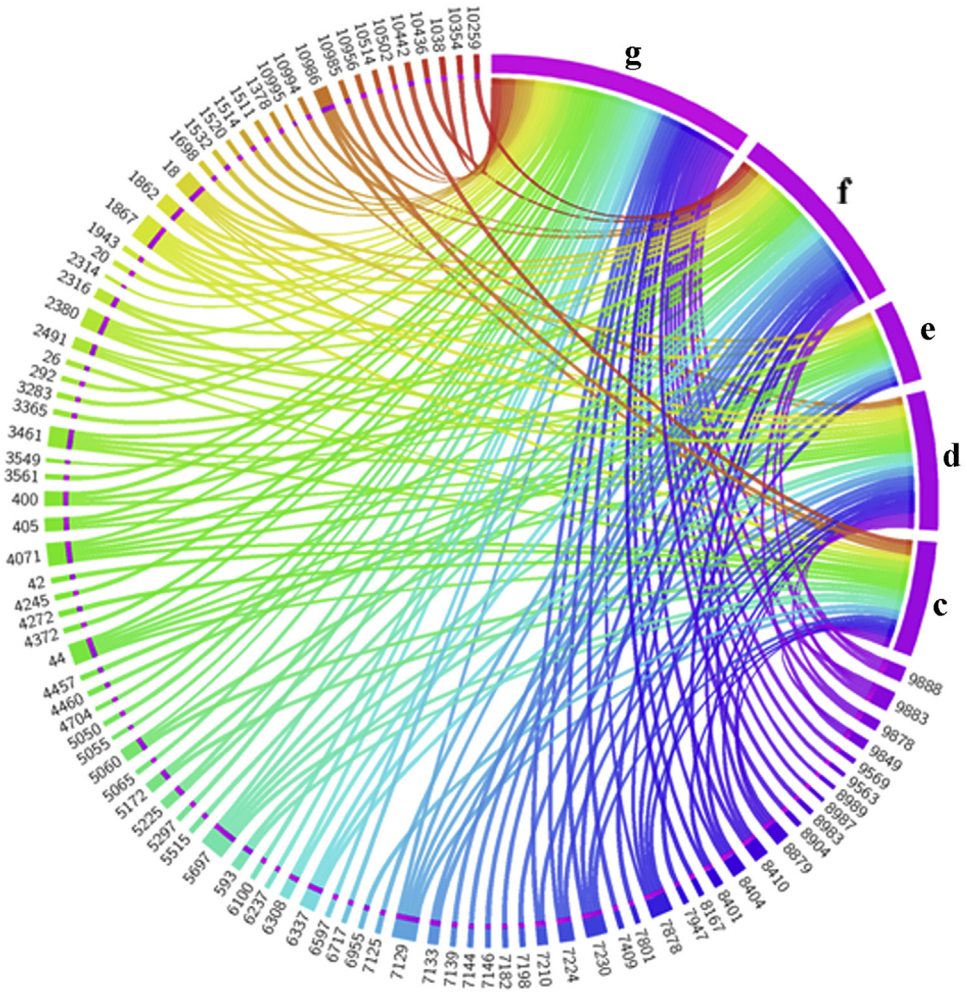

FIGURE 7.8 Visualization of the complexity of mutant spectra (haplotype composition) of five isolates of WNV denoted by c, d, e, f, and g (magenta bars) that have identical consensus sequence. The color lines connect the genomes where the same single nucleotide change occurs. Similarly, color-coded ribbons indicate that the same mutation occurs in two genomes in the same position. Nucleotide positions are numbered next to the outer rim of the circle. *Figure reproduced from Kortenhoeven et al. (2015). BMC Genomics is an open-access journal, and the article can be reproduced under the terms of the Creative Commons Attribution. The figure has been reproduced with the permission of the authors.*

virulence-associated mutations by genetic drift, or (ii) ecological factors such as decline of population numbers of susceptible host species, combined with increased vaccine coverage in horses (Bergren et al., 2014). It is not clear which is the major mechanism behind the decline, and whether it represents a firm, irreversible trend, or genetic and ecological factors may vary to produce a reemergence of this virus as a pathogen.

FMDV, the agent of the economically most important disease of cattle and other farm animals circulated until recently as seven different serotypes termed A, O, C, Asia 1, SAT1, SAT2,

and SAT3, and each serotype as multiple subtypes and antigenic variants (review in Sobrino and Domingo, 2004). Interestingly, in the 1980s, the incidence of serotype C FMDV decreased to the point that at the beginning of the 21st century this FMDV serotypes was considered nearly extinct and its eradication feasible. It cannot be totally excluded, however, that type C FMDV is replicating in some persistently infected ruminant in some remote part of our planet and that the virus reemerges again. If not, its ecological niche has been occupied by FMDVs of other serotypes. This is one important issue behind virus eradication (smallpox in the late 1970s or rinderpest in 2011): the possibility that the niche left by an eradicated pathogen is occupied by a related pathogen. A.E. Gorbalenya, E. Wimmer, and colleagues examined the possible evolutionary origin of present-day PV, and that other picornaviruses might occupy the PV niche in the event of its eradication (Jiang et al., 2007). Their phylogenetic analysis suggests that PV could originate from a C-cluster coxsackie A virus through amino acid substitutions in the capsid that led to a change of receptor specificity (other cases are discussed in Chapter 4). They generated chimeras of PV and its putative ancestors, and some of them were viable and pathogenic for transgenic mice expressing the PV receptor. The authors suggest that in a world without antiPV neutralizing antibodies, coxsackieviruses may mutate to generate a new PV-like agent.

Thus, despite virology being a very recent scientific discipline, there is ample evidence of the emergence of new viral pathogens, as well as some cases of extinction due to human interventions, and possible extinctions by natural influences. Viruses may evolve with regard to the symptoms they inflict upon their hosts. An increase of severity of Dengue virus infection has been observed in some world areas, consisting of neurological manifestations in patients with dengue fever or dengue hemorrhagic fever (Cam et al., 2001), among other examples of human and veterinary viral diseases. The dynamics of extinction of mutant viruses and their replacement by other forms is a continuous process, as the cycles of birth-death for any organism, but in a highly accelerated fashion.

We now turn to the pressing problem of the emergence and reemergence of viral disease.

7.7.1 Factors in viral emergence

New human viral pathogens emerge or reemerge at a rate of about one per year, representing an important concern for public health, alerting established and new organizations of the need of preparedness (Brechot et al., 2018). Emergence is defined as the appearance of a new pathogen for a host, while reemergence often refers to the reappearance of a viral pathogen, following a period of absence. Being a popular topic, the reader will find numerous books and reviews on the subject. It is worth emphasizing that in the 20th century many authors took the lead in emphasizing the problem of viral emergences and the need to investigate the underlying mechanisms, notably S.S. Morse and J. Lederberg [see several chapters of Morse (1993, 1994)]. Given the adaptive capacity of viruses, in particular, the RNA viruses, the reader will certainly suspect that genetic variation of viruses must be one of the factors involved in viral emergences. Indeed, most of the high-impact new viral diseases recorded recently or historically are due to RNA viruses. A statement by J. Lederberg reflects our vulnerability in the face of the nearly unlimited potential of viruses to vary: "Abundant sources of genetic variation exist for viruses to learn new tricks, not necessarily confined to what happens routinely or even frequently" (Lederberg, 1993). The situation is even more complex because a genetic variation of viruses is only one of many ingredients that promote the introduction of new viral pathogens in the human population. A report issued by the U.S. Institute of Medicine in 2003, analyzed and documented 13 factors that, individually or in combination,

participate in the emergence of microbial disease. They include a number of sociological, environmental, and ecological influences that act to promote the emergence and reemergence of viruses, bacteria, fungi, and protozoa (Smolinski et al., 2003) (Box 7.2).

Here, we will deal briefly with those factors of viral emergence related to the virus and host population numbers, in line with the focus of this book. Other aspects have been covered elsewhere (Antia et al., 2003; Haagmans et al., 2009; Wang and Crameri, 2014; Lipkin and Anthony, 2015; among others). The emergence of a viral disease can be regarded as a consequence of virus adaptation to a new environment, therefore, involving the concepts and mechanisms dissected in previous chapters. In particular, a relevant parameter is the variation of viral fitness in different environments (Domingo, 2010; Wargo and Kurath, 2011; Domingo et al., 2019).

Fitness can directly or indirectly impact any of the three steps involved in viral disease emergence or reemergence, which can be summarized as follows:

- Introduction of a virus into a new host species.
- Establishment (replication) of the virus in the new host.
- Dissemination of the virus among individuals of the new host species (transmissibility) to produce outbreaks, epidemics, or pandemics.

For the introduction and establishment steps, replicative fitness is critical while for the dissemination step, epidemiological fitness plays a major role (Chapter 5).

Two population numbers are relevant to the establishment step: the number of infectious viral particles shed by the infected donor host, and the number of potential new hosts that come into contact with the infected donor. We are now aware that even if two viral populations shed by an infected host have an identical number of infectious particles, not all mutant spectra might have the genomes subpopulations to permit the establishment in the human host (Fig. 7.9). There is a natural lottery regarding which quasispecies subpopulations will hit which host. In the words of J.J. Holland and his colleagues: "Although new RNA virus diseases of humans will continue to emerge at indeterminate intervals, the viruses themselves will not really be new, but rather mutated and rearranged to allow infection of new hosts, or to cause new disease patterns. It is important to remember that every quasispecies genome swarm in an infected individual is

BOX 7.2

FACTORS IN THE EMERGENCE OF MICROBIAL DISEASE

- Microbial change and adaptation.
- Human susceptibility to infection: impaired host immunity and malnutrition.
- Climate and weather.
- Changing ecosystems: vector ecology; reservoir abundance; and distribution.
- Human demographics and behavior: population growth; aging; and urbanization.
- Economic development and land use.
- International travel and commerce.
- Technology and industry.
- Breakdown of public health measures.
- Poverty and social inequality.
- War and famine.
- Lack of political will.
- Intent to harm: bioterrorism and agroterrorism.

Points summarized from Smolinski, M.S., Hamburg, M.A., Lederberg, J., 2003. Microbial Threats to Health, Emergence, Detection and Response. The National Academies Press, Washington, DC.

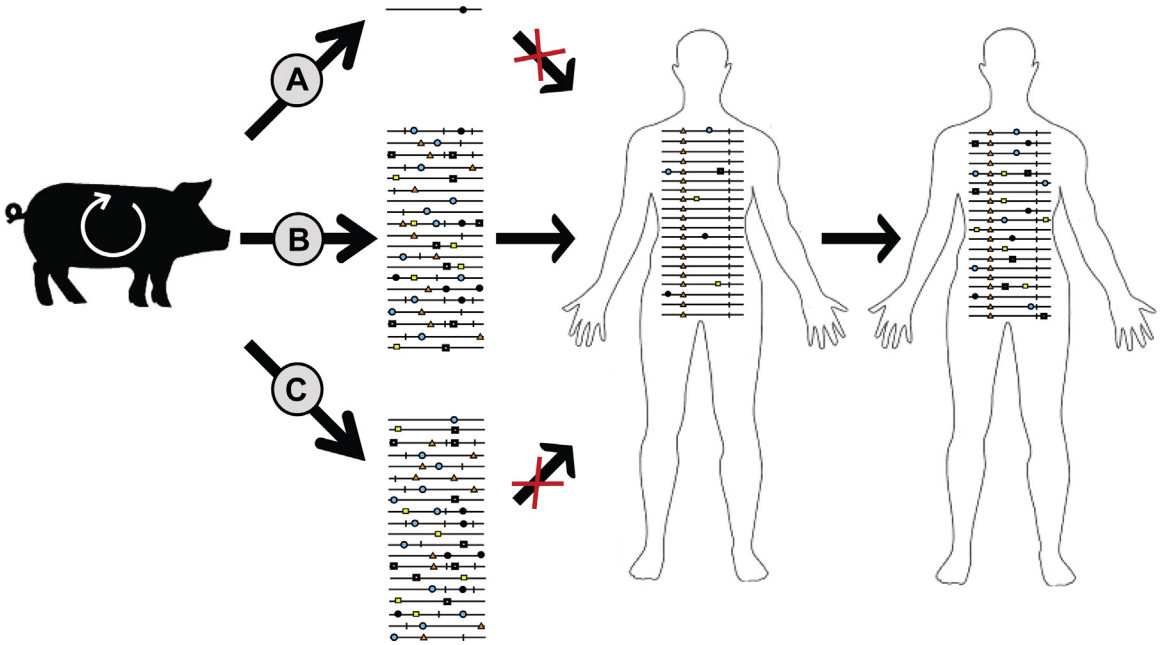

FIGURE 7.9 Relevance of virus population size and mutant spectrum composition in the zoonotic transmission of a virus. Only a subset of the genomes that surface an infected pig may be able to establish an infection in humans. The scheme indicates that a single genome that reached a human was not adequate to establish infection (pathway A). When multiple genomes reached the human (pathways B and C), only those that included a subset that displayed a minimum fitness in humans were able to initiate an infection and expand in the new host (pathway B). For pathways A and C, events are as if the contact between donor and recipient host had not taken place (arrows with cross). See text for implications.

unique and 'new' in the sense that no identical population of RNA genomes has ever existed before and none such will ever exist again" (Holland et al., 1992).

The number of infectious particles shed by an infected host determines the probability of transmission to susceptible hosts (Section 7.2), and of producing an emergence in a new host species. Viral population numbers and the number of transmissible particles can be largely amplified in immunocompromised individuals, and have been termed super-spreaders; they can contribute large amounts of variant viruses to the transmission lottery (Rocha et al., 1991; Paunio et al., 1998; Gavrilin et al., 2000; Khetsuriani et al., 2003; Small et al., 2006; Odoom et al., 2008; Woolhouse, 2017). Concerning the recipient hosts, the higher the number of potentially susceptible hosts that come into contact with an infected donor, the higher the probability of establishment of an emergent infection. It is likely that the advent of agricultural practices some 10,000 years ago, combined with increased contacts between humans and animals, inaugurated a time of new viral emergences. In the new scenario, viruses could shift from a persistent (low interhost transmission) mode into an acute (high interhost transmission) infection mode.

Not only population numbers are important, but the connections between the spatial habitats of a potential donor and recipient hosts are also highly relevant (Fig. 7.10). As correctly emphasized by S.S. Morse, changes in viral traffic may allow viruses to come near potential new hosts that had never been encountered before. Several sociological and ecological factors that

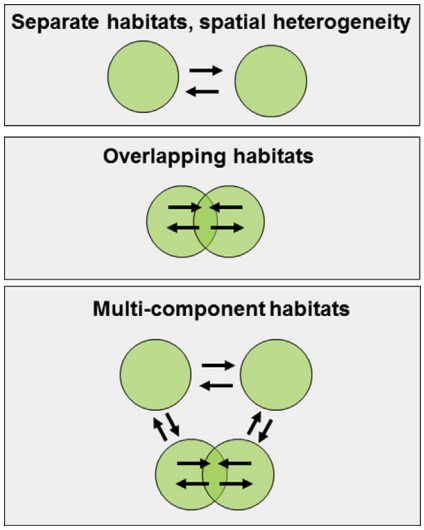

FIGURE 7.10 Types of habitats that may limit or facilitate interaction between hosts that can potentially establish an emergent infection in a local habitat. In separate habitats, contacts are restricted while in overlapping habitats, contacts are facilitated. In most cases, habitats cannot be reduced to the standard extremes, and are multicomponent habitats with various degrees of complexity. See text for implications for viral emergences.

can impact directly or indirectly the accessibility to an infected donor play a role. A typical example that connects several of the points listed in Box 7.2 is provided by the increase of arbovirus vectors during a humid season due to the climate change because insect larvae can proliferate on water reservoirs. Increased travel may put humans infected with arboviruses in contact with the flourishing insect vector population. Climate change may modify the migration routes of some birds, again putting these potential vertebrate hosts in contact with infected animals and insect vectors.

Other points listed in Box 7.2 are worth commenting: close human-to-human contacts are favored by urbanization. In 1975, there were five megacities in the world (meaning cities with more than 10 million human population) while in 2018 the number was 33, with a projection of 43 for 2030. Humans in close contact are, in addition, highly mobile. At present, it is possible to go around the world in about 36 h (if you choose the adequate airports) which represents a 1000-fold increase in spatial mobility of humans relative to the mobility in the year 1800. The 2014–15 Ebola epidemics in Africa was made worse by the breakdown of public health measures, poverty, and lack of political will of local and international agencies to put efforts in stopping transmission; Ebola outbreaks in Africa continue at the time of this writing. Underdeveloped countries are a reservoir of viral infections that represent a global threat due to several of the points listed in Box 7.2 [(Smolinski et al., 2003); several chapters of Singh (2014)].

Concerning the establishment and dissemination steps, the molecular mechanisms of quasispecies optimization in the new host environment apply. The underlying events are those presided by the extended Darwinian concepts of variation, competition, and selection, with the perturbations derived from stochastic effects (treated in different chapters of this book).

It is worth emphasizing parallelism between the steps involved in viral disease emergence and the steps that mediate entrance and establishment of any organism in a new habitat, which is a process well studied in ecology (van Driesche and van Driesche, 2000).

7.7.2 Complexity revisited

The meanings of complexity in virology were discussed in Section 3.9 of Chapter 3, one of them being the inability to explain a whole as the sum of its parts (Solé and Goodwin, 2000). R.V. Solé and B. Goodwin define the sciences of complexity as "the study of those systems in which there is no simple and predictable relationship between levels, between the properties of parts and of wholes." Several levels of complexity can be identified in the events that give rise to the emergence of a viral disease (Domingo, 2010; Sáiz

et al., 2014). One level of complexity concerns the behavior of viral populations which is often determined by interactions among components of mutant spectra in a way that cannot be predicted by the individual components of the population, even if we knew them!

The second level of complexity that can have an impact on the emergence of viral disease stems from the environmental, sociological, and ecological variables that must converge for a virus from some animal reservoir to come into contact and successfully infect a new host, for example, a human. Despite close surveillance, the emergences of viral diseases are unpredictable. Experts anticipate new influenza pandemics to arise somewhere in Asia from the avian reservoirs of IV; however, in 2009, the new influenza pandemic originated in Central America. Paradoxically, despite a general agreement that surveillance of human and zoonotic virus reservoirs should be intensified using new molecular tools (i.e., mutant spectrum analyses to go beyond the consensus sequences), the reality is that what we have learned are the reasons why viral emergences are unpredictable. The "abundant sources of genetic variation" that was emphasized by J. Lederberg (see Section 7.7.1) should be extended to refer to "abundant sources of complexity in viral emergences." For the time being, we have to be ready to react once the emergence has already occurred. Research funding agencies have a lot to do in favoring "reactive" rather than "preventive" antiviral policies.

7.8 Overview and concluding remarks

Viruses have survived because they have undergone multiple rounds of vertical and horizontal transmission in their host organisms and because occasionally they have found new suitable hosts where to replicate. Among the many parameters involved, in this chapter, we have emphasized the relevance of virus and host population numbers for sustained transmissions and the long-term maintenance of viral entities. A point that is often either ignored or not sufficiently emphasized is that the quasispecies nature of viral populations introduces an element of uncertainty regarding which types of mutants are transmitted to new hosts. Despite being a complication that cannot be easily handled, it is a fact that should stimulate new approaches to the surveillance of virus transmission and the identification of the founder viruses in new infections.

A steady accumulation of mutations during the evolution of organisms was a well-accepted proposal that agreed with the neutral theory of molecular evolution developed last century. One suspects that this agreement resulted in a premature preference for a regular clock of steady incorporation of mutations to work in the case of viruses. The evidence is that there are multiple molecular mechanisms that render the operation of a molecular clock for viruses very unlikely, and perhaps fortuitous in some cases. Several possible mechanisms of variable evolutionary rates have been discussed, and further clarification is expected from entire genome sequencing applied to viruses during outbreaks and epidemics. Viruses are probably not the best biological systems to obtain experimental evidence in support of the clock hypothesis.

A puzzling and pendent issue in viral evolution is the interpretation of the widely different number of viral serotypes, despite viruses sharing comparably large mutation rates and frequencies. Different possibilities have been examined, and a slight preference for variable constraints acting on the amino acid residues that determine the antigenic properties of viruses has been expressed. Again, additional work is necessary to solve this interesting problem.

Procedures for sequence alignments and the establishment of phylogenetic relationships among related viruses have been summarized, with some indications to find useful URL sites for sequence alignment and processing of large

data sets. The comparison of genomic sequences (and their encoded amino acids) of new viral isolates with those of the viruses characterized to date is important given the increasing number of new viruses discovered in all kinds of natural habitats.

The important problem of viral emergences and reemergences has been treated with emphasis on the concept of complexity. There are multiple interacting influences that converge to produce the emergence or reemergence of a viral pathogen, one of them being the heterogeneity of viral populations at the genetic and phenotypic level. Despite considerable methodological progress, we are still in the realm of uncertainty regarding the prediction of when and where a new viral pathogen will emerge (see Summary Box).

Summary Box

- Long-term evolution of viruses is the result of a history of virus transmission among hosts. Basic principles of transmission dynamics must take into consideration sampling effects and the inherent heterogeneity of viral populations.
- Rates of evolution of viruses in nature are extremely high as compared to the estimated rate for their host organisms. Contrary to some tenets of neutral evolution, rates of viral evolution are not constant with time. In particular, several mechanisms explain why intrahost virus evolution is faster than interhost evolution.
- Several procedures for sequence alignments and derivation of phylogenetic trees allow a partial description of virus diversification in nature.
- Antigenic diversification of viruses is subjected to constraints that differ among viruses. Some viruses have a single serotype while others have 100 serotypes. Several possible mechanisms may contribute to this difference.
- The emergence and reemergence of new viral pathogens is a multifactorial event with a clear influence of host and virus population numbers. Several levels of complexity participate in the emergence of a new pathogen, rendering the event highly unpredictable.

References

Aiewsakun, P., Katzourakis, A., 2016. Time-dependent rate phenomenon in viruses. J. Virol. 90, 7184–7195.

Ali, A., Melcher, U., 2019. Modeling of mutational events in the evolution of viruses. Viruses 11, E418.

Allen, B., Sample, C., Dementieva, Y., Medeiros, R.C., Paoletti, C., et al., 2015. The molecular clock of neutral evolution can be accelerated or slowed by asymmetric spatial structure. PLoS Comput. Biol. 11, e1004108.

Althaus, C.L., 2014. Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. PLoS Currents Outbreaks. https://doi.org/10.1371/currents.outbreaks.91afb5e0f279e7129e7056095255b288.

Anderson, R.M., May, R.M., 1991. Infectious Diseases of Humans. Oxford University Press, Oxford.

Antia, R., Regoes, R.R., Koella, J.C., Bergstrom, C.T., 2003. The role of evolution in the emergence of infectious diseases. Nature 426, 658–661.

Aswad, A., Katzourakis, A., 2012. Paleovirology and virally derived immunity. Trends Ecol. Evol. 27, 627–636.

Bamford, D.H., Grimes, J.M., Stuart, D.I., 2005. What does structure tell us about virus evolution? Curr. Opin. Struct. Biol. 15, 655–663.

Beale, G., Dellicour, S., Suchard, M.A., Lemey, P., Vrancken, B., 2018. Recent advances in computational phylodynamics. Curr. Opin. Virol. 31, 24–32.

Bergren, N.A., Auguste, A.J., Forrester, N.L., Negi, S.S., Braun, W.A., et al., 2014. Western equine encephalitis virus: evolutionary analysis of a declining alphavirus based on complete genome sequences. J. Virol. 88, 9260–9267.

Bigman, D.Y., Bobrin, B.D., 2018. Von Economo's disease and postencephalitic parkinsonism responsive to carbidopa and levodopa. Neuropsychiatric Dis. Treat. 14, 927–931.

Boege, U., Kobasa, D., Onodera, S., Parks, G.D., Palmenberg, A.C., et al., 1991. Characterization of Mengo virus neutralization epitopes. Virology 181, 1–13.

Botstein, D., 1980. A theory of modular evolution for bacteriophages. Ann. N. Y. Acad. Sci. 354, 484–491.

Botstein, D., 1981. A modular theory of virus evolution. In: Fields, B.N., Jaenisch, R., Fox, C.F. (Eds.), Animal Virus Genetics. Academic Press, New York, pp. 363–384.

Brechot, C., Bryant, J., Endtz, H., Griffin, D.E., Lewin, S.R., et al., 2018. International meeting of the global virus network. Antivir. Res. 163, 140–148.

Brooksby, J.B., 1981. Surveillance and control of virus diseases: Europe, Middle East and Indian sub-continent. In: Gibss, E.P.J. (Ed.), Virus Diseases of Food Animals, International Perspectives, vol. I. Academic Press Inc., London, pp. 69–78.

Buonagurio, D.A., Nakada, S., Desselberger, U., Krystal, M., Palese, P., 1985. Noncumulative sequence changes in the hemagglutinin genes of influenza C virus isolates. Virology 146, 221–232.

Cam, B.V., Fonsmark, L., Hue, N.B., Phuong, N.T., Poulsen, A., et al., 2001. Prospective case-control study of encephalopathy in children with dengue hemorrhagic fever. Am. J. Trop. Med. Hyg. 65, 848–851.

Chakravarty, S., Hutson, A.M., Estes, M.K., Prasad, B.V., 2005. Evolutionary trace residues in noroviruses: importance in receptor binding, antigenicity, virion assembly, and strain diversity. J. Virol. 79, 554–568.

Cheynier, R., Kils-Hutten, L., Meyerhans, A., Wain-Hobson, S., 2001. Insertion/deletion frequencies match those of point mutations in the hypervariable regions of the simian immunodeficiency virus surface envelope gene. J. Gen. Virol. 82, 1613–1619.

Delamater, P.L., Street, E.J., Leslie, T.F., Yang, Y.T., Jacobsen, K.H., 2019. Complexity of the basic reproduction number (R_0). Emerg. Infect. Dis. 25, 1–4.

Domingo, E., 2006. Quasispecies: concepts and implications for virology. Curr. Top. Microbiol. Immunol. 299. Springer, Berlin.

Domingo, E., 2007. Virus evolution. In: Knipe, D.M., Howley, P.M. (Eds.), Fields Virology, fifth ed. Lippincott Williams & Wilkins, Philadelphia, pp. 389–421.

Domingo, E., 2010. Mechanisms of viral emergence. Vet. Res. 41, 38.

Domingo, E., Mateu, M.G., Martínez, M.A., Dopazo, J., Moya, A., et al., 1990. Genetic variability and antigenic diversity of foot-and-mouth disease virus. In: Kurkstak, E., Marusyk, R.G., Murphy, S.A., Van-Regenmortel, M.H.V. (Eds.), Applied Virology Research. Plenum Publishing Co., New York, pp. 233–266.

Domingo, E., Biebricher, C., Eigen, M., Holland, J.J., 2001. Quasispecies and RNA Virus Evolution: Principles and Consequences. Landes Bioscience, Austin.

Domingo, E., de Avila, A.I., Gallego, I., Sheldon, J., Perales, C., 2019. Viral fitness: history and relevance for viral pathogenesis and antiviral interventions. Pathogens and Disease 77 ftz021.

Dopazo, J., Dress, A., von Haeseler, A., 1993. Split decomposition: a technique to analyze viral evolution. Proc. Natl. Acad. Sci. U.S.A. 90, 10320–10324.

Duchene, S., Holmes, E.C., ho, S.Y.W., 2014. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. Proc. Boil. Sci. 281, 20140732.

Duffy, S., Holmes, E.C., 2008. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus tomato yellow leaf curl virus. J. Virol. 82, 957–965.

Eggers, H.J., 2002. History of poliomyelitis and poliomyelitis research. In: Semler, B.L., Wimmer, E. (Eds.), Molecular Biology of Picornaviruses. ASM Press, Washington, DC, pp. 3–14.

Eigen, M., 1992. Steps towards Life. Oxford University Press, Oxford.

Emini, E.A., Jameson, B.A., Lewis, A.J., Larsen, G.R., Wimmer, E., 1982. Poliovirus neutralization epitopes: analysis and localization with neutralizing monoclonal antibodies. J. Virol. 43, 997–1005.

Farheen, N., Sen, N., Nair, S., Tan, K.P., Madhusudhan, M.S., 2017. Depth dependent amino acid substitution matrices and their use in predicting deleterious mutations. Prog. Biophys. Mol. Biol. 128, 14–23.

Feigelstock, D.A., Mateu, M.G., Valero, M.L., Andreu, D., Domingo, E., et al., 1996. Emerging foot-and-mouth disease virus variants with antigenically critical amino acid substitutions predicted by model studies using reference viruses. Vaccine 14, 97–102.

Felsenstein, J., 2004. Inferring Phylogenies. Sinauer Associates, Sunderland, MA.

Firth, C., Kitchen, A., Shapiro, B., Suchard, M.A., Holmes, E.C., et al., 2010. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. Mol. Biol. Evol. 27, 2038–2051.

Feng, D.F., Doolittle, R.F., 1996. Progressive alignment of amino acid sequences and construction of phylogenetic trees from them. Methods Enzymol. 266, 368–382.

Ford, F.R., 1937. Diseases of the Nervous System in Infancy, Childhood and Adolescence. Charles C. Thomas, Springfield, IL.

Fry, E.E., Stuart, D., 2010. Virion structure. In: Eherenfeld, E., Domingo, E., Roos, R.P. (Eds.), The Picornaviruses. ASM Press, Washington, DC, pp. 59–71.

Gaucher, E.A., Gu, X., Miyamoto, M.M., Benner, S.A., 2002. Predicting functional divergence in protein evolution by site-specific rate shifts. Trends Biochem. Sci. 27, 315–321.

Gavrilin, G.V., Cherkasova, E.A., Lipskaya, G.Y., Kew, O.M., Agol, V.I., 2000. Evolution of circulating wild poliovirus and of vaccine-derived poliovirus in an immunodeficient patient: a unifying model. J. Virol. 74, 7381–7390.

Geoghegan, J.L., Holmes, E.C., 2018. Evolutionary virology at 40. Genetics 210, 1151–1162.

Gething, M.J., Bye, J., Skehel, J., Waterfield, M., 1980. Cloning and DNA sequence of double-stranded copies of haemagglutinin genes from H2 and H3 strains elucidates antigenic shift and drift in human influenza virus. Nature 287, 301–306.

Geysen, H.M., Meloen, R.H., Barteling, S.J., 1984. Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. Proc. Natl. Acad. Sci. U.S.A. 81, 3998–4002.

Gilbert, C., Feschotte, C., 2010. Genomic fossils calibrate the long-term evolution of hepadnaviruses. PLoS Biol. 8, e1000459.

Gojobori, T., Yokoyama, S., 1985. Rates of evolution of the retroviral oncogene of Moloney murine sarcoma virus and of its cellular homologues. Proc. Natl. Acad. Sci. U.S.A. 82, 4198–4201.

Gorman, O.T., Bean, W.J., Webster, R.G., 1992. Evolutionary processes in influenza viruses: divergence, rapid evolution, and stasis. Curr. Top. Microbiol. Immunol. 176, 75–97.

Haagmans, B.L., Andeweg, A.C., Osterhaus, A.D., 2009. The application of genomics to emerging zoonotic viral diseases. PLoS Pathog. 5, e1000557.

Hall, B.G., 2001. Phylogenetic Trees Made Easy: A How-To Manual for Molecular Biologists. Sinauer Association Inc., Sunderland, MA.

Hamelaar, J., Elangovan, R., Yun, J., Dickson-Tetteh, L., Fleminger, I., et al., 2019. Global and regional molecular epidemiology of HIV-1, 1990-2015: a systematic review, global survey, and trend analysis. Lancet Infect. Dis. 19, 143–155.

Hanada, K., Suzuki, Y., Gojobori, T., 2004. A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. Mol. Biol. Evol. 21, 1074–1080.

Hedskog, C., Parhy, B., Chang, S., Zeuzem, S., Moreno, C., et al., 2019. Identification of 19 novel hepatitis C virus subtypes–Further expanding HCV classification. Open Forum Infect. Dis. 6, ofz076.

Heffernan, J.M., Smith, R.J., Wahl, L.M., 2005. Perspectives on the basic reproductive ratio. J. R. S. Interface R. Soc. 2, 281–293.

Herbeck, J.T., Nickle, D.C., Learn, G.H., Gottlieb, G.S., Curlin, M.E., et al., 2006. Human immunodeficiency virus type 1 env evolves toward ancestral states upon transmission to a new host. J. Virol. 80, 1637–1644.

Holland, J.J., Domingo, E., de la Torre, J.C., Steinhauer, D.A., 1990. Mutation frequencies at defined single codon sites in vesicular stomatitis virus and poliovirus can be increased only slightly by chemical mutagenesis. J. Virol. 64, 3960–3962.

Holland, J.J., de La Torre, J.C., Steinhauer, D.A., 1992. RNA virus populations as quasispecies. Curr. Top. Microbiol. Immunol. 176, 1–20.

Holmes, E.C., 2008. Comparative studies of RNA virus evolution. In: Domingo, E., Parrish, C.R., Holland, J.J. (Eds.), Origin and Evolution of Viruses, second ed. Elsevier, Oxford, pp. 119–134.

Holmes, E.C., 2009. The evolution and emergence of RNA viruses. Oxford Series in Ecology and Evolution. Oxford University Press, New York.

Huelsenbeck, J.P., Dyer, K.A., 2004. Bayesian estimation of positively selected sites. J. Mol. Evol. 58, 661–672.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R., Bollback, J.P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294, 2310–2314.

Jenkins, G.M., Rambaut, A., Pybus, O.G., Holmes, E.C., 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J. Mol. Evol. 54, 156–165.

Jiang, P., Faase, J.A., Toyoda, H., Paul, A., Wimmer, E., et al., 2007. Evidence for emergence of diverse polioviruses from C-cluster coxsackie A viruses and implications for global poliovirus eradication. Proc. Natl. Acad. Sci. U.S.A. 104, 9457–9462.

Kay, A., Zoulim, F., 2007. Hepatitis B virus genetic variability and evolution. Virus Res. 127, 164–176.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., et al., 2012. Geneius basic: an integrated and extendable desktop software platform for the organization and analysis of sequencing data. Bioinformatics 28, 1647–1649.

Khetsuriani, N., Prevots, D.R., Quick, L., Elder, M.E., Pallansch, M., et al., 2003. Persistence of vaccine-derived polioviruses among immunodeficient persons with vaccine-associated paralytic poliomyelitis. J. Infect. Dis. 188, 1845–1852.

Knowles, N.J., Hovi, T., King, A.M.Q., Stanway, G., 2010. Overview of taxonomy. In: Ehrenfeld, E., Domingo, E., Roos, R.P. (Eds.), The Picornaviruses. ASM Press, Washington, DC, pp. 19–32.

Kortenhoeven, C., Joubert, F., Bastos, A., Abolnik, C., 2015. Virus genome dynamics under different propagation pressures: reconstruction of whole genome haplotypes of west Nile viruses from NGS data. BMC Genomics 16, 118.

Kulkarni, A., Bangham, C.R.M., 2018. HTLV-1: regulating the balance between proviral latency and reactivation. Front. Microbiol. 9, 449.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA 7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.

Kumar, S., Agarwal, S., Ranvijay, 2019. Fast and memory efficient approach for mapping NGS reads to a reference genome. J. Bioinform. Comput. Biol. 17, 1950008.

Lai, D., Verma, M., 2017. Large-scale sequence comparison. Methods Mol. Biol. 1525, 191–224.

Lea, S., Hernández, J., Blakemore, W., Brocchi, E., Curry, S., et al., 1994. The structure and antigenicity of a type C foot-and-mouth disease virus. Structure 2, 123–139.

Lederberg, J., 1993. Viruses and humankind: intracellular symbiosis and evolutionary competition. In: Morse, S.S. (Ed.), Emerging Viruses. Oxford University Press, Oxford, pp. 3–9.

Leslie, A.J., Pfafferott, K.J., Chetty, P., Draenert, R., Addo, M.M., et al., 2004. HIV evolution: CTL escape mutation and reversion after transmission. Nat. Med. 10, 282–289.

Lin, Y.Y., Liu, C., Chien, W.H., Wu, L.L., Tao, Y., et al., 2015. New insights into the evolutionary rate of hepatitis B virus at different biological scales. J. Virol. 89, 3512–3522.

Lipkin, W.I., Anthony, S.J., 2015. Virus hunting. Virology 479–480C, 194–199.

Louie, R.H.Y., Kaczorowski, K.J., Barton, J.P., Chakraborty, A.K., McKay, M.R., 2018. Fitness landscape of the human immunodeficiency virus envelope protein that is targeted by antibodies. Proc. Natl. Acad. Sci. U.S.A. 115, E564–E573.

Lythgoe, K.A., Fraser, C., 2012. New insights into the evolutionary rate of HIV-1 at the within-host and epidemiological levels. Proc. Biol. Sci. 279, 3367–3375.

Martin, D.P., Williamson, C., Posada, D., 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21, 260–262.

Martínez, M.A., Carrillo, C., Gonzalez-Candelas, F., Moya, A., Domingo, E., et al., 1991a. Fitness alteration of foot-and-mouth disease virus mutants: measurement of adaptability of viral quasispecies. J. Virol. 65, 3954–3957.

Martínez, M.A., Hernández, J., Piccone, M.E., Palma, E.L., Domingo, E., et al., 1991b. Two mechanisms of antigenic diversification of foot-and-mouth disease virus. Virology 184, 695–706.

Martínez, M.A., Dopazo, J., Hernandez, J., Mateu, M.G., Sobrino, F., et al., 1992. Evolution of the capsid protein genes of foot-and-mouth disease virus: antigenic variation without accumulation of amino acid substitutions over six decades. J. Virol. 66, 3557–3565.

Mateu, M.G., 1995. Antibody recognition of picornaviruses and escape from neutralization: a structural view. Virus Res. 38, 1–24.

Mateu, M.G., 2017. The foot-and-mouth disease virion: structure and function. In: Sobrino, F., Domingo, E. (Eds.), Foot-and-Mouth Disease Virus. Current Research and Emerging Trends. Caister Academic Press, Norfolk, UK, pp. 61–105.

Mateu, M.G., Da Silva, J.L., Rocha, E., De Brum, D.L., Alonso, A., et al., 1988. Extensive antigenic heterogeneity of foot-and-mouth disease virus of serotype C. Virology 167, 113–124.

Mateu, M.G., Martínez, M.A., Capucci, L., Andreu, D., Giralt, E., et al., 1990. A single amino acid substitution affects multiple overlapping epitopes in the major antigenic site of foot-and-mouth disease virus of serotype C. J. Gen. Virol. 71, 629–637.

McGeoch, D.J., Davison, A.J., 1999. The molecular evolutionary history of the herpesviruses. In: Domingo, E., Webster, R., Holland, J. (Eds.), Origin and Evolution of Viruses. Academic Press, San Diego, pp. 441–465.

McTavish, E.J., Drew, B.T., Redelings, B., Cranston, K.A., 2017. How and why to build a unified tree of life. Bioassays 39. https://doi.org/10.1002/bies.201700114.

Melamed, A., Witkover, A.D., Laydon, D.J., Brown, R., Ladell, K., et al., 2014. Clonality of HTLV-2 in natural infection. PLoS Pathog. 10, e1004006.

Mims, C.A., 1981. Vertical transmission of viruses. Microbiol. Rev. 45, 267–286.

Minor, P.D., Schild, G.C., Bootman, J., Evans, D.M., Ferguson, M., et al., 1983. Location and primary structure of a major antigenic site for poliovirus neutralization. Nature 301, 674–679.

Minor, P.D., Ferguson, M., Evans, D.M., Almond, J.W., Icenogle, J.P., 1986. Antigenic structure of polioviruses of serotypes 1, 2 and 3. J. Gen. Virol. 67, 1283–1291.

Minot, S., Bryson, A., Chehoud, C., Wu, G.D., Lewis, J.D., et al., 2013. Rapid evolution of the human gut virome. Proc. Natl. Acad. Sci. U.S.A. 110, 12450–12455.

Morse, S.S., 1993. Emerging Viruses. Oxford University Press, Oxford.

Morse, S.S., 1994. The Evolutionary Biology of Viruses. Raven Press, New York.

Mount, D.W., 2004. Bioinformatics. Sequence and Genome Analysis. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Muhlemann, B., Margaryan, A., Damgaard, P. de B., Allentoft, M.E., Vinner, L., et al., 2018. Ancient human parvovirus B19 in Eurasia reveals its long-term association with humans. Proc. Natl. Acad. Sci. U.S.A. 115, 7557–7562.

Nakamura, T., Yamada, K.D., Tomii, K., Katoh, K., 2018. Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics 34, 2490–2492.

Nash, A.A., Dalziel, R.G., Fitzgerald, J.R., 2015. Mims' Pathogenesis of Infectious Disease. Elsevier Science & Technology, Edinburgh.

Notredame, C., Higgins, D.G., Heringa, J., 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302, 205–217.

Nowak, M.A., 2006. Evolutionary Dynamics. The Belknap Press of Harvard University Press, Cambridge, Massachusetts and London, England.

Nowak, M.A., May, R.M., 2000. Virus Dynamics. Mathematical Principles of Immunology and Virology. Oxford University Press Inc., New York.

Odoom, J.K., Yunus, Z., Dunn, G., Minor, P.D., Martin, J., 2008. Changes in population dynamics during long-term evolution of sabin type 1 poliovirus in an immunodeficient patient. J. Virol. 82, 9179–9190.

Page, R.D.M., Holmes, E.C., 1998. Molecular Evolution. A Phylogenetic Approach. Blackwell Science Ltd., Oxford.

Parrish, C.R., Kawaoka, Y., 2005. The origins of new pandemic viruses: the acquisition of new host ranges by canine parvovirus and influenza A viruses. Annu. Rev. Microbiol. 59, 553–586.

Paunio, M., Peltola, H., Valle, M., Davidkin, I., Virtanen, M., et al., 1998. Explosive school-based measles outbreak: intense exposure may have resulted in high risk, even among revaccinees. Am. J. Epidemiol. 148, 1103–1110.

Peng, Y., Tang, S., Wang, D., Zhong, H., Jia, H., et al., 2018. Meta PGN: a pipeline for construction and graphical visualization of annotated pangenome networks. GigaScience 7. https://doi.org/10.1093/gigascience/giy121.

Plyusnin, A., Morzunov, S.P., 2001. Virus evolution and genetic diversity of hantaviruses and their rodent hosts. Curr. Top. Microbiol. Immunol. 256, 47–75.

Porto, M., Roman, H.E., Vendruscolo, M., Bastolla, U., 2005. Prediction of site-specific amino acid distributions and limits of divergent evolutionary changes in protein sequences. Mol. Biol. Evol. 22, 630–638.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Ravantti, J., Bamford, D., Stuart, D.I., 2013. Automatic comparison and classification of protein structures. J. Struct. Biol. 183, 47–56.

Redd, A.D., Collinson-Streng, A.N., Chatziandreou, N., Mullis, C.E., Laeyendecker, O., et al., 2012. Previously transmitted HIV-1 strains are preferentially selected during subsequent sexual transmissions. J. Infect. Dis. 206, 1433–1442.

Rocha, E., Cox, N.J., Black, R.A., Harmon, M.W., Harrison, C.J., et al., 1991. Antigenic and genetic variation in influenza A (H1N1) virus isolates recovered from a persistently infected immunodeficient child. J. Virol. 65, 2340–2350.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Rossmann, M.G., 1989. The canyon hypothesis. Hiding the host cell receptor attachment site on a viral surface from immune surveillance. J. Biol. Chem. 264, 14587–14590.

Rowlands, D.J., Clarke, B.E., Carroll, A.R., Brown, F., Nicholson, B.H., et al., 1983. Chemical basis of antigenic variation in foot-and-mouth disease virus. Nature 306, 694–697.

Russell, D.J., 2014. GramAlign: fast alignment driven by grammar-based phylogeny. Methods Mol. Biol. 1079, 171–189.

Sáiz, J.C., Sobrino, F., Sevilla, N., Martin, V., Perales, C., et al., 2014. Molecular and evolutionary mechanisms of viral emergence. In: Singh, S.K. (Ed.), Viral Infections and Global Change. John Wiley & Sons Inc., Hoboken, NJ, pp. 297–325.

Sakaoka, H., Kurita, K., Iida, Y., Takada, S., Umene, K., et al., 1994. Quantitative analysis of genomic polymorphism of herpes simplex virus type 1 strains from six countries: studies of molecular evolution and molecular epidemiology of the virus. J. Gen. Virol. 75 (Pt 3), 513–527.

Salemi, M., Vandamme, A.M., 2004. The phylogeny handbook. A Practical Approach to DNA and Protein Phylogeny. Cambridge University Press, Cambridge.

Salemi, M., Vandamme, A.M., Gradozzi, C., Van Laethem, K., Cattaneo, E., et al., 1998. Evolutionary rate and genetic heterogeneity of human T-cell lymphotropic virus type II (HTLV-II) using isolates from European injecting drug users. J. Mol. Evol. 46, 602–611.

Salemi, M., Lamers, S.L., Yu, S., de Oliveira, T., Fitch, W.M., et al., 2005. Phylodynamic analysis of human immunodeficiency virus type 1 in distinct brain compartments provides a model for the neuropathogenesis of AIDS. J. Virol. 79, 11343–11352.

Schrag, S.J., Rota, P.A., Bellini, W.J., 1999. Spontaneous mutation rate of measles virus: direct estimation based on mutations conferring monoclonal antibody resistance. J. Virol. 73, 51–54.

Sellers, R.F., 1971. Quantitative aspects of the spread of foot-and-mouth disease. Vet. Bull. 41, 431–439.

Sellers, R.F., 1981. Factors affecting the geographical distribution and spread of virus diseases of food animals. In: Gibbs, E.P.J. (Ed.), Virus Diseases of Food Animals, International Perspectives, vol. I. Academis Press Inc, London, pp. 19–29.

Sharp, P.M., Simmonds, P., 2011. Evaluating the evidence of virus/host co-evolution. Curr Opin Virol 1, 436–441.

Sherry, B., Mosser, A.G., Colonno, R.J., Rueckert, R.R., 1986. Use of monoclonal antibodies to identify four neutralization immunogens on a common cold picornavirus, human rhinovirus 14. J. Virol. 57, 246–257.

Simmonds, P., Holmes, E.C., Cha, T.A., Chan, S.W., McOmish, F., et al., 1993. Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. J. Gen. Virol. 74 (Pt 11), 2391–2399.

Simmonds, P., Aiewsakun, P., Katzourakis, A., 2019. Prisoners of war-host adaptation and its constraints on virus evolution. Nat. Rev. Microbiol. 17, 321–330.

Singh, S.K., 2014. Viral Infections and Global Change. John Wiley & Sons Inc., Hoboken, NJ.

Small, M., Tse, C.K., Walker, D.M., 2006. Super-spreaders and the rate of transmission of the SARS virus. Physica D 215, 146–158.

Smith, D.B., Inglis, S.C., 1987. The mutation rate and variability of eukaryotic viruses: an analytical review. J. Gen. Virol. 68, 2729–2740.

Smith, D.B., Bukh, J., Kuiken, C., Muerhoff, A.S., Rice, C.M., et al., 2014. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. Hepatology 59, 318–327.

Smolinski, M.S., Hamburg, M.A., Lederberg, J., 2003. Microbial Threats to Health, Emergence, Detection and Response. The National Academies Press, Washington, DC.

Sobrino, F., Domingo, E., 2004. Foot-and-Mouth Disease: Current Perspectives. Horizon Bioscience, Wymondham, England.

Sobrino, F., Palma, E.L., Beck, E., Dávila, M., de la Torre, J.C., et al., 1986. Fixation of mutations in the viral genome during an outbreak of foot-and-mouth disease: heterogeneity and rate variations. Gene 50, 149–159.

Solé, R., Goodwin, B., 2000. Signs of Life. How Complexity Pervades Biology. Basic Books, New York.

Sorensen, J.H., Mackay, D.K., Jensen, C.O., Donaldson, A.I., 2000. An integrated model to predict the atmospheric spread of foot-and-mouth disease virus. Epidemiol. Infect. 124, 577–590.

Stapleton, J.T., Lemon, S.M., 1987. Neutralization escape mutants define a dominant immunogenic neutralization site on hepatitis A virus. J. Virol. 61, 491–498.

Stec, D.S., Waddell, A., Schmaljohn, C.S., Cole, G.A., Schmaljohn, A.L., 1986. Antibody-selected variation and reversion in Sindbis virus neutralization epitopes. J. Virol. 57, 715–720.

Sullivan, J., 2005. Maximum-likelihood methods for phylogeny estimation. Methods Enzymol. 395, 757–779.

Switzer, W.M., Salemi, M., Shanmugam, V., Gao, F., Cong, M.E., et al., 2005. Ancient co-speciation of simian foamy viruses and primates. Nature 434, 376–380.

Tidesley, M.J., Probert, W.J.M., Woolhouse, M.E.J., 2017. Mathematical models of the epidemiology and control of foot-and-mouth disease. In: Sobrino, F., Domingo, E. (Eds.), Foot-and-mouth Disease Virus. Current Research and Emerging Trends. Caister Academic Press, Norfolk, UK, pp. 385–408.

Tomonaga, K., Suzuki, N., Berkhout, B., 2019. Integration of viral sequences into eukaryotic host genomes: legacy of ancient infections. Virus Res. 262, 1 [Preface of a special issue of Virus Research on integration of viral sequences in host genomes].

Tonkin-Hill, G., Lees, J.A., Bentley, S.D., Frost, S.D.W., Corander, J., 2019. Fast hierarchical Bayesian analysis of population structure. Nucleic Acids Res. 47, 5539–5549.

van Driesche, J., van Driesche, R., 2000. Nature Out of Place: Biological Invasions in the Global Age. Island Press, Washington, DC.

Vilensky, J.A., Gilman, S., McCall, S., 2010. Does the historical literature on encephalitis lethargica support a simple (direct) relationship with postencephalitic Parkinsonism? Mov. Disord. 25, 1124–1130.

Villarreal, L.P., 2005. Viruses and the Evolution of Life. ASM Press, Washington, DC.

Villaverde, A., Martínez, M.A., Sobrino, F., Dopazo, J., Moya, A., et al., 1991. Fixation of mutations at the VP1 gene of foot-and-mouth disease virus. Can quasispecies define a transient molecular clock? Gene 103, 147–153.

Voskarides, K., Christaki, E., Nikolopoulos, G.K., 2018. Influenza virus-host coevolution. A predator-prey relationship? Front. Immunol. 9, 2017. https://doi.org/10.3389/fimmu.2018.02017.

Wang, L.F., Crameri, G., 2014. Emerging zoonotic viral diseases. Rev. Sci. Tech. (Int. Off. Epiz.) 33, 569–581.

Wang, H.Y., Chien, M.H., Huang, H.P., Chang, H.C., Wu, C.C., et al., 2010. Distinct hepatitis B virus dynamics in the immunotolerant and early immunoclearance phases. J. Virol. 84, 3454–3463.

Wargo, A.R., Kurath, G., 2011. In vivo fitness associated with high virulence in a vertebrate virus is a complex trait regulated by host entry, replication, and shedding. J. Virol. 85, 3959–3967.

Webster, R.G., 1999. Antigenic variation in influenza viruses. In: Domingo, E., Webster, R.G., Holland, J.J. (Eds.), Origin and Evolution of Viruses. Academic Press, San Diego, pp. 377–390.

Weddell, G.N., Yansura, D.G., Dowbenko, D.J., Hoatlin, M.E., Grubman, M.J., et al., 1985. Sequence variation in the gene for the immunogenic capsid protein VP1 of foot-and-mouth disease virus type A. Proc. Natl. Acad. Sci. U.S.A. 82, 2618–2622.

Wiktor, T.J., Koprowski, H., 1980. Antigenic variants of rabies virus. J. Exp. Med. 152, 99–112.

Woolhouse, M.E., 2017. Quantifying transmission. Microbiol. Spectr. 5 https://doi.org/10.1128/microbiolspec.MTBP-0005-2016.

Woolhouse, M.E., Webster, J.P., Domingo, E., Charlesworth, B., Levin, B.R., 2002. Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nat. Genet. 32, 569–577.

Worobey, M., 2001. A novel approach to detecting and measuring recombination: new insights into evolution in viruses, bacteria, and mitochondria. Mol. Biol. Evol. 18, 1425–1434.

Xia, X., Xie, Z., 2001. DAMBE: software package for data analysis in molecular biology and evolution. J. Hered. 92, 371–373.

Xing, L., Tjarnlund, K., Lindqvist, B., Kaplan, G.G., Feigelstock, D., et al., 2000. Distinct cellular receptor interactions in poliovirus and rhinoviruses. EMBO J. 19, 1207–1216.

Yang, Z., 2006. Computational Molecular Evolution. Oxford University Press, Oxford.

Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.M., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155, 431–449.

Zimmern, D., 1988. Evolution of RNA viruses. In: Domingo, E., Holland, J.J., Ahlquist, P. (Eds.), RNA Genetics. CRC Press Inc., FL, pp. 211–240.