ELSEVIER

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

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

practical requirement in the development of any EIAV vaccine is compatibility with established regulatory policies and diagnostic assays. The ability of EIAV-infected horses to routinely establish immunologic control over virus replication and disease suggests that an effective vaccine can indeed be developed, if the critical natural immune correlates of protection can be elicited by a candidate vaccine. An attenuated live EIAV vaccine with a reported protection efficacy of about 70% has been used in China since the early 1980s, but the effectiveness of this vaccine remains to be confirmed outside of that country. Evaluation of other candidate EIAV vaccines (live-attenuated, inactivated whole virus, subunit vaccines, synthetic peptides, etc.) under experimental conditions has revealed a spectrum of vaccine efficacy that ranges from 'sterile protection' (prevention of infection upon inoculation with EIAV) to severe elevation of EIAV replication, and exacerbation of disease. These results indicate that immune responses to EIAV are a double-edged sword that can either mediate protection or yield vaccine enhancement. Vaccine enhancement has previously been reported for other viral infections (dengue virus, respiratory syncitial virus, feline infectious peritonitis virus) and is of special concern with macrophage-tropic viruses. Similar examples of vaccine protection and enhancement have been reported in studies of experimental vaccines for other lentiviruses, including feline immunodeficiency virus, caprine arthritis-encephalitis virus, and visna-maedi virus. These observations in several diverse animal lentivirus systems suggest that the potential for immune enhancement may be a general property of lentiviruses, including HIV-1. Current efforts in the production of a commercial EIAV vaccine are focused on the development of a vaccine that can achieve sufficient maturation of immune responses to provide protection from virus infection, but allow the serological differentiation between vaccinated and infected horses. In this regard, DNA vaccine strategies appear to be well suited to accomplish these criteria for a commercial EIAV vaccine.

Future Research

EIAV provides a dynamic system for examining the interaction between virus populations and host immune responses that are evolving in response to each other. In addition, EIAV offers a remarkable model for studying the delicate balance between immune responses to a persistent virus infection that result in disease and those that have beneficial results. A characterization of the nature of protective and enhancing immune responses can provide important information about the mechanisms of lentivirus disease and the type of immune responses to be elicited or avoided by a vaccine. The results of these studies in the EIAV system should be applicable to other lentiviruses, including HIV-1.

See also: Human Immunodeficiency Viruses: Antiretroviral agents; Human Immunodeficiency Viruses: Molecular Biology; Human Immunodeficiency Viruses: Origin; Human Immunodeficiency Viruses: Pathogenesis; Immune Response to viruses: Antibody-Mediated Immunity; Vaccine Strategies; Viral Pathogenesis.

Further Reading

Cook RF, Issel CJ, and Montelaro RC (1996) Equine infectious anemia virus. In: Studdert R (ed.) *Viral Diseases of Equines*, pp. 295–323. Amsterdam: Elsevier.

Cordes TA and Issel CJ (1996) Equine infectious anemia: A status report on its control. USDA Animal and Plant Health Inspection Service Publication No. APHIS 91–55–032.

Montelaro RC, Ball JM, and Rushlow KE (1992) Equine retroviruses. In: Levy J (ed.) *The Retroviridae,* vol. 2, pp. 257–360. New York: Plenum.

Montelaro RC and Bolognesi DP (1995) Vaccines against retroviruses. In: Levy J (ed.) *The Retroviridae,* vol. 2. New York: Plenum.

Sellon DC, Fuller FJ, and McGuire TC (1994) The immunopathogenesis of equine infectious anemia virus. *Virus Research* 32: 111.

Evolution of Viruses

L P Villarreal, University of California, Irvine, Irvine, CA, USA

© 2008 Elsevier Ltd. All rights reserved.

Glossary

Dendrogram A schematic line drawing, often tree-like, that represents evolutionary relationships between species.

Error catastrophe A threshold at which a high error rate of genome replication can no longer maintain the integrity of essential genetic information.

ERV An endogenous retroviral-like genetic element, containing a long terminal repeat that is found in the genomes of many organisms.

Fitness landscape A hypothetical representation of a three-dimensional surface that represents relative fitness of genetic variants.

Muller's ratchet A decrease in fitness resulting from a repetition of a genetically restricted founder population.

Quasispecies A population of viral genomes that are the product of error-prone replication usually envisioned as a swarm or cloud of related genomes.

Red Queen hypothesis An evolutionary concept in which an organism must evolve at high rates in order to maintain its competitive advantage.

Reticulated evolution A pattern of evolution in which elements are derived from different ancestors represented as a tree with cross-connections between distinct branches.

Sequence space A multidimensional representation of all possible sequences for a given genome.

Introduction

The initial study of virus evolution sought to explain how virus variation affects viral and host survival and to understand viral disease. However, we now realize that virus evolution is a basic issue, impacting all life in some way. In general, the principles of virus evolution are very much the same Darwinian principles of evolution for all life, involving genetic variation, natural selection, and survival of favorable types. However, virus evolution also entails features such as high error rates, quasispecies populations, and genetic exchange across vast reticulated gene pools that extend the traditional concepts of evolution. Evolution simply means a noncyclic change in the genetic characteristics of a virus; and viruses are the most rapidly evolving genetic agents for all biological entities. Principles of virus evolution provide an integrating framework for understanding the diversity of viruses and the relationships with host as well as providing explanations for the emergence of new viral disease. Most emerging viral diseases are due to species jumps from persistently infected hosts that have long-term virus–host evolutionary histories (called natural reservoirs). Since early human populations (i.e., small bands of hunter gatherers) could not have supported the great human viral plagues of civilization (e.g., smallpox virus/variola and measles), these viruses must also have originated from species jumps that adapted to humans. In recent history, the emergence of HIV demonstrates that virus evolution continues to impact human populations. Early observation established that viruses often show significant variation in virulence. Such variation was used as an early, yet risky, form of vaccination (i.e., variolation against smallpox). The variation that occurred with passage into alternate tissue and host was also used to make various vaccines (rabies in 1880s) or attenuate viral virulence, such as live yellow fever vaccine. But variation also allows some viruses, such as influenza A, to escape neutralization by vaccination. However, variation in viral disease or virulence does not provide a quantitative basis to study evolution.

The study of virus variation and evolution is an applied science that allows the observation of evolutionary change in real time. For example, human individuals (or populations) infected with either human immunodeficiency virus 1 (HIV-1) or hepatitis C virus (HCV) show progressive or geographical evolutionary adaptation associated with the emergence of specific viral clades that affect disease therapy and progression (such as resistance to antiviral drugs). Figure 1 shows HIV variation in an individual human patient whereas Figure 2 shows HCV variation in the human population. Virus evolution is also important for the commercial growth of various organisms, such as the dairy industry (lactose fermenting bacteria), the brewing industry, agriculture, aquaculture, and farming. In all these applied cases major losses can result from virus adaptation to the cultivated species, often from viruses of wild species. Some organisms appear much less prone to viral adaptations (e.g., nematodes, ferns, sharks). Virus evolution can also be applied to technological innovation, as in phage display. This is a process in which a terminal surface protein of some filamentous bacterial virus can be genetically engineered for novel surface protein expression. By generating diversity *in vitro* (with up to 10^{15} types), and applying the principles of evolution (random variation) to biochemical selection (such as binding to a chemical substrate), a reiterative amplification can find solutions to problems in biochemistry, such as surface interactions or catalytic activity.

Virus Evolution as a Basic Science

Virus variation is a global issue. In the last decade it has been established that viruses are the most numerous biological entities on the planet. The oceans and soil harbor vast numbers of viral-like particles (VLPs), mostly resembling the tailed DNA viruses of bacteria. In addition, some of these environmental viruses are unexpectedly large and complex, such as the phycodnaviruses of algae or mimivirus of amoeba, a 1.2 Mb DNA virus that can encode nearly 1000 genes. Thus viruses represent a vast and diverse source of novel genes. However, the evolutionary dynamics of this population and its effect on hosts is not well understood. It is likely that this virus gene pool also affects host evolution since prophage colonization is

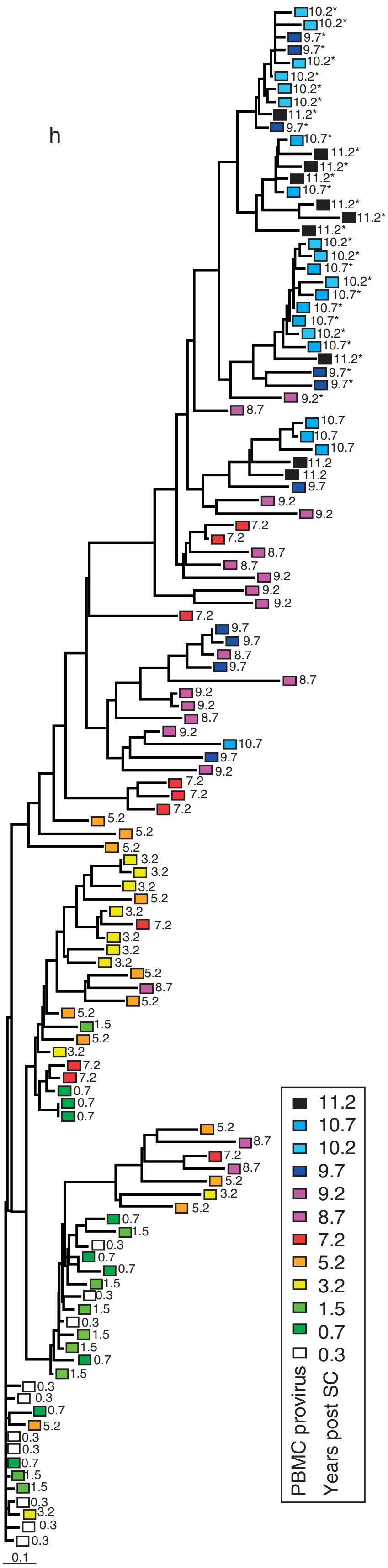

Figure 1 HIV population analysis from an infected individual. Shown is a neighbor-joining phylogram derived from maximum likelihood distance between all sequences. Sequences are represented by a square for PBMC sequences or a triangle for plasma sequences. The arbitrary color gradient corresponds to the time of sampling. Adapted from Shankarappa R, Margolick JB, Gange SJ, *et al.* (1999) Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus Type 1 infection. *Journal of Virology* 73: 10489–10502, with permission from the American Society for Microbiology.

known for all prokaryotic genomes. Thus, this vast pool of viruses connects directly to prokaryotes and the ‘tree of life’. The study of virus evolution has become an extension of all evolution.

Distinctions from Host Evolution

For the most part, virus evolution conforms to the same Darwinian principles as host evolution, involving variation and natural selection. However, viruses have multiple origins, and are thus polyphyletic. There are six major categories of viruses (+RNA, –RNA, dsRNA, retro, small DNA, large DNA) that have no common genes and hence have no common ancestor. However, these categories all have conserved hallmark genes (i.e., capsid proteins, Rd RNA pol, RT, primase, helicase, and RCR initiation proteins) which are all monophyletic and which may trace their origins to primordial host/viral gene pools. Many DNA viruses, for example, have replication strategies and polymerases that are clearly distinct from that of their host, which depends on such hallmark genes. Thus, virus evolution appears ancient but inextricably linked to its host. Also, in contrast to host evolution, viral quasispecies show a population-based adaptability that extends the selection of the fittest to include populations of otherwise unfit genomes (described below). Viruses can clearly cross the usual host-species barriers so that viral evolution can be reticulated in vast genes pools. For example, bacterial DNA viruses and +RNA viruses can show high rates of recombination across viruses that infect numerous host species. Accordingly, tailed DNA phage of bacteria appear to represent one single vast gene pool. Viruses can also violate concepts of death and extinction, reassembling genomes from parts and/or repairing lethal damage by multiplicity reactivation. In addition, damaged (defective) viruses can also affect virus and host evolution. Such defectives are found in many types of viruses and can also be found in many host genomes (as defective prophage or defective endogenous retroviruses). Such defective viruses can clearly affect host survival. In all these characteristics, viruses extend the Darwinian principles of host evolution.

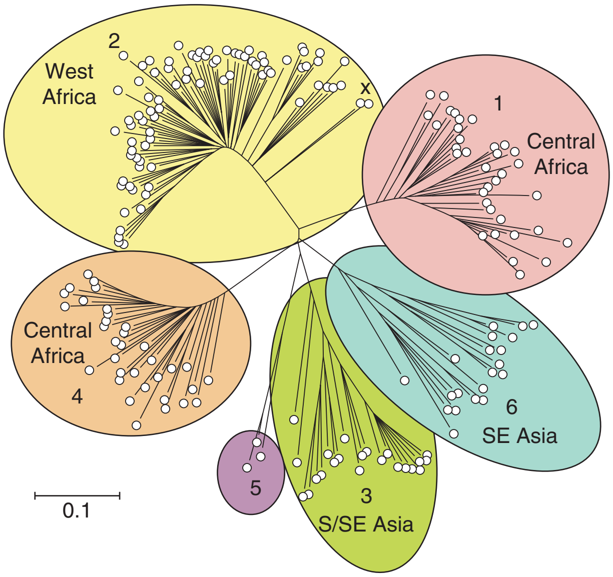

Figure 2 Unrooted phylogenetic analysis of HCV nucleotide sequences from globally distributed human isolates. Adapted from Simmonds P (2001) Reconstructing the origins of human hepatitis viruses. *Philosophical Transactions of the Royal Society of London* 356: 1013–1026, with permission from Royal Society Publishing.

A History of Virus Evolution

The coherent study of virus evolution awaited the development of sequence technology to measure mutations and genetic variation in viral populations. Concepts of natural selection, fitness, and propagation of favorable variation had long been established in the evolutionary biology literature prior to the growth of virology. Thus, mathematical models, such as Fisher population genetics, concerned with gene frequency in a (sexually exchanging) species population, had been well developed and seemed directly applicable to virus evolution, which resemble that of host genes. Here, viral fitness was typically expressed as relative replication rates (replicative fitness) but sometimes host virulence and disease were also used. However, as presented below, a comprehensive definition of viral fitness remains problematic. The first quantitative measurements of the rate of virus mutation was done in the 1940s with bacterial phage. The mutation rates were expressed as a set of ordinary differential equations that were subsequently used to develop the quasispecies equations as applied to error-prone RNA genomes (see below). However, the species definition of a virus poses a problem for evolutionary thinking and challenges how we define kinship in viruses. Unlike the sex-defined host, a virus species is currently defined as a polythetic class: a mosaic of related parts of which not all elements are shared (such as host range, genome relatedness, antigenic properties). No specific defining characteristic or gene exists for a virus species and sexual exchange need not be included. This is an inherently fuzzy definition, like defining a 'heap', which although clear, cannot be specified by its number of parts. The ensuing molecular characterization of many virus populations supports this species definition. The challenge then is to understand viral evolutionary patterns working with such fuzzy definitions. Yet, conserved patterns of virus evolution are still seen, some of which suggest viruses are indeed an ancient lineage, possibly extending into the primordial RNA world.

Error-Prone Replication and Quasispecies

In the 1970s, Manfred Eigen and also Peter Shuster developed a fundamental theoretical model of virus evolution. A set of ordinary differential equations was published that described what was called 'quasispecies'. Starting from measurements of phage mutation rates, they considered the consequences of high error rates as expected from RNA replication (an error-prone noncorrecting replication process). The resulting population shared many properties and was called quasispecies, a society (or community) of individuals that are the error products of replication. The name 'quasispecies' thus describes a chemically diverse set of molecules and was not intended to refer to a biological species (i.e., genetic exchanging). However, as discussed above, the fuzzy definition of virus species and quasispecies overlaps somewhat, which has been a source of confusion. Several premises were used to develop this theory: (1) the individual products ignore one another and interact only as individuals; (2) the system is not at equilibrium and resources are not limiting. Based on relative replication, the growth of favorable types is described which provides a mathematical definition of replicative fitness. The original equations represent an idealized generalized system of infinite population size and are not directly applicable to the real world, although they provide valuable insights into real world systems. The equations do not address variable mortality (longevity), interference, exclusion, competition, complementation, and persistence, or how such issues affect nonreplicative fitness definitions. The issue of mortality and fitness is interesting from the perspective of viruses. For example, an interfering defective virus can be considered dead, but can clearly interfere with and drive the extinction wild-type template replication in quasispecies. In some cases, the quasispecies equations appear to be mathematically equivalent to classical Wright and Fisher population genetics equations as applied by Kimura and Maroyama to asexual haploid populations at the mutation-selection balance. However, these two approaches begin with distinct perspectives, and it was the assumption of high error in the quasispecies equations that had a major impact on experimentation and our current understanding of virus evolution. This has also led to some counter intuitive conclusions, such as the concept of selection of 'the fittest' compared to the consensus character of the master template. Quasispecies from a virus with high error rates (such as HIV-1) might be composed of all mutant progeny RNAs such that the consensus template (the mean, the fittest, or the master template) may not actually exist. With classical population genetics, an asexual clonal population should fix the clonal sequence. With quasispecies, this is not observed. The first laboratory measurements of viral quasispecies were made using Qβ RNA polymerase *in vitro*. Error estimates ranged from 10^{-3} to 10^{-4} substitutions per site per year (an error rate applicable to most RNA viruses). With Qβ, the replication of many nonviable mutants generated a genetic spectra that had a characteristic makeup. For example, separate DNA clones of Qβ were initially distinct from each other but quickly generated the same RNA quasispecies as before cloning. Additional lab measurements have shown that quasispecies can have significant adaptive fitness (above the cloned master template) and display memory; that is, they can retain information of prior selections in a minority of the population. Complementation, interference, competition suppression, and extinction have all been measured in various quasispecies, thus indicating a collective form of evolution and violating an original premise of genome independence in these equations. In addition, the sequence diversity in a quasispecies is now seen as a source of adaptive potential, not simply error (see below). Despite these results, the concept of quasispecies has still been highly useful, and not simply as a theoretical development. For example, the live poliovirus vaccine is clearly a heterogeneous quasispecies. Within this population exists a minority of neurovirulent variants that are suppressed by the majority avirulent virus. A main point of the quasispecies concept is that it provides an understanding of high adaptability from a population that has many, even lethal, mutations. It is interesting that the proposed early RNA world would have also been a collective quasispecies world.

Error Catastrophe, Sequence Space

Quasispecies theory also predicts a situation known as error catastrophe, defined as an error rate threshold at which information is lost and the system decays. If error rates are too high, or the information content (genome length) too extensive, the system will be unable to maintain its information integrity. This predicts a basic limit on the size of RNA genomes, consistent with the observation that the largest RNA genomes are only about 27–32 kbp (coronaviruses). There is a possible therapeutic use of error catastrophe: drugs (possibly Ribavarin and 5-fluorouracil) that increase the error rates of RNA polymerase can potentially push a virus beyond its error threshold and induce a catastrophe. Quasispecies is an inherently fuzzy and dynamic population that has no sharp boundaries or specific members and has been metaphorically referred to as swarms and clouds. Here, cloud is a metaphor for the population landscape that exists in high-dimensional hyperspace and cannot be readily visualized. The concept of sequence space has been used to represent topography of the distribution of all mutants. Kinship relationships between mutants can be measured by Hamming distance; the minimal steps needed to specify the difference between two mutants. In spite of high error and adaptation rates (and sometimes high recombination rates), RNA viruses are not able to explore all potential sequence space. Selection significantly limits the quasispecies, since the potential sequence space is hyperastronomical even for a moderately sized virus. For example, an RNA virus with 10 000 nt would correspond to 10^{6000} possible sequences, well beyond what could be explored by even the potentially vast number of viruses over the lifetime of the world. In addition, there are clearly mechanistic constraints that prevent many possible sequences, such as necessary domains of ± strand RNA folds, physical association with ribonucleotide proteins, virion packaging and assembly – all in addition to usual selection for gene fitness (function) that all severely limit possible adaptations. This creates a multipeak 'fitness' landscape in hyperspace (see Figure 3). Assuming fitness itself can have a single definition (i.e., replicative fitness, not subjected to variable and stochastic competition), we can visualize this space as many steep valleys and ledges (in this case with 10 000 dimensions). Normally we think that adaptation by natural selection is the force to explore and move through fitness landscape. But as the deep valleys are often lethal, they cannot be explored via natural selection. Here we see the major adaptive power of the quasispecies

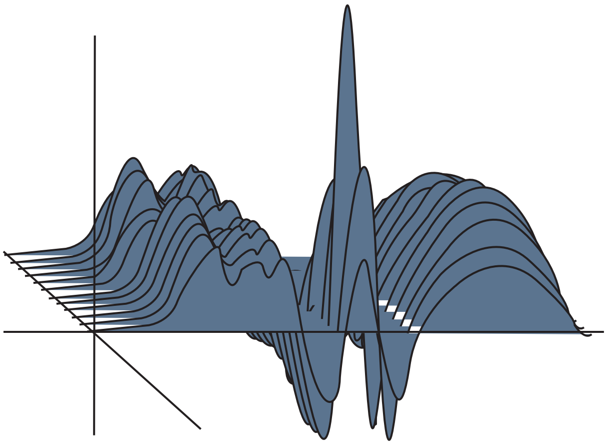

Figure 3 Hypothetical fitness landscape for an RNA virus. Assuming one definition for a nonrelativistic fitness (such as replicative fitness), the coordinates indicate relative fitness. Those below the *y*-axis are interfering or lethal variants.

collective. Since random, even lethal, errors and drift are inherent in a quasispecies, lethal valleys can be readily crossed by such variable genomes, allowing the master genome to adapt by natural selection to a new fitness peak. Thus, error-prone replication and the generation of mutant clouds allows for much better exploration of sequence space and eventual adaptability.

When viruses are transmitted to new hosts, they can experience a genetic bottleneck since a relatively small number of viral genomes could be involved (aka low multiplicity passage). If this process is serially repeated, a phenomenon known as 'Muller's Ratchet' can result in lost competitiveness as the essentially clonal RNA virus accumulates deleterious mutations (sometimes measured as pfu/plaque). However, in lab studies, virus extinction from serial passage does not occur, presumably due to plaque selection for a restored phenotype. Even a single plaque is in reality a small population (due to nonideal particle/pfu ratios and ID_{50}). However, lost competitiveness with other viruses is seen with clonal laboratory passage. However, if a quasispecies population is passed, this generally results in increased competitive fitness. Such passage can produce a seemingly never-ending better version of the virus that outcompetes all prior versions of the same virus (although virion yields and absolute replication are not necessarily improved). This has been likened to the Red Queen hypothesis in that the viruses are evolving at high rates, simply to maintain their competitive position, so as not to be displaced as the dominant viral type. Virus–virus competition is thus a crucial selection.

Never-Ending Adaptation

A real world example of the potentially never-ending virus adaptation is shown in Figure 4. The HA and NA genes of human influenza A virus have been monitored for several decades. As shown, the prevalent master template of the virus circulating in the human population has been continually changing, due to immune selection and stochastic viral immigration, necessitating yearly vaccine changes (also shown). Although such a population dynamic has been stably maintained in the human population, all prior versions have essentially become extinct, as they do not reappear.

Not all RNA virus populations show this dynamic of a continual change or even the diversity expected from quasispecies. Even in influenza virus A, avian isolates from natural host (waterfowl) can be genetically stable. Some RNA (and retro) viruses with high error rates can nevertheless maintain stable populations in specific hosts. For example, measles virus shows much less antigenic drift in human infections compared to influenza virus A. Hepatitis G virus (a human prevalent and distant relative of HCV) shows little variation in even isolated human populations. The filoviruses (Ebola virus and Marburg virus) have shown no genetic variation in Zaire isolates from 20 years apart. Hendra virus, isolated from Australian fruit bats, and Nipah virus from Malaysian fruit bats also show little genetic diversity. Arenaviruses and hantavirus are also genetically stable in their natural rodent host. The reasons for such population stability have not been well evaluated. In some cases (measles), purifying selection would seem likely. In other cases, persistence and low replication rates seem to apply. For example, simian foamy virus (SFV) and human T-lymphotropic virus II (HTLV-II) generate only about 10^{-8} substitutions per site per year, probably due to low replication rates.

Virus–Host Congruence and RNA Stability

There are now many examples of species-specific RNA virus/host coevolution, indicating very slow rates of virus evolution. Since error rates must be similar, this appears to be at odds with the quasispecies theory. For example, Hantavirus (genus *Bunyavirus*) coevolution with its rodent host, suggests a 20 million year association. Arenaviruses (ssRNA bisegmented ambisense) also coevolve in Old/New World rodents. These viruses are of special interest with regard to emergence as they represent the source of five hemorrhagic human fevers (such as Lassa virus). In all these examples, however, it appears that the virus causes a persistent unapparent infection in its natural host and that human disease is due to species jump.

Tools

Although viral genomes were the first to be sequenced, the initial focus was simply to identify similarity between viral genes, not to evaluate distant evolutionary relationships. The most popular tool for finding similarity is BLAST (Basic Local Alignment Search Tool) from the National Centers for Biotechnology Information (NCBI), which calculates similarity between query sequences and infers a probability based on a matrix database. Various versions of BLAST are the most used tool in bioinformatics to trace evolution. Although BLAST will identify similar genes, it is also necessary to compile and evaluate the similarities in sets of the related sequences. Multiple sequence alignment software, such as ClustalW, is used for this purpose. Phylogenetic relationships are then inferred from tree-building software. This software includes maximum parsimony, neighbor-joining, and maximum likelihood methods. The statistical significance of the tree (relative to all possible trees) can then be evaluated by algorithms such as bootstrap. More recently, Bayesian analyses, such as Bali-Phy, which implements a Markov Chain Monte Carlo (MCMC) method and calculates joint posterior

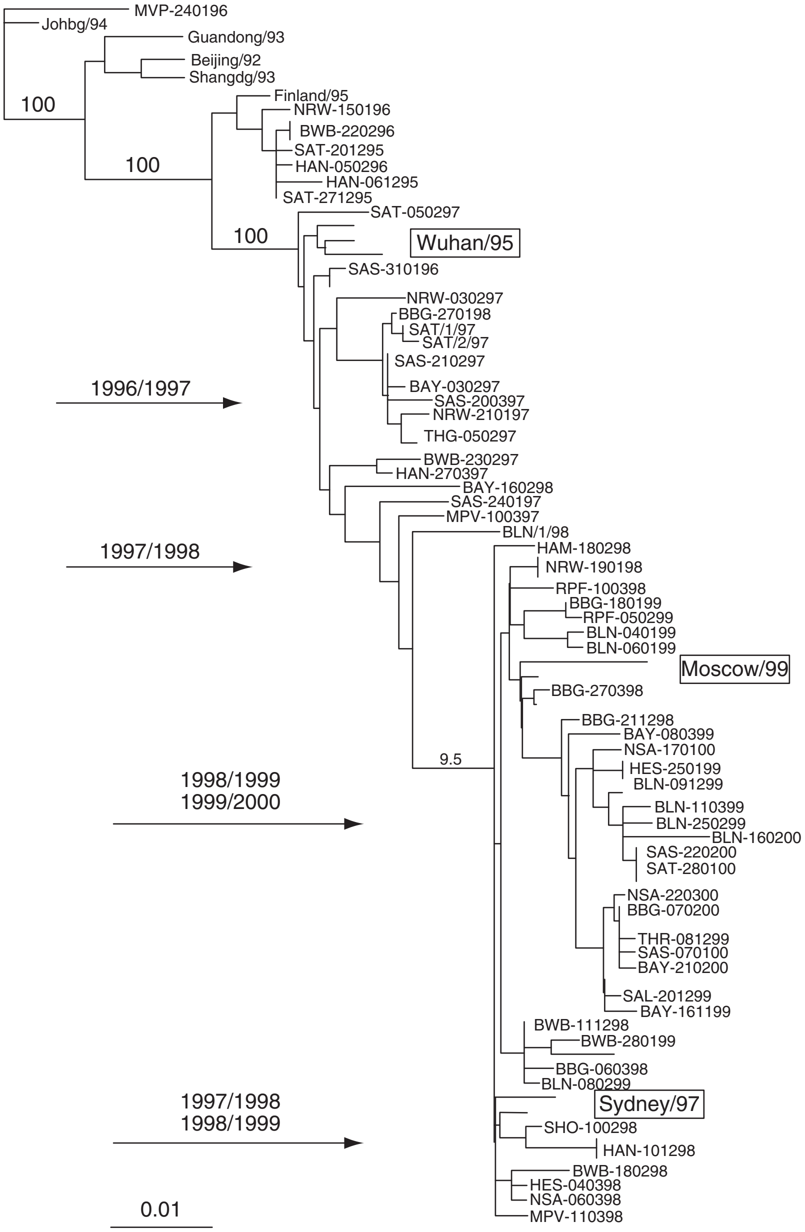

Figure 4 Phylogenetic tree of yearly influenza A/H3N2 viruses variants based on the hemaglutinin gene. Locations of specific vaccine strains are boxed. Reproduced from Schweiger B, Zadow I, and Heckler R (2002) Antigenic drift and variability of influenza viruses. *Medical Microbiology and Immunology* 191: 133–138, with permission from Springer-Verlag.

probabilities of phylogeny and alignment have become popular. This software has the added potential of using sliding windows and evaluating multiple trees, such as in virus–host coevolution. These methods, however, use the single master consensus sequence as its query and do not evaluate quasispecies, collective-based populations. Also, the clearly reticulated or hybrid character of some virus evolution is problematic. Also, distant evolutionary relationships are no longer preserved in the same sequences. Here, conservation of structural motifs, assembly patterns, gene order, and replication strategy are used to identify distant kinship.

Patterns of RNA Virus Evolution

The sometimes extreme variation of RNA virus sequence has led some to propose that most family lineages appear to be only about 10 000 years old, which is clearly at odds with much older estimates. The +RNA viruses in particular show a remarkable diversity of genomes and replicator mechanisms. These families also show much evidence of recombination and a tendency to cross host barriers. About 38 families of +RNA viruses with up to four segments are known. There are four distinct classes of replicase recognized in viruses that also share a common genetic plan. These RNA viruses have three helicase superfamilies, two protease superfamilies, and two jelly roll capsid domains. For the most part, capsid and RdRpol sequences are congruent except for members of the families *Luteoviridae* and *Tetraviridae*, which appear to have undergone recombination between these two gene lineages. The smallest +RNA virus is a member of the bacterial virus family *Leviviridae*, which has only four genes. This simple virus appears to represent the ancestral +RNA virus. Curiously, no RNA virus has yet been found to infect archaebacteria. The largest +RNA viruses belong to the genus *Coronavirus* (27–32 kbp). The most recently described +RNA viruses are members of the family *Marnaviridae* infecting bats and marine organisms sea life, which appear to be basal to evolution of picornaviruses. However, natural populations of some +RNA viruses can be stable. For example, dengue virus (*Flaviviridae*) shows low rates of amino acid substitution (e.g., nonsynonymous to synonymous ratios). Since it is an acute arbovirus infection with high error rates, strong selective constraints likely account for this stability involving multiple (systemic) tissues and vector transmission.

Negative-strand RNA viruses have distinct patterns of evolution, which is traced via their polymerase genes. Gene order tends to be highly conserved. The unsegmented viruses, such as rhabdoviruses, lyssaviruses, and paramyxoviruses, do not undergo significant recombination so their variation tends to be by point mutations and deletions. Although high error rates, variation, and quasispecies generation can be seen in laboratory settings, natural isolates, such as lyssaviruses and measles virus, tend to be relatively homogeneous. For example, lyssaviruses show a slow rate of evolution (5×10^{-5}/site/year). Lyssavirus persistent infections in natural host might contribute to this stability. However, measles virus is a strictly human-specific acute infection so its stability is likely due to purifying selection.

Patterns of DNA Virus Evolution: Tailed Phage

Large DNA viruses of bacteria, archaea, and eukaryotes appear to be evolutionarily linked. Although little sequence conservation can be identified between the T4 phage of bacteria, the halophage of archaea, the members of the family *Phycodnaviridae* infecting algae, and the herpes viruses of vertebrate eukaryotes, all show similarities in their gene programs, DNA polymerase types, capsid structures, and capsid assembly, consistent with a common ancestor. For example, both the Enterobacteria phage T4 (T4) and herpes simplex virus 1 (HSV-1) have $T = 1$ capsid symmetry with 60 copies of capsid protein. The bacterial DNA viruses would appear to represent the ancestor of all these viruses, but the origins of these phages now appear lost in the primordial gene pool. These DNA viruses can have large genomes that could not be sustained by error-prone replication. Thus, many DNA viruses do have error-correcting DNA replication, with error rates that approach or equal those of their host cells (10^{-8}). Giant bacterial phage genomes (*Bacillus megaterium* phage G, of about 600 genes), and algal phycodnaviruses have now been characterized. Even larger DNA viruses of amoeba (acanthamoeba polyphaga mimivirus) coding for more than 1000 genes are known to be abundant in some water habitats.

The tailed phage of bacteria have been called the dark matter of genetics, due to their numerical dominance ($\sim 10^{31}$ *en toto*). This corresponds to about 10^{24} productive infections per second on a global scale. Most host-restricted phage lineages clearly conserve sets of core proteins (especially capsid genes), but others (the broader T-even phages) do not conserve any hallmark genes. Hallmark genes, when present, are usually recognized by conserved domains within proteins, such as replication and structural proteins. Replicator strategy and gene order are also frequently conserved. Phage also tends to conserve genes that are active against other phage (i.e., DNA modification, lambda RexA, T4 rII). With the sequencing of numerous phage genomes, however, a large number of novel genes have been identified. Currently, 350 full genomes of tailed phage and 400 prophage from bacterial genomes have been sequenced. In general, large DNA viruses are tenfold overrepresented in small single-domain genes (~100 aa). Comparative genomics, especially of lactobacterial phages, suggest that most phage genomes evolve as mosaics, with sharp boundaries between genes as well as at protein domains within genes (see Figure 5). Recombination between lytic, temperate, and cryptic prophages appears to account for this gene and subgene domain variation. Some specific phages have mechanisms to generate specific gene diversity (such as bordetella phage using RT for surface receptor diversity), but most diversity is the product of recombination. Two broad patterns of phage variation have been observed corresponding to host-unassociated lytic and host-associated (congruent) temperate phage. In most bacterial genomes (ECOR *Escherichia coli* collection, cyanobacteria, *B. subtilis*), patterns of prophage colonization account for significant genetic distinctions between closely related host strains. The general picture for tailed phage of bacteria is that they are not the products of reduction of host genomes.

Large Eukaryotic DNA Viruses

As noted, evolutionary links between tailed phage and large DNA viruses of eukaryotes are apparent. The phycodnaviruses of unicellular green algae clearly have many phage-like characteristics, including the presence of

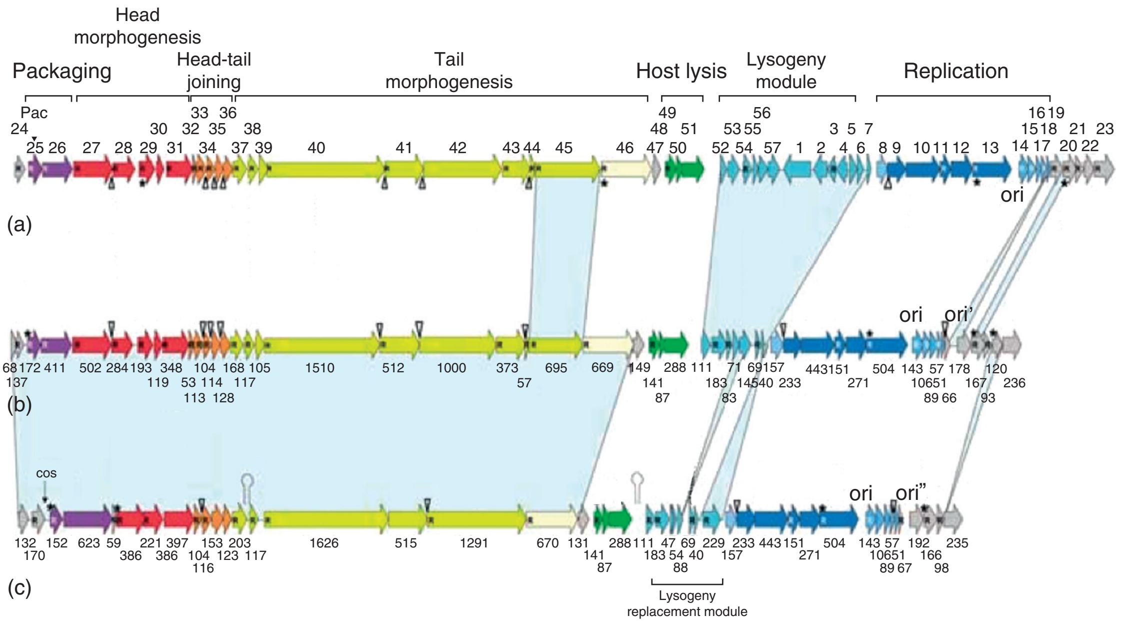

Figure 5 Genome comparison of temperate *S. thermophilus* phage 01205, virulent *S. thermophilus* Sfi11, and the virulent *S. thermophilus* Sfi19. Probable gene functions are indicated and genomes have been divided into functional units. Genes belonging to the same module are indicated with the same color. Areas of shading indicate regions of major difference. From Desiere F, Lucchini S, Canchaya C, Ventura M, and Brüssow H (2002) Comparative genomics of phages and prophages in lactic acid bacteria. *Antonie van Leeuwenhoek* 82: 73–91.

restriction/modification enzymes, homing endonucleases, the injection of viral DNA, and the external localization of the viral capsid. They also have many characteristics of eukaryotic DNA viruses, such as a clearly herpes virus-related DNA polymerase, PCNA proteins, nonintegrating DNA, and numerous signal-transduction proteins. Thus, phycodnaviruses show hybrid characteristics of prokaryotic and eukaryotic viruses.

The evolutionary pattern of the large DNA viruses of eukaryotes is generally best traced by comparing their respective DNA-dependent DNA polymerases (DdDp). These exist in distinct classes that are typically specific for each viral lineage and are usually the most highly conserved of the set of core genes within a viral lineage. However, some viruses, such as the white spot syndrome virus (WSSV) infecting shrimps, have almost no genes in common with other DNA viruses. Generally, the specific set of core genes is clade specific. The first fully sequenced viral genome tree was that of the baculoviruses (see Figure 6). The overall pattern of evolution shows the conservation of the core set in which most clades can be differentiated from one another mainly by acquisition of several novel viral genes (although some lineage-specific gene loss is also apparent). In another example, coccolithoviruses differ from related phycodnaviruses by the acquisition of 100 kbp gene set, including six subunits of DdDp core genes. Similar patterns of divergence can be seen with the herpesvirus family members. In addition, most herpesvirus clades also show coevolution with their host. However, the poxviruses (orthopoxviruses), show a different overall evolutionary pattern and are not congruent with host. The more ancestral orthopoxvirus members, such as cowpox virus and mousepox virus, have greater gene numbers that appear to have been lost in the human-specific and virulent smallpox virus. Avipoxviruses have even greater gene diversity but the entomopoxviruses are the most complex and diverse of all. The complexity and brick shape of the poxviruses originally inspired the view that these viruses might evolve from bacterial cells following the reduction of complexity. However, DNA sequencing makes it clear that viral core genes have no bacterial analogs. In some instances, viral lineages have clearly fused with other viral and host lineages. For example, the baculovirus Autographa californica MNP virus (AcMNPV) has acquired a gypsy-like retrovirus (e.g., TED), an endogenous retrovirus associated with moth development. The polydnaviruses (circular DNA viruses) are fused into their host genomes (as endogenous DNA viruses) of some parasitoid wasps, essential for survival of the wasp larvae.

Small DNA Viruses

The small, double-stranded, circular DNA viruses (*Papillomaviridae* and *Polyomaviridae*) show evolutionary patterns that are highly host linked. Virus and host evolution are mostly congruent, and virus evolution tends to be slow.

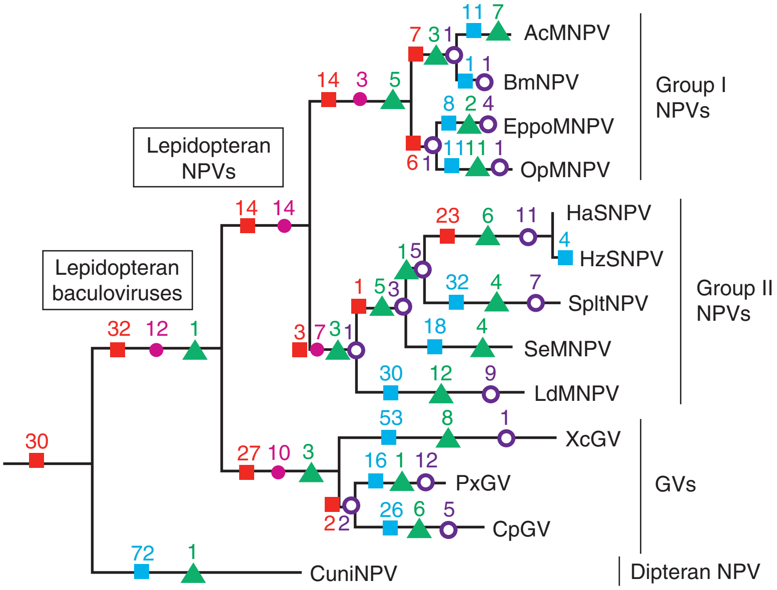

Figure 6 Gene content map of 13 complete sequences of baculoviruses, including the genus *Granulovirus*. The tree shows the most parsimonious hypothesis of changes in gene content during baculovirus evolution. Colors and shapes indicate gene conservation, acquisition, and loss. Reproduced from Herniou EA, Olszewski JA, Cory JS, O'Reilly DR (2003) The genome sequence and evolution of baculoviruses. *Annual Review of Entomology* 48: 211–234, with permission from Annual Reviews.

For example, approximately 100 human papillomaviruses show congruent evolution with human (and primate) host. This seems to be due to both a highly species- and tissue-specific virus replication, as well as a tendency to establish persistent infections. However, the rolling circular replicon (RCR) viruses, such as parvoviruses, can have distinct evolution patterns. Both mouse minute virus (MMV) and canine parvoviruses can have quasispecies-like populations, which can show evolutionary rates at 10^{-4} substitutions per site per year. Such rates are at the lower end of those seen with RNA viruses. In bacteria, RCR viruses and RCR plasmids appear to represent a common gene pool. Other poorly characterized small eukaryotic DNA viruses, such as human torque teno virus, are asymptomatic but show high variation during persistence for unknown reasons.

Endogenous and Autonomous Retroviruses

Retroviruses present a special problem in understanding patterns of eukaryotic virus evolution. Like prophage of bacteria, retroviruses both stably colonize their host as endogenous or genomic retroviruses (ERVs) that are often defective, but may also sometimes emerge from their host (especially rodents) to produce autonomous virus. In addition, retroviruses are polyphyletic and prone to generating quasispecies due to high error rates as well as high rates of recombination. The most common conserved retrovirus genome elements are domains within the long terminal repeats (LTRs), RT, integrase, protease, gag protein, and env protein. Of these, env are the most often altered or deleted in host genomes. In addition, tRNA primer sites (such a lys [K] tRNA) are also often conserved and used for classification (i.e., human endogenous retrovirus, HERV-K family). However, each of these retroviral elements can potentially have distinct patterns of evolution and conservation, generating distinct dendrograms. Vertebrates, especially mammals, seem to host many retroviral elements within their genomes. Their autonomous retroviruses have a tendency to infect cells of the immune system. Murine leukemia virus (MLV) is the best-studied simple autonomous retrovirus, but many endogenous MLV relatives also exist. Retroviruses are present in genomes of early eukaryotes but significantly expanded in vertebrates. Gypsy-like retroviruses (aka chromoviruses, defined via RT and gag similarity) are often found conserved as full-length elements including env genes in most lower eukaryote genomes (e.g., *Caenorhabditis elegans*), but were mostly lost from tetrapods. Some lower eukaryotes clearly prevent colonization by ERVs, such as *Neurospora* fungi (via the RIP exclusion system). Many endogenous retroviruses are congruent with host evolution, whereas other ERVs are recently acquired and highly host specific. In terms of gene diversity, the retroviral *env* are the most diverse. There are five RT-based families recognized such as *Retroviridae, Hepadnaviridae, Caulimovoridae, Pseudoviridae,* and *Metaviridae*, the latter three being especially prevalent as genomic elements in flowering plants (especially Gypsy). Yet not all retroviruses seem able to colonize host germ line. For example, lentiviruses (such as simian immunodeficiency virus (SIV) and HIV) show no examples of endogenization compared to the simpler MLV-related viruses that can be both autonomous (i.e., MLV, Gibbon ape leukemia virus) and endogenous (i.e., *Mus dunni* ERV, koala ERV). The converse can also be true, since no autonomous versions of HERV K, for example, are known.

Early views proposed that retroviruses evolved from nonviral retroposons (LTR RT elements, non-LTR LINE-like elements). These non-LTR elements, have distinct nonretroviral mechanistic features and core protein domains, but retain some virus-like domains of RT; thus, they appeared to predate retroviruses. However, we now know that gypsy-like retroviruses were present in the earliest eukaryotes. In addition, some LTR-containing elements, such as Gypsy, had initially been considered ancestral to retroviruses because all copies seemed to be defective. However, it is now established that complete gypsy retroviruses are conserved as ERVs in some yeast and *Drosophila* strains. Thus, although endogenous and exogenous retroviruses appear to evolve from each other, there is no evidence that exogenous retroviruses have emerged from non-LTR LINE-like elements.

The congruence between ERVs and host eukaryote evolution is sometimes striking. For example, all mammals have acquired their own peculiar versions of ERVs (and LINES). Recently, it has become clear that the placental mammals have conserved several families of ERV-derived *env* genes that provide an essential function for placental tissue (ERV W-syncytin 1, ERV FRD syncytin 2, enJSRV). Clearly, retrovirus evolution is highly intertwined with that of their hosts.

Never-Ending Emergence

A remaining concern of virus evolution is to understand the emergence of new viral pathogens. The unpredictable and stochastic nature of such virulent adaptations makes predictions difficult, as the link between virulence and evolution is vague. For example, the genetic changes that made the SARS virus (persisting in bats) into an acute human pathogen are still not predictable. Viral fitness and selection, and how they change from persistent states with acute species jumps, are not yet defined. However, some variables contribute to the likelihood of viral emergence, such as virus ecology. The population density and dynamics of the new host and the ecological interactions-between new and stable viral host are often crucial. In addition, virus–virus interactions can be important, allowing for recombination and/or reassortments or lowering immunological selective barriers via immunosuppression. The emergence of HIV-1 from different, persistent SIVs of African monkeys through chimpanzees into a new human disease, for example, includes the same issues. Also, the potential emergence of pandemic human influenza from avian (Anatiformes) sources, such as H5N1, remains a great concern. Thus, virus evolution will continue to interest us as we seek to predict, control, or eradicate viral agents of disease.

See also: Antigenic Variation; Coronaviruses: Molecular Biology; Emerging and Reemerging Virus Diseases of Plants; Emerging and Reemerging Virus Diseases of Vertebrates; Emerging Geminiviruses; Origin of Viruses; Phylogeny of Viruses; Picornaviruses: Molecular Biology; Quasispecies; Retrotransposons of Vertebrates; Virus Databases; Virus Evolution: Bacterial Viruses; Virus Species.

Further Reading

Desiere F, Lucchini S, Canchaya C, Ventura M, and Brussow H (2002) Comparative genomics of phages and prophages in lactic acid bacteria. *Antonie Van Leeuwenhoek* 82(1–4): 73–91.

Domingo E (2006) *Current Topics in Microbiology and Immunology, Vol. 299: Quasispecies: Concept and Implications for Virology.* Berlin: Springer.

Herniou EA, Olszewski JA, Cory JS, and O'Reilly DR (2003) The genome sequence and evolution of baculoviruses. *Annual Review of Entomology* 48: 211–234.

Hurst CJ (2000) *Viral Ecology.* San Diego: Academic Press.

Schweiger B, Zadow I, and Heckler R (2002) Antigenic drift and variability of influenza viruses. *Medical Microbiology and Immunology* 191: 133–138.

Shankarappa R, Margolick JB, Gange SJ, *et al.* (1999) Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *Journal of Virology* 73: 10489–10502.

Simmonds P (2001) Reconstructing the origins of human hepatitis viruses. *Philosophical Transactions of the Royal Society of London* 356: 1013–1026.

Roossinck MJ (2005) Symbiosis versus competition in plant virus evolution. *Nature Reviews Microbiology* 3(12): 917–924.

Villarreal LP (2005) *Viruses and the Evolution of Life.* Washington, DC: ASM Press.

Zanotto PM, Gibbs MJ, Gould EA, and Holmes EC (1996) A reevaluation of the higher taxonomy of viruses based on RNA polymerases. *Journal of Virology* 70(9): 6083–6096.